



**UNICAMP**

**UNIVERSIDADE ESTADUAL DE CAMPINAS**

**INSTITUTO DE GEOCIÊNCIAS**

**PÓS-GRADUAÇÃO EM GEOCIÊNCIAS - ÁREA DE  
GEOENGENHARIA DE RESERVATÓRIOS**

**MARCELO LOPES DE OLIVEIRA**

**ANÁLISE DAS INCERTEZAS ENVOLVIDAS NA MODELAGEM  
DE RESERVATÓRIOS NO CONTEXTO GEOESTATÍSTICO**

Dissertação apresentada ao Instituto de Geociências como  
parte dos requisitos para obtenção do título de Mestre em  
“Geociências - Área de Geoengenharia de Reservatórios”

**Orientador:** Professor Dr. Armando Zaupa Remacre

**OL4a**  
**33020/BC**

Este exemplar corresponde a  
redação final da tese defendida  
por Marcelo Lopes de Oliveira  
e aprovada pela Comissão Julgadora  
em 19/21/98

Armando Zaupa Remacre  
ORIENTADOR

**CAMPINAS - SÃO PAULO**

**DEZEMBRO - 1997**



**UNICAMP**

**UNIVERSIDADE ESTADUAL DE CAMPINAS**  
**INSTITUTO DE GEOCIÊNCIAS/ÁREA DE GEOLOGIA DE**  
**PETRÓLEO**

**PÓS-GRADUAÇÃO EM GEOCIÊNCIAS - ÁREA DE**  
**GEOENGENHARIA DE RESERVATÓRIOS**

**MARCELO LOPES DE OLIVEIRA**

**ANÁLISE DAS INCERTEZAS ENVOLVIDAS NA MODELAGEM**  
**DE RESERVATÓRIOS NO CONTEXTO GEOESTATÍSTICO**

Dissertação apresentada ao Instituto de Geociências como requisito  
parcial para obtenção do título de Mestre em Geociências - Área de  
Geoengenharia de Reservatórios.

**Orientador:** Dr. Armando Zaupa Remacre - UNICAMP/IG/AGP

**CAMPINAS - SÃO PAULO**

**DEZEMBRO - 1997**

NÚMERO...  
 \* CHAMADA: 7/UNICAMP  
 0292  
 Es.  
 IMBO B. 38020  
 ROC. 395/98  
 C  D  X  
 RECO R\$ 18,00  
 DATA 10/03/98  
 \* CPD

FICHA CATALOGRÁFICA ELABORADA  
 PELA BIBLIOTECA I.G. UNICAMP

CM-00106395-0

OL4a Oliveira, Marcelo Lopes de  
 Análise das incertezas envolvidas na modelagem de  
 reservatórios no contexto geoestatístico / Marcelo Lopes de  
 Oliveira.- Campinas,SP.: [s.n.], 1997.

Orientador: Armando Zaupa Remacre  
 Dissertação (mestrado) Universidade Estadual de Campinas,  
 Instituto de Geociências

1. Geoestatística. 2. Incerteza. 3. Reservatórios. 4. Petróleo.  
 I. Remacre, Armando Zaupa II. Universidade Estadual de  
 Campinas, Instituto de Geociências. III. Título



UNIVERSIDADE ESTADUAL DE CAMPINAS/  
INSTITUTO DE GEOCIÊNCIAS/ÁREA DE GEOLOGIA DE  
PETRÓLEO  
PÓS-GRADUAÇÃO EM GEOCIÊNCIAS - ÁREA DE  
GEOENGENHARIA DE RESERVATÓRIOS

**AUTOR:** Marcelo Lopes de Oliveira

**TÍTULO DA DISSERTAÇÃO:** ANÁLISE DAS INCERTEZAS ENVOLVIDAS NA  
MODELAGEM DE RESERVATÓRIOS NO CONTEXTO GEOESTATÍSTICO.

**ORIENTADOR:** Dr. Armando Zaupa Remacre

Aprovada em: 19 / 01 / 98

**PRESIDENTE:** Dr. Armando Zaupa Remacre

**EXAMINADORES:**

**Prof. Dr. Armando Zaupa Remacre**

- Presidente

Handwritten signature of Armando Zaupa Remacre in black ink.

**Prof. Dr. Ademir José Petenate**

Handwritten signature of Ademir José Petenate in black ink.

**Prof. Dr. João Felipe C. L. Costa**

Handwritten signature of João Felipe C. L. Costa in black ink.

Campinas, 19 de janeiro de 1998

**À TERESA.**

# **Agradecimentos**

**À Petróleo Brasileiro S.A. - PETROBRÁS - pela oportunidade e apoio financeiro oferecidos.**

**Ao Dr. Armando Zaupa Remacre pela sugestão do tema da dissertação e pelas valiosas orientações.**

**Aos membros das bancas examinadoras de qualificação e defesa pelas valiosas sugestões.**

**Aos professores e funcionários do IG e do DEP - UNICAMP que participaram da formação teórica e prática durante o mestrado.**

**Às bibliotecárias Alice e Dora pela presteza e colaboração.**

**Aos colegas de mestrado que souberam manter a união, colaboração e amizade.**

**À Nery pelo apoio e confiança em meu trabalho e potencial.**

**Aos meus irmãos Geraldo e Cristóvão, pelas oportunidades que me concederam.**

**À minha esposa Teresa pela compreensão, carinho e incentivo**

**À Deus pela concessão de chegar ao final de mais uma jornada.**



UNIVERSIDADE ESTADUAL DE CAMPINAS/  
INSTITUTO DE GEOCIÊNCIAS/ÁREA DE GEOLOGIA DE  
PETRÓLEO

PÓS-GRADUAÇÃO EM GEOCIÊNCIAS - ÁREA DE  
GEOENGENHARIA DE RESERVATÓRIOS

ANÁLISE DAS INCERTEZAS ENVOLVIDAS NA MODELAGEM DE RESERVATÓRIOS  
NO CONTEXTO GEOESTATÍSTICO

RESUMO

DISSERTAÇÃO DE MESTRADO

**Marcelo Lopes de Oliveira**

A análise dos diversos algoritmos de krigagem e simulação estocástica possibilitou o entendimento do potencial e das limitações das ferramentas geoestatísticas na modelagem de reservatórios. Esses algoritmos diferem em suas hipóteses básicas, faixa de aplicação, complexidade e eficiência computacional. Portanto, cada técnica tem seu uso a depender da fase de desenvolvimento do reservatório, dos objetivos do estudo, dos atributos que estão sendo modelados e, conseqüentemente, da quantidade e da qualidade dos dados disponíveis sobre o reservatório: dados sísmicos, geológicos e de produção.

Assim, diante da diversidade de situações encontradas na modelagem estocástica de reservatórios, é imprescindível o entendimento das características das diferentes alternativas de simulação estocástica disponíveis, para que se possa escolher a metodologia mais adequada ao contexto que está sendo analisado. Nesta dissertação foram analisados os principais algoritmos de simulação estocástica e krigagem com o objetivo de facilitar a análise do tema incerteza. É importante ressaltar que as estimativas, simulações estocásticas e as conseqüentes avaliações de incerteza são dependentes do modelo adotado e de seus parâmetros.

Em relação às krigagens, foram construídos intervalos de incerteza a partir de krigagens paramétricas, enfatizando as hipóteses adotadas. Verificou-se também a influência do aumento do número de dados condicionantes e da representatividade dos mesmos na melhor definição dos semivariogramas e na obtenção de estimativas mais representativas.

Quanto a simulação estocástica, foram obtidas representações de incerteza como mapas de quantis, de probabilidade, de dispersão, etc. Foi implementado o algoritmo de simulação campo de probabilidade, sendo proposta uma alternativa específica para distribuição multivariada gaussiana.

Enfim, a análise crítica de tópicos e problemas específicos sobre avaliação de incertezas evidenciam que muitos conceitos devem ser melhor entendidos, para possibilitar melhor utilização das ferramentas geoestatísticas e, conseqüentemente, melhor conhecimento de suas limitações. Dentre os diversos tópicos abordados, destacam-se: equiprobabilidade das realizações, flutuações ergódicas, número de realizações necessárias para cobrir adequadamente o espaço de incertezas, etc. Para alguns destes problemas, a geoestatística não tem uma solução específica.



**UNIVERSIDADE ESTADUAL DE CAMPINAS/  
INSTITUTO DE GEOCIÊNCIAS/ÁREA DE GEOLOGIA DE  
PETRÓLEO**

**PÓS-GRADUAÇÃO EM GEOCIÊNCIAS - ÁREA DE  
GEOENGENHARIA DE RESERVATÓRIOS**

**GEOSTATISTICAL APPROACH OF UNCERTAINTIES ANALYSIS  
RELATED TO RESERVOIR MODELING**

**ABSTRACT**

**MASTER DISSERTATION**

**Marcelo Lopes de Oliveira**

The analysis of several algorithms of kriging and stochastic simulation allowed the understanding of potential and shortcomings of geostatistics tools for reservoir modeling. Those algorithms differ in range of application, underlying assumptions, complexity of usage and computer efficiency. However, each technique has its application, depending on the reservoir development level, the purpose of the study, the modeled parameter, and consequently on the reservoir quality data: seismic, geologic and production data.

Thus, depending on the diversity of the problems faced stochastic reservoir modeling, it is of utmost importance the understanding of the characteristics related to the different algorithms of stochastic simulation available, and the choosing of methodology better applied to the studied case. In this dissertation, the most used algorithms of stochastic simulation and kriging were analyzed within purpose of facilitating the analysis related to uncertainty. It is important to be aware that the estimations, stochastic simulations and uncertainty evaluations are dependents on the adopted model and its parameters.

Concerning the kriging, confidence intervals from parametric kriging were built with emphasis on the adopted hypotheses. The influence of the increasing of number of the conditioning data and its representativity were also verified to achieve an adequate semivariogram and consistent estimation.

Uncertainty representations such as maps of spread, quantile maps, probability maps were obtained. The algorithm of simulation p-field was implemented and a specific alternative for multigaussian distributions was proposed.

At last, a critical analysis of specifics topics and problems related to uncertainty evaluation show that many concepts need to be better understood to enable better application of the geostatistics tools and, consequently, better knowledge of their shortcomings, such as realizations equiprobability, ergodic fluctuations, number of realizations necessary for to adequately cover the uncertainty space. For some problems, geostatistics does not have a specific solution.

# Sumário

Agradecimentos	v
Resumo	vi
Abstract	vii
Índice	viii
Lista de Figuras	x
Lista de Tabelas	xi
Lista de Símbolos	xii
Lista de Abreviaturas	xv
<b>1. Introdução</b>	<b>1</b>
1.1 Motivação	4
1.2 Objetivos da Dissertação	5
1.3 Organização da Dissertação	6
<b>2. Terminologia</b>	<b>7</b>
2.1 Incerteza e Erro	7
2.2 Precisão e Acurácia	7
2.3 Ergodicidade e <i>Integral range</i>	9
2.4 Heterogêneo e Homogêneo	11
2.5 Sinergismo	12
<b>3. Krigagem</b>	<b>13</b>
3.1 Conceitos Iniciais	13
3.2 Krigagem da Média	14
3.3 Krigagem Simples	15
3.4 Krigagem Ordinária	16
3.5 Krigagem Universal	17
3.6 Krigagem Fatorial	19
3.7 Krigagem Com Deriva Externa	20
3.8 Cokrigagem	21
3.8.1 <i>Cokrigagem Collocated</i>	22
3.9 Krigagem Indicatriz	24
3.10 Cokrigagem Indicatriz	26
3.10.1 <i>Cokrigagem Indicatriz Reduzida</i>	27
3.11 Krigagem dos Componentes Principais Indicatrizes	28
3.12 Krigagem com Dados Qualitativos	30
3.12.1 <i>Tipos de Dados</i>	30
3.12.2 <i>Abordagens Alternativas</i>	32
3.12.3 <i>Krigagem Soft : Modelo de Markov-Bayes</i>	32
3.13 A Questão dos Ponderadores Negativos	35
<b>4. Incertezas na Estimativa</b>	<b>39</b>

4.1 Escolha do Tipo de Krigagem	40
4.2 Medidas de Incerteza	41
4.3 Intervalos de Incerteza com Krigagens Paramétricas	43
4.3.1 Abordagem alternativa	45
4.4 Verificação Experimental	46
4.4.1 Dados Básicos	46
4.4.2 Construção do Caso Base	47
4.4.3 Parâmetros das Krigagens	47
4.4.4 Resultados das Krigagens	49
4.4.5 Intervalos de Incerteza	51
4.4.6 Análise Visual	52
4.4.7 Variogramas Cruzados	53
4.4.8 Comparação através de cortes	54
4.4.9 Intervalos de Incerteza - Abordagem de Roth e Armstrong	55
4.5 Comentários	57
<b>5. Simulação Estocástica</b>	<b>59</b>
5.1 Bandas Rotativas	61
5.2 Simulação Gaussiana Sequencial	63
5.2.1 Cossimulação Sequencial Gaussiana	65
5.3 Simulação Indicatriz Sequencial	66
5.4 Simulação Gaussiana Truncada	68
5.4.1 Simulação Plurigaussiana Truncada	70
5.5 Simulação Campo de Probabilidade	71
5.6 Simulated Annealing	74
5.7 Simulações Baseadas em Objetos	77
<b>6. Incertezas na Simulação Estocástica</b>	<b>80</b>
6.1 Comparação de Algoritmos	81
6.2 Imagens equiprováveis	82
6.3 Escolha de um Algoritmo	83
6.4 Representações de Incertezas	86
6.5 Flutuações Ergódicas	89
6.6 Simulação Condicional, Não Condicional e Krigagem	90
6.7 Ergodicidade	92
6.8 Número de Realizações Necessárias	94
6.9 Simulação Campo de Probabilidade	96
6.10 Animação de Incerteza	97
6.11 Proposta Alternativa Para a Simulação P-field	98
6.12 Incertezas nos Parâmetros e no Modelo	101
<b>7. Conclusões</b>	<b>104</b>
<b>Referências Bibliográficas</b>	<b>108</b>
<b>Bibliografia</b>	<b>114</b>

# Lista de Figuras

Figura 1-1: Incertezas na caracterização e simulação de reservatórios.-----	1
Figura 2-1: Precisão e acurácia. -----	8
Figura 3-1: Informações codificadas como cdfs. Os valores 'a' e 'b' são os limites extremos dos dados.-----	33
Figura 4-1: Mapa base com a posição dos poços. -----	46
Figura 4-2: Histograma dos dados originais. -----	46
Figura 4-3 : Caso base -----	47
Figura 4-4: Conjuntos amostrais retirados do caso base -----	48
Figura 4-5: Resultados das krigagens. -----	50
Figura 4-6: Mapas dos percentis 15.87 e 84.13 %. -----	51
Figura 4-7: Variogramas cruzados entre estimativas e caso real. -----	54
Figura 4-8: Distribuições de volumes porosos em relação ao corte de 1.3 m. -----	55
Figura 4-9: Intervalo de incerteza ao longo de uma linha com coordenada Y fixa, passando por dois poços (figura 7-20 de Harbaugh et al., 1995, p.167) -----	58
Figura 5-1:Krigagem x simulação condicional. -----	60
Figura 5-2: Simulação campo de probabilidade. -----	72
Figura 6-1: Realizações estocásticas utilizando SGS. -----	86
Figura 6-2: Média e desvio padrão das 100 simulações SGS. -----	87
Figura 6-3: Coeficiente de variação das 100 Simulações SGS. -----	87
Figura 6-4: Mapas de probabilidade.-----	88
Figura 6-5: Mapas de quantis. -----	89
Figura 6-6: Semivariogramas das simulações e da krigagem.-----	89
Figura 6-7: Comparação entre os mapas krigado e o da média das 100 simulações. -----	90
Figura 6-8: Esquema ilustrativo para cálculo do HCIP.-----	91
Figura 6-9: Resultados do cálculo dos VOIPs simulados e krigados -----	92
Figura 6-10:Ergodicidade versus a) Alcance; b) Modelo de Semivariograma; c) Tamanho do Domínio-----	93
Figura 6-11: Geração de imagens minimamente diferentes -----	97
Figura 6-12: Volumes porosos totais obtidos com a simulação campo de probabilidade. -----	98
Figura 6-13: Realizações <i>P-field</i> minimamente diferentes -----	99
Figura 6-14: Proposta Alternativa para a <i>P-field</i> -----	100
Figura 6-15: Volume poroso total e média/variância do campo uniforme em cada simulação.-----	101

# Lista de Tabelas

Tabela 1-1: Trabalhos apresentados nos Congressos Internacionais de Geoestatística por país. -----	5
Tabela 4-1: Tipos de krigagem, hipóteses e características. -----	42
Tabela 4-2: Variogramas e Vizinhanças Adotadas nas Krigagens -----	49
Tabela 4-3: Resultados das Krigagens -----	49
Tabela 4-4: Intervalos de Incerteza e Erros -----	52
Tabela 4-5: Fator de correlação -----	53
Tabela 4-6: Volumes porosos para diferentes cortes -----	54
Tabela 4-7: Intervalos de Incerteza -----	56
Tabela 4-8: Intervalos de Incerteza (Estimativa Local: $H\phi$ por bloco em m) -----	56
Tabela 6-1: Resultados obtidos por Srivastava (1997) -----	83
Tabela 6-2: Algoritmos de Simulação: Parâmetros de entrada, Hipóteses, Características. -----	84/85
Tabela 6-3: Estatística das Realizações -----	88
Tabela 6-4: Simulação condicional, não condicional e krigagem -----	91
Tabela 6-5: Expressões do <i>integral range</i> em função do alcance. -----	93
Tabela 6-6: Estatística das Realizações -----	96
Tabela 6-7: Estatísticas das Simulações com Abordagem <i>P-field</i> Alternativa -----	100

# Lista de Símbolos

A	<i>Integral range.</i>
$a_t$	Coefficientes do componente t do modelo de deriva.
B	Matriz ortogonal (autovetores) oriunda da decomposição espectral da matriz covariâncias (Análise do componente principal).
$b_{k',k}$	Elementos da matriz ortogonal B.
$C(x_i - x_j)$	Covariância estacionária entre duas VAs nas posições $x_i$ e $x_j$ .
$C(h)$	Covariância estacionária entre duas VAs separadas por h.
$C(Z(x_1), Z(x_2))$	Covariância entre duas VAs $Z(x_1)$ e $Z(x_2)$ .
$C_{F_i F_j}(x, x + h)$	Covariância cruzada entre fácies $F_i$ e $F_j$ separadas por h.
$C_I(h; z_k)$	Covariância estacionária para o corte $z_k$ .
$C_I(h; z_k, z_{k'})$	Covariância estacionária cruzada entre os cortes $z_k$ e $z_{k'}$ .
$C_1(h)$	Covariância de um dos fatores ou estruturas da covariância imbricada.
$C_p(h)$	Covariância do campo de probabilidades.
$C_u(h)$	Covariância do campo uniforme.
$C_Y(h)$	Covariância estacionária da FA residual $Y(x)$ ou covariância da gaussiana $Y(x)$ .
$C_{\alpha\beta}(h)$	Covariância entre uma variável primária e uma secundária.
$D(x)$	Vetor dos componentes principais indicatrizes.
$D_k(x)$	Cada um dos componentes principais indicatrizes.
$E[Z(x)]$	Esperança matemática de $Z(x)$ .
$E[I(x; z_k)]$	Esperança da indicatriz do corte $z_k$ na posição x.
$E[I(x; z_k) (n)]$	Esperança condicional da indicatriz do corte $z_k$ na posição x em relação aos dados indicatrizes da vizinhança.
$F(z_k)$	Função de distribuição acumulada estacionária relativa ao corte $z_k$ .
$F(x; z_k)$	Função de distribuição acumulada não estacionária relativa ao corte $z_k$ .
$F(x; z_k (n))$	Função de distribuição acumulada não estacionária na posição x até o corte $z_k$ condicionada a n dados na vizinhança de x.
$F_Z$	Função de distribuição acumulada univariada, obtida a partir de dados amostrais, representativa da área em estudo.
$F^{-1}(x; p (n))$	Inversa da cdf para o valor de probabilidade p na posição x.
$f_t(x)$	Funções que compõem a deriva.
$G(.)$	Função de distribuição acumulada gaussiana padrão.
$H\phi$	Produto espessura x porosidade (m).

$I_{F_i}(x)$	Indicatriz da Fácies $F_i$ em uma dada posição $x$ .
$I(x; z_k)$	Indicatriz correspondente a um corte $z_k$ na posição $x$ .
$m$	Esperança matemática estacionária de uma VA.
$m(x)$	Esperança matemática não estacionária, deriva.
$m_e$	Esperança matemática da ccdf gaussiana, no caso o valor da estimativa.
$m_{cg}$	Esperança matemática do campo gaussiano, no caso igual a 0.
$n_\alpha$	Número de amostras de um tipo de variável na cokrigagem.
$N(0,1)$	Distribuição gaussiana padrão com média zero e variância unitária.
$N(h)$	Número total de pares de amostras separadas pela distância $h$ .
$O_c$	Componentes da função objetivo.
$P(x)$	Campo de Probabilidade.
$P[Z(x) \leq z_k]$	Probabilidade de $Z(x)$ ser menor ou igual ao corte $z_k$ .
$P[Z(x) \leq z_k   (n)]$	Probabilidade condicional de $Z(x)$ ser menor ou igual ao corte $z_k$ considerando as informações da vizinhança.
$P[Z(x) \leq z   v(x)]$	Probabilidade de $Z(x)$ ser menor ou igual a $z$ dado que a variável secundária $V$ tem o valor $v$ no mesmo ponto $x$ .
$R$	Alcance, amplitude prática, <i>practical range</i> .
$s(x)$	Função da deriva externa.
$U(Z(x))$	Transformação uniforme de $Z(x)$ .
$V(x)$	Variável secundária.
$Var(.)$	Variância.
$x_0$	Posição genérica em que é realizada a estimativa.
$W_c$	Ponderadores da função objetivo.
$y_i$	Parâmetros de corte ( <i>cutoffs</i> ) gaussianos.
$Y(x)$	Componente estocástico com média zero e covariância conhecida; VA correspondente às ccdfs prévias na krigagem de Markov-Bayes; Função gaussiana nas simulações gaussianas.
$Z_{\alpha_1}^*$	Estimativa de uma Variável na cokrigagem.
$z(x_i)$	Valor amostral na posição $x_i$ .
$Z_{jk}$	Variável aleatória correspondente a um atributo $k$ em uma posição $j$ .
$z_{kd}(x)$	Valor obtido através da krigagem dos dados condicionantes (bandas rotativas).
$z_{ksnc}(x)$	Valor obtido através da krigagem dos valores simulados (bandas rotativas) não condicionalmente na mesma posição dos dados originais condicionantes.
$Z_L$	Fator que compõe uma variável aleatória para a KF.
$z_{sc}(x)$	Valor simulado condicional.
$z_{snc}(x)$	Valor simulado não condicional.

$z_{si}(x)$	Resultado de cada uma das $i$ simulações unidimensionais da TB.
$\Lambda$	Matriz diagonal correspondente aos autovalores da decomposição espectral da matriz de covariâncias.
$\forall$	Para todo e qualquer valor.
*	Referente à estimativa.
$\varepsilon_m$	Erro médio da estimativa.
$\varepsilon_{ma}$	Erro médio absoluto da estimativa.
$\xi_i$	Ponderadores da krigagem relativos a uma variável transformada auxiliar.
$\lambda_i$	Ponderadores da krigagem.
$\sigma^2$	Variância.
$\sigma_k^2$	Variância de krigagem.
$\sigma$	Desvio padrão.
$\sigma_e$	Desvio padrão da cdf gaussiana, no caso a raiz quadrada da $\sigma_k^2$ .
$\sigma_{cg}$	Desvio padrão do campo gaussiano, no caso igual a 1.
$2\gamma(\cdot)$	Variograma.
$\gamma_I(h; z_k)$	Semivariograma indicatriz estacionário para o corte $z_k$ .
$\gamma_I(h; z_k, z_{k'})$	Semivariograma indicatriz cruzado estacionário entre os cortes $z_k$ e $z_{k'}$ .
$\gamma_{\alpha\alpha_1}(h)$	Semivariograma cruzado entre duas variáveis na cokrigagem.
$\rho(h)$	Correlograma.
$\mu$	Multiplicador de Lagrange.
$\delta_{kk_0}$	Delta de Kronecker.

# Lista de Abreviaturas

ACP	Análise do Componente Principal
cb	Relativo ao caso base
ccdf	Função de distribuição acumulada condicional.
cdf	Função de distribuição acumulada.
coKI	Cokrigagem Indicatriz
coKIr	Cokrigagem Indicatriz Reduzida
COK	Cokrigagem Ordinária
COKS	Cokrigagem Simples
FA	Função aleatória.
FAIk	Função aleatória intrínseca de ordem k.
FO	Função objetivo.
GSLIB	Geostatistical Software Library
VOIP	Volume de óleo <i>in place</i> .
HCIP	Volume de hidrocarboneto <i>in place</i> .
IC	Intervalo de Incerteza
IPCK	Krigagem dos Componentes Principais Indicatrizes
ISATIS	Software Geoestatístico da Geovariances.
KDE	Krigagem com Deriva Externa
KF	Krigagem Fatorial
KI	Krigagem Indicatriz
KIm	Krigagem Indicatriz Mediana
KM	Krigagem da Média
KO	Krigagem Ordinária
KS	Krigagem Simples
KT	Krigagem com modelo de <i>trend</i> , deriva.
KU	Krigagem Universal
LU	Decomposição LU (lower/upper) de Cholesky
Mm <sup>3</sup>	Volume em milhares de metros cúbicos.
pdf	Função de densidade de probabilidade
<i>P-field</i>	Campo de Probabilidade

RA	Abordagem de Roth e Armstrong (1995) para intervalos de incerteza locais.
SA	<i>Simulated Annealing</i>
SCRF	Publicações de Stanford Center for Reservoir Forecasting
SGS	Simulação Gaussiana Seqüencial
SIS	Simulação Indicatriz Seqüencial
TB	Simulação <i>Turning Bands</i> ou Bandas Rotativas
VA	Variável aleatória.
VPT	Volume poroso total do reservatório.

# 1. Introdução

Muitos e diversificados processos geológicos tais como sedimentação, erosão, migração, diagênese produzem complexas distribuições espaciais das propriedades do reservatório. O desafio atual em caracterização de reservatórios é combinar o maior número de componentes relevantes das incertezas sobre as características do reservatório (volume de hidrocarboneto *in-place*, falhas, continuidade das argilas e areias, tamanho de aquífero, permeabilidades relativas, etc) com intuito de obter uma distribuição de probabilidade da performance de produção.

Mesmo com os grandes avanços em todas as áreas que envolvem a estimativa de performance de reservatórios petrolíferos, muitas incertezas ainda persistem.

Muitos fatores contribuem para a incerteza na estimativa de performance e do volume de hidrocarboneto *in-place* (HCIP). Das quatro incertezas na figura 1.1, as que mais se destacam são as referentes às etapas de caracterização do reservatório e mudança de escala (*upscaling*).

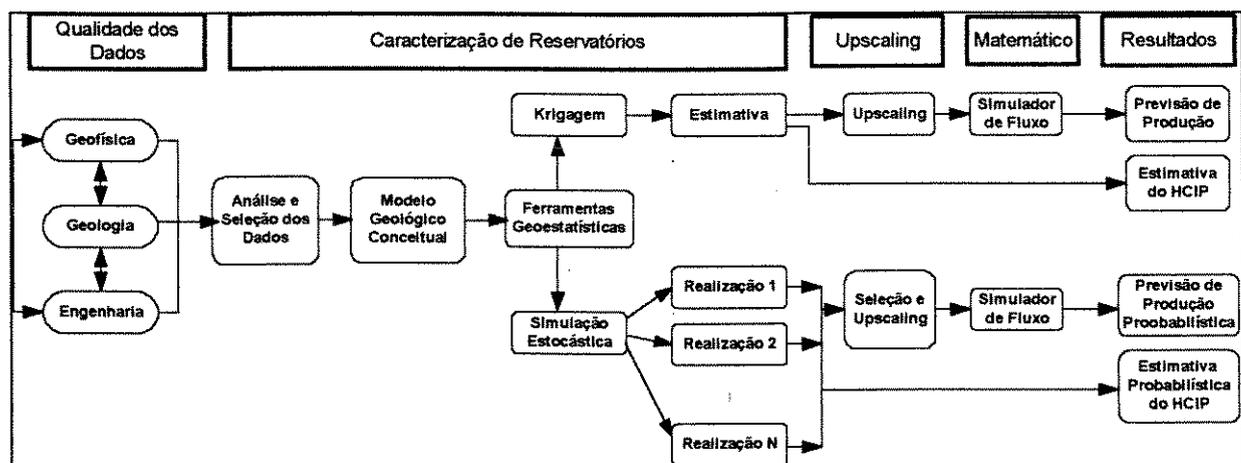


Figura 1-1: Incertezas na caracterização e simulação de reservatórios.

A primeira etapa consiste no levantamento dos dados oriundos da geologia, engenharia e geofísica. Constantes avanços têm possibilitado a obtenção de dados com melhor qualidade e com menor incerteza, por exemplo: melhorias nas técnicas de perfilagem, na execução e interpretação de testes, utilização de sísmica 3D, etc. Ressalta-se que mesmo nesta fase, têm-se muitas incertezas, principalmente porque medidas são obtidas de forma indireta e

as condições de coleta e análise das amostras não representam convenientemente as condições de subsuperfície.

Em seguida, começa a etapa de caracterização de reservatórios propriamente dita. São realizadas análises estatísticas dos dados, correlações rocha-perfil, definição de eletrofácies e das unidades de fluxo. O segundo passo na caracterização de reservatórios é a construção de um modelo geológico conceitual quando são definidos os ambientes deposicionais e diagenéticos e o arcabouço estrutural e estratigráfico da acumulação.

Todas as informações servirão de base para a modelagem do reservatório no que diz respeito à distribuição de fácies e de suas características permoporosas. Procura-se extrapolar características petrofísicas para todas as regiões do reservatório a partir das informações nos poços, dos levantamentos sísmicos, dados de afloramentos, etc. A geoestatística possibilita a integração de dados de diversos tipos, escalas, suportes e graus de confiabilidade. As abordagens geoestatísticas são a krigagem e a simulação estocástica.

A krigagem envolve a obtenção dos atributos em cada célula (bloco) do reservatório para a alimentação dos simuladores de fluxo ou obtenção de mapas para cálculo de volumes. É obtida uma única realização segundo o critério de minimização da variância do erro. Os mapas krigados capturam as grandes variações, mas sempre suavizam e não descrevem realisticamente a variabilidade do fenômeno. Este comportamento pode ser crucial quando as imagens krigadas são submetidas às simulações de fluxo.

Haldorsen e Damsleth (1993) reportaram que os projetos iniciais para campos descobertos, avaliados e explorados no Mar do Norte do Reino Unido apresentaram alto percentual de erros nas estimativas de custos, produção de óleo e água, etc. De uma forma geral apresentaram grandes diferenças, principalmente pelo fato de que o *breakthrough* de água ter ocorrido muito mais cedo do que o previsto nos projetos. Os principais motivos expostos por Haldorsen e Damsleth (1993) para estes resultados foram:

- Qualidade dos dados sísmicos - disponibilidade apenas de sísmica 2D resultando em reduzido número de falhas e mapas estruturais incorretos;
- Modelos de simulação de fluxo com propriedades do reservatório muito homogêneas, com pouco detalhamento de heterogeneidades;
- Exagerada comunicação vertical e lateral nos modelos de simulação.

A necessidade de modelar realisticamente heterogeneidades do reservatório tem levado a adoção da abordagem estocástica. A abordagem estocástica envolve a determinação de inúmeros modelos prováveis para o reservatório sob estudo, em termos de distribuição de

fácies e de suas propriedades petrofísicas. Há duas classes de modelos estocásticos: os discretos e os contínuos. Os discretos foram desenvolvidos para descrever características geológicas de natureza discreta: posição e dimensões dos corpos de areia (por exemplo canal, *crevasses* em um ambiente deposicional fluvial); distribuição e tamanhos dos corpos de argila suspensos no meio das areias; distribuição, orientação e comprimento de fraturas e falhas; modelagem de fácies, etc. Em todos esses casos, um ponto no espaço pertence a apenas um limitado número de classes, e o modelo estocástico controla como as classes interagem em cada ponto.

Já os modelos contínuos foram desenvolvidos para descrever um fenômeno que varia continuamente, por exemplo as propriedades da rocha tais como: porosidade, permeabilidade, velocidades sísmicas, etc.

Para o técnico que faz uso da simulação estocástica, as diferenças entre os diversos algoritmos disponíveis e os efeitos nos resultados da modelagem necessitam ser melhor entendidos para que possa utilizar o algoritmo adequado à situação que ele estiver analisando. Cada algoritmo de simulação estocástica tem seu campo de aplicação, não havendo o melhor para qualquer situação.

As avaliações de incertezas derivadas de análises geoestatísticas devem sempre ser acompanhadas das condições (hipóteses) sob as quais os resultados foram obtidos. Deve-se sempre ter em mente que a incerteza não está apenas na variabilidade entre realizações a modelos fixos e parâmetros fixos, mas também na validade do próprio modelo ou em seus parâmetros. **Um modelo por definição é uma simplificação ou uma aproximação do fenômeno real, e conseqüentemente não pode representar na totalidade todos os seus aspectos particulares.**

A dissertação estará limitada ao estudo das incertezas envolvidas na modelagem do reservatório com as ferramentas geoestatísticas.

É salutar que outras incertezas sejam introduzidas no processo de avaliação da performance do reservatório, destacando-se principalmente a etapa de mudança de escala (*upscaling*) que basicamente é a transferência da malha geoestatística refinada para uma malha mais grosseira para simulação de fluxo. Nesta etapa, grandes incertezas são introduzidas, sendo que muitas vezes, detalhes das heterogeneidades modeladas na fase anterior, são parcialmente perdidos. Outras incertezas são introduzidas quando da seleção das imagens a serem submetidas a simulação de fluxo. Têm-se também as incertezas devido às simplificações envolvidas nos simuladores numéricos. Grandes progressos já foram alcançados e vários

modelos de simuladores foram desenvolvidos: *black-oil*, composicional, modelos térmicos, etc. Uma das limitações em termos de simulação de fluxo é a impossibilidade de utilizar modelos com “bilhões de blocos”, o que descartaria a necessidade da mudança de escala.

Muitos trabalhos têm sido realizados no sentido de identificar e reduzir as incertezas nas várias etapas da caracterização e simulação de fluxo para previsão de comportamento. Contudo, poucos trabalhos têm sido realizados no sentido de explicitar as incertezas intrínsecas da modelagem geoestatística.

## **1.1 Motivação**

Nos últimos congressos de geoestatística, foram apresentados trabalhos voltados para analisar a capacidade da geoestatística enquanto ferramenta de avaliação do espaço de incerteza de um determinado atributo. Em especial, Srivastava (1997) fez uma análise crítica da geoestatística, expondo os principais problemas que devem ser analisados pela comunidade de geoestatísticos. Dentre estes problemas destacam-se: geoestatística e as universidades, softwares geoestatísticos, notação geoestatística e quantificação de incertezas.

Em seu trabalho Srivastava (1997) reporta que ainda são poucas as universidades que têm a disciplina geoestatística a nível de graduação e que deve se dar prioridade na formação de novos professores. O levantamento dos trabalhos publicados nos últimos congressos sobre geoestatística e geologia quantitativa (tabela 1-1) representa um forte indício de que, a nível de Brasil, esta questão também deve ser relevante. Neste levantamento foi considerado apenas a instituição ou universidade do primeiro autor, na época da publicação. Apesar do levantamento incluir trabalhos de universidades, consultores e da indústria, verifica-se claramente a preponderância dos principais centros de referência em geoestatística: França e Estados Unidos. Até mesmo o destaque de outros países como Portugal, Canadá e Noruega pode ser observado. A nível de indústria de petróleo, alguns trabalhos de técnicos e geoestatísticos brasileiros têm sido publicados no SCRF, na *Mathematical Geology* e nos congressos específicos da área.

Uma outra questão levantada por Srivastava (1997) diz respeito aos softwares que, segundo ele, apresentam *bugs* e algoritmos muito rígidos, difíceis de serem otimizados para uma situação específica. O autor também reporta que a notação geoestatística utilizada nas publicações é bastante diversificada, o que dificulta o intercâmbio de idéias e disseminação de novos algoritmos.

**Tabela 1-1: Trabalhos apresentados nos Congressos Internacionais de Geoestatística por país.**

País	Congresso 83	Congresso 88	Congresso 92	Congresso 96	Total
França	20	32	18	16	86
USA	27	14	19	20	80
Portugal	1	6	6	14	27
Canadá	9	2	8	6	25
Noruega	-	1	4	16	21
Inglaterra	3	4	5	3	15
África do Sul	2	3	4	4	13
Austrália	1	2	2	6	11
Espanha	-	1	5	4	10
Holanda	-	3	4	3	10
Brasil	-	2	1	4	7
Outros (15)	2	9	9	15	35
Total	65	79	85	111	340

A questão central levantada por Srivastava (1997) diz respeito ao espaço de incertezas coberto pelos algoritmos de simulação estocástica. Alguns outros trabalhos de Srivastava (1994) e de outros autores como por exemplo Journel (1994 e 1997) enfocam tópicos relativos a incertezas na modelagem estocástica, que muitas vezes são mal compreendidos. A motivação foi basicamente fazer uma análise crítica e uma reflexão sobre conceitos e temas relativos a incertezas, quando da utilização da ferramenta geoestatística na modelagem de reservatórios. Será dado um enfoque às aplicações da área de petróleo, porém muitas das questões abordadas são extensíveis a outros campos de aplicação.

As questões levantadas por Srivastava (1997) devem ser motivo de reflexão por parte da comunidade de geoestatísticos.

## ***1.2 Objetivos da Dissertação***

Assim, o objetivo primordial desta dissertação é fazer uma reflexão e análise crítica, esclarecendo conceitos relativos às incertezas na modelagem geoestatística. Serão utilizados exemplos práticos e alguns disponíveis na literatura. Um objetivo secundário é revisar os principais métodos de estimativa (krigagem) e de simulação estocástica com ênfase na avaliação de incertezas. Para as principais metodologias serão feitos breves resumos para que a discussão sobre incerteza se torne facilitada.

Algumas das questões abordadas não têm uma solução, como por exemplo o número de simulações estocásticas necessárias para cobrir adequadamente o espaço de incertezas. Este, dentre outros conceitos, serão abordados de forma a elucidá-los.

### ***1.3 Organização da Dissertação***

Trata-se de uma dissertação baseada em uma extensa revisão bibliográfica, acrescida de reflexão e análise crítica, expondo as questões relevantes em termos de incertezas envolvidas na utilização da geoestatística como ferramenta auxiliar na caracterização de reservatórios.

O capítulo 2 apresenta um conjunto de termos e conceitos envolvidos na modelagem estocástica de reservatórios que dada a importância dos mesmos no escopo do trabalho optou-se por constituir um capítulo em separado no início da dissertação.

No capítulo 3, são apresentados os principais tipos de krigagem. Serão abordadas as metodologias que possibilitam a utilização de dados qualitativos e a evolução das propostas que enfocam a questão da ocorrência dos ponderadores negativos.

No capítulo 4, são apresentadas as principais incertezas envolvidas na krigagem. São analisados a influência do número de dados condicionantes e do viés na amostragem na obtenção da estimativa. Foram comparadas duas abordagens para geração de intervalos de incerteza a partir das krigagens paramétricas: a tradicional (distribuição multivariada gaussiana dos erros de estimativa) e a metodologia proposta por Roth e Armstrong (1995).

O capítulo 5 compreende uma revisão bibliográfica sobre os principais métodos de simulação estocástica.

No capítulo 6, são apresentadas maneiras de visualização de incerteza referentes a simulação de variáveis aditivas (porosidade, volume poroso, etc). Foi implementado o algoritmo de simulação campo de probabilidade. Serão também abordados temas relacionados à incerteza, tais como: flutuações ergódicas e *integral range*, equiprobabilidade das realizações, incertezas nos parâmetros e no modelo, etc.

No capítulo 7 estão as principais conclusões.

## **2. Terminologia**

A utilização da geoestatística como ferramenta de modelagem de reservatórios tem colocado uma porção de conceitos nem sempre bem entendidos pelos usuários finais. Os termos descritos a seguir fazem parte de um trabalho maior realizado em paralelo à preparação da dissertação. Este trabalho denominado Glossário Geoestatístico Básico (Oliveira; Remacre, 1997) está em fase final de elaboração.

### ***2.1 Incerteza e Erro***

Em sentido lato, erro é a diferença entre um valor medido ou calculado e o valor correto. Contudo erro em problemas de modelagem estocástica não pode ter esta abordagem. Na realidade, quando são usados modelos para obter estimativas não se conhece o valor real, mas apenas valores potenciais para o atributo. Sem uma possibilidade de determinação do erro nestas circunstâncias, pode-se introduzir o termo que é definitivamente melhor - incerteza.

Incerteza é o conceito ou condição de estar em dúvida sobre um valor. Note que na definição de incerteza não há julgamento da exatidão e precisão de um dado valor. Erro, como usado em modelagem matemática, poderia ser definido como um desvio de um resultado de algum valor considerado o valor real para aquele lugar, tempo e sob aquelas condições.

O conhecimento incompleto do fenômeno e a falha na inclusão de todos os fatores pertinentes que influenciam a variável de interesse resultam necessariamente em incertezas. As ferramentas geoestatísticas são usadas com o objetivo de delimitar estas incertezas.

### ***2.2 Precisão e Acurácia***

Na literatura técnica, às vezes, os termos precisão e acurácia são usados com o mesmo significado. Porém, alguns geoestatísticos evidenciam a diferença entre os dois termos. Técnicas geoestatísticas são usadas para construir modelos probabilísticos de incerteza sobre valores verdadeiros não conhecidos. A qualidade desses modelos probabilísticos pode dar as medidas relativas aos conceitos de precisão e acurácia.

A idéia básica da geoestatística é modelar a incerteza a respeito de um valor desconhecido de  $z$  sobre uma variável aleatória (VA)  $Z$  caracterizada por uma específica distribuição de probabilidade. O valor desconhecido pode ser um parâmetro global tal como o fator de recuperação, o tempo de *breakthrough*, ou um atributo local tal como a porosidade e a permeabilidade em uma determinada região ou posição do reservatório.

Uma distribuição é dita acurada se dado algum intervalo de probabilidade, ela contém a resposta verdadeira. Por exemplo Deutsch e Journel (1992) utilizaram intervalo de probabilidade de 95%, enquanto Gotway e Rutheford (1994), 80%.

A precisão é avaliada diretamente da dispersão da distribuição, o que facilita a comparação de algoritmos. Será tanto mais precisa, quanto mais estreita for a distribuição. Contudo é impossível avaliar a acurácia, uma vez que não se conhece a priori, o valor verdadeiro do atributo ou da performance do reservatório que se está simulando. Uma condição mínima para acurácia é que a técnica utilizada leve em consideração todos as informações consideradas relevantes. Na Figura 2-1 estão exemplos de distribuições de resposta com diferentes situações de precisão e acurácia. O ideal é obter uma distribuição precisa e acurada.

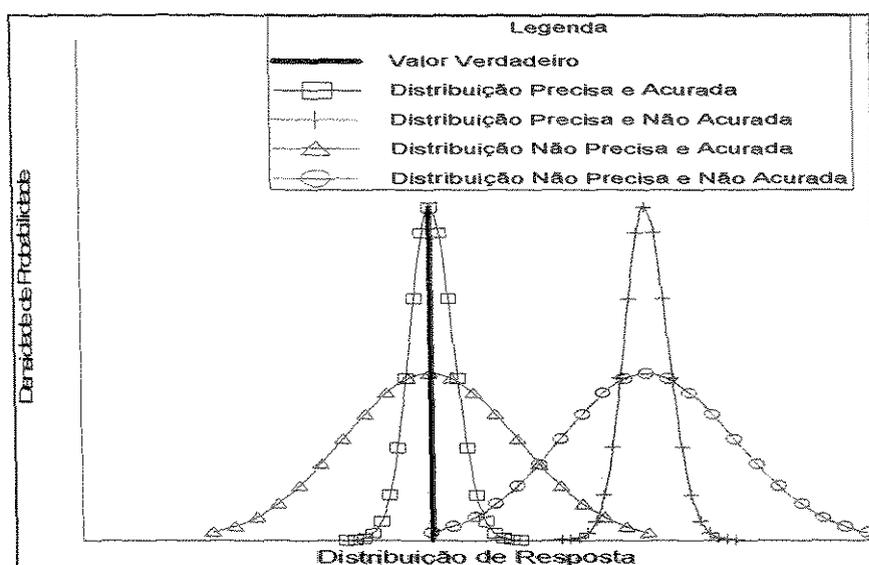


Figura 2-1: Precisão e acurácia.

No contexto de avaliação do modelo probabilístico, Deutsch (1997) propôs definições específicas de acurácia e precisão utilizando uma abordagem tipo validação cruzada, uma vez que a avaliação da acurácia requer o conhecimento do valor real.

## 2.3 Ergodicidade e Integral Range

Ergodicidade está relacionada com a questão de determinar as estatísticas de um processo estocástico  $Z(x)$  a partir de uma única realização. Um processo estocástico é dito ergódico se com probabilidade 1 todas as suas estatísticas podem ser determinadas a partir de uma única realização (Papoulis, 1965, p.327). Ressalta-se que o conceito de ergodicidade foi definido inicialmente para processos estocásticos temporais, sendo estendido posteriormente para processos estocásticos espaciais.

Geralmente não se está interessado em todas as estatísticas do processo, mas em apenas algumas particulares. Assim, pode-se definir ergodicidade com respeito a estas estatísticas. Por exemplo, considere uma função aleatória (FA)  $Z = (Z(x), x \in \mathbb{R}^D)$  com média  $\mathbf{m}$ , variância  $\sigma^2$  e função de correlação  $\rho(h)$ . Dispondo de uma realização desta FA no domínio  $V$ , um estimador natural para a média  $\mathbf{m}$  é:

$$Z(V) = \frac{1}{|V|} \int_V Z(x) dx$$

Este estimador  $Z(V)$  tem variância

$$\text{Var}[Z(V)] = \frac{\sigma^2}{|V|^2} \int_V \int_V \rho(x-y) dx dy$$

onde  $\rho(h) = C(h) / \sigma^2$ .

A precisão deste estimador será tão maior quanto maior for o domínio; será máxima quando o volume  $V$  for infinito. Contudo, os domínios analisados em geoestatística são finitos. Por esta razão é utilizado o conceito de ergodicidade.

Uma FA aleatória é dita de média ergódica se:

$$\lim_{V \rightarrow \infty} \text{Var}\{Z(V)\} = 0$$

Então, a ergodicidade faz com que  $Z(V)$  seja assintoticamente um bom estimador de  $\mathbf{m}$ . Mas a ergodicidade não indica quão grande deverá ser o valor de  $V$  para que a variância de  $Z(V)$  se torne negligenciável. Para isto é utilizado o conceito de *integral range*<sup>1</sup>.

O *integral range* (Lantuéjoul, 1991, p.390) é definido pela fórmula:

$$A = \lim_{V \rightarrow \infty} |V| \frac{\text{Var}\{Z(V)\}}{\sigma^2}$$

---

<sup>1</sup> Alternativas para '*integral range*' são as expressões 'alcance integral' ou 'domínio de correlação'. Optou-se por não utilizá-las na dissertação.

Esta quantidade é não negativa, mas pode ser zero ou infinita. Ela tem a dimensão de um volume-D. Pode ser dada a seguinte interpretação física para o *integral range*: suponha  $0 < A < +\infty$ . Se  $|V|$  é muito maior do que A, então:

$$\text{Var}\{Z(V)\} \approx \frac{\sigma^2 A}{|V|}$$

E pode se determinar um inteiro tal que

$$N \approx \frac{|V|}{A} \text{ e portanto}$$

$$\text{Var}\{Z(V)\} = \frac{\sigma^2}{N}.$$

O lado direito da fórmula acima representa a variância da média de  $n$  observações independentes. O domínio  $V$  pode ser dividido em  $n$  subdomínios independentes e disjuntos de tamanho  $A$ . Interpreta-se assim o *integral range* como a escala do fenômeno. Lantuéjoul (1991) apresentou uma metodologia prática para determinação do domínio  $V$  para se obter uma variância de  $Z(V)$  dentro de limites predeterminados.

O *integral range* também pode ser escrito com a fórmula (Matheron, 1989, p.84):

$$A = \int_{R^D} \rho(h) dh$$

Quando a função de correlação  $\rho(h)$  tem uma baixa taxa de decrescimento, o *integral range* é infinito. Porém, para um modelo de semivariograma que atinge um patamar, pode-se dizer que o *integral range* objetivamente existe. Só que, muitas vezes, o domínio em que será realizada a simulação é pouco maior que o *integral range* e não se observará ergodicidade.

A noção de ergodicidade pode ser estendida para qualquer outro parâmetro do modelo, por exemplo, a função covariância ou o semivariograma.

Em simulação estocástica, utiliza-se a expressão ‘flutuações ergódicas’ para se referir às variações nas estatísticas das várias realizações em relação a estatística do modelo de FA escolhido. Por exemplo, os variogramas e as cdfs (funções de distribuição acumulada condicionais) das realizações apresentam diferenças em relação ao variograma e à cdf (função de distribuição acumulada) que foram utilizados para gerá-las. As flutuações ergódicas aumentam o espaço de incertezas. Realizações que honrem exatamente os modelos de cdf e variograma podem ser consideradas como um subconjunto de todas as realizações.

Algumas flutuações ergódicas são aceitáveis desde que o modelo estatístico esteja afetado por flutuações de amostragem. Em muitas aplicações, os modelos estatísticos são inferidos de amostras esparsas e não podem ser considerados como perfeitamente conhecidos. Portanto, desvios em relação ao modelo estatístico podem ser desejáveis, pois eles se relacionam com o aspecto da incerteza inerente.

A partir das formulações apresentadas por Matheron (1989) e Lantuéjoul (1991) pode-se verificar que o tamanho de um domínio  $V$  suficiente para verificar a ergodicidade depende do alcance<sup>2</sup> do modelo do semivariograma. Quanto mais contínuo for o semivariograma na origem e maior o alcance do semivariograma, maior será o domínio  $V$  necessário para se verificar a ergodicidade.

Ergodicidade é uma propriedade do modelo probabilístico e não característica dos dados ou fenômenos sob estudo. É a propriedade da ergodicidade que permite inferir distribuição e momentos do modelo a partir de uma única realização.

## ***2.4 Heterogêneo e Homogêneo***

Heterogeneidade é equivalente a variabilidade espacial. Significa que o valor de um atributo muda com a sua posição no reservatório. Portanto, quando se considera, para efeito de simulação de fluxo, reservatório homogêneo em relação a algum atributo, está se considerando que este atributo é uniforme no reservatório, isto é, ele não varia com a posição.

Um reservatório apresenta vários níveis de heterogeneidades geológicas, cada um deles influenciando diferentemente o fluxo de fluidos e conseqüentemente a produção de hidrocarbonetos.

De uma forma geral, os sistemas de classificação de heterogeneidades se baseiam na escala de observação, na origem genética e na influência no fluxo de fluidos. Por exemplo, Haldorsen e Lake (1984) analisando as propriedades internas dos reservatórios, que envolvem fatores como continuidade dos folhelhos, valores de porosidade e escala de observação, definiram quatro escalas de heterogeneidades:

- escala microscópica (escala de poros e gargantas -  $\mu\text{m}$ );
- escala macroscópica (escala de *plugs* - cm);
- escala megascópica (escala de blocos para simulação de fluxo - 200 a 400 m);

---

<sup>2</sup> Alcance ou amplitude prática é a distância  $h$ , na qual o semivariograma atinge o patamar, ou 95% do mesmo para os modelos que o atingem assintoticamente. O alcance representa o comprimento de correlação.

- escala gigascópica (escala da formação ou reservatório - maior que 1000 m).

Um outro conceito, algumas vezes mal relacionado com homogeneidade, é a estacionariedade. Madureira et al. (1994, p.55) reportam que “estacionariedade é um outro nome para a noção intuitiva de homogeneidade espacial...”. A noção correta de estacionariedade é a de um processo estocástico que apresenta flutuações regulares em torno de um valor médio, o que necessariamente não implica em um reservatório homogêneo.

## ***2.5 Sinergismo***

Há uma forte tendência na indústria de petróleo para agrupar todos os técnicos em uma equipe. Deve-se ter em mente que não existe o trabalho individual de um profissional, mas sim o trabalho de uma equipe. Cada profissional pode ser um especialista em sua área de atuação, porém não se pode prescindir que este profissional tenha um conhecimento generalizado de todas as etapas envolvidas na caracterização de reservatórios. Assim, o profissional saberá das limitações e das incertezas introduzidas nas etapas anteriores e os possíveis impactos em seu trabalho e conseqüentemente quais novas incertezas são introduzidas na etapa por ele trabalhada. Damsleth e Tjolsen (1993) sugerem que as universidades incluam um pouco mais de disciplinas ligadas à matemática, estatística e computação nos currículos de geologia. A recomendação multidisciplinar desses autores pode ser estendida para cada profissional que compõe a equipe de gerenciamento.

Estes profissionais, sejam engenheiros, geólogos ou geofísicos, devem ter conhecimento multidisciplinar, sem contudo perderem a identidade profissional. Trata-se do sinergismo em gerenciamento de reservatórios. No escopo desta dissertação, várias vezes será utilizado o termo usuário para se referir ao profissional da equipe de gerenciamento que está mais envolvido com a modelagem de reservatórios via geoestatística. É evidente que todos, indistintamente, devem ter embasamento geoestatístico, tão mais especializado quanto maior for o envolvimento na utilização da ferramenta geoestatística para a modelagem de reservatórios e mais importantes forem os estudos desenvolvidos. A utilização dos algoritmos de simulação estocástica requer por parte do usuário, bom conhecimento do processo físico do fenômeno sob estudo, e conseqüentemente do modelo escolhido. Não se trata de aplicar as ferramentas geoestatísticas a qualquer reservatório que se estiver analisando. É preciso identificar qual o tipo de heterogeneidade que mais influencia o fluxo de fluidos no meio poroso. Também é preciso saber quais técnicas geoestatísticas são aplicáveis, se os dados disponíveis são suficientes para a modelagem escolhida e, principalmente, se o objetivo do estudo requer um estudo mais ou menos detalhado.

## 3. Krigagem

Geoestatística é um ramo da estatística que se originou na indústria mineira no início da década de 50 para auxiliar na estimativa de reservas minerais. Nesta área, ressaltam-se os trabalhos estatísticos de Krige sobre jazidas auríferas e os trabalhos de Wijs sobre jazidas uraníferas, ambos na África do Sul.

Nos últimos anos da década de 50, as técnicas adotadas pela escola sul-africana chamaram a atenção de engenheiros franceses do Comissariado de Energia Atômica, e em especial de Georges Matheron, que a partir delas, desenvolveu conceitos inovadores colocando-as dentro de um arcabouço teórico chamado Teoria das Variáveis Regionalizadas

Originalmente desenvolvido para solucionar problemas de estimativas de reservas minerais, a geoestatística tem se expandido para outras áreas das ciências naturais.

Krigagem, procedimento geoestatístico introduzido por Matheron, foi definido por ele como uma estimativa de um atributo (no caso teor de minério), em um volume de suporte, através da ponderação de todas as amostras disponíveis, onde os pesos são obtidos com a restrição de que a somatória dos mesmos seja igual a 1 e a variância de estimativa seja mínima.

A krigagem é originalmente um estimador linear. Nos recentes desenvolvimentos em geoestatística, métodos de estimativas espaciais não lineares têm tomado parte da ‘família de krigagens’.

### 3.1 Conceitos Iniciais

Dispõe-se de  $n$  dados  $z(x_i)$  em posições  $x_i$ . Admite-se que em cada uma das  $i$  posições, tem-se uma variável aleatória (VA)  $Z(x_i)$  e que dada uma posição adicional  $x_0$ , tem-se que  $Z(x_0)$  também é uma VA em  $x_0$ .

Assume-se também que estas VAs são um subconjunto de um função aleatória (FA)  $Z(x)$  com estacionariedade de segunda ordem sobre um domínio  $D$ :

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

$$C[Z(x+h), Z(x)] = C(h)$$

Deseja-se estimar  $Z(x_0)$  em uma posição pontual conhecida  $x_0$ , baseando-se nos dados disponíveis  $z(x_i)$ .

Os diferentes tipos de krigagem podem dar uma falsa impressão de complexidade. Porém, todas as krigagens são técnicas de regressão que diferem apenas nos tipos particulares de funções obtidas a partir dos dados, que estão sendo recombinaadas para a obtenção da estimativa. Assim, as principais técnicas de krigagem serão apresentadas com o objetivo de esclarecer tanto a obtenção da estimativa, quanto dos intervalos de incerteza, e consequentemente, das funções de distribuição acumulada condicional (ccdfs) que são utilizadas em alguns algoritmos de simulação estocástica. As técnicas básicas de krigagem são a krigagem simples e a krigagem ordinária, sendo que as demais constituem variações das mesmas em relação a utilização de mais de um tipo de atributo desde que correlacionados, a modelagem com deriva, ou mesmo transformações realizadas nos dados originais, por exemplo a indicatriz. De uma forma geral cada metodologia constitui uma alteração das abordagens básicas iniciais ora no sentido de incorporar mais informações oriundas de outras fontes de dados, ora no sentido de diminuir o tamanho dos sistemas de krigagens priorizando as informações adicionais que são mais conseqüentes para a obtenção da estimativa e da ccdf em qualquer posição do domínio. Ressalta-se que para a construção de ccdfs locais a partir de krigagens, adota-se o modelo gaussiano. A exceção é com relação às abordagens indicatizes que não necessitam da adoção de nenhum modelo a priori. As principais técnicas de krigagem, priorizando as que são mais utilizadas na simulação estocástica, serão apresentadas em uma seqüência que considera o aumento de informações.

### ***3.2 Krigagem da Média***

O valor médio de um conjunto de amostras pode ser calculado através da média aritmética se as amostras estiverem regularmente distribuídas, ou através da krigagem da média que considera a configuração espacial das amostras e a correlação espacial entre as mesmas. Assim a estimativa da média ( $m^*$ ) pode ser realizada através da combinação linear:

$$m^* = \sum_{i=1}^n \lambda_i Z(x_i) \quad (3.2.1)$$

Com o critério de minimização da variância do erro de estimativa da média

$$\text{Var}(m - m^*) = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(x_i - x_j)$$

e com a condição de não viés (somatória dos pesos igual a 1), é obtido o sistema de equações:

$$\begin{cases} \sum_{j=1}^n \lambda_j C(x_i - x_j) - \mu = 0 \quad \forall i = 1, 2, \dots, n \\ \sum_{j=1}^n \lambda_j = 1 \end{cases} \quad (3.2.2)$$

Resolvendo este sistema de equações, obtêm-se os ponderadores e o multiplicador de Lagrange que corresponde à própria variância de estimativa da média. Com os ponderadores, estima-se a média.

### 3.3 Krigagem Simples

Assume-se que a média  $m$  é conhecida, isto é,  $E[Z(x)]$  é a mesma em qualquer ponto  $x$  do domínio. Então, o estimador de krigagem simples (KS) é definido como:

$$Z^*(x_0) = m + \sum_{i=1}^n \lambda_i(x_0)(Z(x_i) - m) \quad (3.3.1)$$

$$Z^*(x_0) = m[1 - \sum_{i=1}^n \lambda_i(x_0)] + \sum_{i=1}^n \lambda_i(x_0)Z(x_i) = m\lambda_m + \sum_{i=1}^n \lambda_i(x_0)Z(x_i) \quad (3.3.2)$$

Na equação 3.3.2,  $\lambda_m$  corresponde ao peso que a média global  $m$  recebe na estimativa da variável na posição  $x_0$ . O estimador da krigagem simples não inclui restrições quanto aos pesos. O estimador da KS é sempre não enviesado, pois o erro da estimativa, diferença entre o valor estimado e o valor verdadeiro, é sempre nulo na média.

A variância da estimativa pode ser expressa por:

$$\sigma^2(x_0) = \sum_{i=1}^n \sum_{j=1}^n \lambda_i(x_0)\lambda_j(x_0)C(x_i - x_j) + C(x_0 - x_0) - 2 \sum_{i=1}^n \lambda_i(x_0)C(x_i - x_0) \quad (3.3.3)$$

Com o critério de minimização da variância de estimativa, é obtido o sistema de equações da krigagem simples :

$$\sum_{j=1}^n \lambda_j(x_0)C(x_i - x_j) = C(x_i - x_0) \quad \forall i = 1, 2, \dots, n \quad (3.3.4)$$

Da resolução deste sistema são obtidos os pesos ótimos da krigagem  $\lambda_i$ , e consequentemente a estimativa da variável no ponto  $x_0$  a partir da equação 3.3.1.

Com as expressões 3.3.3 e 3.3.4, pode-se obter a variância da krigagem simples:

$$\sigma_{ks}^2(x_0) = C(0) - \sum_{i=1}^n \lambda_i(x_0)C(x_i - x_0) \quad (3.3.5)$$

Como pode ser observado, os valores de  $Z(x)$  não entram na formulação, portanto a variância de krigagem é um indicador da qualidade do arranjo geométrico dos dados. Ela depende apenas da configuração dos dados e do modelo de covariância adotado.

### 3.4 Krigagem Ordinária

A krigagem ordinária (KO) é um dos mais importantes tipos de krigagem. Deseja-se determinar um valor em uma determinada posição  $x_0$  usando os  $n$  dados vizinhos  $Z(x_i)$  através de uma combinação linear com ponderadores  $\lambda_i$ :

$$Z^*(x_0) = \sum_{i=1}^n \lambda_i(x_0) Z(x_i) \quad (3.4.1)$$

Da mesma forma que na KS, deseja-se minimizar a variância do erro, só que na KO é necessário acrescentar a condição não viés, isto é, a somatória dos pesos igual a 1. O sistema é então resolvido introduzindo o multiplicador de Lagrange  $\mu$ . O procedimento fornece um sistema com  $(n+1)$  equações e  $(n+1)$  incógnitas que correspondem aos  $n$  valores dos ponderadores mais o parâmetro de Lagrange. O sistema de krigagem pode ser escrito em função da covariância (3.4.2) ou do semivariograma, pois foi incluída a condição de não viés.

$$\begin{cases} \sum_{j=1}^n \lambda_j(x_0) C(x_i - x_j) - \mu(x_0) = C(x_i - x_0) \quad \forall i = 1, 2, \dots, n \\ \sum_{j=1}^n \lambda_j(x_0) = 1 \end{cases} \quad (3.4.2)$$

O sistema de equações terá uma única solução se, e somente se, a função covariância modelada a partir dos dados for positiva definida. A variância de krigagem é escrita como:

$$\sigma_{ko}^2(x_0) = C(0) + \mu(x_0) - \sum_{i=1}^n \lambda_i(x_0) C(x_i - x_0) \quad (3.4.3)$$

Da mesma forma que na KS, a variância da KO não fornece medida absoluta de confiabilidade. Trata-se de uma medida relativa da qualidade da interpolação para as diferentes regiões, ou a qualidade da configuração geométrica dos pontos. Implicitamente dentro da KO está o cálculo da média dos dados através da KM. Remacre (1995) demonstrou a relação entre os ponderadores obtidos através da KM, KS e KO e o relativo à média na estimativa através da KS:

$$\lambda_i^{KO} = \lambda_m \lambda_i^{KM} + \lambda_i^{KS}$$

Usando vizinhança móvel, para cada posição  $x_0$ , a KO estima a média  $m(x_0)$  a partir dos dados da vizinhança. Assim sendo, a KO com vizinhança móvel já é um algoritmo não

estacionário que pode manusear não homogeneidade dos dados, sendo este o fator de sucesso e durabilidade da KO quando aplicada com vizinhança móvel.

### 3.5 Krigagem Universal

A krigagem universal (KU) difere da krigagem ordinária no modelo do componente da deriva, *trend*. Na KO o componente da deriva é limitado a uma simples constante ( $\mathbf{m}$ ). Na KU, também conhecida como krigagem com modelo de *trend* (KT), o mesmo é modelado por uma expressão analítica que toma diferentes valores no espaço.

Define-se a função aleatória  $Z(x)$  composta por um componente determinístico  $m(x)$ , a deriva, e uma função residual aleatória  $Y(x)$ .

$$Z(x) = m(x) + Y(x)$$

O componente  $Y(x)$  é modelado como uma função aleatória estacionária de segunda ordem com média zero e covariância  $C_Y(h)$ . Assim sendo, tem-se que:

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

O componente de deriva, chamado também de média local, definido como  $m(x)$  é geralmente modelado como um polinômio de baixa ordem das coordenadas do vetor  $x$ .

$$m(x) = \sum_{t=0}^T a_t f_t(x)$$

Na equação acima,  $a_t$  são coeficientes desconhecidos do modelo de deriva e são não nulos;  $f_t(x)$  são funções conhecidas das coordenadas e  $f_0(x) = 1$ . As funções  $f_t(x)$  que definem a deriva devem ser prioritariamente especificadas conforme a física do problema.

Para a krigagem é usada a combinação linear:

$$Z^*(x_0) = \sum_{i=1}^n \lambda_i(x_0) Z(x_i) \quad (3.5.1)$$

A partir da condição de não viés, pode ser obtida a relação entre os ponderadores da krigagem e as funções  $f_t(x)$  da deriva. Esta relação faz parte do sistema de equações da KU.

Desenvolvendo a expressão para a variância de estimativa e minimizando-a, é obtido o sistema da krigagem universal (KU).

$$\begin{cases} \sum_{j=1}^n \lambda_j(x_0) C_Y(x_i - x_j) - \sum_{t=0}^T \mu_t(x_0) f_t(x_i) = C_Y(x_i - x_0) \quad \forall i = 1, 2, \dots, n \\ \sum_{j=1}^n \lambda_j(x_0) f_t(x_j) = f_t(x_0) \quad \forall t = 0, 1, \dots, T \end{cases} \quad (3.5.2)$$

onde  $\mu_i$  são os  $(T+1)$  multiplicadores de Lagrange associados às  $(T+1)$  restrições dos pesos. A variância de estimativa é escrita como :

$$\sigma_{ku}^2(x_0) = C_Y(0) - \sum_{i=1}^n \lambda_i(x_0)C_Y(x_i - x_0) - \sum_{i=0}^T \mu_i(x_0)f_i(x_0) \quad (3.5.3)$$

A inferência sobre o modelo de covariância residual  $C_Y(h)$  a partir dos dados  $Z(x)$  é uma etapa crítica para os algoritmos de KU. Alguns softwares geoestatísticos, como o ISATIS, possuem um procedimento de reconhecimento de deriva onde o usuário pode escolher diversos tipos de funções.

Às vezes, se observam problemas com a utilização da KU. Por exemplo, a remoção da deriva  $m(x)$  dos dados  $Z(x)$  para obtenção dos resíduos  $Y(x)$  pode introduzir correlações espúrias e obtenção da covariância dos resíduos não representativa para o fenômeno.

Matheron (1973, apud Cressie, 1993) propôs uma abordagem para krigagem com existência de deriva através do formalismo da função aleatória intrínseca de ordem  $k$  (FAIk). Com esta abordagem é inferido um modelo de covariância generalizada que permite calcular a variância de qualquer combinação de dados  $Z(x)$  que filtram a deriva  $m(x)$ . As covariâncias generalizadas de ordem  $k$  são definidas como variâncias das diferenças de ordem  $(k+1)$  do modelo inicial da FA  $Z(x)$ . O variograma tradicional corresponde à variância das diferenças de primeira ordem e portanto está relacionado com a covariância generalizada de ordem zero. Assim a utilização do semivariograma filtra qualquer componente polinomial de ordem zero das coordenadas adicionado ao modelo de FA  $Z(x)$ , tal como  $m(x)=\text{constante}$ . Da mesma forma, a covariância generalizada de ordem  $k$  filtra os componentes do polinômio de ordem  $k$  adicionados ao modelo de FA  $Z(x)$ . Assim, a covariância generalizada pode ser usada nas equações de krigagem como se fosse uma covariância ordinária.

Deutsch e Journel (1996, p.14) reportam que “a despeito da importância teórica do formalismo das funções aleatórias intrínsecas de ordem  $k$  (FAIk)”, decidiram não incluir covariâncias generalizadas e os modelos de FAIk na GSLIB. Journel (1986a) refere-se a FAIk como um “formalismo esotérico” e reporta que a krigagem com FAIk resulta exatamente no mesmo sistema de equações da KU. Não obstante a controvérsia, a krigagem com FAIk proposta por Matheron é uma alternativa à KU com um embasamento teórico e matemático muito mais elaborado. Ressalta-se que KU foi proposta por Huijbregts e Matheron (1971, apud David,1988).

### 3.6 Krigagem Fatorial

A krigagem fatorial (KF), metodologia geoestatística desenvolvida por Matheron, permite a decomposição de uma variável regionalizada em diferentes componentes que podem ser mapeados separadamente para análise, possibilitando a eliminação de ruídos, sendo portanto utilizada como filtro. Trabalhos recentes têm retomado a utilização da KF como uma técnica geoestatística de filtragem, principalmente nas aplicações que utilizam dados sísmicos e, por isso, optou-se por descrevê-la sucintamente apesar de não ser uma técnica de krigagem propriamente dita.

Em vez de separar o modelo de FA  $Z(x)$  em uma deriva determinística  $m(x)$  mais o componente estocástico  $Y(x)$ , a KF considera um modelo com dois ou mais componentes estocásticos independentes (também chamados fatores em relação a análise fatorial).

$$Z(x) = Z_0(x) + Z_1(x) + \dots + Z_L(x)$$

A covariância de  $Z(x)$  é então a soma de  $(L+1)$  covariâncias:

$$C_z(h) = \sum_{i=0}^L C_i(h)$$

Por exemplo, as  $(L+1)$  FAs  $Z_i(x)$  podem ser modeladas a partir das  $(L+1)$  estruturas da covariância imbricada  $C_i(h)$  usadas para modelar a covariância dos dados amostrais. Assim sendo, cada estrutura do semivariograma imbricado modelado, corresponde uma variável aleatória 'fictícia'.

A krigagem da soma parcial de qualquer número de componentes  $Z_i(x)$  pode ser obtida por filtrar a contribuição da covariância dos componentes não selecionados. Por exemplo iniciando do sistema da KO, filtrando o primeiro componente, tem-se o seguinte estimador:

$$\left[ \sum_{i=i_0}^L Z_i(x_0) \right]_{KO}^* = \sum_{i=1}^n \lambda_i(x_0) Z(x_i) \quad (3.6.1)$$

Onde os pesos  $\lambda$  são obtidos pelo sistema de krigagem fatorial:

$$\begin{cases} \sum_{j=1}^n \lambda_j(x_0) C_z(x_j - x_i) + \mu_0(x_0) = \sum_{i=i_0}^L C_i(x_0 - x_i) \quad \forall i = 1, \dots, n \\ \sum_{j=1}^n \lambda_j(x_0) = 1 \end{cases} \quad (3.6.2)$$

As equações podem ser rearranjadas para realizar a krigagem isolada de cada componente (fator), o que corresponderia à filtragem de todos os demais.

### 3.7 Krigagem Com Deriva Externa

Quando duas variáveis medidas de forma diferente refletem o mesmo fenômeno e sendo a variável primária exata, porém conhecida em poucas posições, enquanto a variável secundária não pode ser medida com muita exatidão, mas pode estar disponível em um grande domínio espacial, pode-se usar a variável secundária para compor a deriva.

A krigagem com deriva externa (KDE) é uma extensão da krigagem universal. O método de deriva externa consiste em integrar no sistema de krigagem, informações de uma ou mais funções de deriva externa medidas exhaustivamente no domínio. As funções de deriva devem ser conhecidas em todas as posições do *grid*, inclusive nas posições onde se dispõe da informação da variável primária.

Admitindo a existência de apenas uma variável secundária e assumindo o modelo de deriva do tipo  $E(Z(x)) = a_0 + a_1s(x)$ , o sistema da KDE pode ser escrito como:

$$\left\{ \begin{array}{l} \sum_{j=1}^n \lambda_j(x_0)C_Y(x_i - x_j) + \mu_0(x_0) + \mu_1(x_0)s(x_i) = C_Y(x_i - x_0) \quad \forall i = 1, 2, \dots, n \\ \sum_{j=1}^n \lambda_j(x_0) = 1 \\ \sum_{j=1}^n \lambda_j(x_0)s(x_j) = s(x_0) \end{array} \right. \quad (3.7.1)$$

A variância de estimativa para a situação do sistema 3.7.1 é escrita como:

$$\sigma_{KDE}^2(x_0) = C_Y(0) - \sum_{i=1}^n \lambda_i(x_0)C_Y(x_i - x_0) - \mu_0(x_0) - \mu_1(x_0)s(x_0) \quad (3.7.2)$$

O sistema de equações 3.7.1 terá solução única se, e somente se, a função covariância modelada para os resíduos for positiva definida.

A forma da deriva é definida implicitamente em cada posição através das variáveis secundárias que devem refletir a forma da variável primária. As relações entre as variáveis secundárias e a primária devem ter sentido físico preferencialmente. Deutsch e Journel (1996, p.70) reportam que as variáveis externas devem variar suavemente no espaço para não ocorrer problemas na inversão das matrizes quando da resolução do sistema de equações que pode ficar instável.

Quando comparado com a cokrigagem, a KDE não requer inferência das covariâncias das variáveis secundárias e nem das cruzadas. O sistema de equações tem  $(n+2)$  equações no caso de uma variável secundária com modelo de deriva linear. A KDE tem como desvantagens

o fato de não capturar toda a correlação cruzada entre as variáveis como a cokrigagem e também requerer o uso da covariância dos resíduos como a KU.

### 3.8 Cokrigagem

A cokrigagem é uma extensão natural da krigagem quando variogramas e dados multivariados estão disponíveis.

O estimador da cokrigagem ordinária (COK), também conhecida apenas como cokrigagem, é uma combinação linear de pesos  $\lambda_i^\alpha$  com dados de diferentes variáveis localizados em pontos amostrais na vizinhança de um ponto  $\mathbf{x}_0$ . Cada variável é definida como um conjunto de amostras de tamanho  $n_\alpha$  e o estimador é definido como:

$$Z_{\alpha_1}^*(\mathbf{x}_0) = \sum_{\alpha=1}^N \sum_{i=1}^{n_\alpha} \lambda_i^\alpha(\mathbf{x}_0) Z_\alpha(\mathbf{x}_i) \quad (3.8.1)$$

Onde o índice  $\alpha_1$  se refere a uma variável particular do conjunto de  $n$  variáveis, para a qual será feita a estimativa.

Deseja-se estimar uma variável do conjunto de  $n$  variáveis, baseando-se que a estimativa do erro deve ser nula na média. Esta condição é satisfeita por escolher pesos cuja somatória é um para a variável de interesse e zero para as auxiliares (as demais).

$$\sum_{i=1}^{n_\alpha} \lambda_i^\alpha(\mathbf{x}_0) = \delta_{\alpha\alpha_1} = \begin{cases} 1 & \text{Se } \alpha = \alpha_1 \\ 0 & \text{Se } \alpha \neq \alpha_1 \end{cases}$$

O sistema da cokrigagem ordinária é o seguinte:

$$\begin{cases} \sum_{\beta=1}^N \sum_{j=1}^{n_\beta} \lambda_j^\beta(\mathbf{x}_0) \gamma_{\alpha\beta}(\mathbf{x}_i - \mathbf{x}_j) + \mu_\alpha(\mathbf{x}_0) = \gamma_{\alpha\alpha_1}(\mathbf{x}_i - \mathbf{x}_0) \quad \forall \alpha = 1, \dots, N; \quad \forall i = 1, \dots, n_\alpha \\ \sum_{j=1}^{n_\alpha} \lambda_j^\alpha(\mathbf{x}_0) = \delta_{\alpha\alpha_1} \quad \forall \alpha = 1, \dots, N \end{cases} \quad (3.8.2)$$

Como pode ser observado do sistema acima, a COK requer a inferência de todos os semivariogramas diretos e cruzados o que pode tornar a modelagem tediosa. Isto é a principal razão para o pouco uso da COK. A variância da cokrigagem pode ser expressa como:

$$\sigma_{\text{COK}}^2(\mathbf{x}_0) = \sum_{\alpha=1}^N \sum_{i=1}^{n_\alpha} \lambda_i^\alpha(\mathbf{x}_0) \gamma_{\alpha\alpha_1}(\mathbf{x}_i - \mathbf{x}_0) + \mu_{\alpha_1}(\mathbf{x}_0) - \gamma_{\alpha_1\alpha_1}(\mathbf{x}_0 - \mathbf{x}_0) \quad (3.8.3)$$

Com dados amostrados na mesma posição, a COK de um conjunto de variáveis tem a importante vantagem de preservar a coerência dos estimadores em relação a krigagem ordinária em separado de cada variável. Isto pode ser observado quando estima-se por exemplo a soma de várias variáveis:

$$S(x) = \sum_{\alpha=1}^N Z_{\alpha}(x) \rightarrow \begin{cases} S^{\text{COK}}(x) = \sum_{\alpha=1}^N Z_{\alpha}^{\text{COK}}(x) \\ S^{\text{KO}}(x) \neq \sum_{\alpha=1}^N Z_{\alpha}^{\text{KO}}(x) \end{cases}$$

A somatória da krigagem ordinária de cada variável isoladamente é geralmente diferente do resultado da krigagem ordinária direta das variáveis adicionadas. As duas estimativas não são coerentes. Não existe critério para se definir qual é a melhor.

A cokrigagem ordinária não tem sentido quando nenhum dado da variável de interesse está disponível em uma determinada vizinhança. Uma alternativa para esta situação é a cokrigagem simples (COKS). Sendo conhecidas as médias das variáveis, a COKS permite a estimativa de uma variável sem ter uma informação sobre a mesma na vizinhança. Semelhante à KS, a COKS não tem restrições quanto aos pesos.

### 3.8.1 Cokrigagem *Collocated*

Xu et al. (1992) denominaram cokrigagem *collocated*<sup>3</sup> a uma estratégia de definição de vizinhança, em que a vizinhança da variável auxiliar é reduzida drasticamente para um ponto: a posição onde será estimada a variável principal. Portanto esta técnica é uma simplificação da cokrigagem e basicamente consiste em reter a informação dos dados da variável secundária somente onde a variável primária será estimada. Isto reduz o esforço de cálculo no sistema de cokrigagem. Considerando apenas uma variável secundária, o estimador da cokrigagem *collocated* pode ser escrito como:

$$Z_1^*(x_0) - m_1 = \sum_{i=1}^{n_1} \lambda_i^{(1)}(x_0)[Z_1(x_i) - m_1] + \lambda^{(2)}(x_0)[Z_2(x_0) - m_2] \quad (3.8.4)$$

Onde  $m_1 = E\{Z_1(x)\}$  e  $m_2 = E\{Z_2(x)\}$  são duas médias estacionárias.

O sistema de equações da cokrigagem *collocated* é :

$$\begin{cases} \sum_{j=1}^{n_1} \lambda_j^{(1)}(x_0)C_{11}(x_j - x_i) + \lambda^{(2)}(x_0)C_{21}(x_0 - x_i) = C_{11}(x_0 - x_i) \quad \forall i = 1, \dots, n_1 \\ \sum_{j=1}^{n_1} \lambda_j^{(1)}(x_0)C_{12}(x_j - x_0) + \lambda^{(2)}(x_0)C_{22}(0) = C_{12}(0) \end{cases} \quad (3.8.5)$$

<sup>3</sup> Eventualmente, pode ser encontrada, na literatura geoestatística, tradução para cokrigagem *collocated* como cokrigagem colocada. Esta tradução não representa a idéia da cokrigagem *collocated* que é a de posicionar a variável secundária na posição da estimativa junto, ou melhor, ao lado das demais informações sobre a variável primária na vizinhança. Alternativas seriam co-localizadas ou co-locadas. De qualquer forma, optou-se por não traduzir o termo *collocated*.

onde  $C_{11}(h), C_{22}(h), C_{12}(h) = C_{21}(h)$  são as autocovariâncias e as covariâncias cruzadas das variáveis primária e secundária, respectivamente. O sistema de equações não utiliza a covariância direta da variável secundária, mas necessita da inferência e modelagem da covariância cruzada  $C_{12}(h)$ . O sistema da cokrigagem *collocated* tem  $(n+1)$  equações.

Uma aproximação para  $C_{12}(h)$  é utilizar o modelo markoviano de correionalização. Considerando a hipótese markoviana

$$E[Z_2(x)|Z_1(x), Z_1(x+h)] = E[Z_2(x)|Z_1(x)] \quad \forall Z_1(x+h)$$

e considerando que  $Z_1(x)$  e  $Z_2(x)$  são FAs gaussianas  $N(0,1)$ , Xu et al. (1992) demonstraram que a covariância cruzada pode ser adequadamente expressa como:

$$C_{12}(h) = \frac{C_{12}(0)}{C_{11}(0)} C_{11}(h) \quad \forall h \quad \text{ou} \quad \rho_{12}(h) = \rho_{12}(0)\rho_1(h) \quad \forall h \quad (3.8.6)$$

onde  $\rho_1(h) = C_1(h) / C_1(0)$  corresponde ao correlograma da variável principal e  $\rho_{12}(h) = C_{12}(h) / \sqrt{C_1(0)C_2(0)}$ , ao correlograma cruzado entre as duas variáveis.

O modelo de Markov é particularmente adequado no sentido de que fornece o modelo de covariância cruzada como uma versão modificada da covariância direta.

Com o modelo markoviano, o estimador da cokrigagem *collocated* e o sistema de equações podem ser escritos como:

$$\frac{Z_1^*(x_0) - m_1}{\sigma_1} = \sum_{i=1}^{n_1} \lambda_i^{(1)} \left[ \frac{Z_1(x_i) - m_1}{\sigma_1} \right] + \lambda^{(2)} \left[ \frac{Z_2(x_0) - m_2}{\sigma_2} \right] \quad (3.8.7)$$

$$\begin{cases} \sum_{j=1}^{n_1} \lambda_j^{(1)}(x_0) \rho_1(x_j - x_i) + \lambda^{(2)}(x_0) \rho_{12}(0) \rho_1(x_0 - x_i) = \rho_1(x_0 - x_i) \quad \forall i = 1, \dots, n_1 \\ \sum_{j=1}^{n_1} \lambda_j^{(1)}(x_0) \rho_{12}(0) \rho_1(x_j - x_0) + \lambda^{(2)}(x_0) = \rho_{12}(0) \end{cases} \quad (3.8.8)$$

onde  $\sigma_1^2 = \text{Var}\{Z_1(x)\}$  e  $\sigma_2^2 = \text{Var}\{Z_2(x)\}$  são duas variâncias estacionárias.

O algoritmo da cokrigagem *collocated* é fácil de implementar. O modelo markoviano 3.8.6 pode ser checado a partir dos dados.

Semelhante ao sistema da KDE, o sistema da cokrigagem *collocated* requer que a variável secundária seja conhecida em todos os nós onde a estimativa da variável primária será realizada. Porém, a informação extra não é incorporada como uma condição de deriva adicional, mas como pontos de dados adicionais que interagem com outras informações na função covariância.

### 3.9 Krigagem Indicatriz

A krigagem indicatriz foi proposta por Journel (1984) como uma alternativa não paramétrica para as abordagens paramétricas. A expressão ‘não paramétrica’ não significa que o modelo de função aleatória (FA) não tem parâmetros, mas que o mesmo tem mais parâmetros livres do que por exemplo o modelo gaussiano, em que as ccdfs são completamente definidas por dois parâmetros: a média e a variância.

A abordagem indicatriz se baseia na codificação da função aleatória  $Z(x)$  em uma série de indicatrizes (variável aleatória binária) definidas em diferentes parâmetros de corte (*cutoffs*<sup>4</sup>)  $z_k$ :

$$I(x; z_k) = \begin{cases} 1 & \text{se } Z(x) \leq z_k \\ 0 & \text{se } Z(x) > z_k \end{cases} \quad (3.9.1)$$

Esta codificação define uma vetor indicatriz estacionário:

$$I(x; z) = [I(x; z_1) \dots I(x; z_K)]^T \quad \text{onde } K \text{ é o número de cortes.}$$

A krigagem indicatriz (KI) fornece uma estimativa da função de distribuição acumulada condicional (ccdf) para um dado corte :

$$[I(x_0; z_k)]^* = E[I(x_0; z_k)|(n)]^* = P^*[Z(x_0) \leq z_k |(n)] \quad (3.9.2)$$

Onde  $(n)$  representa o condicionamento das informações disponíveis nas vizinhanças da posição  $x_0$ , na qual será realizada a estimativa. O processo da KI é repetido para uma série de  $K$  valores de cortes que subdividem o intervalo de variabilidade de um atributo contínuo, ou para cada uma das classes de uma variável categórica.

A ccdf construída através da união das  $K$  estimativas do tipo 3.9.2 representam um modelo probabilístico para a incerteza sobre um valor não amostrado  $Z(x_0)$ .

Para variáveis categóricas, não é necessária a prévia transformação indicatriz. Se  $Z(x)$  é uma variável contínua, então a correta seleção dos cortes  $z_k$  é uma etapa crucial: quanto mais cortes, maior o número de covariâncias a serem inferidas; quanto menos cortes, mais detalhes da distribuição serão perdidos.

Com os dados referentes às variáveis indicatrizes, pode-se usar tanto a KS como a KO para realizar a estimativa da ccdf em cada posição, relativa a cada corte. O sistema de

---

<sup>4</sup> Com intuito de simplificação da terminologia será utilizado, às vezes, o termo corte(s) para se referir aos parâmetros de corte ou *cutoffs* utilizados nas abordagens indicatrizes.

equações e o estimador da krigagem indicatriz simples são similares a da KS, só que agora assume-se ser conhecida a esperança estacionária das FAs indicatrizes  $I(x; z_k)$ :

$$E[I(x; z_k)] = P[Z(x) \leq z_k] = F(z_k)$$

A formulação mais usual envolve a utilização da KO com vizinhança móvel, o que implica em uma estimativa local das prévias cdfs  $F(z_k)$ . O estimador pode ser escrito como:

$$[I(x_0; z_k)]^* = \sum_{i=1}^n \lambda_i(x_0; z_k) I(x_i; z_k) \quad (3.9.3)$$

onde  $\lambda_i(x_0; z_k)$  são os pesos correspondentes a cada dado indicatriz relativo ao corte  $z_k$  utilizado para a estimativa em uma posição  $x_0$ . Estes pesos são obtidos pelo sistema de equações da KO rescrito como:

$$\begin{cases} \sum_{j=1}^n \lambda_j(x_0; z_k) C_I(x_j - x_i; z_k) + \mu_{z_k} = C_I(x_0 - x_i; z_k), \quad \forall i = 1, \dots, n \\ \sum_{j=1}^n \lambda_j(x_0; z_k) = 1 \end{cases} \quad (3.9.4)$$

O modelo de krigagem indicatriz ordinária, ou simplesmente krigagem indicatriz, resulta em um sistema com  $(n+1)$  equações para cada corte. A KI requer o conhecimento de  $K$  autocovariâncias correspondentes aos  $K$  cortes.

Uma outra possibilidade para a realização da KI é a krigagem indicatriz mediana (KIm) que é adequada a situações em que os variogramas ou as covariâncias indicatrizes são proporcionais entre si, isto é, os correlogramas indicatrizes são todos similares. O correspondente modelo de FA  $Z(x)$  é o denominado modelo mosaico:

$$\rho_Z(h) = \rho_I(h; z_k) = \rho_I(h; z_k, z_{k'}), \quad \forall z_k, z_{k'} \quad (3.9.5)$$

onde  $\rho_Z(h)$  e  $\rho_I(h; z_k, z_{k'})$  são os correlogramas diretos e os correlogramas indicatrizes cruzados da FA contínua  $Z(x)$ .

A krigagem indicatriz mediana (KIm) é um procedimento simples e rápido, pois requer apenas a inferência do semivariograma indicatriz mediano. Assim sendo, um simples sistema de KI deve ser resolvido, com os pesos resultantes sendo usados para todos os cortes.

Qualquer que seja a KI utilizada (simples, ordinária ou mediana), verifica-se a propriedade que a cdf estimada para a posição de um dado *hard*<sup>5</sup> é uma cdf de variância zero identificada com a classe a que pertence o dado *hard*.

---

<sup>5</sup> Neste contexto, dado *hard* se refere a uma informação conhecida com outro grau de precisão e acurácia, como por exemplo o valor de uma variável indicatriz (litologia ou classe de porosidade, permeabilidade, etc) obtido nos poços. Na classificação da página 31, representa o dado *hard* tipo 1.

A partir da obtenção das ccdfs, intervalos de confiança e estimativas podem ser derivados. Na krigagem indicatriz, a determinação da ccdf precede a estimativa. Podem ser obtidas estimativas segundo o critério que se escolher. Por exemplo, o cálculo da estimativa ‘tipo E’ que minimiza a variância de estimativa condicional é aproximado por uma soma discreta de  $(K+1)$  contribuições das classes, sendo portanto influenciado pelo número de cortes usados na discretização do intervalo:

$$Z_E^*(x_0) = \sum_{k=0}^K z'_k(x_0)[F^*(x_0, z_{k+1}|(n)) - F^*(x_0, z_k|(n))]$$

com  $z'_k(x_0) \in [z_k, z_{k+1}]$  sendo característica da  $k$ -ésima classe, geralmente estimada pela média de todos os dados  $z_i(x_0) \in [z_k, z_{k+1}]$  na vizinhança do ponto  $x_0$  ou utilizando procedimentos de interpolação como os propostos na GSLIB, seção V.1.6.

A variância da KI pode ser definida como:

$$\sigma_{KI}^2 = F^*(z_k)[1 - F^*(z_k)]$$

A relação acima é uma característica das variáveis indicatrizes. Observe que variância atinge o máximo de 0.25 quando  $F^*(z_k) = 0.5$ .

Devido à ocorrência dos pesos negativos, as relações de ordem de uma ccdf, isto é  $F_{KI}^*(x; z_{k'}|(n)) \geq F_{KI}^*(x; z_k|(n))$ ,  $\forall z_{k'} > z_k$  e  $F_{KI}^*(x; z_k|(n)) \in [0,1]$ , podem não ser obedecidas. Todas as metodologias que utilizam a abordagem indicatriz estão sujeitas à ocorrência dos problemas de relação de ordem. Deutsch e Journel (1996) reportam como a GSLIB faz as correções necessárias para estes problemas. Segundo esses autores, as correções são geralmente bastante reduzidas. De qualquer forma, os problemas com as relações de ordem constituem uma desvantagem severa da abordagem indicatriz e reavivou a pesquisa em relação à questão dos ponderadores negativos da krigagem a ser analisada alguns tópicos a frente. Uma outra dificuldade com as abordagens indicatrizes é a dificuldade de inferir modelos de covariâncias para os cortes limites.

### ***3.10 Cokrigagem Indicatriz***

A krigagem indicatriz pode ser vista como uma aproximação da coKI, onde foram ignoradas as covariâncias cruzadas. Uma alternativa à KI é considerar todas as  $K$  indicatrizes para a estimativa do valor da ccdf relativa a cada corte  $z_k$ . A implementação desta alternativa é denominada cokrigagem indicatriz e o estimador é definido como:

$$\left[ I(x_0; z_{k_0}) \right]^* = F_{\text{coKI}}^*(x_0; z_{k_0} | (n)) = \sum_{k=1}^K \sum_{i=1}^n \lambda_k(x_i; z_{k_0}) I(x_i; z_k) \quad (3.10.1)$$

Os pesos  $\lambda_k(x_i; z_{k_0})$  são obtidos no sentido de minimizar a variância de estimativa, o que corresponde a resolver o sistema de equações da cokrigagem ordinária:

$$\begin{cases} \sum_{k'=1}^K \sum_{j=1}^n \lambda_{k'}(x_j; z_{k_0}) C_I(x_i - x_j; z_k, z_{k'}) + \mu_k(x_0) = C_I(x_0 - x_i; z_{k_0}, z_{k'}) \quad \forall k = 1, \dots, K; \quad \forall i = 1, \dots, n \\ \sum_{j=1}^n \lambda_k(x_j; z_{k_0}) = \delta_{kk_0} \quad k = 1, 2, \dots, K \end{cases}$$

onde os parâmetros  $\mu_k$  são os multiplicadores de Lagrange e  $\delta_{kk_0}$  corresponde ao delta de Kronecker :

$$\delta_{kk_0} = \begin{cases} 1 & \text{Se } k = k_0 \\ 0 & \text{Se } k \neq k_0 \end{cases}$$

Assim sendo, para cada corte, o estimador da coKI requer a solução de um sistema de equações com  $K(n+1)$  equações.

A coKI faz uso de todas as distribuições bivariadas como discretizadas pelos  $K$  cortes e então requer a modelagem de  $K^2$  covariâncias indicatrizes (auto e cruzada) ou de  $K(K+1)/2$  semivariogramas (cruzados e diretos). Agrupamentos, isto é, amostragens preferenciais em altos ou baixos valores e reduzido número de dados podem tornar difícil a inferência das autocovariâncias e covariâncias cruzadas indicatrizes. Por todos esses motivos, a coKI é raramente utilizada.

Uma opção proposta por alguns autores consiste em usar um modelo de correionalização que envolve a modelagem de todos os semivariogramas como combinações lineares de um conjunto de variogramas considerados básicos.

### 3.10.1 Cokrigagem Indicatriz Reduzida

O algoritmo da krigagem indicatriz mediana pode ser considerado como a máxima aproximação do algoritmo da coKI, onde a matriz original de  $K^2$  covariâncias é reduzida a um único modelo de covariância. Entre estes dois extremos existem muitas outras aproximações possíveis que ignoram elementos específicos da matriz de covariâncias.

Goovaerts (1994) apresentou um algoritmo que ele denominou cokrigagem indicatriz dos *cutoffs* adjacentes, ou seja uma cokrigagem indicatriz reduzida. Baseando-se na hipótese de que a maior parte de informação está nas variáveis indicatrizes definidas para os *cutoffs*

mais próximos ao que está sendo considerado, Goovaerts (1994) negligencia todos os termos da matriz de covariâncias com exceção das  $K$  covariâncias diretas  $C_I(h; z_k, z_k)$  e as  $2(K-1)$  covariâncias cruzadas  $C_I(h; z_k, z_{k+1})$  e  $C_I(h; z_k, z_{k-1})$ . Obtém-se uma matriz de covariância tridiagonal onde somente os termos da diagonal e os imediatamente adjacentes a diagonal são não nulos. Então, para a estimativa em relação a um determinado *cutoff*  $z_{k_0}$ , são utilizados os dados do próprio *cutoff* e dos *cutoffs* adjacentes  $z_{k_{0-1}}$  e  $z_{k_{0+1}}$ . Para os *cutoffs* extremos  $z_0$  e  $z_K$ , apenas uma variável auxiliar é utilizada.

A estimativa através da cokrigagem indicatriz dos *cutoffs* adjacentes (coKI<sub>r</sub>) e o sistema de equações resultantes para a obtenção dos pesos que minimizem a variância de estimativa são:

$$[I(x_0; z_{k_0})]^* = F_{\text{coKI}_r}^*(x_0; z_{k_0} | (n)) = \sum_{k=k_{0-1}}^{k_{0+1}} \sum_{i=1}^n \lambda_k(x_i; z_{k_0}) I(x_i; z_k) \quad (3.10.2)$$

$$\left\{ \begin{array}{l} \sum_{k'=k_{0-1}}^{k_{0+1}} \sum_{j=1}^n \lambda_{k'}(x_j; z_{k_0}) C_I(x_i - x_j; z_k, z_{k'}) + \mu_k(x_0) = C_I(x_0 - x_i; z_k, z_{k_0}) \\ \text{para } k = (k_{0-1}, k_0, k_{0+1}); i = 1, \dots, n \\ \sum_{j=1}^n \lambda_k(x_j; z_{k_0}) = \delta_{kk_0} \text{ para } k = (k_{0-1}, k_0, k_{0+1}) \end{array} \right. \quad (3.10.3)$$

Em comparação com a completa coKI, o número de equações do sistema a ser resolvido para a coKI<sub>r</sub> é inferior  $(3n+3)$  ao da coKI, mas mesmo assim um modelo de correionalização ainda deve ser utilizado.

Teoricamente a cokrigagem indicatriz deveria ser melhor do que os outros algoritmos indicatrizes, pois ela utiliza toda a informação disponível.

### 3.11 Krigagem dos Componentes Principais Indicatrizes

A inferência direta das  $K^2$  covariâncias para a coKI é bastante problemática para a condição de muitos cortes. Uma outra alternativa é utilizar transformadas lineares das variáveis indicatrizes que são menos correlacionadas, tais como as componentes principais indicatrizes.

A análise do componente principal (ACP) é uma técnica algébrica envolvendo a transformação linear de um vetor em outro e não requer nenhuma hipótese estatística prévia sobre os dados. A interpretação estatística da ACP mostra que a covariância cruzada entre as variáveis transformadas é nula.

Adicionalmente, a primeira variável transformada ou o primeiro componente principal tem a máxima variância, o segundo componente principal, a segunda maior variância e assim por diante.

Estas propriedades fazem da ACP um caminho usual para reduzir a dimensão dos dados, selecionando um reduzido número de componentes principais, que mais contribuem para a variância. Portanto a opção por ACP implica na hipótese de que a variância é o principal aspecto da variabilidade e reter as variáveis com as maiores contribuições para a variância pode fornecer uma concisa, mas ainda satisfatória explicação da fonte da variabilidade.

Dado um vetor indicatriz  $I(x; z) = [I(x; z_1)I(x; z_2) \dots I(x; z_k)]^T$  de uma variável contínua  $Z(x)$  tais como a porosidade e a permeabilidade e dada um matriz de covariância indicatriz específica para uma dada distância, é feita a ACP desta matriz para obter o vetor dos componentes principais indicatrizes:

$$D(x) = [D_1(x)D_2(x) \dots D_k(x)]$$

$$D(x) = B^T I(x; z) \tag{3.11.1}$$

A matriz  $B^T$  é a transposta da matriz  $B$  obtida da decomposição espectral da matriz de covariâncias indicatrizes para uma distância específica  $h_0$ :

$$[C_I(h_0, z_k, z_k)] = B \Lambda B^T$$

onde  $\Lambda$  é uma matriz diagonal (autovalores) e  $B$  é a matriz ortogonal (autovetores). Cada elemento do vetor  $D(x)$  pode ser escrito como:

$$D_k(x) = \sum_{k'=1}^k b_{k',k} I(x; z_{k'}) \tag{3.11.2}$$

em que  $b_{k',k}$  e  $I(x; z_{k'})$  são os elementos da matriz  $B$  e do vetor  $I(x; z)$ , respectivamente. Assim sendo, as novas variáveis  $D_k$  são combinações lineares das indicatrizes originais com a propriedade adicional de que as covariâncias cruzadas são nulas para distância igual a da matriz de covariância indicatriz original.

Um conjunto de  $n$  vetores  $D_k(x_\alpha)$  é obtido e um estimador tipo cokrigagem dos componentes principais indicatrizes é implementado:

$$D_{k_0}^*(x_0) = \sum_{k=1}^K \sum_{i=1}^n \lambda_{k_0 k_i}(x_0) D_k(x_i) \tag{3.11.3}$$

Assumindo que a condição de ortogonalidade se mantém verdadeira para todo  $h$  (distância), o estimador pode ser reduzido a versão simples da krigagem ordinária:

$$D_{k_0}^*(x_0) = \sum_{i=1}^n \lambda_{k_{0i}}(x_0) D_{k_0}(x_i) \quad \forall k_0 = 1, 2, \dots, K \quad (3.11.4)$$

onde os pesos  $\lambda_{k_{0i}}$  são obtidos com a resolução do sistema de equações:

$$\begin{cases} \sum_{j=1}^n \lambda_{k_{0j}}(x_0) C_I(x_j - x_i; k_0) + \mu_{k_0}(x_0) = C_I(x_0 - x_i; k_0) \quad \forall i = 1, 2, \dots, n \\ \sum_{j=1}^n \lambda_{k_{0j}}(x_0) = 1 \end{cases} \quad (3.11.5)$$

onde  $C_I(x_i - x_j; k_0) = C(D_{k_0}(x_i), D_{k_0}(x_j))$ .

A transformação inversa da equação 3.11.1 fornece um modelo para a ccdf de  $Z(x)$ :

$$F_{\text{IPCK}}(x; z | \{N\}) = B D^*(x)$$

onde o vetor  $F_{\text{IPCK}}(x; z | \{N\})$  é definido como:

$$F_{\text{IPCK}}(x; z | \{n\}) = [F_{\text{IPCK}}(x; z_1 | \{n\}) \dots F_{\text{IPCK}}(x; z_K | \{n\})]$$

e o vetor  $D^*(x)$  é :

$$D^*(x) = [D_1^*(x) \dots D_K^*(x)]$$

Deutsch e Journel (1996) retiraram da segunda versão da GSLIB a discussão sobre a krigagem dos componentes principais indicatrizes (IPCK) porque, segundo eles, o algoritmo não se firmou na prática. Não obstante, vários trabalhos publicados na literatura e algumas dissertações de mestrado se basearam na IPCK. A grande motivação da IPCK não é a redução do número de indicatrizes, mas sobretudo indiretamente considerar as covariâncias cruzadas entre indicatrizes requerendo apenas a inferência de  $(K-1)$  autocovariâncias das componentes principais indicatrizes. A grande questão da abordagem IPCK é a hipótese de que a ortogonalidade se mantém para qualquer distância  $h$ . Ressalta-se que em alguns trabalhos que utilizaram a IPCK, os autores certificaram que para as situações analisadas nos mesmos, as covariâncias cruzadas entre as componentes principais indicatrizes são negligenciáveis quando comparadas com as autocovariâncias.

## 3.12 Krigagem com Dados Qualitativos

### 3.12.1 Tipos de Dados

Em muitas aplicações na indústria do petróleo, os dados *hard* tais como informações oriundas de poços. Contudo, eles podem ser suplementados por informações *soft*, às vezes disponíveis em razoável quantidade.

De uma maneira geral, quando estimando um atributo sobre um determinado campo, as informações disponíveis podem ser classificadas em três categorias:

- Tipo 1- Dados *hard* que são valores conhecidos do atributo em determinadas posições.  $Z(x_i) = z_i$  para  $i = 1, 2, \dots, n$
- Tipo 2- Dados *hard* tipo desigualdade ou inequações, isto é, o valor de um atributo em determinadas posições é maior ou menor que um valor limite ou está dentro de um intervalo.  $Z(x_i) \in (a_i, b_i)$   $i = n + 1, \dots, n + n' = N$
- Tipo 3- Informações qualitativas *soft*, ou de natureza global tais como a interpretação geológica sobre continuidade espacial, ou de natureza local como interpretação de dobramentos, falhas, limites da acumulação. Ou informações de que um determinado valor  $Z(x_i)$  está provavelmente em um determinado intervalo, semelhante ao dado *hard* tipo 2.

O desafio consiste em obter a estimativa que honre todas as informações disponíveis, *soft* ou *hard*, de natureza global ou local.

Em todas as aplicações, os valores interpolados  $Z^*(x)$  devem obedecer alguns limites, geralmente desigualdades do tipo  $Z^*(x) \geq 0$ , por exemplo para porosidade, permeabilidade, espessura porosa. Alguns atributos também têm limite superior, como é o caso da porosidade.

Um dado *hard*  $Z(x_i) = z_i$  pode ser interpretado como intervalo com largura zero, ou melhor, o limite inferior igual ao superior. Em verdade, sempre haverá uma largura mínima se for incluído o erro na obtenção da medida. Um dado limitado inferiormente  $Z(x_i) > a_i$  é na prática um intervalo com limites  $Z(x_i) \in (a_i, b)$ . Similarmente, um dado limitado superiormente  $Z(x_i) < b_i$  é na prática um intervalo com limites  $Z(x_i) \in (a, b_i)$ .

Pode-se considerar que sempre existe um intervalo limitado para todas informações em qualquer aplicação. Desta forma, a única diferença entre um dado tipo 1 (*hard*) ou tipo 2 (*hard* com desigualdade) é a largura do intervalo.

$$Z(x_i) \in [a(x_i), b(x_i)]$$

As técnicas usuais de krigagem (KO, KU) não permitem levar em consideração dados tipo 2 e 3. Os resultados obtidos a partir das mesmas podem não obedecer aos limites.

### 3.12.2 Abordagens Alternativas

Várias abordagens que não se baseiam simplesmente na positividade dos pesos foram propostas no sentido de garantir a restrição de não negatividade de um atributo, bem como incorporar informações *soft*.

Baseando-se nos algoritmos da programação quadrática e através do método das *splines*, Dubrule e Kostov (1986) propuseram uma solução para o problema mais geral de estimar uma superfície  $\{Z^*(x), x \in A\}$  honrando todos os dados tipo 1 e tipo 2, isto é, dados exatos e do tipo inequações. Eles demonstraram o paralelo entre *splines* e a krigagem.

Barnes e You (1992) propuseram uma metodologia bastante simplificada para realizar a estimativa com dados tipo 2. Partindo das equações da KO, eles adotaram mais dois tipos de equações condicionantes: uma para os limites superiores e outra para os limites inferiores. Assim sendo, no sistema da KO, apareceram mais dois tipos de multiplicadores de Lagrange associados com os limites inferiores e superiores. Caso o valor estimado fique dentro do intervalo permitido, nada é alterado, isto é, a estimativa e a variância de krigagem são iguais a da KO sem a informação do intervalo. Porém, se o valor estimado através da KO está fora do intervalo permitido, a estimativa fica igual ao limite que estiver mais próximo. Isto é garantido através da introdução dos multiplicadores específicos para os limites. Para esta última condição, foram obtidas expressões para determinar as alterações na variância de krigagem que consiste em acrescentar um termo de correção positivo.

Estas e algumas outras abordagens apresentam várias limitações comuns:

- 1- A função estrutural usada,  $C(Z(x), Z(x+h))$  é inferida somente dos dados exatos. Portanto, não extraem informações estruturais fornecidas pelos dados tipo inequações ou *soft*;
- 2- Não permitem o uso de informação prévia sobre o valor mais provável de um valor dentro do intervalo limitante  $[a_i, b_i]$ ;
- 3- Não fornecem intervalos de confiança para os resultados das estimativas.

### 3.12.3 Krigagem *Soft* : Modelo de Markov-Bayes

Journel (1986b) propôs uma metodologia que permite colocar os dados tipo 3 (*soft*) em um intervalo limitado *hard* do tipo  $Z(x_i) \in [a(x_i), b(x_i)]$  com uma distribuição provável do valor do atributo. O formalismo da indicatriz fornece uma solução direta para codificar dados tipo inequações.

A função indicatriz  $I(x_i; z)$  relacionada a um dado *hard*  $Z(x_i) = z_i$  pode ser vista como uma cdf prévia do parâmetro  $Z_i$  com variância zero (figura 3-1.a). Do mesmo modo,

em uma posição  $x_i$  com um dado tipo 2, isto é, intervalo com limites  $Z(x_i) \in (a_i, b_i)$ , a função indicatriz é conhecida apenas fora dos limites do intervalo onde ela ou é 0 ou 1 (3-1.b).

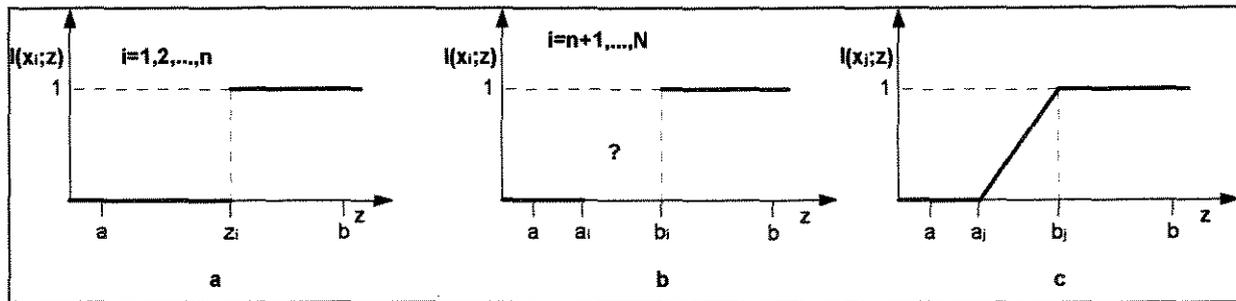


Figura 3-1: Informações codificadas como cdfs. Os valores 'a' e 'b' são os limites extremos dos dados.

Se além do intervalo com limites  $Z(x_i) \in (a_i, b_i)$ , algumas informações prévias permitem quantificar a probabilidade do valor  $Z(x_i)$  dentro do intervalo, a correspondente função indicatriz pode ser completada com dados subjetivos de probabilidade acumulada (3-2.c). A única diferença entre dados indicatrizes oriundos de dados exatos ou do tipo inequações e a informação subjetiva é que os primeiros podem ser apenas 0 e 1, enquanto o último pode assumir valores entre 0 e 1.

Os modelos de variogramas indicatrizes

$$2\gamma_I(h; z, z') = E[[I(x; z) - I(x + h; z)][I(x; z') - I(x + h; z')]]$$

requeridos para resolver os vários sistemas de krigagens indicatrizes podem ser inferidos usando, não apenas os dados indicatrizes resultantes dos dados tipo 1, mas também os dados indicatrizes oriundos de dados tipos 2 e 3.

Em muitas aplicações, importantes informações residem nos dados tipo inequações (tipo 2) ou qualitativos (tipo 3). Dados tipo inequações são informações objetivas e a não utilização dos mesmos pode levar a resultados não acurados ou a estimativas inaceitáveis.

O formalismo da krigagem *soft* permite construir um estimador que honra todas as informações prévias, por exemplo dados tipo 2 ou 3. Na krigagem *soft* introduzida por Journel (1986b) e Kostov e Journel (1986) poder-se-ia usar tanto a coKI como a KI.

Baseando-se na metodologia da krigagem *soft* de Journel (1986b), Journel e Zhu (1990) apresentaram uma alternativa para a cokrigagem denominada pelos autores inicialmente como modelo de *Markov-Bayes*.

A partir de dados oriundos de calibração entre a variável primária  $Z(x)$  e uma secundária  $V(x)$ , esses autores fizeram uma codificação bayesiana da informação *soft*.

Em uma determinada posição onde o atributo  $Z$  não é amostrado, mas o valor da variável secundária  $V$  é disponível, a cdf condicional é extraída da calibração:

$$P[Z(x_0) \leq z | V(x_0) \in [v_j, v_{j+1}]] \quad (3.12.1)$$

Esta cdf representa a informação prévia disponível sobre o valor não amostrado de  $Z(x_0)$ . Em um arcabouço bayesiano, todos os dados (*hard* ou *soft*) devem ser colocados na forma de cdfs prévias locais, como mostrado anteriormente (figura 3-1). Journel e Zhu (1990) denominaram esta cdf local prévia como uma variável  $y(x_i; z)$ :

$$y(x_i; z) = P[Z(x_i) \leq z | (\ell x_i)] \quad (3.12.2)$$

onde  $(\ell x_i)$  representa a informação local em  $x_i$ , isto é, um dado *hard* ou *soft*.

A partir da calibração da informação de uma variável secundária, pode ser obtida uma cdf prévia do tipo  $y(x_1; z) = P[Z(x_1) \leq z | v(x_1)] \in [0,1]$ . Igualmente, qualquer informação resultado de uma interpretação ou de experiência sobre a área em estudo pode ser codificada como uma cdf prévia.

A cdf prévia local  $y(x_i; z)$  toma a forma de um vetor de valores de probabilidade, um vetor para cada posição com  $K$  membros correspondentes aos  $K$  cortes que foram usados para discretizar a variabilidade de  $Z$  formando a VA indicatriz  $I(x_i; z)$  referente aos dados *hard*. Para fornecer um modelo de incerteza em uma posição  $x_0$  qualquer, as cdfs locais prévias  $y(x_0; z)$  deverão ser atualizadas utilizando todas as informações da vizinhança.

Esta informação a priori será atualizada, por algumas outras informações  $Z$  e/ou  $V$  na vizinhança, para compor uma cdf posterior e foi nomeada com uma variável  $y(x_0; z)$ :

$$y(x_0; z) = P[Z(x_0) \leq z | (n)] \in [0,1] \quad (3.12.3)$$

Qualquer cdf local prévia  $y(x_i; z)$  pode ser interpretada como realizações de uma variável aleatória  $Y(x; z)$ , com suas específicas covariâncias diretas e cruzadas com a variável  $I(x; z)$ . O algoritmo da cokrigagem simples pode ser utilizado para atualizar o valor de  $y$  a partir dos valores de seus vizinhos:

$$P^*[Z(x_0) | (n)] - F(z) = \sum_{i=1}^{n_1} \lambda_i(x_0; z)(I(x_0; z) - F(z)) + \sum_{j=1}^{n_2} \xi_j(x_0; z)(y(x_j; z) - E[Y(z)]) \quad (3.12.4)$$

onde  $F(z)$  é a cdf estacionária global prévia,  $E[Y(z)]$  é a esperança estacionária das variáveis *soft*,  $I(x; z)$  são as cdfs prévias geradas pelos dados *hard* em  $n_1$  locações, e  $y(x; z)$  são as cdfs prévias geradas pelas informações *soft* em outras  $n_2$  posições.

Contudo esta abordagem implicaria na modelagem de todas as covariâncias diretas e cruzadas das VAs  $I(x; z)$  e  $Y(x; z)$ , o que é impraticável.

Felizmente sob a hipótese de uma aproximação Markoviana (daí o nome Markov-Bayes) que estabelece que a informação *hard*  $I(x; z)$  sempre prevalece sobre uma informação *soft collocated*  $Y(x; z)$ , consideráveis simplificações podem ser feitas e um modelo de correionalização obtido.

As seguintes relações podem ser obtidas:

- $E[Y(x; z)] = E[P[Z(x) \leq z | v]] = F(z) = E[I(x; z)]$
- $\text{Var}[Y(x; z)] = B(z)\text{Var}[I(x; z)]$
- $C_{Y'}(h; z) = B(z)C_I(h; z)$
- $C_Y(h; z) = B^2(z)C_I(h; z)$

$$\text{Onde } \begin{cases} B(z) = m^{(1)}(z) - m^{(0)}(z) & B(z) \in [-1, 1] \\ m^{(0)}(z) = E[Y(x; z) | I(x; z) = 0] & m^{(0)}(z) \in [0, 1] \\ m^{(1)}(z) = E[Y(x; z) | I(x; z) = 1] & m^{(1)}(z) \in [0, 1] \end{cases}$$

A covariância dos dados *soft* e a covariância cruzada entre dados *hard* e *soft* são deduzidas diretamente do modelo de covariância indicatriz facilitando o sistema de cokrigagem para o estimador da equação 3.12.4. Assim sendo, com a utilização do modelo de Markov-Bayes somente a inferência da covariância indicatriz é requerida, equivalente ao que é requerido pela krigagem simples.

### 3.13 A Questão dos Ponderadores Negativos

A obtenção de valores negativos para a estimativa de um atributo que é necessariamente positivo, como por exemplo porosidade e permeabilidade, suscitou vários estudos quase sempre endereçados no sentido de garantir a positividade da estimativa através da não negatividade dos ponderadores.

Sendo todos os dados positivos, um valor estimado somente pode ser negativo se um ou mais ponderadores da krigagem são negativos. Assim, os primeiros estudos foram no sentido de garantir que todos os ponderadores fossem positivos, e se basearam nas soluções de problemas de otimização com restrições. Por exemplo a resolução de problemas do tipo minimização de uma forma quadrática (variância de estimativa) sujeita a restrições (somatória dos ponderadores igual a um e todos os ponderadores positivos) que pode ser obtida através da programação quadrática. Com esta abordagem destacam-se os trabalhos de Barnes e Johnson (1984) e Limic e Mikelic (1984).

Barnes e Johnson (1984) apresentaram a metodologia que denominaram krigagem positiva com as seguintes considerações:

*Tentar justificar a existência de pesos negativos através do fato de que a krigagem é o melhor estimador linear não enviesado é um argumento vago, um dogma. A única justificativa da filosofia geoestatística é o fato de que ela é excelente para modelar a realidade. Quando a teoria diverge da realidade, é a teoria que deve ser modificada.*

A abordagem proposta por estes autores suscitou debates e controvérsias. Journel (1986b) foi um dos opositores das metodologias que garantiam a não negatividade dos ponderadores. Segundo Journel (1986b), o fato de que os pesos da krigagem possam ser negativos é extremamente positivo, porque garante a possibilidade de uma estimativa não convexa, saindo fora dos limites dos dados.

Para Journel (1986b), o cerne do problema não está nos pesos negativos, mas no fato de que o sistema de krigagem não é informado sobre a limitação que  $Z(x) \geq 0$ , para todo  $x$  no domínio. Enfim, a solução não consiste em forçar que todos os pesos sejam positivos, mas sim forçar que todos os valores estimados sejam positivos ou, mais genericamente, que  $Z^*(x) \in (a, b)$ . As abordagens que não se baseiam na positividade dos ponderadores foram descritas no tópico anterior, referente a krigagens com dados qualitativos.

Mesmo com a polêmica sobre a não negatividade dos ponderadores, abordagens continuaram a ser propostas neste sentido.

Szidarovsky et al. (1987) apresentaram uma metodologia que realiza a krigagem em todos os subconjuntos de amostras possíveis do conjunto total disponível e, a partir desses resultados, são selecionadas as soluções que resultam em positividade de todos os ponderadores. Destas soluções, é escolhida a que apresenta a menor variância de estimativa, que certamente é maior que a mínima variância de krigagem. Como reportado por Szidarovsky et al. (1987), a grande desvantagem da metodologia é a necessidade da resolução de  $(2^n - 1)$  sistemas de krigagem para cada estimativa a ser realizada. Os autores desta proposta não apresentaram nenhuma aplicação prática da metodologia.

Com o aparecimento e desenvolvimento da proposta da krigagem das indicatrizes, novas pesquisas sobre a geração dos pesos negativos têm sido conduzidas. Isto se deve ao fato de que os pesos negativos prejudicam eventual interpretação dos mesmos como percentuais e no caso da krigagens indicatrizes provoca a ocorrência dos problemas de relação de ordem.

Observa-se recentemente uma retomada da consideração da não negatividade dos ponderadores. No último congresso de geoestatística, (Bourgault, 1997) relata que:

*Infelizmente, um problema com a krigagem é a ocorrência dos pesos negativos. Krigagem é um estimador não convexo porque pode gerar pesos negativos e produzir valores fora dos limites dos dados. Pior, com a krigagem é possível gerar estimativas negativas mesmo que todos os dados sejam estritamente positivos.*

A história dos ponderadores se faz assim cíclica, pois novas propostas estão sendo feitas, baseando-se nas idéias colocadas no início da década de 80.

Por exemplo, Froideveaux (1993) propôs uma metodologia a que denominou krigagem restrita, na qual os ponderadores são forçados a serem estritamente positivos a fim de possibilitar a obtenção da função de distribuição acumulada condicional (ccdf) local. A metodologia consiste em zerar todos os pesos negativos e renormalizar os pesos restantes de forma que a somatória dos mesmos seja igual a 1. Como os pesos são todos positivos, os mesmos podem ser encarados como percentuais e podem ser utilizados na construção das ccdfs locais:

$$F(x_0, z_\alpha) = \sum_{j=1}^{\alpha} \lambda'_j$$

onde  $F(z_\alpha, x_0)$  é a probabilidade de que na posição  $x_0$ , o valor real seja menor ou igual a  $z_\alpha$ , e  $\lambda'_j$  correspondem aos ponderadores relativos a dados amostrais com valores inferiores ou iguais a  $z_\alpha$ . Esses ponderadores são obtidos através da krigagem ordinária com a correção de pesos proposta por Froideveaux.

Remacre (1997) propôs uma metodologia que basicamente faz um recálculo dos pesos obtidos com a metodologia de Froideveaux de forma a privilegiar as amostras que estão mais próximas do ponto a ser estimado.

Bourgault (1997) apresentou uma metodologia que gera ponderadores convexos, mantendo a condição da somatória igual a 1, como alternativa para krigagem. Os pesos convexos levam em consideração a redundância dos dados originais e o efeito dos agrupamentos estatísticos, isto é, dados com distâncias inferiores à amplitude prática que caracteriza a continuidade espacial do conjunto de dados. A abordagem de Bourgault apresenta um parâmetro que deve ser ajustado iterativamente, e portanto demanda maior tempo de CPU, para se obter a média desejada para o conjunto de dados. Como não se sabe, a

priori, o valor da média, o ajuste deste parâmetro pode resultar em estimativas errôneas e enviesadas.

Journel e Rao (1996) propuseram uma metodologia em que os pesos da krigagem são corrigidos de forma a serem todos positivos. Em vez de zerar todos os pesos negativos, uma correção tipo translação foi proposta por Journel e Rao (1996), na qual o módulo do maior peso negativo é adicionado a todos os pesos da krigagem e em seguida é feito um reescalonamento de forma a manter a somatória dos pesos igual a 1. Esta correção segundo Journel e Rao (1996) representa a diferença entre a média aritmética dos dados locais e a média estacionária usada na KS ou implicitamente calculada na KO.

A diferença entre as metodologias de Froideveaux (1993) e a de Journel e Rao (1996) é que a primeira resulta em ponderadores que valorizam proporcionalmente mais as informações que estão mais próximas do ponto a ser estimado em relação à segunda abordagem.

Uma outra metodologia foi proposta por Deutsch (1996) que basicamente anula todos os pesos negativos e alguns pesos positivos pequenos relacionados aos negativos, e em seguida renormaliza os demais pesos no sentido de manter a condição de não viés.

É importante ressaltar que em todas as abordagens que asseguram a não negatividade dos pesos e a estimativa convexa, a propriedade perdida é a mínima variância de krigagem. Uma característica muito importante para essas metodologias é que as mesmas devem ter praticidade, isto é, não devem ser computacionalmente intensivas. Ressalta-se que as recentes propostas que consideram a positividade dos ponderadores objetivam mais a determinação da ccdf no âmbito das indicatrizes, do que propor alternativa à krigagem tradicional. Estimativas que honrem informações sobre os limites de ocorrência do atributo podem ser obtidas de forma a não exigir a positividade dos ponderadores, como foi reportado no tópico anterior.

## 4. Incertezas na Estimativa

Implícito em qualquer estudo geoestatístico está a escolha de um modelo. No início da exploração e desenvolvimento de um reservatório, poucos poços são disponíveis. Assim, na modelagem geoestatística, seja utilizando simulação estocástica ou krigagem, são necessárias informações adicionais (por exemplo oriundas da sísmica) para a construção do modelo. Por exemplo, extrair a forma da deriva do atributo de interesse correlacionado com um atributo sísmico qualquer. Nestas situações, a participação de uma variável secundária auxiliar é decisiva na construção do modelo. À medida que novos poços são perfurados, os mesmos podem ser utilizados na construção de um novo modelo, se necessário interagindo com a variável secundária.

A crescente utilização dos algoritmos de simulação estocástica pode levar os mais afoitos a considerarem a krigagem como ferramenta geoestatística ultrapassada. É importante frisar que a krigagem tem seu campo de aplicação a depender dos objetivos do estudo, e ademais, em grande parte dos algoritmos de simulação estocástica, ela é implicitamente realizada. Assim, todas as incertezas que se têm na estimativa de um atributo em uma determinada posição, ocorrerão também nas krigagens implicitamente realizadas nas simulações estocásticas e, portanto, também as influenciará. Por exemplo, nos algoritmos de simulação com abordagem seqüencial, as cdfs são obtidas via krigagens.

Nos textos sobre incerteza na estimativa geoestatística, o primeiro parâmetro que geralmente aparece é a variância de krigagem e, conseqüentemente, os intervalos de confiança derivados da mesma. Contudo, a incerteza na estimativa é muito mais do que isso. Têm-se incertezas no primeiro e segundo momentos oriundos das amostras disponíveis. Várias definições apresentam caráter subjetivo, como por exemplo a escolha do modelo de semivariograma e seus parâmetros (efeito pepita, amplitude prática, patamar) que dependem do entendimento do fenômeno físico e da relação com os parâmetros da modelagem.

Em todo o desenvolvimento geoestatístico está implícita as propriedades de estacionariedade e ergodicidade do modelo probabilístico. Às vezes, a função covariância ou o semivariograma não podem ser completamente determinados, principalmente quando se tem

apenas poucos dados disponíveis. Tem-se o problema de definição de deriva e das anisotropias quando elas ocorrem.

Starks e Fang (1982) analisaram o efeito da deriva no semivariograma experimental. Segundo estes autores, é possível que ocorra uma má interpretação da deriva no semivariograma experimental e, portanto, se consiga ajustar um outro modelo sem deriva.

Englund (1990), em um trabalho intitulado "*A Variance of Geostatisticians*", fez uma interessante pesquisa analisando o caráter subjetivo da escolha de uma determinada abordagem de estimativa e da definição dos parâmetros da mesma, como por exemplo a interpretação da estrutura de correlação espacial. Conjuntos de dados idênticos foram enviados para 12 pesquisadores, que independentemente analisaram os dados e produziram interpolações, não sendo especificado, entretanto, o uso da krigagem.

A maioria dos pesquisadores (dez) consultados por Englund usaram a krigagem. Foram usadas krigagem ordinária, disjuntiva e indicatriz. Quanto à modelagem dos variogramas também foram observadas grandes diferenças: modelos de variogramas esféricos, exponenciais e gaussianos, com e sem efeito pepita, com e sem anisotropias. Foram observadas diferenças nos parâmetros de vizinhança utilizados pelos diversos pesquisadores. Em face das diferenças de como a estimativa foi realizada, os resultados variaram consideravelmente.

#### ***4.1 Escolha do Tipo de Krigagem***

Desde a apresentação inicial da krigagem por Matheron, muitos algoritmos foram desenvolvidos no sentido de obter uma estimativa ótima face a uma diversidade de conjuntos de dados e também de obter medidas de incertezas em relação à estimativa.

O avanço da geoestatística para outros campos além da mineração, como por exemplo para a área de petróleo, propiciou o desenvolvimento de alternativas para trabalhar com dados *hard* escassos, incorporando também as informações qualitativas, a experiência do profissional da área, etc. Os novos tipos de krigagem também incorporam a necessidade de se extrair também medidas de incerteza.

De posse de um conjunto de dados, qual algoritmo de krigagem utilizar? Dependendo do tipo de dado (contínuo ou discreto), da quantidade, da existência de mais de uma variável correlacionada, do objetivo da estimativa, da existência ou não de deriva, a resposta a essa questão pode mudar significativamente.

É essencial que o usuário de qualquer tipo de krigagem saiba quais são as características do modelo de FA, sobre quais hipóteses ela se fundamenta e como elas influenciam o resultado, ou mesmo, o que elas representam.

Na tabela 4.1 está um resumo comparativo dos diversos tipos de krigagem com suas principais hipóteses, características e a forma de obter também medidas de incerteza.

## 4.2 Medidas de Incerteza

Qualquer que seja o método de estimativa escolhido, sempre se terá um erro associado. Não se pode calcular este erro exatamente. Incerteza é, de certa forma, alguma indicação da magnitude desse erro.

A incerteza associada a um valor não amostrado, em uma determinada posição  $\mathbf{x}$ , pode ser calculada através da ccdf (função de distribuição acumulada condicional) obtida diretamente, por exemplo, das krigagens indicatrizes. Para as demais krigagens, a obtenção de uma medida de incerteza requer a adoção de um modelo de distribuição dos erros de estimativa.

A probabilidade do valor verdadeiro de um atributo  $Z$  ser menor do que um determinado valor limite  $z$ , condicionado aos  $n$  dados disponíveis, é igual a:

$$P[Z(\mathbf{x}) \leq z | (\mathbf{n})] = F(\mathbf{x}; z | (\mathbf{n}))$$

onde  $F(\mathbf{x}; z | (\mathbf{n}))$  é a ccdf, e o símbolo  $|(\mathbf{n})$  indica o condicionamento para os  $n$  dados disponíveis. Com essa ccdf, várias medidas de incerteza podem ser derivadas:

- Intervalos de Probabilidade

$$P[Z(\mathbf{x}) \in ]a, b] | (\mathbf{n})] = F(\mathbf{x}; b | (\mathbf{n})) - F(\mathbf{x}; a | (\mathbf{n})) .$$

O mesmo pode ser feito para a probabilidade do valor verdadeiro exceder a um determinado valor limite:  $P[Z(\mathbf{x}) > b | (\mathbf{n})] = 1 - F(\mathbf{x}; b | (\mathbf{n}))$

- Mapas de Quantis

O p-quantil de uma ccdf de  $Z(\mathbf{x})$  é o valor limite  $a_p$  tal que:

$$F(\mathbf{x}; a_p | (\mathbf{n})) = p \in [0,1], \text{ então } a_p(\mathbf{x}) = F^{-1}(\mathbf{x}; p | (\mathbf{n})) .$$

O p-quantil de uma ccdf representa um valor para o qual se tem a probabilidade  $p$  do verdadeiro valor ser menor ou igual a ele, e portanto a probabilidade  $(1-p)$  do valor real excedê-lo.

Tabela 4-1: Tipos de krigagem, hipóteses e características.

Krigagem	Hipóteses	Variável	Características
Simple	Média global conhecida. Estacionariedade de 2ª Ordem.	Primária contínua.	Construção de cdf requer adoção do modelo gaussiano (Variância é fator geométrico). Simplicidade. Só usa covariância.
Ordinária	Hipótese intrínseca. Quase-estacionariedade	Primária contínua.	Construção de cdf requer adoção do modelo gaussiano. Simplicidade. Pode usar variograma ou covariância. Não exige conhecimento da média.
Universal	Quase-estacionariedade do componente do resíduo.	Primária contínua.	Construção de cdf requer adoção do modelo gaussiano. Não exige estacionariedade de 2ª Ordem. O variograma do resíduo pode introduzir viés.
Deriva Externa	Variáveis primária e secundária correlacionadas. Quase-estacionariedade do componente do resíduo.	Primária e secundária contínuas.	Construção de cdf requer adoção do modelo gaussiano. Não requer covariância da variável secundária e nem covariância cruzada. Variável secundária conhecida em todo o domínio. Utiliza o variograma do resíduo.
Krigagem Indicatriz	Quase-estacionariedade. Desprezível as correlações cruzadas entre indicatrizes.	Discreta ou contínua discretizada.	Fornecer cdf. Só utiliza as autocovariâncias. Problemas de relação de ordem. Problemas para processos tipo difusão.
Cokrigagem	Variável secundária correlacionada com a primária. Quase-estacionariedade.	Primária e secundária contínuas.	Construção de cdf requer adoção do modelo gaussiano. Dados isotópicos - preserva coerência dos estimadores. Requer autocovariância da variável secundária e a cruzada.
Cokrigagem <i>Collocated</i>	Variável secundária correlacionada com a primária. Quase-estacionariedade.	Primária e secundária contínuas.	Construção de cdf requer adoção do modelo gaussiano. Não usa covariância direta da variável secundária. Ignora informação dos dados secundários próximos ao nó onde está sendo realizada a estimativa. Variável secundária conhecida em todo o domínio.
Cokrigagem Indicatriz	Quase-estacionariedade.	Discreta ou contínua discretizada.	Fornecer cdf. Faz uso de toda a informação disponível. Problemas de relação de ordem. Necessita das covariâncias cruzadas entre as indicatrizes.
Cokrigagem Indicatriz Reduzida	Quase-estacionariedade. A maior parte da informação está nos <i>cutoffs</i> adjacentes.	Discreta ou contínua discretizada.	Fornecer cdf. Exige apenas as autocovariâncias e as cruzadas com os <i>cutoffs</i> mais próximos. Problemas de relação de ordem.
IPCK	Quase-estacionariedade. Variância é o principal aspecto da variabilidade.	Discreta ou contínua discretizada.	Fornecer cdf. Simplicidade. Problemas de relação de ordem. Transformada lineares das variáveis indicatrizes gerando novas variáveis.
Soft: Markov-Bayes	Markoviana - Os dados <i>hard</i> prevalecem sobre os dados <i>soft</i> . Quase-estacionariedade.	Discreta ou contínua discretizada.	Fornecer cdf. Incorporação de dados <i>soft</i> e informações qualitativas. Utilização de parâmetros de calibração. Exige covariância apenas das indicatrizes dos dados <i>hard</i> . Problemas de relação ordem.

O mesmo pode ser feito para a probabilidade do valor real exceder a um determinado valor limite:

$$P[Z(x) \leq a_p | (n)] = p \quad \text{e} \quad P[Z(x) > a_p | (n)] = 1 - p$$

Pode-se construir mapas de p-quantis, por exemplo, mapas em que os valores estimados em cada posição têm uma determinada probabilidade de serem excedidos.

A qualidade de um modelo de ccdf, ou seja, da cdf posterior depende da adequabilidade do modelo de dependência espacial utilizado e do percentual da informação disponível que foi utilizado na obtenção da mesma.

O modo mais tradicional de construir um modelo de incerteza  $F(x; z|(n))$  é assumir uma forma ou expressão analítica completamente especificada por um número limitado de parâmetros. Por exemplo, se  $F(x; z|(n))$  é assumida ser gaussiana, ela é determinada por dois parâmetros: sua média e sua variância. O problema de determinar os modelos de ccdfs se reduz a determinar os poucos parâmetros correspondentes.

As abordagens não paramétricas não assumem um modelo de expressão analítica para as posteriores cdfs. A obtenção das ccdfs precedem a obtenção de uma estimativa ótima, como reportado nas krigagens indicatrizes.

### ***4.3 Intervalos de Incerteza com Krigagens Paramétricas***

A ccdf gaussiana  $F(x; z|(n))$  é completamente determinada por um sistema de krigagem que requer apenas a inferência do modelo de covariância ou semivariograma.

A variância de krigagem depende apenas da configuração dos dados e do modelo de covariância adotado, não dependendo do valor do atributo na posição. Conseqüentemente, a dispersão da ccdf em volta da média, a estimativa de krigagem, é também independente da mesma. Esta propriedade de independência da variância em relação aos valores dos dados é chamada homocedasticia. Esta propriedade é freqüentemente inadequada.

A justificativa do modelo de distribuição gaussiana dos erros provém dos teoremas gerais da teoria das probabilidades, em especial do Teorema Central do Limite. Neste teorema tem-se que a soma de um grande número de variáveis independentes, com mesma distribuição (não necessariamente gaussiana) tende para uma distribuição gaussiana. Contudo, quando lidando com erros de interpolação, onde as fontes de erros não são necessariamente aditivas e independentes, não há razão para se acreditar que a distribuição dos erros seja gaussiana. Mas,

sem dúvida, o grande apelo desta abordagem é a extrema facilidade para se obter os intervalos de incerteza.

Contudo, alguns cuidados devem ser tomados antes de adotar esta abordagem. Para serem obtidos os intervalos de incerteza, além da hipótese de distribuição gaussiana dos erros, se faz necessária a hipótese de que a variância de krigagem seja uma estimativa precisa da variância dos erros reais. Ao fazer essas hipóteses, pode-se combinar a estimativa  $Z^*$  com a variância de estimativa  $\sigma_k^2$  para produzir intervalos de confiança e limites de confiabilidade.

Por exemplo:

$$1- P[Z^*(x) - 1.96\sigma_k < Z(x) < Z^*(x) + 1.96\sigma_k] = 0.95$$

$$2- P[Z(x) > Z^*(x) + 0.675\sigma_k] = 0.25$$

Com estes intervalos de confiança poderiam ser construídos os mapas de quantis ou intervalos de probabilidade.

Segundo Isaaks e Srivastava (1989), distribuições globais de erros, mesmo para distribuições assimétricas de dados, tendem a ser simétricas. Isto não significa contudo, que elas são necessariamente bem modeladas por uma distribuição gaussiana. Ao adotarmos a hipótese de distribuição normal, é importante considerar a segunda hipótese, isto é, capacidade de estimar a variância dos erros reais. Contudo, a  $\sigma_k^2$  só será uma boa estimativa se o modelo de variograma estiver bem ajustado aos dados experimentais. Por exemplo, um pequeno erro no patamar do variograma não afeta os pesos da krigagem, mas afeta a variância da krigagem. Assim, a adoção da  $\sigma_k^2$  para a estimativa dos intervalos de confiança requer, em parte, que o patamar do variograma seja uma estimativa precisa da variância global do atributo estimado.

Embora as distribuições globais do erro possam ser simétricas, as distribuições locais dos erros não são. Nas áreas de baixos valores do atributo, há uma grande probabilidade de superestimação e, nas áreas de altos valores, uma correspondente tendência de subestimação. Particularmente, nas áreas com valores extremos que geralmente têm grande importância, as distribuições locais dos erros são freqüentemente assimétricas. A nível local, a hipótese de normalidade é ainda mais questionável e isto geralmente inviabiliza a determinação de intervalos de confiança locais, principalmente nas regiões onde se encontram os valores extremos do atributo a ser estimado.

### 4.3.1 Abordagem alternativa

Na indústria do petróleo, as variáveis raramente apresentam distribuição normal. É razoável admitir que, para distribuições assimétricas dos dados, os intervalos de confiança também sejam assimétricos.

Intervalos de confiança para a média de uma distribuição condicional de  $Z_0 | Z_1, Z_2, \dots, Z_n$  devem ter a propriedade de assimetria reversa. Por exemplo, isto significa que valores da cauda superior de uma distribuição com assimetria positiva devem ter intervalos de confiança com assimetria negativa e os da cauda inferior, assimetria positiva.

A nível de informação local, a variância de krigagem fornece uma indicação da precisão do valor estimado, blocos tendo variâncias de estimativa similares sendo muito bem ou pessimamente estimados. Contudo, a variância de estimativa não leva em conta a distribuição assimétrica dos dados. Por esta razão, quando lidando com dados assimétricos, intervalos de confiança (ICs) permitem uma maior apreciação da viabilidade da estimativa do que a real incerteza local. Apesar desta aparente necessidade para a obtenção de ICs realistas a nível de estimação local, muitas aplicações práticas ainda confiam nos ICs baseados em hipóteses, às vezes questionáveis, sobre a normalidade dos erros de estimativa.

Roth e Armstrong (1995) propuseram um método para construir ICs de atributos com distribuição não gaussiana. Eles utilizaram a transformação gaussiana dos dados originais. Intervalos de confiança são obtidos para estas variáveis normalizadas, e transformados de volta para dar os correspondentes intervalos de confiança dos valores estimados. Os autores da proposta frisaram algumas hipóteses para a aplicação da metodologia. Inicialmente o atributo, cujos ICs serão obtidos, deverá apresentar estacionariedade de segunda ordem. Se necessário, os conjuntos de dados devem ser submetidos a uma 'declusterização' (remoção dos agrupamentos) a fim de que o histograma possa ser considerado como representativo da distribuição. Eles ressaltaram a necessidade dos dados transformados apresentarem a propriedade de distribuição multivariada gaussiana para obtenção do IC.

A metodologia proposta por Roth e Armstrong (1995) resulta em intervalos de incerteza com assimetria reversa e que não respeita apenas o erro associado com o processo de estimativa (a variância de krigagem), mas também a distribuição dos valores dos blocos.

## 4.4 Verificação Experimental

### 4.4.1 Dados Básicos

Foi utilizado um conjunto de dados de um campo real fornecido pela AMOCO para os centros de geoestatística para fins de pesquisa. Os dados são de um reservatório carbonático, de idade permiana, produtor de óleo no oeste do estado do Texas, EUA.

Este reservatório representa uma seqüência de plataforma carbonática progradante com arrasamento para o topo, formada em ambiente de moderada a baixa energia.

Estes dados da AMOCO fazem parte do estudo de caso de petróleo do ISATIS. Foram usados também nos trabalhos de Chu et al. (1991 e 1994) e Deutsch e Journel (1992).

No trabalho de Chu et al. (1991), os dados relativos a  $H\phi$  (produto espessura x porosidade) estão dispostos em uma mapa com coordenadas espaciais transformadas para assegurar o sigilo, mas preservando as estruturas de variabilidade. Deste mapa foram obtidos os valores das coordenadas e os respectivos  $H\phi$  dos 74 poços distribuídos em uma área de aproximadamente 15.4 km<sup>2</sup>. Nas figuras 4.1 e 4.2 estão respectivamente o mapa base com posição dos poços e o histograma dos dados relativos ao produto espessura vezes porosidade.

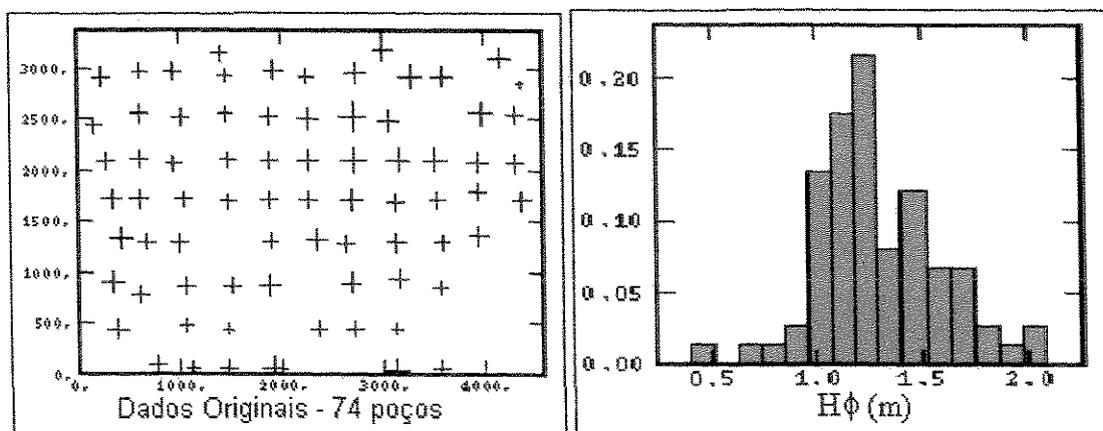


Figura 4-1: Mapa Base com a Posição dos Poços. Figura 4-2: Histograma dos dados originais.

Verifica-se que a distribuição dos dados (poços) é aproximadamente regular, não apresentando agrupamentos, evidenciando apenas algumas regiões com reduzida amostragem como no extremo sudeste. Quanto ao histograma, pode-se observar que trata de uma distribuição com assimetria positiva. A distribuição apresenta média 1.294 m com desvio padrão 0.296 m.

#### 4.4.2 Construção do Caso Base

Com este conjunto de dados, realizou-se krigagem ordinária em um *grid* com 61 blocos na direção X e 45 blocos na direção Y. Cada bloco tem dimensão de 75x75 metros. Este resultado foi considerado como caso base para análise da influência do número de dados iniciais e do variograma na estimativa. O caso base (figura 4.3) tem as seguintes características:

Valor mínimo  $H\phi$  por bloco: 0.9087 m; Valor máximo  $H\phi$  por bloco: 1.8335 m;

Valor médio de  $H\phi$ : 1.2949 m; Desvio Padrão de  $H\phi$ : 0.1760 m;

Volume Poroso Total (VPT) do reservatório: 20 Milhões de  $m^3$ ;

Variograma: uma estrutura gaussiana com alcance 1776.7m e patamar 0.0442 $m^2$ .

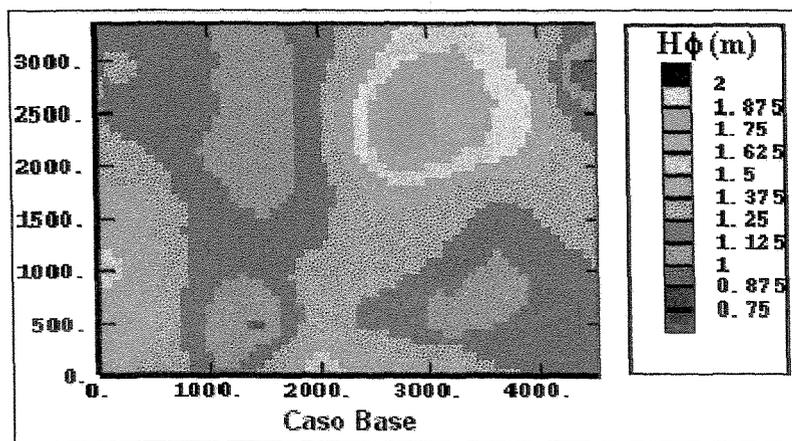


Figura 4-3 : Caso base.

Do caso base, foram retirados quatro conjuntos contendo 20, 48, 63 e 88 dados regularmente espaçados. Foram obtidos também quatro conjuntos de dados irregularmente espaçados, privilegiando ou altos ou baixos valores. Na figura 4.4 estão as distribuições das amostras constituintes dos conjuntos analisados.

#### 4.4.3 Parâmetros das Krigagens

Na tabela 4-2 estão os modelos de variogramas e a vizinhança utilizados para as krigagens em cada conjunto de dados. Os termos 'melhor' e 'pior' referem-se aos conjuntos de dados com amostragem seletiva em melhores e piores situações. Observa-se (tabela 4-2) que à medida que aumenta o número de amostras, mais o modelo de variograma se aproxima do variograma do caso base. As vizinhanças, variando o raio e o número mínimo de amostras necessárias para a estimativa, e o percentual de efeito pepita no modelo de semivariograma (máximo de 8% do patamar) foram escolhidos de forma a obter uma validação cruzada mais robusta e permitir a estimativa de todos os blocos do domínio.

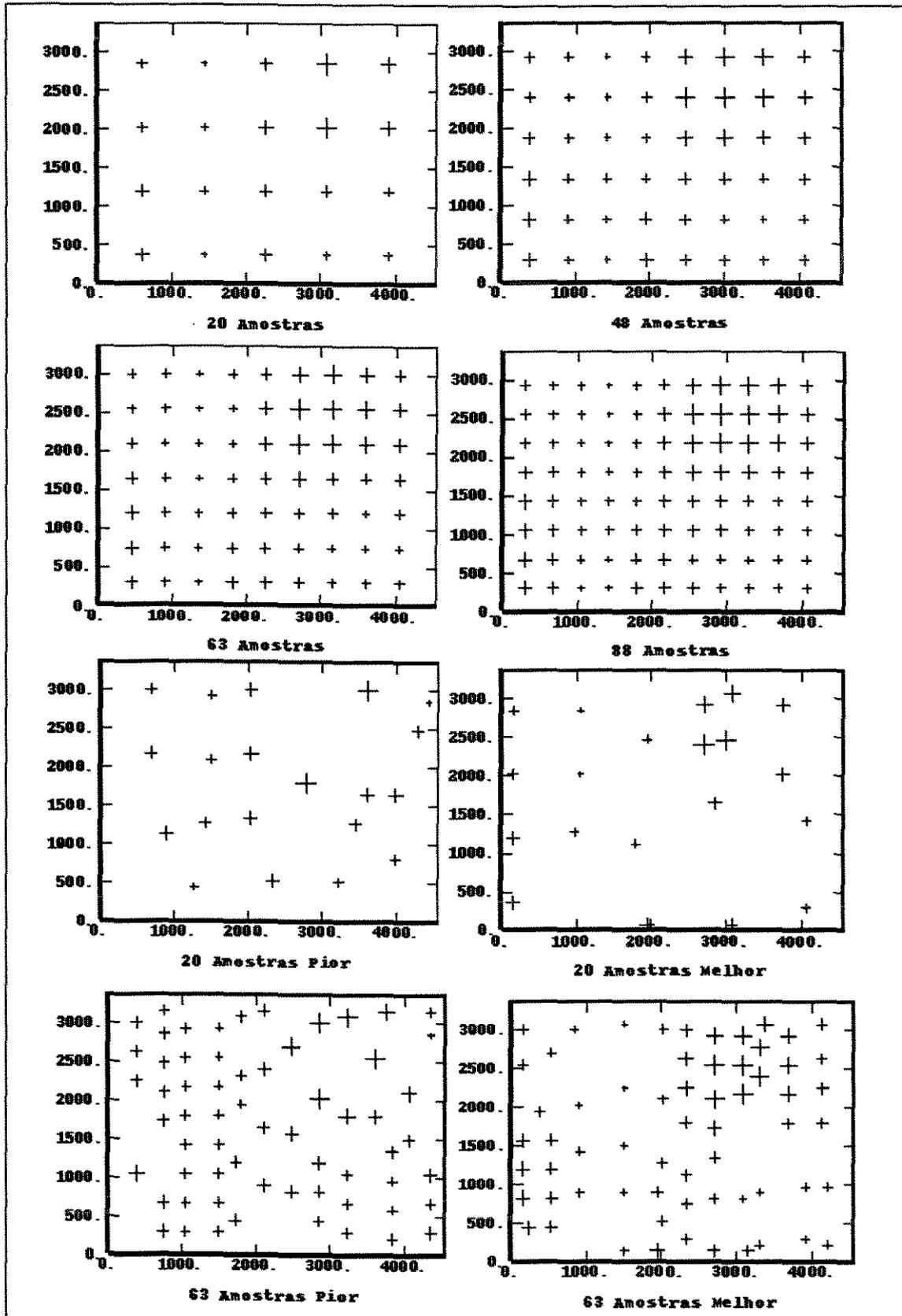


Figura 4-4: Conjuntos amostrais retirados do caso base.

**Tabela 4-2: Variogramas e Vizinhanças Adotadas nas Krigagens**

Conjunto de Dados	Variograma Isotrópico Ajustado			Vizinhança	
	Estrutura 1		Efeito Pepita	Circular	
	Modelo	Alcance (m)	Patamar	Patamar	Raio (m)
Original (Controle)	Gaussiano	1776.7	0.0442	-	-
20 Amostras	Esférico	2647.5	0.0391	0.0032	2000
20 Amostras melhor	Gaussiano	1848.4	0.0581	0.0030	1500
20 Amostras pior	Gaussiano	1942.0	0.0600	-	1600
48 Amostras	Esférico	2182.5	0.0419	-	1800
63 Amostras	Gaussiano	1793.1	0.0441	0.0031	1000
63 Amostras melhor	Gaussiano	1740.5	0.0433	0.0020	900
63 Amostras pior	Gaussiano	2181.9	0.0354	0.0016	1050
88 Amostras	Gaussiano	1839.6	0.0446	0.0008	900

#### 4.4.4 Resultados das Krigagens

Os resultados das krigagens em termos de volume poroso do reservatório estão na tabela 4-3. O cálculo do volume poroso total é obtido a partir da somatória do produto de  $H\phi$  em cada bloco pelas dimensões dos blocos (75m x 75m). Para dados regularmente espaçados, as estimativas de volume poroso total do reservatório apresentaram diferenças inferiores a 1%.

**Tabela 4-3: Resultados das Krigagens**

Conjunto	Volume Poroso do reservatório (Mm <sup>3</sup> )	( $H\phi$ ) médio por bloco (m)
Caso Base - controle	19994	1.2949
Krigagem - 20 Amostras pior	18825	1.2192
Krigagem - 20 Amostras melhor	20851	1.3504
Krigagem - 20 Amostras	19878	1.2874
Krigagem - 48 Amostras	19993	1.2948
Krigagem - 63 Amostras pior	19717	1.2769
Krigagem - 63 Amostras melhor	20161	1.3057
Krigagem - 63 Amostras	20019	1.2965
Krigagem - 88 Amostras	19990	1.2947

Para dados irregularmente espaçados privilegiando áreas de interesse, que é a situação mais comum em termos de engenharia de reservatórios, os resultados das estimativas podem ficar bastante enviesados. Por exemplo para o conjunto de dados com 20 amostras, a diferença entre a estimativa e o valor real apresentou uma diferença da ordem de 5%. Na indústria do petróleo, a depender da dimensão do reservatório e do atributo que está sendo estimado, diferenças de 1% ou 5% podem ter um impacto econômico bastante elevado. Os resultados (mapas krigados) estão na figura 4-5. Ressalta-se que o caso base analisado não apresentava anisotropias ou deriva. A ocorrência das mesmas, dificultaria a modelagem a partir de poucos dados e poderia levar a resultados mais discrepantes em termos de volume poroso total do reservatório.

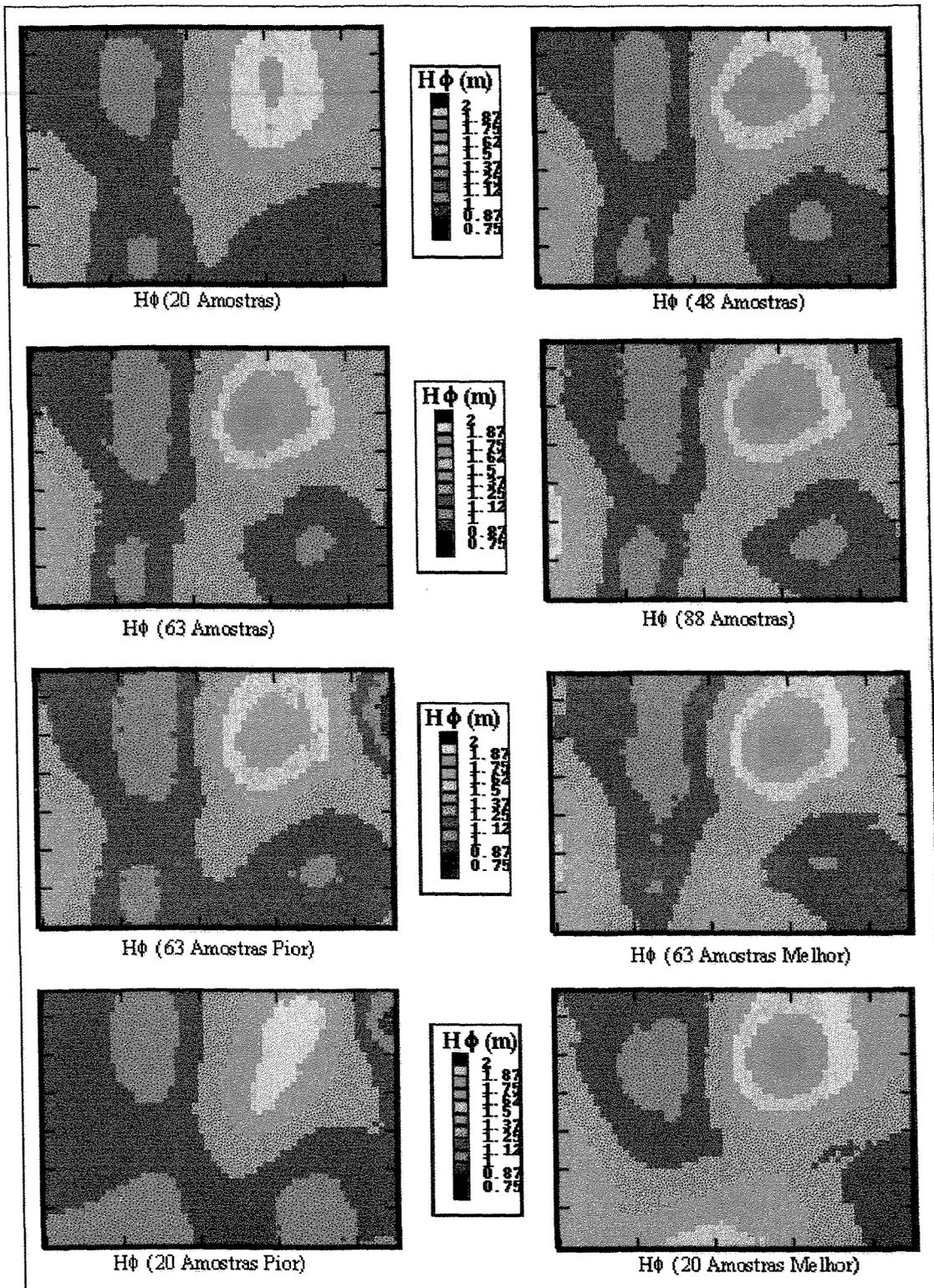


Figura 4-5: Resultados das krigagens.

#### 4.4.5 Intervalos de Incerteza

Admitindo a distribuição gaussiana para os erros, intervalos de incerteza foram obtidos considerando a estimativa mais ou menos um (intervalo 1) e dois desvios padrões (intervalo 2).

Na figura 4.6 estão os mapas considerando a estimativa mais e menos um desvio padrão para o caso de 20 dados amostrais regularmente espaçados. Correspondem aos mapas dos quantis 15.87% e 84.13%, considerando as hipóteses propostas de Isaaks e Srivastava (1989), anteriormente mencionadas.

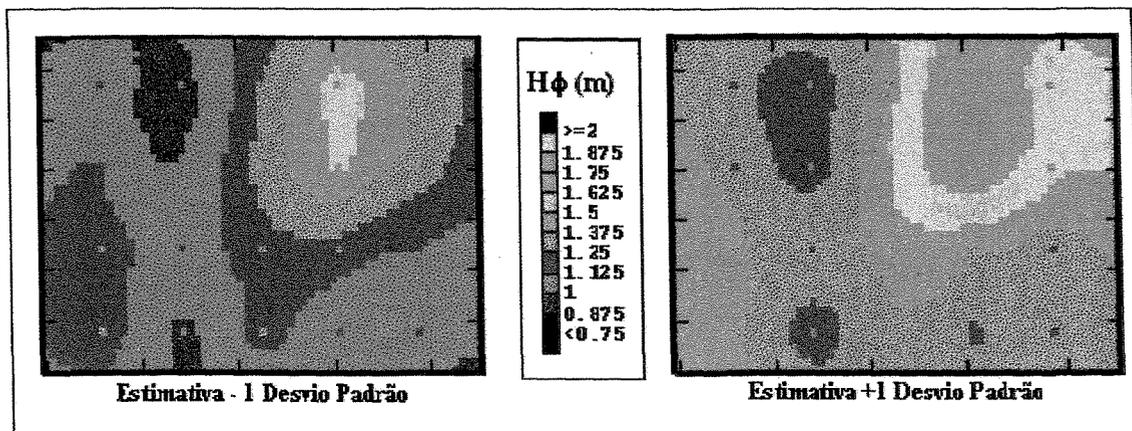


Figura 4-6: Mapas dos percentis 15.87 e 84.13 %.

Na tabela 4-4, além dos resultados relativos a intervalos de confiança, estão os

relativos aos erros:  $\epsilon_m = \frac{1}{2745} \sum_{i=1}^{2745} Z_i^{cb} - Z_i^*$  e  $\epsilon_{ma} = \frac{1}{2745} \sum_{i=1}^{2745} |Z_i^* - Z_i^{cb}|$

onde  $\epsilon_m$  é o erro médio,  $\epsilon_{ma}$  é o erro absoluto médio,  $Z_i^*$  e  $Z_i^{cb}$  são respectivamente o valor estimado de  $H\phi$  e o valor do caso base em um bloco  $i$  do *grid*. Na tabela 4-4, as colunas relativas aos intervalos de incerteza 1 e 2 correspondem em quantos blocos, os valores reais (caso base) ficaram dentro dos respectivos intervalos construídos a partir da estimativa e de sua variância.

Da tabela 4-4, confirma-se que o estimador é não enviesado, pois o erro de estimativa é nulo na média ( $\epsilon_m \cong 0$ ). Para amostras regularmente espaçadas, verifica-se que à medida que aumenta o número das mesmas, ocorre uma diminuição acentuada do erro absoluto médio das estimativas em relação ao caso base. Para amostras irregularmente espaçadas, os erros absolutos médios se apresentaram sempre maiores em relação à situação com mesmo número de dados e amostragem regular.

**Tabela 4-4: Intervalos de Incerteza e Erros**

Estimativa com	Intervalo de Incerteza 1	Intervalo de Incerteza 2	Erro médio	Erro absoluto médio
20 Amostras	2339	2658	0.0075	0.067
20 Amostras pior	1982	2617	0.0757	0.087
20 Amostras melhor	2220	2625	0.0554	0.070
48 Amostras	2530	2670	0.0001	0.030
63 Amostras	2509	2636	-0.0016	0.026
63 Amostras pior	2210	2570	0.0180	0.034
63 Amostras melhor	2439	2656	-0.0108	0.027
88 amostras	753	1118	0.00024	0.020

Como se dispõe dos dados considerados reais, era de se esperar que 1874 e 2608 valores estimados de  $H\phi$  por bloco (o *grid* tem 2745 blocos) ficassem dentro dos intervalos de incerteza 1 e 2, respectivamente, considerando a condição de distribuição gaussiana dos erros. Isto foi verificado com folga para as situações de 20, 48 e 63 amostras regular e irregularmente espaçadas, não ocorrendo para o caso de 88 amostras. Porém, para a condição de 88 amostras, as estimativas ficaram mais próximas do valor real, isto é, foram observados os menores erros. Pode-se concluir que a variância de estimativa ficou reduzida o suficiente, de forma que os intervalos de incerteza ficaram bastante estreitos, a ponto de não incluírem o valor real. Para esta situação, poder-se-ia supor que a melhoria na precisão (variância de estimativa) se deu às custas da acurácia (conter o valor real). Nos casos de quantidades inferiores de amostras, ocorreu exatamente o contrário. Em verdade, é o modelo de distribuição gaussiana para os erros que não é adequado para todos os casos.

Através de testes de hipóteses, verificou-se que todas as distribuições globais dos erros apresentaram grandes discrepâncias em relação a uma distribuição gaussiana, podendo rejeitar a hipótese de distribuição normal ao nível de 5% para todas as condições.

#### 4.4.6 Análise Visual

A análise visual e qualitativa dos mapas gerados levam indubitavelmente a observar uma melhoria acentuada na semelhança com o caso base à medida que se aumenta o número de dados. Ressalta-se que apesar da diferença entre a estimativa do volume poroso total do reservatório para o caso de 20 e 88 amostras ser bem pequena, as diferenças bloco a bloco (figura 4.5) são marcantes.

A comparação quantitativa entre as estimativas e o caso base podem ser feitas usando o coeficiente de correlação. Da Tabela 4-5, verifica-se que, com o aumento do número de

dados regularmente espaçados, ocorre um aumento do fator de correlação. Porém, o impacto de aumentar o número de dados é cada vez menor. O aumento foi mais pronunciado (13.8%) quando se aumenta de 20 para 48 dados.

**Tabela 4-5: Fator de correlação**

Estimativa com	Fator de correlação
20 dados - pior	0.8336
20 dados - melhor	0.8339
20 dados	0.8405
48 dados	0.9566
63 dados - pior	0.9537
63 dados	0.9635
63 dados - melhor	0.9721
88 dados	0.9779

#### 4.4.7 Variogramas Cruzados

Como se dispõe de um caso base, considerado como referência, pode-se quantificar a melhoria da qualidade de uma estimativa com o aumento do número de dados amostrais, a partir do variograma cruzado entre os valores da imagem de referência e os valores de uma imagem krigada. Considere **Z** correspondente ao caso base e **Y** correspondente a uma imagem krigada. O semivariograma cruzado entre **Z** e **Y** é:

$$\gamma_{ZY}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \{ [Z(x_i + h) - Z(x_i)] [Y(x_i + h) - Y(x_i)] \}$$

Quanto mais próximo do caso base a imagem estiver, mais próximo o semivariograma cruzado vai estar do semivariograma do caso base. Na expressão do semivariograma cruzado, com  $Y=Z$ , obtém-se a expressão do semivariograma para **Z**. Para esta análise foram considerados apenas os conjuntos de dados regularmente espaçados.

Na figura 4-7 estão os semivariogramas cruzados e o semivariograma do caso base. Verifica-se melhoria acentuada à medida que aumenta o número de amostras. Concordante com as análises anteriores, o maior percentual de melhoria se dá no aumento de 20 para 48 amostras. O interessante desta visualização é que para valores de **h** (distância) maiores que 1000 metros, o aumento do número de poços continua a reduzir consideravelmente a diferença em relação ao semivariograma do caso base. Esta forma de comparação de estimativas com um caso de referência é mais informativa do que o coeficiente de correlação. Ressalta-se que poderiam ter sido utilizadas outras funções geoestatísticas como, por exemplo, a covariância e o correlograma cruzados. Os resultados seriam similares.

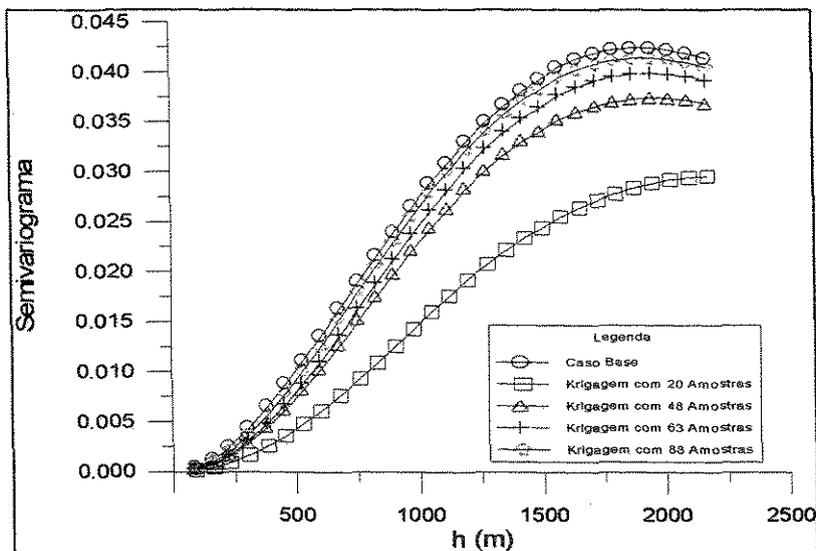


Figura 4-7: Variogramas cruzados entre estimativas e caso real.

#### 4.4.8 Comparação através de cortes

Uma outra maneira de comparar as diversas estimativas realizadas em relação ao caso de referência é a utilização de cortes. Isto corresponde a considerar um determinado parâmetro de corte para o atributo de interesse, no caso  $H_{\phi}$ , e calcular o volume poroso do reservatório só considerando os blocos nos quais a estimativa é maior que o corte. Foram utilizados três cortes correspondentes a  $H_{\phi}$  por bloco igual a 1.1, 1.3 e 1.5 m. Realizou-se então o cálculo do volume poroso considerando apenas os blocos onde a estimativa do atributo  $H_{\phi}$  era superior ao dado corte. Observa-se na tabela 4-6 que os erros podem ser elevados para as condições de poucas amostras, principalmente se irregularmente distribuídas.

Tabela 4-6: Volumes porosos para diferentes cortes

Estimativa com	Corte $H_{\phi}$ 1.1m		Corte $H_{\phi}$ 1.3m		Corte $H_{\phi}$ 1.5m	
	Volume (Mm <sup>3</sup> )	Erro (%)	Volume (Mm <sup>3</sup> )	Erro (%)	Volume (Mm <sup>3</sup> )	Erro (%)
Caso Base	18291	-	9032	-	3464	-
20 Dados melhor	20330	11.1	12991	43.8	3918	13.1
20 Dados	19421	6.2	8401	-7.0	2566	-25.9
20 Dados pior	16697	-8.7	4855	-46.2	1276	-63.2
48 Dados	19009	3.9	9298	2.9	2964	-14.5
63 Dados melhor	19239	5.2	9304	3.0	3400	-1.9
63 Dados	18984	3.8	8826	-2.3	3215	-7.2
63 Dados pior	18517	1.2	7569	-16.2	2873	-17.1
88 Dados	18451	0.9	8886	-1.6	3556	2.7

Para o corte de 1.3m foram refeitos os cálculos individuais dos volumes porosos estimados divididos em quatro classes: abaixo e acima do parâmetro de corte, os falsos positivos e os falsos negativos. Os falsos positivos são blocos que no caso de referência estão

abaixo do parâmetro de corte e na estimativa estão acima. Os falsos negativos correspondem a situação inversa. Na figura 4-8 estão os gráficos para as condições de 20 e 88 dados regularmente espaçados. Verifica-se que ao considerar apenas as estimativas onde realmente o valor real está acima do parâmetro de corte, as diferenças entre as duas situações aumentam significativamente. Como era de se esperar, a condição com menor número de amostras apresenta maior percentual de falsos positivos e falsos negativos. Como em todas as abordagens comparativas utilizadas, evidencia-se melhoria da qualidade da estimativa à medida que se aumenta o número de dados. Com a abordagem de cálculo dos volumes através de cortes esta melhoria ficou mais evidenciada.

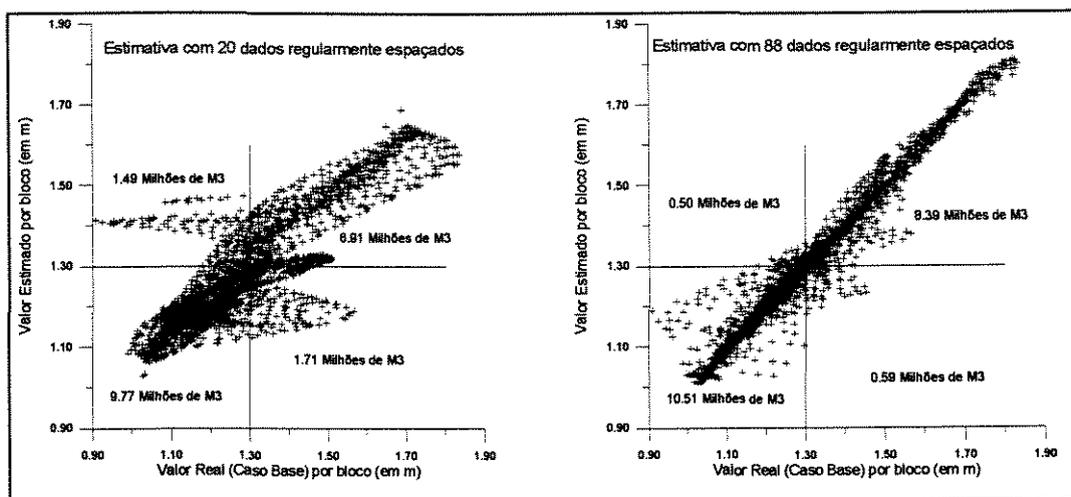


Figura 4-8: Distribuições de volumes porosos em relação ao corte de  $H\phi$  1.3 m.

#### 4.4.9 Intervalos de Incerteza Locais- Abordagem de Roth e Armstrong

Baseando-se na metodologia apresentada por Roth e Armstrong (1995), foram refeitas as krigagens para os conjuntos de dados uniformemente distribuídos.

Para tal, os dados originais foram submetidos a transformação gaussiana sendo obtidos novos semivariogramas. Foram mantidos os alcances de todos os semivariogramas, bem como a vizinhança adotada para cada um dos conjuntos de dados regularmente espaçados. Nas tabelas 4-7 e 4-8 estão os principais resultados obtidos com esta nova abordagem. Adotou-se a sigla RA para fazer referência à abordagem proposta por Roth e Armstrong.

Verifica-se a partir da tabela 4-7, uma tendência de aproximar da distribuição gaussiana dos erros com a abordagem RA, o que corresponde a 1874 e 2608 valores de  $H\phi$  reais dentro dos intervalos de incerteza 1 e 2 respectivamente.

Com relação aos intervalos de incerteza para a estimativa local, confirma-se que a grande maioria dos ICs obtidos com a abordagem RA apresentam assimetria reversa. Assim, estimativas próximas ao limite superior da distribuição apresentaram intervalos de incerteza predominantemente à esquerda da mesma. O contrário foi observado para estimativas próximas ao limite inferior. Ressalta-se que as distribuições originais dos 4 conjuntos de dados são assimétricas positivas.

**Tabela 4-7: Intervalos de Incerteza**

Estimativa com	Valores reais no Intervalo de Incerteza 1	Valores reais no Intervalo de Incerteza 2
20 Amostras	2339	2658
20 Amostras RA	2332	2508
48 Amostras	2530	2670
48 Amostras RA	2463	2621
63 Amostras	2509	2636
63 Amostras RA	2388	2590
88 amostras	753	1118
88 amostras RA	2228	2525

Estimativas que na abordagem tradicional apresentam a mesma variância de krigagem (intervalos de incerteza iguais e simétricos), apresentam largura dos intervalos de incerteza, com a abordagem RA, tanto maior quanto maior for a estimativa (tabela 4-8). A largura de um intervalo de incerteza corresponde à diferença entre os seus dois limites. Para a abordagem tradicional, a largura do IC 2 é o dobro da largura do IC 1.

**Tabela 4-8: Intervalos de Incerteza (Estimativa Local:  $H\phi$  por bloco em m)**

Conjunto de Dados	Coordenadas X/Y do bloco	Estimativa de $H\phi$ (m)	Largura do IC 1		Largura do IC 2	
			Tradicional	RA	Tradicional	RA
20 Amostras	1050/1800	1.16	0.236	0.224	0.472	0.355
	3450/1800	1.47	0.236	0.332	0.472	0.490
88 Amostras	1050/1800	1.06	0.0045	0.032	0.0090	0.064
	3450/1800	1.77	0.0045	0.036	0.0090	0.082

Verificou-se que os intervalos de incerteza obtidos com a abordagem proposta por Roth e Armstrong tendem a ser mais conservativos, isto é, mais largos do que os obtidos com a abordagem tradicional, quanto maior for o valor da estimativa. Adicionalmente verificou-se também que se a variância de krigagem é pequena em relação à estimativa em qualquer posição, como no caso com 88 amostras, a abordagem RA tende a apresentar intervalos de incerteza locais mais conservativos que a abordagem tradicional, qualquer que seja o valor da estimativa local (Tabela 4-8). Isto leva a intervalos de incerteza globais muito mais conservativos com a abordagem RA do que a tradicional. Ressalta-se que a proposta de Roth e Armstrong visa apenas a obtenção de intervalos de incerteza locais representativos.

## 4.5 Comentários

Uma questão fundamental, bem colocada por Isaaks e Srivastava (1989), é se somos capazes de estimar a variância dos dados reais. Para esta condição, o patamar do semivariograma deve ser uma boa estimativa da variância global do atributo a ser estimado. Nas situações analisadas, o patamar dos semivariogramas ficaram sistematicamente acima da variância dos valores estimados por krigagem para cada conjunto de dados em cada situação e da variância dos valores do caso base. Em uma situação real não se dispõe do caso considerado referência. Mas, de qualquer forma, o patamar do semivariograma diferente da variância dos valores estimados em um determinado domínio é um forte e inquestionável indício que provavelmente não se pode estimar a variância dos erros reais.

A abordagem proposta por Roth e Armstrong (1995) gera intervalos de incertezas com assimetria reversa e, portanto, representa um grande avanço no sentido de obter intervalos de incerteza de krigagens paramétricas. Porém requer, como na abordagem tradicional, que as hipóteses de distribuição multivariada gaussiana e a capacidade de avaliar a variância dos erros reais sejam verdadeiras. O principal problema verificado com a abordagem é a tendência de obtenção de intervalos de incerteza locais mais conservativos que a abordagem tradicional, qualquer que seja o valor da estimativa, para as situações em que a variância de krigagem é pequena comparada à estimativa. Isto leva a intervalos de incerteza globais muito mais conservativos.

Várias referências existem na literatura abordando intervalos de incerteza com krigagens paramétricas sob a ótica apenas da distribuição multivariada gaussiana. Poucos reportam e avaliam a capacidade de estimar a variância dos erros reais. Armstrong (1994, p.306) reporta que a idéia de obter intervalos de incerteza a partir da estimativa e da variância de krigagem reaparece regularmente, mesmo em respeitáveis periódicos.

Por exemplo em Harbaug et al. (1995, p.165-166), os autores mostram mapas de espessura estimada (krigagem) mais ou menos 1.63 desvios padrão. Os dados constituem a espessura de um reservatório carbonático de uma área do Oriente Médio. A única hipótese ressaltada pelos autores é que a distribuição dos erros de estimativa é gaussiana. Os autores avançam considerando distribuição gaussiana dos erros locais (Harbaugh et al., 1995, p. 167). Eles construíram uma 'seção' de uma determinada coordenada e afirmam que 'a probabilidade é apenas de 10 % da espessura cair fora do envelope da estimativa mais ou menos 1.63 desvios padrão para qualquer posição' (figura 4-9). Curiosamente na figura mostrada pelos autores, os

intervalos que deveriam ser simétricos em relação à estimativa (uma vez que admitiu-se a distribuição gaussiana local), são apenas aproximadamente simétricos. Observar na figura 4.9, que o trecho entre as coordenadas 20 e 29 (coluna X), próximas aos 2 poços, os limites superiores e inferiores não são simétricos em relação à estimativa. Provavelmente trata-se de um descuido quando da impressão original da citada obra.

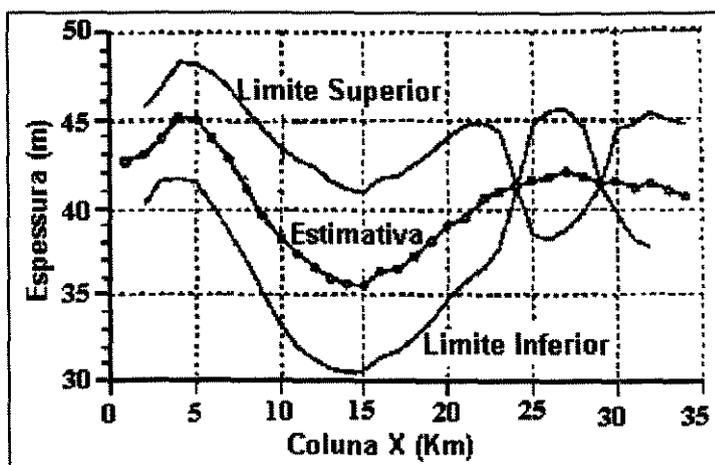


Figura 4-9: Intervalo de incerteza ao longo de uma linha com coordenada Y fixa, passando por dois poços. (Figura 7.20 de Harbaugh et al. 1995, p.167)

Concluindo, a construção de intervalos de incerteza globais e locais a partir de krigagens paramétricas é uma área em que a pesquisa pode avançar. O grande apelo da metodologia da distribuição gaussiana dos erros é a extrema facilidade com que se pode obter os intervalos de incerteza a partir da estimativa e da variância de krigagens paramétricas. Ressalta-se que a abordagem proposta por Roth e Armstrong (1995) segue a mesma linha da construção de cdf a partir de krigagens nas etapas de simulação estocástica gaussiana.

Apesar de não terem sido utilizadas abordagens não paramétricas (krigagem indicatriz) na construção de intervalos de incerteza para variáveis contínuas, algumas considerações podem ser feitas. A grande vantagem sempre reportada pelos autores e usuários da krigagem indicatriz é que ela não requer a adoção de nenhum modelo de distribuição a priori. Esta questionável supremacia da abordagem não paramétrica depende de um elevado número de cortes (*cutoffs*) e da utilização da cokrigagem indicatriz, que são condições difíceis de serem obedecidas. Por exemplo, um reduzido número de cortes implica em uma redução da quantidade e conseqüentemente da representatividade dos diferentes intervalos de incerteza que podem ser obtidos. Ressaltam ainda os problemas de relação de ordem que ocorrem com as abordagens indicatrizes. De qualquer forma, esta metodologia exige um maior esforço computacional e um maior trabalho de inferência de semivariogramas do que a abordagem paramétrica, daí o porquê de muitas vezes se utilizar esta última, mesmo sob hipóteses algumas vezes questionáveis.

## 5. Simulação Estocástica

Atualmente, existe uma grande variedade de algoritmos de simulação estocástica disponíveis que diferem em suas hipóteses básicas, campo de aplicação, dados requeridos, implementação prática, complexidade e eficiência computacional.

A escolha de um método de simulação estocástica deve se basear na disponibilidade de tempo, dinheiro, recursos humanos e computacionais de hardware e software, bem como a complexidade requerida pelo estudo. Parafraseando Srivastava (1994b): nem todos os estudos de modelagem necessitam de uma técnica ‘Cadillac’, algumas vezes um ‘Volkswagen’ ou um bom par de patins pode ser suficiente. Esse bom par de patins pode ser a krigagem que, a depender do objetivo do estudo, pode ser aplicada. Infelizmente também pode ocorrer o inverso. A importância do estudo requerer uma modelagem geoestatística elaborada e se utilizar uma simplificada, ou mesmo não utilizar a geoestatística.

Neste ponto se faz necessário explicitar os principais aspectos que diferem a krigagem da simulação estocástica. O objetivo da krigagem é fornecer uma única estimativa local, sem contudo respeitar a estatística espacial de todas as estimativas consideradas conjuntamente. Assim, o semivariograma de um mapa krigado é diferente do usado na krigagem. Na simulação estocástica, a reprodução de características globais (textura) e estatísticas (histograma e covariância) são priorizadas em relação à acurácia local na obtenção de vários conjuntos alternativos de representações.

A krigagem apresenta a característica de suavização. Só que esta suavização será menos acentuada quanto mais contínuo for o fenômeno que está sendo modelado. A título de demonstração foram construídos, via simulação estocástica, dois casos base, um com modelo de semivariograma gaussiano e outro com modelo de semivariograma exponencial. Destes casos, foram retiradas aleatoriamente 100 amostras. Com estas amostras e os modelos de variogramas obtidos da variografia de cada caso base, foram realizadas uma simulação estocástica com bandas rotativas e a krigagem para cada um dos casos. Na figura 5-1 estão os resultados das simulações e da krigagem. Verifica-se que quanto menos errático, isto é, mais contínuo for a realidade do fenômeno que está sendo modelado, menores serão diferenças entre as diversas realizações da simulação estocástica e a krigagem. Os efeitos de suavização

da krigagem são mais severos quanto mais errático, isto é, menos contínuo for o fenômeno analisado.

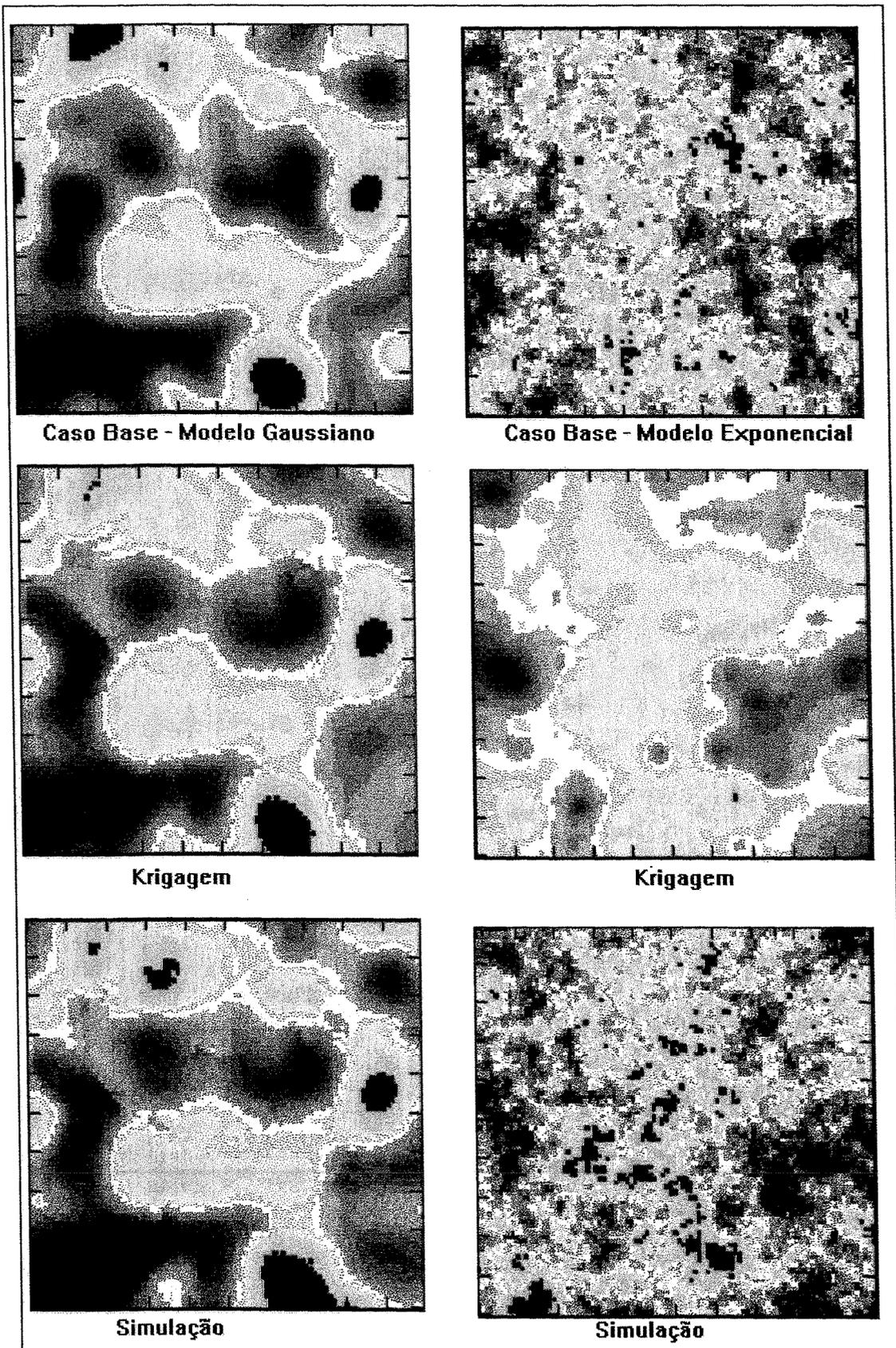


Figura 5-1:Krigagem x Simulação condicional.

Na geoengenharia de reservatórios, quando os resultados obtidos com as ferramentas geoestatísticas são submetidas a funções de transferência, possivelmente não lineares, como por exemplo um simulador de fluxo, deve-se optar pela utilização da simulação estocástica.

A questão central na caracterização de reservatórios é a modelagem realística das heterogeneidades no reservatório. Portanto, todas as etapas envolvidas no processo de simulação estocástica devem ser cuidadosamente verificadas. A escolha do algoritmo de simulação estocástica e até mesmo a escolha do software a ser utilizado podem influenciar os resultados. Existem diferenças importantes na implementação do algoritmo de um mesmo tipo de simulação nos softwares geoestatísticos comerciais: parâmetros que podem ser gerenciados pelos usuários, na forma de apresentação dos resultados e mesmo na metodologia de implementação do algoritmo. Segundo Srivastava (1994b), escolher uma abordagem específica pode se tornar uma tarefa desconcertante para os novos usuários da simulação estocástica. É necessário saber qual método é mais adequado ao particular contexto. Infelizmente, não existe a melhor técnica, qualquer que seja a situação.

Os artigos que propõem as novas metodologias de simulação estocástica geralmente se concentram nas vantagens das mesmas, deixando que as desvantagens sejam publicadas por outros autores que, ou propõem um algoritmo alternativo ou uma melhoria no algoritmo anterior. Assim, cada técnica tem seus próprios defensores, os seus próprios autores.

Várias revisões de métodos de simulação estocástica estão disponíveis na literatura, destacando-se Dubrule (1989), Alabert e Modot (1992), Dowd (1992), Damsleth e Holden (1994), Galli e Beucher (1997), dentre outros. Ressalta-se que a maioria desses trabalhos são direcionados a um leitor especialista em geoestatística. Para o usuário com menor conhecimento geoestatístico, é geralmente difícil entender as diferenças entre os métodos e suas principais características. Com intuito de fornecer uma visão global das técnicas de simulação estocástica mais utilizadas, as mesmas são abordadas a seguir de forma a evidenciar suas linhas mestras de aplicação e principais características.

## ***5.1 Bandas Rotativas***

O algoritmo das bandas rotativas foi o primeiro algoritmo de simulação 3D realmente implementado em grande escala. Neste método são feitas as seguintes hipóteses: o campo a ser simulado é estacionário de segunda ordem e isotrópico com covariância  $C(h)$  conhecida; os valores deste campo são  $N(0,1)$ , se necessário, após a transformação gaussiana; o campo apresenta distribuição multivariada gaussiana.

A originalidade do método desenvolvido por Matheron (1973, apud Journel e Huijbregts, 1978) está na redução das simulações  $n$ -dimensionais em várias e independentes simulações unidimensionais ao longo de linhas que são giradas no espaço  $R^n$ . Um valor simulado em uma posição  $x$  é obtido a partir das contribuições das  $n$  simulações unidimensionais  $z_{si}(x)$ .

$$z_{snc}(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^n z_{si}(x)$$

A realização resultante  $z_{snc}(x)$  é uma realização de uma FA, por exemplo no caso 3D  $Z_s(x) = Z_s(u, v, w)$ , que é estacionária de segunda ordem, com esperança zero e covariância

$$E[Z_s(x)Z_s(x+h)] = C(h)$$

que tende à covariância isotrópica quando o número de linhas tende a infinito. Uma covariância  $C(h)$  anisotrópica pode ser modelada como somatória de modelos isotrópicos imbricados em espaços de dimensões  $n \leq 3$ . Basta então simular independentemente cada um dos componentes do modelo e em seguida somar as realizações em cada ponto.

Na prática, o número  $n$  de linhas não pode ser infinito. O reduzido número de linhas pode provocar o efeito de artefatos (lineamentos) nos campos simulados, gerando anisotropias inexistentes. Além disso, o algoritmo das bandas rotativas somente é capaz de lidar com determinados tipos de função covariância.

O condicionamento para a simulação, via bandas rotativas, trata-se de uma etapa em separado. Para produzir um *grid* com os valores condicionantes e que reproduzam a variabilidade espacial da FA, valores são produzidos através de krigagens, usando os valores simulados não condicionais na posição dos dados originais. Estes novos valores krigados são subtraídos dos valores da simulação não condicional para se obter um *grid* com valores de erros correlacionados. Estes erros correlacionados são adicionados aos valores de uma krigagem realizada considerando os dados originais, de forma a produzir uma simulação condicional. Então:

$$z_{sc} = z_{kd} + (z_{snc} - z_{ksnc})$$

Onde  $z_{sc}$  é a simulação condicional,  $z_{kd}$  é a krigagem obtida a partir dos dados originais reais,  $z_{snc}$  é a simulação não condicional e  $z_{ksnc}$  é a krigagem dos valores simulados não condicionalmente nas mesmas posições dos dados originais condicionantes.

## 5.2 Simulação Gaussiana Seqüencial

Os algoritmos seqüenciais propostos por Journel e Alabert (1989) são uma aplicação do Teorema de Bayes. Os  $n$  eventos independentes  $A_i, i = 1, \dots, n$  podem ser simulados seqüencialmente usando a expressão:

$$P[A_1, A_2, \dots, A_n] = P[A_n | A_1, A_2, \dots, A_{n-1}] P[A_{n-1} | A_1, A_2, \dots, A_{n-2}] \dots P[A_2 | A_1] P[A_1].$$

A técnica requer a inferência de sucessivas  $(n-1)$  distribuições de probabilidades condicionais. Isto pode ser obtido de duas formas. A primeira é por adoção do modelo gaussiano, o que corresponde à simulação gaussiana seqüencial, e a segunda é por inferência direta da distribuição através do uso das indicatrizes correspondendo à simulação indicatriz seqüencial.

No algoritmo de simulação gaussiana seqüencial, cada variável é simulada seqüencialmente de acordo com a sua cdf gaussiana completamente caracterizada através de um sistema de krigagem. Em cada etapa, os dados condicionantes são todos os dados originais mais todos os valores previamente simulados que se encontram dentro da vizinhança da posição a ser simulada.

O algoritmo da Simulação Gaussiana Seqüencial consiste das seguintes etapas:

1. Determinar a cdf univariada  $F_Z$  representativa da área inteira sob estudo e não apenas dos dados amostrais disponíveis. A remoção dos agrupamentos (*declustering*) é necessária se os dados estão preferencialmente localizados;
2. Fazer a transformação gaussiana da cdf univariada de  $F_Z$ ;
3. Calcular um modelo de semivariograma dos dados condicionantes transformados;
4. Definir um caminho aleatório, não necessariamente regular, que visita uma vez todos os nós do *grid*;
5. Em cada nó  $\mathbf{x}$ , reter um determinado número de dados condicionantes vizinhos incluindo os dados originais transformados e os previamente simulados;
6. Usar um tipo de krigagem com o variograma dos dados transformados para obter a estimativa e a variância de krigagem. Construir a cdf gaussiana da FA  $Y(\mathbf{x})$  em cada posição  $\mathbf{x}$ ;
7. Sortear aleatoriamente um valor de  $y(\mathbf{x})$  desta cdf;
8. Adicionar o valor simulado ao conjunto de dados;
9. Voltar ao passo 5 até todos os nós serem simulados;

10. Transformar os valores simulados gaussianos para a ordem de grandeza da variável original.

A transformação gaussiana dos dados originais define uma nova variável  $Y$  que é, por construção, univariada normalmente distribuída. Contudo, todos métodos gaussianos requerem que esta nova variável também seja multivariada normalmente distribuída.

O algoritmo de simulação gaussiana seqüencial como foi descrito, requer a solução de sistemas de krigagem cada vez maiores para o cálculo das probabilidades à medida que o algoritmo progride, pois aumenta o número de dados condicionantes. Para contornar este problema se adota vizinhança móvel, em vez de vizinhança única para os cálculos das cdfs.

A vizinhança deveria ser tão grande quanto o alcance para possibilitar a reprodução dos variogramas. Às vezes, o uso de vizinhança muito grandes torna-se impraticável. Quando é importante a reprodução de amplitudes práticas muito grandes em relação ao espaçamento do *grid*, deve-se utilizar a alternativa da abordagem *multigrid* (Tran, 1993).

A decomposição da função de densidade de probabilidade (pdf) multivariada em produto de pdfs univariadas é independente do ordenamento das  $n$  variáveis aleatórias. Contudo, como a adoção da vizinhança móvel se faz necessária para reduzir o tempo de cálculo das cdfs, uma seqüência que visite nós adjacentes deve ser evitada. Por exemplo, tem-se a estratégia de busca em espiral que é um algoritmo eficiente quando os dados estão ou são reposicionados em um *grid* regular. A idéia é simular os nós em uma seqüência que sempre privilegie o nó mais distante do anteriormente simulado e que ainda não tenha sido simulado.

Geralmente, o número de nós a ser simulado excede em muito o número de dados condicionantes e, conseqüentemente, os dados originais terão sua influência reduzida à medida que o algoritmo progride. Os softwares comerciais, como por exemplo o ISATIS, fornece a opção do usuário limitar o número máximo de nós previamente simulados, mesmo que estejam na área da vizinhança, a serem usados nos sistemas de krigagem da simulação.

As maiores vantagens da simulação seqüencial gaussiana são: fácil condicionamento, anisotropias manuseadas automaticamente, aplicável para qualquer função covariância.

A aparente desvantagem do método é a utilização da distribuição gaussiana intermediária. Vários trabalhos reportam que as simulações gaussianas resultam em pequena continuidade dos valores extremos. Se este comportamento for crucial para a confiabilidade da simulação, alguns autores, como por exemplo Journel e Deutsch (1993), não recomendam a utilização das técnicas gaussianas.

Na utilização das diferentes implementações da simulação seqüencial gaussiana, vários parâmetros devem ser especificados pelo usuário, tais como, krigagem simples versus krigagem ordinária, máximo número de nós simulados mantidos para a krigagem, parâmetros de vizinhança, estratégia de busca e, em particular, limites superior e inferior da escolha de extrapolação a serem usados na transformação gaussiana. A escolha dos diferentes parâmetros podem afetar a eficiência do algoritmo, a natureza das realizações, e o resultado da distribuição de incertezas. Por exemplo, a rigor, a reprodução dos histogramas e dos semivariogramas do modelo só podem ser alcançadas através da utilização da KS no sistema de krigagem da simulação seqüencial gaussiana. Os softwares comerciais, às vezes, só flexibilizam a escolha de alguns desses parâmetros.

### 5.2.1 Cossimulação Seqüencial Gaussiana

Na geoengenharia de reservatórios, as variáveis apresentam muitas vezes correlações e devem ser simuladas em conjunto. Verly (1993) e Gómez-Hernández e Journel (1993) propuseram uma extensão à SGS, denominada cossimulação gaussiana seqüencial, como forma de simular diversas variáveis contínuas. Ela reproduz as distribuições e os variogramas diretos e cruzados das variáveis. O único problema adicional é a inferência e modelagem da matriz de covariâncias cruzadas. A metodologia assume que as variáveis são estacionárias de segunda ordem com covariâncias cruzadas também estacionárias e apresentam distribuição multivariada gaussiana, se necessário após a transformação gaussiana.

O algoritmo de simulação pode ser facilmente generalizado para manusear múltiplas variáveis, representando vários atributos sobre um grande número de posições. Considere a distribuição conjunta de um total de  $n$  variáveis aleatórias. As  $n$  variáveis aleatórias incluem  $K$  diferentes atributos em  $n'$  nós de um *grid*, com  $n = Kn'$ . A notação genérica  $Z_i$ , com  $i \in [1, \dots, n]$  é usada para indicar variável aleatória qualquer que seja sua posição ou tipo de atributo. A específica notação  $Z_{jk}$  com  $j \in [1, \dots, n']$  e  $k \in [1, \dots, K]$  representa uma variável aleatória em uma posição  $j$  e um correspondente atributo  $K$ .

Considere o condicionamento das  $n$  variáveis aleatórias por um conjunto  $n_0$  de dados de qualquer tipo. A correspondente pdf  $n$ -variada condicional pode ser decomposta usando o axioma da probabilidade condicional em um produto de  $n$  pdfs condicionais univariadas:

$$f((z_1, \dots, z_n)|(n_0)) = f((z_n)|(n-1) \cup (n_0)) f((z_{n-1})|(n-2) \cup (n_0)) \dots \\ \dots f((z_2)|(z_1) \cup (n_0)) f(z_1|(n_0))$$

onde, por exemplo,  $f(z_2|z_1 \cup (n_0))$  representa a pdf condicional de  $z_2$  dado a realização  $z_1$  e um dado conjunto de  $n_0$  valores de dados originais.

Esta decomposição permite obter realizações de  $n$  variáveis aleatórias  $Z_i$  condicionadas a  $n_0$  dados originais. O algoritmo é independente da ordem usada para visitar as  $n$  variáveis aleatórias. O único pré-requisito para sua implementação é o cálculo da ccdf em cada passo, fornecidos os dados originais e os dados previamente simulados. As expressões requeridas para as ccdfs condicionais podem ser obtidas por cokrigagem no caso de simulação conjunta de VAs com distribuições multivariadas gaussianas.

As principais desvantagens da metodologia de cossimulação gaussiana seqüencial são: requer inferência das covariâncias diretas e cruzadas e, às vezes, instabilidade dos sistemas de cokrigagem.

Almeida e Journel (1994) apresentaram uma metodologia que permite a cossimulação direta de várias FAs interdependentes, sem a inferência e modelagem da matriz completa de covariância cruzada, a partir de uma alteração da cossimulação seqüencial gaussiana. A abordagem se baseia em duas idéias chaves: a aproximação da cokrigagem *collocated* para reduzir o tamanho do sistema de cokrigagem; modelo de correionalização tipo Markov para modelagem das covariâncias cruzadas. Com a cokrigagem *collocated*, as ccdfs são obtidas seguindo o princípio da simulação seqüencial. O algoritmo considera a correlação espacial entre diferentes variáveis retendo para simulação de qualquer variável primária, o valor *collocated* de todas as demais variáveis previamente simuladas. Deve-se definir uma hierarquia de simulação das variáveis primárias indo da mais importante para a menos importante.

### 5.3 Simulação Indicatriz Seqüencial

O algoritmo da simulação indicatriz seqüencial, como o próprio nome sugere, segue também a abordagem seqüencial, só que nenhuma hipótese é feita sobre a forma da distribuição condicional. Estimativas da distribuição condicional em cada nó são obtidas por krigagem indicatriz, utilizando semivariogramas indicatrizes.

Como visto na KI, a geração da ccdf precede a estimativa.

$$F(x_0; z_k) = [I(x_0; z_k)]^* = E[I(x_0; z_k)|(n)]^* = P^*[Z(x_0) \leq z_k | (n)]$$

Na prática, em cada nó  $x$ ,  $K$  sistemas de krigagens indicatrizes devem ser resolvidos para gerar um versão discretizada da ccdf  $P[Z(x) \leq z_{k_0} | (n)]$ ,  $k_0 = 1, \dots, K$ . Semelhante ao algoritmo da SGS, na SIS também se faz necessário a utilização de algoritmos eficientes

baseados na vizinhança móvel para evitar problemas associados com o aumento do número de dados condicionantes.

Em cada nó, a ccdf obtida via KI pode apresentar os problemas de relação de ordem: ccdFs maiores do que um e menores do que zero, ou não monotônicas. São feitas correções nestes problemas, porém nenhuma pesquisa foi ainda realizada no sentido de verificar quais efeitos estas correções podem ter nas realizações.

Modelos de FAs indicatrizes são adequados para simulação de variáveis categóricas (ou contínuas discretizadas) controladas por estatísticas de dois pontos (covariância). A questão fundamental é saber se a estatística de dois pontos é suficiente para caracterizar a geometria da distribuição das classes ou categorias. Se a resposta for não, deve-se optar por algoritmos que levem em consideração estatísticas de maior ordem como *simulated annealing*, por exemplo.

As principais vantagens da SIS são: fácil condicionamento; não assumir, a priori, nenhum modelo de FA; flexibilidade de permitir diferentes modelos de continuidade espacial para cada parâmetro de corte; permite considerar informações *soft*.

As maiores desvantagens são: ocorrência eventual de viés nas frequências geradas com a simulação; impossibilidade de reproduzir os modelos de variograma de uma variável original contínua; geometrias bem definidas não podem ser reproduzidas; parametrização e inferências são pesadas, principalmente quando muitas indicatrizes estão sendo modeladas; torna-se computacionalmente lenta quando a correlação cruzada entre as indicatrizes é introduzida.

Na simulação indicatriz seqüencial, o uso da KI implica em desconsiderar a covariância cruzada entre as indicatrizes. Porém, as indicatrizes não são independentes. Por exemplo se  $Z(\mathbf{x}) \geq z_{k'}$ , então  $Z(\mathbf{x}) \geq z_k$  para todo  $k < k'$ . Esta dependência entre  $I(\mathbf{x}, z_{k'})$  e  $I(\mathbf{x}, z_k)$  se mantém entre pontos vizinhos ' $\mathbf{x}$ ' e ' $\mathbf{x}+\mathbf{h}$ ', pois  $Z(\mathbf{x}+\mathbf{h})$  tem pouca probabilidade de ser maior que  $z_{k'}$  se  $Z(\mathbf{x}) < z_k$ . O adequado seria utilizar a cokrigagem indicatriz, mas isso exige a inferência e modelagem das covariâncias cruzadas entre as indicatrizes.

Outras abordagens seqüenciais indicatrizes se encontram reportadas na literatura onde se varia o tipo de krigagem indicatriz a ser usada na determinação das ccdFs. Assim, muitos outros algoritmos de simulação de indicatrizes que usam o princípio seqüencial, não o incluem em seu nome, por exemplo a Simulação de Markov-Bayes (Zhu e Journel, 1993) e a

Simulação dos componentes principais indicatrizes (Suro-Pérez e Journel, 1990) que utilizam a krigagem *soft* com modelo de Markov-Bayes e a IPCK respectivamente.

## 5.4 Simulação Gaussiana Truncada

Journel e Isaaks (1984, apud Deutsch e Journel, 1996) apresentaram uma metodologia de simulação indicatriz baseada no truncamento do modelo gaussiano por um único *cutoff* gerando assim a distribuição de duas fácies complementares.

No software Heresim<sup>®</sup> (*heterogeneous reservoir simulation*) desenvolvido pelo Instituto Francês do Petróleo e pelo Centro de Geoestatística de Fontainebleau, a idéia prévia de um único truncamento foi ampliada para múltiplos truncamentos do mesmo campo gaussiano, com curvas de *cutoffs* que variam no espaço (Matheron et al., 1987).

A idéia básica da metodologia proposta por Matheron et al. (1987) é descrever  $n$  fácies usando uma função indicatriz para cada fácies. A indicatriz para a fácies  $F_i$ , definida por meio de uma FA gaussiana  $Y(x)$ , é:

$$I_{F_i}(x) = \begin{cases} 1 & \text{se } Y(x) \in ]y_{i-1}, y_i] \\ 0 & \text{senão} \end{cases}$$

Portanto um ponto  $x$  pertence a fácies  $F_i$  se  $Y(x) \in ]y_{i-1}, y_i]$ . Os  $y_i$  são denominados *cutoffs* gaussianos. A probabilidade para a FA gaussiana  $Y(x)$  ser menor que  $y_i$  e maior que  $y_{i-1}$  é  $P(y_{i-1} \leq Y(x) < y_i) = G(y_i) - G(y_{i-1})$ , onde  $G$  é a distribuição acumulada gaussiana. Se  $Y(x)$  está entre estes dois valores, por definição o ponto  $x$  pertence à fácies  $F_i$ . A probabilidade  $P(y_{i-1} \leq Y(x) < y_i)$  representa a proporção global de fácies  $F_i$  no campo. Observa-se que no caso contínuo, somente fácies  $F_{i-1}$  e  $F_{i+1}$  podem ser contíguas à fácies  $F_i$ .

Assim, a idéia de gerar classes, interpretadas como fácies geológicas é muito simples, sendo realizada por truncamento de uma curva contínua ou superfície.

Os níveis variáveis dos *cutoffs* permitem um controle direto na proporção de cada fácies gerada. O problema está em como controlar as estatísticas espaciais de ordem mais elevada das fácies geradas, por exemplo a probabilidade de transição de fácies  $k$  para fácies  $k'$  em função da separação  $h$ .

A simulação é obtida por uma variável truncada gaussiana como descrita nas seguintes etapas:

- (1) Calcular a proporção de  $p_j$  em cada categoria  $F_j$  dos dados e então determinar os *cutoffs* gaussianos  $y_j$ .

$$\begin{aligned}
y_1 &= G^{-1}(p_1) \\
y_2 &= G^{-1}(p_1 + p_2) \\
y_{N-1} &= G^{-1}(p_1 + p_2 + \dots + p_{N-1})
\end{aligned}$$

- (2) Obter as covariâncias cruzadas e diretas das indicatrizes (fácies).
- (3) Determinar a covariância  $C_Y(h)$  da variável gaussiana usando a relação:

$$C_{F_i F_j}(x, x+h) = \frac{1}{2\pi\sqrt{1-(C_Y(h))^2}} \int_{y_{i-1}}^{y_i} \int_{y_{j-1}}^{y_j} e^{-\frac{u^2+v^2-2uvC_Y(h)}{2(1-(C_Y(h))^2)}} dudv \quad (5.4.2)$$

onde  $C_{F_i F_j}(x, x+h)$  é a covariância cruzada entre as fácies  $F_i$  e  $F_j$  que corresponde à probabilidade de ocorrer fácies  $F_i$  na posição  $x$  e  $F_j$  na posição  $(x+h)$ .

- (4) Nas posições dos dados, gerar valores de uma variável gaussiana com a covariância  $C_Y(h)$  de forma que  $y_j \leq y(x_\alpha) < y_{j+1}$ , onde  $j$  é o índice correspondente à fácies  $F_j$ .
- (5) Simular uma variável gaussiana com covariância  $C_Y(h)$  e condicionar os valores simulados, em uma maneira usual, com os dados transformados na etapa 4.
- (6) Transformar cada valor gaussianamente simulado  $y_{cs}(x)$  para a sua categoria indicatriz condicionalmente simulada por:

$$I_{F_i}(x) = \begin{cases} 1 & \text{se } y_{i-1} < y_{cs}(x) \leq y_i \\ 0 & \text{senão} \end{cases}$$

A determinação de  $C_Y(h)$  é o principal aspecto computacional deste algoritmo. São calculadas todas as covariâncias diretas e cruzadas das diferentes fácies  $s$ . A covariância teórica da FA gaussiana é aceita se, e somente se, o ajuste resultante for satisfatório para as covariâncias diretas e cruzadas experimentais das fácies  $s$ . A equação 5.4.2 mostra que todas as covariâncias indicatrizes estão embutidas na covariância da gaussiana, o que leva a perda do controle espacial das fácies  $s$ .

Apesar desta falta de controle na estatística espacial de cada fácies específica, a idéia do truncamento apresenta as seguintes características positivas: velocidade, pois a simulação de  $n$  fácies no espaço é reduzida para a simulação de uma única gaussiana; sequenciamento de fácies  $s$ , importante principalmente em reservatórios flúvio-deltaicos; informações prévias sobre probabilidade da fácies  $F_i$  prevalecer em um determinada posição  $x$ , também chamadas curvas de proporções  $p_i = P[x \in \text{fácies } F_i]$ , podem ser transformadas em *cutoffs* locais

$y_i(x)$ , permitindo a utilização de dados *soft* e também a intervenção do geólogo com sua experiência na área; fácil condicionamento.

Uma característica interessante da definição de indicatrizes por truncamento é sua flexibilidade em lidar com não estacionariedades. Beucher et al. (1993) demonstraram a utilização do algoritmo de simulação truncado gaussiano com proporções não estacionárias tanto na vertical quanto na horizontal. Eles demonstraram como realizar a análise estrutural, que é possível porque a gaussiana é ainda estacionária, mesmo se as distribuições de fácies apresentarem fortes não estacionariedades.

#### 5.4.1 Simulação Plurigaussiana Truncada

Extensões para o truncado gaussiano foram desenvolvidas no sentido de transpor as limitações relativas a pouca flexibilidade nas transições entre fácies, e a impossibilidade de simular fácies com diferentes anisotropias.

A metodologia consiste em definir fácies por truncamento de duas ou mais funções aleatórias gaussianas que podem ser correlacionadas ou não. Esta extensão preserva o rigor e a capacidade de usar informações externas através de proporções. As propriedades e os exemplos fornecidos por Galli et al.(1994) demonstram o potencial deste método e sua habilidade em produzir realísticas imagens geológicas.

O modelo de FA plurigaussiana truncada é definida baseando-se:

- no número de funções gaussianas  $N(0,1)$ ;
- nas matrizes de covariâncias diretas e cruzadas, que completamente definem o modelo de funções gaussianas;
- na maneira de transformar um conjunto de funções gaussianas em uma única função discreta de fácies  $s$ . É caracterizada pela partição do plano definida pela funções aleatórias gaussianas. Uma fácies é atribuída a cada conjunto desta partição, cuja forma pode ser tão complexa quanto se queira;

Na determinação dos *cutoffs*, a partir dos dados de proporção de cada fácies, tem-se que estudar cuidadosamente com os geólogos as relações espaciais entre as mesmas, isto é, a probabilidade de contato direto, o ordenamento, etc. Isto assegurará uma adequada partição do espaço definido pelas funções gaussianas.

As funções gaussianas podem ser correlacionadas ou não. A escolha do coeficiente de correlação entre as FAs gaussianas afeta os cálculos dos *cutoffs* gaussianos e dos variogramas.

Um alto coeficiente introduzirá um ordenamento entre as fácies, enquanto contatos serão mais severos e desordenados quando o coeficiente de correlação é pequeno.

Devem ser calculados os variogramas diretos e cruzados de cada indicatriz de fácies  $s$ , conhecendo os coeficientes de correlação, os *cutoffs*, e os modelos de variogramas da FAs gaussianas. O ajuste consiste em comparar estes variogramas com os variograma experimentais diretos e cruzados das fácies. Semelhante ao truncado gaussiano, a idéia é iterativamente modificar os parâmetros, obtendo um ajuste satisfatório aos variogramas experimentais. Porém, com o método plurigaussiano, tem-se mais liberdade no ajuste, pois se dispõe de mais parâmetros. Por exemplo, a anisotropia pode diferir entre as duas FAs gaussianas, e até mesmo pode-se alterar o coeficiente de correlação entre as FAs reconstruindo o modelo desde o início.

## 5.5 Simulação Campo de Probabilidade

Simulação campo de probabilidade (*p-field*) foi introduzida por Srivastava (1992) e Froideveaux (1993). Recentemente tem se observado uma retomada da utilização desta abordagem em conjunto com a krigagem fatorial para a simulação de atributos, usando informações obtidas de levantamentos sísmicos. A estimativa das ccdfs locais é separada da amostragem destas distribuições. Assim, a abordagem *p-field* oferece um método de simulação estocástica que é fácil de entender. É um algoritmo simples que pode pós processar os resultados dos procedimentos de estimativa ou de outras simulações estocásticas para produzir realizações prováveis.

Resumidamente, a implementação da simulação *p-field* envolve as seguintes etapas:

1. Definir um *grid* cobrindo a área de interesse;
2. Obter para cada nó do *grid*, a ccdf local  $F(x,z)$  do atributo a ser simulado. As ccdfs podem ser resultado de um processo de simulação estocástica ou de krigagem, ou definidas empiricamente;
3. Calcular o modelo de variograma da transformada uniforme do atributo a ser simulado. Assumir que a covariância de  $U(Z(x))$  e do campo de probabilidades  $P(x)$  são similares:  $C_u(h) \approx C_p(h)$ ;
4. Assumir que o campo de probabilidade  $P(x)$  segue uma distribuição uniforme;
5. Gerar, em cada nó do *grid*, uma simulação estocástica não condicional de  $P(x)$  honrando a distribuição uniforme prévia e a função covariância  $C_p(h)$ ;

6. Em cada nó do *grid*, obter uma realização  $z_s(x)$  da cdf local usando o valor do campo de probabilidade  $p(x)$ :  $z_s(x) = F^{-1}[x, p(x)]$ . (Figura 5-2);
7. Repetir os passos 5 e 6 para cada imagem a ser gerada.

Uma etapa crucial na simulação do campo de probabilidade (*p-field*) é estabelecer as distribuições de probabilidades condicionais em cada posição em que um valor simulado é requerido. As distribuições de probabilidades condicionais locais são simplesmente as possíveis realizações com suas respectivas probabilidades.

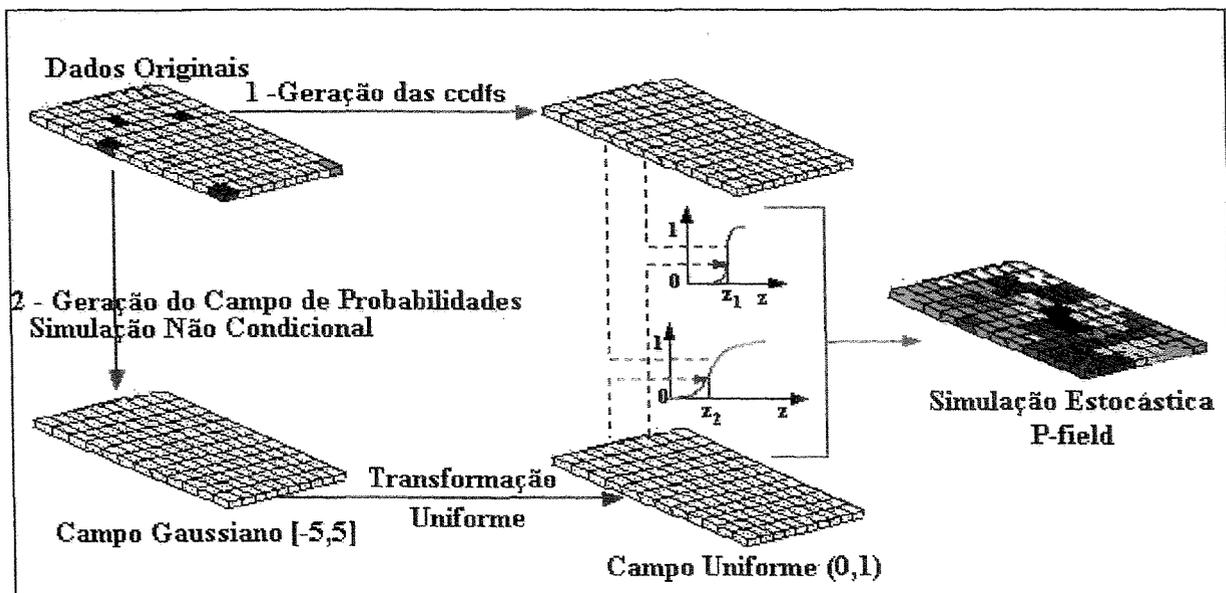


Figura 5-2: Simulação campo de probabilidade.

Embora a simulação campo de probabilidade não controle como as cdfs são construídas, a relevância dos resultados depende da relevância das mesmas. Um outro ponto crítico é a correta inferência das características univariadas e bivariadas da distribuição de  $P(x)$ , isto é, a forma da distribuição de probabilidade e suas características de continuidade espacial. Assim, a direção de máxima continuidade, alcance, razão de anisotropia e efeito pepita relativo, são similares para a transformada uniforme do atributo  $Z(x)$  e para seu campo de probabilidade  $P(x)$ . A prática comum é gerar um campo de probabilidade que tem uma distribuição global uniforme entre 0 e 1 e com suas autocorrelações se ajustando à transformada uniforme das variáveis a serem simuladas.

As cdfs locais colapsam em uma distribuição de largura zero para as posições das amostras; qualquer percentil retido da distribuição honrará necessariamente o valor original da amostra.

Posto que a FA da qual os valores de probabilidade devem ser retirados é certamente uniforme em cada posição, a distribuição dos valores de probabilidade dentro de um campo finito pode não ser uniforme. Froidevaux (1993) mostrou que a distribuição observada dos valores do campo de probabilidade depende do tamanho do campo que está sendo simulado em relação ao alcance. Para campos muito maiores que o *integral range*, a distribuição  $p(x)$  dentro do campo será muito mais próxima da distribuição uniforme. À medida que o *integral range* torna-se semelhante ao tamanho do campo, a distribuição dos valores de probabilidade se tornam menos uniformes.

A criação de um campo de probabilidade pode ser realizada de várias maneiras. O método mais comum é gerar uma simulação gaussiana não condicional com uma função de correlação igual a dos dados originais condicionantes. Só que não se completa a simulação fazendo a transformação de volta para o campo da variável simulada, mas realiza-se uma transformação dos valores simulados normalmente distribuídos  $N(0,1)$  para valores uniformes, por exemplo, usando aproximações polinomiais para a função de distribuição gaussiana acumulada disponíveis na literatura.

Segundo Srivastava (1992), umas das possibilidades para futuras investigações é a possibilidade de gerar diretamente realizações de um campo multivariado uniforme em vez de utilizar um campo intermediário gaussiano.

Uma das principais vantagens da abordagem *p-field* para a simulação é sua velocidade computacional. O gargalo do tempo computacional está no cálculo das ccdfs locais. Pode-se escolher um algoritmo tão rápido quanto se queira para gerá-las.

Uma vez calculadas e armazenadas as ccdfs, diferentes realizações podem ser criadas tão rapidamente quanto se cria um campo de probabilidade. Uma outra vantagem da abordagem é que ela separa o problema de estimar ccdfs locais do problema de criar uma simulação condicional.

A técnica pode ser aplicada para qualquer conjunto de ccdfs locais, não importando como foram obtidas, podendo ser inclusive de métodos não geoestatísticos. Srivastava (1992) fornece um exemplo em que as distribuições são obtidas a partir da interpretação sísmica, produzindo vários mapas do topo de um reservatório. Esta flexibilidade é uma das principais vantagens práticas da abordagem *p-field*. Consequentemente, a simulação *p-field* permite uma fácil integração de dados *hard* muito esparsos com exaustivas informações *soft*.

Devido a forma como o campo de probabilidade interage com a ccdf local, verifica-se que os dados condicionantes formam extremos locais. A razão para este artefato é devido ao

fato de que um dado condicionante será muito próximo da mediana da cdf de seus vizinhos imediatos. O método da simulação *p-field* também se aplica a variáveis categóricas. Na verdade, ele trabalha melhor com variáveis categóricas, pois o artefato descrito não faz sentido nesta situação.

Embora falte à simulação *p-field* uma base teórica claramente articulada, vários estudos de caso documentam seu sucesso prático. Entretanto, é fundamental um aprofundamento teórico do método para verificar sobre quais condições ele não trabalha adequadamente. De qualquer forma, o usuário deve checar se a realização honra a distribuição e o padrão de continuidade espacial.

## 5.6 *Simulated Annealing*

Na abordagem *annealing* para simulação estocástica, a criação de realizações é formulada como um problema de otimização a ser resolvido com relaxação estocástica ou técnica *annealing*. A técnica tem o potencial de combinar a reprodução da estatística de dois pontos com estatísticas multipontos implícitas, por exemplo, em formas geométricas. A técnica de otimização mais frequentemente usada para obter tais realizações é baseada no processo físico de *annealing* (recozimento).

A idéia básica do *simulated annealing* (SA) é perturbar continuamente uma imagem inicial até que ela se ajuste a algumas características preestabelecidas, escritas como componentes de uma função objetivo (FO).

A função objetivo (ou energia) é uma medida da diferença entre as características espaciais desejadas e as da realização candidata. A característica básica dos métodos de relaxação estocástica é iterativamente perturbar (relaxar) a realização candidata e então aceitar ou rejeitar a perturbação segundo alguma regra de decisão. A regra de decisão é baseada em quanto a perturbação traz a imagem candidata para mais próximo das propriedades desejadas.

O real interesse do *simulated annealing* está na sua habilidade de melhor incorporar complexas estruturas geológicas e dados de engenharia.

As simulações estocásticas convencionais estão limitadas a estatística de dois pontos. *Simulated annealing* não tem esta limitação: a função objetivo pode ser definida com componentes relativos a propriedades geológicas, estatísticas ou de engenharia.

Em geral a função objetivo (FO) é uma soma ponderada de C componentes:

$$FO = \sum_{c=1}^C w_c O_c$$

onde  $w_c$  e  $O_c$  são os pesos e os componentes da função objetivo, respectivamente. Um componente pode ser, por exemplo, a medida da diferença entre um variograma modelado a partir dos dados experimentais e um outro obtido a partir da realização. Um segundo componente pode ser uma medida de reprodução de uma particular estatística multiponto. Um terceiro componente poderia medir a fidelidade em relação a uma propriedade equivalente derivada de testes em poços, envolvendo muitos valores simulados simultaneamente.

Os pesos são estabelecidos de forma que, em média, cada componente contribua igualmente para a mudança na função objetivo ou da forma que o usuário desejar. Por exemplo, pode-se usar pesos diferentes, quando se deseja que uma realização obedeça mais a um dos componentes.

A taxa de decrescimento de cada componente não é considerada pelos pesos. Um componente pode rapidamente ir para zero, enquanto outro ainda está muito longe. Consequentemente, os pesos precisam ser reavaliados periodicamente durante a simulação.

Muitas variações do mesmo algoritmo básico são possíveis, considerando funções objetivo, caminhos para criar a imagem inicial, mecanismos de perturbação e regras de decisão diferentes.

As técnicas *annealing* baseiam em muitas, freqüentemente milhões de perturbações para atingir uma realização aceitável. Isto implica que cada componente da função objetivo (FO) deva ser razoavelmente simples e que a FO não tenha muitos componentes, nem componentes conflitantes.

Vários são os mecanismos de perturbação no *simulated annealing*. Por exemplo, pode-se optar por trocar os valores de dois nós escolhidos aleatoriamente, modificar o valor de um nó de acordo com uma função de distribuição de probabilidade local, etc.

A perturbação é aceita ou rejeitada, baseando-se em uma regra de decisão. Uma possibilidade é rejeitar todas as perturbações que levem a um aumento na função objetivo. Isto corresponde a abordagem *steepest descent*.

Considerando o algoritmo de Metropolis, se a FO após a perturbação é maior que a FO anterior, ela será aceita ou rejeitada de acordo com a probabilidade  $P(\Delta FO) = \exp((FO_{ant} - FO_{após}) / T)$ , onde T é análogo à temperatura do processo real do *annealing* e tem a mesma unidade da função objetivo. Computacionalmente este passo consiste em gerar um número aleatório  $z$  no intervalo  $[0,1]$ . Se  $z < \exp((FO_{ant} - FO_{após}) / T)$ , então a perturbação é aceita; do contrário é rejeitada.

Com temperatura alta constante, o erro total permanece grande, pois muitas perturbações serão aceitas. Por outro lado, temperatura baixa constante leva a solução ser ‘congelada’ em um mínimo local. Disto resulta que se faz necessário uma adequada programação de ‘resfriamento’. O sucesso do método depende de um lento ‘resfriamento’ das realizações controlado pela temperatura que deve decrescer com o tempo. Ressalta-se, porém, que se a temperatura for reduzida muito lentamente, a convergência para uma imagem ótima pode ser bastante lenta. Tem assim que ser definida a programação (*schedule*) do *annealing* que corresponde quando e quanto reduzir a temperatura. Deutsch e Journel (1992) usaram uma programação de *annealing* empírica, cuja idéia básica é começar com uma temperatura elevada e reduzi-la, multiplicando por um fator de redução  $\lambda$  ( $0 < \lambda < 1$ ), quando muitas perturbações tiverem sido aceitas, ou muitas tiverem sido tentadas. O algoritmo é interrompido quando mais esforços para reduzir a função objetivo ficarem desencorajadores ou já se atingiu um valor aceitável para a função objetivo.

*Simulated annealing* apresenta artefatos conhecidos como efeitos de borda que ocorrem principalmente quando se utiliza uma programação rápida de ‘resfriamento’ ou quando a distribuição univariada é assimétrica. Os valores extremos tendem a se concentrar nas bordas, porque assim participam de menos pares quando da atualização dos componentes da função objetivo.

Uma simples maneira de fazer com uma realização com a SA honre as informações locais *hard* é nunca perturbá-las. Esta solução, entretanto, introduz discontinuidades próximas aos dados condicionantes. Isto ocorre porque os dados condicionantes passam a não contribuir igualmente para a função objetivo global desde que eles não têm chance de serem perturbados. Uma correção proposta por Deutsch e Cockerham (1994) corresponde basicamente em dar maior peso aos pares que tem pelo menos um dado condicionante.

A principal vantagem da SA é que ela pode incorporar na função objetivo qualquer tipo de informação que estiver disponível sobre o reservatório.

Uma das principais desvantagens é o fato de ser uma metodologia ainda relativamente nova para a área de petróleo. Há pouca literatura que forneça detalhes de como preparar uma função objetivo adequada e principalmente fazer a programação (*schedule*) do *annealing*. Uma outra desvantagem é que o algoritmo torna-se cada vez mais lento, à medida que aumenta a complexidade da função objetivo.

A grande desvantagem teórica da SA é a falta de um arcabouço matemático consistente. O algoritmo básico é um procedimento de otimização e isto leva a muitos questionamentos sobre o espaço de incerteza que está sendo explorado.

## ***5.7 Simulações Baseadas em Objetos***

As técnicas baseadas em objetos modelam a distribuição de objetos geológicos com geometria bem definida, tais como canais fluviais ou turbidíticos, frente deltaica, barras de desembocadura, lóbulos arenosos ou falhas. Os objetos são geralmente descritos com formas muito simples, tais como blocos retangulares ou elipsóides. A variabilidade geológica é modelada através da distribuição das formas, tamanhos e orientações dos objetos, cujas posições são aleatórias dentro do reservatório.

A idéia básica é que o modelo de reservatório deve ser criado a partir de objetos que têm algum significado genético. É selecionada para cada litofácia, a forma básica que adequadamente descreve sua geometria. Por exemplo, as areias de um canal em um sistema turbidítico parecem meias elipses, ou as de leques deltaicos parecem cunhas triangulares.

Atributos da forma (comprimento, largura, altura, orientação) são caracterizados por distribuições que podem ser independentes ou correlacionadas. Estes atributos podem ser dependentes da posição para levar em consideração tendências geológicas conhecidas. A posição da distribuição é, em muitos casos, puramente aleatória: posições dos centróides são independentes umas das outras, e a densidade do objeto (média de objetos por unidade de volume) é constante. Nestes casos, as posições dos centróides são ditas seguir um processo pontual de Poisson estacionário e o modelo estocástico resultante é o modelo booleano.

Haldorsen e MacDonald (1987) publicaram um dos primeiros trabalhos com este tipo de abordagem de simulação estocástica, no qual eles modelaram a distribuição de corpos de areia em um reservatório de ambiente fluvial. Esses autores consideraram os canais como barras retangulares dispersas em uma matriz argilosa.

O modelo booleano também pode ser não estacionário para considerar as variações da densidade de objetos no reservatório. Modelagem de objetos é melhorada levando em conta a interação entre as posições dos objetos: fatores de repulsão ou atração são introduzidos no modelo probabilístico da posição dos centróides. Tais modelos permitem gerar vários canais dentro de um cinturão de canais, ou associações espaciais de diferentes tipos de objetos geológicos, por exemplo, associações canais-*crevasses*. Assim, alguns algoritmos sofisticados permitem escolher as regras que descrevem como as várias formas são posicionadas

relativamente umas às outras; estas regras podem exigir que determinada forma não possa ocorrer dentro de uma distância mínima de outras formas, ou que certas formas devam sempre estar justapostas.

Numerosas melhorias tem sido propostas desde Haldorsen e MacDonald (1987) focalizando principalmente os seguintes aspectos: a interação dos objetos, a sinuosidade dos objetos em terceira dimensão, condicionamento aos dados dos poços.

Escolhidos os parâmetros da distribuição e as regras de posicionamento, o seguinte algoritmo simplificado captura as etapas essenciais no procedimento de modelagem estocástica baseada em objetos:

1. Preencher o modelo de reservatório com alguma litofácia de fundo (por exemplo, folhelho);
2. Aleatoriamente escolher uma posição no modelo de reservatório;
3. Aleatoriamente escolher uma das formas de litofácies e selecionar das distribuições desta forma, um tamanho apropriado e orientação;
4. Checar se há conflito entre a forma escolhida e algum dado condicionante ou com outras formas que foram previamente simuladas. Se não houver conflito, aceitar a forma. Caso contrário, rejeitá-la e voltar ao passo 3;
5. Checar se todas as proporções globais das várias formas foram alcançadas; se não voltar ao passo 2.

Os dados de poços devem ser honrados primeiro. Quando a região entre poços é simulada, deve-se tomar cuidado para evitar conflitos com a seqüência conhecida de litologias nos poços. As muitas implementações desta abordagem diferem bastante em como assegurar que as regras de posicionamento relativo são honradas. Alguns algoritmos simplesmente rejeitam o objeto quando ele não se ajusta às regras, outros tentam ser mais eficientes e checam se pequenas alterações no reposicionamento ou no tamanho do mesmo, permitem que as regras de posicionamento sejam obedecidas.

Os métodos baseados em objetos são fáceis de entender, estão mais próximos do ‘sentimento’ geológico, isto é, apresentam um maior apelo visual do que as técnicas baseadas em *pixel*.

As técnicas baseadas em objetos apresentam algumas desvantagens. A limitação mais crítica é relativa ao condicionamento aos dados de poços que é muitas vezes manuseado inadequadamente. Isto resulta em sérios problemas de implementação, particularmente quando o espaçamento dos poços é menor do que a dimensão dos objetos e vários poços podem

interceptar o mesmo objeto. As formas dos objetos são, às vezes, irrealisticamente simples. Inferência dos parâmetros requeridos é impossível a partir de dados de poços geralmente esparsos. Usuários devem então confiar no altamente incerto e, às vezes, enviesado conjunto de dados de afloramento, supostamente de formação similar. Regras de erosão entre objetos necessitam ser especificadas podendo implicar em problemas na implementação. Por último, modelos com interações requerem algoritmos iterativos que podem convergir muito lentamente quando o domínio for grande com elevado número de poços.

No último Congresso de Geoestatística (Wollongong'96), alguns trabalhos de grupos de pesquisadores noruegueses, como por exemplo Johnsen et al. (1997), reportaram a utilização da sísmica na simulação estocástica baseada em objetos. Porém, nenhuma referência foi encontrada na literatura a respeito da utilização de dados dinâmicos, isto é, dados de produção, testes de formação, etc. Por exemplo, se dois intervalos em um mesmo poço ou em poços diferentes, reconhecidamente não pertencentes a um mesmo canal, estão hidraulicamente conectados, informação obtida dos dados de testes ou de produção, o modelo booleano deveria de alguma forma obedecer a essa restrição e construir um caminho contínuo para permitir a ocorrência de fluxo entre os dois intervalos, seja através da geração de objetos de mesmo tipo (canais justapostos), ou através de geração de objetos de tipos diferentes (associação canais-*crevasses*), mas ambos caminhos devem ser permeáveis. Mais importante que a geração de unidades genéticas ou corpos geológicos é que os modelos resultantes desta abordagem, quando submetidos à simulação de fluxo sejam representativos e coerentes. Um modelo que não inclui a comunicação hidráulica entre os referidos intervalos, quando se tem esta informação a priori, é um modelo incorreto para o reservatório.

## 6. Incertezas na Simulação Estocástica

Grande parte do crescimento da popularidade das simulações estocásticas condicionais na indústria do petróleo é devido ao fato de que elas acrescentam um apelo de realismo visual ao modelo. Por vezes, esse realismo visual obscurece o entendimento dos engenheiros e geólogos sobre as incertezas envolvidas.

A simulação estocástica fornece maneiras de incorporar vários tipos de incertezas na previsão de performance de um reservatório. Imagens estocásticas, obtidas via algum algoritmo de simulação geoestatística, servem como parâmetros de entrada para o simulador de fluxo, obtendo-se assim a previsão de comportamento do reservatório para cada uma das realizações. Como as realizações caracterizam a incerteza espacial do parâmetro de interesse, a distribuição dos resultados (previsão de performance) refletem esta incerteza. Várias são as respostas que podem ser analisadas em cada uma das realizações: tempo de *breakthrough*, produção acumulada de água e hidrocarbonetos, fator de recuperação, etc.

Decisões gerenciais importantes tais como a localização de poços horizontais, estratégia de desenvolvimento de um campo de petróleo podem ser baseadas na distribuição de incerteza da performance do reservatório.

Ressalta-se que para um mesmo algoritmo, diferentes softwares poderão levar a diferenças consideráveis nas imagens geradas. Para o usuário da simulação estocástica, as diferenças entre os diversos algoritmos e suas diferentes implementações, e os efeitos na incerteza da previsão de performance do reservatório necessitam ser melhor entendidos para que possa utilizar o algoritmo e implementação adequados à situação que estiver analisando.

Muito pouco tem sido feito para comparar os algoritmos que estão disponíveis nos softwares comerciais. Cada metodologia tem seus próprios defensores que, a priori, são os seus criadores, e como seus idealizadores, reforçam apenas as suas vantagens.

Neste capítulo, diversos temas referentes a incertezas na simulação estocástica serão abordados. A questão principal é fomentar o debate sobre incertezas na modelagem.

Geoestatísticos e usuários da geoestatística deveriam ser extremamente criteriosos quando assumirem que podem quantificar incertezas, especialmente se poucos dados são

disponíveis. É imprescindível que todas as hipóteses feitas para a escolha do modelo sejam identificadas, uma vez que as mesmas podem estar erradas, e isto ser observável quando mais dados se tornarem disponíveis. Ressalta-se que a principal incerteza pode não estar na variabilidade entre as realizações de um modelo fixo, com parâmetros fixos, mas na validade do próprio modelo ou dos parâmetros deste.

## ***6.1 Comparação de Algoritmos***

Algumas poucas publicações comparam alguns algoritmos de simulação estocástica, sem contudo entrar nas diferenças de implementação a nível de software. Por exemplo, a SGS implementada no ISATIS e na GSLIB apresentam diferentes parâmetros que podem ser manuseados pelo usuário. Mesmo baseando-se nos mesmos parâmetros e metodologia, diferenças nos softwares poderão levar a resultados diferentes.

Journal e Deutsch (1993) compararam a SIS e a SGS chegando à conclusão de que, em determinadas situações não completamente definidas e delimitadas, o algoritmo de simulação seqüencial gaussiana apresenta distribuição de incerteza não acurada, isto é, que não contém o valor real. Esses autores estenderam estes resultados a todos os métodos gaussianos. Journal e Deutsch (1993) concluíram ainda que a modelagem seqüencial indicatriz é imprecisa no sentido de gerar distribuições de respostas com alto grau de variabilidade, mas esta imprecisão permite acurácia no sentido de que as distribuições incluem o valor verdadeiro. Para Journal e Deutsch (1993) a máxima entropia do modelo de FA não implica necessariamente em máxima entropia da distribuição de resposta.

Esta afirmação de Deutsch e Journal (1993) em relação às simulações gaussianas pode ser encontrada em vários outros trabalhos destes e de outros autores. Curiosamente, no último congresso de Geoestatística, Journal (1997) apresentou um trabalho enfocando o princípio da parcimônia, isto é, para um dado objetivo, deve-se utilizar o modelo com o menor número de parâmetros e mais fáceis de serem inferidos. Esta colocação aparentemente está contradizendo a posição do referido autor em relação às simulações gaussianas e em defesa da abordagem indicatriz. O modelo gaussiano é bastante parcimonioso e requer apenas dois parâmetros fáceis de serem inferidos. Contudo, Journal (1997) reporta que o critério de modelo parcimonioso não deve ser seguido às custas de inconsistência dos dados, fazendo uma referência clara aos modelos gaussianos. Como formador de opinião, as colocações de Journal e outros a respeito da simulação gaussiana podem confundir o usuário de simulação estocástica.

Um outro interessante trabalho comparativo de alguns algoritmos de simulação foi realizado por Gotway e Rutheford (1994) que utilizaram diferentes conjuntos de dados de forma a cobrir uma diversidade de situações. Entre os vários resultados e conclusões obtidos por estes autores devem ser destacados: algoritmos de Bandas Rotativas e Decomposição LU apresentam menos agrupamentos de valores similares (*clusters*) do que o algoritmo de simulação gaussiana seqüencial; algoritmos de simulação estocástica que usam os mesmos parâmetros (por exemplo mesmos semivariogramas) podem resultar em distribuições de incerteza bastante diferentes. Gotway e Rutheford (1994) chegaram a resultados de distribuição de incerteza mais acurados para os métodos gaussianos se comparados com a simulação indicatriz, diferentemente do obtido por Journel e Deutsch (1993).

Os trabalhos que reportam as deficiências da SGS e das simulações gaussianas geralmente se baseiam em distribuições que apresentam fortes correlações de valores extremos (por exemplo baixas e altas permeabilidades). De concreto, esta é a única limitação a utilização do modelo gaussiano. Estudos devem ser conduzidos no sentido de melhor delimitar as fronteiras de utilização do modelo gaussiano.

## **6.2 *Imagens equiprováveis***

Uma importante questão é a equiprobabilidade das imagens. Sabe-se que o número de imagens possíveis é, via de regra, exorbitante. Em trabalho recente, Srivastava (1997) aplicou diversas técnicas de simulação em um *grid* 5x4 com uma variável binária e uma determinada função de correlação. Para a situação em questão, só eram possíveis 276 imagens diferentes. Srivastava (1997) então realizou 10000 simulações com diferentes algoritmos e verificou que em algumas técnicas de simulação estocástica, as 276 imagens não eram equiprováveis.

Admitindo-se que elas fossem equiprováveis, então em um conjunto de 10000 realizações, cada uma tem a probabilidade de 99% de ocorrer pelo menos 18 vezes e não mais que 54 vezes. Os resultados listados na tabela 6-1 indicam que para algumas abordagens de simulação estocástica não se verifica a equiprobabilidade das realizações, como por exemplo a seqüencial indicatriz.

Se faz necessário desenvolver teoria e procedimentos práticos que verifiquem a equiprobabilidade das realizações. Em situações reais, corre-se o risco do algoritmo de simulação estocástica privilegiar um determinado conjunto de imagens que, por exemplo, pode

ser constituído de imagens bastantes similares e, portanto, não cobrirem o espaço de incerteza, resultando em distribuições de incertezas não representativas.

**Tabela 6-1: Resultados obtidos por Srivastava (1997)**

Tipo de Simulação	Realização menos comum		Realização mais comum	
	Número	Frequência	Número	Frequência
<i>Annealing</i> Clássico	79	19	217	54
<i>Annealing</i> 2 *	107	11	11	175
<i>Annealing</i> 3*	214	7	218	108
Seqüencial Indicatriz	51	7	262	85
Seqüencial Gaussiana	40	20	184	58

Obs: \* refere-se a diferentes implementações da *simulated annealing*.

### 6.3 Escolha de um Algoritmo

Nenhum algoritmo de simulação é o melhor, qualquer que seja o conjunto de dados disponíveis e as situações que queremos simular. Como forma de fornecer uma diretriz básica para a escolha de um algoritmo de simulação, na tabela 6.2 estão os principais métodos de simulação com as principais características, hipóteses, parâmetros de entrada. Nesta tabela estão outras metodologias não apresentadas no capítulo anterior como a decomposição LU, campos aleatórios markovianos, simulação gaussianas fractais e algoritmo genético, que foram analisadas no decorrer do trabalho mas, devido ao caráter da dissertação, procurou-se centrar nas metodologias mais usadas. O conhecimento das técnicas disponíveis, um claro entendimento de suas limitações, e o impacto potencial na descrição do reservatório são indiscutivelmente imprescindíveis para o sucesso do trabalho de modelagem.

Em qualquer estudo, os resultados só são representativos quando as metodologias são aplicadas corretamente. Muitas vezes uma metodologia menos adequada e flexível, porém cuidadosamente utilizada, com reconhecimento e entendimento das limitações, pode dar melhores resultados do que uma metodologia mais adequada e abrangente, porém mal utilizada. Ressalta-se que todos os algoritmos de simulação tentam reproduzir um modelo de reservatório que é intrinsecamente o resultado de processos complexos que muitas vezes não são considerados pelos mesmos. Para os algoritmos baseados em *pixel* isto é bastante óbvio. Galli et al. (1997) ressalta que isto também é válido para as metodologias baseadas em objetos: “Por exemplo, canais não são simples objetos, eles são o resultado da evolução temporal de uma canal primário sujeito a condições variáveis (aporte de sedimentos, erosões, avulsões, etc)”. O essencial é o entendimento de que qualquer que seja a abordagem escolhida para a modelagem do reservatório, importantes hipóteses e simplificações serão adotadas.

Tabela 6-2: Algoritmos de Simulação: Parâmetros de entrada, Hipóteses, Características.

Método/Atributo	Entrada	Hipóteses	Características
Gaussiano Bandas Rotativas (Contínuo)	Variogramas	Multivariada gaussiana; estacionariedade de 2ª ordem.	Velocidade; primeiro método de simulação, portanto muito difundido; artefatos de lineamento; dificuldade de manusear anisotropias; restrição a determinados modelos de covariâncias; condicionamento em separado; simplicidade do modelo gaussiano.
Gaussiano Decomposição LU (Contínuo)	Variogramas	Multivariada gaussiana; estacionariedade de 2ª ordem.	Velocidade; facilidade de implementação; limitado a pequenos <i>grids</i> ; imprecisões numéricas se a matriz de covariância for esparsa; simplicidade do modelo gaussiano.
Gaussiana Sequencial (SGS) (Contínuo)	Variogramas	Multivariada gaussiana; estacionariedade de 2ª ordem	Velocidade; simplicidade do modelo gaussiano.
Gaussiano Co-SGS (Contínuo)	Variogramas diretos e cruzados	Multivariada gaussiana; estacionariedade de 2ª ordem	Simulação de atributos interdependentes; simplicidade do modelo gaussiano; instabilidade dos sistemas de cokrigagem; inferência de um elevado número de variogramas
Gaussiano Co-SGS <i>Collocated</i> (Contínuo)	Variogramas	Multivariada gaussiana; estacionariedade de 2ª ordem	Simulação de atributos interdependentes; estabilidade do sistema de cokrigagem; simplicidade do modelo gaussiano; necessidade de definir hierarquia entre as variáveis; dados das variáveis secundárias conhecidos em todo o domínio.
Gaussiano Fractal (Contínuo)	Variograma de potência; dimensão fractal.	Multivariada gaussiana; autossimilaridade.	Dependência espacial em todas as escalas; dificuldade de incorporar dados <i>soft</i> ; condicionamento é uma etapa em separado; limitado ao variograma de potência
Gaussiano Truncado Gaussiano (Contínuo ou discreto)	Variogramas, cur- vas de proporções verticais e hori- zontais	Estacionariedade de 2ª ordem da gaussianas.	Sequenciamento de fácies; velocidade; utilização de dados <i>soft</i> ; permite intervenção do geólogo; modelo teórico bem fundamentado; habilidade de lidar com não estacionariedades; impossibilidade de gerar fácies com diferentes anisotropias e complexas transições entre as mesmas; complexidade na determinação do variograma da gaussiana (método iterativo).
Plurigaussiano Truncado(Contínuo ou discreto)	Variogramas, pro- porções verticais e horizontais	Estacionariedade de 2ª ordem das gaussianas.	Simulação de fácies com diferentes anisotropias e complexas transições; complexidade na determinação do variograma da gaussiana (método iterativo).
Sequencial Indicatriz (SIS e derivações) (Contínuo ou discreto)	Variogramas das indicatrizes; a dependem da abor- dagem variograma cruzado entre as indicatrizes.	Estacionariedade de 2ª ordem.	Não assumir um modelo prévio de distribuição; permitir diferentes modelos de continuidade para os diferentes <i>cutoffs</i> ; permitir utilizar informações <i>soft</i> ; problemas de relação de ordem; impossibilidade de reproduzir variogramas de variáveis contínuas; parametrização e inferência bastante pesadas; lentidão se acrescentar o semivariograma cruzado;

Continuação da Tabela 6-2: Algoritmos de Simulação: Parâmetros de entrada, Hipóteses, Vantagens, Desvantagens.

Método/Atributo	Entrada	Hipóteses	Características
Simulated <i>Annealing</i> (Contínuo ou discreto)	Variogramas; estatísticas multiponto; proporções, etc...	Dependentes da função objetivo utilizada.	Capacidade de modelar complexas estruturas geológicas; pode incluir estatística multiponto; pode ser usada como um pós-processador; pode incorporar informações <i>soft</i> e informações em diferentes suportes (testes e sísmica); ocorrência de artefatos (efeitos de borda); descontinuidade próxima aos dados condicionantes; pouca informação a respeito da escolha da programação de resfriamento e como preparar a função objetivo; lentidão se a função objetivo for complexa; falta de arcabouço teórico consistente; dúvidas sobre o espaço de incerteza coberto.
Algoritmo Genético (Contínuo ou discreto)	Idêntico ao SA	Dependente da função de adaptação escolhida.	Idêntico ao SA; permite a paralelização; pouca informação sobre a definição das taxas de probabilidade de mutação, <i>crossing-over</i> e seleção; elevada dependência da população inicial; elevado tempo CPU.
Métodos Iterativos (Contínuo ou discreto)	Variogramas e histogramas	Estacionaridade de 2ª ordem.	Velocidade; convergência não é garantida; realizações muito similares (espaço de incerteza); possibilidade de não honrar os histogramas.
Campo de Probabilidade (Contínuo ou Discreto)	Variograma; ccdfs geradas por outros algoritmos geoestatísticos ou não geoestatísticos.	Similaridade entre o variograma do campo uniforme e o de probabilidades; hipóteses relacionadas ao método de geração das ccdfs.	Rapidez; possibilidade de animação de incerteza; fácil integração de dados <i>soft</i> ; pode ser usada como um pós-processador; possibilidade de utilizar ccdfs geradas por algoritmos não geoestatísticos (sísmica); separa a estimação das ccdfs da amostragem da mesmas; forte dependência da qualidade das ccdfs obtidas previamente; dados condicionantes tornam-se extremos locais (variável contínua); falta de um embasamento teórico bem fundamentado.
Campos Markovianos (Contínuo ou discreto)	Probabilidades condicionais (de mudança)	A ocorrência de um estado depende basicamente das fácies que ocorrem em uma determinada vizinhança.	Capacidade de modelar complexos padrões de heterogeneidades; facilidade de implementação de associação de fácies; parâmetros de entrada difíceis de serem manipulados por um não especialista; inferência dos parâmetros a partir de dados esparsos é difícil; difícil condicionamento para dados <i>soft</i> ; complexidade do modelo aumenta exponencialmente com o aumento do número de fácies e do tamanho da vizinhança.
Booleanos (Discreto)	Atributos de forma e orientação dos objetos; curvas de proporção; regras de atração, repulsão e erosão.	Representatividade dos dados de afloramento ou reservatório similar; os poços interceptam os objetos em posições aleatórias; formas bem definidas para os objetos.	Simulação de geometrias complexas; grande apelo visual; parâmetros de entrada são mais próximos do trabalho do geólogo; em algumas implementações, o conhecimento geoestatístico exigido é mínimo (sic); formas dos objetos bastante simplificadas; dependência da representatividade dos dados de afloramento; reduzida velocidade quando o conjunto de dados condicionantes é elevada; dificuldade de definir as regras de interação entre os objetos por um não especialista (algoritmos mais complexos); dificuldades em integrar dados oriundos da sísmica e de testes.

## 6.4 Representações de Incertezas

Várias opções são convencionalmente usadas para apresentar a informação da incerteza de um conjunto de realizações possíveis.

Neste tópico serão abordadas algumas figuras de incerteza que são mais adequadas às propriedades do reservatório consideradas aditivas, por exemplo, relativas ao volume de rocha, ou poroso, ou de hidrocarbonetos.

Com o intuito de utilizar as diversas representações de incertezas oriundas das simulações estocásticas, será utilizado o mesmo conjunto de dados da AMOCO (74 poços) que deram origem ao caso base na análise da incerteza na krigagem.

Portanto será utilizada a variável  $H\phi$  (produto espessura x porosidade), mas poderia ter sido a espessura, a porosidade, a saturação de fluidos ou outra combinação das mesmas.

Inicialmente foi utilizado o algoritmo da SGS condicional do ISATIS para se obter 100 realizações do atributo  $H\phi$ . A primeira idéia que se tem de incerteza quando da utilização da simulação estocástica é a diferença apresentada entre diversas realizações (figura 6.1). Os mapas da figura 6.1 correspondem a quatro das 100 realizações incluindo as que resultaram no maior e no menor volume poroso total (VPT) para o reservatório.

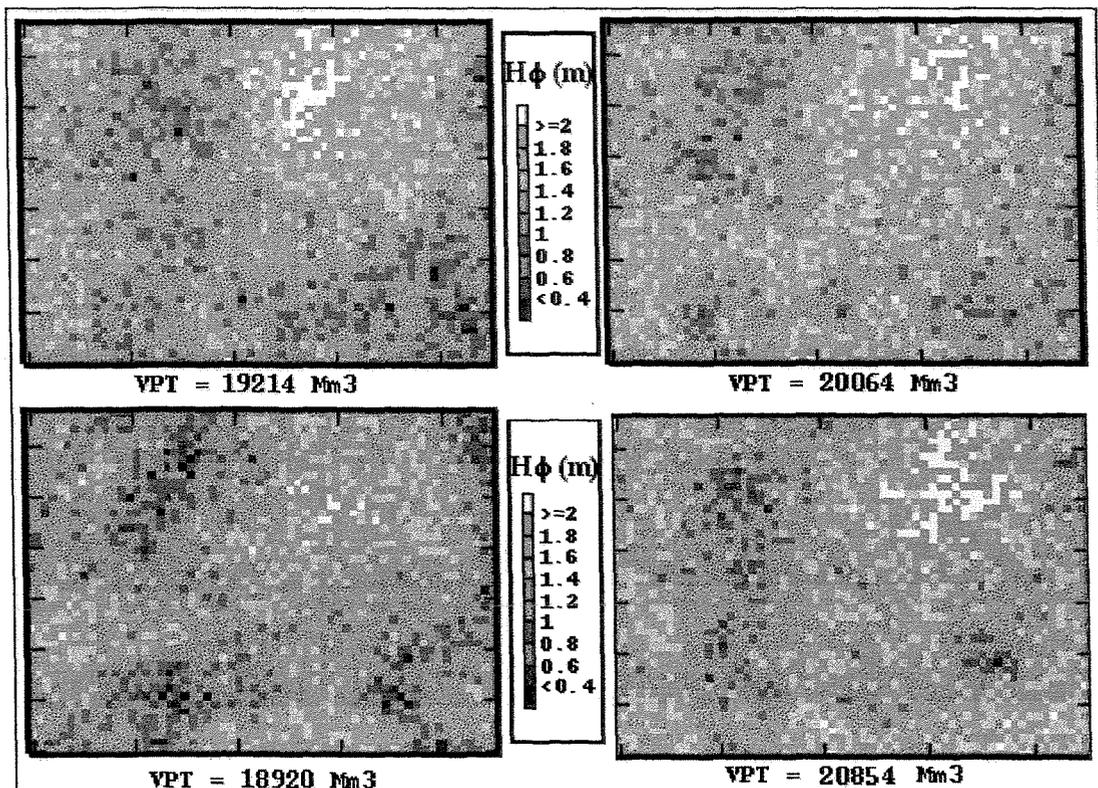


Figura 6-1: Realizações estocásticas utilizando SGS.

Os softwares comerciais possibilitam a utilização de representações alternativas de incerteza. Por exemplo, com o ISATIS pode-se obter a média e variância em cada bloco (*pixel*) das 100 realizações obtidas. A variância do valor simulado em cada bloco tem sido utilizada como uma estimativa da incerteza local. Na figura 6.2 estão a média e o desvio padrão (raiz quadrada da variância) das 100 realizações. No mapa do desvio padrão, observa-se que as áreas com os maiores valores correspondem às áreas com um número menor de dados na vizinhança, por exemplo, o extremo sudeste.

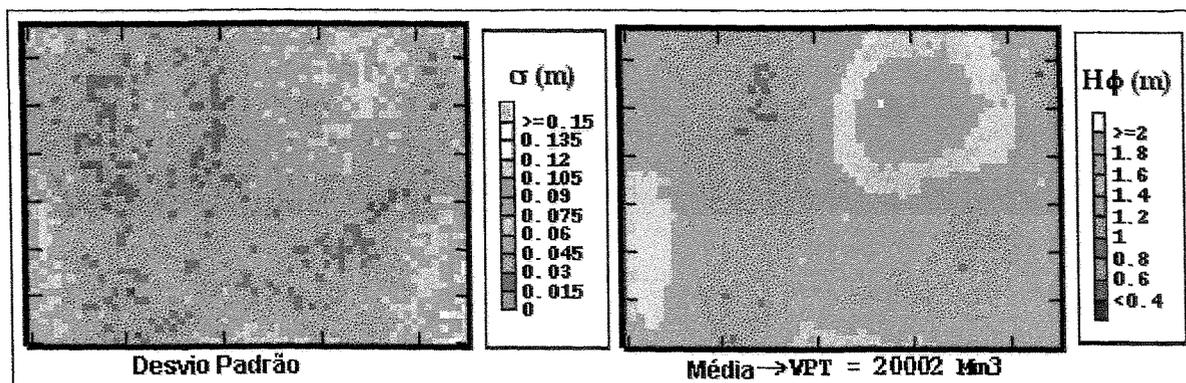


Figura 6-2: Média e desvio padrão das 100 simulações SGS.

Uma outra possibilidade é a construção do mapa do coeficiente de variação (razão entre o desvio padrão e a média) que representaria uma medida de dispersão relativa. Na figura 6-3 está o mapa do coeficiente de variação. Neste mapa, os dados condicionantes são facilmente visualizados, pois eles constituem ‘ilhas’ de coeficiente zero. Neste caso, as áreas de elevado coeficiente de variação correspondem a áreas com poucos dados condicionantes ou com valor baixo na média das 100 simulações.

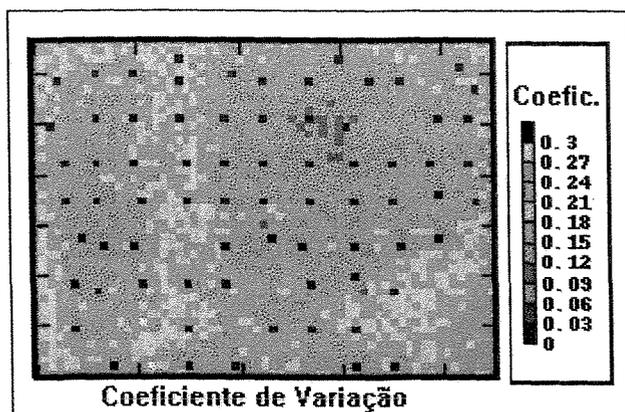


Figura 6-3: Coeficiente de variação das 100 Simulações SGS.

Na tabela 6-3 estão as estatísticas da média, da pior e da melhor realização. Os termos ‘pior’ e ‘melhor’ significam as realizações que resultaram respectivamente no menor e

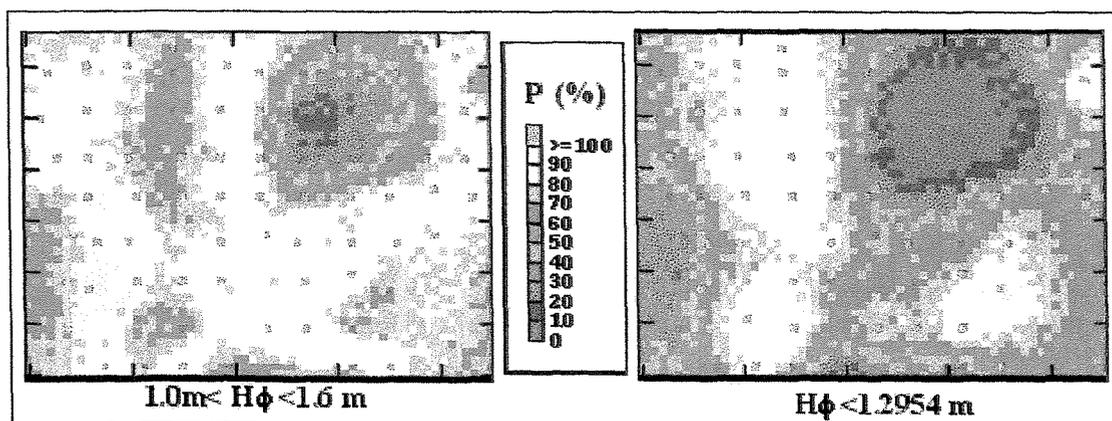
maior volume poroso total para o reservatório. Verifica-se que as estatísticas da média das 100 realizações são bem similares às obtidas através da krigagem correspondente ao caso base.

**Tabela 6-3: Estatística das Realizações**

Realizações	Volume (Mm <sup>3</sup> )	(Hφ) <sub>b</sub> mínimo	(Hφ) <sub>b</sub> máximo	(Hφ) <sub>b</sub> médio	Variância (Hφ) <sub>b</sub>
Pior Realização	18920	0.3589	2.0950	1.2253	0.0737
Média das 100	20002	0.4023	2.0818	1.2954	0.0338
Melhor Realização	20854	0.3718	2.1024	1.3506	0.0771
Krigagem	19994	0.9087	1.8335	1.2949	0.0310

Obs: (Hφ)<sub>b</sub> é o valor do atributo (Hφ) por bloco em metros.

Uma outra alternativa é a utilização de mapas de probabilidade (figura 6.4) que indicam a probabilidade do valor simulado em cada bloco do modelo ser superior/inferior que um determinado valor limite. No caso em análise, os mapas de probabilidade reportam em cada bloco, o percentual de imagens que apresentaram valores superiores ou inferiores a um determinado valor de *cutoff*. Os mapas de probabilidade intervalares que correspondem à probabilidade de, em cada bloco (*pixel*), o valor simulado estar entre dois limites.



**Figura 6-4: Mapas de probabilidade.**

Os mapas de quantis (figura 6-5), por sua vez, apresentam o valor correspondente a um determinado quantil (percentil) em cada bloco do modelo. Pode-se também construir mapas de interquantis ou interquartis, que em cada bloco (*pixel*), apresentam a diferença entre dois valores simulados correspondentes a dois percentis preestabelecidos.

Os mapas de desvio padrão, variância, coeficiente de variação ou interquantis são todos medidas da dispersão que refletem algum aspecto da incerteza. Onde essas estatísticas são baixas é porque se tem um alto grau de certeza sobre o atributo e vice-versa.

Segundo Srivastava (1994a), embora as representações acima citadas sejam convencionais em geoestatística, a maior falha reside no fato de que seus formatos não são intuitivos. Isto é, requerem por parte dos usuários do modelo um entendimento de como podem representar a incerteza.

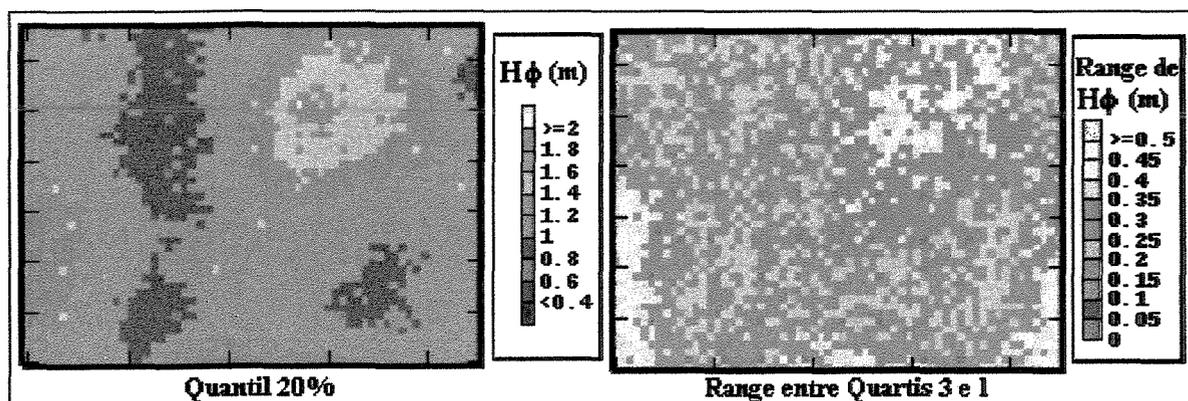


Figura 6-5: Mapas de quantis.

Embora não muito comuns, a adoção de representações híbridas pode tornar o entendimento da incerteza mais intuitivo. Por exemplo sobrepor aos mapas das realizações, um mapa que ressalte a informação de incerteza, como por exemplo uma medida de dispersão. Com os recentes recursos gráficos, esta alternativa fica bastante facilitada: as duas informações podem ser visualizadas simultaneamente, facilitando a identificação das áreas com as maiores e menores incertezas.

## 6.5 Flutuações Ergódicas

A ocorrência das flutuações ergódicas, amplamente reportadas na literatura, poucas vezes tem sua importância relacionada à incerteza. Na figura 6.6 estão os semivariogramas do modelo inicial, das 100 imagens simuladas, da média das 100 simulações e do mapa krigado.

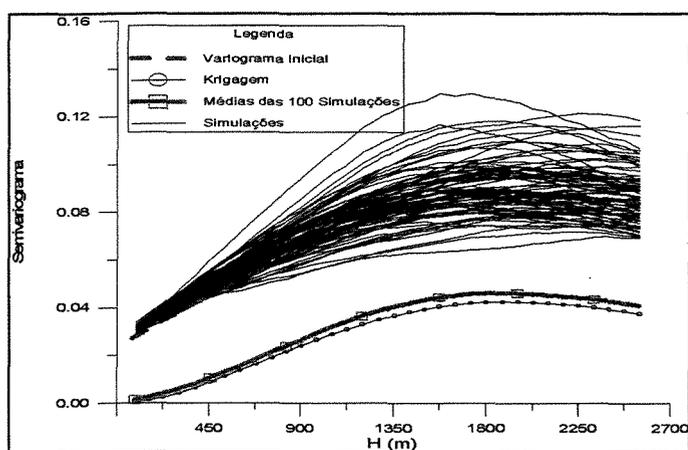


Figura 6-6: Semivariogramas das simulações e da krigagem.

Verifica-se que algumas imagens apresentam semivariogramas muito diferentes do utilizado para gerá-las, isto é, do modelo inicial. Essas flutuações, denominadas ergódicas, ocorrem porque o tamanho do domínio a ser simulado é da ordem de grandeza do *integral range* para o modelo de semivariograma adotado. Um critério que deveria ser adotado quando

da seleção de imagens é a verificação se o semivariograma da imagem simulada não se distancia muito do considerado representativo. As flutuações ergódicas podem ser consideradas úteis no sentido de cobrir alguma incerteza que se tem no modelo de semivariograma escolhido. Porém, quanto maior a confiabilidade que se tem no modelo e nos parâmetros do semivariograma escolhido, menores deverão ser as flutuações ergódicas aceitas.

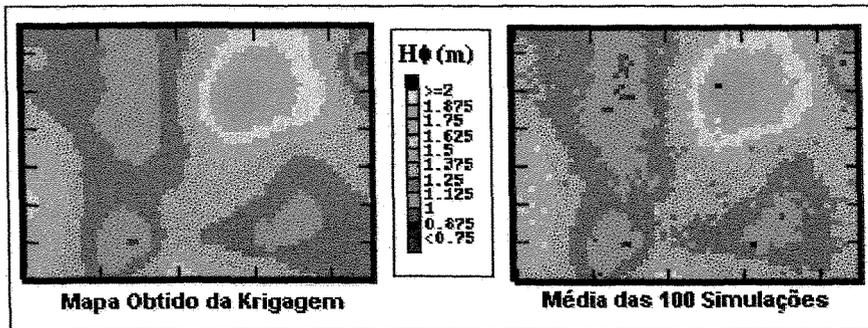


Figura 6-7: Comparação entre os mapas krigado e o da média das 100 simulações.

Para o caso analisado, o semivariograma da média das simulações, bloco a bloco, tende ao semivariograma da imagem krigada, à medida que aumenta o número de realizações. Também a média das simulações tende à imagem krigada (figura 6-7), o que resulta em volumes porosos totais para o reservatório bastante similares para as duas condições.

## 6.6 Simulação Condicional, Não Condicional e Krigagem

Os resultados semelhantes entre a média das simulações condicionais e a krigagem, como obtido no item anterior, têm muitas vezes confundido o usuário que, às vezes, incorretamente generaliza o resultado para todas as situações. Exemplos podem ser encontrados na literatura, para os quais esta relação não é verdadeira.

Tem-se o exemplo da estimativa do comprimento de um cabo submarino a ser lançado dispondo de alguns valores das cotas batimétricas do fundo do mar e da correlação espacial da mesma. Construiu-se um caso base em que os pontos extremos distam 10 km. Deste caso base foram retiradas 21 amostras distanciadas de 50m. Foram utilizadas a krigagem, simulação estocástica condicional e não condicional. Os resultados estão listados na tabela 6.4.

A estimativa obtida através da krigagem resultou bastante diferente dos valores obtidos através da simulações condicionais e não condicionais que apresentaram resultados mais próximos ao considerado verdadeiro. A krigagem apresentou a tendência de suavização da estimativa do relevo do fundo do mar, o que levou a subestimativas para o comprimento do cabo a ser lançado.

**Tabela 6-4: Simulação condicional, não condicional e krigagem**

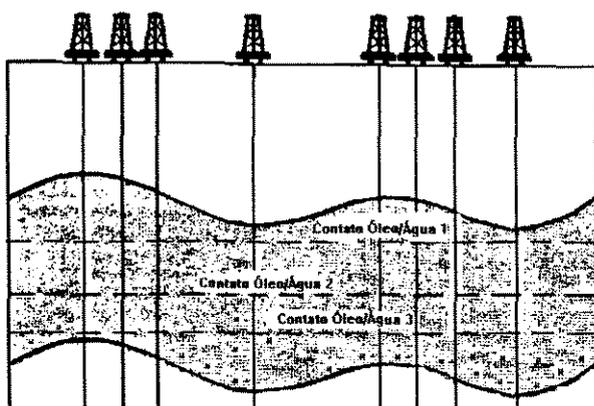
Resultados de 20 Simulações	Simulação	
	Condicional	Não Condicional
Máximo	17.66	17.98
Média	17.11	17.26
Mínimo	16.31	16.21
Desvio Padrão	0.28	0.49

Obs: Valores em km. Krigagem 10.29 km e Valor de Referência 16.64 km.

Conclui-se que a média das simulações tenderá ao valor obtido através da krigagem a depender dos objetivos para o qual o modelo é construído. Caso o objetivo do estudo fosse a determinação da cota batimétrica média naquela mesma região, a média das simulações tenderia à realização da krigagem, à medida que se aumentasse o número de realizações.

Neste exemplo acadêmico ainda deve ser ressaltado que se o único objetivo do trabalho fosse determinar o comprimento do cabo a ser lançado, a simulação não condicional substituiria a condicional com mesma performance em termos de resultados, sem considerar a maior rapidez com que se pode obtê-la.

Para o exemplo do cabo submarino, foi aplicada uma função de transferência que consistia em realizar uma integral de linha ao longo da superfície do fundo do mar, que foi realmente o atributo simulado ou estimado através da krigagem.

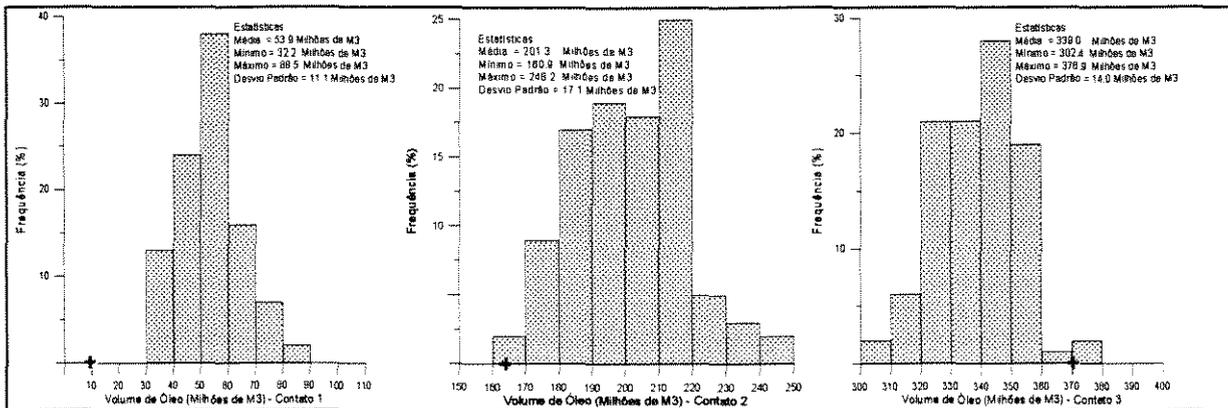


**Figura 6-8: Esquema ilustrativo para cálculo do HCIP.**

Exemplos semelhantes podem ser encontrados na literatura geoestatística voltada para reservatórios petrolíferos. Com os mesmos dados das 100 simulações SGS e da krigagem da variável  $H\phi$  correspondente ao produto porosidade x espessura foi construído um exemplo demonstrativo. Foi suposto que a variável krigada e simulada correspondesse a profundidade do topo de um reservatório (em km) com espessura constante e igual a 200 metros. Em seguida, foram adotados 3 contatos óleo/água diferentes (figura 6.8) a 1100, 1300 e 1500 m, sendo calculado o volume de óleo *in place* (VOIP) para as imagens simuladas e a krigada. Para

efeito de cálculo, foram utilizadas porosidade de 20 % e saturação de óleo acima do contato óleo/água igual a 75%. Os resultados estão apresentados nos histogramas da figura 6.9, onde o valor obtido através da krigagem está indicado com uma cruz. Em nenhuma das situações a krigagem corresponde à média das simulações. Pode também ser observado o efeito de suavização da krigagem. Para o contato óleo/água muito alto, a krigagem tendeu a indicar um VOIP muito baixo, no caso inferior ao mínimo obtido através das simulações estocásticas. Já para o contato óleo/água muito baixo o resultado com a krigagem ficou próximo ao maior valor obtido através das simulações.

Nestes dois exemplos foram aplicadas funções de transferência a um conjunto de imagens simuladas e à imagem krigada. A idéia preconcebida de que o resultado obtido através da krigagem deve ficar em torno da média dos resultados obtidos através de todas as simulações é, muitas vezes, incorreta.



**Figura 6-9: Resultados do cálculo dos VOIPs simulados e krigados**

Por exemplo, Rodrigues (1996) submeteu ao simulador de fluxo 30 imagens de permeabilidade obtidas através de SGS e uma imagem krigada de um campo real. As funções de transferência envolvidas no simulador de fluxo são bem mais complexas que as dos dois exemplos acima citados. Na análise de Rodrigues, os resultados obtidos com a imagem krigada não correspondem à média dos resultados obtidos com as imagens estocásticas quando submetidas ao simulador de fluxo.

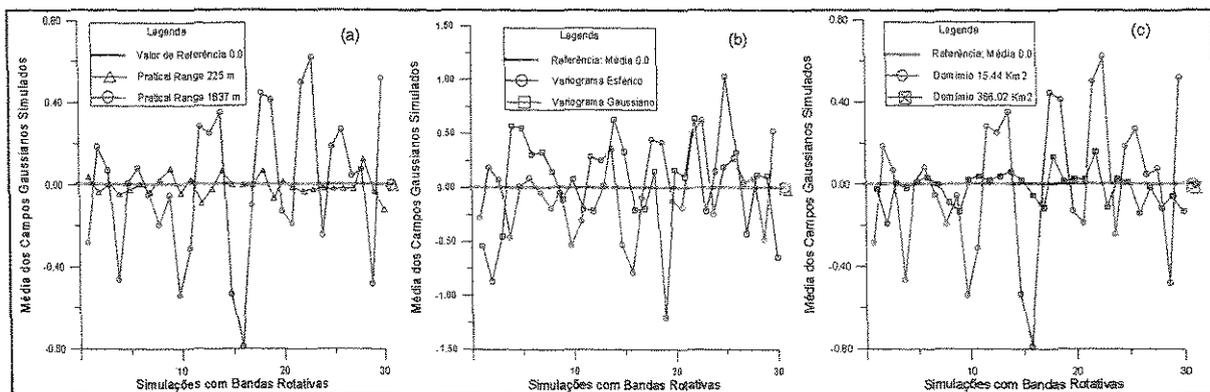
## 6.7 Ergodicidade

Um conceito muitas vezes mal entendido pelo usuário de simulação estocástica é a ergodicidade e como ela se relaciona com os parâmetros escolhidos para o modelo.

Para exemplificar como os parâmetros do modelo determinam a ocorrência de flutuações ergódicas em relação à media, foram realizadas 30 simulações com bandas rotativas

não condicionais em um domínio igual ao utilizado nas simulações condicionais, utilizando um semivariograma com uma estrutura esférica com alcance de 1837 m. Espera-se que os campos simulados gaussianos apresentem distribuição  $N(0,1)$ . Porém, são observadas flutuações nestas estatísticas devido a não ergodicidade do modelo. Para avaliar o impacto de alguns parâmetros do modelo na ocorrência das flutuações ergódicas foram realizadas 30 simulações não condicionais para cada uma das seguintes condições: redução do alcance de 1837 para 225 m; alteração do modelo de semivariograma de esférico para gaussiano; aumento do domínio a ser simulado em 25 vezes, isto é 5 vezes em cada direção do domínio bidimensional.

Nas figuras 6-10a/c estão os gráficos resultantes destas comparações que confirmam que quanto mais contínuo for o modelo de semivariograma, maior for o alcance e menor for o domínio, maiores serão as flutuações ergódicas.



**Figura 6-10: Ergodicidade versus a) Alcance; b) Modelo de Semivariograma; c) Tamanho do Domínio**

A partir do modelo de semivariograma utilizado, pode-se determinar o *integral range* que é um parâmetro intimamente relacionado à ergodicidade. A rigor, o domínio a ser simulado tem que ser muito maior que o *integral range* para que se verifique ergodicidade da média.

**Tabela 6-5: Expressões do *integral range* em função do alcance.**

Modelo de Semivariograma	<i>Integral range</i>	
	Domínio 2D	Domínio 3D
Esférico	$(\pi/5)R^2$	$(\pi/6)R^3$
Exponencial	$(2\pi/9)R^2$	$(8\pi/27)R^3$
Gaussiano	$(\pi/3)R^2$	$(\pi/3)^{3/2}R^3$

Para alguns modelos de semivariogramas existem expressões analíticas para o *integral range*. Na tabela 6-5 estão algumas expressões considerando modelo de semivariograma isotrópico com alcance R.

Quando não se observa ergodicidade da média, não será observada a ergodicidade do semivariograma ou de qualquer outro momento de ordem superior. Nas modelagens de reservatórios petrolíferos, a ocorrência das flutuações ergódicas tanto da média quanto do semivariograma é bastante normal, uma vez que o *integral range* dos semivariogramas modelados são muitas vezes da mesma ordem de grandeza do tamanho do domínio a ser simulado.

## **6.8 Número de Realizações Necessárias**

Um interessante e curioso aspecto na determinação do número de simulações estocásticas necessárias para cobrir o espaço de incertezas é a pouca referência às flutuações ergódicas. Lantuéjoul (1994) reporta que quanto maior for o *integral range*, maior será este número. Isto porque maior será a magnitude das flutuações ergódicas. Contudo, não é respondida a questão principal: quantas imagens são necessárias? Os trabalhos na literatura geralmente utilizam números cabalísticos como 100 e 50 imagens.

Porém, são encontradas várias referências que apresentam resultados e conclusões, baseando-se em um número bastante limitado de imagens (inferiores a 10), sem se preocupar sobre a representatividade das imagens em termos de parâmetros estatísticos (histograma e semivariograma) e se as mesmas cobriam o espaço de incerteza do atributo. Por exemplo, Alabert e Modot (1992) compararam a SGS, Truncada Gaussiana e a SIS com dados de um reservatório turbidítico do Gabão. As conclusões reportadas por esses autores foram baseadas em apenas 5 simulações para cada uma das metodologias. De qualquer forma, quando da análise de incerteza do próprio atributo simulado ou mesmo quando as imagens simuladas são submetidas à simulação de fluxo, selecionar imagens segundo critério de flutuações ergódicas não tem sido uma prática comum.

Trabalhos mais recentes já demonstram uma preocupação com o número de imagens e os critérios de seleção a serem adotados.

No último congresso de Geoestatística, Journel (1997) reportou a metodologia por ele empregada para a questão. Inicialmente ele utiliza um conjunto de 100 realizações, para as quais são avaliados os resultados. Journel claramente reporta que o número 100 é cabalístico, e sua escolha é meramente para suscitar menos questionamentos do que por exemplo 99 ou 76. Em seguida são obtidas mais 100 realizações e os resultados destas são comparadas com os do primeiro conjunto. Se os resultados se mantiverem constantes, devem ser usadas as 200 realizações. Caso contrário, o processo é reiniciado gerando 200 realizações e assim

sucessivamente. Journal esclarece que a metodologia se baseia em um princípio elementar de que o tempo de CPU não é problema. Isto não se verifica, principalmente quando as realizações serão submetidas a simuladores de fluxo.

Algumas propostas de seleção das imagens representativas e a determinação do número de imagens necessárias para se cobrir o espaço de incerteza da previsão de comportamento do reservatório envolvem a utilização de parâmetros oriundos do ajuste de histórico e/ou da própria previsão de comportamento.

Por exemplo, Rodrigues (1996) utilizou como critério para determinação do número de imagens estocásticas, a estabilização da variância da produção acumulada final de água na simulação de fluxo a que as imagens estocásticas eram submetidas. Para o caso analisado por Rodrigues (1996) foram necessárias 30 realizações. Para a seleção das imagens representativas, Rodrigues (1996) utilizou parâmetros comparativos entre a resposta através da simulação de fluxo e o histórico de produção para compor uma função objetivo. As imagens estocásticas que apresentaram os menores valores da função objetivo foram selecionadas para prosseguir a análise e realizar a previsão de produção. Rodrigues dividiu o histórico de produção em duas etapas, e verificou que imagens selecionadas considerando a parte inicial do histórico de produção apresentavam resultados discrepantes, em termos de performance de produção, quando considerando todo o histórico. Esses resultados reforçam a idéia de se utilizar uma abordagem probabilística do reservatório e que a depender da disponibilidade de novos dados sejam eles poços, interpretação sísmica, dados de produção, a modelagem do reservatório e as previsões de comportamento derivadas da mesma devem ser revistas.

Campozana (1997) utilizou uma metodologia semelhante à Rodrigues (1996) para determinação do número de imagens através da avaliação da estabilização da diferença entre o maior e o menor fator de recuperação e da variância do mesmo, considerando o aumento do número de simulações estocásticas submetidas ao simulador de fluxo.

Em um caso real, para o qual ainda não se dispõe de um considerável conjunto de dados de produção, parâmetros da previsão de produção podem ser usados para determinação do número de imagens necessárias.

De qualquer forma, os estudos apontam geralmente para soluções específicas para cada caso que está sendo analisado. Não se vislumbra, por enquanto, a definição de uma regra universal que sirva para qualquer situação que estiver sendo modelada com diferentes mecanismos de produção (capa de gás, influxo de água, expansão de fluidos, etc) e diferentes métodos de recuperação secundária e terciária (injeção de água, gás, vapor, etc). A

transferência das incertezas que se tem na fase de construção dos modelos estocásticos do reservatório para a previsão de performance é uma questão que envolve uma grande diversidade de parâmetros.

## 6.9 Simulação Campo de Probabilidade

Com o mesmo conjunto de dados do reservatório da AMOCO utilizado para a SGS, foi implementada a *simulação p-field* conforme proposta por Srivastava (1992). Os resultados das 100 SGS condicionais foram utilizadas para compor as cdfs em cada bloco. Para efeito de simplificação, as cdfs foram discretizadas em 1,2,5,8,...,92,95,98,100%.

Para obtenção dos campos uniformes, foram realizadas 30 simulações de bandas rotativas não condicionais (ISATIS) com o mesmo semivariograma utilizado para gerar as 100 simulações SGS condicionais. Foi elaborado um programa em FORTRAN que inicialmente faz a conversão dos campos gaussianos obtidos das simulações das bandas rotativas para o campo uniforme.

Construídos os campos uniformes/campos de probabilidades, foi implementado uma rotina que faz a amostragem das cdfs, baseando-se no valor do campo de probabilidade do respectivo bloco. Como as cdfs estavam discretizadas, foi necessário realizar interpolação durante a amostragem das mesmas, que por motivo de simplicidade foi considerada linear.

Devido ao fato de que o domínio a ser simulado é da ordem de grandeza do *integral range* do semivariograma modelado, flutuações ergódicas ocorreram, como já evidenciado na simulação gaussiana seqüencial. As estatísticas das 30 realizações estão na tabela 6-6.

Tabela 6-6: Estatística das Realizações

Realizações	Volume (Mm <sup>3</sup> )	(Hφ) <sub>b</sub> mínimo	(Hφ) <sub>b</sub> máximo	(Hφ) <sub>b</sub> médio	Variância (Hφ) <sub>b</sub>
Pior Realização	17931	0.3600	2.0818	1.1613	0.0550
Média das 30	20046	0.4023	2.0818	1.2982	0.0346
Melhor Realização	21871	0.3591	2.0988	1.4165	0.1123
Krigagem	19994	0.9087	1.8335	1.2949	0.0310

Obs: (Hφ)<sub>b</sub> é o valor do atributo Hφ por bloco em metros.

Em comparação com a SGS, observa-se que a simulação *p-field* apresentou um intervalo de incerteza (diferença entre o maior e o menor volume) bem mais dilatado. Isto também reforça a idéia de que diferentes algoritmos de simulação levam a diferentes distribuições de incerteza. Em relação à média das 30 simulações, verifica-se uma coerência com a média obtida das 100 simulações SGS e com a krigagem.

## 6.10 Animação de Incerteza

Uma possibilidade para facilitar a visualização da incerteza é adotar uma sistemática que mostre um conjunto maior de realizações simultaneamente sem estar limitado ao tamanho da página. Uma opção é utilizar recursos de animação de forma que várias realizações minimamente diferentes sejam mostradas em uma rápida sucessão como em um desenho animado. Esta animação facilitaria a visualização das áreas do reservatório onde se tem as maiores incertezas. A possibilidade de geração de imagens minimamente diferentes é um dos subprodutos da técnica de simulação estocástica campo de probabilidades.

Para permitir a geração de imagens minimamente diferentes, adotou-se uma sistemática igual a abordagem descrita no item anterior só alterando o tamanho do *grid* do campo de probabilidades para 261x245 blocos.

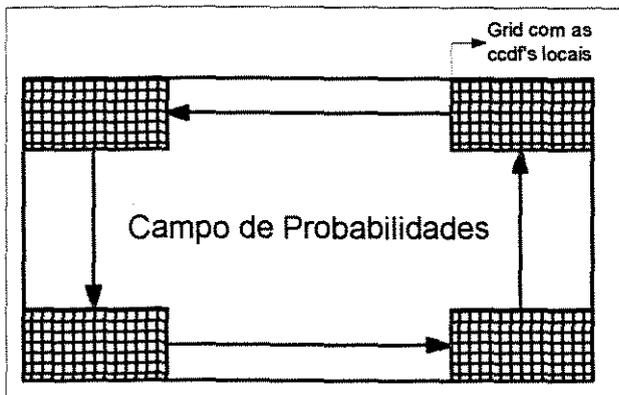
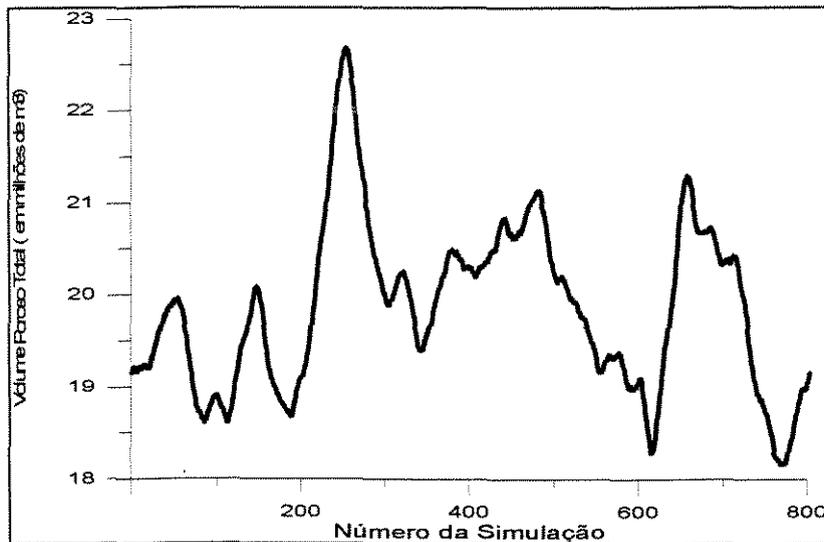


Figura 6-11: Geração de imagens minimamente diferentes.

Foi implementada uma rotina em FORTRAN que faz o *grid* de tamanho real 61x45 blocos (ccdf em cada bloco) deslizar sobre um outro de tamanho maior (campo de probabilidades). Assim, deslocando coluna a coluna (deslocamento em *x*) e linha a linha (deslocamento em *y*) foram construídas 200 realizações em cada uma das direções paralelas às bordas do *grid* do campo de probabilidades, obtendo um total de 800 simulações (figura 6-11). Não foi implementado nenhum caminho com direções aleatórias em cada um dos movimentos sobre o *grid* maior.

Para se obter uma realização após uma movimentação sobre o *grid* com o campo de probabilidades, o valor correspondente a probabilidade em cada bloco é levado ao *grid* com as cdfs para a realização da amostragem. Na figura 6-12, está o gráfico dos volumes porosos totais no reservatório obtido a partir de 800 simulações com variações mínimas entre cada realização.



**Figura 6-12: Volumes Porosos Totais obtidos com a Simulação *P-field*.**

O intervalo de incerteza do volume poroso total obtido a partir das 800 simulações minimamente diferentes ( $4.51 \text{ MMm}^3$ ) é maior que o obtido a partir das 30 simulações com diferentes campos de probabilidade ( $3.94 \text{ MMm}^3$ ).

Na figura 6-13 estão 8 mapas de realizações minimamente diferentes, parte do conjunto total de 800 realizações. Como tratava-se de um domínio 2D, optou-se por utilizar representações tridimensionais em que o atributo simulado é a terceira dimensão.

Pode ser visualizado (seta) um dado condicionante que fica destacado e passa a constituir um extremo local. Observa-se que o volume poroso total é ligeiramente diferente em cada realização. Com recursos de computação gráfica cada vez mais avançados, poderia se fazer a animação das imagens minimamente diferentes.

A geração de imagens minimamente diferentes só faz sentido no contexto de animação de incertezas para visualização. Em quaisquer outras situações, deseja-se amostrar convenientemente o espaço de incertezas com realizações bem diferentes umas das outras.

### ***6.11 Proposta Alternativa Para a Simulação *P-field*.***

Uma proposta alternativa à simulação *p-field* (Oliveira e Remacre, 1997) é a adoção do próprio campo gaussiano para a amostragem das ccdfs construídas via algoritmo de simulação estocástica ou krigagem paramétrica (figura 6-14). A abordagem se baseia na hipótese de que o campo a ser simulado apresenta distribuição multivariada gaussiana, isto é, podem ser obtidas ccdfs gaussianas completamente caracterizadas por sistemas de krigagens paramétricas como na SGS.

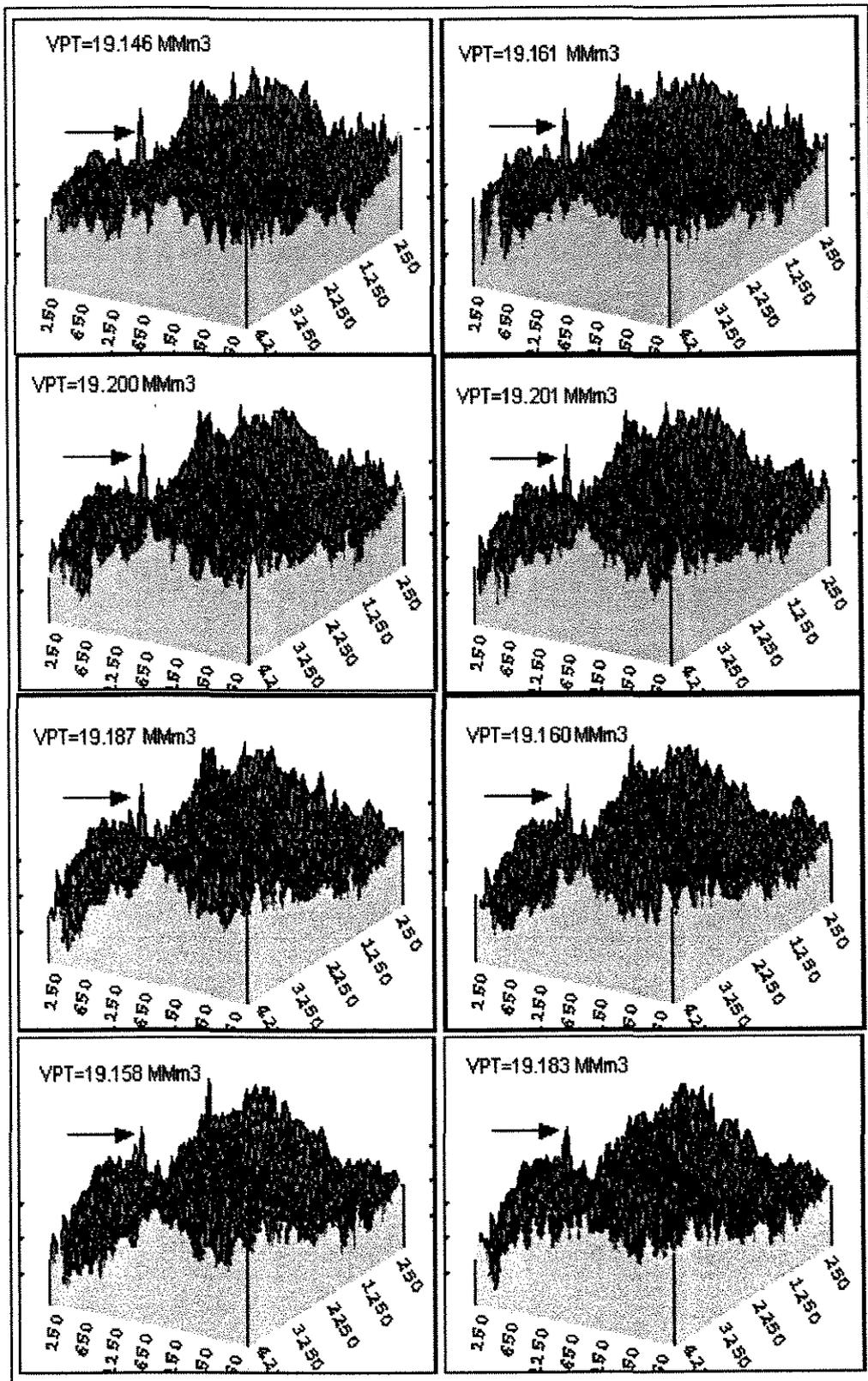


Figura 6-13: Realizações *P*-field minimamente diferentes.

A idéia básica do método é a seguinte:

- 1 - Obter uma cdf que possa ser expressa por dois parâmetros: a estimativa (a média) e um desvio padrão (a dispersão).

2 - Fazer a amostragem desta ccdf, baseando-se no campo gaussiano.

$$\frac{Z_s - m_e}{\sigma_e} = \frac{Z_{cg} - m_{cg}}{\sigma_{cg}} = Z_{cg}$$

$$Z_s = m_e + \sigma_e Z_{cg}$$

onde  $Z_s$  é o valor simulado,  $m_e$  e  $\sigma_e$  são respectivamente a média o desvio padrão que representam a ccdf em cada nó; onde  $Z_{cg}$  é o valor do campo gaussiano a ser utilizado para a amostragem,  $m_{cg}$  e  $\sigma_{cg}$  são respectivamente a média e o desvio padrão do campo gaussiano que é considerado  $N(0,1)$ .

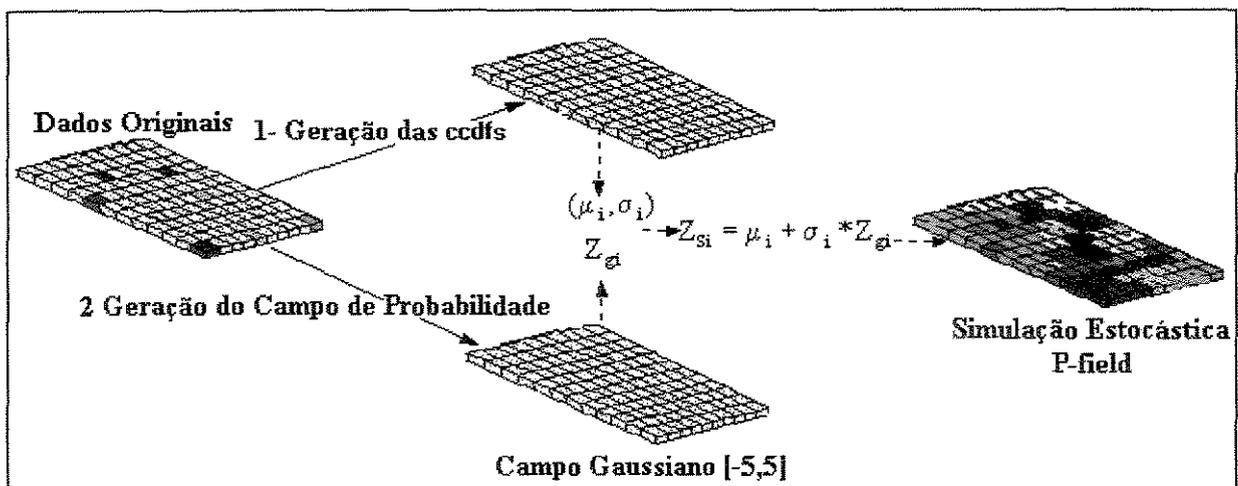


Figura 6-14: Proposta Alternativa para a P-field.

A proposta se baseia na construção de uma ccdf local a partir da estimativa e do desvio padrão da krigagem dos dados transformados. De qualquer forma, a proposta implica em assumir que os dados apresentam características multivariadas gaussianas. Após a amostragem das ccdfs transformadas, os valores sofrem a transformação de volta. Os resultados com esta abordagem estão na tabela 6-7.

Tabela 6-7: Estatísticas das Simulações com Abordagem P-field Alternativa

Realizações	Volume (Mm <sup>3</sup> )	(Hφ) <sub>b</sub> mínimo	(Hφ) <sub>b</sub> máximo	(Hφ) <sub>b</sub> médio	Variância (Hφ) <sub>b</sub>
Pior Realização	17843	0.3602	2.0697	1.1556	0.0518
Média das 30	20008	0.8919	1.8259	1.2958	0.0327
Melhor Realização	21909	0.3589	2.1433	1.4190	0.1123
Krigagem	19994	0.9087	1.8335	1.2949	0.0310

Obs: (Hφ)<sub>b</sub> é o valor do atributo Hφ por bloco em metros.

As estatísticas de cada realização obtida com a abordagem alternativa, usando campo gaussiano, são bastante similares às estatísticas da respectiva realização obtida com a abordagem tradicional, com os campos de probabilidade oriundos da transformação uniforme

dos mesmos campos gaussianos. Na figura 6-15 estão os valores simulados de volume poroso total para o reservatório em 30 simulações com as duas abordagens.

Uma observação importante é a extrema dependência dos resultados da simulação campo de probabilidade, tanto da abordagem alternativa quanto a tradicional, em relação ao campo de probabilidade uniforme/gaussiano. Na figura 6-15 estão os gráficos dos valores de volume poroso total médio obtidos das simulações com as duas abordagens para cada um dos 30 campos de probabilidade construídos e o gráfico da média do campo de probabilidade usado. Verifica-se fator de correlação da ordem de 0.97 entre os valores totais médios em cada simulação e a média do campo uniforme (gaussiano) para as duas abordagens. Isto evidencia que desde que as cdfs locais sejam representativas, os resultados das abordagens de simulação campo de probabilidade modificada e tradicional são dominados pelo campo gaussiano e uniforme, respectivamente.

Após a esquematização e testes da proposta alternativa para a *p-field*, foram encontradas referências que enfocam parte da mesma. Bourgault e Journel (1996) e Bourgault (1996), utilizaram cdfs construídas a partir de krigagens paramétricas dos dados originais transformados (gaussiana) e o conceito do campo de probabilidades com objetivo de gerar uma estimativa alternativa à cokrigagem e pós processar realizações que apresentem ruídos, respectivamente. Assim, a real inovação da abordagem alternativa proposta se reduz apenas a utilização do próprio campo gaussiano para realizar a amostragem de cdfs baseadas em modelos gaussianos como os da SGS.

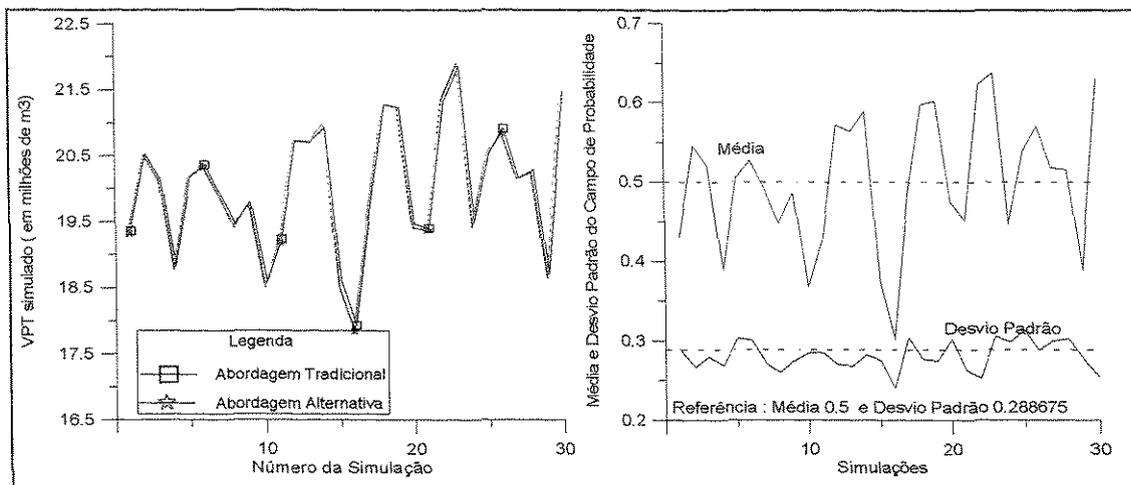


Figura 6-15: Volume poroso total e média/variância do campo uniforme em cada simulação.

## 6.12 Incertezas nos Parâmetros e no Modelo

Um problema comum na obtenção da distribuição univariada das propriedades do reservatório é a ocorrência de agrupamentos (*clustering*). Os poços são perfurados

preferencialmente nas áreas onde se acredita, através do modelo geológico, nas melhores possibilidades de sucesso. Mas mesmo nestas áreas, o espaçamento entre poços é geralmente muito grande impossibilitando a definição do comportamento do semivariograma na origem. Verticalmente, as amostras retiradas dos testemunhos para análises petrofísicas geralmente privilegiam as melhores fácies, deixando as piores precariamente amostradas. O efeito da amostragem preferencial é que a distribuição univariada não corresponderá à do reservatório, tendo um alto primeiro momento (média elevada, maior densidade de poços nas áreas de maiores espessuras permoporosas) e um baixo segundo momento (baixa variabilidade das amostras).

Outro problema comum é que os dados disponíveis são insuficientes para inferir as estatísticas das propriedades a serem simuladas. Geralmente é grande a incerteza nos parâmetros do semivariograma (alcance, patamar e efeito pepita) e até no modelo do mesmo. Toledo (1990) reporta a importância do aumento do número de dados condicionantes e da melhor definição dos parâmetros do semivariograma na obtenção de modelos geoestatísticos mais representativos do reservatório.

Em seu estudo, Rodrigues (1996), por sua vez, reporta que não ocorreram variações significativas nas respostas de fluxo das imagens estocásticas em função de alterações no modelo variográfico, quando comparado com os resultados de todas as imagens simuladas. Segundo, Rodrigues (1996) este comportamento pode ser atribuído ao fato de que, em sua interpretação, as alterações no modelos eram limitadas e consistiram basicamente em modificar o alcance e o patamar nos modelos representativos dos variogramas experimentais, dentro dos limites de incertezas existentes.

Campozana (1997) demonstrou através da utilização de dados sintéticos a influência do aumento do número de dados condicionantes, admitindo que o semivariograma já era conhecido. Campozana (1997) construiu um modelo sintético para fazer esta análise. Constatou-se melhoria no fator de correlação entre as imagens simuladas e o caso base (real) com o aumento do número de dados condicionantes. Com relação à distribuição de resposta das realizações quando submetidas ao simulador de fluxo, Campozana (1997) também verificou que o aumento do número de dados condicionantes, no caso poços, também reduz a incerteza. Campozana (1997) reportou melhoria na obtenção de intervalos de incertezas mais estreitos quando se considera informações adicionais vindas de outras fontes como a de dados de testes.

É razoável acreditar que maior integração de dados leva a realizações que são mais próximas da realidade. Seria natural também esperar que a medida que mais dados são disponibilizados e integrados ao algoritmo de modelagem, menores seriam os intervalos de incertezas, conforme reportado por Campoza (1997). Para Galli et al. (1997), esta hipótese nem sempre se verifica. Dubrule (1994) esclarece de forma bastante didática esta questão. Uma vez que se tem definido um modelo probabilístico com os seus parâmetros bem identificados, à medida que aumenta o número de dados, a variabilidade entre as realizações diminui e conseqüentemente o intervalo de incertezas. São encontradas várias referências na literatura que reportam a influência do aumento do número de dados condicionantes e a diminuição das incertezas, só que na maioria das vezes baseando-se em um modelo probabilístico fixo. Dubrule (1994) explica que em muitas situações reais, a disponibilidade de novas informações leva a intervalos de incerteza maiores, porque algumas vezes implica em alteração drástica no modelo ou em seus parâmetros.

A necessidade de incorporar dados de diferentes fontes e qualidade tem resultado em uma gama de propostas de alterações nos antigos e novos algoritmos de simulação estocástica visando principalmente dados oriundos de levantamentos sísmicos. Trata-se de utilizar toda a informação disponível oriunda da geofísica, que até bem pouco tempo era voltada exclusivamente para a arquitetura externa do reservatório, com intuito de obter descrições e modelagens de propriedades internas do reservatório cada vez mais representativas. Quanto à utilização de dados dinâmicos de produção (por exemplo, testes em poços), as principais aplicações reportadas na literatura abordam a questão com utilização da *simulated annealing*.

Portanto, quando da utilização de simulação estocástica para modelagem de reservatórios é crucial a integração de toda a informação disponível considerada relevante. Assim, a seleção do algoritmo de simulação estocástica depende, dentre outras coisas, da habilidade das pessoas envolvidas em lidar com as informações oriundas da geofísica, da geologia ou da engenharia no software geostatístico disponível. O mal estar gerado por uma distribuição de incerteza sobre a resposta do reservatório mais imprecisa quando da adição de novas informações, como por exemplo, um novo levantamento sísmico, pode levar algumas pessoas a questionar a modelagem estocástica, pois de um ponto de vista imediatista isto é contraproducente. É preciso que os usuários de simulação estocástica se convençam e saibam convencer as pessoas que fazem tais argumentações que a adição de mais informações pode resultar em alterações do próprio modelo utilizado na simulação estocástica e que, desprezar esta informação é adotar um modelo que, agora, é sabidamente errado.

## 7. Conclusões

Apesar do inquestionável sucesso da geoestatística enquanto ferramenta de modelagem de reservatórios, não se pode esquivar de uma reflexão sobre suas falhas ou limitações. Tópicos e problemas específicos sobre avaliação de incertezas foram analisados nesta dissertação. A partir de uma análise crítica e de uma profunda reflexão sobre incertezas na modelagem de reservatórios com utilização das ferramentas geoestatísticas, algumas conclusões se fazem pertinentes. Para alguns dos problemas analisados, a geoestatística não tem uma solução específica. Ao chamar a atenção para estes problemas, eu espero simplesmente catalisar ações, novos estudos ou no mínimo promover o debate.

### **Em relação à geoestatística:**

a) A diversidade de notação geoestatística prejudica o entendimento e a disseminação das novas propostas. Não existe consenso da notação, bem como da terminologia geoestatística. A comunidade geoestatística deve envidar esforços no sentido de uniformizar notações e terminologias.

b) A krigagem além de ser uma ferramenta geoestatística de estimativa, que pode ter sua utilização, a depender dos objetivos do estudo, também é parte integrante dos vários algoritmos de simulação estocástica que se baseiam em variogramas.

c) Cada metodologia de simulação estocástica e estimativa tem seus ferrenhos defensores, os seus próprios criadores, que na maioria das vezes só ressaltam as vantagens, deixando que as desvantagens e possibilidades de melhoria sejam investigadas e publicadas por outros autores.

d) Nenhum algoritmo de simulação estocástica ou estimativa é o melhor qualquer que seja a situação que se deseja analisar. O conhecimento das técnicas disponíveis, suas limitações e faixa de aplicação são essenciais para o sucesso do trabalho de modelagem. A escolha de um particular algoritmo geoestatístico deve basear principalmente no objetivo do estudo, na qualidade dos dados disponíveis, na habilidade do usuário em utilizar a metodologia escolhida, da disponibilidade de tempo e recursos computacionais.

### **Em relação à krigagem:**

a) O grande número de tipos de krigagem, algumas vezes obscurecidas por notações matemáticas complexas, pode mascarar a simplicidade do algoritmo básico: todas são técnicas de regressão que diferem apenas nos tipos particulares de funções, obtidas a partir dos dados, que estão sendo recombinaadas para a obtenção da estimativa.

b) Intervalos de incerteza derivados de krigagem paramétricas exercem um enorme fascínio devido à extrema facilidade para gerá-los. É de domínio dos geoestatísticos, as hipóteses sobre as quais eles se fundamentam, porém muitas vezes, isto é pouco transparente para o usuário. Apesar do conhecimento, a nível da comunidade geoestatística, sobre as limitações da abordagem paramétrica dos intervalos de incerteza, ainda são encontradas referências recentes na bibliografia, baseando-se em hipóteses algumas vezes questionáveis, principalmente a nível local, e o que é pior, sem enfatizá-las e sem explicitar o significado das mesmas. Intervalos de incerteza derivados de krigagens não paramétricas baseadas em indicatrizes são mais trabalhosos. É questionável a vantagem de não assumir nenhuma distribuição a priori.

c) A abordagem proposta por Roth e Armstrong (1995) para obtenção de intervalos de incerteza locais a partir de krigagens paramétricas resultam em ICs com assimetria reversa e dependentes do valor da estimativa. A nível de incerteza local, a abordagem de Roth e Armstrong (1995) é um grande avanço em relação à abordagem tradicional. Verificou-se também que, para as situações em que a variância de krigagem é pequena em relação à estimativa, a utilização desta abordagem leva a intervalos de incerteza globais muito mais conservativos que a abordagem tradicional.

d) Com o desenvolvimento e aumento da utilização das técnicas indicatrizes, observa-se a retomada das pesquisas direcionadas à correção dos pesos negativos da krigagem, com várias alternativas sendo propostas nos últimos anos. Ressalta-se que estas novas alternativas visam prioritariamente obter ponderadores que possam ser utilizados como percentuais na construção das ccdfs. Não se objetiva criar um substituto para a estimativa através da krigagem tradicional. Uma característica reforçada por alguns autores é que estas abordagens devem ter praticidade, isto é, não devem ser computacionalmente intensivas. Em todas as abordagens que garantem a positividade dos ponderadores o que é perdido é a condição de mínima variância de krigagem.

### **Em relação à simulação estocástica:**

a) A simulação *p-field* é basicamente um pós processador de outros métodos de simulação estocástica e estimativa. É bastante fácil de implementar a partir dos softwares geoestatísticos comerciais. Ela apresenta como subproduto a possibilidade de gerar realizações minimamente diferentes. A abordagem alternativa para a simulação *p-field* basicamente elimina a conversão do campo gaussiano não condicional em campo de probabilidades e possibilita a utilização de krigagens paramétricas usuais, porém é restrita a domínios gaussianos.

b) Apesar do tema ergodicidade estar difundido entre os geoestatísticos, poucos são os trabalhos que verificam a reprodução das estatísticas do modelo nas realizações obtidas. Na maioria das vezes, as realizações são submetidas a um simulador de fluxo sem o cuidado adicional de verificar a representatividade da mesma frente às estatísticas do modelo inicial. Pode-se usar as eventuais flutuações ergódicas das estatísticas para considerar as incertezas que se tem nos parâmetros do modelo. Porém, algumas realizações poderão apresentar flutuações ergódicas das estatísticas do modelo inicial (média e semivariograma) muito além da incerteza que se tem na definição desses parâmetros.

c) Ergodicidade é uma característica do modelo probabilístico e não dos dados ou fenômeno sob estudo. Quanto mais contínuo for o modelo de semivariograma e quanto menor for a relação entre o *integral range* e as dimensões do domínio onde se realizará a simulação, maiores serão as flutuações ergódicas verificadas.

d) Se faz necessários estudos que chequem a equiprobabilidade das realizações obtidas com os diferentes algoritmos de simulação estocástica.

### **Em relação à utilização de dados sísmicos e de produção:**

a) Vários são os trabalhos e algoritmos geoestatísticos que envolvem a utilização de dados sísmicos na caracterização de reservatórios. Trata-se de aplicar os dados de levantamento sísmicos para mapear propriedades internas do reservatório.

b) A inclusão dos dados dinâmicos de produção na modelagem estocástica tem sido realizada sobretudo com a adoção da técnica de *simulated annealing* ou alguma técnica de otimização considerando uma formulação de transferência de escala baseada principalmente nas médias potenciais. Se faz necessário que novas pesquisas se realizem no sentido de ampliar os horizontes da utilização dos dados dinâmicos face a importância dos mesmos, pois representam um retrato atualizado do comportamento do reservatório em situações reais de produção.

### **Em relação à simulação estocástica e a modelagem de reservatórios:**

a) A adoção de critérios relacionados com a distribuição de resposta da performance das imagens estocásticas submetidas aos simuladores de fluxo tem dado bons resultados. Portanto, constitui uma abordagem válida para a definição do número de imagens necessárias para cobrir adequadamente o espaço de incertezas. Não se vislumbra, por enquanto, a determinação de uma regra universal para a determinação do número de imagens necessárias para cobrir o espaço de incertezas seja de um atributo, seja da resposta do reservatório submetido ao simulador de fluxo.

b) É essencial o entendimento de que, qualquer que seja a abordagem escolhida para a modelagem estocástica do reservatório, importantes hipóteses e simplificações serão adotadas. É importante que o usuário saiba como as diferenças nas hipóteses podem influenciar a modelagem do reservatório. É imprescindível também que todas as hipóteses feitas para a escolha do modelo sejam documentadas, uma vez que as mesmas podem estar equivocadas, e isto ser observável quando mais dados se tornarem disponíveis.

c) Durante todo o processo de caracterização e simulação de reservatórios várias fontes de variabilidade envolvem o julgamento subjetivo por parte do usuário: seja na coleta e utilização dos dados considerados relevantes, seja nos diferentes usos que se pode dar aos mesmos dados como por exemplo diferentes processamentos sísmicos, análise de testes, abordagens de simulação geoestatística e numérica, etc. Pode-se dizer que a primeira grande incerteza é a do próprio usuário.

d) Em sentido holístico, pode-se dizer que a incerteza é devido a nossa ignorância e inabilidade para modelar com suficientes precisão e acurácia as distribuições das propriedades de rocha e fluidos do reservatório, e os fenômenos físicos e químicos que envolvem o fluxo de fluidos. O maior desafio dos profissionais envolvidos em um trabalho de caracterização e simulação de reservatórios é minimizar a incerteza que é inevitável. Esta tarefa requer equipe coesa e com sinergismo entre os profissionais das diversas especialidades que atuam na área.

# Referências Bibliográficas

- ALABERT, F. G.; MODOT, V. Stochastic Models of Reservoir Heterogeneity Impact on Connectivity and Average Permeabilities. In: **ANNUAL TECHNICAL CONFERENCE AND EXHIBITION OF THE SPE 67**, Washington, USA, 1992. p.355-370. (SPE Paper 24893)
- ALMEIDA, A. S.; JOURNEL, A. G. Joint simulation of Multiple Variables with a Markov-Type Coregionalization Model. **Mathematical Geology**, v.26, n.5, p.565-588, 1994.
- ARMSTRONG, M. Is Research in Mining Geostats as Dead as a Dodo. In: DIMITRAKOPOULUS, R. (ed.), **Geostatistics for the Next Century**. Dordrecht, Holland: Kluwer Academics Publishers, 1994. p.303-312.
- BARNES, R. J.; JOHNSON, T. B. Positive Kriging. In VERLY, G. et al. (eds.), **Geostatistics for Natural Resources Characterization**. Dordrecht, Holland: D. Reidel, 1984. p.231-244.
- BARNES, R. J.; YOU, K. Adding Bounds to Kriging. **Mathematical Geology**, V.24, n.2, p.171-176, 1992.
- BEUCHER, H.; GALLI, A.; LE LOC'H, G.; RAVENNE, C. Including a Regional Trend in Reservoir Modeling Using the Truncated Gaussian Method. In: SOARES, A. (ed.), **Geostatistics Tróia'92**. Dordrecht, Holland: Kluwer Academics Publishers, 1993. p.555-566.
- BOURGAULT, G. Robustness of Noise Filtering by Kriging Analysis. In: **Stanford Center for Reservoir Forecasting**. Stanford, USA: Stanford University, 1995. (Report 8)
- BOURGAULT, G. Probability Field for the Post-processing of Stochastic Simulations. **Mathematical Geology**, V.28, n.6, p.723-734, 1996.
- BOURGAULT, G.; JOURNEL, A. G. Unsmoothed Estimation with Dense Secondary Information Using Probability Field Technique. In: **Stanford Center for Reservoir Forecasting**. Stanford, USA: Stanford University, 1996. (Report 9)
- BOURGAULT, G. Statistical Declustering and Convex Estimation Using Determinant of Redundancy Matrix. In: BAAFI, E. Y.; SCHOFIELD, N. A. (eds.), **Geostatistics Wollongong'96**. Dordrecht, Holland: Kluwer Academics Publishers, 1997. p. 103-114.
- CAMPOZANA, F. P. **Incorporating Dynamic Data Into Geostatistical Reservoir Modeling**. Austin, Texas, 1997. 281p. Dissertation Ph.D., University of Texas.
- CHU, J.; XU, W.; ZHU, H.; JOURNEL, A. G. The Amoco Case Study. In: **Stanford Center for Reservoir Forecasting**. Stanford, USA: Stanford University, 1991. (Report 4)

- CHU, J.; XU, W.; ZHU, H.; JOURNEL, A. G. 3-D Implementation of Geostatistical Analyses - The Amoco Case Study. In: YARUS, J. M.; CHAMBERS, R. L. (eds.), **Stochastic Modeling and Geostatistics**. Tulsa, USA: AAPG, 1994. p.201-216. (Computer Applications in Geology, n.3)
- CRESSIE, N. **Statistics for Spatial Data**. New York, USA: Wiley Interscience, 1993. 900p. (Wiley Series in Probability and Mathematical Statistics)
- DAMSLETH, E.; HOLDEN, L. Mixed Reservoir Characterization Methods, In: **TULSA CENTENNIAL PETROLEUM ENGINEERING SYMPOSIUM**, Tulsa, USA, 1994. p.129-143. (SPE Paper 27969)
- DAVID, M. **Handbook of Applied Advanced Geostatistical Ore Reserve Estimation**. New York, USA: Elsevier Scientific Publishing Company, 1988. 216p. (Developments in Geomathematics 6)
- DEUTSCH, C. V. The Relationship Between Universal Kriging, Kriging With an External Drift, and Cokriging. In: **Stanford Center for Reservoir Forecasting**. Stanford, USA: Stanford University, 1991. (Report 4)
- DEUTSCH, C. V.; JOURNEL, A. G. *Annealing* Techniques Applied to the Integration of Geological and Engineering Data, **Stanford Center for Reservoir Forecasting**. Stanford, USA: Stanford University, 1992. (Report 5)
- DEUTSCH, C. V.; COCKERHAM, P.W. Practical Considerations in the Application of Simulated *Annealing* to Stochastic Simulation. **Mathematical Geology**, V.26, n.1, p.67-82, 1994.
- DEUTSCH, C. V.; JOURNEL, A. G. **GSLIB Geostatistical Software Library and User's Guide**. 2. ed., New York, USA: Oxford University Press, 1996. 360p.
- DEUTSCH, C. V. Correcting For Negative Weights in Ordinary Kriging. **Computers & Geosciences**, v.22, n.7, p.765-773, 1996.
- DEUTSCH, C. V. Direct Assessment of Local Accuracy and Precision. In: BAAFI, E. Y.; SCHOFIELD, N. A. (eds.), **Geostatistics Wollongong'96**. Dordrecht, Holland: Kluwer Academics Publishers, 1997. p. 115-125.
- DOWD, P.A. A Review of Recent Development in Geostatistics. **Computers & Geosciences**, v.17, n.10, p.1481-1500, 1992.
- DUBRULE, O.; KOSTOV, C. An Interpolation Method Taking Into Account Inequality Constraints I. Methodology. **Mathematical Geology**, v.18, n.1, p.33-51, 1986.
- DUBRULE, O. A Review of Stochastic Models for Petroleum Reservoirs. In: ARMSTRONG, M. (ed.), **Geostatistics**. Dordrecht, Holland: Kluwer Academics Publishers, 1989, p.493-506.

- DUBRULE, O. Estimating or Choosing a Geostatistical Model?. In: DIMITRAKOPOULUS, R. (ed.), **Geostatistics for the Next Century**. Dordrecht, Holland: Kluwer Academics Publishers, 1994. p.3-14.
- ENGLUND, E. J. A Variance of Geostatisticians. **Mathematical Geology**, v.22, n.4, p.417-455, 1990.
- FROIDEVEAUX, R. Constrained Kriging as an Estimator of Local Distribution Functions. In: CAPASSO, V. et al. (eds.), **Statistics of Spatial Processes: Theory and Applications**, 1993. p.106-118.
- FROIDEVEAUX, R. Probability Field Simulation. In: SOARES, A. (ed.), **Geostatistics Tróia'92**. Dordrecht, Holland: Kluwer Academics Publishers, 1993. p.73-83.
- GALLI, A.; BEUCHER, H.; LE LOC'H, G.; DOLIGEZ, B. The Pros and Cons of the Truncated Gaussian Method. In: ARMSTRONG, M.; DOWD, P.A. (eds.), **Geostatistical Simulations**. Dordrecht, Holland: Kluwer Academics Publishers, 1994. p.217-233.
- GALLI, A.; BEUCHER, H. Stochastic Models for Reservoir Characterization a User-Friendly Review. In: **LACPEC 5**, Rio de Janeiro, 1997. (SPE Paper 38999)
- GEOVARIANCES. Manuais do Software ISATIS, versão 3.1. Fontainebleau, France: Geovariances, 1996.
- GÓMEZ-HERNÁNDEZ, J. J.; JOURNAL, A. G. Joint Simulation of Multigaussian Fields. In: SOARES, A. (ed.), **Geostatistics Tróia'92**. Dordrecht, Holland: Kluwer Academics Publishers, 1993. p.85-94.
- GOOVAERTS, P. Comparative Performance of Indicator Algorithms for Modeling Conditional Probability Distribution Functions. **Mathematical Geology**, v.26, n.3, p.389-411, 1994.
- GOTWAY, C. A.; RUTHERFORD B. M. Stochastic Simulation for Imaging Spatial Uncertainty Comparison and Evaluation of Available Algorithms. In: ARMSTRONG, M.; DOWD, P.A. (eds.), **Geostatistical Simulations**. Dordrecht, Holland: Kluwer Academics Publishers, 1994. p.1-21.
- HALDORSEN, H. H.; LAKE, L. W. A New Approach to Shale Management in Field-Scale Models. **SPE Journal**, p.447-457, Aug. 1984.
- HALDORSEN, H. H.; MACDONALD, C. J. Stochastic Modeling of Underground Reservoir Facies (SMURF). **ANNUAL TECHNICAL CONFERENCE AND EXHIBITION OF THE SPE 62**, Dallas, USA, 1987, p.99-113. (SPE Paper 16751)
- HALDORSEN, H. H.; DAMSLETH, E. Challenges in Reservoir Characterization. **AAPG BULLETIN**, v.77, n.4, p.541-551, 1993.

- HARBAUGH, J. W.; DAVIS, J. C.; WENDEBOURG, J. **Computing Risk for Oil Prospects; principles e programs**. New York, USA: Pergamon, 1995. 452p. (Computer Methods in the Geosciences)
- ISAAKS, E. H.; SRIVASTAVA, R. M. **Applied Geostatistics**. New York, USA: Oxford University Press, 1989. 561p.
- JOHNSEN, G.; TJOLSEN, C.; RYSETH, A.; DAMSLETH, E. Use of Seismic Data as a Guide in Stochastic Modelling. In: BAAFI, E. Y.; SCHOFIELD, N. A. (eds.), **Geostatistics Wollongong'96**. Dordrecht, Holland: Kluwer Academics Publishers, 1997. p.362-373.
- JOURNEL, A. G.; HUIJBREGTS, C. J. **Mining Geostatistics**. London, England: Academic Press, 1978. 600p.
- JOURNEL, A. G. The Place of Non-Parametric Geostatistics. In: VERLY, G. et al. (eds.), **Geostatistics for Natural Resources Characterization**. Dordrecht, Holland: D. Reidel, 1984. p.307-355.
- JOURNEL, A. G. Geostatistics Models and Tools for the Earth Sciences. **Mathematical Geology**, v.18, n.1, p.119-140, 1986a.
- JOURNEL, A. G. Constrained Interpolation And Qualitative Information - The Soft Kriging Approach. **Mathematical Geology**, v.18, n.3, p.269-286, 1986b.
- JOURNEL, A. G.; ALABERT, F. Non-Gaussian Data Expansion in the Earth Sciences. **Terra Nova**, n.1, p.123-134, 1989.
- JOURNEL, A. G.; ZHU, H. Integrating Soft Seismic Data Markov-Bayes Updating, an Alternative to Cokriging and Tradicional Regression. In: **Stanford Center for Reservoir Forecasting**. Stanford, USA: Stanford University, 1990. (Report 3)
- JOURNEL, A. G.; DEUTSCH, C. V. Entropy and Spatial Disorder. **Mathematical Geology**, v.25, n.3, p.329-355, 1993.
- JOURNEL, A. G. Modeling Uncertainty Some Conceptual Thoughts. In: DIMITRAKOPOULUS, R. (ed.), **Geostatistics for the Next Century**. Dordrecht, Holland: Kluwer Academics Publishers, 1994, p.30-43.
- JOURNEL, A. G.; RAO, S. E. Deriving Conditional Distributions form Ordinary Kriging. **Stanford Center for Reservoir Forecasting**. Stanford, USA: Stanford University, 1996. (Report 9)
- JOURNEL, A. G. The Abuse of Principles in Model Building and Quest for Objectivity, In: BAAFI, E. Y.; SCHOFIELD, N. A. (eds.), **Geostatistics Wollongong'96**. Dordrecht, Holland: Kluwer Academics Publishers, 1997. p.3-14.
- KOSTOV, C.; JOURNEL, A. G. Coding e Extrapolating Expert Information for Reservoir Description. In: LAKE, L. W.; CARROL, H. B. (eds.), **Reservoir Characterization**. San Diego, USA: Academic Press, 1986. p.249-264.

- LANTUÉJOUL, C. Ergodicity e Integral Range. **Journal of Microscopy**, v.161, p.387-403, 1991.
- LANTUÉJOUL, C. Non Conditional Simulation of Stationary Isotropic Multigaussian Random Functions. In: ARMSTRONG, M.; DOWD, P.A. eds, **Geostatistical Simulations**. Dordrecht, Holland: Kluwer Academics Publishers, 1994. p.147-177.
- LIMIC, N.; MIKELIC, A. Constrained Kriging Using Quadratic Programming. **Mathematical Geology**, v.16, n.4, p.423-429, 1984.
- MADUREIRA, C. M.; CAVALHEIRO, A. A. T.; FIUZA, A. M. A. Geostatistics as the Art of Making Statistical Inference Upon a Single Sample, **Geosistemas**, n.3, p.49-67, 1994.
- MATHERON, G. et al. Conditional Simulation of the Geometry of Fluvio-Deltaic Reservoirs. **ANNUAL TECHNICAL CONFERENCE AND EXHIBITION OF THE SPE 62**, Dallas, 1987. p.123-131. (SPE Paper 16753)
- MATHERON, G. **Estimating and Choosing**. Tradução por A. M. Hasofer, Berlin, Germany: Springer-Verlag, 1989. 141p.
- OLIVEIRA, M. L.; REMACRE, A. Z. **Glossário Geoestatístico Básico**. Campinas, UNICAMP/AGP, 1997. (publicação interna, preprint)
- OLIVEIRA, M. L.; REMACRE, A. Z. Simulação P-field: uma proposta alternativa. **WORKSHOP SOBRE CARACTERIZAÇÃO E ENGENHARIA DE RESERVATÓRIOS 2**. Campinas, UNICAMP, 1997. p.12-14.
- PAPOULIS, A. **Probability, Random Variables and Stochastic Processes**. New York, USA: McGraw-Hill Book Company, 1965. 583p.
- REMACRE, A. Z. **Krigagem da Média, Krigagem Simples e Krigagem Ordinária**. Campinas, UNICAMP, 1995. p.1-6. (Notas de aula)
- REMACRE, A. Z. Correção dos Pesos da Krigagem Comparação da Proposição de Journel, Froideveaux e uma nova Proposta. **WORKSHOP SOBRE CARACTERIZAÇÃO E ENGENHARIA DE RESERVATÓRIOS 2**. Campinas, UNICAMP, 1997. p.7-9.
- RODRIGUES, L. G. **Consideração das Incertezas na Caracterização de Reservatórios Utilizando Ajuste de Histórico e Simulação Estocástica**. Campinas, 1996. 130p. Dissertação de Mestrado em Geoengenharia de Reservatórios, UNICAMP.
- ROTH, C.; ARMSTRONG, M. Confidence Intervals for Local estimation Application to the Witwatersrand Basin. **APCOM Proceedings 26**. Fontainebleau, France: Centre de Géostatistique, 1995.
- SRIVASTAVA, R. M. Reservoir Characterization With Probability Field Simulation. **ANNUAL TECHNICAL CONFERENCE AND EXHIBITION OF THE SPE 67**, Washington, USA, 1992. p.927-938. (SPE Paper 24753)

- SRIVASTAVA, R. M. The Visualization of Spatial Uncertainty. In: YARUS, J. M.; CHAMBERS, R. L. (eds.), **Stochastic Modeling and Geostatistics**. Tulsa, USA: AAPG, 1994a. p.339-345. (Computer Applications in Geology, n.3)
- SRIVASTAVA, R. M. An Overview of Stochastic Methods for Reservoir Characterization. In: YARUS, J. M.; CHAMBERS, R. L. eds, **Stochastic Modeling and Geostatistics**. Tulsa, USA: AAPG, 1994b. p.3-16. (Computer Applications in Geology, n.3)
- SRIVASTAVA, R. M. Matheronian Geostatistics Where is it Going. In: BAAFI, E. Y.; SCHOFIELD, N. A. (eds.), **Geostatistics Wollongong'96**. Dordrecht, Holland: Kluwer Academics Publishers, 1997. p.53-68.
- STARKS, T. H.; FANG, J. H. The Effect of the Drift on the Experimental Semivariogram. **Mathematical Geology**, v.14, n.4, p.309-319, 1982.
- SURO-PÉREZ, V.; JOURNEL, A. G. Stochastic Simulation of Lithofacies for Reservoir Characterization. In: **Stanford Center for Reservoir Forecasting**. Stanford, USA: Stanford University, 1990. (Report 3)
- SURO-PÉREZ, V.; JOURNEL, A. G. Indicator Principal Component Kriging. **Mathematical Geology**, v.23, n.5, p.759-788, 1991.
- SZIDAROVSKY, F.; BAAFI, E. Y.; KIM, Y. C. Kriging without Negative Weights. **Mathematical Geology**, v.19, n.6, p.549-559, 1987.
- TOLEDO, J. P. **Caracterização Geostatística em Simuladores Numéricos de Reservatório**. Campinas, 1990. 117p. Dissertação de Mestrado em Engenharia de Petróleo, UNICAMP.
- TRAN, T. Variogram Reproduction On Dense Simulation Grids. In: **Stanford Center for Reservoir Forecasting**. Stanford, USA: Stanford University, 1993. (Report 6)
- VERLY, G. Sequential Gaussian Co-Simulation A Simulation Method Integrating Several Types of Information. In: SOARES, A. (ed.), **Geostatistics Tróia'92**. Dordrecht, Holland: Kluwer Academics Publishers, 1993. p.543-554.
- XU, W.; TRAN, T.; SRIVASTAVA, R. M.; JOURNEL, A. G. Integrating Seismic Data in Reservoir Modeling. **ANNUAL TECHNICAL CONFERENCE AND EXHIBITION OF THE SPE 67**, Washington, USA, 1992. p.887-902. (SPE Paper 24742)
- XU, W.; JOURNEL, A. G. GTSIM - Gaussian Truncated Simulations of Lithofacies, In: **Stanford Center for Reservoir Forecasting**. Stanford, USA: Stanford University, 1993. (Report 6)
- ZHU, H.; JOURNEL, A. Formating and Integrating Soft Data Stochastic Imaging via the Markov-Bayes Algorithm. In: SOARES, A. (ed.), **Geostatistics Tróia'92**. Dordrecht, Holland: Kluwer Academics Publishers, 1993. p.1-12.

# Bibliografia

- CRESSIE, N. The Origins of Kriging. **Mathematical Geology**, v.22, n.3, p.239-252, 1990.
- DAVID, M. **Geostatistical Ore Reserve Estimation**. New York, USA: Elsevier Scientific, 1977. 364p. (Developments in Geomathematics 2)
- DAVIS, J. C. **Statistics and Data Analysis**. 2. ed. New York, USA: John Wiley & Sons, 1973. 646p.
- DOWD, P.A. Some Observations on Confidence Intervals and Kriging Errors. In: ARMSTRONG, M. (ed.), **Geostatistics**, 1988. Dordrecht, Holland: Kluwer Academics Publishers, 1989. p.861-874.
- DUBRULE, O. Introducing More Geology in Stochastic Reservoir Modeling. In: SOARES, A. (ed.), **Geostatistics Tróia'92**. Dordrecht, Holland: Kluwer Academics Publishers, 1993. p.351-370.
- FAYERS, F. J.; HEWETT, T. A. A Review of Current Trends in Petroleum Reservoir Description and Assessment of the Impacts on Oil Recovery. In: **Stanford Center for Reservoir Forecasting**. Stanford, USA: Stanford University, 1993. (Report 6)
- GALLI, A.; NEUILLET, F. G.; DADOU, C. Factorial Kriging Analysis A Substitute to Spectral Analysis of Magnetic Data. In: VERLY, G. et al. (eds.), **Geostatistics for Natural Resources Characterization**. Dordrecht, Holland: D. Reidel, 1984. p.543-558.
- GEOVARIANCES. Manuais do Software ISATIS, versão 2.2. Fontainebleau, France: Geovariances, 1994.
- GÓMEZ-HERNÁNDEZ, J. J.; CASSIRAGA, E. F. Theory and Practice of Sequential Simulation. In: ARMSTRONG, M.; DOWD, P.A. (eds.), **Geostatistical Simulations**. Dordrecht, Holland: Kluwer Academics Publishers, 1994. p.111-124.
- HEWETT, T. A. **Geostatistical Reservoir Characterization**. Campinas, UNICAMP, 1992. (Apostila de Seminário)
- HOHN, M. E. **Geostatistics and Petroleum Geology**. New York, USA: Van Nostrand Reinhold, 1988. 264p. (Computer Methods in the Geosciences)
- JOURNEL, A. G. **Geostatistics for the Environmental Sciences**. Stanford, USA: Stanford University, Applied Earth Science Department, 1987. 135p.
- JOURNEL, A. G. **Fundamentals of Geostatistics in Five Lessons**. Stanford, USA: Stanford University, Applied Earth Science, 1988.
- JOURNEL, A. G.; ROSSI, M. E. When do We Need a Trend Model in Kriging. **Mathematical Geology**, v.21, n.7, p.715-739, 1989.

- JOURNAL, A. G.; XU, W.; TRAN, T. Integrating Seismic Data in Reservoir Modeling The Collocated Cokriging Alternative. In: **Stanford Center for Reservoir Forecasting**. Stanford, USA: Stanford University, 1992. (Report 5)
- JOURNAL, A. Geostatistics Roadblocks and Challenges. In: SOARES, A. (ed.), **Geostatistics Tróia'92**. Dordrecht, Holland: Kluwer Academics Publishers, p.213-224. 1993.
- KOVÁCS, Z. L. **Teoria da Probabilidade e Processos Estocásticos**. São Paulo: Edição Acadêmica, 1996. 120p.
- LANTUÉJOUL, C. **Random Sets and Functions from Models to Conditional Simulations**. Campinas, UNICAMP, 1995. 141p. (Apostila de Seminário)
- REMACRE, A. Z. **Coestimativas**. Campinas, UNICAMP, 1995, p.1-13. (Notas de aula)
- RIPLEY, B. D. **Stochastic Simulation**. New York, USA: John Wiley & Sons 1987. 237p. (Wiley Series in Probability and Mathematical Statistics)
- SEN, M. K.; STOFFA, P.L.; POPE, G. A. Stochastic reservoir Modeling Using Simulated *Annealing* and Genetic Algorithm. **ANNUAL TECHNICAL CONFERENCE AND EXHIBITION OF THE SPE 67**, Washington, USA, 1992. p.939-950. (SPE Paper 24754)
- STOYAN, D.; KENDALL, W. S.; MECKE, J. **Stochastic Geometry and its Applications**. Berlin, Germany: John Wiley & Sons, 1987. 345p. (Probability and Mathematical Statistics Applied)
- WACKERNAGEL, H. **Multivariate Geostatistics; an introduction with applications**. Berlin, Germany: Springer-Verlag, 1995. 257p.
- WALPOLE, R. E.; MYERS, R. H. **Probability and Statistics for Engineers and Scientists**. 3. ed. New York, USA: MacMillan Publishing Company, 1985. 639p.
- ZIMMERMAN, D. A.; WILSON, J. L. **Description of and User's Manual for TUBA a Computer Code for Generating Two-dimensional Random fields via the Turning Bands Method**. Albuquerque, USA: SEASOFT, 1990. 114p.