

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
Escola de Engenharia
Programa de Pós-Graduação em Engenharia de Minas, Metalúrgica e Materiais
PPGEM

AUMENTO DA EFICIÊNCIA DOS MÉTODOS SEQUENCIAIS DE SIMULAÇÃO
CONDICIONAL

Gustavo Grangeiro Pilger

Tese para obtenção do título de
Doutor em Engenharia

Porto Alegre, RS
2005

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
Escola de Engenharia
Programa de Pós-Graduação em Engenharia de Minas, Metalúrgica e Materiais
PPGEM

AUMENTO DA EFICIÊNCIA DOS MÉTODOS SEQUENCIAIS DE SIMULAÇÃO
CONDICIONAL

Gustavo Grangeiro Pilger

Engenheiro de Minas

Tese apresentada ao programa de Pós-Graduação em Engenharia de Minas, Metalúrgica e dos Materiais - PPGEM, como parte dos requisitos para a obtenção do título de Doutor em Engenharia.

Área de concentração: Metalurgia Extrativa e Tecnologia Mineral

Porto Alegre, RS

2005

Esta Tese foi julgada adequada para a obtenção do Título de Doutor em Engenharia e aprovada em sua forma final pelo Orientador e pela Banca Examinadora do Curso de Pós-Graduação.

Orientador: Prof. Dr. João Coimbra Leite Costa

Banca Examinadora:

Prof. Dr. André Cezar Zingano

Prof. Dr. Marcelo C. Godoy

Prof. Dr. Giorgio Francesco Cesare de Tomi

Prof. Dr. Antônio Cezar Faria Vilela
Coordenador do PPGEM

"A análise dos dados é uma ferramenta para separar, da ganga dos dados, o diamante puro da verdade".

Jean - Paul Benzécri

À Ana Paula, pelo constante
companheirismo, dedicação e amor,
fundamentais à conclusão desse
trabalho.

AGRADECIMENTOS

Ao Professor Dr. João Felipe C. L. Costa pela orientação, confiança e amizade.

Ao LPM (Laboratório de Pesquisa e Planejamento Mineiro) e à todo o pessoal que o constitui. Agradeço em especial aos professores Jair C. Koppe e André Zingano. E aos colegas: Enrique Munaretti, José Antônio Kurcewicz, Sérgio Klein, Luis Eduardo de Souza, Fernando Gambin, Leandro José de Oliveira, Keli Cristiane Corrêa, Mario Goulart, Vanessa C. Koppe, Daiane Folle e Simone Picarelli pelo companheirismo e bons momentos no trabalho diário.

À UFRGS, em especial ao DEMIN, por oferecerem condições para o desenvolvimento de pesquisa.

Ao Centro de Geoestatística de Fontainebleau da ENSMP, em especial aos professores: Jacques Rivoirard, Christian Lantuéjoul, Jean-Paul Chilès, Gaëlle Le L'och e Chantal de Fouquet.

Aos meus pais, pelo constante e insubstituível apoio, por terem proporcionado a educação e incentivo.

Ao CNPq pela bolsa concedida.

Índice

Lista de figuras	xii
Lista de quadros e tabelas	xxv
Resumo	xxix
Abstract	xxx
Capítulo 1 – Introdução	1
1.1 Abordagem probabilística e o estimador ótimo.....	2
1.2 Simulação Geoestatística.....	4
1.2.1 Simulação geoestatística: histórico.....	5
1.2.2 Simulação geoestatística: últimas pesquisas.....	6
1.2.3 Simulação geoestatística na mineração	7
1.3 Problema	8
1.4 Meta.....	9
1.5 Objetivos	10
1.6 Organização da Tese	11
Capítulo 2 – Conceitos introdutórios de geoestatística	12
2.1 Modelos probabilísticos	12
2.2 Variáveis regionalizadas	13
2.3 Função aleatória e variável aleatória.....	13
2.4 Hipóteses estacionária e intrínseca.....	14
2.5 Covariância espacial	16
2.6 Variância de combinações lineares autorizadas, CLA.....	17
2.7 Equações de krigagem e suas relações com as hipóteses estacionária e intrínseca.....	19

2.7.1	Krigagem ordinária, KO	19
2.7.2	Krigagem simples, KS.....	19
2.8	Comentários.....	20
Capítulo 3	– Simulação condicional seqüencial.....	21
3.1	Métodos de Monte Carlo: conceitos	21
3.2	Simulação estocástica.....	22
3.3	Simulação condicional.....	23
3.3.1	Simulação condicional seqüencial	24
3.3.1.1	Simulação seqüencial Gaussiana	26
3.3.1.2	Simulação seqüencial dos indicadores	27
3.4	Distribuições condicionais locais de probabilidade, dclp.....	27
3.5	O modelo multiGaussiano – método paramétrico	29
3.5.1	Anamorfose Gaussiana	32
3.5.2	Verificação de binormalidade	35
3.5.3	Discretização da dclp cumulativa experimental	35
3.5.4	Considerações finais sobre a anamorfose Gaussiana	41
3.6	Método dos Indicadores – método não-paramétrico	41
3.6.1	Codificação binária das VA.....	43
3.6.2	Krigagem dos Indicadores	45
3.6.2.1	Krigagem Ordinária dos Indicadores.....	46
3.6.2.2	Krigagem Simples dos Indicadores.....	47
3.6.3	Cokrigagem dos Indicadores.....	47
3.6.3.1	Cokrigagem Ordinária dos Indicadores	47
3.6.3.2	Cokrigagem Simples dos Indicadores	50
3.6.3.3	Krigagem da probabilidade	51
3.6.4	KI X CKI.....	52
3.6.4.1	Modelo Linear de Correionalização	52
3.6.4.2	Simplificação da cokrigagem	55
3.6.4.2.1	Krigagem dos indicadores da mediana.....	55
3.6.4.2.2	Krigagem fatorial multivariada.....	56
3.6.4.2.3	Krigagem dos componentes principais dos indicadores.....	58
3.6.5	Correção para os desvios de relação de ordem.....	62
3.6.6	Discretização da dclp experimental cumulativa não-paramétrica.....	64
3.7	Algoritmos de ssG e ssi	66
3.7.1	Algoritmos de ssG.....	66
3.7.2	Algoritmos de ssi	67

3.8 Comentários	67
Capítulo 4 – Incerteza e funções de transferência	69
4.1 Incerteza local x incerteza espacial.....	69
4.2 Krigagem multiGaussiana	71
4.3 Simulação estocástica x krigagem multiGaussiana	72
4.4 Suporte das medidas x suporte dos modelos de incerteza	73
4.5 Funções de transferência	77
4.6 Propagação da incerteza por meio de funções de transferência	78
4.6.1 Latin Hypercube Sampling (LHS)	79
4.7 LHS x simulação de Monte Carlo	81
4.8 Aumento da eficiência das técnicas LHS e de Monte Carlo.....	82
4.9 Avaliação da qualidade dos modelos locais de incerteza.....	84
4.10 Reprodução dos momentos estatísticos da distribuição experimental	89
4.11 Índices locais de incerteza	91
4.11.1 Variância Condicional	91
4.11.2 Coeficiente de variação.....	92
4.11.3 Intervalo entre quartis.....	92
4.12 Medida da incerteza a partir de krigagem.....	94
4.13 Comentários	94
Capítulo 5 – LHS aplicado aos algoritmos de simulação seqüencial.....	96
5.1 LHS: histórico	96
5.2 Quadrados Latinos (<i>Latin Squares</i>).....	97
5.3 Latin Hypercube Sampling, LHS	99
5.4 Monte Carlo nos algoritmos simulação condicional seqüencial.....	104
5.5 LHS nos algoritmos simulação condicional seqüencial	106
5.5.1 Variações no algoritmo.....	109
5.5.1.1 LHS determinístico	110
5.5.1.2 LHS cíclico	110
5.5.1.3 LHS com caminho determinístico	111
5.5.2 Implementação do algoritmo	112
5.5.2.1 Definição das probabilidades $plhs_{ad}$	112
5.5.2.2 Inicialização da simulação.....	112
5.5.2.3 Implementação do algoritmo cíclico	113
5.5.2.4 Tiragens de classes.....	113
5.5.2.5 Definição das probabilidades $plhs$	114

5.5.2.6	Determinação dos valores simulados.....	114
5.5.2.7	Associação dos valores simulados.....	114
5.5.3	Verificações teóricas.....	115
5.6	Comentários.....	124
Capítulo 6 –	Estudo de caso: LHS aplicado.....	125
6.1	Banco de dados.....	125
6.2	Análise exploratória dos dados.....	126
6.2.1	Modelo de variabilidade espacial.....	127
6.2.2	Malha de simulação.....	128
6.3	Anamorfose Gaussiana.....	128
6.3.1	Verificação de biGaussianidade.....	129
6.3.2	Modelo de variabilidade espacial Gaussiano.....	131
6.3.3	Validação do modelo de variabilidade espacial e estratégia de busca.....	132
6.4	Simulação sequencial condicional.....	133
6.4.1	Algoritmo original.....	135
6.4.2	Algoritmo LHS clássico.....	138
6.4.2.1	Validação prática do algoritmo LHS clássico.....	144
6.4.3	Algoritmo LHS cíclico clássico.....	146
6.4.3.1	M = 4, 25 ciclos.....	147
6.4.3.2	M = 5, 20 ciclos.....	149
6.4.3.3	M = 6, 16 ciclos.....	150
6.4.3.4	M = 7, 14 ciclos.....	151
6.4.3.5	M = 8, 12 ciclos.....	153
6.4.3.6	M = 9, 11 ciclos.....	154
6.4.3.7	M = 10, 10 ciclos.....	156
6.4.3.8	M = 11, 9 ciclos.....	157
6.4.3.9	M = 12, 8 ciclos.....	158
6.4.3.10	M = 13, 7 ciclos.....	160
6.4.3.11	M = 14, 7 ciclos.....	161
6.4.3.12	M = 15, 6 ciclos.....	162
6.4.3.13	M = 16, 6 ciclos.....	164
6.4.3.14	M = 17, 5 ciclos.....	165
6.4.3.15	M = 18, 5 ciclos.....	166
6.4.3.16	M = 19, 5 ciclos.....	167
6.4.3.17	M = 20, 5 ciclos.....	169
6.4.3.18	Validação prática do algoritmo LHS cíclico clássico.....	170

6.4.4	Algoritmo LHS determinístico	171
6.4.4.1	Validação prática do algoritmo LHS determinístico	174
6.4.5	Algoritmo LHS cíclico determinístico	175
6.4.6	Estratificação da dclp	177
6.4.7	Algoritmo LHS com caminho determinístico	183
6.4.8	Comparativos entre os algoritmos de simulação (MC x LHS)	185
6.4.8.1	Reprodutibilidade dos resultados	191
6.4.8.2	Caracterização da incerteza x curvas de parametrização.....	194
6.4.8.3	Caracterização da incerteza x curvas de parametrização – Fe.....	196
6.5	Comentários	200
Capítulo 7 – Conclusões e recomendações		201
7.1	Resumo dos primeiros capítulos	201
7.2	Meta e objetivo x resultados teóricos e práticos	202
7.3	Recomendações.....	205
Referências bibliográficas.....		207
Apêndice A – Decomposição de matrizes e ACP		223
A.1	Valores próprios ou autovalores.....	223
A.2	Vetores próprios ou autovetores	223
A.3	Decomposição em valores e vetores próprios	224
A.4	Análise dos componentes principais	224
A.5	Correlação das variáveis com os fatores	225
A.6	Decomposição em valores singulares.....	225
Apêndice B – Reprodução da função de covariância espacial.....		226
B.1	Algoritmo original (MC).....	226
B.2	Algoritmo modificado (LHS).....	227

Lista de figuras

Figura 1.1	Estimativas de teor por krigagem ordinária (KO) (gráfico da esquerda) e suas variâncias (gráfico da direita), diminuindo a medida que se aproxima dos dados. Fonte: Goovaerts, 1997 – p. 181	4
Figura 3.1	Curva da distribuição Gaussiana.....	29
Figura 3.2	Distribuições condicionais locais de probabilidade (dclp) cumulativas para o caso de $y(u)$ ser conhecido (a) e para o caso onde as informações circunvizinhas de u estiverem além da distância de correlação espacial (b)	31
Figura 3.3	Função de anamorfose. Fonte: Rivoirard, 1991 p. 40.....	33
Figura 3.4	Funções de repartição $F(z)$ e $G(y)$, mostrando que o valor de $y(u)$ associado ao de $z(u)$ é tal que $F(u) = G(u)$. Fonte: Rivoirard, 1991 p. 40	33
Figura 3.5	Transformação inversa, limitada pelo nº de informações z	36
Figura 3.6	Anamorfose empírica	36
Figura 3.7	Modelo potência de acordo o valor atribuído a w	37
Figura 3.8	Modelo hiperbólico, de acordo o valor atribuído a w	38
Figura 3.9	Modelo de incerteza para a localização u (dclp) a partir do método não-paramétrico. Dez valores limite z_k ($k=1,\dots,10$) foram usados para a construção da dclp	42
Figura 3.10	Esquema representando a relação de correlação espacial entre os fatores regionalizados.....	57

Figura 4.1	Modelos de incerteza local inferidos por krigagem multiGaussiana (linha tracejada) e por meio de simulação estocástica (linha pontilhada). Fonte: Goovaerts (2001)	73
Figura 4.2	Gráficos de exatidão, considerando intervalos simétricos de probabilidade (p-IP) (a) e intervalos ilimitados de probabilidade (b). Fonte: Goovaerts (2001)	87
Figura 4.3	Gráfico mostrando a amplitude média dos intervalos de valores de acordo com o intervalo simétrico de probabilidade p esperada. Fonte: Goovaerts (2001)	88
Figura 4.4	Flutuações na reprodução do momento de 2ª ordem da distribuição condicionante nas direções de maior (a) e de menor continuidade espacial (b). Fonte: Pilger (2000)	89
Figura 4.5	Flutuações ergódicas na reprodução do momento de 1ª ordem da distribuição condicionante (histograma central). Fonte: Pilger (2000)	90
Figura 4.6	Concentração (ppm) de acordo com a distância linear entre as localizações medidas (figura superior esquerda), juntamente com três gráficos comparativos de variabilidade em cada posição do perfil. Fonte: Goovaerts (1997 p. 339)	93
Figura 5.1	Exemplo de LHS, com $M = 6$, para duas VAs (A e B) independentes e uniformes $U[0;1]$	100
Figura 5.2	LHS para duas VAs Gaussianas dependentes (Tabela 5.2), de acordo com o procedimento proposto por Stein (1987). Amostragem aleatória simples (●) translada para LHS (●), onde as linhas contínuas entre símbolos indicam a translação	104
Figura 5.3	Tiragem aleatória de um n^o Gaussiano, associado à localização u_j , a partir da 4ª classe da distribuição $G(y)$; $M = 14$	108
Figura 5.4	Quadro de tiragens de classes (A, B e C) em um caso em que $N = 16$ e $M = L = 3$ (esquerda); domínios das classes na dp $G(y)$ (direita).....	115
Figura 5.5	Quadrados Latinos para pares de localizações do exemplo da Figura 5.4	115
Figura 5.6	Quadro de tiragens de classes (A, B e C) em um caso em que $j = 16$ e $M = L = 3$, para dois ciclos distintos	116
Figura 5.7	Representação gráfica dos resultados apresentados na Tabela 5.3, obtidos segundo o procedimento descritivo	117

Figura 5.8	Representação gráfica dos resultados apresentados na Tabela 5.4, obtidos segundo o procedimento clássico	117
Figura 5.9	Distribuições experimentais de probabilidade Gaussiana a partir de 10 realizações, de acordo com o método de Monte Carlo (cima esquerda), LHS (cima direita) e LHS descritivo (baixo esquerda). Gráfico de probabilidade, segundo os três métodos (baixo direita)	118
Figura 5.10	Distribuições experimentais de probabilidade Gaussiana a partir de 15 realizações, de acordo com o método de Monte Carlo (cima esquerda), LHS (cima direita) e LHS descritivo (baixo esquerda). Gráfico de probabilidade, segundo os três métodos (baixo direita)	119
Figura 5.11	Distribuições experimentais de probabilidade Gaussiana a partir de 20 realizações, de acordo com o método de Monte Carlo (cima esquerda), LHS (cima direita) e LHS descritivo (baixo esquerda). Gráfico de probabilidade, segundo os três métodos (baixo direita)	120
Figura 5.12	Distribuições experimentais de probabilidade Gaussiana a partir de 30 realizações, de acordo com o método de Monte Carlo (cima esquerda), LHS (cima direita) e LHS descritivo (baixo esquerda). Gráfico de probabilidade, segundo os três métodos (baixo direita)	121
Figura 5.13	Distribuições experimentais de probabilidade Gaussiana a partir de 50 realizações, de acordo com o método de Monte Carlo (cima esquerda), LHS (cima direita) e LHS descritivo (baixo esquerda). Gráfico de probabilidade, segundo os três métodos (baixo direita)	122
Figura 5.14	Distribuições experimentais de probabilidade Gaussiana a partir de 100 realizações, de acordo com o método de Monte Carlo (cima esquerda), LHS (cima direita) e LHS descritivo (baixo esquerda). Gráfico de probabilidade, segundo os três métodos (baixo direita)	122
Figura 5.15	Distribuições experimentais de probabilidade Gaussiana, a partir de processos de simulação distintos de 20 realizações, de acordo com o método de Monte Carlo (esquerda) e LHS (direita).....	123
Figura 6.1	Mapa de localização dos furos de sondagem, com símbolos proporcionais aos teores (ppm) associados a variável V.....	125
Figura 6.2	Histograma e parâmetros estatísticos da distribuição da variável V (ppm).....	126

Figura 6.3	Mapa de localização dos furos de amostragem, de acordo com os quartis da distribuição da variável V (ppm).....	127
Figura 6.4	Variogramas experimentais (D) e modelos (M) para quatro direções. D1 e M1: azimute 158°; D2 e M2: azimute 113°; D3 e M3: azimute 68° e D4 e M4: azimute 23°	127
Figura 6.5	Anamorfose Gaussiana, empírica (linha espessa preta) e modelada (linha fina azul).....	128
Figura 6.6	Histograma e parâmetros estatísticos da distribuição da variável Gaussiana	129
Figura 6.7	Teste de binormalidade da razão madograma/semi-variograma para os madogramas/semi-variogramas onidirecionais. A linha vermelha representa o valor ideal para a relação (Equação 3.25).....	129
Figura 6.8	Histogramas de frequência cumulativa e parâmetros estatísticos para as distribuições da diferença de pares de dados, separados pelo vetor h onidirecional	130
Figura 6.9	Variogramas experimentais (D) e modelos (M) da variável Gaussiana, para quatro direções. D1 e M1: azimute 158°; D2 e M2: azimute 113°; D3 e M3: azimute 68° e D4 e M4: azimute 23°	131
Figura 6.10	Histograma da distribuição dos erros normalizados sobre os desvios padrão das estimativas (esquerda) e diagrama de dispersão entre os valores verdadeiros e estimados (direita)	132
Figura 6.11	Mapa proporcional do erro bruto, indicando as localizações onde os dados foram subestimados e superestimados, segundo o processo de validação cruzada.....	133
Figura 6.12	Esquema da estratégia de busca utilizada nos processos de simulação	134
Figura 6.13	Arquivo de parâmetros dos processos de simulação (padrão GSLIB)	134
Figura 6.14	Médias (a) e desvio padrão (b) das 200 realizações obtidas por meio do algoritmo original – no espaço normal	135
Figura 6.15	Histogramas das médias (esquerda) e dos desvios padrão (direita) das 200 realizações obtidas por meio do algoritmo original – no espaço normal.....	135
Figura 6.16	Médias (a) e desvio padrão (b) das 200 realizações obtidas por meio do algoritmo original – no espaço original.....	136

Figura 6.17	Histogramas das médias (esquerda) e dos desvios padrão (direita) das 200 realizações obtidas por meio do algoritmo original – no espaço original	136
Figura 6.18	Média (a) e CV (b) da distribuição das médias das realizações, de acordo com a rotina <i>sgsim</i> original.....	138
Figura 6.19	Médias (esquerda) e desvios padrão (direita) das realizações, resultantes dos processos de simulação a partir de 100 classes de discretização (a) e de 50 classes de discretização (b) – no espaço normal	139
Figura 6.20	Médias (esquerda) e desvios padrão (direita) das realizações, resultantes dos processos de simulação a partir de 100 classes de discretização (a) e de 50 classes de discretização (b) – no espaço original.....	140
Figura 6.21	Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS clássico; a linha vermelha representa os valores considerados estáveis	141
Figura 6.22	Médias (esquerda) e desvios padrão (direita) das realizações, resultantes dos processos de simulação a partir de 15 classes de discretização, no espaço normal	142
Figura 6.23	Médias (esquerda) e desvios padrão (direita) das realizações, resultantes dos processos de simulação a partir de 15 classes de discretização, no espaço original	143
Figura 6.24	Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS clássico. A linha vermelha representa os valores considerados estáveis	144
Figura 6.25	Histograma de frequência de tiragens de classes ao longo de uma realização ...	145
Figura 6.26	Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de 25 ciclos de quatro realizações, no espaço normal (a) e espaço original (b).....	147
Figura 6.27	Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (25 ciclos de quatro realizações). A linha vermelha representa os valores considerados estáveis.....	148
Figura 6.28	Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de 20 ciclos de cinco realizações, no espaço normal (a) e espaço original (b).....	149

Figura 6.29	Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (20 ciclos de cinco realizações). A linha vermelha representa os valores considerados estáveis.....	149
Figura 6.30	Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de 16 ciclos de seis realizações, no espaço normal (a) e espaço original (b).....	150
Figura 6.31	Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (16 ciclos de seis realizações). A linha vermelha representa os valores considerados estáveis.....	151
Figura 6.32	Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de 14 ciclos de sete realizações, no espaço normal (a) e espaço original (b).....	152
Figura 6.33	Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (14 ciclos de sete realizações). A linha vermelha representa os valores considerados estáveis.....	153
Figura 6.34	Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de 12 ciclos de oito realizações, no espaço normal (a) e espaço original (b).....	153
Figura 6.35	Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (12 ciclos de oito realizações). A linha vermelha representa os valores considerados estáveis.....	154
Figura 6.36	Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de 11 ciclos de nove realizações, no espaço normal.....	154
Figura 6.37	Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de 11 ciclos de nove realizações, no espaço original.....	155
Figura 6.38	Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (11 ciclos de nove realizações). A linha vermelha representa os valores considerados estáveis.....	155
Figura 6.39	Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de 10 ciclos de 10 realizações, no espaço normal (a) e espaço original (b).....	156

Figura 6.40	Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (10 ciclos de 10 realizações). A linha vermelha representa os valores considerados estáveis.....	157
Figura 6.41	Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de nove ciclos de 11 realizações, no espaço normal (a) e espaço original (b).....	157
Figura 6.42	Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (nove ciclos de 11 realizações). A linha vermelha representa os valores considerados estáveis.....	158
Figura 6.43	Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de oito ciclos de 12 realizações, no espaço normal (a) e espaço original (b).....	159
Figura 6.44	Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (oito ciclos de 12 realizações). A linha vermelha representa os valores considerados estáveis.....	159
Figura 6.45	Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de sete ciclos de 13 realizações, no espaço normal (a) e espaço original (b).....	160
Figura 6.46	Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (sete ciclos de 13 realizações). A linha vermelha representa os valores considerados estáveis.....	161
Figura 6.47	Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de sete ciclos de 14 realizações, no espaço normal (a) e espaço original (b).....	161
Figura 6.48	Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (sete ciclos de 14 realizações). A linha vermelha representa os valores considerados estáveis.....	162
Figura 6.49	Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de seis ciclos de 15 realizações, no espaço normal.....	162
Figura 6.50	Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de seis ciclos de 15 realizações, no espaço original.....	163

Figura 6.51	Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (seis ciclos de 15 realizações). A linha vermelha representa os valores considerados estáveis.....	163
Figura 6.52	Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de seis ciclos de 16 realizações, no espaço normal (a) e espaço original (b).....	164
Figura 6.53	Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (seis ciclos de 16 realizações). A linha vermelha representa os valores considerados estáveis.....	165
Figura 6.54	Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de cinco ciclos de 17 realizações, no espaço normal (a) e espaço original (b).....	165
Figura 6.55	Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (cinco ciclos de 17 realizações). A linha vermelha representa os valores considerados estáveis.....	166
Figura 6.56	Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de cinco ciclos de 18 realizações, no espaço normal.....	166
Figura 6.57	Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de cinco ciclos de 18 realizações, no espaço original.....	167
Figura 6.58	Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (cinco ciclos de 18 realizações). A linha vermelha representa os valores considerados estáveis.....	167
Figura 6.59	Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de cinco ciclos de 19 realizações, no espaço normal (a) e espaço original (b).....	168
Figura 6.60	Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (cinco ciclos de 19 realizações). A linha vermelha representa os valores considerados estáveis.....	168
Figura 6.61	Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de cinco ciclos de 20 realizações, no espaço normal (a) e espaço original (b).....	169

Figura 6.62	Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (cinco ciclos de 20 realizações). A linha vermelha representa os valores considerados estáveis.....	170
Figura 6.63	Distribuição experimental de probabilidade Gaussiana, associada à localização u_1 , para o caso no qual $M = 10$, resultante da aplicação do algoritmo LHS cíclico clássico, em três ciclos, totalizando 30 tiragens	171
Figura 6.64	Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação (LHS determinístico) a partir de 20 realizações ($M = 20$), no espaço normal (a) e espaço original (b).....	172
Figura 6.65	Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação (LHS determinístico) a partir de 60 realizações ($M = 60$), no espaço normal (a) e espaço original (b).....	173
Figura 6.66	Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS determinístico. A linha vermelha representa os valores considerados estáveis	174
Figura 6.67	Distribuição experimental de probabilidade Gaussiana, associada à localização u_1 , para o caso no qual $M = 10$, resultante da aplicação do algoritmo LHS determinístico	175
Figura 6.68	Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação (LHS cíclico determinístico) a partir de 10 ciclos de 10 realizações, no espaço normal (a) e espaço original (b)	176
Figura 6.69	Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico determinístico (10 ciclos de 10 realizações). A linha vermelha representa os valores considerados estáveis.....	176
Figura 6.70	Modelos de incerteza (dclp) associados à localização inserida em uma zona de baixo teor, de acordo com os algoritmos LHS clássico (a) e com o algoritmo original (b), nos espaços normal (esquerda) e original (direita)	177
Figura 6.71	Distribuições experimentais de probabilidade Gaussiana, de acordo com o algoritmo LHS clássico (a) e original (b), utilizadas para a determinação dos valores simulados da Figura 6.70	178
Figura 6.72	Modelos de incerteza (dclp) associados à localização inserida em uma zona de médio teor, de acordo com os algoritmos LHS clássico (a) e com o algoritmo original (b), nos espaços normal (esquerda) e original (direita)	179

Figura 6.73	Distribuições experimentais de probabilidade Gaussiana, de acordo com o algoritmo LHS clássico (a) e original (b), utilizadas para a determinação dos valores simulados da Figura 6.72	180
Figura 6.74	Modelos de incerteza (dclp) associados à localização inserida em uma zona de alto teor, de acordo com os algoritmos LHS clássico (a) e com o algoritmo original (b), nos espaços normal (esquerda) e original (direita).....	180
Figura 6.75	Distribuições experimentais de probabilidade Gaussiana, de acordo com o algoritmo LHS clássico (a) e original (b), utilizadas para a determinação dos valores simulados da Figura 6.74	181
Figura 6.76	Modelos de incerteza (dclp) associados à localização inserida em uma zona de médio teor, de acordo com os algoritmos LHS cíclico clássico (a) e LHS determinístico (b), nos espaços normal (esquerda) e original (direita)	182
Figura 6.77	Modelos de incerteza (dclp) associados à localização u_1 , de acordo com o algoritmo LHS clássico com caminho determinístico, nos espaços normal (a) e original (b)	184
Figura 6.78	Modelos de incerteza (dclp) associados à localização u_1 , de acordo com o algoritmo LHS clássico com caminho determinístico, nos espaços normal (a) e original (b)	185
Figura 6.79	Média (a) e CV (b) da distribuição das médias das realizações, de acordo com os algoritmos original e LHS clássico. A linha horizontal vermelha indica os respectivos valores considerados estáveis	186
Figura 6.80	Média (a) e CV (b) da distribuição das médias das realizações, de acordo com os algoritmos original e LHS determinístico. A linha horizontal vermelha indica os respectivos valores considerados estáveis	187
Figura 6.81	Média (a) e CV (b) da distribuição das médias das realizações, de acordo com os algoritmos original e LHS cíclico clássico (25 ciclos de quatro realizações). A linha horizontal vermelha indica os respectivos valores considerados estáveis	188
Figura 6.82	Média da distribuição das médias das realizações, de acordo com os algoritmos original e LHS cíclico clássico (12 ciclos de oito realizações). A linha horizontal vermelha indica os respectivos valores considerados estáveis ...	188

Figura 6.83	CV da distribuição das médias das realizações, de acordo com os algoritmos original e LHS cíclico clássico (12 ciclos de oito realizações). A linha horizontal vermelha indica os respectivos valores considerados estáveis.....	189
Figura 6.84	Média (a) e CV (b) da distribuição das médias das realizações, de acordo com os algoritmos original e LHS cíclico clássico (10 ciclos de 10 realizações). A linha horizontal vermelha indica os respectivos valores considerados estáveis ...	189
Figura 6.85	Média (a) e CV (b) da distribuição das médias das realizações, de acordo com os algoritmos original e LHS cíclico clássico (seis ciclos de 15 realizações). A linha horizontal vermelha indica os respectivos valores considerados estáveis ...	190
Figura 6.86	Modelos de incerteza (dclp) associados à localização inserida em uma zona de médio teor, de acordo com os algoritmos LHS clássico, LHS cíclico clássico, LHS determinístico e original (MC), nos espaços normal (a) e original (b).....	191
Figura 6.87	Média (a) e CV (b), da distribuição das médias das realizações, segundo os processos de simulação (LHS clássico) e de acordo os modelos de discretização.....	192
Figura 6.88	Erro médio apresentado pela média e CV, da distribuição das médias das realizações, segundo os modelos de discretização.....	193
Figura 6.89	Soma dos erros médios absolutos da média e do CV, da distribuição das médias das realizações, segundo os modelos de discretização.....	193
Figura 6.90	Curvas de parametrização ($t \times cutoff$) produzidas a partir de 14 realizações geradas por meio dos algoritmos original (a) e LHS clássico (b).....	194
Figura 6.91	Amplitude das curvas de parametrização ($t \times cutoff$), segundo os algoritmos original (linha preta) e LHS clássico (linha vermelha).....	194
Figura 6.92	CV da distribuição dos 14 valores de tonelagem, de acordo com os algoritmos original e LHS clássico.....	195
Figura 6.93	CV da distribuição de 15 valores de tonelagem, de acordo com os algoritmos original e LHS clássico.....	195
Figura 6.94	CV da distribuição de 20 valores de tonelagem, de acordo com os algoritmos original e LHS clássico.....	195
Figura 6.95	CV da distribuição de 60 valores de tonelagem, de acordo com os algoritmos original e LHS clássico.....	196

Figura 6.96	Curvas de parametrização ($t \times cutoff$) produzidas a partir de 20 realizações geradas por meio dos algoritmos original (a) e LHS clássico (b) – estudo de caso: Fe.....	197
Figura 6.97	CV da distribuição dos 20 valores de tonelagem, de acordo com os algoritmos original e LHS clássico - estudo de caso: Fe.....	197
Figura 6.98	CV da distribuição dos 14 valores de tonelagem, de acordo com os algoritmos original e LHS clássico - estudo de caso: Fe.....	198
Figura 6.99	CV da distribuição dos 10 valores de tonelagem, de acordo com os algoritmos original e LHS clássico - estudo de caso: Fe.....	198
Figura 6.100	Histograma cumulativo dos 20 valores de tonelagem, segundo os algoritmos MC e LHS clássico, de acordo com o <i>cutoff</i> de 48 % - estudo de caso: Fe.....	199
Figura 6.101	Histograma cumulativo dos 14 valores de tonelagem, segundo os algoritmos MC e LHS clássico, de acordo com o <i>cutoff</i> de 48 % - estudo de caso: Fe.....	199
Figura 6.102	Histograma cumulativo dos 10 valores de tonelagem, segundo os algoritmos MC e LHS clássico, de acordo com o <i>cutoff</i> de 48 % - estudo de caso: Fe.....	200
Figura B.1	Semi-variogramas experimentais, associados à direção de menor variabilidade espacial (azimute 158°), referentes a cinco realizações produzidas por meio do algoritmo original, superpostas ao respectivo semi-variograma condicionante.	227
Figura B.2	Semi-variogramas experimentais, associados à direção de maior variabilidade espacial (azimute 68°), referentes a cinco realizações produzidas por meio do algoritmo original, superpostas ao respectivo semi-variograma condicionante. ..	227
Figura B.3	Semi-variogramas experimentais, associados aos azimutes 158° (a) e 68° (b), referentes a cinco realizações produzidas por meio do algoritmo modificado (LHS), a partir de 10 classes de discretização ($M = 10$), superpostas aos respectivos semi-variogramas condicionantes	228
Figura B.4	Semi-variogramas experimentais, associados aos azimutes 158° (a) e 68° (b), referentes a cinco realizações produzidas por meio do algoritmo modificado (LHS) a partir de 14 classes de discretização ($M = 14$), superpostas aos respectivos semi-variogramas condicionantes	228

Figura B.5 Semi-variogramas experimentais, associados aos azimutes 158° (a) e 68° (b), referentes a cinco realizações produzidas por meio do algoritmo modificado (LHS) a partir de 20 classes de discretização ($M = 20$), superpostas aos respectivos semi-variogramas condicionantes. 229

Lista de quadros e tabelas

Quadro 3.1	Relação do valor dos indicadores segundo cada valor limite z_k , para as quatro informações $z(u_\alpha)$	44
Tabela 3.1	Intervalos de probabilidade p-IP, indicando as probabilidades p e seus intervalos IP associados	85
Quadro 4.1	Valores assumidos por PI, conforme as características da dclp.....	82
Tabela 5.1	Exemplo numérico da Figura 5.1, LHS de dimensão 2, para duas VAs independentes e uniformes $U[0; 1]$	101
Tabela 5.2	LHS de dimensão 2, para duas VAs Gaussianas ($VA \sim N[0;1]$), correlacionadas (correlação = -0,53)	103
Tabela 5.3	Exemplo de três tiragens de $plhs_{ad}$, e valores Gaussianos x_{plhs_r} , de acordo com o procedimento descritivo.....	116
Tabela 5.4	Exemplo de três tiragens de $plhs$, e valores Gaussianos x_{plhs_r} , de acordo com o procedimento clássico	117
Tabela 6.1	Parâmetros da estratégia de busca	132
Tabela 6.2	Comparação estatística entre os erros robustos e brutos, após o processo de validação cruzada	132
Tabela 6.3	Média, desvio padrão e CV dos conjuntos de médias das realizações (c).....	137
Tabela 6.4	Média das médias e dos desvios padrão para 10 processos de simulação, conforme o número de classes M – espaço normal	138

Tabela 6.5	Média das médias e dos desvios padrão para 10 processos de simulação, conforme o número de classes M – espaço original	140
Tabela 6.6	Média, desvio padrão e CV dos conjuntos de médias das realizações, para 10 processos de simulação.....	141
Tabela 6.7	Média das médias e dos desvios padrão para 15 processos de simulação, conforme o número de classes M , no espaço normal e no original. Valores condicionantes entre parênteses	142
Tabela 6.8	Média, desvio padrão e CV dos conjuntos de médias das realizações, para 17 processos de simulação.....	143
Tabela 6.9	Parâmetros de simulação para a localização $j = 1$ da malha ($L = M = 15$).....	144
Tabela 6.10	Procedimento prático de tiragens de classes, segundo as 10 localizações extremas do caminho aleatório de uma realização.....	146
Tabela 6.11	Número de ciclos e número total de realizações de acordo a discretização da distribuição G	146
Tabela 6.12	Média, desvio padrão e CV dos conjuntos de médias das realizações, para 25 ciclos de quatro realizações	148
Tabela 6.13	Média, desvio padrão e CV dos conjuntos de médias das realizações, para 20 ciclos de cinco realizações	150
Tabela 6.14	Média, desvio padrão e CV dos conjuntos de médias das realizações, para 16 ciclos de seis realizações	151
Tabela 6.15	Média, desvio padrão e CV dos conjuntos de médias das realizações, para 14 ciclos de sete realizações.....	152
Tabela 6.16	Média, desvio padrão e CV dos conjuntos de médias das realizações, para 12 ciclos de oito realizações	154
Tabela 6.17	Média, desvio padrão e CV dos conjuntos de médias das realizações, para 11 ciclos de nove realizações.....	155
Tabela 6.18	Média, desvio padrão e CV dos conjuntos de médias das realizações, para 10 ciclos de 10 realizações	156
Tabela 6.19	Média, desvio padrão e CV dos conjuntos de médias das realizações, para nove ciclos de 11 realizações	158

Tabela 6.20	Média, desvio padrão e CV dos conjuntos de médias das realizações, para oito ciclos de 12 realizações	159
Tabela 6.21	Média, desvio padrão e CV dos conjuntos de médias das realizações, para sete ciclos de 13 realizações	160
Tabela 6.22	Média, desvio padrão e CV dos conjuntos de médias das realizações, para sete ciclos de 14 realizações	162
Tabela 6.23	Média, desvio padrão e CV dos conjuntos de médias das realizações, para seis ciclos de 15 realizações	163
Tabela 6.24	Média, desvio padrão e CV dos conjuntos de médias das realizações, para seis ciclos de 16 realizações	164
Tabela 6.25	Média, desvio padrão e CV dos conjuntos de médias das realizações, para cinco ciclos de 17 realizações.....	166
Tabela 6.26	Média, desvio padrão e CV dos conjuntos de médias das realizações, para cinco ciclos de 18 realizações.....	167
Tabela 6.27	Média, desvio padrão e CV dos conjuntos de médias das realizações, para cinco ciclos de 19 realizações.....	168
Tabela 6.28	Média, desvio padrão e CV dos conjuntos de médias das realizações, para cinco ciclos de 20 realizações.....	169
Tabela 6.29	Classes e probabilidades (plhs), tiradas na localização u_1 , para a obtenção dos valores Gaussianos $x_{p_{lhs}}$, de acordo com três ciclos de 10 realizações ($M = 10$) ...	170
Tabela 6.30	Média das médias e dos desvios padrão para 19 processos de simulação, conforme o número de classes M , no espaço normal e no original. Valores condicionantes entre parênteses	172
Tabela 6.31	Média, desvio padrão e CV dos conjuntos de médias das realizações, para 19 processos de simulação.....	173
Tabela 6.32	Parâmetros de simulação para a localização $j = 1$ da malha ($L = M = 10$).....	174
Tabela 6.33	Média das médias e dos desvios padrão para oito processos de simulação, conforme o número de classes M , no espaço normal e no original. Valores condicionantes entre parênteses	175
Tabela 6.34	Classes não amostradas da Figura 6.70, em termos de probabilidade, segundo os dois algoritmos (LHS e MC)	178

Tabela 6.35	Classes não amostradas da Figura 6.72, em termos de probabilidade, segundo os dois algoritmos (LHS e MC)	179
Tabela 6.36	Classes não amostradas da Figura 6.75, em termos de probabilidade, segundo os dois algoritmos (LHS e MC)	181
Tabela 6.37	Classes não amostradas da Figura 6.76, em termos de probabilidade, segundo os algoritmos LHS cíclico clássico (20 realizações, $M = 4$) e LHS determinístico ($M = 20$)	182
Tabela 6.38	Parâmetros de simulação para a localização $j = 1$ da malha ($L = M = 20$), por meio da aplicação do algoritmo LHS clássico com caminho determinístico	183
Tabela 6.39	Classes não amostradas, duplamente amostradas e triplamente amostradas da Figura 6.77, em termos de probabilidade, segundo o algoritmo LHS clássico com caminho determinístico	184
Tabela 6.40	Ordem seqüencial de simulação (caminho) de acordo com os algoritmos original, LHS clássico e LHS determinístico	185
Tabela 6.41	Média e CV, da distribuição das médias das realizações, segundo os processos de simulação (LHS clássico) e de acordo os modelos de discretização.....	192
Tabela 6.42	Domínio de variação de tonelagem e diferença de amplitude dos domínios, para os três processos de simulação, segundo os algoritmos MC e LHS clássico, de acordo com o <i>cutoff</i> de 48 %	198

Resumo

O algoritmo de simulação seqüencial estocástica mais amplamente utilizado é o de simulação seqüencial Gaussiana (ssG). Teoricamente, os métodos estocásticos reproduzem tão bem o espaço de incerteza da VA $Z(u)$ quanto maior for o número L de realizações executadas. Entretanto, às vezes, L precisa ser tão alto que o uso dessa técnica pode se tornar proibitivo. Essa Tese apresenta uma estratégia mais eficiente a ser adotada. O algoritmo de simulação seqüencial Gaussiana foi alterado para se obter um aumento em sua eficiência. A substituição do método de Monte Carlo pela técnica de *Latin Hypercube Sampling* (LHS), fez com que a caracterização do espaço de incerteza da VA $Z(u)$, para uma dada precisão, fosse alcançado mais rapidamente. A técnica proposta também garante que todo o modelo de incerteza teórico seja amostrado, sobretudo em seus trechos extremos.

Abstract

Sequential simulation is probably the most used algorithm in geostatistical simulation, specially the sequential Gaussian algorithm. In theory, this method maps the space of uncertainty as the number realizations increase. However, some times the number of simulations needs to be large which makes the procedure prohibitive. This Thesis presents a more efficient strategy. The idea is to replace the Monte Carlo simulation by the *Latin Hypercube Sampling* (LHS) technique in order to improve the efficiency of the algorithm. The use of the modified algorithm showed that the space of uncertainty related to the random variable modeled was faster obtained than the traditional Monte-Carlo simulation for a given degree of precision. This approach also ensures that the model of uncertainty is better represented in its entirety.

Capítulo 1

Introdução

O sucesso de qualquer empreendimento depende do uso racional dos recursos financeiros. O cenário econômico atual exige que as empresas apliquem seus capitais visando obter os melhores resultados pretendidos a um menor custo possível, eliminando custos supérfluos, sem comprometer a qualidade dos produtos. As empresas de mineração não são contrárias a isso e também sentem a necessidade de seguir essa tendência global.

A indústria mineral do Brasil passa por um momento de euforia, visualizando um futuro próximo muito promissor (Minérios & Minerales, ed. 273, p. 6-21 e Minérios & Minerales, ed. 276, p. 8-13). Assim, as empresas de mineração se mostram interessadas em expandir suas frentes de lavra. Mas para isso, os futuros investidores exigem que os depósitos (recursos minerais) potencialmente mineráveis sejam reavaliados, considerando a incerteza em torno das fontes de risco inerentes desse tipo de investimento. A identificação e o gerenciamento dessas fontes de risco são fundamentais para a viabilidade de um projeto de mina.

A avaliação em relação à viabilidade econômica de um projeto passa pela análise do risco financeiro. Em mineração, o risco financeiro se caracteriza pelo fluxo de caixa do empreendimento, o qual engloba várias fontes de risco. O fluxo de caixa considera, entre outros fatores, o mercado econômico do bem mineral, custos operacionais, investimentos, empréstimos e custos ambientais (relacionados às medidas mitigadoras, que deverão ser implantadas em face aos danos ambientais causados pelas operações mineiras). As estimativas de recursos minerais, o planejamento de mina e de produção, assim como o projeto da usina de tratamento, lidam com parâmetros que contém incertezas geológicas. Esses parâmetros se constituem como fontes de risco de caráter técnico, das quais o engenheiro de minas é responsável. A incerteza em torno dessas fontes de risco, envolvidas nas aplicações de engenharia que definem o projeto de mina, deve ser considerada.

Dentre as etapas de caráter técnico que constituem o projeto de mina, a estimativa de recurso e reservas, juntamente com a classificação estéril/minério são cruciais. Essas etapas representam duas das mais importantes para viabilizar o sucesso de uma operação mineira. Por exemplo, a classificação de reservas, a partir das estimativas, pode se transformar em um fator limitante em relação à viabilidade econômica do projeto. Isso acontece em situações em que as reservas classificadas como *medidas* não são suficientes para deflagrar o desenvolvimento do projeto.

As incertezas geológicas, inerentes aos parâmetros utilizados para a realização dessas etapas, devem ser identificadas e avaliadas para auxiliar no processo de tomada de decisão. Vários autores (Ravenscroft, 1992; Dowd, 1994; Rossi e Parker, 1994; Vallee, 1999; Dimitrakopoulos *et al.*, 2002) afirmam que as incertezas geológicas são as principais responsáveis por empreendimentos mineiros apresentarem desempenho aquém das expectativas (e estimativas). Diferenças de quantidade (tonelagem) e concentração (teor) são os maiores contribuintes para a configuração desse comportamento. Lassonde (1990) mostra que dentre 40 projetos fracassados, nos EUA e Canadá, a causa de 73% dos fracassos estava associada a incertezas geológicas, devido aos erros cometidos no processo de avaliação de reservas – superestimando reservas ou teores. Em mineração, a importância da quantificação das incertezas geológicas é reconhecida desde a década de 70 (David *et al.*, 1974; Journel, 1974; Matheron, 1976).

A identificação e o gerenciamento da incerteza, em relação às variáveis geoposicionadas, não é de interesse exclusivo das geociências. Qualquer atividade industrial que utilize informações geoposicionadas passa, atualmente, a conviver com essas exigências. Por exemplo, em engenharia ambiental, a classificação errônea de áreas contaminadas pode acarretar em graves prejuízos financeiros, e ambientais para com a sociedade (Chien, 1998; Garcia e Froidevaux, 1997; van Meirvenne e Goovaerts, 2001; Goovaerts e van Meirvenne, 2001). Em agronomia, o conhecimento sobre a variabilidade da qualidade do solo (entre outros fatores, como clima, etc.) pode auxiliar na obtenção de boas safras (Pachepsky e Acock, 1998).

Por isso, devido às características atuais de mercado econômico, análises de sensibilidade têm sido cada vez mais solicitadas, exigindo aplicação de metodologias dinâmicas, que forneçam respostas satisfatórias em relação à variabilidade das mais variadas fontes de risco. Nesse contexto, a busca por métodos eficientes, que atinjam os objetivos propostos com agilidade, é crescente.

1.1. Abordagem probabilística e o estimador ótimo

O grau de complexidade, envolvido nos fenômenos geológicos que originam os depósitos minerais, impõe uma grande dificuldade na correta estimativa dos atributos de interesse nas localizações não amostradas. Ainda, freqüentemente, a amostragem realizada não se mostra capaz de descrever adequadamente esses processos geológicos. Não obstante à complexidade inerente à gênese desses tipos de fenômenos geológicos, o sistema de amostragem, incluindo a coleta, análise e interpretação, está sujeito a vários tipos de incerteza. Todo esse contexto impossibilita a utilização de um modelo determinístico para caracterizar esses depósitos. Ao contrário dos modelos determinísticos, os modelos probabilísticos permitem que essas incertezas fundamentais sejam consideradas. Por isso, a geoestatística utiliza modelos probabilísticos para descrever a distribuição espacial de qualquer atributo geológico, visando obter estimativas com o menor erro possível.

O processo de interpolação conhecido como krigagem se constitui, atualmente, no principal método de interpolação espacial. Esse estimador linear é teoricamente construído para oferecer estimativas ótimas e não tendenciosas, a partir de um modelo de covariância. Estimativas ótimas e não tendenciosas significam que as mesmas são obtidas de maneira a apresentar mínimo erro médio e mínima variância do erro. O método de krigagem foi concebido por Matheron (1963) a partir das idéias elaboradas por Krige (1951), na busca de um estimador para resolver suficientemente bem o problema de controle e estimativa de teores nas minas de ouro da África do Sul. Entretanto, antes de Matheron, Kolgomorov (1941), Ito (1954), Matern (1960) e Gandin (1963) apresentaram métodos similares, em outras áreas de aplicação. Na verdade, esses métodos pertencem à área dos estimadores ótimos, que têm sua origem no século XIX (1810), quando o matemático francês Pierre Simon de Laplace desenvolveu seu trabalho sobre o método dos mínimos quadrados¹.

O método de krigagem, incluindo suas versões simples (krigagem simples, KS) e ordinária (krigagem ordinária, KO), tem sido utilizado com sucesso em várias áreas de aplicação. Contudo, no decorrer das décadas de 70 e 80, foram desenvolvidas algumas variações em torno desse método. Principalmente, essas variações foram desenvolvidas visando dois objetivos distintos:

- i. estimativas pontuais ou do valor médio de blocos (krigagem pontual ou de blocos);
- ii. estimativas de distribuições locais de probabilidade.

Dentre os métodos que visam realizar estimativas pontuais ou do valor médio de blocos (i), destacam-se:

- i.i. krigagem universal (Huijbregts e Matheron, 1970);
- i.ii. cokrigagem (Marechal, 1970) e
- i.iii. krigagem lognormal (Journel e Huijbregts, 1978, p. 570-572).

Nesse período, os principais métodos desenvolvidos para realizar estimativas de distribuições locais de probabilidade (ii) são:

- ii.i. krigagem disjuntiva (Matheron, 1976);
- ii.ii. krigagem dos incadotes (Journel, 1983);
- ii.iii. krigagem multiGaussiana (Verly, 1983, 1984);
- ii.iv. krigagem da probabilidade (Sullivan, 1984) e

¹ Na realidade, o método dos mínimos quadrados foi desenvolvido pelo matemático alemão Karl Friedrich Gauss (1777 - 1855), em 1801. Entretanto, somente em 1809 Gauss publicou um artigo, demonstrando que a melhor maneira de determinar um parâmetro desconhecido de uma equação de condições é minimizando a soma dos quadrados dos resíduos. Mais tarde, o método foi chamado de Mínimos Quadrados pelo matemático francês Adrien-Marie Legendre (1752 - 1833). Em abril de 1810, Pierre Simon de Laplace (1749 - 1827) apresenta no *memoir* da Academia de Paris, a generalização para problemas com vários parâmetros desconhecidos.

ii.v. krigagem biGaussiana (Marcotte e David, 1985).

Detalhes do método de krigagem e de todas suas variações podem ser encontrados em Goovaerts (1997) e Chilès e Delfiner (1999).

1.2. Simulação Geoestatística

Até a metade da década de 80, a variância de krigagem σ_k^2 era considerada uma boa medida de qualidade das estimativas. Contudo, Journel (1986a) e mais tarde Brus e Gruijter (1993), começam a questionar o uso desse parâmetro como índice de qualidade das estimativas. Journel (1986a) demonstrou que a σ_k^2 incorpora somente as características geométricas dos padrões de amostragem, considerando exclusivamente o posicionamento espacial das medidas e o modelo de continuidade espacial associado. Ou seja, Journel (1986a) demonstrou que a σ_k^2 é independente dos dados e da própria estimativa. Esse comportamento é denominado como homoscedasticidade. Portanto, como mostra Goovaerts (1997, p. 189), a σ_k^2 não sofre qualquer interferência dos valores das medidas; isto é, a σ_k^2 não relaciona as variabilidades locais. Dessa maneira, podem ocorrer casos em que a σ_k^2 seja idêntica para cenários (teores na vizinhança local) completamente diferentes, desde que o modelo de covariância seja o mesmo (Figura 1.1). Assim, a σ_k^2 se mostra como um parâmetro inadequado para a medida de qualidade das estimativas.

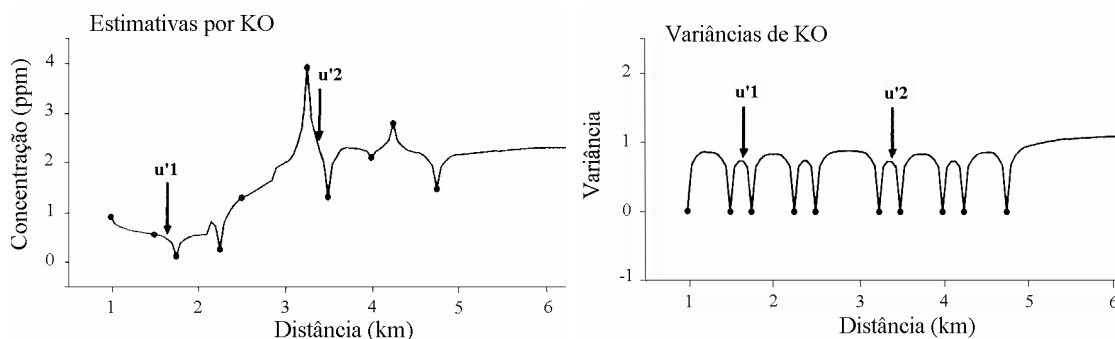


Figura 1.1: Estimativas de teor por krigagem ordinária (KO) (gráfico da esquerda) e suas variâncias (gráfico da direita), diminuindo a medida que se aproxima dos dados. Fonte: Goovaerts, 1997 – p. 181.

Por ser uma técnica de regressão linear, a krigagem tende a superestimar zonas (blocos) de baixa magnitude e subestimar zonas (blocos) de alta magnitude. Em função disso, a distribuição dos resultados de krigagem (estimativas) gera um modelo que subestima a variância real da variável em estudo. Isso porque, o objetivo do método de krigagem é produzir uma estimativa local ótima, minimizando o erro médio e a variância do erro da estimativa. Por isso, independente da área de aplicação, os resultados dados por métodos de krigagem não permitem realizar análises de sensibilidade em torno das estimativas.

Ao contrário dos métodos de krigagem, as técnicas de simulação condicional estocástica permitem que a incerteza seja avaliada. Os métodos de simulação estocástica foram inicialmente desenvolvidos para corrigir o efeito de suavização e outros artefatos mostrados nos mapas produzidos via krigagem (Deutsch e Journel, 1998). De maneira oposta à interpolação, a aplicação dos métodos de simulação estocástica não resulta em uma única estimativa do mapa da variável de interesse. Essas técnicas possibilitam a criação de vários cenários equiprováveis (realizações) da distribuição espacial da variável em estudo, reproduzindo algumas das características da distribuição experimental (amostras). Particularmente, as realizações reproduzem (em média) características importantes da população amostrada, como por exemplo, o histograma e a continuidade espacial. Assim, a partir da análise do conjunto de realizações, é possível medir a incerteza local ou global associada, de forma realista, considerando a heteroscedasticidade inerente aos dados.

A análise das diferentes realizações permite quantificar a incerteza associada à estimativa. Ainda, o conjunto de valores simulados, em uma dada localização, representa uma amostragem da distribuição local de probabilidade que descreve a incerteza em torno do valor real associado a essa localização. A variabilidade apresentada por esses valores é avaliada por meio do cálculo de índices de incerteza. Esses índices são calculados após o processo de simulação estocástica, por meio da compilação e avaliação dos vários valores equiprováveis. Alguns dos índices de incerteza utilizados são:

- i. variância condicional;
- ii. coeficiente de variação condicional;
- iii. intervalo entre quartis (quantis);
- iv. entropia.

A disponibilidade de múltiplas imagens (modelos) da distribuição espacial dos atributos geológicos envolvidos, faz com que seja possível realizar análises de sensibilidade em torno de processos de engenharia, como por exemplo, planejamento de mina (projeto de cava) e de produção. Esse processo de análise é conhecido como análise de risco (David, 1988; Journel, 1992; Rossi e van Brunt, 1997; Dimitrakopoulos, 1998; Dimitrakopoulos *et al.*, 2002). Os vários modelos gerados alimentam funções (funções de transferência) que representam os processos de engenharia. Dessa forma, as incertezas geológicas originais, associadas aos atributos geológicos (teores, por exemplo), são propagadas para refletir a incerteza sobre as aplicações de engenharia.

1.2.1. Simulação geoestatística: histórico

As pesquisas mais recentes na área de geoestatística (realizadas desde o final dos anos 80 até os dias atuais) envolvem o desenvolvimento e aplicação de métodos de simulação estocástica. Embora o conceito de simulação geoestatística tenha sido desenvolvido no início da década de 70 por Matheron (1972, 1973) e Journel (1974), as técnicas de simulação não foram largamente exploradas devido às altas exigências computacionais envolvidas na sua aplicação. No final da década de 80, com

os avanços na tecnologia dos microprocessadores, o interesse por esses métodos foi novamente despertado. Ao longo dos últimos 32 anos foram propostos vários métodos de simulação geoestatística, dirigidos tanto para variáveis contínuas como categóricas. Em ordem cronológica, destacam-se:

- i. método das bandas rotativas de simulação não-condicional, para variáveis contínuas (Matheron, 1972, 1973);
- ii. método das bandas rotativas de simulação condicional, dirigido à variáveis contínuas (Journel, 1974);
- iii. simulação fractal, para variáveis contínuas (Hewett, 1986);
- iv. método LU, desenvolvido para variáveis contínuas ou categóricas (Davis, 1987);
- v. simulação Gaussiana truncada, para variáveis categóricas (Matheron *et al.*, 1987);
- vi. simulação seqüencial dos indicadores, aplicado para variáveis contínuas ou categóricas (Alabert, 1987b);
- vii. simulação seqüencial Gaussiana, aplicado à variáveis contínuas (Isaaks, 1990);
- viii. simulação não-estacionária, para variáveis contínuas (Dimitrakopoulos, 1990);
- ix. simulação seqüencial *annealing*, designado à variáveis contínuas ou categóricas (Deutsch, 1992);
- x. simulação Booleana (Damsleth *et al.*, 1992);
- xi. métodos dos campos de probabilidade (Srivastava, 1992);
- xii. simulação truncada pluri-Gaussiana, destinado à variáveis categóricas (Galli *et al.*, 1994; Armstrong *et al.*, 2001);
- xiii. algoritmos de simulação iterativa (Lantuéjoul, 1997);
- xiv. simulação seqüencial direta, para variáveis contínuas ou categóricas (Caers, 2000b, 2000b);
- xv. simulação direta de blocos, desenvolvido para variáveis contínuas (Godoy, 2002);
- xvi. simulação condicional multi-pontos, simulação de objetos (Strebelle, 2001; Strebelle e Zhang, 2004; Srivastava *et al.*, 2004);
- xvii. simulação seqüencial Gaussiana generalizada, desenvolvido para variáveis contínuas (Dimitrakopoulos e Luo, 2004).

Dowd (1992), Gotway e Rutherford (1994), Srivastava (1994), Chilès (1995), Goovaerts (1997), Costa (1997) e Dimitrakopoulos (1998) discutem os métodos mais utilizados. Luster (1985) e Lantuéjoul (2002) apresentam vários algoritmos de simulação não-condicional e condicional.

1.2.2. Simulação geoestatística: últimas pesquisas

Os últimos avanços em simulação geoestatística convergem para a modelagem de objetos, incentivada pelas aplicações em modelamento de reservatórios na indústria do petróleo. Essas pesquisas têm o objetivo de desenvolver metodologias capazes de modelar estruturas geológicas

complexas. Principalmente, são métodos que visam reproduzir as estatísticas de multi-pontos, a partir de imagens modelo de referência (Strebelle, 2001; Strebelle e Zhang, 2004; Srivastava *et al.*, 2004; Journel, 2004; Arpat e Caers, 2004; Zhang *et al.*, 2004). Outra linha de pesquisa que tem se destacado é o que visa desenvolver algoritmos de simulação direta.

Dimitrakopoulos e Luo (2004) desenvolveram uma variação do método de simulação seqüencial Gaussiana, chamada simulação seqüencial Gaussiana generalizada. Esse método visa à avaliação de grandes áreas e é sugerido para malhas que contenham mais do 10⁸ nós (pontos ou blocos). O método alia a idéia do método LU com o método de simulação seqüencial Gaussiana, sendo eficiente para alguns modelos de covariância espacial. Os autores sugerem que esse método pode ser até 50 vezes mais rápido (computacionalmente) do que o método original, com pouca perda de exatidão.

Embora a idéia de simulação direta tenha sido desenvolvida nos meados dos anos 90 (Journel e Xu, 1994), sua implementação, na forma de um algoritmo seqüencial, foi apresentada pela primeira vez por Caers (2000a, 2000b). Recentemente, vários algoritmos sobre simulação seqüencial direta, com algumas variações, têm sido propostos e aplicados (Soares, 2001; Tran *et al.* 2001; Godoy, 2002; Oz *et al.*, 2003; Manchuk *et al.*, 2004; Gross e Caers, 2004). Godoy (2002) apresenta um método de simulação condicional direta de blocos, desenvolvido para uma aplicação mineira de controle de teores. Leuangthong (2004) mostra as vantagens e alerta para as carências teóricas ainda existentes nesses algoritmos.

A aplicabilidade dos métodos de simulação seqüencial direta ainda sofre, por parte da comunidade científica, uma certa resistência e descrença quanto a sua funcionalidade e eficiência. Isso se deve, principalmente, aos problemas de reprodução dos histogramas condicionantes e a incapacidade de considerar a heteroscedasticidade espacial inerente aos dados (efeito proporcional). Por isso, a simulação seqüencial direta é considerada um dos principais ramos de pesquisa da geoestatística para os próximos anos.

1.2.3. Simulação geoestatística na mineração

Na indústria mineira, a aplicação de técnicas de simulação geoestatística é crescente. Essas técnicas têm sido usadas tanto nas fases de pré-viabilidade/viabilidade econômica, quanto nas fases de planejamento de mina (longo e médio prazo) e de produção (curto prazo). Principalmente, essas técnicas têm sido aplicadas em estudos de sensibilidade, que relacionam os parâmetros envolvidos (teores, principalmente) e seus efeitos sobre o valor presente líquido (VPL).

Os métodos de simulação estocástica permitem quantificar a variabilidade em relação à atributos geológicos como concentração (teor) ou a quaisquer outros atributos relevantes de um dado projeto. Esses métodos possibilitam auxiliar nas decisões cotidianas em relação à lavra e mesmo na adequação ou regulação da planta de beneficiamento. O controle de variabilidade de teores pode ser muito útil em minerações seletivas e em estratégias de blendagem, por exemplo.

As primeiras aplicações de simulação geoestatística em mineração remontam os anos 70 e início dos 80 (David, 1973; David *et al.*, 1974; Clark e White, 1976; Journel e Huijbregts, 1978; Dagbert, 1981; Chilès, 1984; Deraysme *et al.*, 1984). Aplicações recentes compreendem os problemas relacionados a controle de teores, análise de risco, planejamento de mina e projeto de amostragem. Por exemplo, Costa (1997), Costa *et al.* (1997); Dimitrakopoulos (1997, 1998); Bonato (2000c); Viscarra Rossel *et al.* (2001); Godoy *et al.* (2001), Dimitrakopoulos *et al.*, (2002), Godoy, (2002) e Richmond (2003, 2004), aplicaram métodos de simulação geoestatística para a resolução de problemas mineiros relacionados ao controle de teores, planejamento de mina e análise de risco. Chien (1998), Chien e Journel (1998), van Groenigen *et al.* (1999), van Groenigen (1999, 2000), Pilger (2000), Pilger *et al.* (2000, 2001, 2002) e Gorla *et al.* (2001), aplicaram técnicas de simulação geoestatística para projetar campanhas de amostragem adicional e avaliar o impacto (benefício) provocado.

1.3. Problema

Os métodos de simulação geoestatística são construídos para gerar L ($L = 1, \dots, L$) modelos estocásticos de qualquer atributo de um depósito mineral, reproduzindo seus momentos estatísticos de 1ª e 2ª ordem. Geralmente, os métodos estocásticos utilizam a técnica de simulação de Monte Carlo (Halton, 1970 e Isaaks, 1990) para a construção dos modelos numéricos. Nessa técnica, os modelos de incerteza dos atributos de interesse, nas localizações não medidas, são amostradas aleatoriamente L vezes, construindo o conjunto de valores simulados. Uma amostragem desse tipo em cada localização da malha resulta em uma realização (um modelo estocástico) da distribuição espacial da variável de interesse. A repetição desse processo garante o acesso a vários modelos numéricos ou realizações do atributo mineral estudado.

Teoricamente, os métodos estocásticos reproduzem tão bem o espaço de incerteza da variável de interesse quanto maior for o número L de realizações executadas. Porém, na prática, esse número é função da distribuição estatística dos dados originais, necessitando ser maior ou menor, conforme as características intrínsecas dos dados. O importante é que o número L de tiragens aleatórias seja grande o suficiente para garantir que o espaço de incerteza do atributo medido seja caracterizado, atendendo ao objetivo almejado. Entretanto, freqüentemente, L precisa ser tão grande que essa técnica pode se tornar computacionalmente muito intensa. Dependendo da situação, esse procedimento pode exigir muito tempo de processo e armazenamento em disco (HD), tornando-se proibitivo. Nesses casos, ainda existem problemas em relação à manipulação dos arquivos resultantes, pois o procedimento de simulação é apenas uma etapa do processo de análise de risco.

Embora o uso das técnicas de simulação geoestatística seja crescente, em determinadas situações industriais, a aplicação correta dessas técnicas pode ser dificultada. O caráter dimensional

do problema (2D ou 3D), o tamanho da área a ser avaliada (número de pontos ou blocos da malha), o número de dados condicionantes, ou as características intrínsecas dos dados (dependência espacial ou variabilidade) podem se tornar barreiras para o uso adequado das técnicas de simulação geoestatística. Os estudos de caso de Godoy (2002) e Santos (2002) confirmam essa característica.

Geralmente, quando se trabalha com informações resultantes de investigações geofísicas, o número excessivo de dados, aliado ao caráter dimensional (3D) do problema, pode dificultar a aplicação correta dessas técnicas. Nesses casos, a execução de apenas uma imagem equiprovável do atributo geofísico estudado pode demandar horas de cálculo computacional – dependendo das características dos dados e do processador utilizado. Koppe *et al.* (2004) simularam, em 3D, o comportamento da velocidade de onda sônica que viaja através de estratos sedimentares. Nesse estudo, devido à quantidade de dados condicionantes e ao tamanho da malha, foram necessárias muitas horas de CPU (Pentium 4, 2.8 MHz, 1024 Mb) para gerar uma única realização. Em função disso, a análise de risco de projetos, como o do exemplo acima, pode ser prejudicada.

Na indústria, o caráter dinâmico da realização de projetos e o tempo para a tomada de decisões podem não permitir a execução de um número adequado de realizações. Dessa forma, a variabilidade do atributo mineral estudado pode não ser avaliada corretamente.

1.4. Meta

Embora os avanços na tecnologia dos microprocessadores estejam acessíveis à maioria das empresas, simular um depósito mineral não é uma tarefa de rápida e simples execução. Frequentemente, os depósitos minerais são representados por vários milhões de nós (ou blocos), o que significa que são necessários vários milhões de procedimentos de cálculo por realização. Nesses casos, o tempo de processo para a execução de um número adequado de realizações, capaz de caracterizar suficientemente bem o espaço de incerteza da variável em estudo, pode ser significativo – visto que o tempo de processo é diretamente proporcional ao número de realizações². Nesse contexto, a velocidade de execução e o espaço de armazenamento em disco representam problemas para a geração de um número adequado de realizações. Conseqüentemente, o gerenciamento dos arquivos resultantes também se torna um problema, pois os mesmos são necessários na validação das realizações e para as etapas subseqüentes do estudo.

Esse trabalho visa o desenvolvimento de um algoritmo de simulação geoestatística capaz de reduzir o tempo computacional e o espaço virtual requerido nas suas aplicações. Dessa forma, a meta dessa Tese é a de propiciar, à comunidade industrial, um algoritmo de simulação geoestatística eficiente, que forneça resultados satisfatórios com agilidade.

² Além do número de realizações, o tempo computacional de um procedimento de simulação estocástica também é função do número de dados condicionantes e da dimensão da malha.

1.5. Objetivos

Dentre os métodos de simulação geoestatística, citados na seção 1.2.1, destacam-se os seqüenciais, paramétricos ou não-paramétricos. A grande diferença entre esses dois grupos está no procedimento de estimativa dos modelos de incerteza: paramétrico x não-paramétrico. Como o modelo paramétrico mais amplamente utilizado é o multivariado Gaussiano, o método de simulação seqüencial paramétrico mais difundido utiliza o formalismo multiGaussiano – simulação seqüencial Gaussiana (Isaaks, 1990). O método de simulação seqüencial dos indicadores (Alabert, 1987b), não-paramétrico, usa o formalismo homônimo.

Os métodos seqüenciais, de simulação de geoestatística, utilizam a técnica de simulação de Monte Carlo para gerar vários modelos estocásticos da distribuição da variável em estudo, que honram suas características estatísticas uni e bivariadas. Como abordado acima, nos métodos estocásticos, o número de realizações deve ser tal que garanta que o espaço de incerteza da variável de interesse seja suficientemente caracterizado. Contudo, dependendo das características do banco de dados, o número de tiragens aleatórias precisa ser muito elevado. Isso faz com que o uso adequado das técnicas de simulação estocástica seja dificultado ou até, algumas vezes, impraticável.

O objetivo dessa Tese é desenvolver uma estratégia de tiragens aleatórias mais eficiente, a ser empregada nos métodos seqüenciais de simulação estocástica. A substituição do método de tiragem aleatória de Monte Carlo pelo de *Latin Hypercube Sampling*, LHS, (McKay *et al.*, 1979) pode aumentar a eficiência do processo. Essa substituição pretende fazer com que a caracterização do espaço de incerteza da variável de interesse, para uma dada precisão, seja alcançada mais rapidamente. Com isso, o uso adequado das técnicas de simulação geoestatística na indústria pode ser facilitado. Isso porque, o número de realizações necessárias para se caracterizar adequadamente a variabilidade dos atributos minerais de interesse pode ser reduzido – fazendo com que o tempo de processo seja reduzido. Conseqüentemente, a redução do número de realizações também proporcionará mais agilidade na manipulação dos arquivos produzidos, uma vez que estes ocuparão menos espaço em disco. Dimitrakopoulos e Luo (2004) desenvolveram uma pesquisa visando o mesmo objetivo, porém com idéias diferentes.

Para alcançar o objetivo proposto, cinco etapas precisam ser desenvolvidas.

- i. Desenvolver o algoritmo de tiragens aleatórias, partindo-se da idéia e do embasamento teórico da técnica LHS.
- ii. Verificar a funcionalidade teórica do algoritmo desenvolvido na etapa (i).
- iii. Inserir o algoritmo desenvolvido na etapa (i) nos algoritmos de simulação seqüencial estocástica.
- iv. Verificar a eficiência e funcionalidade prática dos algoritmos modificados.
- v. Quantificar o impacto provocado pela alteração dos algoritmos modificados em problemas reais de mineração.

As etapas (iv) e (v) serão realizadas a partir de dois estudos de caso que verificarão a eficiência e aplicabilidade do algoritmo proposto.

1.6. Organização da Tese

Essa Tese será desenvolvida em seis capítulos. De forma resumida, os capítulos seguintes estão organizados da seguinte maneira:

- O Capítulo 2 revisa os conceitos básicos necessários à abordagem geoestatística. São definidas as hipóteses intrínsecas e estacionárias, assim como suas relações com as equações de krigagem. Também são apresentadas as propriedades elementares da função de covariância espacial e algumas propriedades estatísticas de combinações lineares.
- O Capítulo 3 apresenta o conceito de simulação condicional seqüencial, apresentando seus principais métodos e as técnicas envolvidas. São revisados, de forma teórica e prática, os procedimentos paramétricos e não-paramétricos para a estimativa de modelos locais de incerteza, necessários à implementação do algoritmo de simulação seqüencial. Ambas as técnicas são apresentadas em detalhe, destacando suas aplicabilidades e possíveis simplificações.
- O Capítulo 4 estuda as ferramentas matemáticas que permitem extrair o máximo de informações dos modelos de incerteza. O capítulo aborda o conceito de função de transferência e mostra como a incerteza é propagada por meio dessas funções, apresentando como construir de maneira adequada os modelos de incerteza resultantes. São apresentadas ferramentas matemáticas e técnicas que permitem avaliar a qualidade dos modelos de incerteza.
- O Capítulo 5 se constitui no desenvolvimento da metodologia proposta. Nesse capítulo, são apresentados os conceitos teóricos do método LHS, assim como o desenvolvimento do algoritmo a ser inserido nas rotinas de simulação seqüencial estocástica. Testes de verificação teórica também são conduzidos nesse capítulo.
- O Capítulo 6 analisa a eficiência e a aplicabilidade do algoritmo proposto por meio de dois estudos de caso.
- O Capítulo 7 inclui conclusões e recomendações para trabalhos futuros.

Capítulo 2

Conceitos introdutórios de geoestatística

Em geoestatística linear, as variáveis sempre são tratadas como estacionárias ou intrínsecas, conforme seus comportamentos estatísticos espaciais. Este capítulo tem a função de apresentar as hipóteses intrínsecas e estacionárias, assim como suas relações com as equações de krigagem. O capítulo inicia com uma breve revisão dos conceitos de variáveis regionalizadas e funções aleatórias. A seguir, as hipóteses citadas acima são definidas, seguida pela apresentação das propriedades elementares da função de covariância espacial. Como os estimadores de (co)krigagem são constituídos de combinações lineares ponderadas, se torna importante conhecer algumas propriedades estatísticas de combinações lineares. Nesse sentido, é estudado o comportamento da variância de combinações lineares, definindo o conceito de combinações lineares autorizadas (CLA). A relação das hipóteses intrínsecas e estacionárias com as equações de (co)krigagem encerra o capítulo.

2.1. Modelos probabilísticos

A maioria dos fenômenos geológicos não é suficientemente conhecido para permitir a utilização de modelos determinísticos na estimativa de um atributo qualquer em locais não amostrados (interpolação). Em geociências, sempre existe incerteza sobre como o fenômeno se comporta entre os pontos amostrais. Por essas razões, os fenômenos devem ser descritos por meio de modelos probabilísticos, os quais consideram essas incertezas fundamentais. Esses modelos proporcionam meios para realizar estimativas em locais não amostrados, utilizando funções aleatórias estacionárias e variáveis aleatórias.

Os modelos adotados, probabilísticos ou não, não podem ser prejudgados como corretos ou errados. Só é possível provar sua aplicabilidade por meio de informações adicionais que validem a escolha. Entretanto, esses modelos podem ser denominados como apropriados ou não. Esse julgamento deve considerar os objetivos do estudo e todo o tipo de informação qualitativa disponível (Isaaks e Srivastava, 1989, p. 198).

No início da década de 60, Georges Matheron desenvolveu a "Teoria das Variáveis Regionalizadas" (Matheron, 1965; 1970) e criou o termo geoestatística para designar o desenvolvimento de modelos de campos aleatórios (Kolmogorov, 1950; Yaglom, 1962; Whittle, 1963)

para descrever fenômenos naturais espaciais, caracterizados pela distribuição no espaço de um ou mais atributos.

Em geoestatística, a abordagem probabilística pode ser aceita na medida em que os eventos geológicos que originam os depósitos minerais representam o resultado de muitos processos que podem ser tão complexos, em relação ao nosso entendimento, que podemos tratá-los como processos aleatórios. Assim, o uso de modelos probabilísticos permite acesso a métodos de estimativa, que na prática produzem bons resultados, e ainda possibilitam medir a acuracidade e precisão das estimativas.

2.2. Variáveis regionalizadas

O termo variável regionalizada foi escolhido por Matheron (Matheron, 1965; 1970) para salientar os dois aspectos aparentemente contraditórios desse tipo de variável:

- i. um aspecto aleatório, associado às irregularidades locais;
- ii. um aspecto estruturado, que reflete as tendências do fenômeno em grande escala.

Para representar os fenômenos geológicos de forma mais realista, a geoestatística associa o aspecto aleatório às flutuações em torno de uma deriva (média). Essas flutuações representam as características intrínsecas do fenômeno geológico, que possui estruturação própria.

2.3. Função aleatória e variável aleatória

Nos modelos probabilísticos, o valor de qualquer atributo, em qualquer localização, conhecido ou não, pode ser interpretado probabilisticamente como o resultado de uma variável aleatória (VA). O valor observado em cada localização u é considerado uma realização $z(u)$ de uma variável aleatória (VA) $Z(u)$. Sua média (esperança matemática) em u é representada por $m(u)$. Matematicamente, o conjunto de todas essas VAs constitui uma função aleatória (FA). A relação entre uma FA e uma de suas realizações é a mesma que a de uma VA e uma de suas tiragens. Nesses modelos, a distribuição espacial dos atributos é vista como uma realização espacial do modelo de FA que caracteriza o fenômeno geológico (natural) considerado (Matern, 1960; Matheron, 1963; Matheron, 1965). Nesse contexto, as n informações medidas disponíveis $z(u_\alpha)$ ($\alpha = 1, \dots, n$) são consideradas como uma realização de um conjunto de VAs $Z(u_\alpha)$ ¹, espacialmente dependentes e originadas por uma função aleatória FA $Z(u)$.

A aparente aleatoriedade dos valores assumidos por $Z(u)$ é explicada pela variância de cada VA, enquanto que a continuidade espacial entre esses valores é caracterizada pela função de

¹ A partir desse momento, o conjunto de VAs $Z(u_\alpha)$ será referido como $Z(u)$.

correlação espacial $C(h)$ (covariância) do modelo de FA $Z(u)$. Essas propriedades do modelo de FA são controladas pela distribuição de probabilidade (dp) cumulativa multivariada², independente do número N de localizações u_j ($j = 1, \dots, N$) considerado:

$$F(u_1, \dots, u_N; z_1, \dots, z_N) = \text{Pr ob}\{Z(u_1) \leq z_1, \dots, Z(u_N) \leq z_N\} \quad (2.1)$$

onde $F(\cdot)$ representa a dp da FA $Z(u)$. A dp determina a probabilidade multivariada da VA $Z(u)$ ser inferior aos vários teores de corte z_c ($c = 1, \dots, N$). É a dp que governa o modelo de FA $Z(u)$ (Isaaks, 1990, p. 9).

Uma FA é caracterizada por sua lei espacial, isto é, pelo conjunto das leis das realizações da VA $Z(u)$ para qualquer localização u_j ($j = 1, \dots, N$) dentro do domínio espacial em estudo. Ou seja, o conjunto de realizações, descrevendo os valores assumidos pelas VAs e suas probabilidades correspondentes, é chamado de lei de probabilidade espacial do modelo de FA. Se a dp de uma FA for conhecida, podem-se calcular parâmetros estatísticos que descrevem as variáveis consideradas. Parâmetros como média e variância podem não fornecer uma descrição completa da distribuição, mas freqüentemente compõem valiosas informações sobre o comportamento das VAs.

A dp (e conseqüentemente a lei de probabilidade espacial) de uma FA também pode ser dada em termos da função de densidade de probabilidade (fdp) multivariada, isto é:

$$f(u_1, \dots, u_N; z_1, \dots, z_N) = \frac{\partial \text{Pr ob}\{Z(u_1) \leq z_1, \dots, Z(u_N) \leq z_N\}}{\partial z} \quad (2.2)$$

logo,

$$\int_{-\infty}^z f(u_1, \dots, u_N; z_1, \dots, z_N) dz = F(u_1, \dots, u_N; z_1, \dots, z_N)$$

onde $f(\cdot)$ e $F(\cdot)$ representam a fdp multivariada e a dp da FA $Z(u)$, respectivamente. Os valores assumidos pelas VAs nas localizações u_j ($j = 1, \dots, N$) são representados por z_j ($j = 1, \dots, N$).

Evidentemente, seria impossível utilizar um modelo de FA sem fazer algumas hipóteses sobre as características dessas leis. Em particular, na medida em que é disponível somente uma realização, é preciso fazer hipóteses com relação à sua estacionaridade. Por exemplo, o que seria possível dizer sobre uma VA com base em uma única realização, $z(u) = 4$?

2.4. Hipóteses estacionária e intrínseca

Em estatística, é comum supor que uma FA é estacionária, isto é, que sua lei espacial não varia por translação. Dessa forma, subentende-se que uma FA estacionária se repete no espaço. Essa propriedade faz com seja possível fazer inferências estatísticas sobre a FA. O modelo de FA $Z(u)$ é considerado estritamente estacionário se sua fdp multivariada (todos os momentos) for invariável por

² Nessa Tese, dp designa distribuição de probabilidade cumulativa multivariada.

translação (Bartlett, 1966). Em geoestatística, essa translação é caracterizada por um vetor \vec{h} ³. Portanto, se $Z(u)$ e $Z(u+h)$ compartilham a mesma fdp multivariada, independente do vetor h , o modelo de FA $Z(u)$ é estritamente estacionário:

$$f(u_1, \dots, u_N; z_1, \dots, z_N) = f(u_1 + h, \dots, u_N + h; z_1, \dots, z_N), \quad \forall h \quad (2.3)$$

onde $f(\cdot)$ caracteriza a fdp multivariada da FA $Z(u)$ e $u_j + h$ ($j = 1, \dots, N$) representam as localizações transladadas.

Se as VAs não compartilham a mesma fdp multivariada, então o modelo de FA é considerado não-estacionário (Dimitrakopoulos, 1990; Chilès e Delfiner, 1999). Entretanto, na prática, isso não é possível de ser verificado a partir de um número limitado de dados experimentais. Por isso, geralmente, somente é necessário que os dois primeiros momentos (média e covariância) sejam invariáveis por translação. Assim, em geoestatística, o grau de estacionaridade requerido é menos exigente, chamado de estacionaridade de ordem 2 (ST2). Em resumo, para uma VA ser considerada ST2 é necessário que:

- i. a média da FA $Z(u)$ seja constante, independente de u

$$E[Z(u)] = m(u) = m, \quad \forall u \quad (2.4)$$

onde $E[Z(u)]$ e m representam a esperança matemática e a média da VA $Z(u)$.

- ii. a função de covariância entre duas localizações (u e $u+h$) seja independente de u , dependendo somente do vetor h :

$$\text{Cov}(Z(u), Z(u+h)) = C(h) = E[Z(u) \cdot Z(u+h)] - m^2, \quad \forall u \quad (2.5)$$

onde $\text{Cov}(Z(u), Z(u+h))$ representa a covariância da VA $Z(u)$ em relação a duas localizações distintas, caracterizadas por u e $u+h$.

Em particular, para $h = 0$, a covariância é igual à variância de $Z(u)$, que também deve ser constante.

Na prática, existem casos em que essas condições não são satisfeitas. Por exemplo, pode ser evidenciada a presença de uma tendência, tornando-se inadequado supor que a média seja constante. Nesses casos, essa tendência pode ser considerada, no momento das estimativas, usando-se técnicas da geoestatística não-estacionária. Ainda, às vezes, ocorrem casos em que a covariância pode não existir, mesmo que a média seja constante (estacionária). Nesses casos, se torna interessante (e útil) fazer com que a condição relacionada à função de covariância $C(h)$ seja menos rígida.

A hipótese intrínseca (Matheron, 1971) supõe que os incrementos da FA sejam estacionários de ordem 2; isto é, essa hipótese supõe que a média e a variância dos incrementos da FA ($Z(u+h) - Z(u)$) existam e que sejam independentes da localização u . Em particular, a variância dos incrementos

³ De agora em diante, o vetor \vec{h} será representado somente por h .

$(Z(u+h)-Z(u))$ da FA pode ser escrita em termos de uma função dependente do vetor h , denominada como semi-variograma ($\gamma(h)$). Dessa maneira:

$$\begin{aligned} E[Z(u+h) - Z(u)] &= 0 \quad \therefore E[Z(u+h)] = E[Z(u)] \text{ e} \\ \text{Var}[Z(u+h) - Z(u)] &= E\{[Z(u+h) - Z(u)]^2\} = 2\gamma(h) \end{aligned} \quad (2.6)$$

onde $E[Z(u+h) - Z(u)]$ e $\text{Var}[Z(u+h) - Z(u)]$ representam a média e a variância dos incrementos $(Z(u+h)-Z(u))$ da FA $Z(u)$, respectivamente.

As variáveis regionalizadas estacionárias sempre verificam a hipótese intrínseca, mas a recíproca não é verdadeira.

Em função da hipótese intrínseca (condição menos rígida relacionada à função de covariância $C(h)$) a classe de funções que podem representar um semi-variograma $\gamma(h)$ é muito mais ampla quando comparada com aquelas que podem representar as covariâncias. Portanto, o uso de variáveis regionalizadas intrínsecas, e não somente estacionárias, aumenta o campo de modelos possíveis. Por isso, o semi-variograma $\gamma(h)$ é a ferramenta básica para a interpretação estrutural e para as estimativas decorrentes do fenômeno geológico.

Na prática, o semi-variograma $\gamma(h)$ é utilizado até uma certa distância, sendo necessário que o fenômeno geológico seja estacionário num raio igual a essa distância. Nesse momento, o problema passa a ser encontrar uma série de vizinhanças móveis dentro das quais a média e o semi-variograma $\gamma(h)$ possam ser considerados estacionários. Essa hipótese de estacionaridade local é função da escala de homogeneidade do fenômeno e da densidade de amostras disponíveis.

2.5. Covariância espacial

Propriedades elementares da covariância espacial:

- i. $C(0) = \text{Var}[Z(u)] = \text{Var}[Z(u+h)]$;
- ii. $C(h) = C(-h)$;
- iii. $|C(h)| \leq C(0)$ (relação de Schwartz (Armstrong e Carignan, 1997)).

É importante salientar que não existe função equivalente à covariância $C(h)$ para as variáveis que são intrínsecas, mas não estacionárias.

A relação entre o semi-variograma $\gamma(h)$ e a covariância $C(h)$, no caso estacionário, é a seguinte:

$$\begin{aligned} 2\gamma(h) &= \text{Var}[Z(u+h) - Z(u)] = 2C(0) - 2C(h), \text{ resultando em} \\ \gamma(h) &= C(0) - C(h) \end{aligned} \quad (2.7)$$

Essa relação também explica porque a utilização do prefixo "semi" para designar a função $\gamma(h)$. A partir das Equações 2.6 e 2.7 é verificado que:

$$\begin{aligned}\text{Var}[Z(u+h) - Z(u)] &= E[\{Z(u+h) - Z(u)\}^2] = 2[C(0) - C(h)] \\ \therefore \gamma(h) &= 1/2 \cdot E[\{Z(u+h) - Z(u)\}^2]\end{aligned}\quad (2.8)$$

Portanto, o semi-variograma $\gamma(h)$ mede a semi-variância dos incrementos $(Z(u+h)-Z(u))$, que é caracterizada por sua semi-média quadrática.

A relação entre o semi-variograma $\gamma(h)$ e a covariância $C(h)$ só é válida para as variáveis cujo variograma apresenta um limite superior (patamar). As variáveis estacionárias sempre têm variogramas que apresentam essa característica. Em consequência, os variogramas que não apresentam limite superior (patamar) são oriundos de variáveis intrínsecas, mas não estacionárias.

De maneira geral, se $\lim_{h \rightarrow \infty} \frac{\gamma(h)}{h^2} = 0$ a variável regionalizada é intrínseca. Se $\gamma(h)$ aumenta mais rapidamente do que o quadrado de h , variável regionalizada não é estacionária.

2.6. Variância de combinações lineares autorizadas, CLA

Como a maioria dos estimadores utilizados na indústria mineira são constituídos basicamente de combinações lineares dos dados, é preciso estudar suas propriedades. Por exemplo, a variância de uma combinação linear de uma VA $Z(u)$ estacionária:

$$\begin{aligned}Z^*(u) &= \sum_{\alpha} \lambda_{\alpha} z(u_{\alpha}) \\ \text{Var}[Z^*(u)] &= \sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \lambda_{\beta} C(u_{\alpha} - u_{\beta})\end{aligned}\quad (2.9)$$

onde λ_{α} e λ_{β} representam os pesos atribuídos aos dados amostrados $z(u_x)$ ($x = \alpha = \beta = 1, \dots, n$) e a função $C(u_{\alpha} - u_{\beta})$ representa a função covariância $C(h)$.

Na verdade:

$$E[Z^*(u)] = E[\sum_{\alpha} \lambda_{\alpha} z(u_{\alpha})] = \sum_{\alpha} \lambda_{\alpha} E[z(u_{\alpha})] = \sum_{\alpha} \lambda_{\alpha} m \quad (2.10)$$

Por definição:

$$\begin{aligned}\text{Var}[Z^*(u)] &= E[Z^* - E(Z^*(u_{\alpha}))]^2 \\ &= E[\sum_{\alpha} \lambda_{\alpha} z(u_{\alpha}) - m]^2 \\ &= \sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \lambda_{\beta} C(u_{\alpha} - u_{\beta})\end{aligned}\quad (2.11)$$

Essa variância não pode ser negativa, independente das localizações u e dos ponderadores λ . Uma função covariância $C(h)$ que satisfaça essa condição é chamada **definida positiva** ou simplesmente **tipo positiva**. Reciprocamente, qualquer função tipo positiva pode ser considerada como a covariância de uma FA estacionária.

Entretanto, a variância de uma combinação linear envolvendo uma VA intrínseca, mas não estacionária, pode não existir. Em razão da hipótese intrínseca, só é possível afirmar que existe a variância da combinação linear dos incrementos $(Z(u+h)-Z(u))$ dessa VA. Essas combinações, que são chamadas autorizadas, são caracterizadas por:

$$\sum_{\alpha} \lambda_{\alpha} = 0 \quad (2.12)$$

onde λ_{α} representa os pesos da combinação linear. Portanto, qualquer combinação linear cuja soma dos pesos é 0 é chamada de **combinação linear autorizada (CLA)**. Por exemplo, a seguinte combinação linear não é uma CLA:

$$z^*(u_0) = 1 \cdot z(u_1) - 0,5 \cdot z(u_2) - 1 \cdot z(u_3) \quad (2.13)$$

pois $\sum_{\alpha=1}^3 \lambda_{\alpha} = -0,5 \neq 0$. Portanto não há qualquer garantia que sua variância exista.

Evidentemente, essa condição é verificada para toda combinação linear dos incrementos de qualquer VA, desde que cada incremento introduz os pesos +1 e -1. Reciprocamente, qualquer combinação linear que satisfaça essa condição ($\sum_{\alpha} \lambda_{\alpha} = 0$) pode ser escrita como uma combinação linear dos incrementos.

Assim, a variância de qualquer CLA existe e pode ser calculada da seguinte maneira:

$$\text{Var}[\sum \lambda_{\alpha} z(u_{\alpha})] = -\sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \lambda_{\beta} \gamma(u_{\alpha} - u_{\beta}) \quad (2.14)$$

onde $\gamma(u_{\alpha} - u_{\beta})$ representa o semi-variograma $\gamma(h)$ em relação a pares de dados amostrais $(z(u_{\alpha})$ e $z(u_{\beta}))$ separados por um vetor h . Os pesos atribuídos às localizações amostradas u_x ($x = \alpha = \beta = 1, \dots, n$) são representados por λ_{α} e λ_{β} .

Em resumo, se uma VA regionalizada for estacionária (preenchendo os requisitos necessários descritos anteriormente), há equivalência entre o semi-variograma $\gamma(h)$ e a covariância $C(h)$, permitindo que seja possível calcular a variância de qualquer CLA em função de $\gamma(h)$ ou de $C(h)$. Por outro lado, a variância de uma CLA de uma VA intrínseca deve ser calculada exclusivamente em função do semi-variograma $\gamma(h)$.

A variância descrita acima deve ser sempre positiva ($-\sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \lambda_{\beta} \gamma(u_{\alpha} - u_{\beta}) \geq 0$). Para isso, $-\gamma$ deve ser **condicionalmente definido positivo**. Para garantir que essa última condição seja satisfeita, é preciso utilizar modelos da função $\gamma(h)$ que sejam autorizados, ou seja, modelos nos quais utilizem funções **tipo positivas**. Geralmente, são utilizados os seguintes modelos:

- i. Efeito pepita.
- ii. Modelo esférico.
- iii. Modelo exponencial.
- iv. Modelo Gaussiano.
- v. Função potência (VA intrínsecas).

vi. Modelo cúbico.

2.7. Equações de krigagem e suas relações com as hipóteses estacionária e intrínseca

A krigagem é um estimador linear, sem viés global, que fornece a média ponderada dos valores do atributo medido, de modo que a variância do erro da estimativa seja mínima. Ou seja, no método de krigagem os ponderadores são calculados de tal forma que:

- i. $E[Z^* - Z] = 0$
- ii. e que a $\text{Var}[Z^* - Z]$ seja mínima.

2.7.1. Krigagem ordinária, KO

Em KO, como o erro da estimativa (ϵ) é uma CLA, a variância do erro $\text{Var}[Z^* - Z]$ de uma VA estacionária $Z(u)$ pode ser calculada em função do semi-variograma $\gamma(h)$ ou da covariância $C(h)$. Isto é:

$$\epsilon = Z^* - Z = \sum \lambda_\alpha Z(u_\alpha) - Z(u_0) \quad , \quad \text{com } \sum \lambda_\alpha = 1 \quad (2.15)$$

de maneira que

$$E[Z^* - Z] = E[\sum \lambda_\alpha Z(u_\alpha) - Z(u_0)] = m(\sum \lambda_\alpha - 1) = 0 \quad (2.16)$$

Portanto, o somatório dos pesos da combinação linear (ϵ) é nulo (Equação 2.16), assim o erro da estimativa (ϵ) é uma CLA. Desse modo, as equações de KO, para VA estacionárias, podem ser escritas em termos de $\gamma(h)$ ou $C(h)$. Entretanto, se a VA $Z(u)$ for intrínseca, mas não estacionária, a variância do erro deve ser calculada exclusivamente em função do semi-variograma $\gamma(h)$. Isso porque, nesse caso, não há $C(h)$ equivalente. Assim, para variáveis regionalizadas desse tipo, as equações de KO só podem ser escritas em termos do semi-variograma $\gamma(h)$.

Portanto, pela característica de erro da estimativa (CLA), as equações de KO se aplicam às variáveis intrínsecas e estacionárias.

2.7.2. Krigagem simples, KS

Em KS, o erro da estimativa (ϵ) não é uma CLA; ou seja:

$$\epsilon = Z^* - Z = [\sum \lambda_\alpha Z(u_\alpha) + (1 - \sum \lambda_\alpha) \cdot m] - Z(u_0) \quad (2.17)$$

de maneira que

$$\begin{aligned} E[Z^* - Z] &= E[\sum \lambda_\alpha Z(u_\alpha) + (1 - \sum \lambda_\alpha) \cdot m - Z(u_0)] \\ &= E[\sum \lambda_\alpha (Z(u_\alpha) - m) + m - Z(u_0)] \end{aligned} \quad (2.18)$$

Nesse caso, não existe restrição quanto ao somatório dos pesos $\sum_\alpha \lambda_\alpha$ atribuídos aos dados amostrais, e portanto o erro da estimativa (ϵ) não é uma CLA. Contudo, agora não é necessário qualquer

condição em relação aos ponderadores, pois o requisito de ausência de viés é naturalmente preenchido:

$$E[Z^* - Z] = \sum \lambda_\alpha m + (1 - \sum \lambda_\alpha) \cdot m - m = \sum \lambda_\alpha (m - m) + m - m = 0 \quad (2.19)$$

Assim, a variância do erro (ϵ) da estimativa ($\text{Var}[Z^* - Z]$) de uma VA estacionária $Z(u)$ deve ser calculada em função da covariância $C(h)$. Por esse motivo, a KS se aplica exclusivamente às VA estacionárias.

2.8. Comentários

Nesse capítulo foram revisados alguns dos conceitos básicos de geoestatística linear, fundamentais para o entendimento e desenvolvimento matemático dos seus métodos associados. Foram apresentadas as hipóteses intrínseca e estacionária de FA, assim como suas relações com as equações de krigagem. O capítulo iniciou com a apresentação e definição da abordagem probabilística aos problemas enfrentados em geociências. Foram revisados os conceitos de variáveis regionalizadas, função aleatória (FA) e variáveis aleatórias (VAs), básicos para o entendimento dos métodos de simulação estocástica. Principalmente, esse capítulo apresentou a definição de combinações lineares autorizadas (CLA) e suas relações com as hipóteses intrínsecas e estacionárias, decisivas para a aplicação das equações de (co)krigagem.

O Capítulo 3 apresenta uma revisão dos métodos de simulação condicional seqüencial, objeto direto dessa Tese.

Capítulo 3

Simulação condicional seqüencial

Esse capítulo tem o objetivo de apresentar o conceito de simulação condicional seqüencial, apresentando seus principais métodos e as técnicas envolvidas. O capítulo inicia com uma breve revisão do conceito dos métodos de Monte Carlo. Em seguida, é apresentada a idéia básica de simulação e condicionamento. É ressaltada a utilidade e a necessidade dos modelos locais de incerteza quando se deseja medir adequadamente a variabilidade em torno das estimativas de atributos em localizações não medidas. A seguir, são revisados, de forma teórica e prática, os procedimentos paramétricos e não-paramétricos para a estimativa de modelos locais de incerteza, necessários à implementação do algoritmo de simulação seqüencial. Ambas as técnicas são apresentadas em detalhe, destacando suas aplicabilidades e possíveis simplificações.

3.1. Métodos de Monte Carlo: conceitos

Os métodos de Monte Carlo foram originalmente desenvolvidos no Laboratório Nacional Los Alamos (EUA) após o término da II Guerra Mundial. Nessa época, os EUA tinham recém concluído a construção do seu primeiro computador eletrônico (o ENIAC) e os cientistas de Los Alamos estavam estudando como usá-lo para auxiliar na construção da bomba de H (Veach, 1997). No final de 1946, um dos pesquisadores do laboratório (Stanislaw Ulam) sugeriu o uso de amostragem aleatória para simular trajetos de nêutrons (em um estudo sobre bombardeamento de partículas). No final dos anos 40, Metropolis e Ulam (1949) publicaram um resumo das idéias e teorias desenvolvidas pelo Laboratório Nacional Los Alamos, despertando um grande interesse de pesquisa nos anos 50 (Meyer, 1956). O nome do método foi sugerido por Metropolis em homenagem a capital de Mônaco, famosa por seus cassinos (Veach, 1997).

Entretanto, muito antes do surgimento dos conceitos teóricos elaborados pelo Laboratório Nacional Los Alamos, amostragem aleatória já era utilizada para resolver problemas numéricos (Kalos e Whitlock, 1986). As origens dos métodos de Monte Carlo foram discutidas por Ulam (1987). Hammersley e Handscomb (1964), Rubinstein (1981), assim como Kalos e Whitlock (1986) também se constituem em referências significantes dos métodos de Monte Carlo.

Os métodos de Monte Carlo podem ser definidos, de uma maneira simplista, como quaisquer métodos que utilizam números aleatórios para solucionar um problema. Segundo Hammersley e Handscomb (1964), os métodos de Monte Carlo estão inseridos em um ramo da matemática aplicada (experimental) que realiza experimentos sobre números aleatórios. De acordo com Halton (1970), os métodos de Monte Carlo apresentam a solução de um problema como um parâmetro de uma população hipotética, usando uma seqüência de números para construir uma amostra da população a partir da qual podem ser realizadas estimativas dos seus parâmetros estatísticos.

Os métodos de Monte Carlo têm sido usados com respaldo científico em muitas áreas de aplicação como mecânica estatística, física nuclear e economia (Isaaks, 1990, p. 7). Esses métodos são aplicados tanto para a resolução de problemas determinísticos como para probabilísticos. Por exemplo, no campo determinístico, a aplicação de métodos de Monte Carlo pode auxiliar na resolução de uma integral envolvendo uma função determinística. No caso de um problema probabilístico, o método de Monte Carlo funciona a partir da observação de números aleatórios de forma que esses próprios números possam simular diretamente o processo aleatório físico que descreve o problema original (Hammersley e Handscomb, 1964). A partir do comportamento desses números aleatórios é possível inferir a solução almejada, de acordo com o problema. Em geoestatística, esses métodos são utilizados para estimar a incerteza associada a fenômenos espacialmente distribuídos.

Segundo Veach (1997), os métodos de Monte Carlo apresentam três grandes vantagens:

- i. O método converge a uma razão de $N^{-1/2}$. Isso significa que a média, por exemplo, será reproduzida, com desvio padrão de $N^{-1/2}$, quando são realizados conjuntos de N tiragens aleatórias.
- ii. Simplicidade do método. De fácil entendimento e implementação, o método se tornou rapidamente popular.
- iii. O método de Monte Carlo é geral. Fundamentado em amostragem aleatória, o método pode ser aplicado em quaisquer situações, em problemas multidimensionais, sem exigências teóricas.

3.2. Simulação estocástica

Em geoestatística, independente do campo de aplicação, muitas vezes se torna necessário avaliar e quantificar a magnitude de incerteza associada às estimativas do atributo físico-químico de interesse. Assim, a investigação em torno das incertezas locais deve ser implementada em toda a área de estudo para que seja possível avaliar, de uma forma matematicamente embasada, os setores potencialmente problemáticos (erráticos) da área. Dessa maneira, a busca por ferramentas matemáticas que possam ser utilizadas para esse fim se torna fundamental.

Os métodos de simulação estocástica foram inicialmente desenvolvidos para corrigir o efeito de suavização e outros artefatos mostrados nos mapas produzidos via krigagem (Deutsch e Journel, 1998). De maneira oposta à interpolação, a aplicação dos métodos de simulação estocástica não resulta em uma única estimativa do mapa da variável de interesse. A idéia básica de simulação estocástica é a de gerar várias realizações do modelo de FA $Z(u)$ por meio de tiragens aleatórias (simulação de Monte Carlo) de valores a partir da função de densidade de probabilidade (fdp) multivariada da FA $Z(u)$. Entretanto, os valores dos atributos simulados devem representar o processo aleatório físico que caracteriza o fenômeno geológico (natural) original. Nesse momento, uma questão é naturalmente inserida no contexto: quais são os parâmetros estatísticos do modelo de FA $Z(u)$ que melhor representam as características desse processo físico? Ainda: essas características podem ser impostas sobre as realizações, por meio de um modelo? Por exemplo, se for assumido que a distribuição univariada (histograma) e a função de correlação espacial caracterizam adequadamente os processos aleatórios físicos originais, então todas as realizações devem reproduzir esses momentos da distribuição original (Isaaks, 1990, p. 9).

3.3. Simulação condicional

Simulação condicional foi inicialmente proposto por Journel (1974) como um método (aplicado em duas etapas) capaz de quantificar a variabilidade e a incerteza de atributos geológicos espacialmente distribuídos. Mais especificamente, visava obter informações sobre concentrações de minério em uma malha compatível com o planejamento de lavra em curto prazo e com as previsões sobre desempenho da planta de beneficiamento. Esse método foi desenvolvido a partir das idéias elaboradas por Matheron (1972, 1973), sobre simulação por bandas rotativas e foi denominado como: simulação condicional por bandas rotativas. Journel e Huijbretgs (1978, capítulo 7) apresentam as idéias básicas desse método e de simulação condicional em geral, juntamente com alguns estudos de caso de aplicações mineiras.

O método de simulação condicional por bandas rotativas é aplicado em duas etapas, pois o condicionamento é realizado após a geração de simulações não condicionais. Primeiramente, são geradas simulações não condicionais a partir da projeção de simulações unidimensionais (linhas) no espaço tridimensional. Esse processo de simulação é sucedido pelo condicionamento, realizado por meio da adição dos erros de simulação às estimativas por krigagem. Luster (1985) ilustra um processo de condicionamento análogo, juntamente com vários algoritmos de simulação condicional.

Simulação condicional em uma única etapa (condicionamento simultâneo) foi originalmente proposto por Davis (1987) com o desenvolvimento do método LU (*lower-upper*). Esse método é baseado na decomposição da matriz de covariâncias, em um processo conhecido como decomposição de Cholesky ou decomposição LU (Anderson, 1984; Press *et al.*, 1992, p. 96-98). No método LU, as

anisotropias são consideradas intrinsecamente, por meio da matriz de covariâncias. Apesar de eficiente, esse método é limitado à FAs que não excedam poucas centenas de localizações u_j ($j = 1, \dots, N$) (Alabert, 1987a) – devido a dimensão da matriz a ser decomposta.

Mais tarde, os métodos seqüenciais viriam amenizar as dificuldades encontradas na aplicação dos métodos pioneiros de simulação condicional, citados acima.

3.3.1. Simulação condicional seqüencial

Os métodos de simulação condicional seqüencial são baseados na decomposição da função de densidade de probabilidade (fdp) multivariada do modelo de FA $Z(u)$. A idéia básica desses métodos consiste em amostrar, seqüencialmente, as distribuições condicionais univariadas originadas dessa decomposição. Mais especificamente, a fdp multivariada é decomposta em um produto de séries de distribuições condicionais univariadas (Johnson, 1987).

Sendo $Z(u)$ um modelo de FA com fdp multivariada $f(u_1, \dots, u_N; z_1, \dots, z_N)$, então, essa fdp multivariada pode ser expressa como um produto de N distribuições condicionais univariadas (Isaaks, 1990, p. 15):

$$\begin{aligned} f(u_1, \dots, u_N; z_1, \dots, z_N) &= f(u_1; z_1) \\ &* f(u_2; z_2 | Z(u_1) = z_1) \\ &* \dots * f(u_N; z_N | Z(u_j) = z_j, j = 1, \dots, N - 1) \end{aligned} \quad (3.1)$$

onde z_j é o valor atribuído a VA Z , na localização u_j . Logo, os índices j , atribuídos às localizações u , representam os N nós de grid a serem simulados ($j = 1, \dots, N$). Se todas as distribuições condicionais univariadas da equação acima são conhecidas, então uma realização da FA $Z(u)$ pode ser construída. Isso pode ser realizado por meio de uma seqüência de tiragens aleatórias de cada distribuição condicional univariada. Portanto, para uma realização da FA $Z(u)$ são executadas as seguintes etapas (Isaaks, 1990, p. 15):

- i. Realização z_1 da VA $Z(u_1)$ por meio da tiragem aleatória de um valor a partir da distribuição de densidade $f(u_1; z_1)$.
- ii. A realização z_1 da VA $Z(u_1)$ passa a ser condicionante da distribuição univariada da VA $Z(u_2)$.
- iii. Realização z_2 da VA $Z(u_2)$ por meio da tiragem aleatória de um valor a partir da distribuição marginal condicional $f(u_2; z_2 | Z(u_1) = z_1)$.
- iv. As realizações z_1 e z_2 são usadas como condicionantes da distribuição univariada da VA $Z(u_3)$.
- v. Realização z_3 da VA $Z(u_3)$ por meio da tiragem aleatória de um valor a partir da distribuição marginal condicional univariada $f(u_3; z_3 | Z(u_2) = z_2, Z(u_1) = z_1)$.
- vi. O processo continua em todas as demais localizações u_j ($j = 4, \dots, N$) por meio de tiragens aleatórias e subseqüentes condicionamentos até que a última distribuição $f(u_N; z_N | Z(u_j) = z_j, j = 1, \dots, N - 1)$ for totalmente condicionada.

vii. Realização z_N da VA $Z(u_N)$ por meio da tiragem aleatória de um valor a partir da distribuição marginal condicional univariada $f(u_N; z_N | Z(u_j) = z_j, j = 1, \dots, N - 1)$.

De fato, a distribuição univariada da etapa (i) seria condicionada aos n dados experimentais originais $(z(u_\alpha), \alpha = 1, \dots, n)$. Assim, na localização do grid (não amostrada) u_1 ($j = 1$), uma realização da VA $Z(u_1)$ seria aleatoriamente tirada da distribuição $f(u_1; z_1 | Z(u_\alpha) = z_\alpha, \alpha = 1, \dots, n)$. Subseqüentemente, as distribuições univariadas de todas as demais localizações u_j ($j = 2, \dots, N$) seriam condicionadas aos dados amostrais e as realizações precedentes das VA $Z(u_j)$.

Esse algoritmo exige que todas as N distribuições condicionais univariadas, associadas a fdp multivariada da FA $Z(u)$ sejam conhecidas. Ainda, esse algoritmo seqüencial prevê que todos os dados experimentais (amostras) $z(u_\alpha)$ ($\alpha = 1, \dots, n$) e todos os atributos simulados $z(u_j)$ ($j = 1, \dots, N$) sejam condicionantes das distribuições univariadas nas localizações não simuladas. Ou seja, a construção do algoritmo, embasado na decomposição da fdp multivariada da FA $Z(u)$, exige um número de dados condicionantes progressivamente maior a cada etapa (a cada localização ou nó de grid). O número de dados condicionantes aumenta de n a $n + (N - 1)$, onde n é o número de dados experimentais e N é o número de nós de grid ou localizações a serem simuladas. Essa característica pode dificultar a aplicabilidade do algoritmo, de acordo com sua construção estrita e teoricamente correta. Por isso, na prática, esse algoritmo é empregado de modo a seguir um caminho aleatório. Assim, em cada localização u_j visitada, uma vizinhança local (incluindo dados originais e simulados) é utilizada como condicionante.

Na prática, esse algoritmo seqüencial é implementado sobre distribuições condicionais locais de probabilidade (dclp) cumulativas $F(u; z | (n))$, que indicam a probabilidade das VAs $Z(u)$ assumirem valores inferiores ou iguais a certos valores limite z . Ou seja:

$$F(u; z | (n)) = \Pr \{Z(u) \leq z | (n)\} \quad (3.2)$$

onde a notação " $| (n)$ " representa o condicionamento por n dados.

Essas funções caracterizam e governam as distribuições condicionais univariadas locais $[f(u; z | (n))]$ das VAs $Z(u)$ referidas no algoritmo seqüencial, pois:

$$\int_{-\infty}^z f(u; z | (n)) dz = F(u; z | (n)) = \Pr \{Z(u) \leq z | (n)\} \quad (3.3)$$

O conjunto de valores simulados $\{z(u_j), j = 1, \dots, N\}$, produzido por meio da implementação desse algoritmo seqüencial, representa uma realização da FA $Z(u)$ no domínio físico definido pelas informações geoposicionadas originais $z(u_\alpha)$ ($\alpha = 1, \dots, n$). A repetição de L vezes desse algoritmo garante a geração de L realizações equiprováveis do modelo de FA $Z(u)$ $\{z^{(l)}(u_j), j = 1, \dots, N; l = 1, \dots, L\}$. Logo, em cada localização u_j ($j = 1, \dots, N$), o processo de simulação gera uma distribuição, composta por L valores, que pode ser considerada uma aproximação numérica (método estocástico) da distribuição condicional local de probabilidade (dclp); ou seja:

$$F(u; z | (n)) \approx \frac{1}{L} \sum_{l=1}^L i^{(l)}(u; z) \quad (3.4)$$

onde

$$i^{(l)}(u; z) = \begin{cases} 1 & \text{se } z^{(l)}(u) \leq z \\ 0 & \text{se não} \end{cases} \quad \text{com } l = 1, \dots, L$$

O valor da VA Z em cada localização u_j ($j = 1, \dots, N$) é simulado dentro do domínio de variação da VA $Z(u)$, por meio de um procedimento aleatório, a partir das dclp que caracterizam as distribuições condicionais univariadas locais. O algoritmo seqüencial, apresentado acima, utiliza simulação de Monte Carlo para efetuar as tiragens aleatórias que caracterizam os valores simulados. Nesse processo, são sorteados números, definidos no domínio $[0;1]$, que representam probabilidades p . Essas probabilidades alimentam as funções inversas $F^{-1}(u; p | (n))$ para a obtenção dos valores simulados $z(u)$. Ou seja:

$$F^{-1}(u; p | (n)) = z(u) \quad (3.5)$$

Imagens estocásticas do modelo de FA $Z(u)$ podem ser geradas por meio de várias técnicas que utilizam o algoritmo seqüencial. As variações em torno desse algoritmo podem ser distinguidas pelo modelo de FA que for assumido para caracterizar o fenômeno geológico (natural) em estudo. Nas seções seguintes, serão apresentadas as duas técnicas de simulação condicional seqüencial mais amplamente utilizadas:

- i. simulação seqüencial Gaussiana e
- ii. simulação seqüencial dos indicadores.

3.3.1.1. Simulação seqüencial Gaussiana

Se a distribuição de probabilidade (dp) multivariada descreve as propriedades do modelo de FA $Z(u)$, então, antes de produzir várias realizações desse modelo de FA, sua dp multivariada deve ser totalmente caracterizada. Infelizmente, o histograma e a função de correlação espacial da distribuição experimental (amostras) não são suficientes para caracterizar a dp multivariada da FA $Z(u)$. A caracterização desse tipo de função exige muito mais conhecimento do que os momentos de 1ª e 2ª ordem da distribuição original. Contudo, para um caso muito particular do modelo Gaussiano de FA, o conhecimento de apenas um desses momentos é suficiente (Anderson, 1984). O modelo Gaussiano padrão de FA é totalmente caracterizado por um único parâmetro: a função de correlação espacial $C(h)$. O conhecimento dessa função garante a inferência da dp multivariada do modelo de FA $Z(u)$. Entretanto, como descrito acima (Equação 3.1), o conhecimento das distribuições condicionais univariadas permite a geração de realizações do modelo de FA por meio do algoritmo seqüencial.

O algoritmo seqüencial apresentado na seção 3.3.1, aplicado a um modelo de FA Gaussiano, caracteriza o método de simulação seqüencial Gaussiana (ssG) proposto por Isaaks (1990). Em outras palavras, o método de ssG é caracterizado pela aplicação do algoritmo seqüencial às distribuições condicionais univariadas locais (resultantes da decomposição de uma particular fdp multivariada

Gaussiana), controlada por uma distribuição de probabilidade (dp) multivariada Gaussiana, caracterizada por uma função de covariância $C(h)$.

3.3.1.2. Simulação seqüencial dos indicadores

As limitações de aplicabilidade dos métodos de simulação condicional originalmente propostos (bandas rotativas e LU) incentivaram o desenvolvimento do método não-paramétrico de simulação seqüencial dos indicadores (ssi) (Alabert, 1987b). Esse método de consiste na aplicação do algoritmo seqüencial a um modelo de FA, cujas distribuições condicionais univariadas locais são aproximadas por séries de krigagem dos indicadores (Journel, 1989). A seção 3.6 apresenta as técnicas utilizadas na obtenção das dclp, que caracterizam as distribuições condicionais univariadas, por meio de (co)krigagem.

3.4. Distribuições condicionais locais de probabilidade, dclp

A implementação do algoritmo de simulação seqüencial, apresentado na seção 3.3.1, requer a inferência de distribuições locais de probabilidade (dlp) que sejam condicionadas aos dados experimentais e aos valores simulados nas etapas precedentes. Esse é o grande desafio para aplicabilidade do algoritmo seqüencial e constitui a diferença entre os métodos de simulação condicional seqüencial. Existem várias formas para se determinar uma distribuição local de probabilidade (dlp). Essas formas variam de acordo com nível de conhecimento da VA, mas a idéia é fazer com que a metodologia de sua construção seja dependente da localização e dos próprios valores das informações locais (condicionamento). Assim, almeja-se que o algoritmo de construção dessas distribuições locais de probabilidade seja espacialmente dependente e que considere o conhecimento de que uma dada localização alvo esteja situada numa zona de alta ou de baixa variância.

Qualquer VA possui uma distribuição de probabilidade teórica, porém desconhecida. Raramente, a distribuição de probabilidade pode ser determinada experimentalmente. Na maioria das vezes, ela é aproximada por meio de uma distribuição empírica. A utilização de modelos nos quais suas distribuições teóricas são conhecidas e que as mesmas possam ser identificadas pelas distribuições empíricas pode ser de grande valia (geoestatística não-linear - Chilès e Delfiner, 1999, capítulo 6).

Considerando que a VA $Z(u)$ é conhecida somente nas localizações medidas $(z(u_\alpha), \alpha = 1, \dots, n)$ no domínio de interesse A , a proporção do subdomínio D ($D \subset A$) ocupada pelas localizações onde $Z(u)$ é inferior a um dado valor limite z é calculada como:

$$I(z) = \frac{1}{|D|} \int_D 1_{Z(u) < z} dz \quad (3.6)$$

Deseja-se estimar a proporção de D, considerando como valor limite todos os valores experimentais $z(u_\alpha)$ disponíveis. Para essa finalidade, a esperança condicional $E[I(z)|(n)]$ se apresenta como um estimador ótimo. Na notação da esperança condicional, n representa todas as informações condicionantes $[z(u_\alpha), \alpha = 1, \dots, n]$. A determinação da esperança condicional de $I(z)$ conduz à determinação da distribuição condicional de probabilidade (dcp):

$$F_{\underline{u}}(z | (n)) = \text{Pr ob}\{Z(\underline{u}) < z | (n)\} \quad (3.7)$$

onde \underline{u} representa uma localização aleatória dentro do subdomínio D. Da mesma maneira, $I(z)$ depende diretamente da estimativa de $1_{z(u)<z}$ em qualquer localização u contida em D.

Geralmente, a diferença entre o caso onde D representa uma localização e o caso onde D representa um volume é similar à diferença existente entre krigagem pontual e krigagem de blocos (Chilès e Delfiner, 1999, p. 379). Dessa forma, se o domínio D representa uma localização u, então a função $I(z)$ pode ser simplificada por: $I(z) = 1_{z(u)<z}$. Assim,

$$\text{Pr ob}\{Z(\underline{u}) < z | (n)\} = E[1_{z(u)<z} | (n)] = E[I(z) | (n)] \quad (3.8)$$

Em outras palavras, busca-se uma aproximação para a distribuição condicional

$$F(u; z | (n)) = \text{Pr ob}\{Z(u) < z | (n)\} \quad (3.9)$$

Nesse sentido, a função de repartição $F(z)$ de uma VA localizada em u pode ser considerada como a sua distribuição local de probabilidade condicionada pelas informações circunvizinhas de u. Dessa forma, a probabilidade da VA $Z(u)$ ser superior a um dado valor limite z é calculada como:

$$\text{Pr ob}\{Z(u) > z | (n)\} = 1 - F(u; z | (n)) \quad (3.10)$$

onde a probabilidade é dependente (condicionamento) das n informações disponíveis, que por sua vez são dependentes da localização u. Dessa maneira, as distribuições de probabilidade associadas às VAs passam a ser chamadas de **distribuições condicionais locais de probabilidade, dclp**. Assim, para qualquer distribuição experimental constituída por n medidas, existirão tantas dclp quanto forem as N localizações desconhecidas e existirão modelos de continuidade espacial capazes de medir a correlação espacial entre as informações. Portanto, essas dclp devem ser vistas como modelos (ou estimativas) de incerteza associados às localizações desconhecidas.

A maneira mais simples de se determinar as dclp consiste em assumir um modelo teórico para toda lei espacial (distribuição multivariada) da FA (Goovaerts, 1997, p. 265). Esse modelo deve ser suficientemente robusto para garantir que todas as dclp tenham a mesma expressão analítica e que estas sejam definidas por meio de poucos parâmetros. Nesse contexto, o modelo multivariado Gaussiano de FA é o modelo paramétrico mais utilizado, pois permite o acesso direto às dclp por meio de parâmetros dados por krigagem. As dclp também podem ser inferidas diretamente, sem a utilização de nenhum modelo, por meio do método dos indicadores.

Nas seções seguintes (3.5 e 3.6), serão apresentadas em detalhes as técnicas envolvidas na construção das distribuições condicionais locais de probabilidade (dclp) de acordo com os métodos paramétrico (modelo multiGaussiano) e não-paramétrico (indicadores).

3.5. O modelo multiGaussiano – método paramétrico

A distribuição Gaussiana (Figura 3.1) é uma distribuição muito utilizada em aplicações estatísticas. Ela pode ser prontamente definida e proporciona simetria e outras propriedades que a tornam muito conveniente para o cálculo de estimativas teóricas. A função de densidade de probabilidade Gaussiana $[g(\cdot)]$ é dada por:

$$g(y) = \frac{1}{\sqrt{2\pi}} e^{\left[-\frac{1}{2}y^2\right]} \quad (3.11)$$

onde y é uma variável Gaussiana situada no intervalo $[-\infty; +\infty]$.

Uma distribuição Gaussiana (normal) pode ocorrer se o fenômeno que originou a VA for o resultado do somatório de várias causas independentes. Todas as distribuições normais são centradas no valor médio e são simétricas em torno desse parâmetro estatístico. De modo que, para uma distribuição normal, a média e a mediana são sempre coincidentes.

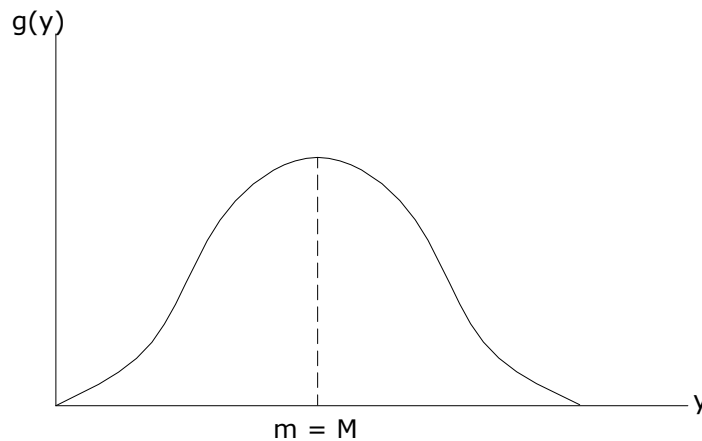


Figura 3.1: Curva da distribuição Gaussiana.

Para facilitar a comparação entre distribuições, é calculada a distribuição cumulativa de probabilidade ou a função de repartição $G(y)$ da VA Gaussiana Y :

$$\int_{-\infty}^y g(y) dy = G(y) \quad (3.12)$$

Essa função não possui uma expressão analítica simples. Na prática, utilizam-se tabelas ou, às vezes, por necessidade computacional, de fórmulas de aproximação. Qualquer variável normalmente distribuída (não reduzida) pode ser transformada linearmente em uma variável normal padrão (reduzida), isto é, que seja normalmente distribuída apresentando $m=0$ e $\sigma^2=1$. Para uma VA Gaussiana X não reduzida (com média (m) e variância (σ^2) diferentes de zero e um, respectivamente) essa transformação é dada por:

$$X = m + \sigma.Y \quad (3.13)$$

onde Y é uma variável Gaussiana padrão [$Y \sim N(0;1)$] dada por:

$$Y = \frac{X - m}{\sigma} \quad (3.14)$$

Essa expressão é muito conveniente, pois permite a determinação de X por meio de m , σ e de uma variável Gaussiana padrão associada (Y), de função de repartição $G(\cdot)$ conhecida. Assim, para qualquer VA normalmente distribuída, a função de densidade de probabilidade Gaussiana pode ser reescrita em função de dois parâmetros: a média m e a variância σ^2 :

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\left[-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right]} \quad (3.15)$$

De acordo com Anderson (1958), se $Y(u)$ ($u \in A$) é uma FA multivariada Gaussiana padrão com função de covariância $C_Y(h)$, então as cinco declarações abaixo são válidas:

- i. Todos os subconjuntos dessa FA também são multivariados normais, por exemplo: $\{Y(u), u \in D \subset A\}$ é multivariado Gaussiano.
- ii. Qualquer combinação linear de primeira ordem da VA é normal. Por exemplo:

$V(u) = \sum_{\alpha=1}^n \lambda_{\alpha} Y(u_{\alpha})$ é normalmente distribuído para qualquer localização $u \in A$ e qualquer conjunto de ponderadores λ_{α} .

- iii. A distribuição constituída por qualquer conjunto de pares das VAs $Y(u)$ e $Y(u+h)$ é normal e totalmente determinada pela função de covariância $C_Y(h)$.
- iv. Se duas VAs $Y(u)$ e $Y(u')$ não apresentam correlação, isto é, se $\text{Cov}\{Y(u), Y(u')\} = 0$, elas também são independentes.
- v. As distribuições condicionais de qualquer subconjunto da FA $Y(u)$ são multivariadas normais. Em particular, a distribuição condicional de uma VA $Y(u)$ é normal e totalmente caracterizada por sua média e variância. Sendo esses dois parâmetros condicionados à VA $Y(u)$ por meio das $n(u)$ informações $y(u_{\alpha})$ ($\alpha = 1, \dots, n$), logo:

$$[G(u; y | (n))] = G\left(\frac{y - E\{Y(u) | (n)\}}{\sqrt{\text{Var}\{Y(u) | (n)\}}}\right) \quad (3.16)$$

onde $G(\cdot)$ é a distribuição cumulativa Gaussiana de probabilidade e y é um quantil da distribuição da FA $Y(u)$.

De acordo com o conceito do modelo multiGaussiano, a média condicional e a variância condicional das distribuições condicionais univariadas locais (Equação 3.16) são iguais, respectivamente, a sua estimativa $y^*(u)$ por krigagem simples (KS) e a sua variância da estimativa $\sigma_{KS}^2(u)$ por KS. Onde $y^*(u)$ e $\sigma_{KS}^2(u)$ são calculados a partir das $n(u)$ informações $y(u_{\alpha})$ ($\alpha = 1, \dots, n$) circunvizinhas de u (Journal e Huijbregts, 1978, p. 566-567), ou seja:

$$E[Y(u) | (n)] = \sum_{\alpha=1}^n \lambda_{\alpha} \cdot y(u_{\alpha}) + (1 - \sum_{\alpha=1}^n \lambda_{\alpha}) \cdot E[Y(u)] = y^*(u) \quad (3.17)$$

$$\text{Var}[Y(u) | (n)] = \text{Var}[Y(u)] - \sum_{\alpha=1}^n \lambda_{\alpha} \cdot \text{Cov}(u, u_{\alpha}) = \sigma_{KS}^2(u) \tag{3.18}$$

onde λ_{α} são os pesos atribuídos aos dados $y(u_{\alpha})$ ($\alpha = 1, \dots, n$), calculados por meio da resolução de um sistema de KS; $\text{Cov}(u, u_{\alpha})$ são as covariância espaciais em relação à localização alvo u e as localizações u_{α} .

Portanto, a inferência das dclp passa pela resolução de um sistema de KS e as mesmas podem ser modeladas como:

$$[G(u; y | (n))]_{KS}^* = G\left(\frac{y - y_{KS}^*(u)}{\sigma_{KS}(u)}\right) \tag{3.19}$$

Dessa forma, para uma VA Gaussiana padrão $Y [Y \sim N(0;1)]$, em uma localização não medida u , se as informações disponíveis estiverem fora do alcance de correlação espacial mostrado pelo semi-variograma das n informações $y(u_{\alpha})$ ($\alpha = 1, \dots, n$), as estimativas serão dadas por:

$$y_{KS}^*(u) = m = 0 \quad \text{e}$$

$$\sigma_{KS}(u) = 1$$

$$\text{logo, } [G(u; y | (n))]_{KS}^* = G(y).$$

Por outro lado, como os interpoladores de krigagem são exatos (KS, nesse caso), se $y(u)$ for conhecido:

$$y_{KS}^*(u) = y(u)$$

$$\sigma_{KS}(u) = 0$$

$$\text{logo, } [G(u; y | (n))]_{KS}^* = \begin{cases} 1 & \text{se } y \geq y(u) \\ 0 & \text{se } y < y(u) \end{cases}$$

A Figura 3.2 ilustra as situações particulares expostas acima.

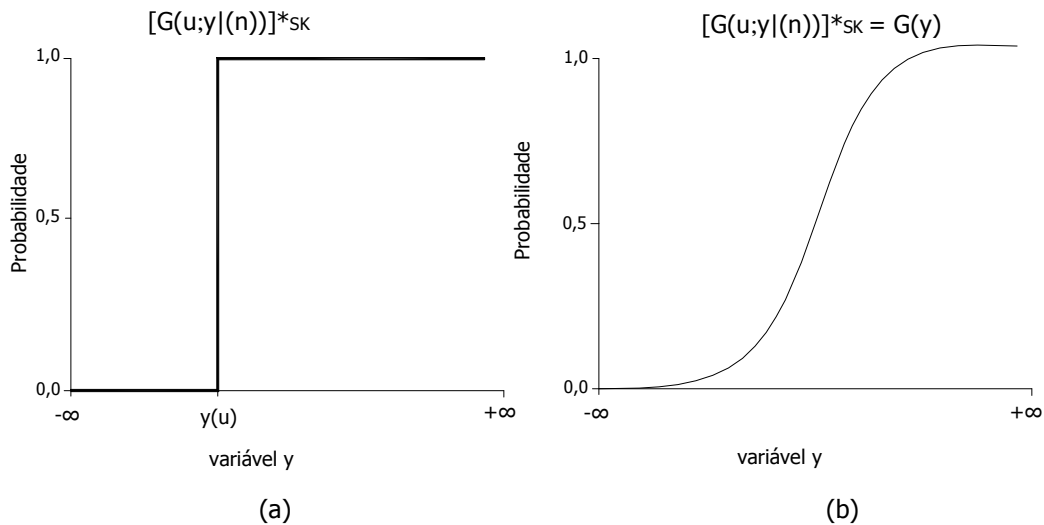


Figura 3.2: Distribuições condicionais locais de probabilidade (dclp) cumulativas para o caso de $y(u)$ ser conhecido (a) e para o caso onde as informações circunvizinhas de u estiverem além da distância de correlação espacial (b).

A hipótese fundamental do modelo multiGaussiano assume que os $(n+1)$ vetores dimensionais $(Y(u), Y(u_1), \dots, Y(u_n))$ apresentam o comportamento de uma distribuição multivariada Gaussiana (Chilès e Delfiner, 1999, p. 381). Essa hipótese garante que a distribuição condicional de $Y(u)$ seja Gaussiana; ou seja:

$$[Y(u) | \{Y(u_\alpha), \alpha = 1, \dots, n\}] = [Y(u)|(n)] = Y_{KS}^*(u) + \sigma_{KS} \cdot y \quad (3.20)$$

onde $Y_{KS}^*(u)$ é a estimativa de $Y(u)$ por meio de krigagem simples, σ_{KS} é seu desvio padrão de krigagem associado à estimativa, e y é uma variável normal padrão (que apresenta $m = 0$, $\sigma = 1$ e de função de repartição $G(\cdot)$ conhecida).

Embora o embasamento teórico do formalismo multiGaussiano exija a condição de estacionaridade, isto é, $E[Y(u)] = E(Y)$, $\forall u \in A$ (daí a utilização de KS), segundo Goovaerts (1997) é possível considerar o comportamento não estacionário. Por exemplo, as dclp podem ser determinadas com auxílio de outros algoritmos geoestatísticos, como krigagem ordinária (KO), que considera as médias locais através de toda a área em estudo. Também pode ser empregado o algoritmo de krigagem em presença de deriva. Entretanto, a aplicação desses algoritmos deve ser muito cuidadosa, uma vez que a variância da dclp deve ser identificada com a σ_{KS}^2 teoricamente correta (Journel, 1980). A utilização de interpoladores diferentes de KS pode introduzir tendências irreversíveis na metodologia, pois é incorporada de uma inconsistência teórica, na medida em que, de acordo com o conceito do modelo multiGaussiano, a esperança condicional é igual à estimativa por KS ($E[Y(u)|(n)] = Y_{KS}^*(u)$). Por exemplo, se $Z(u)$ for localmente estacionário e optar-se pelo uso de KO, o viés pode ser corrigido assumindo-se que a distribuição de $Z(u)$ é espacialmente lognormal. Desse modo, a esperança condicional pode ser considerada igual à estimativa por krigagem lognormal ordinária, KLO, e de variância igual a σ_{KS}^2 , isto é: $E[Y(u)|(n)] = Y_{KLO}^*(u)$ e $\sigma^2 = \sigma_{KS}^2$.

Como descrito acima, o uso do modelo multivariado Gaussiano de FA para a construção da dclp, em qualquer localização u não medida, se mostra muito prático. Entretanto, o modelo exige que a distribuição das n informações $z(u_\alpha)$ ($\alpha = 1, \dots, n$) siga um comportamento Gaussiano, isto é, que a distribuição experimental esteja no espaço normal – Gaussiano. Esse é o requisito necessário à praticidade de aplicação do modelo multiGaussiano para a determinação das dclp.

3.5.1. Anamorfose Gaussiana

Na maioria dos casos em geociências, a distribuição experimental original não apresenta um comportamento Gaussiano. Assim, a distribuição experimental das n informações $z(u_\alpha)$ ($\alpha = 1, \dots, n$) é transformada em uma distribuição de n valores $y(u_\alpha)$ ($\alpha = 1, \dots, n$) por meio da função de transformação normal $\varphi(\cdot)$ (também chamada de função de anamorfose Gaussiana), tal como a expressão abaixo:

$$z(u_\alpha) = \varphi(y(u_\alpha)) \quad (3.21)$$

A Figura 3.3 apresenta a representação gráfica da função $\phi(\cdot)$ que relaciona as variáveis z e y .

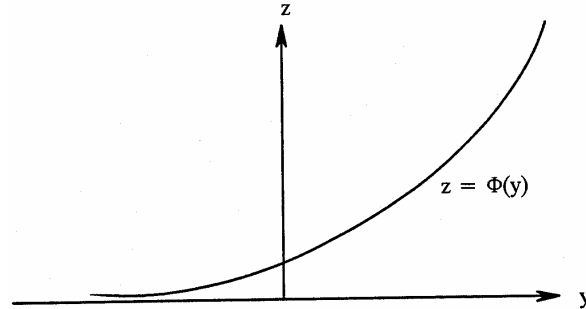


Figura 3.3: Função de anamorfose. Fonte: Rivoirard, 1991 p. 40.

O conhecimento da função de anamorfose de $Z(u)$ é equivalente ao conhecimento de sua distribuição. Em particular, a anamorfose incorpora todas as irregularidades que a distribuição experimental possa apresentar.

Em seguida, o modelo multiGaussiano pode ser aplicado sobre as n informações $y(u_\alpha)$ ($\alpha = 1, \dots, n$) de modo que:

$$G(u; y | (n)) = \text{Pr ob}\{Y(u) \leq y | (n)\} \tag{3.22}$$

Pelo fato de que a função de anamorfose Gaussiana $\phi(\cdot)$ é bijetiva (monotônica crescente), a dclp da variável original pode ser diretamente acessada, pois

$$\begin{aligned} F(u; z | (n)) &= \text{Pr ob}\{Z(u) \leq z | (n)\} = \text{Pr ob}\{Y(u) \leq y | (n)\} \\ &= G(u; \phi(z) | (n)) \end{aligned} \tag{3.23}$$

Ou seja, as duas FAs $[F(z)$ e $G(y)]$ que caracterizam as duas VAs $Z(u)$ e $Y(u)$ apresentam a mesma probabilidade cumulativa para os mesmos quantis (Figura 3.4).

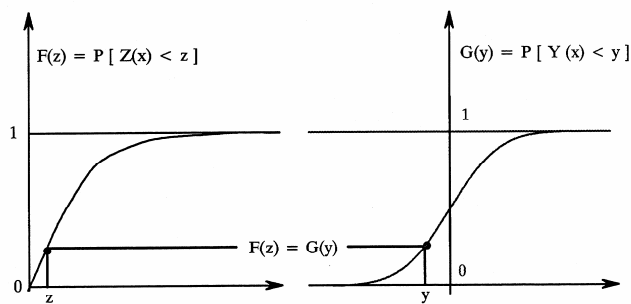


Figura 3.4: Funções de repartição $F(z)$ e $G(y)$, mostrando que o valor de $y(u)$ associado ao de $z(u)$ é tal que $F(u) = G(u)$. Fonte: Rivoirard, 1991 p. 40.

Na prática, o procedimento da transformação normal se faz por meio de três etapas (Goovaerts, 1997, p.268):

- i. As n informações $z(u_\alpha)$ ($\alpha = 1, \dots, n$) devem ser classificadas em ordem crescente:

$$[z(u_\alpha)]^{(1)} \leq \dots \leq [z(u_\alpha)]^{(k)}, \quad \alpha = 1, \dots, n.$$

onde o índice k indica a ordem de classificação da informação $z(u_\alpha)$. Como a função $G(\cdot)$ dever ser monotônica crescente, se existirem informações de valores iguais, deve-se adotar um procedimento de desempate desses valores. Esse procedimento pode ser aleatório ou de acordo com a média local no entorno das $n(u)$ medidas $z(u_\alpha)$ circunvizinhas de u . A última solução parece ser mais coerente, na medida em que as informações localizadas em áreas de altos valores receberão índices de classificação superiores.

ii. A freqüência cumulativa das n medidas $z(u_\alpha)$ de índice de classificação k são calculadas

como $p_k^* = \frac{1}{n} \left(k - \frac{1}{2} \right)$, se as medidas forem regularmente espaçadas e o histograma

original for representativo de toda a área em estudo. Por outro lado, se houver a necessidade da utilização de pesos de desaninhamento (w_α), as freqüências cumulativas

são calculadas como: $p_k^* = \sum_{i=1}^k w_i - \frac{1}{2} w_k$; onde w_k é o peso de desaninhamento associado à

amostra de índice k e w_i ($i = 1, \dots, k$) são os pesos de desaninhamento associados às amostras de índices inferiores ou iguais a k .

iii. A anamorfose das z -informações em y -informações é obtida por meio dos quantis associados a p_k^* da distribuição cumulativa Gaussiana.

$$y(u_\alpha) = G^{-1} \left[F^* (Z(u_\alpha)) \right] = G^{-1} (p_k^*) \quad (3.24)$$

Graficamente, os y -valores são extraídos diretamente no eixo das abscissas do gráfico da distribuição cumulativa Gaussiana a partir de cada quantil $q_{p_k^*}^*$ (no eixo das ordenadas),

como mostra a Figura 3.4. Como exercício, as y -informações podem ser obtidas por meio

de tabelas da função de repartição $G(y)$, onde a cada quantil $q_{p_k^*}^*$ de probabilidade p_k^* , é

associado um valor (y) dentro do domínio $[-\infty; +\infty]$. Os valores de $y(u_\alpha)$ sempre estarão

dentro desse intervalo, na medida em que a probabilidade de p_1^* e p_n^* em $G(y)$ é sempre diferente de zero e um, respectivamente.

Para que se possa utilizar o modelo multiGaussiano de FA na construção das dclp, é necessário que a FA que exprime a distribuição experimental apresente um comportamento multinormal. Isto é, é necessário que a distribuição constituída pelos n valores $y(u_\alpha)$ ($\alpha = 1, \dots, n$) apresente um comportamento multivariado Gaussiano. Porém, a simples normalização da distribuição original não garante que o modelo de FA seja multinormal. Esse procedimento garante apenas a normalidade univariada da FA. A multinormalidade da FA só é assegurada se as distribuições bivariadas, trivariadas, ..., n -variadas das y -informações seguirem um comportamento Gaussiano.

Contudo, verificar a multinormalidade de uma distribuição é uma tarefa muito difícil, senão impossível, de modo que a garantia de binormalidade se torna suficiente para a aplicabilidade do modelo multiGaussiano de FA.

3.5.2. Verificação de binormalidade

Na prática, a constatação da binormalidade valida a utilização do modelo multiGaussiano $G(u;y|(n))$. Existem vários testes que podem ser utilizados para verificar a binormalidade da FA. Um deles consiste em verificar se a distribuição da combinação linear entre qualquer conjunto de pares de y -valores separados pelo mesmo vetor h caracteriza uma distribuição Gaussiana. Ou seja, se, por exemplo, a distribuição constituída por $Y(u) - Y(u+h)$, $\forall u, \forall h$, também for normal (Deutsch e Journal, 1998, p. 142-144). Outro teste de verificação da binormalidade da FA considera a relação madograma/semi-variograma. Essa relação (Equação 3.25) deve se aproximar à constante 0,564 e deve ser verificada para diversas distâncias h (passos).

$$\frac{\gamma_1(h)}{\sqrt{\gamma_2(h)}} = \frac{1}{\sqrt{\pi}} = 0,564 \quad (3.25)$$

onde:

$$2 \gamma_1(h) = E[| Y(u+h) - Y(u) |]$$

$$2 \gamma_2(h) = E\{[Y(u+h) - Y(u)]^2\}$$

Deutsch e Journal (1998, p. 231-232) propuseram um teste de binormalidade fundamentado na comparação entre a função covariância de uma distribuição cumulativa bivariada normal calculada para certos quantis e o covariograma dos indicadores calculado para os mesmos quantis (porém, com a distribuição dos dados originais). Isto é, esse procedimento consiste na comparação entre os covariogramas teóricos e os covariogramas experimentais dos indicadores. O covariograma dos indicadores deve reproduzir esse modelo de função covariância da distribuição cumulativa bivariada. Esse teste normalmente é realizado utilizando os quartis inferior, superior e a mediana.

3.5.3. Discretização da dclp cumulativa experimental

Como descrito anteriormente, o modelo de incerteza (dclp) da VA $Z(u)$, $\forall u \in A$, é determinado por meio do modelo de incerteza da VA Y , no espaço normal, que descreve cada localização u . Assim, a dclp no espaço original é construída em cada localização alvo, de modo que para cada valor de y , de probabilidade p , seja associado um valor de z com a mesma probabilidade p ; isto é:

$$F(u;z | (n)) = G(u;y | (n)) \quad (3.26)$$

Entretanto, essa associação é limitada pela densidade da informação original; pois só é possível fazê-la para os valores de z cujas probabilidades são conhecidas – Figura 3.5.

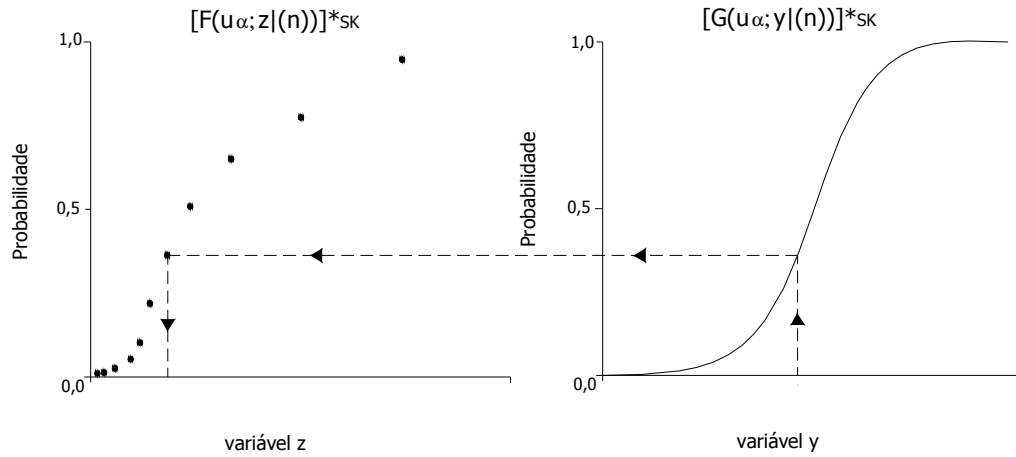


Figura 3.5: Transformação inversa, limitada pelo nº de informações z.

Na verdade, só se conhece a função de repartição empírica $\hat{F}(z)$, que é constituída pelo número n de medidas disponíveis. A função de repartição empírica $\hat{F}(z)$ é definida por:

$$\hat{F}(z) = \frac{1}{n} \sum_{\alpha=1}^n 1_{z(u_\alpha) < z} \tag{3.27}$$

Dessa forma, a função de anamorfose que relaciona as VAs Z e Y é na verdade uma função de anamorfose empírica $\hat{\phi}(y)$ (Figura 3.6), dada por:

$$\hat{\phi}(y) = \sum_{\alpha=1}^n z_\alpha \cdot 1_{y \in I_\alpha} \quad \text{sendo que:} \quad 1_{y \in I_\alpha} = \begin{cases} 1, & \text{se } y \in I_\alpha \\ 0, & \text{se } y \notin I_\alpha \end{cases}$$

onde I_α são os segmentos $\left[G^{-1}\left(\frac{\alpha-1}{n}\right), G^{-1}\left(\frac{\alpha}{n}\right) \right]$ e z_α ($\alpha = 1, 2, \dots, n$) é tal que $z_1 < z_2 < \dots < z_n$.

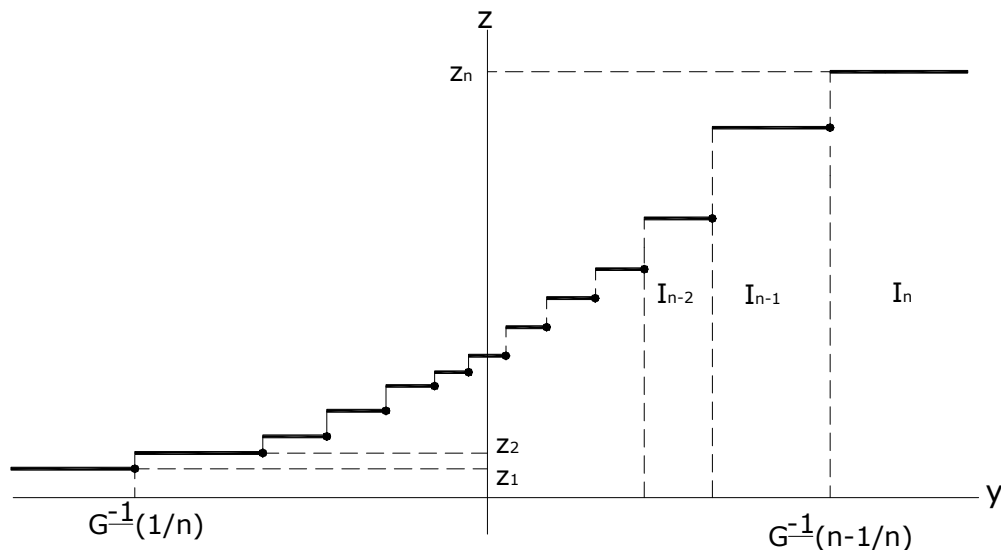


Figura 3.6: Anamorfose empírica.

Assim, a inferência da incerteza associada a um dado valor z diferente de qualquer valor $z(u_\alpha)$ não pode ser diretamente determinada, pois $\hat{\phi}(y)$ não é bijetiva (monotônica crescente). Portanto, a função de repartição empírica $\hat{F}(z)$ (e conseqüentemente $\hat{\phi}(y)$) deve ser discretizada para aumentar sua resolução e alcançar a forma de ϕ mostrada na Figura 3.3.

A partir do procedimento de anamorfose Gaussiana (procedimento gráfico) descrito anteriormente, a resolução da função de repartição empírica $\hat{F}(z)$ pode ser aumentada por meio da interpolação dentro de cada classe de valores $(z_{k-1}, z_k]$ e extrapolação das classes extremas. Deutsch e Journal (1998) propõem vários modelos de interpolação e extrapolação para cada parte da distribuição: trecho extremo inferior, classes $(z_{k-1}, z_k]$ e trecho extremo superior.

O modelo linear de interpolação, geralmente, é empregado entre os valores experimentais, isto é, dentro das classes de valores $(z_{k-1}, z_k]$. Esse modelo, que proporciona uma distribuição uniforme entre os valores experimentais $z(u_\alpha)$, é dado por:

$$[F(z)]_{Lin} = F^*(z_{k-1}) + \left[\frac{z - z_{k-1}}{z_k - z_{k-1}} \right] \cdot [F^*(z_k) - F^*(z_{k-1})], \quad \forall z \in (z_{k-1}, z_k] \quad (3.28)$$

O modelo potência também pode ser utilizado para a interpolação entre os valores experimentais. Esse modelo é um caso particular do modelo linear:

$$[F(z)]_{Pot} = F^*(z_{k-1}) + \left[\frac{z - z_{k-1}}{z_k - z_{k-1}} \right]^w \cdot [F^*(z_k) - F^*(z_{k-1})], \quad \forall z \in (z_{k-1}, z_k] \quad (3.29)$$

onde a potência w deve ser obrigatoriamente superior a zero. Vários modelos de interpolação podem ser inferidos variando o valor de w (Figura 3.7):

- i. $w=1$ corresponde ao modelo linear;
- ii. $w<1$ gera distribuições de assimetria positiva;
- iii. $w>1$ gera distribuições de assimetria negativa.

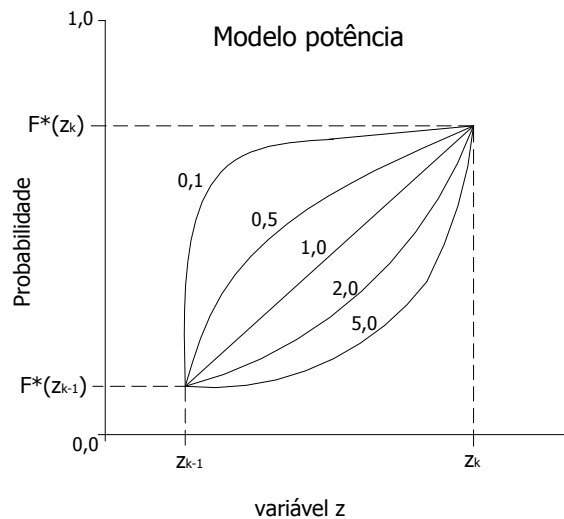


Figura 3.7: Modelo potência de acordo o valor atribuído a w .

Esse modelo também pode ser usado para modelar o trecho inferior e superior da distribuição cumulativa de probabilidade dos n valores de z . Nesse caso, os valores extremos z_{\min} e z_{\max} devem ser pré-determinados. Por exemplo, o trecho inferior de uma distribuição pode ser modelado como:

$$[F(z)]_{\text{Pot}} = \left[\frac{z - z_{\min}}{z_1 - z_{\min}} \right]^{w > 1} \cdot F^*(z_1) , \quad \forall z \in (z_{\min}, z_1] \quad (3.30)$$

pois $F(z_{\min}) = 0$, assim como $F(z_{\max}) = 1$.

Na modelagem do trecho superior por meio do modelo potência, a escolha de z_{\max} é muitas vezes arbitrária. Como alternativa, o modelo hiperbólico permite uma aproximação assintótica da probabilidade máxima. Isto é: $F(z \rightarrow \infty) = 1$. Essa aproximação é realizada de acordo com uma distribuição de assimetria positiva. Assim o modelo hiperbólico é definido como:

$$[F(z)]_{\text{Hip}} = 1 - \frac{\lambda}{z^w} , \quad \forall z > z_k \text{ e } w \geq 1 \quad (3.31)$$

Nesse caso, o parâmetro w controla a taxa de aumento que faz a distribuição cumulativa de probabilidade experimental alcançar a probabilidade máxima (Figura 3.8). O parâmetro λ é definido por:

$$\lambda = z_k^w \cdot [1 - F^*(z_k)] \quad (3.32)$$

De modo que se $z = z_k$, o modelo hiperbólico se identifica com a probabilidade cumulativa experimental $F^*(z_k)$.

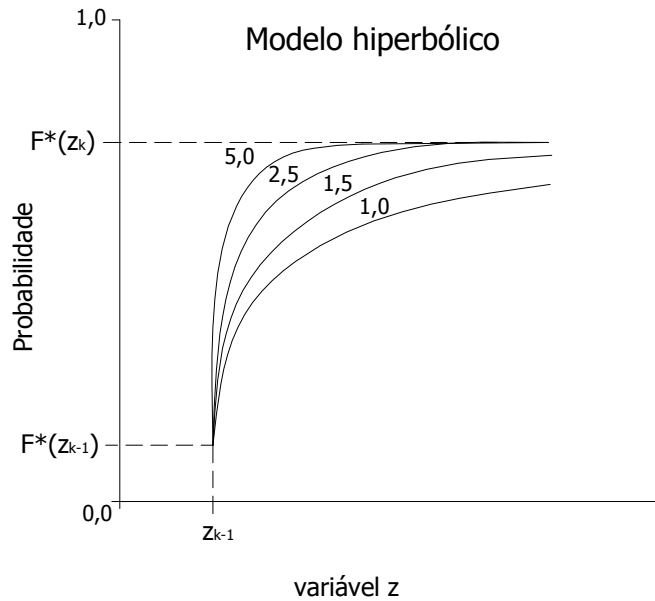


Figura 3.8: Modelo hiperbólico, de acordo o valor atribuído a w .

Outro procedimento utilizado para aumentar a resolução da função de anamorfose empírica $\hat{\varphi}(y)$ é desenvolvê-la com auxílio dos polinômios de Hermite. Em outras palavras, a função $\hat{\varphi}(y)$ pode ser discretizada (e modelada) por meio do ajuste de um modelo analítico polinomial de ordem k . Isso somente é possível porque as distribuições bivariadas do modelo Gaussiano são isofatoriais.

Em geral, as distribuições bivariadas da FA $Y(u)$, de função de repartição $G(y)$, constituem um modelo isofatorial se a distribuição bivariada do par $(Y(u), Y(u+h))$, $\forall u$ e $\forall h$, puder ser fatorizada. A principal propriedade de um modelo isofatorial está no fato de que existe um conjunto de funções $\{\chi_k(y): k = 0, 1, \dots\}$ no qual pode ser utilizado para decompor uma função de $Y(u)$ [$\varphi(y)$] em uma série como:

$$\sum_{k=0}^{\infty} \frac{\varphi_k(y)}{k!} \cdot \chi_k(y) \tag{3.33}$$

onde os coeficientes φ_k são dados por:

$$\varphi_k = \int_{-\infty}^{+\infty} \varphi(y) \cdot \chi_k(y) \cdot g(y) \cdot dy \tag{3.34}$$

Na verdade, as k funções $\chi_k(y)$ são FAs que representam os fatores dos modelos isofatoriais. Esses modelos são denominados isofatoriais, pois os fatores χ_k empregados na decomposição da distribuição bivariada do par $(Y(u), Y(u+h))$ são independentes do vetor h ($\chi_k(y)$ são FAs espacialmente independentes). Chilès e Delfiner (1999, p. 396) apresentam algumas propriedades dos modelos isofatoriais.

Toda a função de quadrado integrável, com relação à densidade Gaussiana $g(\cdot)$, pode ser desenvolvida em polinômios de Hermite (Lantuéjoul 1994, p. 24). Como $\int_{-\infty}^{+\infty} \varphi^2(y) \cdot g(y) \cdot dy < +\infty$, a função de anamorfose empírica $\hat{\varphi}(y)$ pode ser expandida em uma série infinita como:

$$\hat{\varphi}(y) = \sum_{k=0}^{\infty} \frac{\varphi_k}{k!} H_k(y) \tag{3.35}$$

onde os coeficientes φ_k são calculados a partir de :

$$\varphi_k = \int_{-\infty}^{+\infty} \varphi(y) \cdot H_k(y) \cdot g(y) \cdot dy \tag{3.36}$$

Nesse caso, as funções $\chi_k(y)$ são constituídas pelos polinômios de Hermite $H_k(y)$, ou seja: $\chi_k(y) = H_k(y)$.

A integral dos coeficientes φ_k pode ser resolvida por meio de métodos numéricos de integração ou pela expansão em um somatório discreto, envolvendo os valores experimentais z_{α} , a densidade Gaussiana $g(\cdot)$ e os polinômios de Hermite:

$$\varphi_k = \sum_{\alpha=1}^n z_{\alpha} \Delta_{\alpha}[H_{k-1}g] \tag{3.37}$$

onde Δ_α representam os incrementos entre valores $y_{\alpha-1} = G^{-1}((\alpha-1)/n)$ e $y_\alpha = G^{-1}(\alpha/n)$, e são definidos como:

$$\Delta_{\alpha[H_{k-1}g]} = H_{k-1}(y_\alpha) \cdot g(y_\alpha) - H_{k-1}(y_{\alpha-1}) \cdot g(y_{\alpha-1}) \quad (3.38)$$

A grande vantagem de se utilizar os polinômios de Hermite, para modelar a função de anamorfose empírica, é que os coeficientes são invariantes, independente do grau do polinômio. Ao contrário, em um ajuste por meio de um polinômio comum, os coeficientes dos monômios $(1, y, y^2, \dots)$ variam de acordo com o grau do polinômio.

Os polinômios de Hermite são definidos por meio da fórmula de Rodrigues (Chilès e Delfiner, 1999, p. 399):

$$H_n(y) = \frac{1}{g(y)} \frac{d^n g(y)}{dy^n} = \frac{n^{\text{ésima derivada da densidade Gaussiana}}}{\text{densidade Gaussiana}} \quad (3.39)$$

A partir do cálculo da primeira e segunda derivada, os dois primeiros monômios são definidos como: $H_0(y) = 1$ e $H_1(y) = -y$. Os demais monômios são calculados por meio da seguinte fórmula de recorrência:

$$H_{n+1}(y) = -y H_n(y) - n H_{n-1}(y) \quad (3.40)$$

Dessa forma:

$$H_2(y) = y^2 - 1,$$

$$H_3(y) = -y^3 + 3y, \text{ e assim por diante.}$$

Porém, a expansão de $\hat{\varphi}(y)$ em polinômios de Hermite não a torna bijetiva, mas garante sua continuidade (em termos de polinômios de Hermite). Entretanto, é possível torná-la bijetiva a partir de um truncamento (dentro do domínio das informações de y) no termo k da série $\varphi^*(y) = \sum_{k=0}^k \frac{\varphi_k}{k!} H_k(y)$.

A discretização de $\varphi^*(y)$ aumenta proporcionalmente com k , fazendo com que $\varphi^*(y)$ (modelo polinomial) se ajuste gradativamente à função de anamorfose empírica $\hat{\varphi}(y)$. Contudo, a utilização de um número elevado de monômios e coeficientes (por exemplo, $k = 1000$) pode comprometer a bijetividade de $\varphi^*(y)$. Wackernagel (2002) recomenda a utilização de 30 a 50 coeficientes. Segundo David (1988, p. 177), para esse método, os seguintes critérios são considerados para avaliar o desempenho do processo de normalização:

- i. o primeiro coeficiente (φ_0) deve se aproximar à média experimental do conjunto de informações, pois teoricamente (e facilmente demonstrado): $E[Z(x)] = \varphi_0$
- ii. o modelo $\varphi^*(y)$ deve se ajustar à função de anamorfose empírica $\hat{\varphi}(y)$;
- iii. a distribuição transformada $y(u_\alpha) = \varphi^{-1}(z(u_\alpha))$ deve seguir um comportamento estritamente normal $N(0,1)$.

3.5.4. Considerações finais sobre a anamorfose Gaussiana

A normalização por meio da anamorfose Gaussiana é um processo simples, entretanto sua aplicação nem sempre é possível. Por exemplo, se a VA $Z(u)$ apresenta 50% de seus valores nulos (o que corresponderia a 50% de valores negativos no espaço Gaussiano), seria necessário classificá-los em ordem crescente por meio de um dos processos especiais (de desempate) descritos anteriormente. Isso acarretaria em estimativas fortemente tendenciosas ou até aleatórias, comprometendo a fidedignidade da metodologia geoestatística. Ainda, essa classificação especial poderia degradar a biGaussianidade ou multiGaussianidade teoricamente assumida pelo método.

Portanto, a anamorfose Gaussiana pode ser adequadamente empregada somente quando os histogramas das distribuições experimentais revelam um comportamento contínuo e não muito assimétrico. O mesmo se aplica ao procedimento de anamorfose por meio da expansão limitada dos polinômios de Hermite. Dessa forma, para os casos onde:

- i. a distribuição experimental das n medidas $z(u_{\alpha})$ ($\alpha = 1, \dots, n$) apresenta uma freqüência significativa de valores iguais (ou quase iguais) ou com forte assimetria;
- ii. a distribuição transformada pela anamorfose Gaussiana apresenta leis bivariadas que não são biGaussianas;

é necessário modificar a estratégia de tratamento das informações e partir em direção aos modelos não-paramétricos (indicadores), os quais não impõem nenhuma forma pré-determinada nem expressões analíticas para as distribuições condicionais.

3.6. Método dos Indicadores – método não-paramétrico

O método dos indicadores (Journel, 1983) é um método não-paramétrico no qual a dclp $F(u; z|(n))$ não é determinada com auxílio de modelos (por exemplo, o modelo multiGaussiano, onde a média e a variância são suficientes para a modelagem da dclp em qualquer localização u , $u \in A$). Nesse método, a construção dos modelos locais de incerteza (dclp) é realizada diretamente a partir das informações, com auxílio de um conjunto de K valores limite z_k dentro do domínio de variação de $Z(u)$ (Figura 3.9) onde:

$$F(u; z_k | (n)) = \text{Pr ob}\{Z(u) \leq z_k | (n)\} \quad k = 1, \dots, K \quad (3.41)$$

Para cada localização u ($u \in A$) a dclp é constituída pela probabilidade referente a cada valor limite z_k . Desse modo, o modelo de incerteza local será tão discretizado quanto maior for o número K de valores limite z_k utilizado. Portanto, a construção das dclp depende do número de informações disponíveis, nesse caso, os valores z_k . Assim, como no caso do método paramétrico, a dclp deve ser interpolada dentro de cada classe $(z_k; z_{k+1}]$ e extrapolada além dos dois valores limite extremos (z_1 e z_k).

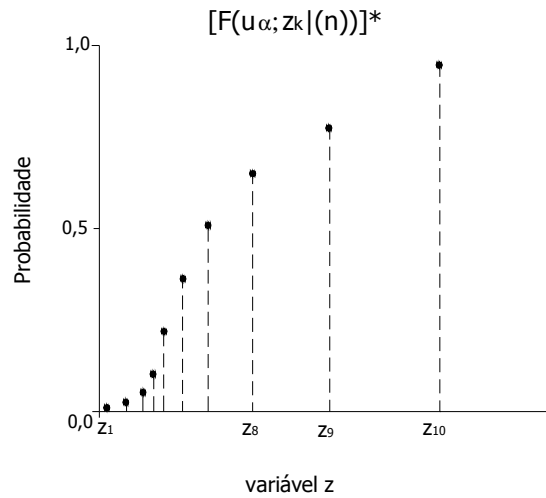


Figura 3.9: Modelo de incerteza para a localização u (dclp) a partir do método não-paramétrico. Dez valores limite z_k ($k=1, \dots, 10$) foram usados para a construção da dclp.

A construção da dclp em qualquer localização u ($u \in A$) só é possível porque no método dos indicadores a distribuição marginal da VA $I(u; z_k)$ é totalmente caracterizada por um único parâmetro: sua esperança matemática. Ainda, esse parâmetro é igual à distribuição local de probabilidade (dlp) da VA $Z(u)$ para um dado valor limite z_k . Isto é:

$$E[I(u; z_k)] = \text{Pr ob}\{Z(u) \leq z_k\} \tag{3.42}$$

$$E[I(u; z_k)] = F(u, z_k) \quad k = 1, \dots, K$$

Portanto, de forma análoga, a esperança condicional da VA $I(u; z_k)$ é igual a distribuição condicional local de probabilidade (dclp) da VA $Z(u)$:

$$E[I(u; z_k | (n))] = \text{Pr ob}\{Z(u) \leq z_k | Z(u_\alpha) = z_\alpha, \alpha = 1, \dots, n\} \tag{3.43}$$

$$E[I(u; z_k | (n))] = \text{Pr ob}\{Z(u) \leq z_k | (n)\} \quad k = 1, \dots, K$$

Em outras palavras, a probabilidade condicional da VA $Z(u)$ ser inferior ou igual a um dado valor limite z_k é igual à esperança condicional da VA indicadora $I(u; z_k)$.

Para construir a VA indicadora $I(u; z_k)$, as n informações disponíveis sobre as VAs $Z(u_\alpha)$ ($\alpha = 1, \dots, n$) devem ser previamente transformadas em VAs $I(u_\alpha; z_k)$. A resposta de $I(u_\alpha; z_k)$ é binária, valendo 1 se $Z(u_\alpha) \leq z_k$ e 0 se não. Os valores indicadores $i(u; z_k)$ são interpretados como resultados ou realizações de um conjunto de VAs $I(u; z_k)$ que constituem um modelo de FA indicador $I(u; z_k)$.

Como cada informação condicionante $z(u_\alpha)$ ($\alpha = 1, \dots, n$) é codificada de acordo com K valores limite e transformada em K valores indicadores $i(u_\alpha; z_k)$, então, na verdade, a esperança matemática da VA $I(u; z_k)$ da Equação 3.43 é condicionada aos valores indicadores locais $i(u_\alpha; z_k)$:

$$\text{Pr ob}\{Z(u) \leq z_k | (n)\} = E[I(u; z_k) | I(u_\alpha; z_k) = i(u_\alpha; z_k), \alpha = 1, \dots, n \text{ e } k = 1, \dots, K] \tag{3.44}$$

Entretanto, na prática, somente as VAs indicadoras $I(u_\alpha; z_k)$ do valor limite z_k corrente (analisado no momento) são utilizadas como dados condicionantes. Dessa maneira, a equação passa a ser escrita de acordo com a seguinte aproximação:

$$\text{Pr ob}\{Z(u) \leq z_k \mid (n)\} \approx E[I(u; z_k) \mid I(u_\alpha; z_k) = i(u_\alpha; z_k), \quad \alpha = 1, \dots, n] \quad (3.45)$$

Portanto, essa aproximação reduz o número de dados indicadores condicionantes de $n * K$ para n . De certa maneira, essa aproximação abdica das correlações cruzadas entre as VAs indicadoras $I(u_\alpha; z_k)$ definidas para diferentes valores limite z_k .

A esperança condicional da VA indicadora $I(u; z_k)$ da equação acima pode ser escrita em termos de uma série expandida (Bahadur, 1961, Journal e Alabert, 1989):

$$\begin{aligned} E[I(u; z_k) \mid I(u_\alpha; z_k) = i(u_\alpha; z_k), \quad \alpha = 1, \dots, n] = & a_0 + \sum_{\alpha=1}^n a_1(u_\alpha) \cdot i(u_\alpha, z_k) \\ & + \sum_{\alpha=1}^n \sum_{\alpha' \neq \alpha}^n a_2(u_\alpha, u_{\alpha'}) \cdot i(u_\alpha, z_k) \cdot i(u_{\alpha'}, z_k) \\ & + \dots + a_n \prod_{\alpha=1}^n i(u_\alpha, z_k) \end{aligned} \quad (3.46)$$

Uma aproximação de primeira ordem da esperança condicional da VA indicadora $I(u; z_k)$ da Equação 3.45 pode ser obtida considerando os primeiros $n+1$ termos:

$$\text{Pr ob}\{Z(u) \leq z_k \mid (n)\} \approx a_0 + \sum_{\alpha=1}^n a_1(u_\alpha) \cdot i(u_\alpha, z_k) \quad (3.47)$$

O conjunto de $n+1$ coeficientes (a_0, \dots, a_n) da aproximação de 1ª ordem da dclp da VA $Z(u)$ são determinadas por meio de uma múltipla regressão linear dos dados indicadores desconhecidos $i(u; z_k)$ (nas localizações não amostradas) sobre os n dados indicadores $i(u_\alpha; z_k)$ experimentais. De acordo com o teorema da projeção (Luenberger, 1969, p. 49), a estimativa por mínimos quadrados (krigagem) dos indicadores $i(u; z_k)$ é igual a estimativa por mínimos quadrados de sua esperança condicional (Goovaerts, 1997, p. 284). Dessa forma, as probabilidades condicionais que constituem a dclp $F(u; z \mid (n))$ podem ser determinadas por (co)krigagem dos indicadores em qualquer localização u não medida.

3.6.1. Codificação binária das VA

Cada informação conhecida $z(u_\alpha)$ ($\alpha = 1, \dots, n$) é confrontada com os K valores limite z_k . Assim, para cada $z(u_\alpha)$ existe um conjunto de K valores binários que são (co)krigados para a determinação de cada probabilidade referente à z_k em qualquer localização u desconhecida. O conjunto de K informações binárias pode ser representado por uma matriz coluna (vetor) de dimensão igual a K .

$$I(u_\alpha; z_k) = \begin{bmatrix} I(u_\alpha; z_1) \\ I(u_\alpha; z_2) \\ \vdots \\ I(u_\alpha; z_{K-1}) \\ I(u_\alpha; z_K) \end{bmatrix}, \text{ com } k = 1, \dots, K, \text{ e } \alpha = 1, \dots, N$$

onde:

$$I(u_\alpha; z_k) = i(u_\alpha; z_k) = 1_{z(u_\alpha) \leq z_k} = \begin{cases} 1, & \text{se sim} \\ 0, & \text{se não} \end{cases}, \text{ com } k = 1, \dots, K$$

A codificação dos n valores assumidos pelas VAs $Z(u_\alpha)$ pode ser esquematizada em um quadro (Quadro 3.1) que relaciona as n de informações e os K valores limites z_k . Por exemplo, em uma situação onde foram pré-estabelecidos cinco valores limite z_k ($K = 5$) e existem quatro informações:

$$\begin{aligned} z_k &=> z_1 = 0\%, z_2 = 7\%, z_3 = 10\%, z_4 = 12\% \text{ e } z_5 = 100\% \\ z(u_\alpha) &=> z(u_1) = 11\%, z(u_2) = 3\%, z(u_3) = 10\% \text{ e } z(u_4) = 17\% \end{aligned}$$

Quadro 3.1: Relação do valor dos indicadores segundo cada valor limite z_k , para as quatro informações $z(u_\alpha)$.

$\alpha \backslash k$	z_1	z_2	z_3	z_4	z_5
$z(u_1)$	0	0	0	1	1
$z(u_2)$	0	1	1	1	1
$z(u_3)$	0	0	1	1	1
$z(u_4)$	0	0	0	0	1
Soma	0	1	2	3	4
Frequência	0	1/4	1/2	3/4	1

Nesse quadro, a razão entre o somatório dos elementos de cada coluna e o número n de informações resulta na proporção de valores $z(u_\alpha)$ ($\alpha = 1, \dots, n$) abaixo do valor limite z_k . Esse conjunto de frequências representa cinco pontos discretos da função de repartição empírica $\hat{F}(z)$ (dependente do número n de medidas disponíveis) desde que

$$\hat{F}(z) = \frac{1}{n} \sum_{\alpha=1}^n 1_{z(u_\alpha) < z} \tag{3.48}$$

O primeiro passo no processo de codificação binária das VAs $Z(u_\alpha)$ é a escolha do número K de valores limite z_k . Segundo Goovaerts (1997, p. 285), o número K, raramente deve ser maior do que 15. Por outro lado, o mesmo autor sugere que K não deve ser menor do que cinco, para garantir que o modelo de incerteza local seja suficientemente discretizado.

Após a escolha de K, parte-se para a escolha dos valores z_k . Geralmente, esses valores são escolhidos de modo que o intervalo de variação de $z(u_\alpha)$ ($\alpha = 1, \dots, n$) seja dividido em (K+1) classes com frequências aproximadamente iguais. Contudo, se existir algum teor crítico z_c (ou teor de corte)

que deva ser considerado nas etapas seguintes do estudo, recomenda-se que algum valor limite z_k coincida com z_c .

Da mesma forma, é conveniente que as classes da distribuição experimental que despertam maior interesse sejam mais discretizadas. Entretanto, não é recomendado discretizar a distribuição antes e depois do primeiro e nono decil, respectivamente. Isso porque, o cálculo dos semi-variogramas experimentais pode se tornar muito difícil em virtude da disponibilidade de um número reduzido de pares. Dependendo da quantidade de informações, 10% do número total pode ser insuficiente para o cálculo dos semi-variogramas. Em alguns casos, por insuficiência de dados, os semi-variogramas ligados a valores limite extremos precisam ser inferidos.

A partir da definição dos valores limite z_k , a próxima etapa consiste no cálculo dos semi-variogramas experimentais dos indicadores para cada z_k . Ou seja, são necessários K modelos de continuidade espacial dos indicadores para a estimativa de $I(u; z_k)$ por krigagem e K^2 para a mesma estimativa por cokrigagem. Como descrito anteriormente, o cálculo de $[I(u; z_k)]_{(co)krigagem}^*$ garante o conhecimento de $[F(u; z_k | (n))]^*$, pois

$$[F(u; z_k | (n))]^* = [I(u; z_k)]_{(co)krigagem}^* = [E[I(u; z_k) | (n)]]^* \quad k = 1, \dots, K \quad (3.49)$$

Van Meirvenne e Goovaerts (2001) propuseram um procedimento alternativo de codificação que permite que os erros analíticos (erros de medida de laboratório) e de amostragem sejam considerados. Essa codificação é baseada principalmente na consideração do coeficiente de variação dos resultados do processo analítico (obtidos por repetibilidade). A metodologia também se baseia na substituição dos valores experimentais $z(u_\alpha)$ por uma distribuição Gaussiana centrada em $z(u_\alpha)$ (assumindo a hipótese de não tendenciosidade) e com um desvio padrão $\sigma(u_\alpha)$ calculado da seguinte forma:

$$\sigma(u_\alpha) = CV \cdot z(u_\alpha) \quad \alpha = 1, \dots, n \quad (3.50)$$

onde CV é coeficiente de variação do processo analítico. Assim, esse procedimento de codificação alternativo é realizado por meio da seguinte equação:

$$i(u_\alpha; z_k) = G[(z_k - z(u_\alpha)) / \sigma(u_\alpha)] \quad k = 1, \dots, K \quad (3.51)$$

onde $G[.]$ é distribuição cumulativa de probabilidade Gaussiana padrão. A codificação apresentada acima, deixa de mostrar uma característica binária, uma vez que esse procedimento produz VAs indicadoras com valores dentro do intervalo definido por $[0;1]$. Essa codificação é conhecida como *soft indicators* (Journel, 1986a).

3.6.2. Krigagem dos Indicadores

A probabilidade estimada $[F(u; z_k | (n))]^*$, associada aos valores limite z_k , é determinada por meio de uma combinação linear entre os elementos indicadores da circunvizinhança de cada localização u alvo. Os ponderadores dos indicadores retidos nessa vizinhança centrada em u podem

ser determinados por krigagem. Assim, a estimativa do modelo de incerteza local (dclp) nas localizações não medidas u é composto pelo resultado de K combinações lineares. Os algoritmos de krigagem mais comumente usados são o de krigagem simples dos indicadores (KSI) e krigagem ordinária dos indicadores (KOI). Como será visto, a estimativa de probabilidade $[F(u; z_k | (n))]^*$ é o resultado da combinação linear de $n(u)$ e $n(u) + 1$ elementos, quando aplicado o algoritmo de KOI e KSI, respectivamente.

3.6.2.1. Krigagem Ordinária dos Indicadores

Como para o caso de krigagem ordinária, o algoritmo de KOI é aplicado quando a média m [nesse caso m_k ou $F(z_k)$] é desconhecida. Esse estimador é definido como:

$$[I(u; z_k)]_{KOI}^* = \sum_{\alpha=1}^{n(u)} \lambda_{\alpha}(u; z_k) \times I(u_{\alpha}; z_k) \quad k = 1, \dots, K. \quad (3.52)$$

Os pesos λ_{α} atribuídos aos valores binários das VAs indicadora $I(u_{\alpha}; z_k)$ circunvizinhos de u são determinados pela resolução do seguinte sistema de krigagem:

$$\begin{cases} \sum_{\beta=1}^{n(u)} \lambda_{\beta}(u; z_k) \cdot C_I(u_{\alpha} - u_{\beta}; z_k) + \mu_{KO}(u; z_k) = C_I(u_{\alpha} - u; z_k) & \alpha = 1, \dots, n(u) \\ \sum_{\beta=1}^{n(u)} \lambda_{\beta}(u; z_k) = 1 \end{cases} \quad (3.53)$$

onde $C_I(h; z_k)$ é a função de covariância da VA $I(u; z_k)$, referente ao valor limite z_k , e μ_{KO} é o Lagrangeano. No caso de KOI, a estimativa da dclp é determinada por meio de K combinações lineares de $n(u)$ elementos.

Os algoritmos de KO em geral (dentre eles KOI) permitem que as flutuações locais em torno da média da variável alvo sejam consideradas. Isso é realizado por meio da limitação do domínio de estacionaridade, de acordo com a circunvizinhança de cada localização u a ser estimada, de modo que a média global (desconhecida) seja filtrada. Por esse motivo, os sistemas de KO (em geral) também podem ser expressos em termos da função semi-variograma $\gamma(h)$ (no caso de KOI: $\gamma_I(h; z_k)$). Isso só é possível porque o erro da estimativa (ϵ) é uma combinação linear autorizada (CLA), o que permite que a variância do erro $\text{Var}(\epsilon)$ seja expressa em termos de $C(h)$ ou de $\gamma(h)$. Dessa forma, as equações de KO se aplicam a variáveis estacionárias e intrínsecas – Capítulo 2. Como será visto a seguir, o mesmo não é válido para o algoritmo de KSI (KS, em geral), que somente pode ser expresso termos de covariâncias. Entretanto, por razões de eficiência computacional, os sistemas de krigagem (em geral) são resolvidos em termos de covariâncias ou pseudocovariâncias (quando o modelo variográfico empregado não apresentar patamar – sill) (Goovaerts, 1997, p. 135).

3.6.2.2. Krigagem Simples dos Indicadores

Ao contrário do algoritmo de KOI, o de KSI incorpora a média global dos indicadores, segundo cada valor limite z_k :

$$[I(u; z_k)]_{KSI}^* = \sum_{\alpha=1}^{n(u)} \lambda_{\alpha}(u; z_k) \times I(u_{\alpha}; z_k) + \left[1 - \sum_{\alpha=1}^{n(u)} \lambda_{\alpha}(u; z_k) \right] F(z_k) \quad k = 1, \dots, K. \quad (3.54)$$

onde $E[I(u; z_k)] = F(z_k)$. Ou seja, $F(z_k)$ é simplesmente a proporção de informações inferiores ou iguais a z_k (isso no caso em que a função de repartição empírica $\hat{F}(z)$ for suficientemente representada para permitir que $\hat{m} \approx E[I(u; z_k)]$). No Quadro 3.1, se $\hat{F}(z)$ fosse representativa, $F(z_k)$ poderia ser lido diretamente como sendo a freqüência referente à z_k . Portanto, a estimativa da dclp por meio do algoritmo de KSI é determinada pelo resultado de K combinações lineares de $n(u)+1$ elementos.

No algoritmo de KSI, os pesos λ_{α} são calculados a partir da resolução do seguinte sistema de krigagem:

$$\sum_{\beta=1}^{n(u)} \lambda_{\beta}(u; z_k) \cdot C_I(u_{\alpha} - u_{\beta}; z_k) = C_I(u_{\alpha} - u; z_k) \quad \alpha = 1, \dots, n(u) \quad (3.55)$$

Nesse caso, o erro da estimativa ε não é uma CLA e por isso a variância do erro $\text{Var}(\varepsilon)$, que dá origem às equações de krigagem, só pode ser escrita em termos de covariâncias $C(h)$. Dessa maneira, o sistema de KS (em geral) é exclusivamente dirigido às variáveis estacionárias – Capítulo 2.

3.6.3. Cokrigagem dos Indicadores

Além dos estimadores convencionais de krigagem, ainda há a possibilidade de cokrigar os indicadores (CKI) e utilizar outros algoritmos, como krigagem de probabilidade (KP) ou krigagem da mediana (KM). Os algoritmos de cokrigagem podem considerar todas as informações disponíveis entre os K valores limite z_k . Dessa maneira, o conhecimento de que uma determinada informação $z(u_{\alpha})$ é inferior a um dado z_k e superior a $z_{k'}$ (informações dependentes) pode melhorar a coerência entre as estimativas. É possível cokrigar essas informações utilizando os algoritmos de cokrigagem simples dos indicadores (CKSI) ou o de cokrigagem ordinária dos indicadores (CKOI).

3.6.3.1. Cokrigagem Ordinária dos Indicadores

A estimativa da probabilidade $[F(u; z_k | (n))]^*$ é calculada por meio de CKOI da seguinte forma:

$$\begin{aligned} [F(u; z_{k_0} | (n))]_{CKOI}^* &= [I(u; z_{k_0})]_{CKOI}^* \\ &= \sum_{k=1}^K \sum_{\alpha=1}^{n(u)} \lambda_{\alpha k}(u; z_{k_0}) \cdot I(u_{\alpha}; z_k) \end{aligned} \quad (3.56)$$

onde os ponderadores $\lambda_{\alpha k}$ são os pesos atribuídos às informações indicadoras de cada localização $[I(u_{\alpha}; z_k)]$ e k_0 é a faixa indicadora alvo. Nesse algoritmo, a estimativa da probabilidade associada aos valores limite z_k (em cada localização desconhecida) é o resultado da combinação linear de $K \cdot n(u)$

elementos. Assim, a estimativa da dclp por meio de CKOI é determinada pelo resultado de K combinações lineares de $K \cdot n(u)$ elementos.

O valor limite z_{k0} representa a faixa indicadora da probabilidade a ser estimada na localização desconhecida u . Os ponderadores $\lambda_{\alpha k}$ são dependentes da localização u e do valor limite z_{k0} – faixa indicadora alvo. Como mostrado no sistema cokrigagem (3.57), em virtude dessa dependência, há uma diferença em relação às condições impostas aos ponderadores das demais faixas indicadoras. O sistema de CKOI para o cálculo dos ponderadores $\lambda_{\alpha k}$, composto por $(K \cdot n(u)) + K$ equações, é dado abaixo:

$$\left\{ \begin{array}{l} \sum_{k'=1}^K \sum_{\beta=1}^{n(u)} \lambda_{\beta k'}(u; z_{k_0}) \cdot \gamma_I(u_{\alpha}, u_{\beta}; z_k, z_{k'}) + \mu_k = \gamma_I(u, u_{\alpha}; z_k, z_{k_0}), \quad \alpha = 1, \dots, n(u) \\ \sum_{\alpha=1}^{n(u)} \lambda_{\alpha k}(u; z_{k_0}) = \delta \quad \quad \quad k = 1, \dots, K \end{array} \right. \quad (3.57)$$

onde $\gamma_I(h; z_k, z_{k'})$ é a função semi-variograma (direta ou cruzada) entre as VAs indicadoras $I(u_{\alpha}; z_k)$, referente aos valores limite z_k e $z_{k'}$, separados por uma distância representada pelo vetor h . Na segunda equação, o valor de δ é 0, exceto quando $k = k_0$, onde $\delta=1$. Ou seja, o somatório dos pesos atribuídos às VAs $I(u_{\alpha}; z_k)$ associadas à faixa indicadora na qual se deseja estimar a probabilidade deve ser igual a um. Enquanto que o somatório dos outros pesos deve ser nulo. Essas condições são atribuídas para que a média da VA a ser estimada seja filtrada. Isso porque, como em qualquer estimador de KO, as esperanças matemáticas das variáveis alvo são consideradas desconhecidas.

Como discutido anteriormente, os sistemas de krigagem ordinária (em geral) podem ser expressos em termos de semi-variogramas. O mesmo pode ser válido para os sistemas de cokrigagem ordinária (CKOI) se for assumido que $C_{ij}(h) = C_{ji}(h)$. Na realidade, essa igualdade não é verdadeira, pois:

$$C_{ij}(h) = C_{ji}(-h) \quad ,$$

Assim como:

$$C_{ij}(h) \neq C_{ij}(-h)$$

Essa característica é chamada de "efeito de retardo". Contudo, na prática essa assimetria é freqüentemente ignorada pelas seguintes razões (Goovaerts, 1997, p. 73):

- i. as ferramentas descritivas usuais de correlação espacial são os semi-variogramas diretos e cruzados, que são simétricos;
- ii. o "efeito de retardo" reflete flutuações experimentais que podem ser o resultado da disponibilidade de um número reduzido de pares de informações;
- iii. a modelagem de covariâncias cruzadas assimétricas é bastante difícil (Journel *et al.*, 1978, p. 173).

Portanto, ao assumir que $C_{ij}(h) = C_{ji}(h)$ o número de covariogramas, diretos e cruzados, necessários para a resolução dos sistemas cokrigagem é reduzido de K^2 para $K \cdot (K+1)/2$. Mesmo

assim, cokrigar as VAs indicadoras $I(u_\alpha; z_k)$ pode ser uma tarefa muito trabalhosa. Por exemplo, para o caso apresentado acima (codificação de $Z(u)$), onde existem cinco valores limite z_k ($K = 5$) e quatro informações disponíveis $z(u_\alpha)$ ($\alpha = 1, \dots, 4$), a estimativa da dclp em uma localização u na vizinhança dessas informações, por meio de CKOI, seria realizada da seguinte forma:

$$[F(u; z_{k_0} | (n))]_{CKOI}^* = \sum_{k=1}^5 \sum_{\alpha=1}^4 \lambda_{\alpha k}(u; z_{k_0}) \cdot I(u_\alpha; z_k) \quad k_0 = 1, \dots, 5 \quad (3.58)$$

Isto é, a dclp em u seria composta pelo resultado de cinco combinações lineares de 20 elementos. Sendo que os pesos $\lambda_{\alpha k}$ atribuídos aos valores assumidos pelas VAs $I(u_\alpha; z_k)$ seriam determinados por meio da resolução de cinco sistemas de cokrigagem compostos por 25 equações. Por exemplo, para a estimativa de $[I(u; z_3)]^*$ em função das VAs indicadoras $[I(u_\alpha; z_k)]$ ($k = 1, \dots, 5$ e $\alpha = 1, \dots, 4$):

$$\begin{cases} \sum_{\beta=1}^{n(u)} \lambda_{\beta 1} \gamma_1(u_\alpha, u_\beta; z_1, z_1) + \sum_{\beta=1}^{n(u)} \lambda_{\beta 2} \gamma_1(u_\alpha, u_\beta; z_1, z_2) + \sum_{\beta=1}^{n(u)} \lambda_{\beta 3} \gamma_1(u_\alpha, u_\beta; z_1, z_3) + \sum_{\beta=1}^{n(u)} \lambda_{\beta 4} \gamma_1(u_\alpha, u_\beta; z_1, z_4) + \sum_{\beta=1}^{n(u)} \lambda_{\beta 5} \gamma_1(u_\alpha, u_\beta; z_1, z_5) + \mu_1 = \gamma_1(u, u_\alpha; z_1, z_3) \\ \sum_{\beta=1}^{n(u)} \lambda_{\beta 1} \gamma_1(u_\alpha, u_\beta; z_2, z_1) + \sum_{\beta=1}^{n(u)} \lambda_{\beta 2} \gamma_1(u_\alpha, u_\beta; z_2, z_2) + \sum_{\beta=1}^{n(u)} \lambda_{\beta 3} \gamma_1(u_\alpha, u_\beta; z_2, z_3) + \sum_{\beta=1}^{n(u)} \lambda_{\beta 4} \gamma_1(u_\alpha, u_\beta; z_2, z_4) + \sum_{\beta=1}^{n(u)} \lambda_{\beta 5} \gamma_1(u_\alpha, u_\beta; z_2, z_5) + \mu_2 = \gamma_1(u, u_\alpha; z_2, z_3) \\ \sum_{\beta=1}^{n(u)} \lambda_{\beta 1} \gamma_1(u_\alpha, u_\beta; z_3, z_1) + \sum_{\beta=1}^{n(u)} \lambda_{\beta 2} \gamma_1(u_\alpha, u_\beta; z_3, z_2) + \sum_{\beta=1}^{n(u)} \lambda_{\beta 3} \gamma_1(u_\alpha, u_\beta; z_3, z_3) + \sum_{\beta=1}^{n(u)} \lambda_{\beta 4} \gamma_1(u_\alpha, u_\beta; z_3, z_4) + \sum_{\beta=1}^{n(u)} \lambda_{\beta 5} \gamma_1(u_\alpha, u_\beta; z_3, z_5) + \mu_3 = \gamma_1(u, u_\alpha; z_3, z_3) \\ \sum_{\beta=1}^{n(u)} \lambda_{\beta 1} \gamma_1(u_\alpha, u_\beta; z_4, z_1) + \sum_{\beta=1}^{n(u)} \lambda_{\beta 2} \gamma_1(u_\alpha, u_\beta; z_4, z_2) + \sum_{\beta=1}^{n(u)} \lambda_{\beta 3} \gamma_1(u_\alpha, u_\beta; z_4, z_3) + \sum_{\beta=1}^{n(u)} \lambda_{\beta 4} \gamma_1(u_\alpha, u_\beta; z_4, z_4) + \sum_{\beta=1}^{n(u)} \lambda_{\beta 5} \gamma_1(u_\alpha, u_\beta; z_4, z_5) + \mu_4 = \gamma_1(u, u_\alpha; z_4, z_3) \\ \sum_{\beta=1}^{n(u)} \lambda_{\beta 1} \gamma_1(u_\alpha, u_\beta; z_5, z_1) + \sum_{\beta=1}^{n(u)} \lambda_{\beta 2} \gamma_1(u_\alpha, u_\beta; z_5, z_2) + \sum_{\beta=1}^{n(u)} \lambda_{\beta 3} \gamma_1(u_\alpha, u_\beta; z_5, z_3) + \sum_{\beta=1}^{n(u)} \lambda_{\beta 4} \gamma_1(u_\alpha, u_\beta; z_5, z_4) + \sum_{\beta=1}^{n(u)} \lambda_{\beta 5} \gamma_1(u_\alpha, u_\beta; z_5, z_5) + \mu_5 = \gamma_1(u, u_\alpha; z_5, z_3) \\ \sum_{\alpha=1}^{n(u)} \lambda_{\alpha 1}(u; z_3) = 0 \\ \sum_{\alpha=1}^{n(u)} \lambda_{\alpha 2}(u; z_3) = 0 \\ \sum_{\alpha=1}^{n(u)} \lambda_{\alpha 3}(u; z_3) = 1 \\ \sum_{\alpha=1}^{n(u)} \lambda_{\alpha 4}(u; z_3) = 0 \\ \sum_{\alpha=1}^{n(u)} \lambda_{\alpha 5}(u; z_3) = 0 \end{cases}$$

Matricialmente (Myers, 1982):

$$\begin{array}{cccccccccc|c|c}
 \underline{\gamma_{11}} & \underline{\gamma_{12}} & \underline{\gamma_{13}} & \underline{\gamma_{14}} & \underline{\gamma_{15}} & 1 & 0 & 0 & 0 & 0 & \underline{\lambda_1} & \underline{\gamma_{01}} \\
 \underline{\gamma_{21}} & \underline{\gamma_{22}} & \underline{\gamma_{23}} & \underline{\gamma_{24}} & \underline{\gamma_{25}} & 0 & 1 & 0 & 0 & 0 & \underline{\lambda_2} & \underline{\gamma_{02}} \\
 \underline{\gamma_{31}} & \underline{\gamma_{32}} & \underline{\gamma_{33}} & \underline{\gamma_{34}} & \underline{\gamma_{35}} & 0 & 0 & 1 & 0 & 0 & \underline{\lambda_3} & \underline{\gamma_{03}} \\
 \underline{\gamma_{41}} & \underline{\gamma_{42}} & \underline{\gamma_{43}} & \underline{\gamma_{44}} & \underline{\gamma_{45}} & 0 & 0 & 0 & 1 & 0 & \underline{\lambda_4} & \underline{\gamma_{04}} \\
 \underline{\gamma_{51}} & \underline{\gamma_{52}} & \underline{\gamma_{53}} & \underline{\gamma_{54}} & \underline{\gamma_{55}} & 0 & 0 & 0 & 0 & 1 & \underline{\lambda_5} & \underline{\gamma_{05}} \\
 \hline
 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu_1 & 0 \\
 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu_2 & 0 \\
 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu_3 & 1 \\
 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \mu_4 & 0 \\
 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & \mu_5 & 0
 \end{array}
 \times =$$

Evidentemente, para a resolução desse sistema de cokrigagem seriam necessários o cálculo e modelagem de 15 semi-variogramas, entre diretos e cruzados.

3.6.3.2. Cokrigagem Simples dos Indicadores

Assim, como para o algoritmo de KSI (ou KS em geral), a aplicação do algoritmo de CKSI exige que a distribuição das VAs indicadoras $I(u_\alpha; z_k)$ seja suficientemente representativa para permitir que $\hat{m}_k \approx E[I(u; z_k)]$. Isso porque, o estimador considera o peso das médias m_k , para o cálculo de $[I(u; z_{k_0})]_{CKSI}^*$. Ou seja:

$$[F(u; z_{k_0} | (n))]_{CKSI}^* = \sum_{k=1}^K \sum_{\alpha=1}^{n(u)} \lambda_{\alpha k} (u; z_{k_0}) \cdot I(u_\alpha; z_k) + \lambda_0 \tag{3.59}$$

$$e \lambda_0 = m_{k_0} - \sum_{k=1}^K \sum_{\alpha=1}^{n(u)} \lambda_{\alpha k} \cdot m_k$$

onde $m_{k_0} = F(z_{k_0})$.

Assim, para cada faixa indicadora, os pesos $\lambda_{\alpha k}(u; z_{k_0})$ são determinados a partir de um sistema de cokrigagem da seguinte forma:

$$\sum_{k'=1}^K \sum_{\beta=1}^{n(u)} \lambda_{\beta k} (u; z_{k_0}) \cdot C_I(u_\alpha, u_\beta; z_k, z_{k'}) = C_I(u, u_\alpha; z_k, z_{k_0}) \tag{3.60}$$

$\alpha = 1, \dots, n(u)$
 $k = 1, \dots, K$

Em ambos algoritmos de cokrigagem (CKOI e CKSI), a faixa indicadora da probabilidade a ser estimada z_{k_0} define a relação de covariâncias a ser empregada entre as localizações conhecidas e a localização incógnita. Isto é, essas matrizes de covariâncias variam de acordo com a estimativa de

probabilidade de cada faixa indicadora z_{k0} . Ao contrário, a matriz de covariâncias entre as localizações medidas é independente do valor limite z_{k0} e permanece a mesma para a estimativa de probabilidade de todos os K valores limite z_k .

A desvantagem de CKOI em relação a CKSI é que a influência das variáveis secundárias tende a ser minimizada pelo fato do somatório dos pesos dados a essas variáveis ser nulo. Essa condição também aumenta a chance de ocorrência de pesos negativos e, por conseqüência, de estimativas inaceitáveis, como probabilidades $[I(u; z_{k_0})]_{CKOI}^*$ negativas ou maiores a 100 % - problemas de relação de ordem (seção 3.6.5).

3.6.3.3. Krigagem da probabilidade

Um caso particular do estimador de CKOI é o algoritmo de krigagem da probabilidade (KP) (Sullivan, 1984). Esse algoritmo utiliza as VAs indicadoras $I(u_\alpha; z_k)$ ($\alpha = 1, \dots, n$) (para um dado valor limite z_k) e as informações originais $z(u_\alpha)$, constituindo um sistema de cokrigagem que envolve somente duas variáveis. Isto é, a cokrigagem das VAs indicadoras $I(u_\alpha; z_k)$ usando as informações originais $z(u_\alpha)$ como variável secundária. Entretanto, se a diferença de magnitude entre as VAs indicadoras (binária) e as informações originais for muito elevada, podem ocorrer problemas de instabilidade na resolução do sistema de cokrigagem, especialmente na inversão da matriz de covariâncias entre as localizações medidas. Para contornar esse problema, uma prática comum consiste em transformar as informações originais em uma distribuição situada dentro do domínio definido por $[0;1]$. Essa transformação pode ser realizada por meio da razão entre os respectivos valores de ordenamento $o(u_\alpha)$ (a partir da distribuição cumulativa original) e o número de informações originais, ou seja:

$$t(u_\alpha) = o_q(u_\alpha)/n \quad q = 1, \dots, n \quad (3.61)$$

onde, $o_{q-1}(u_\alpha) < o_q(u_\alpha)$ e $o_q(u_\alpha) \in [1;n]$.

Assim, o estimador de KP é escrito da seguinte forma:

$$\begin{aligned} [F(u; z_k | (n))]_{KP}^* &= [I(u; z_k)]_{KP}^* \\ &= \sum_{\alpha=1}^{n(u)} \lambda_\alpha(u; z_k) \cdot I(u_\alpha; z_k) + \sum_{\alpha=1}^{n(u)} v_\alpha(u; z_k) \cdot t(u_\alpha) \end{aligned} \quad k = 1, \dots, K \quad (3.62)$$

Os ponderadores $\lambda_\alpha(u_\alpha; z_k)$ e $v_\alpha(u_\alpha; z_k)$ são determinados por meio da resolução do seguinte sistema de cokrigagem ordinária, composto por $2 \cdot n(u) + 2$ equações:

$$\begin{cases} \sum_{\beta=1}^{n(u)} \lambda_\beta(u; z_k) \cdot C_I(u_\alpha, u_\beta; z_k) + \sum_{\beta=1}^{n(u)} v_\beta(u; z_k) \cdot C_{It}(u_\alpha, u_\beta; z_k) + \mu_I(u; z_k) = C_I(u, u_\alpha; z_k) \\ \sum_{\beta=1}^{n(u)} \lambda_\beta(u; z_k) \cdot C_{It}(u_\alpha, u_\beta; z_k) + \sum_{\beta=1}^{n(u)} v_\beta(u; z_k) \cdot C_t(u_\alpha, u_\beta) + \mu_t(u; z_k) = C_{It}(u, u_\alpha; z_k) \\ \sum_{\alpha=1}^{n(u)} \lambda_\alpha(u; z_k) = 1 \\ \sum_{\alpha=1}^{n(u)} v_\alpha(u; z_k) = 0 \end{cases} \quad (3.63)$$

$\alpha = 1, \dots, n(u)$

onde:

$C_I(h; z_k)$ é covariância espacial dos indicadores;

$C_t(h; z_k)$ é covariância espacial dos dados transformados e

$C_{It}(h; z_k)$ é covariância espacial cruzada entre os indicadores e dos dados transformados.

Assim, a estimativa da dclp em cada localização não medida u requer a resolução K de sistemas de CKO (de acordo com o apresentado acima). Para isso, o algoritmo de KP exige o cálculo e a modelagem de $2 \cdot K + 1$ semi-variogramas, entre diretos e cruzados (assumindo que $C_{ij}(h) = C_{ji}(h)$, como discutido anteriormente).

3.6.4. KI X CKI

Teoricamente, os estimadores de cokrigagem dos indicadores são melhores do que os de krigagem dos indicadores, pelo fato de considerarem todas as K informações disponíveis em cada localização medida. Por esse motivo, a variância do erro da estimativa por cokrigagem sempre é inferior ou igual à variância do erro por krigagem, ou seja:

$$[\text{Var}(Z^*-Z)]_{CK} \leq [\text{Var}(Z^*-Z)]_K$$

Porém, a prática tem mostrado que o uso de cokrigagem dos indicadores melhora apenas um pouco os resultados. Segundo Goovaerts (1997, p. 298) isso é principalmente devido a três razões:

- i. O resultado da codificação binária já incorpora informações significativas entre os K valores limite z_k .
- ii. Em uma condição isotópica (como é o caso de CKI), para CKOI, a estimativa da probabilidade do valor limite z_{k0} tende a inibir a influência dos demais valores limite z_k .
- iii. CKI gera mais problemas de relação de ordem do que os algoritmos de KI. Contudo, de acordo com o problema em mãos, esse ganho diferencial na qualidade das estimativas pode ser significativo, justificando o emprego de CKI. Porém, como mostrado anteriormente, os algoritmos de CKI são mais complexos do que os de KI, pois requerem a inversão de grandes matrizes, além de exigir o cálculo e o ajuste (simultâneo) de $K \cdot (K+1)/2$ semi-variogramas (entre diretos e cruzados). Sendo que a determinação do modelo linear de correção regionalização (MLC) pode ser muito trabalhosa.

3.6.4.1. Modelo Linear de Correção Regionalização

No modelo linear de correção regionalização (MLC), as estruturas diretas e cruzadas das diferentes VAs são obtidas linearmente a partir das mesmas estruturas de base, sendo definido como:

$$F_{ij}(h) = \sum_s B^s \cdot F^s(h) \quad (3.64)$$

onde B^s são as matrizes de correção regionalização, constituídas pelas variâncias e covariâncias b_{ij}^s , definidas para cada estrutura s de função $F(h)$. A função $F(h)$ pode ser representada pela:

- i. função semi-variograma $\gamma(h)$, $\gamma_{ij}(h) = \sum_s B^s \cdot \gamma^s(h)$;
- ii. função covariância $C(h)$, $C_{ij}(h) = \sum_s B^s \cdot C^s(h)$;
- iii. função correlograma $\rho(h)$, $\rho_{ij}(h) = \sum_s B^s \cdot \rho^s(h)$;

Logo, o MLC é totalmente caracterizado pelas estruturas de base [$\gamma(h)$, $C(h)$ ou $\rho(h)$] e pelas matrizes de correionalização (B^s). Contudo, para que o MLC seja autorizado é necessário que cada matriz B^s seja definida positivamente (determinante ≥ 0). Em particular, a correlação (definida pelo coeficiente de correlação estrutural, $\rho^s = \frac{b_{ij}}{\sqrt{b_{ii} \cdot b_{jj}}}$) entre duas variáveis deve estar definida no

intervalo $[-1;+1]$; condição que é garantida se $|b_{ij}^s| \leq \sqrt{b_{ii}^s \cdot b_{jj}^s}$. Sendo que, o semi-variograma cruzado entre pares de VAs deve se situar dentro do intervalo definido por $-\sum_s \sqrt{b_{ii}^s \cdot b_{jj}^s} \cdot \gamma^s(h)$ e $+\sum_s \sqrt{b_{ii}^s \cdot b_{jj}^s} \cdot \gamma^s(h)$.

Entretanto, por definição, cada matriz B^s é automaticamente definida positivamente se as VAs envolvidas [$Z_i(u)$, $i = 1, \dots, N$] permitirem serem decompostas em VAs $Z_i^s(u)$ espacialmente independentes:

$$Z_i(u) = m_i + \sum_s Z_i^s(u) \quad (3.65)$$

[onde m_i são as médias ($m_i = E[Z_i(u)]$) das N VAs $Z_i(u)$], de modo que os componentes $Z_i^s(u)$ possam ser decompostos como: $Z_i^s(u) = \sum_q a_{iq}^s Y_q^s(u)$. Sendo que, $Y_q^s(u)$ são FAs independentes, de média nula

($E[Y_q^s(u)] = 0$, $\forall q, s$) caracterizadas espacialmente por $\gamma^s(h)$. Os elementos $Y_q^s(u)$ são chamados de fatores regionalizados. Então, o MLC é definido como:

$$Z_i(u) = m_i + \sum_s \sum_q a_{iq}^s Y_q^s(u) \quad (3.66)$$

onde os elementos a_{iq}^s , da matriz A , são os coeficientes dos fatores regionalizados $Y_q^s(u)$. Assim, a partir de (3.66), o semi-variograma de $Z_i(u)$ é definido como:

$$\gamma_i(h) = \sum_s \sum_q (a_{iq}^s)^2 \cdot \gamma^s(h) = \sum_s b_{ii}^s \cdot \gamma^s(h) \quad (3.67)$$

ou seja: $b_{ii}^s = \sum_q (a_{iq}^s)^2$. Portanto, o MLC interpreta cada VA $Z_i(u)$ como sendo uma combinação linear das FAs $Y_q^s(u)$ espacialmente independentes, sendo que as matrizes de correionalização B^s são derivadas a partir dos coeficientes a_{iq}^s .

De maneira análoga, o semi-variograma cruzado é definido como:

$$\gamma_{ij}(h) = \sum_s \sum_q a_{iq}^s a_{jq}^s \cdot \gamma^s(h) = \sum_s b_{ij}^s \cdot \gamma^s(h) \quad (3.68)$$

de modo que $b_{ij}^s = \sum_q a_{iq}^s a_{jq}^s$.

A decomposição das VAs $Z_i(u)$, ($i = 1, \dots, N$) em fatores regionalizados independentes $\gamma_q^s(u)$ pode ser melhor compreendida por meio do seguinte exemplo. Se duas VAs ($i = 2$), $Z_1(u)$ e $Z_2(u)$, apresentassem o MLC abaixo

$$\gamma_{ij}(h) = B^1 \cdot \gamma^1(h) + B^2 \cdot \gamma^2(h)$$

onde:

$$\begin{aligned} \gamma_1(h) &= \text{sph}(h/20 \text{ m}) \\ \gamma_2(h) &= \text{sph}(h/100 \text{ m}) \end{aligned} \quad B^1 = \begin{pmatrix} b_{11}^1 = 16 & b_{12}^1 = 12 \\ b_{21}^1 = 12 & b_{22}^1 = 25 \end{pmatrix}; \quad B^2 = \begin{pmatrix} b_{11}^2 = 4 & b_{12}^2 = 0 \\ b_{21}^2 = 0 & b_{22}^2 = 0 \end{pmatrix}$$

e

esse MLC seria autorizado, pois $|b_{12}^1| \leq \sqrt{b_{11}^1 \cdot b_{22}^1}$, com $\rho^1 = 0,6$ e $|b_{12}^2| \leq \sqrt{b_{11}^2 \cdot b_{22}^2}$, com $\rho^2 = 0$.

Porém, o MLC é acima tudo autorizado porque permite que as VAs $Z_1(u)$ e $Z_2(u)$ sejam decompostas em fatores $\gamma_q^s(u)$. Por exemplo:

$$\begin{aligned} Z_1(u) &= \underbrace{4\gamma_1^1(u)}_{Z_1^1(u)} + \underbrace{0\gamma_2^1(u)}_{Z_2^1(u)} + \underbrace{2\gamma_1^2(u)}_{Z_1^2(u)} + m_1 & b_{11}^1 &= 4^2 + 0^2 = 16; & b_{11}^2 &= 2^2 = 4 \\ & & b_{22}^1 &= 3^2 + 4^2 = 25; & b_{22}^2 &= 0^2 = 0 \\ Z_2(u) &= \underbrace{3\gamma_1^1(u)}_{Z_1^2(u)} + \underbrace{4\gamma_2^1(u)}_{Z_2^2(u)} + \underbrace{0\gamma_1^2(u)}_{Z_1^2(u)} + m_2 & b_{12}^1 &= 4 \cdot 3 + 0 \cdot 4 = 12; & b_{12}^2 &= 2 \cdot 0 = 0 \end{aligned}$$

É importante salientar que a decomposição das matrizes de corre regionalização B^s em fatores $\gamma_q^s(u)$ (de coeficientes a_{iq}^s , representados pela matriz A^s) não é realizada de uma única maneira. No exemplo acima, foi utilizado o método dos componentes principais, onde que para cada estrutura s as VAs $Z_i^s(u)$ estão em correlação intrínseca. A análise regionalizada dos componentes principais consiste na decomposição espectral de cada matriz de corre regionalização B^s em valores próprios e vetores próprios (Apêndice A):

$$B^s = A^s \cdot (A^s)^t, \quad \text{com} \quad A^s = Q^s \sqrt{\Lambda^s} \quad \text{e} \quad Q^s \cdot (Q^s)^t = I$$

onde Q^s é a matriz ortogonal dos vetores próprios e Λ^s é a matriz diagonal constituída pelos valores próprios [representa a matriz diagonal das variâncias e covariâncias dos fatores regionalizados $\gamma_q^s(u)$] da matriz de corre regionalização B^s . As matrizes A^s representam os coeficientes do MLC, sendo que os elementos a_{iq}^s são as covariâncias regionalizadas entre as VAs $Z_i(u)$ e os fatores regionalizados $\gamma_q^s(u)$

(Wackernagel, 1993, p. 14). Segundo Rivoirard (2002), o número máximo de fatores a ser empregado é igual ao produto do número de variáveis pelo número de estruturas de base utilizado.

3.6.4.2. Simplificação da cokrigagem

A cokrigagem entre as K FAs indicadoras $I(u_{\alpha_i}; z_k)$, definidas pelos valores limite z_k , pode ser grandemente simplificada se essas FAs forem correlacionadas intrinsecamente (modelo de correlação intrínseca). Isto é, se a correlação entre as VAs for independente do vetor h e, portanto da escala de medida. Para que as K VAs indicadoras $I(u_{\alpha_i}; z_k)$ estejam correlacionadas intrinsecamente, todos os $K*(K+1)/2$ semi-variogramas simples e cruzados devem ser idênticos (proporcionais). Ou seja, todos os semi-variogramas envolvidos devem compartilhar as mesmas s estruturas de base, sendo que os s coeficientes de correlação estruturais ρ_s devem ser idênticos. Na verdade, esse modelo é um caso particular do MLC. Nesse caso, as K VAs indicadoras $I(u_{\alpha_i}; z_k)$ são autokrigáveis e conseqüentemente $[I(u; z_k)]_{CK}^* = [I(u; z_k)]_K^*$, ou seja: a cokrigagem se identifica com a krigagem.

Se as K FAs indicadoras estiverem correlacionadas intrinsecamente, o algoritmo de krigagem dos indicadores da mediana (KIM) também pode ser aplicado, o que pode simplificar ainda mais o processo de estimativa da dclp $F(u; z|(n))$. Entretanto, se esse não for o caso, outros algoritmos podem ser aplicados visando à simplificação do procedimento de cokrigagem. A seguir serão apresentados os métodos isofatoriais e de krigagem dos componentes principais dos indicadores, além do de KIM.

3.6.4.2.1. Krigagem dos indicadores da mediana

Como comentado na seção anterior, a aplicação desse método exige que as K FAs indicadoras estejam em correlação intrínseca. Isso porque, nesse algoritmo, todas as K VAs $I(u_{\alpha_i}; z_k)$ compartilham o semi-variograma dos indicadores referente ao valor limite z_k mediano, ou seja, $z_M = F^{-1}(0,5)$. Assim, embora os coeficientes b_{ij} de cada semi-variograma não sejam constantes, o algoritmo considera somente aquele de maior representatividade: b_{z_M} . Dessa forma, em cada localização, todas as K VAs indicadoras recebem os mesmos pesos. Ou seja, os pesos passam a ser independentes dos valores limite z_k e exclusivamente dependentes do semi-variograma mediano e da configuração espacial das informações retidas em cada localização a ser estimada. Assim, o algoritmo se torna fácil e compacto, na medida em que são necessários a modelagem e a resolução de um único semi-variograma e sistema de krigagem, respectivamente. Contudo, há a perda da incorporação das diferenças existentes entre os semi-variogramas dos valores limite z_k . Essa metodologia pode ser aplicada tanto nos algoritmos de krigagem simples como nos de krigagem ordinária dos indicadores. Por exemplo, o estimador $[I(u; z_k)]_{KIM}^*$ por meio do algoritmo de krigagem ordinária dos indicadores (KOI) é dado por:

$$\begin{aligned} [F(u; z_k | (n))]_{KIM}^* &= [I(u; z_k)]_{KIM}^* \\ &= \sum_{\alpha=1}^{n(u)} \lambda_{\alpha}^{KIM}(u) \cdot I(u_{\alpha}; z_k) \end{aligned} \quad k = 1, \dots, K \quad (3.69)$$

sendo que os ponderadores λ_{α}^{KIM} são determinados pela resolução do seguinte sistema de KOI:

$$\begin{cases} \sum_{\beta=1}^{n(u)} \lambda_{\beta}^{KIM}(u) \cdot C_{IM}(u_{\alpha} - u_{\beta}) + \mu_{KIM}(u) = C_{IM}(u_{\alpha} - u) & \alpha = 1, \dots, n(u) \\ \sum_{\beta=1}^{n(u)} \lambda_{\beta}^{KIM}(u) = 1 \end{cases} \quad (3.70)$$

onde $C_{IM}(h)$ representa a covariância espacial das VAs $I(u_{\alpha}; z_k)$, de acordo o valor limite z_k mediano; e μ_{KIM} representa o lagrangeano, segundo as equações de KIM.

3.6.4.2.2. Krigagem fatorial multivariada

A decomposição das VAs indicadoras $I(u; z_k)$ em fatores regionalizados $Y_q^s(u)$ pode ser de grande utilidade quando for verificado que a correlação ($h = 0$) entre o conjunto de K VAs indicadoras $I(u_{\alpha}; z_k)$ é dependente da escala de medida. A correlação entre as K VAs pode ser o resultado da mistura entre a correlação existente em diferentes escalas de medida. Por exemplo, a correlação medida a grande escala, entre certos pares de VAs indicadoras, pode ser maior quando comparada com a correlação original [escalas misturadas, todas K VAs $I(u_{\alpha}; z_k)$]. Nesse caso, ao invés de fazer um tratamento geostatístico utilizando simplesmente o modelo de continuidade espacial resultante, seria interessante aproveitar da melhor maneira possível essa maior correlação apresentada em grande escala. Isso se torna possível a partir do momento em que os fatores regionalizados $Y_q^s(u)$ sejam estimados, unicamente, em função das características espaciais de cada estrutura s envolvida no MLC.

O procedimento conhecido como **krigagem fatorial multivariada (KFM) ou análise krigante** (Matheron, 1982) permite estimar a probabilidade associada a cada valor limite z_k em função da estimativa, por CKOI, dos fatores regionalizados de cada escala de medida s . Ou seja, cada matriz de correogionalização B^s é analisada separadamente, de modo que as demais estruturas presentes no MLC sejam filtradas. Assim, o estimador de $[I(u; z_k)]^*$ é dado por:

$$[I(u, z_k)]_{KFM}^* = m_k + \sum_s \sum_q a_{kq}^s [Y_q^s(u)]_{CKOI}^* \quad (3.71)$$

onde:

m_k é a média $\{E[I(u, z_k)]\}$ da VA indicadora, de acordo com a faixa indicadora k ;

$Y_q^s(u)$ são os q fatores regionalizados das s estruturas e

a_{kq}^s são os coeficientes dos fatores regionalizados $Y_q^s(u)$.

Os fatores regionalizados $\gamma_q^s(u)$ são mutuamente independentes, porém o conjunto de K fatores (de mesmo índice q e s) que ajudam a decompor as K VAs indicadoras são espacialmente correlacionados. Por exemplo, a Figura 3.10 mostra a relação existente entre os fatores regionalizados do exemplo apresentado anteriormente (MLC).

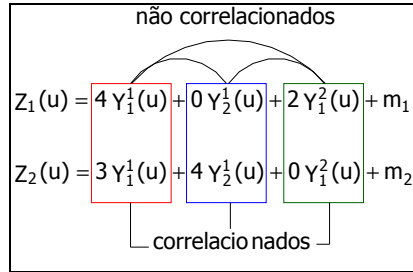


Figura 3.10: Esquema representando a relação de correlação espacial entre os fatores regionalizados.

Assim, os correspondentes fatores regionalizados $\gamma_q^s(u)$ de diferentes VAs são calculados por CKOI a partir das K VAs indicadoras $I(u_\alpha; z_k)$:

$$[Y_q^s(u)]_{CKOI}^* = \sum_{k=1}^K \sum_{\alpha=1}^{n(u)} \lambda_{q\alpha k}^s(u) \cdot I(u_\alpha, z_k) \quad q = 1, \dots, K \quad (3.72)$$

onde os ponderadores $\lambda_{q\alpha k}^s(u)$, associados aos q fatores regionalizados das s estruturas, são determinados por meio da resolução do seguinte sistema de CKOI:

$$\begin{cases} \sum_{k=1}^K \sum_{\beta=1}^{n(u)} \lambda_{q\alpha k}^s(u) \cdot C(u_\alpha, u_\beta; z_k, z_{k'}) + \mu_{kq}^s = a_{kq}^s \cdot C^s(u, u_\alpha) & \alpha = 1, \dots, n(u) \\ & k = 1, \dots, K \\ \sum_{\beta=1}^{n(u)} \lambda_{q\alpha k}^s(u) = 0 & k = 1, \dots, K \end{cases} \quad (3.73)$$

onde:

$C(h; z_k, z_{k'})$ é a covariância espacial (direta ou cruzada) entre as VAs indicadoras $I(u_\alpha; z_k)$, referente aos valores limite z_k e $z_{k'}$, separados por uma distância representada pelo vetor h;

$a_{kq}^s C^s(u, u_\alpha)$ representa a covariância espacial entre as K VAs indicadoras e os fatores regionalizados $\gamma_q^s(u)$ e

μ_{kq}^s é o lagrangeano do sistema de cálculo dos q pesos associados às s estruturas das K VAs indicadoras

Como qualquer estimador de krigagem, o estimador $[Y_q^s(u)]_{CKOI}^*$ não deve ser tendencioso. Para isso, a esperança matemática do erro $E\left([Y_q^s(u)] - [Y_q^s(u)]_{CKOI}^*\right)$ deve ser nula. Assim, como os fatores regionalizados $\gamma_q^s(u)$ são FAs de média nula, para que essa condição seja respeitada o

somatório de todos os ponderadores deve ser nulo. Ou seja, a condição de não tendenciosidade filtra as médias das K VAs indicadoras $I(u_{\alpha}; z_k)$.

No sistema de CKOI apresentado acima, o termo $a_{kq}^s C^s(u, u_{\alpha})$ representa a covariância espacial entre as K VAs indicadoras e os fatores regionalizados $\gamma_q^s(u)$. Entretanto, essa covariância não possui nenhum significado físico, uma vez os fatores regionalizados são meramente matemáticos.

Portanto, os modelos isofatoriais também podem ser utilizados para simplificar o procedimento de cokrigagem, especialmente se houverem mais de três variáveis (como é o caso de cokrigagem dos indicadores). Por exemplo, se $K = 10$, a estimativa da dclp em uma localização u por CKOI seria o resultado de 10 combinações lineares de $10 \cdot n(u)$ elementos (Equação 3.56). Sendo que, os ponderadores seriam calculados por meio de 10 sistemas de CKOI compostos por $10 \cdot n(u) + 10$ equações. Se KFM fosse aplicada (a partir do MLC, decomposto por exemplo em $\gamma_1^1(u)$, $\gamma_2^1(u)$ e $\gamma_1^2(u)$, como apresentado na Figura 3.10), a estimativa da dclp em u seria o resultado de 10 combinações lineares de quatro elementos (Equação 3.71). Logo, com o emprego de KFM seria necessária a resolução de somente três sistemas de cokrigagem, que definiriam os três fatores $\gamma_q^s(u)$. Contudo, esse procedimento não dispensa o cálculo do MLC.

A cokrigagem de duas VAs $Z_i(u)$ também pode se simplificada, se o MLC se enquadrar no chamado "Modelo do Resíduo" ou "Modelo de Markov" (Rivoirard, 2001 e 2002). Nesse caso, uma das duas variáveis, juntamente com o resíduo $[R(u) = Z^*(u) - Z(u)]$, são autokrigáveis. Assim, a VA $Z_i(u)$ que não é autokrigável é estimada em função da krigagem do resíduo e da outra variável. Por exemplo, o MLC exemplo apresentado acima (p. 54 e Figura 3.10) se enquadra no modelo do resíduo. Nesse caso a VA $Z_2(u)$ é autokrigável e a VA $Z_1(u)$ poderia ser determinada, por cokrigagem, dessa forma:

$$[Z_1(u)]_{CK}^* = a \cdot [Z_2(u)]_k^* + b + [R(u)]_k^* \quad (3.74)$$

$$\text{onde } a = \frac{\text{Cov}[Z_1(u), Z_2(u)]}{\text{Var}[Z_1(u)]} \text{ e } b = m_1 - a \cdot m_2$$

Os termos m_1 e m_2 são as médias das VAs $Z_1(u)$ e VA $Z_2(u)$, respectivamente. Porém essa simplificação só é aplicada nos casos onde existem duas variáveis, o que não é o caso de cokrigagem de K VAs indicadoras $I(u_{\alpha}; z_k)$.

3.6.4.2.3. Krigagem dos componentes principais dos indicadores

O método de krigagem dos componentes principais dos indicadores (KCPI) (Suro-Pérez e Journel, 1991) agrega a Análise dos Componentes Principais (ACP) (Matheron, 1982; Davis e Greenes, 1983) e o formalismo dos indicadores para modelar as dclp. Nesse algoritmo, as dclp são modeladas a partir da estimativa dos componentes principais dos indicadores $Y_k(u)$, ao invés da estimativa dos próprios indicadores $I(u_{\alpha}; z_k)$. O método de KCPI é considerado uma simplificação dos algoritmos de

cokrigagem, pois permite que as dclp sejam modeladas com a mesma eficiência proporcionada pelos algoritmos de CKI e com o mesmo esforço despendido para KI. Isso porque, a ACP faz com que as VAs indicadoras $I(u_{\alpha}; z_k)$ sejam transformadas em componentes $Y_k(u)$ “independentes”, resumindo as covariâncias diretas e cruzadas dos indicadores em covariâncias diretas dos componentes principais. Entretanto, como será visto a seguir, as covariâncias cruzadas dos componentes principais não são necessariamente nulas, mas são insignificantes quando comparadas com as covariâncias diretas – isso explica porque o termo independente foi escrito entre aspas na frase acima. Isso faz com que os componentes principais dos indicadores $Y_k(u)$ possam ser considerados como sendo autokrigáveis, isto é: $[Y_k(u)]_{CK}^* = [Y_k(u)]_K^*$.

A Análise dos Componentes principais é uma técnica algébrica de transformação linear de vetores. Em duas dimensões, ACP pode ser visto como uma rotação de eixos ortogonais, de modo que o ângulo de rotação seja tal que o espalhamento da primeira variável transformada, ao longo do primeiro eixo, seja máximo, enquanto que o espalhamento da segunda variável transformada, ao longo do segundo eixo, seja mínimo (Anderson, 1984).

ACP é um procedimento de ortogonalização que não requer nenhuma hipótese estatística *a priori* sobre as informações. Entretanto, a aplicação estatística dessa ortogonalização mostra que as covariâncias cruzadas entre as VAs transformadas $Y_k(u)$ são estritamente nulas (Suro-Pérez e Journal, 1991). Ou seja:

$$\text{Cov}(Y_i Y_j) = E(Y_i Y_j) - E(Y_i) \cdot E(Y_j) = 0, \quad \forall j \neq i \quad (3.75)$$

Outra característica importante apresentada pela ACP é que a variância dos componentes principais (Y_k) é decrescente, isto é:

$$\text{Var}(Y_1) > \text{Var}(Y_2) > \text{Var}(Y_3) > \dots > \text{Var}(Y_{k-1}) > \text{Var}(Y_k)$$

Sendo que, na maioria das vezes, as variâncias apresentadas pelos últimos componentes principais são quase nulas. Portanto, a aplicação da ACP permite que a dimensão do problema seja reduzida por meio dos componentes principais que mais contribuem para variância.

A determinação dos componentes principais $Y_k(u)$ pode ser realizada a partir da decomposição espectral da matriz de covariâncias espaciais $C(h; z_k, z_k')$ das K VAs indicadoras $I(u_{\alpha}; z_k)$:

$$C(h; z_k, z_k') = Q \cdot \Lambda \cdot Q^t, \quad \text{sendo que} \quad Q \cdot Q^t = I$$

Como anteriormente, a matriz Q representa os vetores próprios de $C(h; z_k, z_k')$ e Λ representa a matriz diagonal constituída pelos seus valores próprios (que representam as variâncias dos componentes principais dos indicadores). Na prática, essas matrizes podem ser obtidas com o auxílio de qualquer rotina de decomposição espectral, por exemplo, o algoritmo de decomposição do valor singular (DVS) – Apêndice A.

A partir daí, o vetor dos componentes principais dos indicadores é calculado por meio de uma simples multiplicação de matrizes, em um processo chamado de ortogonalização:

$$Y(u) = Q^t \cdot I(u; z) \quad (3.76)$$

Ou seja, cada componente principal $Y_k(u)$ é calculado da seguinte forma:

$$Y_k(u_\alpha) = \sum_{k'=1}^K q_{k'k} \cdot I(u_\alpha; z_{k'}) \quad \begin{array}{l} \alpha = 1, \dots, n(u) \\ k = 1, \dots, K \end{array} \quad (3.77)$$

onde $q_{k'k}$ são os elementos da matriz ortogonal Q , enquanto que $I(u_\alpha; z_{k'})$ são os elementos do vetor $I(u; z)$. Portanto, os componentes principais $Y_k(u_\alpha)$ são o resultado da combinação linear das K VAs indicadoras $I(u_\alpha; z_k)$ com os vetores próprios resultantes da decomposição de $C(h; z_k, z_{k'})$.

Entretanto, a ortogonalização, apresentada acima, não garante que as covariâncias cruzadas $C_Y(h; z_k, z_{k'})$ sejam nulas para vetores h diferentes daqueles empregados no momento da ortogonalização. Porém, foi provado empiricamente (Suro-Pérez e Journal, 1991) que, nesses casos, a magnitude de correlação é tão insignificante que pode ser desprezada. Assim, a matriz de variâncias e covariâncias dos componentes principais dos indicadores $Y_k(u)$ pode ser considerada como sendo uma matriz diagonal, independente do vetor h . Porém, é conveniente que essa condição sempre seja testada. Wackernagel (1994), reforça exatamente isso, alertando para o fato de que os componentes principais possam apresentar um certo grau de correlação em curtas distâncias.

Em princípio, a ortogonalização pode ser realizada para qualquer vetor h . Porém, na prática, essa distância deve ser tal que as correlações espaciais entre os componentes principais sejam tão mínimas quanto possíveis, de modo que possam ser ignoradas. Em especial, essas correlações devem ser insignificantes quando comparadas com as covariâncias diretas. Na prática, uma primeira aproximação consiste em utilizar o menor passo (*lag distance*) do semi-variograma experimental dos indicadores, pois as covariâncias cruzadas são decrescentes para a maioria dessas distribuições. Como alternativa, também pode-se utilizar $h=0$ ou $h \rightarrow 0$. Nesse caso, a grande vantagem é que os semi-variogramas diretos e cruzados dos indicadores se tornam desnecessários para a determinação de seus componentes principais. Isto é, a decomposição espectral pode ser realizada a partir da matriz de variâncias e covariâncias dos indicadores.

A partir da transformação linear das K VAs indicadoras $I(u_\alpha; z_k)$ em componentes principais $Y_k(u)$ "independentes", os componentes principais nas localizações não medidas são calculados a partir da seguinte combinação linear:

$$[Y_k(u)]_{KCPI}^* = \sum_{\alpha=1}^{n(u)} \lambda_\alpha^{KCPI} \cdot Y_k(u_\alpha) \quad k = 1, \dots, K \quad (3.78)$$

onde os ponderadores λ_α^{KCPI} dos componentes principais Y_k nas localizações u , são determinados pela resolução do seguinte sistema de KO:

$$\left\{ \begin{array}{l} \sum_{\beta=1}^{n(u)} \lambda_{\beta z_k}^{KCPI}(u) \cdot C_Y(u_\alpha - u_\beta, z_k) + \mu_{z_k}^{KCPI}(u) = C_Y(u_\alpha - u; z_k) \\ \sum_{\beta=1}^{n(u)} \lambda_{\beta z_k}^{KCPI}(u) = 1 \end{array} \right. \quad \alpha = 1, \dots, n(u) \quad (3.79)$$

onde:

$C_Y(h; z_k)$ é a covariância espacial entre os componentes principais $Y_k(u_\alpha)$ e

$\mu_{z_k}^{KCPI}$ é o lagrangeano do sistema CPI.

Após o cálculo da estimativa dos componentes principais nas localizações desconhecidas, as probabilidades condicionais podem ser determinadas por meio da seguinte transformação inversa:

$$\begin{aligned} [F(u; z_k | (n))]_{KCPI}^* &= [I(u; z_k)]_{KCPI}^* \\ &= Q \cdot Y^*(u) \end{aligned} \quad k = 1, \dots, K \quad (3.80)$$

onde Q representa os vetores próprios de $C(h; z_k, z_k)$. Em cada localização não medida u, o vetor Y^* é dado por:

$$Y^*(u) = [Y_1^*(u) \dots Y_k^*(u)]^t \quad (3.81)$$

onde, t indica transposta. Em termos de somatório:

$$[F(u; z_k | (n))]_{KCPI}^* = \sum_{k'=1}^K q_{k k'} \cdot Y_{z_{k'}}(u) \quad k = 1, \dots, K \quad (3.82)$$

Assim, a dclp associada a cada localização desconhecida fica sendo definida pelo seguinte vetor:

$$[F(u; z | (n))]_{KCPI}^* = [F(u; z_1 | (n)) \dots F(u; z_k | (n))]^t \quad (3.83)$$

No exemplo apresentado por Suro-Pérez e Journel (1991), a matriz de covariâncias dos indicadores foi ortogonalizada para um vetor h unitário. Nesse exemplo, havia nove VAs indicadoras $I(u_\alpha; z_k)$ ($K = 9$) e, portanto, nove componentes principais $Y_k(u)$ ($k = 1, 2, \dots, 9$). Como os autores esperavam, as covariâncias cruzadas espaciais ($h \neq 0$) entre os componentes principais dos indicadores foram nulas para alguns pares de componentes e insignificantes para outros. Dessa forma, teoricamente, seriam necessários o cálculo e o ajuste de nove semi-variogramas experimentais, assim como a resolução de nove sistemas de krigagem. Contudo, somente os três primeiros componentes principais apresentaram alguma estruturação espacial (sendo que as variâncias dos últimos componentes principais foram quase nulas). Os outros seis componentes principais tiveram seus semi-variogramas modelados como efeito pepita puro. Assim, para a modelagem da dclp em cada localização não medida, foi necessário resolver somente três sistemas de krigagem. Nesse caso, o algoritmo de KCPI simplificou enormemente a CKI e também significativamente a KI, sendo que os resultados obtidos foram muito similares àqueles obtidos a partir do algoritmo de CKI (teoricamente melhor do que o de KI).

Portanto, além de transformar problemas de K^2 dimensões (ou $K*(K+1)/2$, considerando as covariâncias como sendo simétricas) para problemas de K dimensões, o algoritmo de KCPI ainda mostra a característica de desestruturação dos semi-variogramas dos componentes principais dos indicadores com o aumento de K . Além disso, em relação a KI, o cálculo e o ajuste dos semi-variogramas experimentais dos componentes principais dos indicadores se torna mais fácil. Isso porque, os componentes principais dos indicadores são variáveis contínuas e, portanto os semi-variogramas calculados a partir dessas variáveis são mais representativos – contando com um maior número de pares.

3.6.5. Correção para os desvios de relação de ordem

No método não-paramétrico (indicadores), a (co)krigagem dos indicadores $I(u_\alpha; z_k)$ pode gerar problemas de relação de ordem. Esses problemas são caracterizados pela estimativa de probabilidades incoerentes (fora domínio limitado por $[0,1]$), ou ainda aqueles relacionados com a não obediência de uma relação crescente de probabilidades de acordo com o aumento dos valores limite z_k (Journel e Posa, 1990). Ou seja: se $F(u, z_{k-1}|(n)) > F(u; z_k |(n))$. Por isso, após a realização das estimativas das dclp $F(u; z|(n))$, é preciso verificar se não existem problemas de relação de ordem, visto que, essas funções devem ser caracterizadas por um comportamento monotônico crescente. Isto é:

$$[F(u; z_k |(n))]^* \leq [F(u; z_{k'} |(n))]^* \quad \forall z_{k'} > z_k \quad (3.84)$$

sendo que: $[F(u; z_k |(n))]^* \in [0, 1]$.

Os problemas relacionados à característica monotônica crescente da dclp $F(u; z|(n))$ podem ocorrer porque as K probabilidades $F(z_k)$ são estimadas independentemente, a partir de semi-variogramas diferentes. Outro motivo que pode aumentar a chance de ocorrência desse tipo de problema de relação de ordem, é a falta de informações $z(u_\alpha)$ em determinadas classes de valores limite $z_{k'}$ (z_{k-1}, z_k). Isso porque, o conjunto de informações indicadoras retidas para a estimativa de $F(z_k)$ e de $F(z_{k-1})$ seria o mesmo. Dessa forma, somente não haveria chance de ocorrer problemas associados ao comportamento monotônico crescente da dclp $F(u; z|(n))$ se os semi-variogramas dos indicadores dos valores limite z_k e z_{k-1} fossem iguais.

Algumas estimativas de probabilidade podem ocorrer fora do domínio $[0;1]$ porque a resolução dos sistemas de (co)krigagem pode fazer com que o peso atribuído a algumas informações $z(u_\alpha)$ seja negativo. Na verdade, as estimativas por meio dos algoritmos de (co)krigagem são o resultado de combinações lineares não convexas das informações condicionantes. Segundo Deutsch e Journel (1998, p. 78), a probabilidade de ocorrência de pesos negativos é maior quando se usa algoritmos de (co)krigagem ordinária, quando comparado com os algoritmos de (co)krigagem simples. Isso é explicado pelas condições impostas aos pesos das variáveis secundárias. A condição de que o somatório dos pesos dessas variáveis seja nulo aumenta as chances de ocorrência de pesos negativos.

Algumas medidas podem ser tomadas para reduzir a ocorrência ou mesmo atenuar a proporção e a magnitude dos desvios de relação de ordem, como:

- i. No emprego de algoritmos de CKOI e KP, usar somente uma condição de não tendenciosidade para diminuir a proporção de pesos negativos.
- ii. Evitar mudanças significativas nos semi-variogramas dos indicadores de valores limite z_k consecutivos. Preferivelmente, os parâmetros desses semi-variogramas devem variar suavemente de modo que se evidencie um comportamento contínuo da variabilidade espacial conforme os valores limite z_k . Como alternativa, é possível usar o mesmo semi-variograma dos indicadores para representar a variabilidade espacial apresentada para os K valores limite z_k (como no caso de krigagem dos indicadores da mediana).
- iii. Escolher valores limite z_k de modo que exista a ocorrência de pelos menos uma informação $z(u_\alpha)$ dentro de cada classe definida por $(z_{k-1}, z_k]$.

Para aumentar a influência das demais VAs indicadoras $I(u_\alpha; z_k)$ (com $z_k \neq z_{k0}$, onde z_{k0} representa a faixa indicadora alvo), e conseqüentemente diminuir a ocorrência de pesos negativos, o sistema de cokrigagem pode passar a ter somente uma equação de restrição imposta aos ponderadores, ou seja:

$$\sum_{k=1}^K \sum_{\alpha=1}^{n(u)} \lambda_{\alpha k} (u; z_k) = 1 \quad (3.85)$$

Assim, o sistema de cokrigagem passa a possuir $K \cdot n(u) + 1$ equações ao invés de $(K \cdot n(u)) + K$ equações. Com essa modificação, o algoritmo de CKOI passa a ser escrito da seguinte maneira:

$$[F(u; z_{k_0} | (n))]_{\text{CKOI}}^* = \sum_{k=1}^K \sum_{\alpha=1}^{n(u)} \lambda_{\alpha k} (u; z_{k_0}) \cdot I(u_\alpha; z_k) + \lambda_0 \quad (3.86)$$

onde $\lambda_0 = [m_{k_0} - m_k] \cdot \sum_{\alpha=1}^{n(u)} \lambda_{\alpha k}$ com $k = 1, \dots, K$ sendo que $k \neq k_0$

Nesse caso, as médias m_k [ou suas probabilidades globais $F(z_k)$] devem ser conhecidas. Isso implica que as mesmas devem ser suficientemente representativas para permitir que $\hat{m}_k \approx E[I(u; z_k)]$. Contudo, isso não significa que as variações locais em torno das médias sejam desconsideradas. Como no algoritmo de CKOI tradicional, a condição de estacionaridade continua sendo limitada pelas informações da circunvizinhança da localização a ser estimada.

Os problemas de relação de ordem podem ser evitados por meio da incorporação, no código fonte do algoritmo de (co)krigagem, das duas condições apresentadas acima. Por exemplo, Barnes e Johnson (1984) propõem que os sistemas de krigagem resultem em, somente, pesos positivos. Alguns softwares de geoestatística aplicada (como o Isatis) permitem que essas condições sejam adicionadas facilmente. Contudo, uma técnica comumente aplicada é a da correção posterior das $dclp$ $F(u; z|(n))$. Embora a taxa de ocorrência de problemas de relação ordem seja alta, a experiência mostra que a grande maioria desses problemas é de pequena magnitude, podendo ser facilmente corrigidos.

Existem vários algoritmos para a correção dos problemas de relação de ordem, porém o mais simples e direto é proposto por Deutsch e Journal (1998, p. 81). O algoritmo consiste na média entre as probabilidades corrigidas em direção a probabilidades ligadas a valores limite z_k imediatamente inferiores e superiores. O algoritmo pode ser dividido em três etapas:

i. Correção para a probabilidade ligada ao valor limite z_{k-1} , de acordo com o seguinte subalgoritmo:

- reajuste das probabilidades fora do intervalo [0;1] de variação:

$$\begin{aligned} [F(u; z_k | (n))]_{INF}^{**} &= 0 & \text{se } [F(u; z_k | (n))]^* < 0 & \quad e \\ [F(u; z_k | (n))]_{INF}^{**} &= 1 & \text{se } [F(u; z_k | (n))]^* > 1 & \end{aligned}$$

- reescalamento para a probabilidade $F[(u; z_{k-1}|(n))]^*$

$$\begin{aligned} [F(u; z_k | (n))]_{INF}^{**} &= [F(u; z_k | (n))]^* & \text{se } [F(u; z_k | (n))]^* \geq [F(u; z_{k-1} | (n))]^* \\ &= [F(u; z_{k-1} | (n))]^* & \text{se não.} \end{aligned}$$

ii. Correção para a probabilidade ligada ao valor limite z_{k+1} , de acordo com o seguinte subalgoritmo:

- reajuste das probabilidades fora do intervalo [0;1] de variação:

$$\begin{aligned} [F(u; z_k | (n))]_{SUP}^{**} &= 0 & \text{se } [F(u; z_k | (n))]^* < 0 & \quad e \\ [F(u; z_k | (n))]_{SUP}^{**} &= 1 & \text{se } [F(u; z_k | (n))]^* > 1 & \end{aligned}$$

- reescalamento para a probabilidade $F[(u; z_{k+1}|(n))]^*$

$$\begin{aligned} [F(u; z_k | (n))]_{SUP}^{**} &= [F(u; z_k | (n))]^* & \text{se } [F(u; z_k | (n))]^* \leq [F(u; z_{k+1} | (n))]^* \\ &= [F(u; z_{k+1} | (n))]^* & \text{se não.} \end{aligned}$$

iii. O conjunto final de K probabilidades corrigidas é resultado da média entre as probabilidades corrigidas para $F[(u; z_{k-1}|(n))]^{**}$ e $F[(u; z_{k+1}|(n))]^{**}$:

$$[F(u; z_k | (n))]^{**} = \frac{[F(u; z_k | (n))]_{INF}^{**} + [F(u; z_k | (n))]_{SUP}^{**}}{2}, \quad k = 1, \dots, K.$$

3.6.6. Discretização da dclp experimental cumulativa não-paramétrica

Como descrito anteriormente, nos métodos não-paramétricos, a dclp em cada localização não medida u , é construída de acordo com a probabilidade condicional de cada valor limite z_k . Assim, a discretização da dclp cumulativa experimental em u é diretamente proporcional ao número K de valores limite z_k empregado. Por outro lado, a tarefa de estimativa das dclp também se torna mais trabalhosa a medida em que K aumenta, sobretudo se o algoritmo escolhido for CKI e nenhuma simplificação possa ser empregada.

O número K de valores limite z_k deve ser escolhido de modo que permita que o modelo de incerteza local seja suficientemente discretizado. De maneira geral, K não precisa ser tão elevado (para que a probabilidade de ocorrência de problemas de relação de ordem não seja alta) nem deve

ser tão baixo. É sabido que K não precisa ser maior do que 15 para que o modelo de incerteza local seja bem modelado. Entretanto, é recomendado que K não seja inferior a cinco. Geralmente, para a maioria das distribuições experimentais, utilizam-se nove valores limite z_k correspondentes aos nove decis da distribuição experimental.

De qualquer forma, para que a dclp estimada por meio dos métodos não-paramétricos possa ser utilizada de maneira adequada, é necessário que sua resolução seja aumentada, como no caso do método paramétrico. Como anteriormente, a resolução das dclp pode ser melhorada por meio da interpolação dentro de cada classe $(z_{k-1}; z_k]$ e extrapolação além dos dois valores limite extremos (z_1 e z_k). Os três modelos de interpolação e extrapolação das dclp utilizados para o método paramétrico (modelos linear, potência e hiperbólico) também são aplicáveis para o aumento da resolução das dclp via métodos não-paramétricos. Entretanto, a simples interpolação e extrapolação das probabilidades condicionais não é suficiente para que a dclp seja suficientemente detalhada. Isso porque, do ponto de vista de discretização de uma função, o número K de valores limite z_k ($5 \leq K \leq 15$) é muito pequeno. Assim, na prática, a informação disponível a partir da função de repartição empírica da VA $Z(u)$ [n informações $z(u_\alpha)$] também pode ser utilizada para auxiliar no aumento da resolução intra-classes.

Cada classe $(z_{k-1}; z_k]$ pode ser dividida em $L^{(k)}$ subclasses $(z_k^{(l-1)}; z_k^{(l)})$, de acordo com o número de informações $z(u_\alpha)$ contidas no respectivo intervalo. Por exemplo, se três informações $z(u_\alpha)$ estivessem contidas em uma determinada classe $(z_{k-1}; z_k]$, a mesma poderia ser dividida em quatro subclasses ($L^{(k)} = 4$): $(z_{k-1} = z_k^{(0)}; z_k^{(1)})$, $(z_k^{(1)}; z_k^{(2)})$, $(z_k^{(2)}; z_k^{(3)})$ e $(z_k^{(3)}; z_k^{(4)} = z_k]$. A seguir, o interior de cada subclasse é linearmente interpolado, de acordo com a função de repartição empírica. A partir daí, o resultado dessa interpolação (probabilidade incondicional) é utilizado para auxiliar na interpolação linear das probabilidades condicionais associadas aos valores da VA $Z(u)$ contidos em cada classe $(z_{k-1}; z_k]$. Goovaerts (1997, p. 327-328) descreve esse processo em três etapas:

- i. Divisão de cada classe $(z_{k-1}; z_k]$ em $L^{(k)}$ subclasses $(z_k^{(l-1)}; z_k^{(l)})$.
- ii. Interpolação linear dentro de cada subclasse, conforme a seguinte equação:

$$[F(z)]_{\text{Lin}} = F^*(z_k^{(l-1)}) + \left[\frac{z - z_k^{(l-1)}}{z_k^{(l)} - z_k^{(l-1)}} \right] \cdot [F^*(z_k^{(l)}) - F^*(z_k^{(l-1)})], \quad \forall z \in (z_k^{(l-1)}, z_k^{(l)}], \quad (3.87)$$

$l = 1, \dots, L^{(k)}$

- iii. As probabilidades incondicionais calculadas em (ii), para cada um dos L conjuntos, são transformadas em probabilidades condicionais, de modo que as probabilidades condicionais associadas à z_{k-1} e z_k sejam respeitadas. Isso é realizado por meio do reescalonamento dos resultados obtidos em (ii):

$$[F(u; z | (n))]_{\text{Lin}} = [F(u; z_{k-1} | (n))]^* + \varphi \cdot [F(z)]_{\text{Lin}} - F^*(z_{k-1}) \quad (3.88)$$

Sendo que

$$\varphi = \frac{[F(u; z_k | (n))]^* - [F(u; z_{k-1} | (n))]^*}{F^*(z_k) - F^*(z_{k-1})} \quad (3.89)$$

3.7. Algoritmos de ssG e ssi

Como descrito nas seções anteriores, os algoritmos de simulação seqüencial estocástica, dois grandes grupos podem ser distinguidos:

- i. simulação seqüencial Gaussiana e
- ii. simulação seqüencial dos indicadores.

A grande diferença entre esses dois grupos está no procedimento de estimativa das dclp: paramétrico x não-paramétrico. O método de simulação seqüencial Gaussiana (ssG) utiliza o formalismo multi-Gaussiano, enquanto que o método de simulação seqüencial dos indicadores (ssi) usa o formalismo homônimo. Uma vez definidos os modelos de incerteza locais (dclp), a aplicação do algoritmo seqüencial, apresentado na seção 3.3.1, garante a obtenção de tantas realizações do modelo de FA $Z(u)$ quanto desejado.

3.7.1. Algoritmos de ssG

O algoritmo de ssG é implementado da seguinte maneira:

- i. transformação da distribuição experimental $z(u_\alpha)$ ($\alpha = 1, \dots, n$) em uma distribuição de n valores $y(u_\alpha)$ ($\alpha = 1, \dots, n$) por meio da função de transformação normal $\varphi(\cdot)$ (também chamada de função de anamorfose Gaussiana) – seção 3.5.1;
- ii. ajuste de um modelo de covariância para a distribuição dos dados Gaussianos, a partir do cálculo do covariograma experimental: $c_y(h) = \frac{1}{N(h)} \sum_{\alpha=1}^{N(h)} [y(u_\alpha) \cdot y(u_\alpha + h)]$;
- iii. definição de um caminho aleatório, de modo que cada localização u_j ($j = 1, \dots, N$) desconhecida (célula ou bloco da malha) seja visitada somente uma única vez;
- iv. construção do modelo de incerteza (dclp) na localização u_j – via KS, condicionado às n informações experimentais na circunvizinhança de u_j ;
- v. simulação, por meio de tiragem aleatória da dclp (simulação de Monte Carlo), de um valor da VA $Y(u_j)$, isto é: $y^{(l)}(u_j)$ ($l = 1, \dots, L$);
- vi. inclusão de $y^{(l)}(u_j)$ no banco de dados, representando uma informação condicional adicional a ser utilizada nas demais localizações a serem visitadas;
- vii. repetição das etapas (iv) a (vi) até que seja associada uma simulação a cada uma das N localizações;
- viii. repetição dos passos (iii) a (vii) para gerar L realizações equiprováveis da distribuição espacial da FA $Y(u)$.

- ix. transformação inversa $\varphi^{-1}(\cdot)$ dos valores simulados Gaussianos $y(u_j)$ ($j = 1, \dots, N$) em valores simulados no espaço da variável original $z(u_j)$ ($j = 1, \dots, N$).

3.7.2. Algoritmos de ssi

O algoritmo de ssi pode ser resumido nas seguintes etapas:

- i. definição do número K de valores limite z_k , discretizando $(K+1)$ classes com frequências aproximadamente iguais;
- ii. codificação binária distribuição experimental $z(u_\alpha)$ ($\alpha = 1, \dots, n$) em um vetor $i(u_\alpha; z_k)$, sendo que: ($\alpha = 1, \dots, n$) e ($k = 1, \dots, K$) – seção 3.6.1,
- iii. cálculo e ajuste dos covariogramas dos indicadores $C_i(h, z_k)$ para os K de valores limite z_k , de acordo com o método escolhido (KI ou CKI);
- iv. definição de um caminho aleatório, de modo que cada localização u_j ($j = 1, \dots, N$) desconhecida (célula ou bloco da malha) seja visitada somente uma única vez;
- v. construção do modelo de incerteza (dclp) na localização u_j a partir do cálculo (via (co)krigagem) das probabilidades associadas a cada valor limite z_k $[F(u; z_k | (n))]^*$ (condicionado às n informações experimentais na circunvizinhança de u_j);
- vi. simulação, por meio de tiragem aleatória da dclp (simulação de Monte Carlo), de um valor da VA $Z(u_j)$, isto é: $z^{(l)}(u_j)$ ($l = 1, \dots, L$);
- vii. inclusão de $z^{(l)}(u_j)$ no banco de dados, representando uma informação condicional adicional a ser utilizada nas demais localizações a serem visitadas;
- viii. repetição das etapas (v) a (vii) até que seja associada uma simulação a cada uma das N localizações;
- ix. repetição dos passos (iv) a (viii) para gerar L realizações equiprováveis da distribuição espacial da FA $Z(u)$.

3.8. Comentários

Esse capítulo consistiu em uma revisão dos principais métodos de simulação condicional seqüencial, apresentando em detalhes as características teóricas e práticas envolvidas na aplicação de cada método. Inicialmente, foi revisado o conceito dos métodos de Monte Carlo, envolvido nos métodos de simulação estocástica. Em seguida foram apresentados os conceitos de simulação e condicionamento, juntamente com a definição do algoritmo seqüencial. Foram introduzidos os modelos de incerteza ou distribuições condicionais locais de probabilidade (dclp), que constituem o alicerce dos métodos condicionais seqüenciais. Finalmente, os algoritmos de ssG e ssi foram apresentados, de acordo com suas implementações práticas.

Principalmente, nas seções 3.5 e 3.6, foram revisados os métodos paramétricos e não-paramétricos de obtenção das dclp, mostrando suas características principais. Foi mostrado, que embora o método paramétrico seja mais prático, nem sempre o mesmo pode ser adequadamente aplicado. Nessa situação a abordagem não-paramétrica deve ser utilizada.

Na seqüência, o Capítulo 4 apresenta como os modelos de incerteza podem ser adequadamente utilizados, de acordo com o objetivo almejado.

Capítulo 4

Incerteza e funções de transferência

Este capítulo apresenta as ferramentas matemáticas que permitem extrair o máximo de informações inclusas e, aparentemente ocultas, dos modelos de incerteza. Primeiramente, é apresentada a diferença entre incerteza local e espacial. Seu entendimento é fundamental para a construção adequada dos modelos de incerteza, de acordo com a finalidade desejada. A seguir, é apresentado o algoritmo de krigagem multiGaussiana e sua relação com simulação seqüencial estocástica, destacando suas diferenças e aplicabilidades. O conceito de função de transferência e propagação da incerteza, por meio dessas funções, são abordados nesse capítulo, mostrando como construir de maneira adequada os modelos de incerteza resultantes. Na seqüência, são apresentadas ferramentas matemáticas e técnicas que permitem avaliar a qualidade dos modelos de incerteza. Finalmente, são apresentados os índices de incerteza que possibilitam medir a magnitude de incerteza associada às estimativas da FA $Z(u)$ (ou de uma FA resposta).

4.1. Incerteza local x incerteza espacial

Como discutido no capítulo anterior, a incerteza associada à estimativa em localizações específicas é modelada por meio das distribuições locais de probabilidade, as dclp (modelos de incerteza). Essas distribuições indicam a probabilidade condicional de o valor da VA Z , na localização u , ser inferior ou igual a quaisquer valores limites z , ou seja:

$$F(u; z | (n)) = \text{Pr ob}\{Z(u) \leq z | (n)\} \quad (4.1)$$

Em algumas aplicações, além da estimativa local da probabilidade condicional da VA $Z(u)$ em relação a um dado valor limite z , pode ser muito importante verificar a probabilidade condicional de um conjunto de localizações u_j ($j = 1, \dots, J$) ultrapassar o valor limite z . Por exemplo, em uma aplicação ambiental, para delinear o caminho de percolação de contaminantes, é crucial estimar a probabilidade de que a concentração, em um conjunto de localizações, extrapole um dado valor limite z . Nessas aplicações, se torna necessário modelar a incerteza espacial de $Z(u)$. Por exemplo, a modelagem da incerteza espacial em duas localizações u_a e u_b , exigiria a inferência da probabilidade condicional conjugada (em u_a e u_b), ou seja, seria necessária a estimativa da seguinte dclp:

$$F(u_a, u_b; z_a, z_b | (n)) = \text{Prob}\{Z(u_a) \leq z_a, Z(u_b) \leq z_b | (n)\} \quad (4.2)$$

Se as localizações u_a e u_b não são espacialmente correlacionadas (se, por exemplo, o vetor h entre u_a e u_b for maior do que o alcance do semi-variograma) a estimativa dessa dclp pode ser diretamente determinada por meio do produto das dclp locais em u_a e u_b . Nesse caso:

$$F(u_a, u_b; z_a, z_b | (n)) = F(u_a; z_a | (n)) \cdot F(u_b; z_b | (n)) \quad (4.3)$$

Entretanto, na presença de correlação espacial, a probabilidade condicional conjugada deve ser estimada a partir de métodos analíticos ou estocásticos.

Nos métodos analíticos, uma função conhecida é adotada para modelar a distribuição multivariada da FA $Z(u)$. Para isso, comumente é empregado o modelo multiGaussiano. Sob esse modelo, a distribuição de probabilidade condicional conjugada é totalmente caracterizada pela média e a variância de cada dclp, mais a covariância entre u_a e u_b (Goovaerts, 2001). Entretanto, a modelagem da incerteza espacial a partir dos métodos analíticos é, principalmente, dificultada por dois fatores:

- i. o número N de localizações consideradas; pois a estimativa da incerteza espacial se torna muito complexa a medida que o número N aumenta;
- ii. o processo de validação da hipótese de biGaussianidade assumida pelo modelo; na prática essa tarefa se torna muito trabalhosa e difícil.

A estimativa da incerteza espacial, considerando a interdependência das localizações u_a e u_b , a partir do método estocástico, consiste na análise de um conjunto de L pares de valores correlacionados $\{z^{(l)}(u_a), z^{(l)}(u_b)\}$, $l = 1, \dots, L$. A partir desse conjunto de valores, a probabilidade condicional conjugada é estimada da seguinte forma:

$$F(u_a, u_b; z_a, z_b | (n)) \approx \frac{1}{L} \sum_{l=1}^L i^{(l)}(u_a; z_a) \cdot i^{(l)}(u_b; z_b) \quad (4.4)$$

sendo que:

$$i^{(l)}(u_j; z_j) = \begin{cases} 1 & \text{se } z^{(l)}(u_j) < z_j \\ 0 & \text{se não} \end{cases} \quad \text{com } j = a, b \quad \text{e } l = 1, \dots, L$$

De maneira geral, é recomendado que o número L de pares de valores necessários para modelar adequadamente a probabilidade condicional conjugada não seja inferior a 100 ($L \geq 100$).

Geralmente, a modelagem da incerteza espacial, com respeito a poucas localizações (como no exemplo acima) é de pouco interesse. Na maioria das vezes, essa incerteza é modelada com respeito a um número significativo de localizações. Dessa forma, o procedimento de estimativa da incerteza espacial, por meio do método estocástico, pode ser diretamente expandido para qualquer número J de localizações:

$$\text{Prob}\{Z(u_j) \leq z_j; j = 1, \dots, J | (n)\} \approx \frac{1}{L} \sum_{l=1}^L \prod_{j=1}^J i^{(l)}(u_j, z_j) \quad (4.5)$$

Os valores $z^{(l)}(u_j)$ ($j = 1, \dots, J$ e $l = 1, \dots, L$) são simulações do (desconhecido) valor real da VA Z nas respectivas J localizações. Normalmente, a distribuição espacial de $Z(u)$ é simulada em todo domínio de interesse, ou seja, em todas as N células ou blocos u_j ($j = 1, \dots, N$) contidos na área A . Assim, a estimativa da incerteza espacial, com respeito a um conjunto de J localizações, pode ser prontamente determinada.

Mas como é possível obter o conjunto de L valores nas N localizações desejadas, de modo que a correlação apresentada entre as localizações seja mantida? Essa tarefa pode ser realizada por meio dos métodos de simulação seqüencial estocástica. Esses métodos produzem um conjunto de mapas consistentes com os dados amostrados, de modo que os conjuntos de dados simulados (que constituem os mapas) obedecem ao mesmo modelo de correlação espacial dos dados condicionantes. Assim, a partir da análise do conjunto de dados simulados, é possível medir a incerteza associada, de forma realista, considerando a heteroscedasticidade inerente aos dados.

Na verdade, os métodos de simulação seqüencial estocástica fazem uso de um algoritmo no mínimo curioso, na medida em que o objetivo é determinar um modelo de incerteza estocástico, conhecendo-se (de antemão) o modelo teórico de incerteza, por meio de krigagem.

4.2. Krigagem multiGaussiana

Se o objetivo almejado não requerer a inferência da incerteza espacial e for somente estimar-se a incerteza em localizações específicas u , o uso do método de krigagem multiGaussiana (Verly, 1983), em um contexto paramétrico, pode ser mais adequado do que ssG. Isso porque esse método, como será mostrado a seguir, produz modelos locais de incerteza muito similares àqueles obtidos por meio de ssG, com muito menor esforço computacional. A krigagem multiGaussiana utiliza o embasamento teórico da lei Gaussiana condicional local, permitindo o cálculo exaustivo das dclp $F(u; z|(n))$, por meio da krigagem dos valores experimentais Gaussianos $y(u_\alpha)$ ($\alpha = 1, \dots, n$), de $Y(u)$ a partir da transformação Gaussiana de $Z(u)$ – Capítulo 3, item 3.5.1.

Sabendo que a função de anamorfose $\varphi(\cdot)$ é bijetiva para qualquer quantil q_p (p definido no intervalo $[0;1]$), logo

$$Z_{q_p}(u) = \varphi(Y_{q_p}(u)) \quad (4.7)$$

ou seja: o quantil no espaço original de $Z(u)$ é equivalente ao mesmo quantil no espaço Gaussiano, de acordo com a função de anamorfose $\varphi(\cdot)$. O quantil Gaussiano é conhecido e se deduz classicamente da lei condicional:

$$y_{q_p}(u) = y^*(u) + y \cdot \sigma_k(u) \quad (4.8)$$

onde,

$y^*(u)$ é o valor estimado por krigagem, a partir das informações transformadas (Gaussianas) $y(u_\alpha)$

$$(\alpha = 1, \dots, n),$$

$\sigma_k(u)$ é o desvio padrão de krigagem e

Y é obtido pelo quantil Gaussiano, a partir da dp Gaussiana padrão G .

O valor assumido por y representa uma realização da VA Y padrão [$Y \sim N(0,1)$], de função de repartição $G(\cdot)$ conhecida. Os valores y , referentes a quaisquer quantis, podem ser imediatamente inferidos por meio de tabelas da distribuição cumulativa normal (Equação 3.12), disponível em grande parte da literatura especializada (exemplo em Myers, 1996, p. 137-139). Assim, para qualquer probabilidade p (dentro do domínio $[0;1]$) existe um valor Gaussiano y associado, conhecido de acordo com $G^{-1}(p)$.

Portanto, o intervalo de probabilidade $[0;1]$ da dclp em u pode ser densamente discretizado, de modo que a mesma possa ser exaustivamente inferida, a partir da estimativa de $y^*(u)$ via krigagem (KS) e de seu desvio padrão de krigagem.

4.3. Simulação estocástica x krigagem multiGaussiana

Os métodos estocásticos reproduzirão tão bem o espaço de incerteza (representado pela dclp) quanto maior for o número L de realizações executadas. Porém, na prática, esse número é função da distribuição experimental, necessitando ser maior ou menor, conforme as características intrínsecas dos dados. O importante é que L seja tal que descreva suficientemente bem o espaço de incerteza, dependendo do objetivo almejado. Bonato (2000c), Bonato *et al.* (2000a) e Bonato *et al.* (2000b) apresentam um estudo de caso que relaciona o número L de realizações com a caracterização do espaço de incerteza.

Teoricamente, a medida que o número de realizações tende ao infinito ($L \rightarrow \infty$), a distribuição local dos valores simulados, por métodos estocásticos, tende à distribuição obtida por meio de krigagem. A Figura 4.1 mostra um caso em que a dclp foi modelada por meio de krigagem (krigagem multiGaussiana) e via simulação estocástica (ssG). Nesse exemplo, a dclp via krigagem multiGaussiana (exaustiva) foi aproximada numericamente por meio da geração de 100 simulações (ssG). A figura mostra que a diferença entre os dois métodos é mínima.

Portanto, fica claro que ambos procedimentos constroem modelos de incerteza local muito similares. Porém, pela facilidade, a determinação das dclp por meio de krigagem sempre leva vantagem quando o objetivo for estimar a probabilidade condicional local. Para esses casos, a aplicação do método de simulação estocástica acarretaria em desperdícios de tempo de processo e de armazenagem em disco. Entretanto, existem duas situações em que o emprego do método de simulação estocástica se torna vantajoso em relação à krigagem:

- i. como no caso descrito acima, no qual se deseja estimar a probabilidade condicional conjugada (incerteza espacial);

- ii. quando se requer a modelagem da incerteza sobre um suporte muito maior do que o suporte de medida.

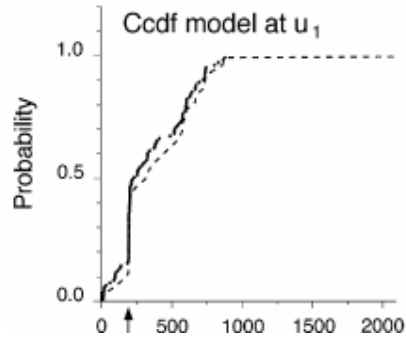


Figura 4.1: Modelos de incerteza local inferidos por krigagem multiGaussiana (linha tracejada) e por meio de simulação estocástica (linha pontilhada). Fonte: Goovaerts (2001).

Por praticidade, a partir desse momento, sempre será referido o termo *estocástico* quando as dclp forem inferidas a partir dos métodos de simulação estocástica.

4.4. Suporte das medidas x suporte dos modelos de incerteza

Freqüentemente, em muitas aplicações, o espaço de incerteza precisa ser estimado em suportes muito maiores em relação ao suporte de medida, seja para estimar a incerteza em relação ao teor de um bloco de minério ou para verificar a probabilidade condicional de uma área exceder uma dada concentração de um certo contaminante. Nesses casos, o objetivo é estimar a probabilidade condicional de que o valor médio de um bloco de tamanho V , centrado na localização u , seja inferior a dado valor limite z . Portanto, é preciso estimar a dcp (distribuição condicional de probabilidade) em relação ao bloco V :

$$F_V(u; z | (n)) = \text{Pr ob}\{Z_V(u) < z | (n)\} \quad (4.9)$$

Para a realização dessa tarefa, três técnicas podem ser utilizadas:

- i. por meio da determinação das dcp via krigagem de bloco;
- ii. por meio do método de correção de variância (Isaaks e Srivastava, 1989, p. 468) e posterior aplicação dos métodos baseados em krigagem (pontual) de determinação de dcp;
- iii. por meio dos métodos estocásticos.

A técnica (i) é realizada de maneira análoga àquela aplicada para a determinação das dclp no mesmo suporte das medidas. Contudo, agora, a KS é substituída por krigagem simples de bloco. Se os modelos de incerteza forem baseados no modelo multi-Gaussiano, os mesmos passam a ser caracterizados pela krigagem simples de bloco (média) e pela variância dessa krigagem (variância) (Chilès e Delfiner, 1999). Entretanto, essa técnica é limitada aos casos em que os atributos em estudo variam linearmente no espaço, o que pode não ser muito freqüente.

A técnica (ii) não apresenta a limitação mostrada pela técnica (i). Nesse caso, a variância da dclp $F(u; z | (n))$ é corrigida com relação à suposta variância da dcp de bloco $F_V(u; z | (n))$. Essa correção envolve a redução da variância, enquanto que a forma de distribuição pode se manter inalterada ou não, dependendo do algoritmo empregado para a correção da variância.

A razão entre as variâncias medidas, no suporte original ($\text{Var}[Z(u)|(n)]$) e no suporte de bloco ($\text{Var}[Z_V(u)|(n)]$), é chamado de fator de redução de variância f .

$$f = \frac{\text{Var}[Z_V(u) | (n)]}{\text{Var}[Z(u) | (n)]} \quad (4.10)$$

Pela relação de Krige (Armstrong *et al.*, 1997, p. 63), a variância total é igual à adição entre a variância intrablocos e a variância entre blocos:

$$\sigma^2(. / V) = \sigma^2(. / v) + \sigma^2(v / V) \quad (4.11)$$

onde:

$\sigma^2(. / v)$ Variância das medidas em relação aos blocos;

$\sigma^2(v / V)$ Variância dos blocos em relação ao todo;

$\sigma^2(. / V)$ Variância das medidas (suporte original) em relação ao todo.

Dessa forma, f pode ser calculado da seguinte maneira:

$$f = \frac{\sigma^2(v / V)}{\sigma^2(. / V)} = \frac{\sigma^2(. / V) - \sigma^2(. / v)}{\sigma^2(. / V)} = 1 - \frac{\sigma^2(. / v)}{\sigma^2(. / V)} \quad (4.12)$$

Ou

$$f = 1 - \frac{\bar{\gamma}(V, V)}{\bar{\gamma}(A, A)} \leq 1$$

onde $\bar{\gamma}(V, V)$ representa o semi-variograma médio em relação a pares aleatoriamente posicionados dentro do bloco de dimensão V . De forma similar, $\bar{\gamma}(A, A)$ representa o semi-variograma médio em relação a área em estudo A . Na maioria das vezes, se a área em estudo A é grande em relação ao alcance do semi-variograma, $\bar{\gamma}(A, A)$ pode ser aproximado ao patamar do semi-variograma dos dados originais $z(u_\alpha)$ ($\alpha = 1, \dots, n$). Portanto, o fator f pode ser estimado com o auxílio do modelo de continuidade espacial das informações originais.

Dependendo da magnitude do fator de redução de variância f , a dcp de bloco pode ser inferida, a partir da dclp, por meio de diferentes algoritmos que diferem no grau de simetria assumida pela dcp de bloco. Geralmente, dois algoritmos são empregados:

- ii.i. correção afim;
- ii.ii. correção lognormal indireta.

Ambos algoritmos [(ii.i.) e (ii.ii.)] compartilham a característica de preservar a média da distribuição experimental (Isaaks e Srivastava, 1989, p. 468). O efeito de simetrização (Myers, 1996, p. 314) está associado ao processo de redução de variância, de acordo com o algoritmo utilizado. Esse

efeito se refere à redução da assimetria da distribuição após a mudança de suporte, aproximando-se de uma distribuição normal. O algoritmo de correção afim apresenta a característica de preservar a forma da distribuição. Por outro lado, o algoritmo de correção lognormal indireta assume que existe uma relação entre a redução da variância e o aumento da simetria da distribuição.

A simetria também é relacionada à homogeneidade da distribuição (dados espacialmente desaninhados). O aumento da simetria em relação à diminuição da variância ocorre mais rapidamente em populações homogêneas. A simetria é alcançada mais lentamente em distribuições cujos valores extremos estão aninhados em uma região particular da área em questão, em relação a outros onde esses dados estão mais homogeneamente distribuídos no espaço. Ainda, a homogeneidade da distribuição influi na taxa de redução da variância, ocorrendo mais lentamente em distribuições contínuas do que em distribuições erráticas. Assim, para distribuições menos contínuas o efeito da mudança de suporte sobre o espalhamento da distribuição é maior. Portanto, o grau de continuidade espacial, evidenciado pelo semi-variograma, pode ser usado como indicativo sobre o impacto do suporte original com respeito à redução da variância e simetria da distribuição.

Uma das maiores vantagens do algoritmo de correção afim é a sua simplicidade matemática, apresentada abaixo:

$$x' = \sqrt{f}(x - m) + m \quad (4.13)$$

onde:

f é o fator de redução da variância;

x e x' são os decis ou percentis, originais e após a transformação, respectivamente;

m é a média da distribuição original.

O algoritmo de correção afim reduz a variância da distribuição por meio da compressão da distribuição em torno da média, preservando a forma da distribuição original. Esse algoritmo apresenta a característica de atribuir valores muito maiores para os menores valores da distribuição experimental e vice-versa, podendo provocar classificações equivocadas em situações onde o teor de corte (TC) for muito distante da média. Essa característica pode ser muito indesejada em alguns casos, como por exemplo, quando se trata de dados de concentração de contaminantes. Além disso, a característica de manter inalterada a forma da distribuição só é realística para blocos de pequena dimensão, onde a diminuição da variância é baixa. De maneira geral, esse algoritmo se mostra apropriado quando o TC for próximo à média e a redução da variância for inferior a 30 %, isto é, $f > 0,70$.

O algoritmo de correção lognormal indireta, apresentado a seguir, tenta minimizar as características indesejáveis apresentadas pelo algoritmo de correção afim, sendo de utilização mais adequada em situações onde o valor limite z estiver inserido no quartil inferior da distribuição experimental. A correção lognormal indireta mantém o valor mínimo; entretanto, altera a forma da distribuição, tornando-a mais simétrica. Esse algoritmo ajusta a distribuição, conforme a variância de blocos, em um processo de duas etapas, pois o método não garante que a média da distribuição

ajustada seja a mesma da original. Essa condição só é garantida se a distribuição a ser ajustada for exatamente lognormal. A correção lognormal indireta é dada pelas seguintes equações:

$$x' = ax^b \quad (4.14)$$

onde:

$$a = \frac{m}{\sqrt{f CV^2 + 1}} \left(\frac{\sqrt{CV^2 + 1}}{m} \right)^2$$

$$b = \sqrt{\frac{\ln(f CV^2 + 1)}{\ln(CV^2 + 1)}}$$

sendo que:

f é o fator de redução da variância, e

CV é coeficiente de variação.

Como descrito anteriormente, para situações em que distribuição original não seguir uma distribuição lognormal, faz-se necessário o reescalamento dos valores oriundos da transformação dada pela equação acima. Esse reescalamento é realizado por meio da seguinte equação:

$$x'' = \frac{m}{m'} x' \quad (4.15)$$

sendo,

m a média da distribuição original;

m' a média da distribuição após a transformação pela Equação 4.14;

x'' e x' os decis ou percentis reescalados e após a transformação pela Equação 4.14, respectivamente.

A magnitude desse reescalamento é inversamente proporcional à similaridade da distribuição original a uma distribuição lognormal, com a razão m/m' mais próxima de um à medida que a distribuição original se aproxima de uma distribuição lognormal. Com esse procedimento, a distribuição é ajustada de acordo com o suporte experimental, tendo como média o mesmo valor da distribuição original.

A técnica (iii), que utiliza o método estocástico, é uma técnica mais versátil em relação às outras duas apresentadas acima, embora requeira mais esforço computacional. Nesse método, as dcp teóricas de bloco são aproximadas numericamente por meio do conjunto de L valores médios $z_v^{(l)}(u)$ ($l = 1, \dots, L$) obtidos pela média dos valores simulados (no suporte original das medidas) no interior dos blocos. Isto é, os blocos de dimensão V são discretizados em M localizações u_β , de modo que o valor assumido pelos blocos seja o resultado da média aritmética dos valores simulados intrablocos; ou seja:

$$z_v(u) = \frac{1}{M} \sum_{\beta=1}^M z(u_\beta) \quad (4.16)$$

A dcp estocástica de bloco, em cada localização u , é estimada pela repetição de L vezes o processo descrito acima; isto é:

$$F_V(u; z | (n)) \approx \frac{1}{L} \sum_{l=1}^L i_V^{(l)}(u; z) \quad (4.17)$$

onde

$$i_V^{(l)}(u; z) = \begin{cases} 1 & \text{se } z_V^{(l)}(u) \leq z \\ 0 & \text{se não} \end{cases} \quad \text{com } l=1, \dots, L$$

Segundo Goovaerts (2001), as principais características que tornam a aplicação do método estocástico mais vantajosa são:

- i. o método permite considerar funções de ponderação não-linear, como média harmônica ou geométrica;
- ii. ausência de hipóteses com relação ao impacto da mudança de suporte sobre a forma e a variância da distribuição;
- iii. a maneira de construção dos modelos de incerteza de blocos permite com que a metodologia seja facilmente aplicada para blocos de diferentes formas e dimensões.

Portanto, pelas características citadas acima, o método estocástico se apresenta como uma boa alternativa para os casos em que o espaço de incerteza necessita ser medido em um suporte muito maior do que suporte de medida. Além disso, dentre as três técnicas relacionadas, o método estocástico é o de mais simplicidade teórica, não assumindo hipóteses e sem restrições (teóricas) de aplicação. Contudo, dependendo do número de nós discretizando os blocos e da característica dimensional do problema em mãos (2-D ou 3-D), o emprego do método pode ocasionar dificuldades operacionais (por exemplo: elevado tempo de processo e alto consumo de espaço em disco). Nesses casos, a aplicação das duas primeiras técnicas relacionadas deve ser reavaliada.

4.5. Funções de transferência

Raramente a estimativa do espaço de incerteza, em relação aos valores desconhecidos do atributo medido, é o objetivo de um estudo geoestatístico. Ao contrário, a inferência desses modelos constitui uma etapa da análise geoestatística. Frequentemente, os modelos de incerteza alimentam funções de transferência (FTs), cujos resultados auxiliam ou determinam o posterior processo de tomada de decisão. Por exemplo, a relação estéril/minério (em uma situação mineira) e o pH (em uma situação ambiental) podem ser considerados como sendo funções de transferência. Nesses exemplos, a relação estéril/minério auxiliaria na delimitação de blocos economicamente viáveis à lavra, enquanto que o pH auxiliaria como um indicativo da qualidade do solo.

Uma FT se caracteriza como um modelo numérico ou analítico que descreve ou representa algum sistema ou operação real (Isaaks, 1990, p. 7). Uma operação mineira pode ser considerada

como uma FT, onde certas variáveis podem ser processadas para produzir variáveis resposta que configuram a operação mineira. Por exemplo, na indústria de petróleo, um simulador de fluxo pode ser uma função de transferência (Hewett, 1986) que descreve um sistema real de um reservatório de petróleo com poços de produção. Essa FT poderia ser alimentada pela distribuição espacial de permeabilidade, porosidade ou saturação. Nesse caso, as respostas seriam: taxa de produção, tempo de recarga e eficiência (Journel e Gomez-Hernandez, 1989). Exemplos de operações mineiras que podem ser modeladas por FTs incluem:

- i. controle de teores (Isaaks, 1990; Srivastava *et al.*, 1992; Glacken, 1996),
- ii. determinação de dimensões ótimas de unidades seletivas de lavra e números de faces simultâneas (Remacre, 1992),
- iii. custos de diluição mediante variações de parâmetros de lavra (Dimitrakopoulos, 1997),
- iv. planejamento a curto prazo (Costa *et al.*, 1997),
- v. reservas recuperáveis (Costa, 1997),
- vi. determinação de cava ótima (Rossi e Van Brunt, 1997),
- vii. programas de amostragem (sondagem) adicional (Pilger, 2000; Pilger *et al.*, 2000; Pilger *et al.*, 2001 e Pilger *et al.*, 2002).

Os algoritmos de simulação condicional permitem a criação de vários cenários equiprováveis, condicionados a distribuição original (amostras). Essas simulações condicionais alimentam funções de transferência (FTs) que permitem gerar várias distribuições resposta $T(u)$. A análise dessas distribuições possibilita quantificar a incerteza em relação à FA que representa a FT, de acordo com a incerteza das variáveis originais. Associado à FT, também existe uma lei espacial, representada por um modelo FA, que caracteriza o processo físico descrito pela FT. Assim, os valores assumidos pelas variáveis resposta nas N localizações u_j ($j = 1, \dots, N$) do grid também são considerados realizações de VAs. Como cada valor que alimenta a FT é equiprovável e condicionado aos dados originais, cada resposta (constituente da distribuição resposta) também é equiprovável e condicional aos dados, exceto quando a FT é probabilística.

4.6. Propagação da incerteza por meio de funções de transferência

Sabendo que existe incerteza associada aos valores desconhecidos do atributo investigado nas localizações alvo, é natural esperar que essa incerteza se propague através das funções de transferência. Conseqüentemente, as respostas (resultados das funções de transferência) também conterão um certo grau de incerteza. Duas classes de funções de transferência podem ser distinguidas:

- i. funções locais de transferência;
- ii. funções globais de transferência.

As **funções locais de transferência** são aquelas que são aplicadas em localizações específicas u ou blocos $V(u)$. As **funções globais de transferência** são mais complexas, apresentando a característica de serem alimentadas por valores simulados $z^{(l)}(u_j)$ ($j = 1, \dots, J$ e $l = 1, \dots, L$) de $Z(u)$ de diferentes localizações ao mesmo tempo.

A propagação da incerteza através dessas funções pode ser estimada por meio de métodos analíticos ou estocásticos. Heuvelink e Pebesma (1999) apresentam detalhes sobre a propagação da incerteza local estimada por meio de métodos analíticos. Embora possível, a utilização de métodos analíticos para a análise da propagação de incerteza associada a FTs globais não é recomendada. Geralmente, a análise da propagação da incerteza é realizada por meio dos métodos estocásticos, que não assumem hipóteses sobre a forma das distribuições dos resultados das funções de transferência. Esses métodos utilizam o conjunto constituído por L valores simulados $z^{(l)}(u)$ ($l = 1, \dots, L$) para alimentar as funções de transferência. Assim, com as distribuições compostas pelos L valores resposta $T^{(l)}(u)$ ($l = 1, \dots, L$) nas localizações alvo constrói-se os modelos locais de incerteza das VAs resposta, associadas ao atributo resposta de interesse. Quando se trata de funções globais de transferência, as FTs são alimentadas por várias realizações da distribuição espacial da FA $Z(u)$. Ou seja, os valores simulados $z^{(l)}(u_j)$ ($j = 1, \dots, J$ e $l = 1, \dots, L$) alimentam as funções globais de transferência (como simuladores de fluxo em engenharia do petróleo, por exemplo) para a construção da distribuição dos valores resposta.

Geralmente, os métodos estocásticos utilizam a técnica de simulação de Monte Carlo (Isaaks, 1990, p. 7; Capítulo 3, seção 3.1), na qual as dclp das VAs Z nas localizações u são amostradas aleatoriamente L vezes, construindo o conjunto de valores simulados $\{z^{(l)}(u) \mid l = 1, \dots, L\}$. Como abordado anteriormente, no método estocástico de construção das dclp, o número L de tiragens aleatórias deve ser tal que garanta que o espaço de incerteza da VA resposta seja caracterizado. Entretanto, freqüentemente, L precisa ser tão grande que essa técnica pode se tornar computacionalmente muito intensa, exigindo muito tempo e espaço de CPU, até que o espaço de incerteza seja suficientemente caracterizado.

4.6.1. Latin Hypercube Sampling (LHS)

Como alternativa ao método de Monte Carlo, uma estratégia mais eficiente pode ser adotada para analisar a propagação da incerteza associada a FTs locais. Para uma dada precisão, a aplicação do método *Latin Hypercube Sampling* (LHS) (McKay *et al.*, 1979) garante que toda a dclp seja representada, reproduzindo a construção do conjunto de L valores resposta $\{T^{(l)}(u) \mid l = 1, \dots, L\}$, em muito menos tempo do que o método de Monte Carlo. O método LHS consiste na tiragem aleatória, sem substituição (amostragem aleatória estratificada sem substituição), de M valores da VA $Z(u)$ a partir das dclp. Contudo, esses M valores devem ser oriundos de M classes equiprováveis distintas. Isto é, primeiramente cada dclp é dividida em M classes equiprováveis disjuntivas. O procedimento de tiragem respeita a equação abaixo:

$$Z^{(m)}(u) = [F^{-1}(u, ((m - 1) + R_m)/M)]^* \quad m = 1, \dots, M \quad (4.18)$$

onde R_m é um número aleatório do domínio $[0;1]$.

A técnica LHS faz com que a dclp seja inteiramente representada, sobretudo em seus trechos extremos. Desse modo, com a aplicação dos M valores na função de transferência, são criados conjuntos de M valores resposta $\{T^{(m)}(u) \ (m = 1, \dots, M)\}$ que caracterizam o espaço de incerteza da VA resposta nas localizações alvo.

A incerteza em relação a outras Q VAs $Y(u)$ envolvidas na função transferência também pode ser considerada. Basta que o espaço de incerteza associado aos valores desconhecidos dessas variáveis, nas localizações alvo, seja caracterizado. De maneira análoga, o método LHS também pode ser empregado para que o conjunto de valores simulados $\{y^{(m)}(u) \ (m = 1, \dots, M)\}$ seja construído. A partir daí, os conjuntos de valores simulados $\{z^{(m)}(u) \ (m = 1, \dots, M)\}$ e $\{y_{(q)}^{(m)}(u) \ (m = 1, \dots, M), \ (q = 1, \dots, Q)\}$ são combinados na alimentação da função de transferência, de modo que possa ser gerado os M^{Q+1} possíveis cenários em relação à VA resposta. Entretanto, é recomendado que o número Q de variáveis auxiliares não seja superior a cinco, visto que a quantidade de combinações se torna rapidamente proibitiva.

Quando os parâmetros variáveis da função de transferência são correlacionados, suas distribuições não podem ser amostradas independentemente, pois a correlação entre os mesmos deve ser reproduzida. Nesses casos, técnicas particulares de amostragem (Viscarra Rossel *et al.*, 2001) devem ser utilizadas. Entretanto, Davis (1987) propôs uma metodologia para fazer com que a correlação entre os parâmetros seja mantida pelos valores aleatoriamente tirados das respectivas distribuições. Essa metodologia é descrita por meio das três seguintes etapas:

- i. as distribuições experimentais dos Q parâmetros, que devem seguir um comportamento normal ($Y \sim N(0,1)$), são amostradas segundo o método LHS para que sejam geradas M^Q combinações. Se, por exemplo, dois parâmetros variáveis estivessem envolvidos, a e b , dois conjuntos de M valores aleatórios seriam combinados em M^2 pares $[v^{(l)}, w^{(l)}]$, $l = 1, \dots, M^2$;
- ii. a matriz de variância-covariância C dos parâmetros é decomposta no produto das matrizes triangulares superior e inferior: $C = LU$ (decomposição de Cholesky (Press *et al.*, 1992, p. 96-98));
- iii. os M^Q valores $[a^{(l)}, b^{(l)}]$, $l = 1, \dots, M^Q$ (correlacionados) são obtidos por meio da multiplicação dos pares simulados $[v^{(l)}, w^{(l)}]$ pela matriz triangular inferior (L) mais o vetor das médias dos parâmetros. Seguindo o exemplo acima, quando $Q = 2$:

$$\begin{bmatrix} a^{(l)} \\ b^{(l)} \end{bmatrix} = \begin{bmatrix} L_{11} & 0 \\ L_{12} & L_{22} \end{bmatrix} \cdot \begin{bmatrix} v^{(l)} \\ w^{(l)} \end{bmatrix} + \begin{bmatrix} m_a \\ m_b \end{bmatrix} \quad (4.19)$$

Portanto, essa metodologia transforma os pares de valores simulados $[v^{(l)}, w^{(l)}]$, independentemente sorteados por meio do método LHS, em pares $[a^{(l)}, b^{(l)}]$, $l = 1, \dots, M^Q$ que

reproduzem a média de cada parâmetro, assim como a correlação entre eles. A única hipótese assumida é a de que os parâmetros sigam uma distribuição Gaussiana.

Nos casos em que se torna necessário trabalhar com FTs globais, a aplicação de método LHS é teoricamente inviável devido à correlação entre os dados. Nesses casos, a caracterização do espaço de incerteza da FA $T(u)$ deve ser realizada com auxílio do método de Monte Carlo, sendo exclusivamente, função do número L de realizações geradas. Dessa forma, como não existe garantia de que o processo de simulação de Monte Carlo, a partir de L tiragens aleatórias, consiga representar todo o espaço de incerteza presente nas diferentes localizações, o número L deve ser determinado empiricamente. Goovaerts (1999), Bonato (2000c), Bonato *et al.* (2000a), Bonato *et al.* (2000b) e Deutsch (2002) descrevem procedimentos para avaliar o impacto do número L de realizações sobre a exatidão e a precisão da distribuição resposta estimada.

Na comparação entre o método paramétrico e o não-paramétrico de estimativa de d_{clp} , a d_{clp} mais precisa, obtida por meio de um método, não necessariamente reflete essa maior precisão quando se analisa o espalhamento da variável resposta (resultado da função de transferência). Evidentemente, para que se possa medir a precisão dos modelos de incerteza é necessário que os valores da VA investigada $Z(u)$ e da VA resposta, nas localizações alvo, sejam conhecidos – como será abordado no item 4.10.

4.7. LHS x simulação de Monte Carlo

Sempre que for preciso estimar o espaço de incerteza em relação à FA resposta, oriundo da propagação da incerteza de $Z(u)$ como parâmetro de funções de transferência locais, o método **LHS** (combinado com a modelagem da incerteza via krigagem multiGaussiana) deve ser a primeira alternativa. Nesses casos, esse método produz resultados similares àqueles obtidos por meio do procedimento de **simulação de Monte Carlo**, consumindo muito menos tempo e espaço de CPU. Entretanto, a técnica de simulação de Monte Carlo se apresenta como uma boa alternativa nas três seguintes situações:

- i. como descrito na seção 4.6.1, quando funções globais de transferência necessitam ser aplicadas;
- ii. quando as funções de transferência são conhecidas e disponíveis para um suporte muito diferente do que o suporte de medida;
- iii. quando for solicitado o cálculo do histograma dos valores resposta médios (e-type) em relação a toda a área em estudo.

4.8. Aumento da eficiência das técnicas LHS e de Monte Carlo

Na caracterização do espaço de incerteza da variável resposta, é normal que seja dada atenção especial aos valores extremos que essa variável pode assumir. Normalmente, esses valores são derivados de valores situados nos trechos extremos das dclp das variáveis parâmetros da função de transferência. Seguindo esse raciocínio, Manteufel (2000) propôs que o tratamento oferecido às dclp das variáveis parâmetros seja diferenciado, de acordo com o trecho e a característica (normal, lognormal, exponencial, etc.) das funções de transferência, FTs.

O algoritmo do processo de simulação de Monte Carlo consiste na tiragem aleatória de L números r do domínio [0;1] (representando o domínio de probabilidade) da função de repartição [F(u, z| (n))]. A partir daí, os valores das VAs parâmetros são obtidos por meio da aplicação do valor de r na função de repartição inversa, ou seja:

$$Z^{(l)}(u) = [F^{-1}(u, r_l)] \quad l = 1, \dots, L \quad (4.20)$$

Manteufel (2000) sugere que a importância dos trechos das dclp seja considerada por meio da transformação do número r sorteado. Essa transformação gera um novo número s aleatório, também definido entre zero e um. Assim, a inversão da função de repartição é realizada a partir de s [s(r)]. O autor propõe que essa transformação seja baseada em uma função polinomial, apresentada abaixo:

$$s_l(u) = r_l \left(r_l - 1 \right)^2 \cdot \left. \frac{ds}{dr} \right|_0 + r_l^2 \left(3 - 2r_l \right)^2 + r_l^2 \left(r_l - 1 \right)^2 \cdot \left. \frac{ds}{dr} \right|_1 \quad l = 1, \dots, L \quad (4.21)$$

Nessa equação, o termo $\left. \frac{ds}{dr} \right|_{\beta}$ ($\beta = 0,1$) representa o parâmetro de importância (PI) e varia de

acordo com a característica das distribuições (do modelo de FT). Por exemplo, o quadro abaixo (Quadro 4.1) apresenta os valores assumidos por PI para as distribuições que seguem um comportamento normal e lognormal.

Quadro 4.1: Valores assumidos por PI, conforme as características da dclp.

Distribuição \ PI	$\left. \frac{ds}{dr} \right _0$	$\left. \frac{ds}{dr} \right _1$
Normal	0	0
Lognormal	2,7	0

Para as funções de transferência que contam com múltiplos parâmetros variáveis Q, PI também pode ser usado para a determinação dos pesos dos vetores que alimentam as funções. Para cada parâmetro, o peso w de cada valor aleatoriamente tirado é calculado segundo a equação abaixo:

$$w_l(u) = \left(3r_l^2 - 4r_l + 1 \right) \cdot \left. \frac{ds}{dr} \right|_0 + r_l \left(6 - 6r_l \right) + \left(3r_l^2 - 2r_l \right) \cdot \left. \frac{ds}{dr} \right|_1 \quad l = 1, \dots, L \quad (4.22)$$

Em seguida, o cálculo do peso dos vetores é realizado por meio da multiplicação de cada peso individual de cada parâmetro; ou seja:

$$\hat{w}^j(u) = \prod_q w_q^j(u) \quad \text{sendo que, } j = 1, \dots, L^Q \text{ e } q = 1, \dots, Q \quad (4.23)$$

Contudo, para a aplicação nas funções de transferência, o peso atribuído aos vetores deve ser normalizado pelo somatório dos pesos de todos os vetores (todas as L^Q combinações possíveis); isto é:

$$w^j(u) = \frac{\hat{w}^j(u)}{\sum_j \hat{w}^j(u)} \quad \text{com } j = 1, \dots, L^Q \quad (4.24)$$

O procedimento descrito acima (consideração do parâmetro de importância PI) também pode ser combinado com a técnica LHS. Nesse caso, os números r são obtidos de maneira aleatória dentro do domínio definido por cada classe: $(i/M; i+1/M]$, $i = 0, \dots, M-1$.

O uso combinado dessas técnicas (LHS com PI e simulação de Monte Carlo com PI) pode ser muito eficaz, fazendo com que a convergência da média e da variância dos modelos (realizações) seja mais rápida. Ou seja, essa combinação possibilita reduzir o número de tiragens aleatórias (L ou M) necessárias à caracterização adequada do espaço de incerteza das distribuições das VAs resposta. Essas técnicas combinadas mostram resultados muito satisfatórios para funções de transferência de caráter multiplicativo, sobretudo se a função for de característica lognormal. Contudo, de maneira geral, o emprego de LHS sempre apresenta melhores resultados do que aqueles obtidos por meio da técnica de simulação de Monte Carlo. Segundo o estudo de caso de Manteufel (2000), a técnica LHS combinada com PI leva vantagem em relação à técnica LHS, pois, para um mesmo nível de exatidão, foram necessárias cerca de 10 vezes menos tiragens quando considerada a convergência da variância dos modelos.

Como alternativa à técnica LHS clássica, uma pequena variação pode ser adotada: fazer com que o processo de tiragem, dentro de cada uma das M classes seja determinístico, ou não aleatório. Ou seja, pode-se fazer, por exemplo, com que r seja a mediana de cada classe. De maneira geral:

$$r_m = \left[\left((m-p) / M \right) \right] \quad m = 1, \dots, M \quad (4.25)$$

onde p representa proporção da classe. Para a mediana de cada classe $p = 0,5$. A partir daí, a dclp estocástica é definida pela aplicação das M probabilidades r_m na função inversa:

$$z^{(m)}(u) = \left[F^{-1}(u, r_m) \right] \quad m = 1, \dots, M \quad (4.26)$$

Essa modificação torna o procedimento mais prático e, às vezes, até mais eficaz. Saliby (1980) propôs uma técnica muito semelhante, denominada de Amostragem Descritiva (AD). Essa técnica questiona o paradigma de que para descrever um comportamento aleatório as tiragens devam ser obrigatoriamente aleatórias, declarando que às vezes um processo de tiragem determinístico pode ser mais eficiente. A utilidade da técnica foi comprovada por meio de uma série de testes e

comparações (Saliby, 1990 e Saliby, 1997) mostrando resultados muito satisfatórios, tanto em termos estatísticos como computacionais.

4.9. Avaliação da qualidade dos modelos locais de incerteza

Os modelos locais de incerteza podem ser obtidos por meio de técnicas paramétricas ou não-paramétricas. A escolha do método para a estimativa das dclp pode ser determinante, pois para um mesmo banco de dados e para uma mesma localização espacial u , os modelos de incerteza resultantes podem ser muito diferentes. O exemplo apresentado por Goovaerts (2001) retrata essa situação. Essa escolha está associada às características intrínsecas da distribuição experimental ou com a qualidade das dclp, sendo que a técnica escolhida deve ser adequada às características estatísticas dos dados.

A qualidade da modelagem será tão boa quanto mais preciso for o modelo local de incerteza. A precisão de uma dclp modelada a partir de técnicas paramétricas é principalmente controlada pela variância de krigagem, que por sua vez é controlada pela configuração espacial das informações. Enquanto que a precisão das dclp, derivadas por métodos não-paramétricos, é principalmente controlada pelo domínio de valores $[z_{\min}; z_{\max}]$ na circunvizinhança de cada localização u a ser estimada.

Mas como é possível avaliar a qualidade dos modelos locais de incerteza? A qualidade relativa dos modelos só pode ser verificada se os valores verdadeiros da variável alvo, nas localizações a serem estimadas, forem conhecidos. À primeira vista, não é possível afirmar que uma técnica seja melhor do que a outra, até porque não existe “o melhor” modelo de incerteza. Contudo, existem técnicas que possibilitam medir a qualidade das dclp. Como na interpolação espacial, é possível comparar os modelos de incerteza com as próprias medidas. Dois métodos podem ser aplicados:

- i. validação-cruzada, onde as informações são temporariamente removidas, ou
- ii. *jackknife*, onde certas informações são excluídas do banco de dados.

Porém, a aplicação da segunda técnica requer um grande número n de medidas $z(u_\alpha)$ ($\alpha = 1, \dots, n$) disponíveis. Isso porque as informações remanescentes devem ser suficientes para que o modelo de continuidade espacial e as estimativas decorrentes e sejam representativos. Dessa forma, antes de se fazer qualquer uso das dclp, é conveniente que esses modelos de incerteza sejam testados, no sentido de avaliar o quão bem eles são capazes de capturar a incerteza associada às estimativas da VA Z nas localizações u desconhecidas.

O confronto dos modelos de incerteza $F[(u_\alpha; z|(n))]^*$ com o atributo $z(u_\alpha)$ conhecido em cada localização medida (no caso de validação-cruzada, $\alpha = 1, \dots, n$) ou nas localizações de teste (no caso de *jackknife*, $\alpha = 1, \dots, J$), permite a análise de um conjunto de intervalos simétricos de probabilidade p (p -IP), limitados pelos quantis $(1-p)/2$ e $(1+p)/2$ da dclp; isto é: $(F^{-1}(u_\alpha; (1-p)/2 | (n)), F^{-1}(u_\alpha; (1+p)/2 | (n)))$.

|(n))]. A Tabela 4.1 apresenta as probabilidades p e seus respectivos intervalos simétricos de probabilidade IP. Essa tabela mostra que a probabilidade p reflete a amplitude do intervalo IP, centrado na mediana.

Tabela 4.1: Intervalos de probabilidade p-IP, indicando as probabilidades p e seus intervalos IP associados.

P(%)	Intervalo p-IP
10	(0,45; 0,55]
20	(0,40; 0,60]
30	(0,35; 0,65]
40	(0,30; 0,70]
50	(0,25; 0,75]
60	(0,20; 0,80]
70	(0,15; 0,85]
80	(0,10; 0,90]
90	(0,05; 0,95]
100	(0,00; 1,00]

Por praticidade, a partir de agora, o conjunto de localizações onde Z(u) é conhecido e seus modelos de incerteza associados serão denotados pelo índice j e limitados por N (j = 1,...,N), independente da técnica de comparação utilizada (validação-cruzada ou jacKnife).

Como exemplo, o intervalo simétrico de probabilidade 0,40 (0,40-IP) é limitado pelos quantis q_{0,30} e q_{0,70}, ou seja: (F⁻¹(u_j; 0,30|(n)), F⁻¹(u_j; 0,70|(n))). Nesse caso, quando os modelos de incerteza F[(u_j; z|(n))] de cada localização medida ou de teste forem confrontados com as medidas de Z(u), a incerteza estaria corretamente modelada se 40% das informações medidas se inserissem no intervalo limitado por (F⁻¹(u_j; 0,30|(n)), F⁻¹(u_j; 0,70|(n))) das dclp. Ou seja, as probabilidades p são comparadas com a proporção $\bar{\xi}(p)$ de valores medidos que se inserem nos intervalos locais.

Na verdade, a análise dos intervalos simétricos de probabilidade p é uma maneira de comparar os modelos locais de incerteza com a função de repartição empírica da distribuição. A proporção de valores conhecidos que se inserem nos intervalos simétricos de probabilidade p pode ser calculada a partir da análise do conjunto de informações conhecidas de Z(u) e das dclp estimadas, disponíveis nas N localizações u_j. O cálculo da proporção $\bar{\xi}(p)$ é realizado da seguinte forma:

$$\bar{\xi}(p) = \frac{1}{N} \sum_{j=1}^N \xi(u_j; p) \quad \forall p \in [0,1] \tag{4.27}$$

sendo que,

$$\xi(u_j; p) = \begin{cases} 1 & \text{se } F^{-1} * (u_j; (1 - p) / 2) < z(u_j) \leq F^{-1} * (u_j; (1 + p) / 2) \\ 0 & \text{se não} \end{cases} \quad \forall p \in [0,1]$$

A solução da Equação 4.27 permite que as probabilidades teóricas sejam ser comparadas com as probabilidades empíricas, calculadas a partir da análise dos intervalos simétricos de probabilidade p (p-IP). Um gráfico de dispersão, entre essas probabilidades (esperadas e estimadas) pode ser

construído para verificar qualidade dos modelos de incerteza. Esse gráfico de dispersão é conhecido como gráfico de exatidão (*accuracy plots*, Deutsch, 1997). Nesse gráfico, em uma visão conservadora, as dclp modelam a incerteza corretamente se os pontos estiverem sobre a bissetriz do quadrante (reta $\bar{\xi}(p) = p$). Isto é, se os pontos estiverem acima ou abaixo dessa reta, significa que os modelos probabilísticos não capturam adequadamente a incerteza associada à VA $Z(u)$, refletindo a inexactidão desses modelos. Contudo, os modelos que apresentam $\bar{\xi}(p) \geq p$, são geralmente interpretados como sendo adequados. Se, por exemplo, os pontos estiverem abaixo da reta ($\bar{\xi}(p) < p$), a proporção de valores conhecidos que se inserem no intervalo é menor do que a probabilidade esperada. Ou seja: os intervalos simétricos de probabilidade p (p -IP), derivados a partir dos modelos locais de incerteza, contém uma proporção menor de valores verdadeiros.

Seguindo o mesmo raciocínio, Goovaerts (2001) propôs que, para aplicações onde o principal enfoque está relacionado ao fato de $Z(u)$ ultrapassar certos valores limites específicos (*threshold values*), os intervalos simétricos (centrados na mediana) fossem substituídos por intervalos ilimitados de probabilidade ($F^{-1*}(u_j; p|(n)), +\infty$]. Porém, na prática, esses intervalos são limitados pelo valor máximo da distribuição experimental (ou o máximo valor admitido), isto é: $[F^{-1*}(u_j; p|(n)), z_{\max}]$. Nesse caso, a modelagem correta da incerteza local associada à VA $Z(u)$ garantiria que, sobre a área em estudo, uma proporção $\bar{\xi}(p)$ de valores verdadeiros extrapolaria o quantil p (q_p) das dclp. Então, nesse caso, a proporção de valores conhecidos inseridos no intervalo é calculada como:

$$\bar{\xi}(p) = \frac{1}{N} \sum_{j=1}^N \xi(u_j; p) \quad \forall p \in [0,1] \quad (4.28)$$

sendo que

$$\xi(u_j; p) = \begin{cases} 1 & \text{se } F^{-1*}(u_j; (1-p)) < z(u_j) \\ 0 & \text{se não} \end{cases} \quad \forall p \in [0,1]$$

O gráfico de dispersão entre os intervalos de probabilidade p e a proporção $\bar{\xi}(p)$ de valores verdadeiros que se inserem no intervalo ($F^{-1*}(u_j; p|(n)), +\infty$] é um tipo particular do gráfico de exatidão, chamado de gráfico da extrapolação de probabilidade. A Figura 4.2 mostra exemplos de um gráfico de exatidão e de um de extrapolação de probabilidade. Nessa figura, as técnicas paramétricas e não-paramétricas são comparadas.

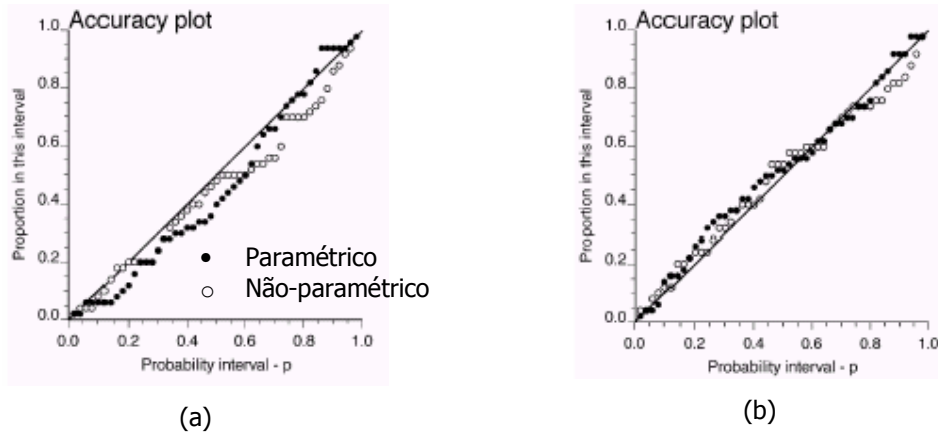


Figura 4.2: Gráficos de exatidão, considerando intervalos simétricos de probabilidade (p-IP) (a) e intervalos ilimitados de probabilidade (b). Fonte: Goovaerts (2001).

Deutsch (1997) propôs o cálculo de um índice para medir a proximidade entre as proporções esperadas e estimadas:

$$G = 1 - \int_0^1 [3 \cdot a(p) - 2] \cdot [\bar{\xi}(p) - p] \cdot dp \tag{4.29}$$

onde a função indicadora $a(p)$ é definida como,

$$a(p) = \begin{cases} 1 & \text{se } \bar{\xi}(p) \geq p \\ 0 & \text{se não} \end{cases}$$

Portanto, no cálculo desse índice, o peso atribuído para o caso exato é unitário, enquanto que é dado o dobro de importância, $|3 \cdot a(p) - 2| = 2$, para o caso inexato (quando $\bar{\xi}(p) < p$). Como comentado acima, os modelos que apresentam $\bar{\xi}(p) \geq p$ são considerados adequados, no sentido de tolerar uma margem na qual a proporção de valores verdadeiros inseridos no intervalo simétrico p-IP seja maior do que a esperada. Evidentemente, a qualidade dos modelos locais de incerteza, medidos por esse índice, aumenta à medida que G assume valor próximo a um ($G \rightarrow 1$).

Idealmente, os valores verdadeiros devem estar inseridos nos intervalos de probabilidade p (p-IP), de acordo com a probabilidade p esperada. Mas também, os intervalos de valores das dclp devem ser tão estreitos quanto possíveis, de modo que a incerteza local seja a menor possível. Para medir o espalhamento das dclp, vários índices de incerteza podem ser aplicados, como: variância, CV, entropia e intervalo entre quartis (IEQ). Além desses índices de incerteza, Goovaerts (2001) propõe que seja avaliada a amplitude média $\bar{W}(p)$ dos intervalos de probabilidade p (p-IP) que incluem valores conhecidos. Dessa forma, a qualidade dos modelos locais de incerteza pode ser avaliada por meio de um gráfico entre as probabilidades p e as respectivas amplitudes médias $\bar{W}(p)$. Para cada probabilidade p, as amplitudes médias $\bar{W}(p)$ são calculadas da seguinte maneira:

$$\bar{W}(p) = \frac{1}{N \cdot \xi} \sum_{j=1}^N \xi(u_j; p) \cdot [F^{-1}(u_j; (1+p)/2) - F^{-1}(u_j; (1-p)/2)] \quad \forall p \in [0,1] \quad (4.30)$$

O mesmo índice pode ser calculado considerando os intervalos ilimitados de probabilidade, ou seja: $(F^{-1}(u_j; p|(n)), +\infty]$. Nesse caso, $\bar{W}(p)$ passa a ser calculado como:

$$\bar{W}(p) = \frac{1}{N \cdot \xi} \sum_{j=1}^N \xi(u_j; p) \cdot [z_{\max} - F^{-1}(u_j; (1-p)/2)] \quad \forall p \in [0,1] \quad (4.31)$$

A Figura 4.3 mostra um gráfico entre as probabilidades esperadas e as respectivas amplitudes médias dos intervalos de valores. Como na figura anterior, o espalhamento dos modelos é avaliado para dclp obtidas por meio de técnicas paramétricas e não-paramétricas.

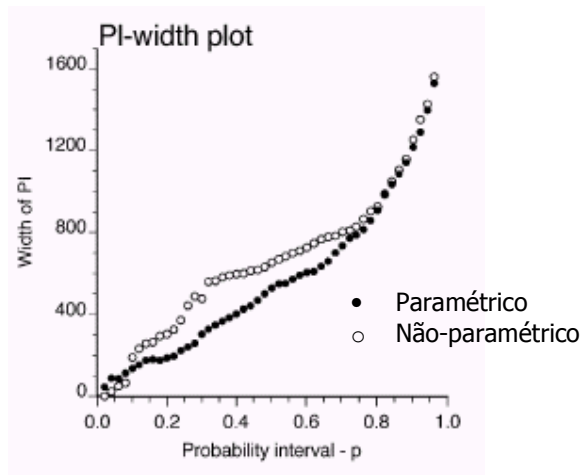


Figura 4.3: Gráfico mostrando a amplitude média dos intervalos de valores de acordo com o intervalo simétrico de probabilidade p esperada. Fonte: Goovaerts (2001).

As ferramentas apresentadas acima são válidas à avaliação da qualidade dos modelos locais de incerteza (dclp), quando utilizadas as técnicas de validação-cruzada ou *jackknife*. Porém, não existe técnica equivalente para avaliar se os modelos de incerteza são capazes de medir adequadamente a incerteza espacial. Para essa tarefa, geralmente, é verificada a qualidade de reprodução do histograma e do semi-variograma (da distribuição experimental) pelas realizações, embora a reprodutibilidade dos mesmos não garante a qualidade do conjunto de realizações como um todo. Portanto, não é possível obter um indicativo da qualidade dos modelos em relação à incerteza espacial por meio de métodos analíticos (como para a avaliação dos modelos locais). Para isso, é preciso utilizar métodos estocásticos, no qual verificam numericamente a reprodutibilidade dos momentos de primeira e segunda ordem da distribuição experimental.

4.10. Reprodução dos momentos estatísticos da distribuição experimental

Todos os algoritmos de simulação geoestatística compartilham o mesmo objetivo: gerar L realizações equiprováveis da distribuição espacial da FA $Z(u)$. A verificação de que esse objetivo foi alcançado adequadamente consiste na comparação das possíveis distribuições com a distribuição experimental. Na medida em que essas realizações são possíveis cenários da FA $Z(u)$, é natural esperar que cada realização reproduza o histograma e o semi-variograma da distribuição experimental. Porém, na prática, essa reprodutibilidade é raramente atingida devido às chamadas flutuações ergódicas (Deutsch e Journel, 1998, p. 128-132). Vários fatores controlam a magnitude das flutuações ergódicas apresentadas pelas realizações (Goovaerts, 1997, p. 426):

- i. o algoritmo empregado. Algoritmos de simulação seqüencial paramétricos e não-paramétricos, reproduzem o modelo de continuidade espacial da distribuição experimental somente na média de muitas realizações;
- ii. o número de informações experimentais. A similaridade entre as distribuições aumenta com o número de informações condicionantes originais;
- iii. as características do modelo de continuidade espacial em relação ao tamanho da área estudada. A magnitude das flutuações ergódicas, em relação aos momentos de segunda ordem das realizações, tende a ser significativa quando o alcance do modelo variográfico for grande em relação ao tamanho da área a ser simulada, particularmente se o efeito pepita for pequeno.

A Figura 4.4 apresenta um exemplo com os semi-variogramas nas direções de máxima e mínima continuidade espacial (caso em 2D) de quatro realizações de um processo de simulação, juntamente com seus semi-variogramas condicionantes (experimentais e teóricos). Como observado na figura, a reprodutibilidade do semi-variograma condicionante (teórico) ocorre na média das realizações analisadas.

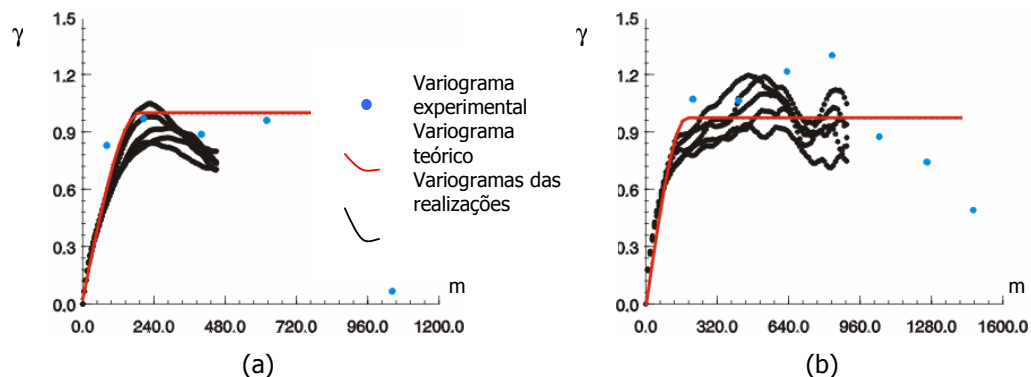


Figura 4.4: Flutuações na reprodução do momento de 2ª ordem da distribuição condicionante nas direções de maior (a) e de menor continuidade espacial (b). Fonte: Pilger (2000).

A Figura 4.5 apresenta os histogramas das mesmas realizações da figura anterior, juntamente com o histograma condicionante. Nesta figura, também podem ser observadas flutuações ergódicas, mas em média o histograma condicionante é reproduzido. Ocorrem flutuações nos valores de média, de variância e dos quartis. Os únicos valores fixos são os de mínimo e máximo.

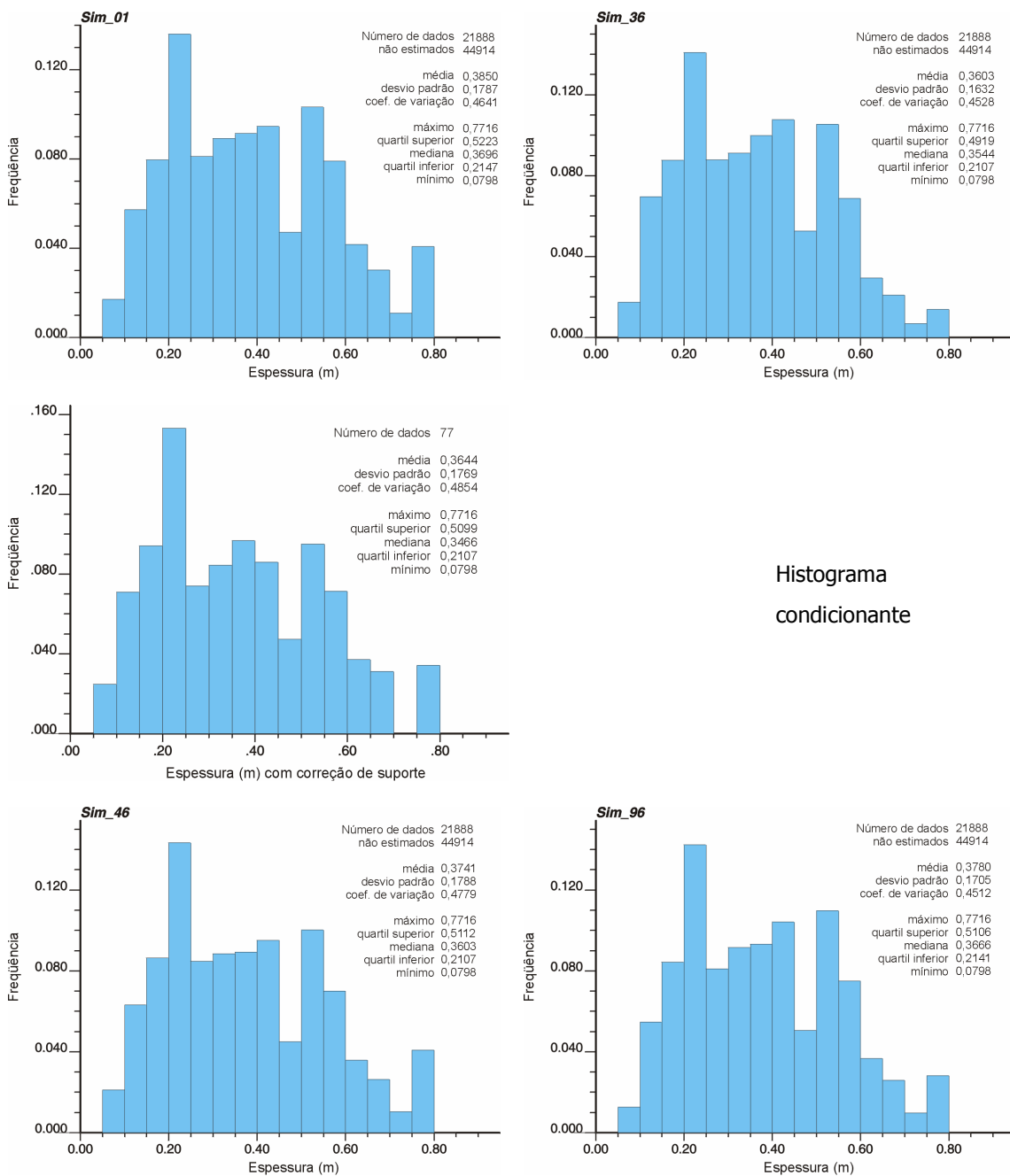


Figura 4.5: Flutuações ergódicas na reprodução do momento de 1ª ordem da distribuição condicionante (histograma central).

Fonte: Pilger (2000).

4.11. Índices locais de incerteza

A estimativa da dclp por meio dos métodos estocásticos permite que a variabilidade local, em relação aos possíveis resultados que a FA $Z(u)$ pode assumir, seja quantificada. A partir do conhecimento estocástico (numérico) da dclp em uma determinada localização u é possível estimar a incerteza em torno do valor desconhecido $z(u)$ antes e independentemente da escolha de uma estimativa $Z^*(u)$. Os principais métodos (índices de incerteza) utilizados na avaliação da magnitude da incerteza por meio de simulação estocástica são:

- i. intervalo de probabilidade;
- ii. entropia local;
- iii. variância condicional;
- iv. coeficiente de variação;
- v. intervalo entre quartis (IEQ).

Dentre esses cinco métodos citados, os três últimos foram considerados no estudo de caso apresentado por Pilger (2000) e Pilger *et al.* (2000).

4.11.1. Variância Condicional

A variância condicional mede o espalhamento da dclp em torno da média e é dada pela seguinte equação:

$$\sigma^2(u) = \int_0^{+\infty} [z - z_E^*(u)]^2 \cdot f(u, z | (n)) \cdot dz \quad (4.32)$$

onde:

$z_E^*(u)$ é a média esperada da dclp (E-type de $Z(u)$) e

$f(u, z | (n))$ é a função densidade de probabilidade local da VA $Z(u)$.

Entretanto, a equação acima pode ser aproximada, numericamente, por meio da equação abaixo:

$$\sigma^2(u) \cong \sum_{k=1}^{K+1} [\bar{z}_k - z_E^*(u)]^2 \cdot [F(u; z_k | (n)) - F(u; z_{k-1} | (n))] \quad (4.33)$$

onde,

z_k , $k = 1, \dots, K$, são os K valores limite discretizando o intervalo de variação da VA $Z(u)$,

\bar{z}_k é média da classe (z_{k-1} , z_k], dependente do modelo de interpolação para os valores disponíveis para cada classe (modelo linear: média aritmética),

$F(u; z | (n))$ é a dclp, onde a notação $| (n)$ expressa o condicionamento às informações locais (circunvizinhos ao nó de grid ou bloco a ser simulado).

O valor esperado $z_E^*(u)$ da dclp é definido, numericamente, pela seguinte equação:

$$z_E^*(u) \equiv \sum_{k=1}^{K+1} \bar{z}_k \cdot [F(u; z_k | (n)) - F(u; z_{k-1} | (n))] \quad (4.34)$$

A variância condicional define o espalhamento em torno da média, que por sua vez é dependente da média das classes \bar{z}_k . A média e as classes extremas da dclp são muito sensíveis à escolha do modelo de extrapolação da distribuição (Goovaerts, 1997, p. 337). Por isso, é preciso ter atenção quanto ao uso da variância condicional frente a um determinado modelo de extrapolação da dclp.

4.11.2. Coeficiente de variação

O coeficiente de variação (CV) é um método um pouco mais robusto que o de variância condicional, embora apresente a mesma característica com relação à escolha do modelo de extrapolação. Esse método deriva do anterior:

$$CV = \frac{\sqrt{\sigma^2(u)}}{z_E^*(u)} \equiv \frac{\sqrt{\sum_{k=1}^{K+1} [\bar{z}_k - z_E^*(u)] \cdot [F(u; z_k | (n)) - F(u; z_{k-1} | (n))]}{\sum_{k=1}^{K+1} \bar{z}_k \cdot [F(u; z_k | (n)) - F(u; z_{k-1} | (n))]} \quad (4.35)$$

onde os termos de (4.35) foram apresentados anteriormente.

4.11.3. Intervalo entre quartis

O método de intervalo entre quartis (IEQ) se apresenta como um método de avaliação da incerteza mais robusto do que os anteriores. O IEQ é principalmente indicado para medidas de espalhamento local de distribuições assimétricas porque não considera os valores extremos (classes extremas), sendo menos sensível quanto à escolha de um determinado modelo de extrapolação para os extremos da dclp. O uso desse método permite que o contraste entre áreas de alta variabilidade esperada e áreas de baixa variabilidade esperada seja mais saliente (Goovaerts, 1997, p. 339; Pilger, 2000; Pilger *et al.*, 2001 e Pilger *et al.*, 2002). O IEQ é definido como a diferença entre os quartis superior e inferior das dclp, ou seja:

$$IEQ(u) = q_{0,75}(u) - q_{0,25}(u) \tag{4.36}$$

De maneira análoga, esse índice de incerteza também pode ser adaptado para quaisquer diferenças entre quantis. Por exemplo, dependendo da situação, pode ser interessante calcular a diferença entre os quantis $q_{0,90}$ e $q_{0,10}$ das dclp de uma determinada FA $Z(u)$. Por isso, de uma forma genérica, esse índice poderia ser denominado de intervalo entre quantis. Pilger *et al.* (2001) e Pilger *et al.* (2002) apresentam exemplos de aplicação desse índice de incerteza.

A Figura 4.6 apresenta um perfil de terreno com 10 amostras, onde é observada a variabilidade em relação aos teores de acordo com o espaçamento entre elas. Na mesma figura, é mostrada a diferença entre os métodos de avaliação da magnitude de incerteza apresentados acima. Nessa figura, é possível observar a deficiência ou não dos métodos em estimar a incerteza em torno das estimativas. A partir de uma análise da Figura 4.6, verifica-se que os métodos de variância condicional e intervalo entre quartis (IEQ) incorporam de uma forma mais realística a incerteza existente entre as medidas. Portanto, esses dois métodos se mostram adequados para medir a variabilidade dos modelos de incerteza. O método de estimativa da incerteza por meio da medida do CV, embora não mostrado graficamente, também apresenta bons resultados (similares aos de variância condicional).

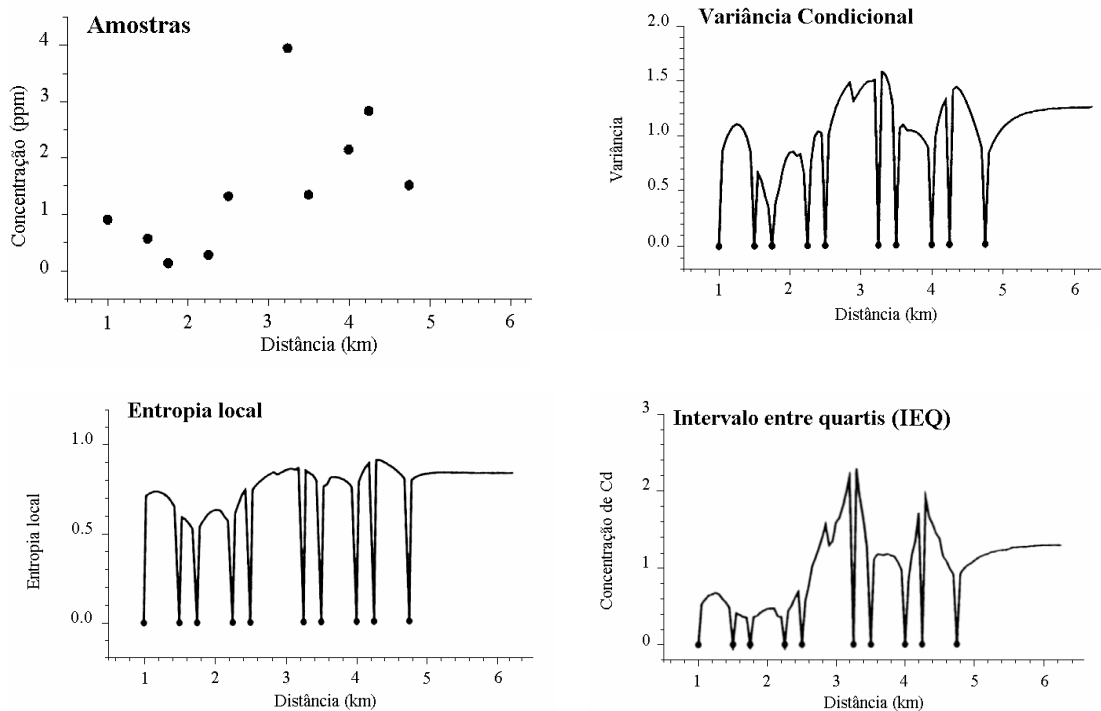


Figura 4.6: Concentração (ppm) de acordo com a distância linear entre as localizações medidas (figura superior esquerda), juntamente com três gráficos comparativos de variabilidade em cada posição do perfil. Fonte: Goovaerts (1997 p. 339).

4.12. Medida da incerteza a partir de krigagem

Se a incerteza local for modelada por meio do método paramétrico, a avaliação do espalhamento dos modelos de incerteza (dclp), por meio de IEQ, pode ser grandemente simplificada. Para esses casos, a utilização do método de krigagem multiGaussiana, embasada na lei Gaussiana condicional local, permite que os índices de incerteza locais [IEQ(u)] sejam calculados por meio de krigagem, sendo desnecessário a realização dos modelos estocásticos. Por exemplo, o valor de y associado ao quartil inferior ($q_{0,25}$) é $-0,6745$, da mesma forma que $+0,6745$ está associado ao quartil superior ($q_{0,75}$).

Portanto, a Equação 4.8 permite que seja realizado o cálculo do valor Gaussiano local $Y(u)$, associado ao quantil desejado, por meio de krigagem, usando a estimativa local e o desvio padrão. Evidentemente, a distribuição experimental deve ser previamente normalizada, de acordo com os procedimentos descritos no capítulo anterior.

Logo, se for desejado medir a magnitude do espalhamento dos modelos de incerteza por meio do método de IEQ, a krigagem dos valores Gaussianos $y(u_\alpha)$ ($\alpha = 1, \dots, n$) pode fornecer os quartis $Z_{q_{0,25}}(u_j)$ e $Z_{q_{0,75}}(u_j)$ para sejam calculados os índices de IEQ(u_j) ($j = 1, \dots, N$). Para isso, o seguinte algoritmo pode ser empregado:

- i. estimativa de $Y^*(u_j)$ ($j = 1, \dots, N$) por krigagem, a partir das informações Gaussianas circunvizinhas de u_j , retendo a variância de krigagem $\sigma_k^2(u_j)$;
- ii. cálculo dos quartis $Y_{q_p}(u_j)$ ($j = 1, \dots, N$) no espaço Gaussiano, por meio da lei Gaussiana condicional – Equação 4.8;
- iii. transformação inversa, ao espaço original das informações, por meio da função de anamorfose – Equação 4.7;
- iv. cálculo dos índices locais IEQ(u_j) ($j = 1, \dots, N$).

O emprego dessa metodologia apresenta a grande vantagem de não ser necessário a realização de L simulações $z^{(l)}(u_j)$ ($l = 1, \dots, L$ e $j = 1, \dots, N$) de $Z(u)$ para o cálculo de IEQ(u). Contudo, se além de IEQ(u), for desejado o cálculo de outros índices de incerteza, a inferência dos modelos estocásticos se torna necessária.

4.13. Comentários

Esse capítulo tinha como meta principal apresentar as utilidades dos modelos de incerteza, mostrando como é possível medir a magnitude de incerteza associada diretamente aos atributos medidos ou a partir de funções de transferência. Para isso, foi definido o conceito de função de transferência (FT) e mostrado como a incerteza pode se propagar por meio delas.

O capítulo iniciou discutindo a distinção entre incerteza local e espacial, pois esse entendimento é essencial para a construção adequada dos modelos de incerteza, conforme os propósitos desejados. A seguir, foi discutida o uso dos algoritmos de simulação seqüencial estocástica e de krigagem multiGaussiana, aplicados de acordo com o tipo de incerteza associada (local ou espacial). Foi salientado a importância do uso dos algoritmos de simulação seqüencial estocástica, sobretudo quando é necessário estimar os modelos de incerteza em unidades de tamanho diferentes daquelas das unidades de medida e para reproduzir as características estatísticas de 1ª e 2ª ordem dos dados. A seção 4.9 alerta para que, na medida do possível, os modelos de incerteza sejam avaliados de modo que se possam fazer inferências aproximadas sobre a FA de interesse.

O capítulo também apresentou uma ferramenta muito eficaz para a construção dos modelos de incerteza das variáveis resposta das FTs: o método *Latin Hypercube Sampling*, LHS. Seu estudo despertou a curiosidade e motivação para a concretização dessa Tese. Como será visto nos próximos capítulos, a aplicação desse método constitui o cerne dessa Tese.

O capítulo seguinte mostra como a técnica de *Latin Hypercube Sampling* pode ser combinada nos algoritmos de simulação condicional seqüencial visando aumentar seus índices de eficiência.

Capítulo 5

LHS aplicado aos algoritmos de simulação seqüencial

Esse capítulo é dedicado à apresentação formal da técnica LHS (*Latin Hypercube Sampling*), assim como sua inclusão nos métodos seqüenciais de simulação estocástica. O capítulo inicia com um breve histórico do método, apresentando o problema motivador de sua elaboração e os métodos similares antecessores. A seguir, são apresentadas as características da técnica LHS, além de seu desenvolvimento matemático. É realçada a potencialidade de aplicação da técnica LHS nas rotinas de simulação seqüencial de dados geoposicionados. Na seqüência, uma etapa fundamental para alcançar o objetivo proposto nessa Tese, é realizada: o desenvolvimento de um algoritmo, envolvendo a técnica LHS, para inclusão nos métodos seqüenciais de simulação estocástica. Testes teóricos de verificação encerram o capítulo.

5.1. LHS: histórico

O método LHS foi originalmente proposto por McKay *et al.* (1979), visando reduzir a variância do resultado de funções de transferência (FT) com relação a sua reprodutibilidade. Curiosamente, essa pesquisa também foi desenvolvida pelo Laboratório Nacional Los Alamos, responsável pela introdução dos conceitos dos métodos de Monte Carlo. Ou seja, se desejava um método de amostragem (de valores de variáveis envolvidas em FT) que permitisse com que os resultados fossem pouco variáveis quando o processo de amostragem fosse repetido. Nesse contexto, o método LHS é visto como um método redutor de variância.

O método LHS pode ser considerado uma extensão da técnica de amostragem por quota, apresentado por Steinberg (1963) e pode ser visto como uma extensão de dimensão K ($K = n^\circ$ de variáveis envolvidas) da técnica *latin square sampling* (Raj, 1968; Pres *et al.*, 1992). Essa última técnica, tem sido utilizada no projeto de experimentos estatísticos desde os anos 20 (pelo menos), em pesquisas relacionadas à agronomia (Fisher, 1925 e Fisher, 1926). Yates (1953) e Patterson (1954) estenderam a aplicabilidade da técnica *latin square sampling*, analisando suas propriedades de

redução de variância. Segundo McKay *et al.* (1979), o método LHS pertence a uma família que inclui outros três métodos de amostragem:

- i. Balanço aleatório (*random balance design*), discutido por Satterthwaite (1959), Budne (1959), Youden *et al.* (1959) e Anscombe (1959).
- ii. Fatores altamente fracionalizados (*highly fractionalized factorial design*), discutido por Enrenfeld e Zacks (1951 e 1967), Dempster (1960 e 1961) e Zacks (1963 e 1964).
- iii. Amostragem *lattice* (*lattice sampling*), discutido por Jensen (1975).

O método LHS está inserido em uma classe de técnicas nas quais a amostragem é realizada de maneira uniforme sobre o domínio considerado. Essa classe inclui, além do método LHS, os métodos de: (i) amostragem estratificada, (ii) amostragem ortogonal seqüencial e (iii) métodos quasi-Monte Carlo. Essas técnicas assumem que o domínio de amostragem é caracterizado por um cubo unitário de dimensão k , $[0;1]^k$. Veach (1997) apresenta uma introdução dessas técnicas, destacando suas aplicabilidades, vantagens e desvantagens.

Antes de iniciar a apresentação formal da técnica LHS, será introduzido o conceito da teoria dos Quadrados Latinos (*Latin quares*), utilizada no desenvolvimento teórico da técnica.

5.2. Quadrados Latinos (*Latin Squares*)

A técnica dos Quadrados Latinos (*Latin Squares*) desempenha uma função importante na teoria estatística de projetos de experimentos. Essa técnica tem sido utilizada com o objetivo de controlar a variação em um experimento (FT) com relação às suas variáveis de entrada (*inputs*). Os valores assumidos pelas variáveis podem ser visualizados matricialmente, onde as diferentes combinações de valores são dadas pelas linhas e colunas da matriz.

A teoria dos Quadrados Latinos foi desenvolvida no final do século XVIII, pelo matemático suíço Leonhard Euler (1707 – 1783) (Euler, 1782), como “uma nova espécie de quadrados mágicos” (*“un nouveau espèce de carrés magiques”*). O nome da técnica tem origem nos símbolos representados matricialmente por Euler. O matemático utilizava letras do alfabeto latino para representar os elementos de suas matrizes. Contudo, esses elementos podem ser representados por quaisquer símbolos, sejam eles números, letras gregas, cores ou objetos.

Um Quadrado Latino de ordem N se caracteriza como uma seqüência (combinação) $N \times N$ (matriz quadrada) de N símbolos, no qual cada símbolo ocorre exatamente uma vez em cada linha ou coluna da matriz. Dessa maneira, os símbolos são ordenados de modo que cada linha ou coluna da matriz contenha cada um dos N símbolos. Por exemplo, um Quadrado Latino de ordem três pode ser representado da seguinte forma:

$$\begin{bmatrix} A & B & C \\ C & A & B \\ B & C & A \end{bmatrix} \quad (5.2)$$

Quando N for igual a três, existem seis permutações possíveis (ABC, ACB, BAC, BCA, CAB e CBA), entretanto é possível construir 12 diferentes Quadrados Latinos de ordem três. Isso é possível por meio da permutação de linhas, colunas, e/ou símbolos em qualquer combinação. Existem várias possibilidades para a construção de Quadrados Latinos, exceto para os casos de ordem dois, onde existem somente duas possibilidades. Por exemplo, para o Quadrado Latino de ordem três apresentado acima, é possível criar outro, simplesmente permutando as duas últimas linhas. Ou seja:

$$\begin{bmatrix} A & B & C \\ B & C & A \\ C & A & B \end{bmatrix} \quad (5.3)$$

O número de possíveis Quadrados Latinos de ordem $N = 1, 2, 3, 4, 5, \dots$ são: 1, 2, 12, 576, 161280, ... (Riordan, 1958). Por exemplo, dois dos 576 possíveis Quadrados Latinos de ordem quatro são dados por:

$$\begin{bmatrix} A & B & C & D \\ B & A & D & C \\ C & D & A & B \\ D & C & B & A \end{bmatrix} \text{ e } \begin{bmatrix} A & B & C & D \\ C & D & A & B \\ D & C & B & A \\ B & A & D & C \end{bmatrix} \quad (5.4)$$

Um par de Quadrados Latinos de ordem N é dito ortogonal se um pode ser sobreposto no outro, de modo que cada uma das N^2 combinações de símbolos ocorre exatamente uma vez nas N^2 células da matriz. Por exemplo, os seguintes Quadrados Latinos, constituídos por dois conjuntos de três elementos (A, B, C e a, b, c)

$$\begin{bmatrix} C & B & A \\ B & A & C \\ A & C & B \end{bmatrix} \text{ e } \begin{bmatrix} b & c & a \\ a & b & c \\ c & a & b \end{bmatrix} \quad (5.5)$$

são ortogonais, pois,

$$\begin{bmatrix} Cb & Bc & Aa \\ Ba & Ab & Cc \\ Ac & Ca & Bb \end{bmatrix} \quad (5.6)$$

Isto é, cada um dos nove pares preenche uma das nove células da matriz.

Os pares de Quadrados Latinos ortogonais também são chamados de Quadrados Greco-Latinos, pois freqüentemente o segundo conjunto de N elementos era representado por letras gregas. Um exemplo ilustrativo de um Quadrado Greco-Latino de ordem quatro é formado pelas figuras de um jogo de cartas (ás, rei, dama e valete) e seus respectivos naipes (copas, ouros, espadas e paus). Na prática, um dos conjuntos pode ser representado por valores de uma determinada variável que se deseja confrontar com um segundo conjunto de valores assumidos por outra variável.

O número de Quadrados Greco-Latinos possíveis de serem construídos a partir de pares de Quadrados Latinos de ordem $N = 1, 2, 3, 4, \dots$ são: 0, 0, 36, 3456, (Riordan, 1958).

5.3. Latin Hypercube Sampling, LHS

Suponhamos que se deseja realizar M tiragens aleatórias a partir de um domínio Ω . A idéia de amostragem estratificada é subdividir (arbitrariamente) esse domínio Ω em M regiões Ω_i ($\Omega_1, \dots, \Omega_m$), sem sobreposição, de modo que:

$$\bigcup_{i=1}^m \Omega_i = \Omega \quad (5.7)$$

Cada região Ω_i representa de um estrato do domínio Ω . A partir daí, um número determinado de amostras é tirado de cada estrato Ω_i , de acordo com uma dada função de densidade de probabilidade, fdp. Como descrito na seção 5.1, amostragem estratificada faz parte de um conjunto de técnicas que assumem que o domínio de amostragem seja caracterizado por um cubo unitário de dimensão k , $[0;1]^k$.

A técnica *Latin Hypercube Sampling* (LHS), também se baseia na subdivisão do domínio Ω . Entretanto, diferente da técnica de amostragem estratificada, a técnica LHS divide o domínio $[0;1]^k$ em M estratos (classes) iguais, ao longo de cada dimensão k , garantindo que uma tiragem seja realizada em cada classe. Desse modo, todas as classes são representadas pelas M amostras, pois cada amostra será proveniente de um intervalo diferente, de probabilidade M^{-1} . O nome da técnica tem origem no caráter k dimensional da amostragem de um cubo unitário. O termo "*Latin*" advém dos Quadrados Latinos, de Euler.

O procedimento da técnica LHS é realizado por meio de K permutações π_1, \dots, π_k de classes $\{1, \dots, M\}$, assumindo que a localização da amostra (tiragem), em termos de probabilidade, seja dada por:

$$p_i^k = M^{-1} \cdot (\pi_{ik} - R_{ik}) \quad i = 1, \dots, M \text{ e } k = 1, \dots, K \quad (5.8)$$

ou

$$p_i^k = M^{-1} \cdot (\pi_{ik} - 1 + R_{ik}) \quad i = 1, \dots, M \text{ e } k = 1, \dots, K \quad (5.9)$$

onde:

p_i^k representa a probabilidade da i -ésima amostra da variável k ,

R_{ik} é um número aleatório independente e uniformemente distribuído no domínio $]0;1[$,

M é o número total de classes,

π_{ik} representa as permutações de classes.

As permutações π_{ik} assumem valores inteiros, representando as M classes $\{1, \dots, M\}$ de cada variável k . Nessas equações, π_{ik} determina a classe da tiragem, enquanto que R_{ik} determina a posição da

tiragem, dentro dessa classe. As permutações π_{ik} e os números aleatórios R_{ik} apresentam a característica de serem mutuamente independentes.

As duas equações apresentadas acima (Equações 5.8 e 5.9) são válidas, podendo ser aplicadas em algoritmos envolvendo a técnica LHS, pois garantem os mesmos resultados. Para $R_{ik} = 0,5$, as duas equações são idênticas, entretanto essas equações trabalham em sentido inverso quando $R_{ik} \neq 0,5$. Por exemplo, quando $R_{ik} = 0,0001$, a Equação 5.9 resulta no limite inferior de cada classe, enquanto que a Equação 5.8 garante o limite superior. Quando $R_{ik} = 0,9999$, ocorre o inverso.

Em um exemplo bi-dimensional, o padrão de amostragem resultante da técnica LHS corresponde à ocorrência de um único símbolo em um Quadrado Latino. Isto é, uma seqüência $M \times M$ de M símbolos, de forma que o símbolo não apareça duas vezes em uma mesma linha ou coluna. Por exemplo, se $M = 6$, utilizando Θ como símbolo:

$$\begin{bmatrix} - & - & - & \Theta & - & - \\ - & - & - & - & - & \Theta \\ \Theta & - & - & - & - & - \\ - & - & \Theta & - & - & - \\ - & \Theta & - & - & - & - \\ - & - & - & - & \Theta & - \end{bmatrix} \tag{5.10}$$

onde os caracteres “_” representam espaços vazios na matriz.

Se um vetor aleatório X de dimensão k (X_1, X_2, \dots, X_k), representando K VAs independentes, for caracterizado por K dp F_k , então, a partir da Equação 5.8 ou 5.9, M amostras dessas VAs podem ser obtidas por meio da seguinte equação:

$$x_{ik} = F_k^{-1}(p_i^k) \tag{5.11}$$

onde $F_k(\cdot)$ representa a dp da variável k .

Um clássico esquema ilustrativo da técnica LHS é apresentado na Figura 5.1. Essa figura ilustra o LHS de duas VAs (A e B) independentes e uniformemente distribuídas no quadrado unitário $U[0;1]$, considerando seis classes ($M = 6$). É possível verificar que cada linha ou coluna (de amplitude $1/6$) contém exatamente um símbolo, representando uma única observação (tiragem).

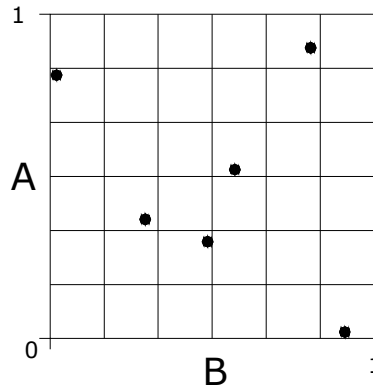


Figura 5.1: Exemplo de LHS, com $M = 6$, para duas VAs (A e B) independentes e uniformes $U[0;1]$.

Um exemplo numérico desse esquema é dado na Tabela 5.1. Nesse exemplo, se as VAs forem Gaussianas [$VA \sim N(0,1)$], os valores assumidos pelas amostras (x_{ik}) podem ser obtidos por meio da tabela de distribuição cumulativa normal (Equação 3.12), disponível em grande parte da literatura especializada (exemplo em Myers, 1996, p. 137-139).

Tabela 5.1: Exemplo numérico da Figura 5.1, LHS de dimensão 2, para duas VAs independentes e uniformes $U[0; 1]$.

i	k	π_{ik} (classe)	Amplitude (prob.)	R_{ik}	$\pi_{ik} - 1 + R_{ik}$	p_i^k
1	1	2	[0,17;0,33)	0,6890	1,6890	0,2815
2	1	6	[0,83;1,00]	0,4154	5,4154	0,9026
3	1	1	[0,00;0,17)	0,2818	0,2818	0,0470
4	1	4	[0,50;0,67)	0,2099	3,2099	0,5350
5	1	5	[0,67;0,83)	0,9302	4,9302	0,8217
6	1	3	[0,33;0,50)	0,2163	2,2163	0,3694
1	2	3	[0,33;0,50)	0,9450	2,9450	0,4908
2	2	5	[0,67;0,83)	0,8390	4,8390	0,8065
3	2	6	[0,83;1,00]	0,4466	5,4466	0,9078
4	2	4	[0,50;0,67)	0,3872	3,3872	0,5645
5	2	1	[0,00;0,17)	0,2876	0,2876	0,0479
6	2	2	[0,17;0,33)	0,7449	1,7449	0,2908

Portanto, LHS é uma técnica de amostragem aleatória estratificada sem substituição, na qual M amostras de K variáveis (contínuas) são tiradas, de modo que para cada variável, a amostragem seja (marginalmente) maximamente estratificada (McKay *et al.*, 1979). A amostragem será maximamente estratificada quando o número de classes for igual ao número de amostras. Ou, em outras palavras, quando a probabilidade de que uma amostra seja tirada de cada classe seja igual a M^{-1} .

McKay *et al.* (1979) apresentaram a técnica LHS considerando apenas o caso onde as K VAs envolvidas forem independentes. Oito anos depois, Stein (1987) propôs uma modificação no algoritmo, proporcionando que a técnica LHS fosse capaz de ser utilizada com sucesso em variáveis dependentes.

McKay *et al.* (1979) compararam o desempenho de três técnicas de amostragem em relação à reprodutibilidade da média e da variância dos resultados de uma FT, por meio da realização de 50 procedimentos de amostragem distintos. As técnicas foram as seguintes:

- i. amostragem aleatória;
- ii. amostragem estratificada; e
- iii. LHS.

Os resultados foram muito favoráveis em relação ao método LHS, mostrando redução significativa de variância para os dois parâmetros estatísticos testados. McKay *et al.* (1979) realizaram 50 repetições de amostragem, para quatro variáveis, de acordo com as três técnicas supracitadas. Em cada repetição, os autores utilizaram 16 amostras, calculando a média e a variância dos resultados de uma determinada FT envolvendo as variáveis. Em comparação com o método de amostragem aleatória, o método LHS mostrou variação, com relação à média, quatro vezes inferior. McKay *et al.* (1979) concluem que o método LHS é mais preciso em relação aos outros dois testados, principalmente quando a FT for monotônica e de caráter aditivo.

Stein (1987) também apresentou um experimento no qual a variância, em relação à média, foi significativamente reduzida por meio do uso da técnica LHS. Esse mesmo autor propôs uma variação na técnica LHS, a qual permite amostrar adequadamente variáveis dependentes. Essa variação é capaz de gerar amostras que conservam a correlação existente entre as variáveis, sobretudo quando o número de amostras for elevado. Iman e Conover (1982) também descreveram uma metodologia similar, entretanto, a mesma apresentou falhas para a produção um número elevado de amostras. O procedimento de amostragem proposto por Stein (1987) é o seguinte:

- i. Amostrar aleatoriamente a fdp multivariada, obtendo s_{ik} ($i = 1, \dots, M$ e $k = 1, \dots, K$) amostras.
- ii. Classificar os vetores de amostras $s_k = (s_{1k}, \dots, s_{Mk})$, com relação aos seus valores, constituindo K conjuntos $o_k = (o_{1k}, \dots, o_{Mk})$ formados pelo número de classificação das amostras s_{ik} .
- iii. Realizar a tiragem de $M \cdot K$ números aleatórios R_{ik} independentes, a partir de $U[0;1]$.
- iv. Obter p_i^k , que representa a localização da i -ésima amostra da VA k , em termos de probabilidade, por meio da Equação 5.8 ou 5.9, substituindo π_{ik} por o_{ik} .
- v. Determinar o valor da amostra x_{ij} ($i = 1, \dots, M$ e $k = 1, \dots, K$) a partir da Equação 5.11.

Nesse algoritmo, os conjuntos $o_k = (o_{1k}, \dots, o_{Mk})$ exercem a função das K permutações de classes $\{1, \dots, M\}$ do algoritmo LHS original. Isso porque, esses conjuntos são constituídos por M números inteiros, ordenados de acordo com o número de classificação das amostras s_i ($i = 1, \dots, M$).

Essa variação na técnica LHS pode proporcionar uma translação de classes, visto que, os valores aleatoriamente amostrados na etapa (i) são reposicionados, de modo a ocorrer somente uma tiragem por classe. Entretanto, a correlação entre os K vetores amostrais $x_k = (x_{1k}, \dots, x_{Mk})$, permanece

inalterada, pois a ordem de classificação das amostras é preservada. Um exemplo numérico, para uma distribuição bivariada, é apresentado na Tabela 5.2. Nesse exemplo, duas distribuições Gaussianas são amostradas seis vezes ($M = 6$) de acordo com a técnica LHS modificada, proposta por Stein (1987). Uma ilustração desse exemplo é dada na Figura 5.2.

Tabela 5.2: LHS de dimensão 2, para duas VAs Gaussianas ($VA \sim N[0;1]$), correlacionadas (correlação = -0,53).

i	K	s_{ik}	o_{ik}	Amplitude (classe)	R_{ik}	$o_{ik} - 1 + R_{ik}$	p_i^k	x_{ik}
1	1	0,4043	6	[0,97;+∞]	0,9605	5,9605	0,9934	2,4795
2	1	0,0567	5	[0,43;0,97)	0,3146	4,3146	0,7191	0,5802
3	1	-0,0282	4	[0,00;0,43)	0,6012	3,6012	0,6002	0,2539
4	1	-1,1387	2	[-0,97;-0,43)	0,5971	1,5971	0,2662	-0,6244
5	1	-0,0565	3	[-0,43;0,00)	0,1956	2,1956	0,3659	-0,3426
6	1	-1,3538	1	[-∞;-0,97)	0,5937	0,5937	0,0989	-1,2876
1	2	-1,8744	1	[-∞;-0,97)	0,1339	0,1339	0,0223	-2,0082
2	2	0,4081	6	[0,97;+∞]	0,8411	5,8411	0,9735	1,9353
3	2	-0,6378	3	[-0,43;0,00)	0,9650	2,9650	0,4942	-0,0146
4	2	-0,3931	4	[0,00;0,43)	0,7395	3,7395	0,6233	0,3140
5	2	-0,9640	2	[-0,97;-0,43)	0,2576	1,2576	0,2096	-0,8078
6	2	0,1016	5	[0,43;0,97)	0,4833	4,4833	0,7472	0,6658

Os conjuntos de amostras s_1 e s_2 , obtidos por meio de amostragem aleatória simples, são anticorrelacionados, apresentando coeficiente de correlação igual a -0,53. O procedimento de amostragem proposto por Stein (1987) não só garante que os conjuntos de amostras x_1 e x_2 conservem essa correlação, como também garante que a amostragem resultante seja estratificada (Figura 5.2). Entretanto, Stein (1987) mostrou que o processo reproduz assintoticamente (para M elevado) a correlação original. O autor enfatiza que a correlação da amostragem resultante pode ser significativamente distorcida se M não for elevado o suficiente.

Na Figura 5.2, as linhas contínuas entre símbolos indicam as translações de classes sofridas por cada amostra individual, a partir da mudança de amostragem aleatória simples para LHS (de s_{ik} para x_{ik}). As linhas contínuas delgadas representam os limites de classes.

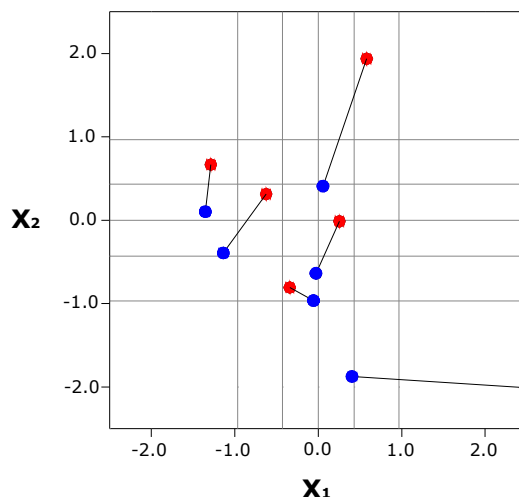


Figura 5.2: LHS para duas VAs Gaussianas dependentes (Tabela 5.2), de acordo com o procedimento proposto por Stein (1987). Amostragem aleatória simples (●) translada para LHS (●), onde as linhas contínuas entre símbolos indicam a translação.

O LHS para as VAs x_1 e x_2 , da figura acima, também pode ser representado por um Quadrado Latino de um único símbolo (Θ):

$$\begin{bmatrix}
 _ & _ & _ & _ & \Theta & _ \\
 \Theta & _ & _ & _ & _ & _ \\
 _ & \Theta & _ & _ & _ & _ \\
 _ & _ & _ & \Theta & _ & _ \\
 _ & _ & \Theta & _ & _ & _ \\
 _ & _ & _ & _ & _ & \Theta
 \end{bmatrix}
 \tag{5.12}$$

onde os caracteres “_” representam espaços vazios na matriz.

5.4. Monte Carlo nos algoritmos simulação condicional seqüencial

Os métodos de simulação geoestatística permitem a criação de L ($L = 1, \dots, L$) imagens equiprováveis do atributo considerado, de modo que os dados simulados (em cada realização) reproduzam as características estatísticas de 1ª e 2ª ordem da distribuição original. A cada ponto ou bloco da área são associados L valores equiprováveis dos atributos de interesse.

Os métodos de simulação condicional estocástica mais utilizados são os de simulação seqüencial Gaussiana (Isaaks, 1990), simulação por bandas rotativas (Matheron, 1973) e simulação seqüencial dos indicadores (Alabert, 1987b). Esses algoritmos estão disponíveis na maioria dos softwares direcionados a geoestatística, como por exemplo, na GSLIB (*Geostatistical Software's Library*) (Deutsch e Journel, 1998) ou no Isatis®. Dentre os métodos citados acima, destacam-se os seqüenciais, paramétricos ou não-paramétricos (Capítulo 3).

Geralmente, os métodos estocásticos utilizam a técnica de simulação de Monte Carlo (Capítulo 3, seção 3.1) para a construção dos modelos numéricos. Nessa técnica, os modelos de incerteza (dclp) da VA Z nas localizações u_j ($j = 1, \dots, N$) são amostradas aleatoriamente L vezes, construindo o conjunto de valores simulados $\{z^{(l)}(u_j) \mid j = 1, \dots, N \text{ e } l = 1, \dots, L\}$. Nesses métodos, o número L de realizações é determinante para a caracterização do espaço de incerteza da VA $Z(u)$. Quanto maior for o número L , maior o número de amostras disponíveis para a construção dos modelos estocásticos, que caracterizam o espaço de incerteza. O número mínimo de realizações necessárias, para se caracterizar suficientemente bem o espaço de incerteza, depende das características intrínsecas dos dados e do objetivo de pesquisa. Existem casos em que L precisa ser tão elevado que essa técnica pode se tornar computacionalmente muito intensa, exigindo muito tempo de processamento e espaço em disco (gerando arquivos de difícil manipulação).

Como descrito no Capítulo 3, o embasamento teórico do método ssG assume que toda distribuição multivariada dos dados siga um comportamento Gaussiano. Dessa forma, a aplicação desse método exige que a distribuição experimental da VA $Z(u)$ seja transformada de acordo com uma distribuição Gaussiana. Ou seja, a VA $Z(u)$ deve ser transformada em uma VA $Y(u)$. A hipótese multiGaussiana é muito conveniente, pois permite que os modelos de incerteza (dclp) sejam derivados por meio de parâmetros obtidos por krigagem (Equação 3.19). A partir da construção das dclp, o conjunto de dados simulados $\{y^{(l)}(u_j) \mid j = 1, \dots, N \text{ e } l = 1, \dots, L\}$ é obtido por meio da técnica de tiragem de simulação de Monte Carlo, na qual as dclp da VA Y nas localizações u são amostradas aleatoriamente L vezes. O conjunto $\{y^{(l)}(u_j) \mid j = 1, \dots, N\}$ representa uma realização da FA $Y(u)$ no domínio físico definido pelas informações geoposicionadas no espaço normal $y(u_\alpha)$ ($u_\alpha = 1, \dots, n$). Ao passo que, o conjunto $\{y^{(l)}(u_j) \mid l = 1, \dots, L\}$ representa L simulações da VA Y na localização u_j ($j = 1, \dots, N$). Posteriormente, o conjunto de dados simulados $\{y^{(l)}(u_j) \mid j = 1, \dots, N \text{ e } l = 1, \dots, L\}$ é transformado ao espaço original da VA $Z(u)$. Isso é realizado por meio da aplicação da função de anamorfose (Equação 3.21). Assim:

$$z^{(l)}(u_j) = F^{-1}\left[G^*(u_j; y^{(l)} \mid (n))\right] \quad u_j = 1, \dots, N \text{ e } l = 1, \dots, L \quad (5.13)$$

onde,

$z^{(l)}(u_j)$ é o valor simulado, no espaço original, associado à localização u_j da l -ésima realização,

$y^{(l)}$ é o valor simulado, no espaço normal, associado à localização u_j da l -ésima realização,

$F(\cdot)$ representa a dp da FA $Z(u)$ e

$G^*(\cdot)$ representa a dp estimada da FA $Y(u)$.

Portanto, o valor da VA Z , em cada localização u_j ($j = 1, \dots, N$), é simulado dentro do domínio de variação da VA $Z(u)$, por meio de um procedimento aleatório, a partir das dclp. Em cada localização, o processo de simulação gera uma distribuição, composta por L valores, que pode ser considerada uma aproximação numérica (método estocástico) da dclp; ou seja:

$$F(u; z | (n)) \approx \frac{1}{L} \sum_{l=1}^L i^{(l)}(u; z) \quad (5.14)$$

onde

$$i^{(l)}(u; z) = \begin{cases} 1 & \text{se } z^{(l)}(u) \leq z \\ 0 & \text{se não} \end{cases} \quad \text{com } l=1, \dots, L$$

Na verdade, os métodos de simulação seqüencial estocástica fazem uso de um algoritmo no mínimo curioso, na medida em que o objetivo é determinar um modelo de incerteza estocástico, conhecendo-se (de antemão) o modelo teórico de incerteza, por meio de krigagem.

O algoritmo de simulação seqüencial Gaussiana está implementado para obter o valor simulado $y^{(l)}(u_j)$ (no espaço normal) por meio da seguinte equação:

$$y^{(l)}(u_j) = xp^{(l)}(u_j) \cdot \sigma_{KS}(u_j) + y^*_{KS}(u_j) \quad u_j = 1, \dots, N \text{ e } l = 1, \dots, L \quad (5.15)$$

Nessa equação, $\sigma_{KS}(u_j)$ e $y^*_{KS}(u_j)$ representam o desvio padrão de krigagem (KS) e a estimativa por KS, respectivamente (esses valores não são fixos devido à característica do método seqüencial de percorrer um caminho aleatório em cada realização, e de adicionar dados simulados no banco de dados original). Enquanto que $xp^{(l)}(u_j)$ representa um valor Gaussiano, derivado da distribuição de probabilidade Gaussiana padrão $G(y)$, associado à localização u_j . Isto é:

$$xp^{(l)}(u_j) = [G^{-1}(u_j, p)] \quad u_j = 1, \dots, N \text{ e } l = 1, \dots, L \quad (5.16)$$

onde p representa uma probabilidade, definida no domínio $[0;1]$, tirada aleatoriamente por meio da técnica de Monte Carlo.

Para uma dada precisão, a aplicação do método LHS garante que toda a dclp (teórica) seja representada, reproduzindo a construção do conjunto de L valores simulados $\{z^{(l)}(u_j) \mid j = 1, \dots, N \text{ e } l = 1, \dots, L\}$, em muito menos tempo (realizações) do que o método de Monte Carlo. Por conseguinte, a redução de L também faz com que a manipulação dos arquivos produzidos seja facilitada, uma vez que estes ocuparão menos espaço em disco.

5.5. LHS nos algoritmos simulação condicional seqüencial

A meta dessa Tese se constitui no desenvolvimento de um algoritmo de simulação condicional seqüencial, capaz de reduzir o tempo computacional (diretamente proporcional a L) e o espaço virtual requerido nas suas aplicações. Essa meta pretende ser alcançada por meio da substituição do método de Monte Carlo pelo de LHS. Dessa forma, a eficiência do processo de caracterização do espaço de incerteza da VA $Z(u)$, para uma dada precisão, pode ser aumentada. Conseqüentemente, os processos pós-simulação também serão beneficiados, pois serão desenvolvidos a partir de arquivos menores (em termos de M bytes).

A técnica LHS proposta por McKay *et al.* (1979) foi desenvolvida para VAs contínuas e independentes, que constituem os argumentos de funções de transferência (FTs). Nesse contexto, na

formulação original, K representa o número de VAs envolvidas na FT. Na adaptação da técnica LHS para VAs espaciais, o número de localizações u_j ($j = 1, \dots, N$) da malha representa o número K das VAs da formulação original. Portanto, para VAs espaciais, o índice k é substituído pelo índice j nas Equações 5.8 e 5.9. Entretanto, nessa adaptação, ainda existe a questão quanto à independência das VAs.

Em geociências, geralmente o conjunto de VAs $Z(u)$, das diferentes localizações u_j , apresenta correlação. Contudo, a proposta de substituição do método de Monte Carlo pelo método LHS, nos algoritmos de simulação condicional seqüencial, não envolve diretamente essas VAs. Na proposta dessa Tese, a técnica LHS está envolvida na tiragem dos números aleatórios e independentes p da distribuição uniforme $U[0;1]$. A proposta consiste em realizar a tiragem de M valores da VA Gaussiana padrão Y [$Y \sim N(0;1)$], a partir da distribuição $G(y)$, de forma aleatória, sem substituição e de M classes equiprováveis distintas. Isto é, primeiramente a distribuição $G(y)$ é dividida em M classes equiprováveis disjuntivas. A partir daí, são tirados, aleatoriamente, um valor de cada classe M , de cada localização u_j ($j = 1, \dots, N$) para cada realização L ($l = 1, \dots, L$) (Figura 5.3). O procedimento de tiragem respeita a equação abaixo:

$$xp_{\text{lhs}}^{(l)}(u_j) = [G^{-1}(u_j, \text{plhs}^{(i)}(u_j))] \quad j = 1, \dots, N, i = 1, \dots, M \text{ e } l = 1, \dots, L \quad (5.17)$$

onde $xp_{\text{lhs}}^{(l)}(u_j)$ representa um valor Gaussiano, derivado da distribuição de probabilidade Gaussiana padrão $G(y)$, associado à localização u_j . Esse termo é análogo ao $xp^{(l)}(u_j)$ da Equação 5.16. A probabilidade, de classes distintas, representada por $\text{plhs}^{(i)}$ pode ser definida como:

$$\text{plhs}^{(i)}(u_j) = M^{-1} \cdot [\pi_i(u_j) - 1 + R_i(u_j)] \quad j = 1, \dots, N \text{ e } i = 1, \dots, M \quad (5.18)$$

ou

$$\text{plhs}^{(i)}(u_j) = M^{-1} \cdot [\pi_i(u_j) + R_i(u_j)] \quad j = 1, \dots, N \text{ e } i = 1, \dots, M \quad (5.19)$$

onde:

$\text{plhs}^{(i)}(u_j)$ representa a probabilidade da i -ésima realização, associada à localização u_j ,

M é o número total de classes,

$\pi_i(u_j)$ representa as M permutações de classes, associadas à localização u_j e

$R_i(u_j)$ é um número aleatório independente e uniformemente distribuído no domínio $[0;1]$.

As Equações 5.18 e 5.19 são análogas às Equações 5.8 e 5.9, a diferença é o caráter espacial dado às variáveis, representado pelas localizações u_j . É importante salientar que, nas Equações 5.17, 5.18 e 5.19, i é aleatório, podendo assumir qualquer valor inteiro do intervalo $[1;M]$. Ao contrário, o valor assumido por l é seqüencial, iniciando por 1 e avançando até M .

Assim, em cada localização u_j ($j = 1, \dots, N$), para que seja tirado um valor Gaussiano $xp_{\text{lhs}}^{(l)}$ de cada classe M , o número L de realizações (tiragens aleatórias) deve ser igual ao número de classes M que discretiza a distribuição $G(y)$. Ou seja: $L = M$. A Figura 5.3 ilustra o procedimento.

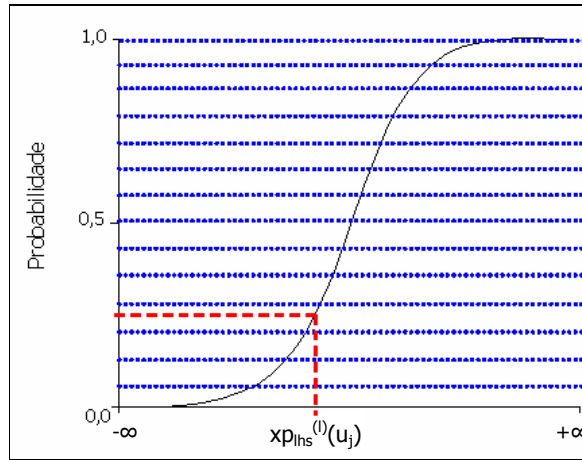


Figura 5.3: Tiragem aleatória de um n^o Gaussiano, associado à localização u_j , a partir da 4ª classe da distribuição $G(y)$; $M = 14$.

De acordo com um caminho (ordem) aleatório, cada localização u_j ($j = 1, \dots, N$) é visitada para a resolução de um sistema de KS visando a obtenção dos valores de $\sigma_{KS}(u_j)$ e $y^*_{KS}(u_j)$. A partir daí, uma classe M é aleatoriamente sorteada (sem substituição) e um valor é tirado aleatoriamente (técnica de Monte Carlo) dessa classe. Esse valor representa a probabilidade $plhs^{(l)}$, que alimenta a Equação 5.17 para a determinação do valor Gaussiano $xplhs^{(l)}(u_j)$. Esse valor Gaussiano, juntamente com aqueles calculados por KS é substituído na equação abaixo para se obter os valores simulados $y^{(l)}(u_j)$ ($j = 1, \dots, N$ e $l = 1, \dots, L$).

$$y^{(l)}(u_j) = xplhs^{(l)}(u_j) \cdot \sigma_{KS}(u_j) + y^*_{KS}(u_j) \quad j = 1, \dots, N \quad \text{e} \quad l = 1, \dots, L \quad (5.20)$$

onde os termos envolvidos foram previamente definidos.

Dessa forma, para o algoritmo modificado, além de informar os parâmetros habituais, o usuário deve informar o número M de classes (discretizando a distribuição $G(y)$) desejado. Portanto, com a modificação proposta, os algoritmos de simulação seqüencial estocástica ficariam resumidos nas seguintes etapas:

- i. definição das M classes disjuntivas equiprováveis;
- ii. definição de um caminho aleatório, de modo que cada localização u_j ($j = 1, \dots, N$) desconhecida (célula ou bloco da malha) seja visitada somente uma única vez;
- iii. resolução de um sistema de KS para a estimativa de $\sigma_{KS}(u_j)$ e $y^*_{KS}(u_j)$ na localização u_j visitada, a partir dos $n(u_j)$ dados experimentais e dos valores previamente simulados na circunvizinhança de u_j ;
- iv. sorteio, sem substituição (amostragem aleatória sem substituição), de uma das classes definidas em (i);
- v. cálculo de uma probabilidade ($plhs^{(l)}(u_j)$), dentro da classe definida em (iv), a partir da Equação 5.18 ou 5.19;
- vi. cálculo do valor Gaussiano $xplhs^{(l)}(u_j)$, a partir da Equação 5.17;

- vii. simulação de um valor equiprovável da VA $Y(u)$ (isto é: $y^{(l)}(u_j)$ sendo que $j = 1, \dots, N$ e $l = 1, \dots, M = L$) por meio da Equação 5.20;
- viii. inclusão de $y^{(l)}(u_j)$ no banco de dados, representando uma informação condicional adicional às demais localizações a serem visitadas;
- ix. repetição das etapas (iii) a (viii) até que seja associada uma simulação a cada uma das N localizações;
- x. repetição dos passos (ii) a (ix) para gerar L ($L = M$) realizações equiprováveis da FA $Y(u)$.

Logo, com essa modificação mais um componente aleatório é introduzido no algoritmo. Além do sorteio das probabilidades $[plhs^{(l)}(u_j)]$, existe também a tiragem (sem substituição) das classes disjuntivas equiprováveis M .

Pebesma *et al.* (1999), também, aplicaram a técnica LHS em simulação condicional seqüencial, contudo, com objetivo diferente em relação ao proposto nessa Tese. Esses autores utilizaram o algoritmo proposto por Stein (1987) para re-amostrar as dclp nas localizações u_j ($j = 1, \dots, J$), visando alimentar FTs globais. Pebesma *et al.* (1999) aplicaram esse algoritmo no conjunto de valores simulados $z^{(l)}(u_j)$ ($l = 1, \dots, L$) $\{z^{(1)}(u_j), \dots, z^{(L)}(u_j)\}$, deslocando valores entre classes, para obter simulações locais estratificadas. A partir daí, foi testada a reprodutibilidade da média e dos percentis p_{05} e p_{90} de uma determinada FT, em relação às técnicas de amostragem Monte Carlo e LHS. A reprodutibilidade foi testada a partir de 100 processos de simulação de 20, 50 e 100 realizações. Os resultados mostraram a eficiência do método LHS, especialmente em relação à reprodutibilidade da média. Para esse parâmetro, foi mostrado que seriam necessárias mais de 5000 realizações pelo método de Monte Carlo para alcançar os resultados obtidos a partir de 20 realizações ($L = 20$) pela técnica LHS. Entretanto, Pebesma *et al.* (1999) alertam que o semi-variograma pode não ser integralmente reproduzido se número de realizações (L) não for elevado o suficiente. Isso porque, quanto menor for o número L (classes M), maior serão os deslocamentos necessários, fazendo com que ocorra uma desestruturação da correlação, especialmente em curtas distâncias.

A técnica desenvolvida por Stein (1987), para VAs correlacionadas, se constitui em uma significativa opção frente aquela que envolve a decomposição de Cholesky, proposta por Davis (1987), apresentada no Capítulo 4 (seção 4.6.1, p. 80). Essa técnica é especialmente útil para aplicações que envolvem FTs espaciais (globais), isto é, aquelas que são alimentadas por valores dependentes associados a várias localizações u_j ($j = 1, \dots, J$) (por exemplo, um simulador de fluxo, ou qualquer FT relacionada a um painel de lavra).

5.5.1. Variações no algoritmo

O algoritmo apresentado na seção anterior representa a base do algoritmo de simulação condicional seqüencial, envolvendo a técnica LHS, proposto e utilizado nessa Tese. Contudo, algumas características desse algoritmo podem ser modificadas e outras podem ser inseridas. A seguir serão apresentadas três variações:

- i. LHS determinístico;
- ii. LHS cíclico e
- iii. LHS com caminho determinístico.

5.5.1.1. LHS determinístico

Como descrito no Capítulo 4 (Seção 4.8), o processo de tiragem, dentro de cada classe, pode ser determinístico. Ou seja, o algoritmo pode ser modificado para que as tiragens sejam realizadas a partir de uma proporção predefinida (e fixa) de cada classe. Esse procedimento torna desnecessária a utilização do termo aleatório $R_i(u_j)$ nas Equações 5.18 e 5.19. Nessas equações, o termo $R_i(u_j)$ é substituído por uma constante p , que representa uma proporção. Por exemplo, se for desejado que as tiragens sejam realizadas a partir da mediana de cada classe, $p = 0,5$. Dessa forma, as Equações 5.18 e 5.19 da etapa (v) (do algoritmo) seriam substituídas pelas seguintes:

$$plhs_{ad}^{(i)}(u_j) = M^{-1} \cdot [\pi_i(u_j) - p] \quad j = 1, \dots, N \quad e \quad i = 1, \dots, M \quad (5.21)$$

ou

$$plhs_{ad}^{(i)}(u_j) = M^{-1} \cdot [\pi_i(u_j) - 1 + p] \quad j = 1, \dots, N \quad e \quad i = 1, \dots, M \quad (5.22)$$

onde:

- $plhs_{ad}^{(i)}(u_j)$ representa a probabilidade da i -ésima realização, associada à localização u_j ,
 M é o número total de classes,
 $\pi_i(u_j)$ representa as l permutações de classes, associadas à localização u_j e
 p representa a proporção da classe.

A Equação 5.21 oferece a probabilidade que corresponde à proporção $(1 - p)$ de cada classe, enquanto que a Equação 5.22 resulta na probabilidade que corresponde à própria proporção (p) das classes.

Portanto, com essa modificação, as probabilidades $plhs^{(i)}(u_j)$ são substituídas por $plhs_{ad}^{(i)}(u_j)$, onde "ad" vem de "amostragem descritiva".

5.5.1.2. LHS cíclico

O algoritmo de simulação, envolvendo a técnica LHS, pode se repetir em um mesmo processo de simulação. Quando todas M classes da distribuição $G(y)$, associada a cada localização u_j , forem amostradas, novos ciclos de amostragem (de dimensão M) podem ser realizados. Para isso, o número L de realizações deve ser maior e múltiplo de M . Por exemplo, se $L = 100$ e $M = 20$, serão realizados cinco ciclos do algoritmo proposto. Dessa maneira, as M classes podem ser melhor representadas. O algoritmo cíclico é o seguinte:

- i. definição das M classes disjuntivas equiprováveis;
- ii. definição de um caminho aleatório, de modo que cada localização u_j ($j = 1, \dots, N$) desconhecida (célula ou bloco da malha) seja visitada somente uma única vez;

- iii. resolução de um sistema de KS para a estimativa de $\sigma_{KS}(u_j)$ e $y_{KS}^*(u_j)$ na localização u_j visitada, a partir dos $n(u_j)$ dados experimentais e dos valores previamente simulados na circunvizinhança de u_j ;
- iv. sorteio, sem substituição (amostragem aleatória sem substituição), de uma das classes definidas em (i);
- v. cálculo de uma probabilidade ($plhs^{(l)}(u_j)$), dentro da classe definida em (iv), a partir da Equação 5.18 ou 5.19;
- vi. cálculo do valor Gaussiano $x_{plhs}^{(l)}(u_j)$, a partir da Equação 5.17;
- vii. simulação de um valor equiprovável da VA $Y(u)$ (isto é: $y^{(l)}(u_j)$ sendo que $j = 1, \dots, N$ e $l = 1, \dots, L$) por meio da Equação 5.20;
- viii. inclusão de $y^{(l)}(u_j)$ no banco de dados, representando uma informação condicional adicional às demais localizações a serem visitadas;
- ix. repetição das etapas (iii) a (viii) até que seja associada uma simulação a cada uma das N localizações;
- x. repetição dos passos (ii) a (ix) para gerar M realizações equiprováveis da FA $Y(u)$.
- xi. repetição de $(L/M - 1)$ vezes das etapas (ii) a (x).

Esse algoritmo também pode ser combinado com o LHS determinístico, apresentado na seção anterior.

5.5.1.3. LHS com caminho determinístico

No algoritmo originalmente proposto, como a técnica LHS é aplicada antes da etapa de condicionamento, as $dclp$ empíricas resultantes (tanto no espaço normal, como no original) não apresentarão necessariamente a característica de estratificação. As mesmas terão mais chance de serem estratificadas em localizações que apresentarem baixa variabilidade.

As $dclp$ resultantes serão mais estratificadas se o caminho (ordem) seqüencial for determinístico, pois assim, os parâmetros obtidos por KS serão aproximadamente constantes. Nesse caso, a chance de os valores assumidos por σ_{KS} e y_{KS}^* nas localizações u_j ($j = 1, \dots, N$) serem constantes é maior, pois os valores previamente simulados serão os únicos dados variáveis nas vizinhanças das localizações u_j . Para que essa característica seja incorporada no algoritmo (proposto na seção 5.4.3), basta modificar a etapa (x). Essa etapa deve garantir que todas as L realizações utilizem o mesmo caminho definido na etapa (ii). Dessa forma, a etapa (x) é substituída pela seguinte:

- x. repetição dos passos (iii) a (ix) para gerar L ($L = M$) realizações equiprováveis da FA $Y(u)$.

5.5.2. Implementação do algoritmo

O algoritmo proposto (incluindo suas variações) foi desenvolvido em linguagem de programação Fortran, a mesma linguagem das rotinas da GSLIB. A implementação do algoritmo, na rotina de simulação condicional seqüencial, pode ser resumida em sete etapas principais:

- i. definição das probabilidades $plhs_{ad}$, para o caso determinístico;
- ii. inicialização da simulação, com associações às localizações u_j e às classes M ;
- iii. laço, para a implementação do algoritmo cíclico;
- iv. tiragens de classes, associada às localizações u_j e às classes M ;
- v. definição das probabilidades $plhs$, para o caso clássico;
- vi. determinação dos valores simulados;
- vii. associação dos valores simulados ao número da realização e à localização u_j (posição).

Essas sete etapas serão desenvolvidas nas subseções seguintes.

5.5.2.1. Definição das probabilidades $plhs_{ad}$

Essa etapa foi realizada por meio da Equação 5.21, a partir do número de classes M e da proporção das classes p , definidas no arquivo de parâmetros. Em linguagem de programação Fortran, as probabilidades $plhs_{ad}$ são definidas por meio de um laço (*do*):

```
do i = 1, M
     $plhs_{ad}(i) = (i - p)/M$ 
end do
```

onde i são números inteiros e representam as classes M . A definição das probabilidades $plhs_{ad}$ é realizada de maneira independente das localizações u_j .

5.5.2.2. Inicialização da simulação

Nessa etapa, os endereços de armazenamento dos valores simulados são inicializados, de acordo com as localizações u_j e as classes M . Em Fortran, esse procedimento é realizado por meio de um laço concatenado:

```
do i = 1, M
    do j = 1, N
         $y(i,j) = UNEST$ 
    end do
end do
```

sendo que, $UNEST$ é uma variável macro que assume o valor de -99. Portanto, a matriz de simulação $[M \times N]$, representando todos os valores a serem simulados, é preenchida, inicialmente, pelo valor -99.

5.5.2.3. Implementação do algoritmo cíclico

Para que o algoritmo seja cíclico, a matriz de simulação $[M \times N]$ deve ser inicializada a cada ciclo de M realizações. Para isso, quando o número da realização corrente (l) for igual ao somatório entre o número de classes M e uma variável temporária ($temp$), a matriz é inicializada, em um procedimento idêntico ao apresentado na seção anterior (seção 5.5.2.3). Em linguagem Fortran:

```

temp = 1
do l = 1, L
    if(l.eq.temp+M) then
        do i = 1, M
            do j = 1, N
                y(i,j) = UNEST
            end do
        end do
        temp = l
    end if
    .
    .
    .
end do

```

onde l é o número da realização corrente, $temp$ é uma variável temporária, o termo $(l.eq.temp+M)$ significa $(l = temp + M)$ e o laço (*do*) externo é referente ao número de realizações. Se, para uma dada realização l , a condição $(l.eq.temp+M)$ for verdadeira, então $temp$ assume o valor de l .

Por exemplo: se $L = 45$ e $M = 15$, ocorrem três ciclos, compreendendo 15 realizações cada. O primeiro ciclo ocorre entre a 1ª e 15ª realização; o segundo entre a 16ª e 30ª realização e o terceiro entre a 31ª e 45ª realização. Assim, a cada ciclo, M probabilidades ($plhs$ ou $plhs_{ad}$), associadas às N localizações u_j , serão tiradas de M classes distintas a partir da distribuição padrão $G(y)$.

5.5.2.4. Tiragens de classes

A etapa (iv) do algoritmo (p. 108) consiste na tiragem aleatória de uma classe (π), associada a cada localização u_j , dentre as M classes disponíveis. Na prática, esse procedimento é realizado em duas etapas. Primeiramente, um número (*random*) é aleatoriamente tirado do domínio $]0;1[$; após, esse valor é multiplicado por 100 e a variável pi armazena sua parte inteira. Em seguida, é verificado se a classe sorteada é válida e se a mesma já foi tirada, para uma dada localização u_j . Em linguagem de programação Fortran:

```

pi = 0
random = 0
5000 random = rand(idum)

```

```

pi = int((random)*100 + 1)
  if(pi.gt.M.or.pi.eq.0.or.y(pi,index).gt.-99) go to 5000
  y(pi,index) = EST

```

onde o número aleatório *random* é obtido por meio da sub rotina `rand(.)`. O argumento dessa função (`idum`) é a semente da sub-rotina geradora de números aleatórios. A variável *index* representa as localizações u_j , aleatoriamente ordenadas, de acordo com um caminho. O termo condicional `(pi.gt.M.or.pi.eq.0.or.y(pi,index).gt.-99)` significa: ($\pi > M$ ou $\pi = 0$ ou $y(\pi, index) > -99$). Se essa condição for verdadeira, o programa retoma a linha comando 5000, em um laço. Se nenhum valor simulado estiver associado a classe π e a posição *index*, o valor assumido por π é aceito. Nesse momento, um valor (EST) é associado à variável $y(\pi, index)$ correspondente. O valor EST não é o valor simulado. Esse valor representa somente um *flag* para indicar que a posição correspondente da matriz simulação $[M \times N]$ já está preenchida.

5.5.2.5. Definição das probabilidades plhs

As probabilidades *plhs* são definidas por meio da Equação 5.18, a partir da classe (π) e de um número aleatório (R). Em Fortran, com a classe π já definida, basta tirar um número aleatório e aplicar a equação:

```

R = rand(idum)
plhs = [(pi-1)+R]/M

```

No caso descritivo:

```

plhs = plhsad(pi)

```

sendo que os termos envolvidos foram previamente definidos.

5.5.2.6. Determinação dos valores simulados

Com *plhs* já definido, basta definir o valor Gaussiano x_p e aplicá-lo, juntamente com os parâmetros obtidos por KS, na Equação 5.20. Logo:

```

call gauinvv(plhs,xp)
y(index) = xp * cstdev + cmean

```

onde a função `gauinvv(.)` representa $G^{-1}(\cdot)$; *cstdev* e *cmean* são o desvio padrão e o valor estimado, respectivamente, obtidos por KS.

5.5.2.7. Associação dos valores simulados

Após a determinação dos valores simulados, os mesmos $[y(index)]$ são associados à realização de origem:

```

y(l,index) = y(index)

```

Assim, é possível extrair todos os M ($L = M$) valores simulados associados a cada localização u_j (*index*).

5.5.3. Verificações teóricas

A partir da implementação do algoritmo na rotina de simulação condicional seqüencial, é preciso verificar sua funcionalidade quanto às características pretendidas pelo algoritmo. Cada classe da dp Gaussiana padrão $G(y)$ [$Y \sim N(0;1)$] deve estar representada nas tiragens dos valores Gaussianos $x_{p_{lhs}}$, associados a cada localização u_j ($j = 1, \dots, N$). Alguns testes foram realizados para verificar esse comportamento. Por exemplo, em uma situação em que existem 16 localizações a serem simuladas ($N = 16$) e desejar-se discretizar a dp padrão $G(y)$ em três classes ($M = 3$), três realizações ($L = 3$) devem ser executadas para garantir ao menos uma tiragem de cada classe (em cada localização). A aplicação da rotina, nesse exemplo, mostra que os resultados são como os esperados. A Figura 5.4 mostra que todas as classes são representadas, nas 16 localizações (index), no processo de simulação. Principalmente, é observado que nenhuma classe aparece duas vezes na mesma linha (processo de simulação); isso prova que as tiragens de classes são realizadas sem substituição. Nesse exemplo, as classes são representadas pelas letras A, B e C; as localizações por index e as realizações por R1, R2 e R3. O gráfico da direita da Figura 5.4, mostra os domínios das classes na dp $G(y)$, em termos de probabilidade e valores Gaussianos.

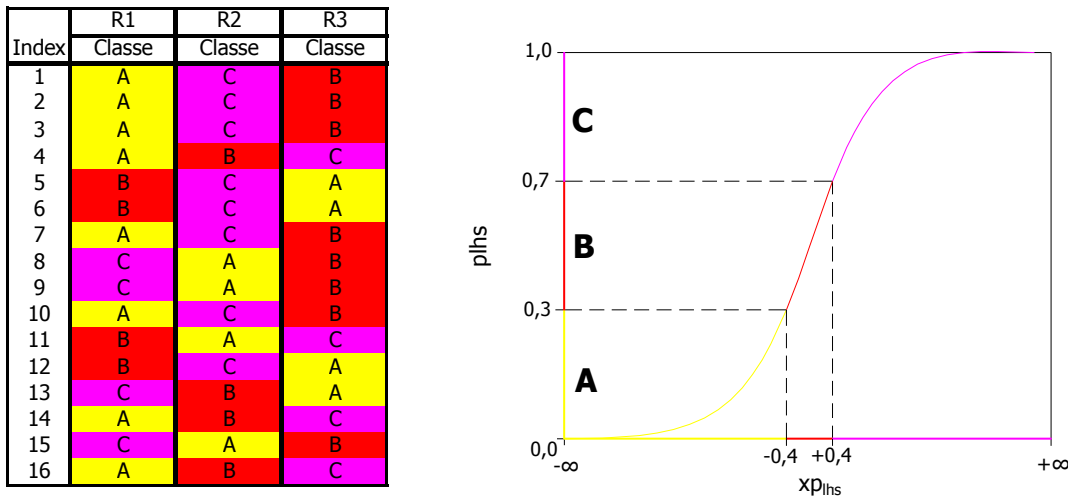


Figura 5.4: Quadro de tiragens de classes (A, B e C) em um caso em que $N = 16$ e $M = L = 3$ (esquerda); domínios das classes na dp $G(y)$ (direita).

Ainda, comprovando a funcionalidade da rotina, a Figura 5.5 mostra que as localizações do exemplo acima (Figura 5.4), quando comparadas em pares, formam Quadrados Latinos de um único símbolo (X). São apresentadas as tiragens de classes dos pares (12;13), (7;15) e (11,16).

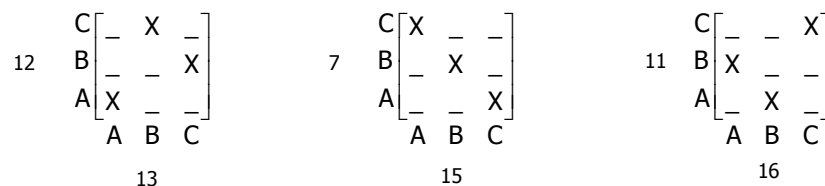


Figura 5.5: Quadrados Latinos para pares de localizações do exemplo da Figura 5.4.

A rotina simples, com um único ciclo foi aprovada. Na seqüência, é preciso verificar sua funcionalidade quanto à característica cíclica do algoritmo. Utilizando o exemplo anterior ($j = 16$ e $M = 3$), para que essa característica seja comprovada, no mínimo seis realizações devem ser executadas. Isso garante a ocorrência de um ciclo adicional no processo de simulação. A Figura 5.6 comprova a eficiência da rotina. Essa figura, que complementa a Figura 5.4, mostra que no segundo ciclo todas as classes também são tiradas sem substituição. Isso prova que, a cada ciclo, a matriz de simulação $[M \times N]$ é inicializada novamente, permitindo um novo processo de seleção de classes.

Index	Ciclo 1			Ciclo 2		
	R1	R2	R3	R4	R5	R6
	Classe	Classe	Classe	Classe	Classe	Classe
1	A	C	B	C	A	B
2	A	C	B	B	A	C
3	A	C	B	A	C	B
4	A	B	C	B	C	A
5	B	C	A	A	B	C
6	B	C	A	C	B	A
7	A	C	B	C	A	B
8	C	A	B	C	B	A
9	C	A	B	A	C	B
10	A	C	B	A	B	C
11	B	A	C	B	A	C
12	B	C	A	C	A	B
13	C	B	A	B	C	A
14	A	B	C	C	A	B
15	C	A	B	A	B	C
16	A	B	C	B	A	C

Figura 5.6.: Quadro de tiragens de classes (A, B e C) em um caso em que $j = 16$ e $M = L = 3$, para dois ciclos distintos.

Qualquer par de localizações, em um mesmo ciclo, configura um Quadrado Latino; todas as N localizações, configuram um Hiper-cubo Latino (*Latin hypercube*).

Com a tiragem de classes sendo realizada com sucesso, falta verificar a eficiência quanto à tiragem das probabilidades $[plhs^{(i)}(u_j)$ e $plhs_{ad}^{(i)}(u_j)]$ de cada classe i ($i = 1, \dots, M$). Para isso, serão consideradas as três realizações, do primeiro ciclo, das três últimas localizações do exemplo anterior. Ou seja, nesse caso, $j = 3$ ($j = 14, 15, 16$) e $L = M$ é mantido inalterado ($M = L = 3$). Em cada localização, uma tiragem ($plhs^{(i)}$ e $plhs_{ad}^{(i)}$) deve ser realizada de dentro das classes A, B e C.

Para o caso descritivo, os três valores assumidos por $plhs_{ad}^{(i)}$ ($i = 1, 2, 3$) (Equação 5.21) são compartilhados nas três localizações, ao longo do processo de simulação. Nesse caso, será amostrada a mediana (proporção, $p = 0,5$) de cada classe, de amplitude $1/3$ (em termos de probabilidade). A Tabela 5.3 apresenta o resultado das tiragens, mostrando os valores Gaussianos x_{plhs} associados (Equação 5.17).

Tabela 5.3: Exemplo de três tiragens de $plhs_{ad}$, e valores Gaussianos x_{plhs} , de acordo com o procedimento descritivo.

index	R1			R2			R3		
	classe	$plhs_{ad}$	x_{plhs}	classe	$plhs_{ad}$	x_{plhs}	classe	$plhs_{ad}$	x_{plhs}
14	A	0,17	-0,97	B	0,50	0,00	C	0,83	0,97
15	C	0,83	0,97	A	0,17	-0,97	B	0,50	0,00
16	A	0,17	-0,97	B	0,50	0,00	C	0,83	0,97

Os valores assumidos pelas probabilidades $plhs_{ad}^{(i)}(u_j)$ e pelos números Gaussianos $xplhs(u_j)$, são coincidentes nas três localizações u_j , ao longo do processo de simulação (Figura 5.7). A ordem de ocorrência desses valores se constitui no único fator variável com as realizações.

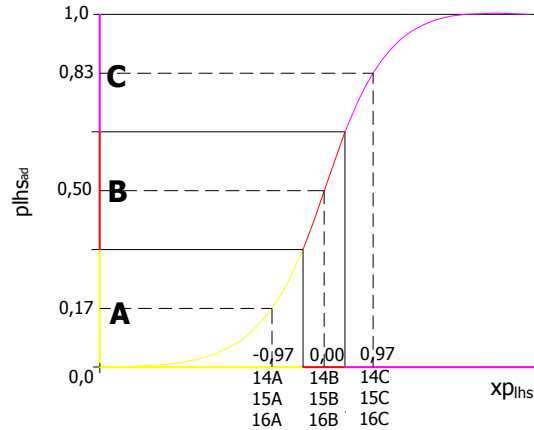


Figura 5.7: Representação gráfica dos resultados apresentados na Tabela 5.3, obtidos segundo o procedimento descritivo.

Para fins de comparação, o caso clássico será testado utilizando o mesmo exemplo do caso descritivo, apresentado acima. Com o procedimento clássico, as probabilidades $plhs^{(i)}$ variam de acordo com as localizações u_j (Equação 5.18). Nessa situação, as classes A, B e C, devem ser amostradas aleatoriamente ao longo do processo de simulação, nas três localizações u_j ($j = 14, 15, 16$). Com as probabilidades $plhs$ definidas, os valores Gaussianos $xplhs$ associados são obtidos por meio da Equação 5.17. Os resultados, apresentados na Tabela 5.4 e na Figura 5.8, foram como os esperados.

Tabela 5.4: Exemplo de três tiragens de $plhs$, e valores Gaussianos $xplhs$, de acordo com o procedimento clássico.

index	R1			R2			R3		
	classe	plhs	$xplhs$	classe	plhs	$xplhs$	classe	plhs	$xplhs$
14	A	0,16	-0,99	B	0,58	0,20	C	0,93	1,48
15	C	0,81	0,88	A	0,29	-0,55	B	0,47	-0,07
16	A	0,01	-2,33	B	0,38	-0,31	C	0,90	1,28

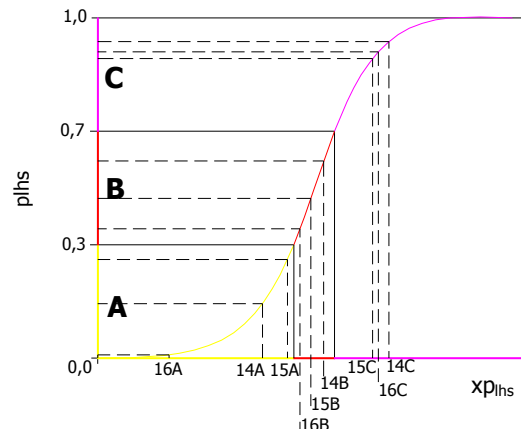


Figura 5.8: Representação gráfica dos resultados apresentados na Tabela 5.4, obtidos segundo o procedimento clássico.

Em ambos os casos (descritivo e clássico), nas três localizações u_j ($j = 14, 15, 16$), foram amostrados números Gaussianos $x_{p_{lhs}}$ oriundos das três classes (A, B e C). Em cada localização individual, não houve a ocorrência de sobreposição de classes. Os valores Gaussianos $x_{p_{lhs}}$ foram amostrados como o esperado, a partir das probabilidades $p_{lhs_{ad}}$ e p_{lhs} , de acordo com o algoritmo proposto. É importante salientar que é permitida a sobreposição de classes, para diferentes localizações, em uma mesma realização.

O uso da técnica LHS faz com que a distribuição Gaussiana padrão $G(y)$ seja melhor representada, sobretudo em seus trechos extremos. A técnica LHS faz com que a distribuição $G(y)$ seja representada em todas as M classes desejadas. Como resultado, a distribuição experimental $G(y)$ não apresenta lacunas de amostragem, como aquelas evidenciadas pelo resultado da aplicação da técnica de Monte Carlo. As figuras seguintes apresentam um comparativo entre as técnicas de tiragem de Monte Carlo e LHS (clássico e descritivo), de acordo com a rotina implementada. Essas figuras mostram a distribuição experimental $G(y)$, obtida por meio dessas técnicas de amostragem, segundo diferentes processos de simulação de L realizações, em uma localização qualquer u . Nessas figuras, para a técnica LHS descritiva, utilizou-se proporção $p = 0,5$ (mediana de classe). Em cada figura, também é apresentado um gráfico de probabilidade, relacionando as realizações com as probabilidades tiradas (ordenadas em ordem crescente). Nesse gráficos, uma reta indica que a distribuição $G(y)$ foi bem representada. A Figura 5.9 apresenta os resultados obtidos para 10 realizações ($L = 10 = M$).

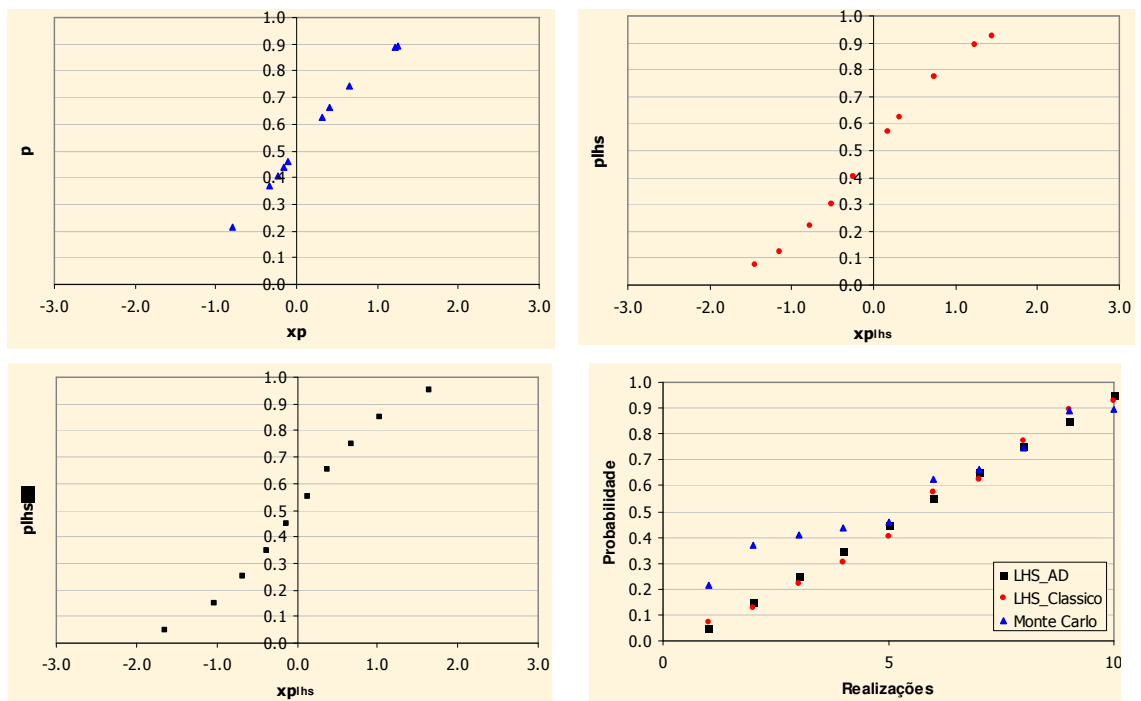


Figura 5.9: Distribuições experimentais de probabilidade Gaussiana a partir de 10 realizações, de acordo com o método de Monte Carlo (cima esquerda), LHS (cima direita) e LHS descritivo (baixo esquerda). Gráfico de probabilidade, segundo os três métodos (baixo direita).

Nessa mesma Figura 5.9, a distribuição experimental correspondente à técnica de Monte Carlo, mostra que os trechos extremos da dp padrão $G(y)$ não são devidamente representados. Essa curva apresenta agrupamento de pontos e lacunas de amostragem. Esse comportamento não é observado nos gráficos correspondentes à técnica LHS. Embora a dp padrão $G(y)$ não tenha sido muito bem representada, devido ao pequeno número de classes, os resultados obtidos a partir da técnica LHS são mais satisfatórios. As linhas de grade dos gráficos, ajudam a observar que uma tiragem é realizada em cada uma das 10 classes. Para o caso descritivo, uma tiragem é realizada na mediana de cada classe.

O gráfico de probabilidade reflete os resultados verificados nas curvas experimentais. O eixo das abscissas mostra as realizações, enquanto que o eixo das ordenadas apresenta as probabilidades. Nesse gráfico (e no das figuras seguintes), os valores das abscissas não são associados aos números das realizações. Esses valores são associados ao ordenamento das probabilidades (eixo y). Cada um desses valores é associado a uma determinada realização, dentre as L (M) disponíveis.

A Figura 5.10 ilustra o gráfico de probabilidade e as distribuições experimentais $G(y)$, resultantes de 15 realizações ($L = 15 = M$), de acordo com as técnicas utilizadas.

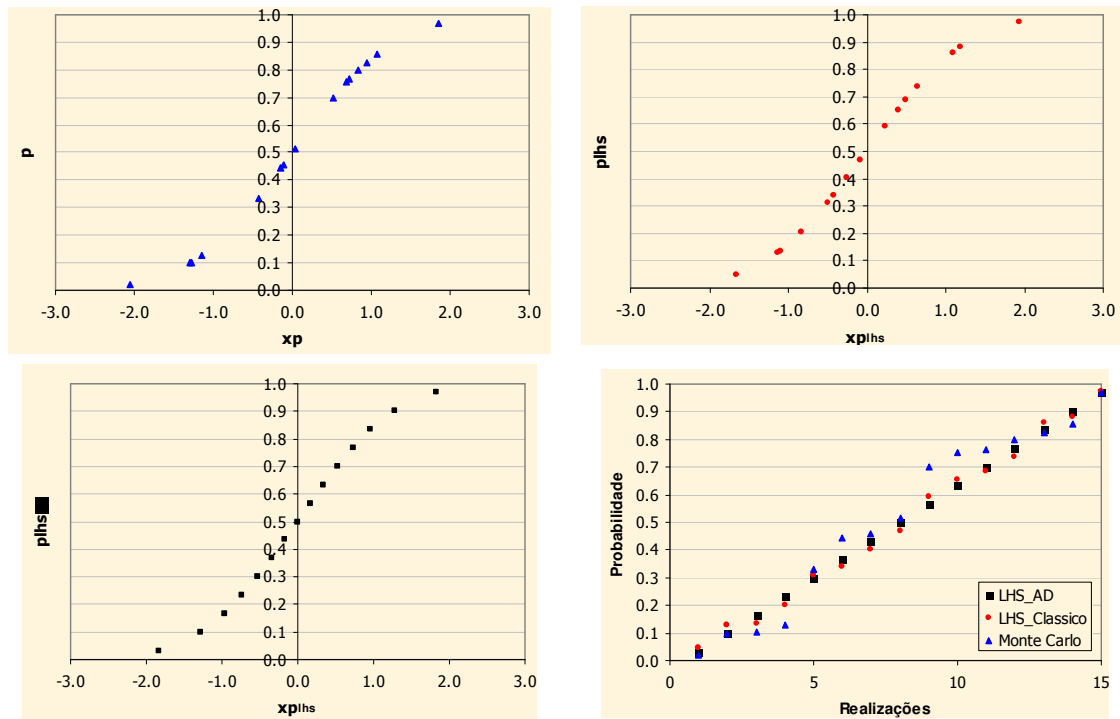


Figura 5.10: Distribuições experimentais de probabilidade Gaussiana a partir de 15 realizações, de acordo com o método de Monte Carlo (cima esquerda), LHS (cima direita) e LHS descritivo (baixo esquerda). Gráfico de probabilidade, segundo os três métodos (baixo direita).

Os gráficos resultantes mostram que as curvas obtidas a partir da técnica LHS representam melhor a dp padrão $G(y)$. Entretanto, a figura mostra que, mesmo utilizando a técnica LHS clássica, pode ocorrer o agrupamento de pares pontos. O segundo e o terceiro pontos, embora tenham sido

tirados de classes distintas, são muito próximos. Isso acontece quando são amostrados os limites superiores e inferiores de classes adjacentes.

A Figura 5.11 apresenta os resultados obtidos a partir de três processos de simulação de 20 realizações ($L = 20 = M$). Nesse caso, a metade inferior da dp padrão $G(y)$ foi suficientemente bem representada pela dp experimental oriunda da técnica de Monte Carlo. Porém, o mesmo não pode ser afirmado quanto à metade complementar, como pode ser verificado no gráfico de probabilidade. Ambas as curvas obtidas a partir da técnica LHS representam melhor a dp padrão $G(y)$.

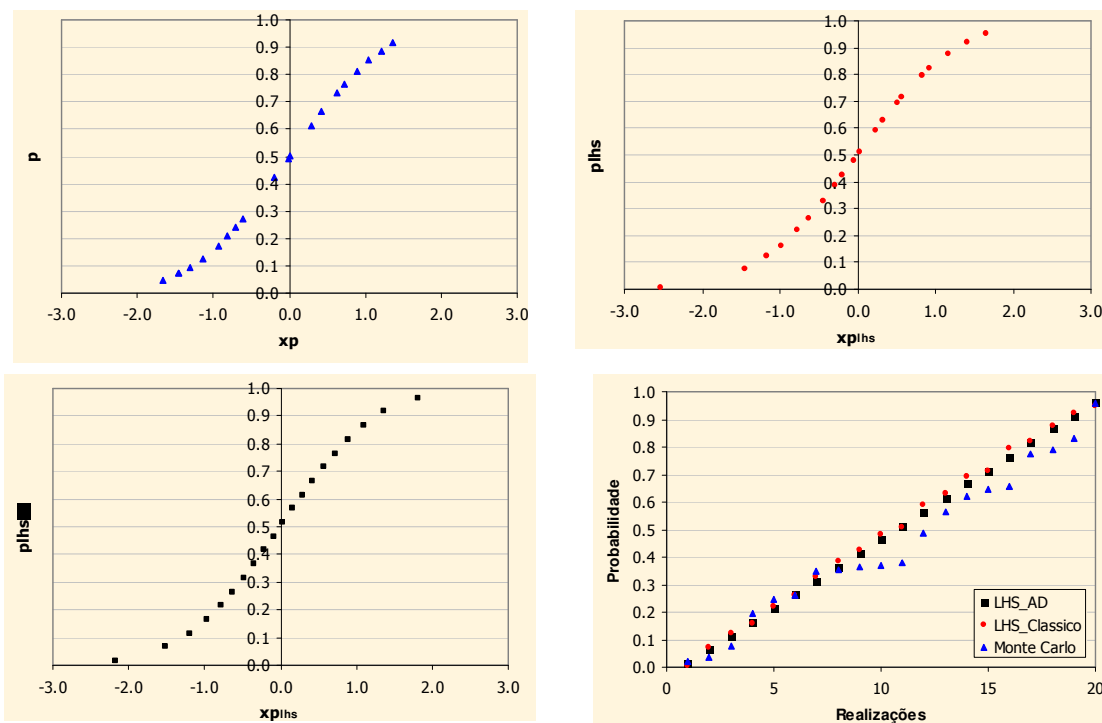


Figura 5.11: Distribuições experimentais de probabilidade Gaussiana a partir de 20 realizações, de acordo com o método de Monte Carlo (cima esquerda), LHS (cima direita) e LHS descritivo (baixo esquerda). Gráfico de probabilidade, segundo os três métodos (baixo direita).

O gráfico de probabilidade e as distribuições experimentais da dp padrão $G(y)$, obtidos a partir de 30 realizações, são apresentados na Figura 5.12. Em todos os trechos da curva experimental, construída a partir do método de Monte Carlo, é possível verificar vários agrupamentos de pontos. Como consequência, a dp padrão $G(y)$ é mal representada. Isso pode ser verificado no gráfico de probabilidade. No gráfico de probabilidade, embora os pontos referentes à técnica LHS clássica, não configurem exatamente uma reta, é evidente que seu desempenho é melhor em relação àquele apresentado pela técnica de Monte Carlo.

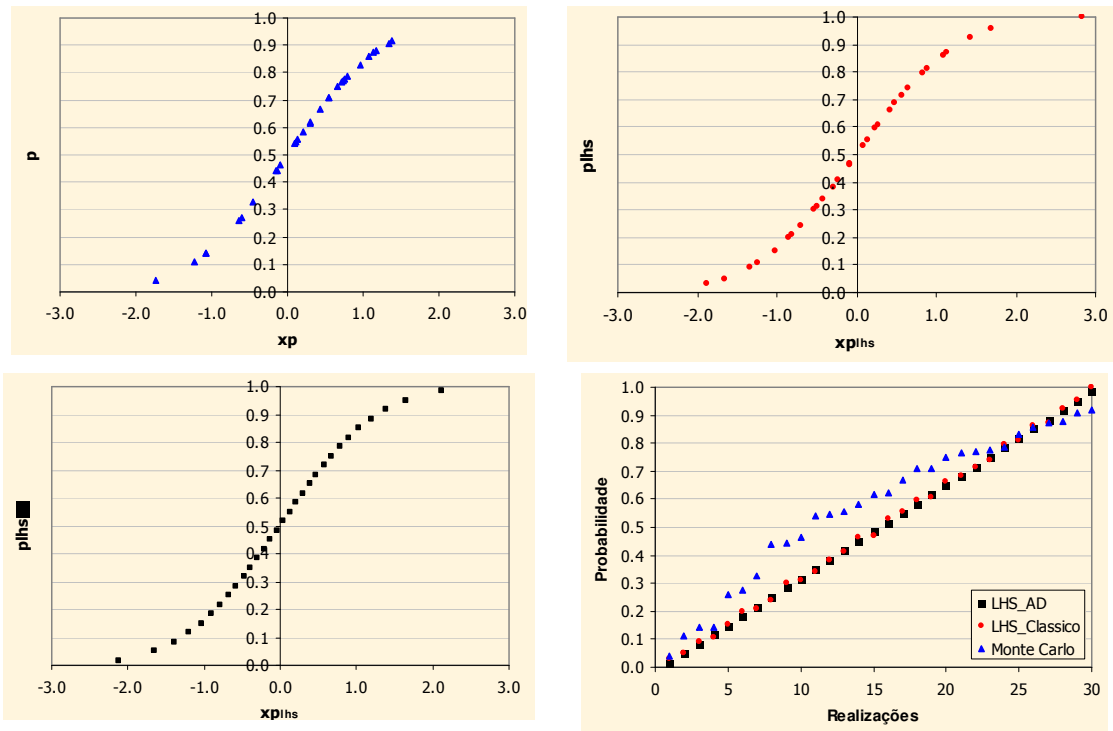


Figura 5.12: Distribuições experimentais de probabilidade Gaussiana a partir de 30 realizações, de acordo com o método de Monte Carlo (cima esquerda), LHS (cima direita) e LHS descritivo (baixo esquerda). Gráfico de probabilidade, segundo os três métodos (baixo direita).

As Figuras 5.13 e 5.14 apresentam os gráficos de probabilidade e as distribuições experimentais $G(y)$, resultantes de 50 e 100 realizações, respectivamente. Em ambas as figuras, embora a aparência das curvas experimentais seja similar, a diferença entre os métodos de amostragem é grande, como mostram os gráficos de probabilidade. Nesses gráficos, os pontos oriundos das técnicas LHS (clássica e descritiva) configuram uma reta, ao contrário dos pontos relacionados à técnica de Monte Carlo. No gráfico de probabilidade, associado à técnica de Monte Carlo para 100 realizações, é possível observar vários patamares. Essa característica indica as porções da curva onde existem agrupamentos de tiragens.

Com a utilização da técnica LHS, a amplitude das classes se torna menor a medida em que o número de classes de discretização (M) aumenta. Isso faz com que o componente aleatório, presente no algoritmo, se torne cada vez mais restrito. Por exemplo, quando $M = 100$, as classes apresentam amplitude de 1%. Esse percentual representa o grau de aleatoriedade da técnica LHS clássica para tal discretização. Portanto, a técnica LHS clássica praticamente se identifica com a técnica LHS descritiva quando $M \geq 100$. As curvas da Figura 5.14 ilustram esse comportamento.

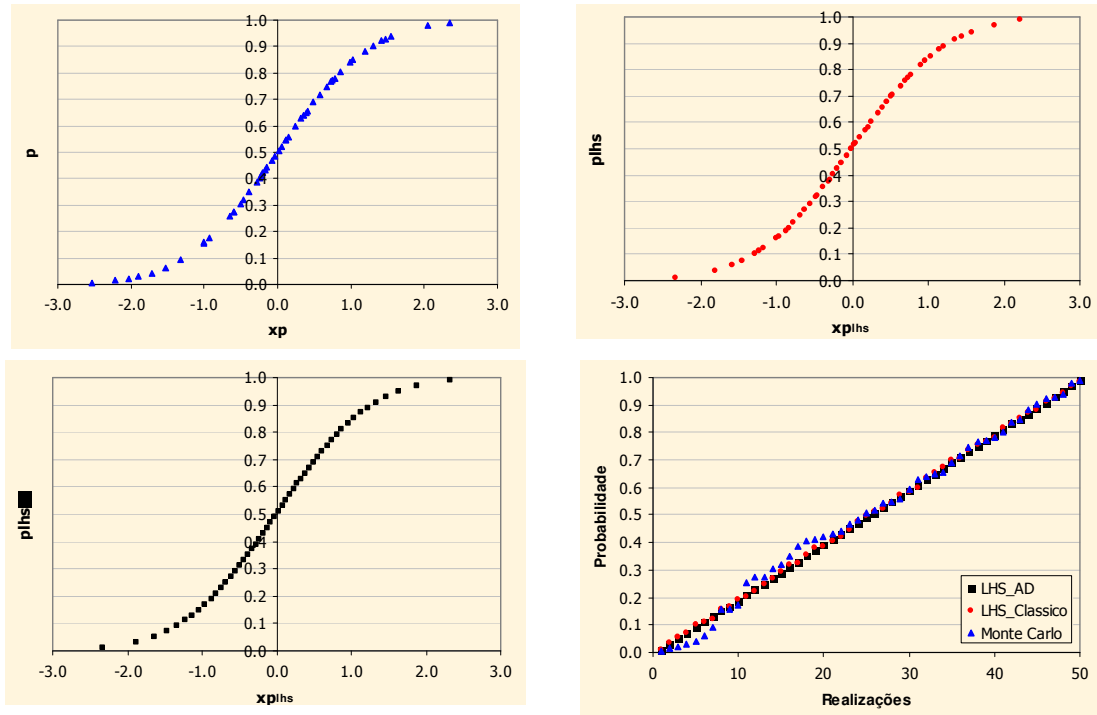


Figura 5.13: Distribuições experimentais de probabilidade Gaussiana a partir de 50 realizações, de acordo com o método de Monte Carlo (cima esquerda), LHS (cima direita) e LHS descritivo (baixo esquerda). Gráfico de probabilidade, segundo os três métodos (baixo direita).

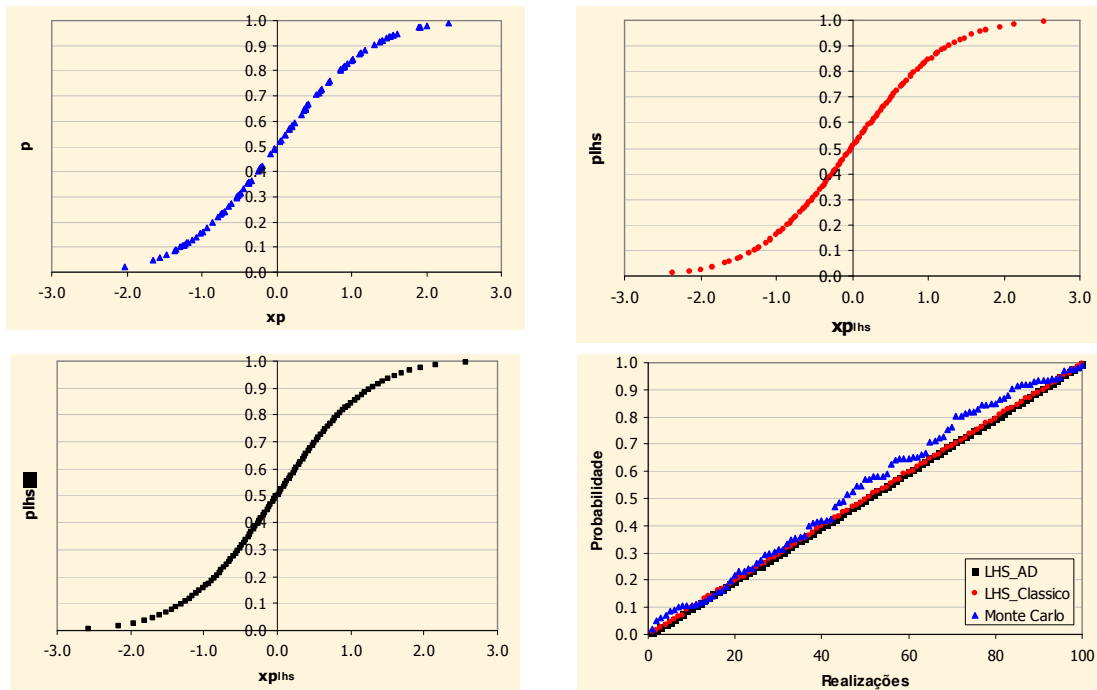


Figura 5.14: Distribuições experimentais de probabilidade Gaussiana a partir de 100 realizações, de acordo com o método de Monte Carlo (cima esquerda), LHS (cima direita) e LHS descritivo (baixo esquerda). Gráfico de probabilidade, segundo os três métodos (baixo direita).

As figuras anteriores (Figuras 5.9 a 5.14) ilustraram um comparativo entre as técnicas de amostragem de Monte Carlo e LHS (clássica e descritiva), para processos de simulação de L realizações. Se esses testes fossem repetidos, os resultados seriam os mesmos? Para $M \geq 15$, os resultados obtidos por meio da técnica LHS clássica seriam similares; com a similaridade aumentando com a diminuição do grau de aleatoriedade (inversamente proporcionais). Entretanto, as distribuições experimentais $G(y)$, derivadas por meio da técnica de Monte Carlo, seriam completamente diferentes (a não ser por coincidência). A Figura 5.15 apresenta três processos de simulação de 20 realizações, em uma determinada localização u , de acordo com essas duas técnicas de amostragem.

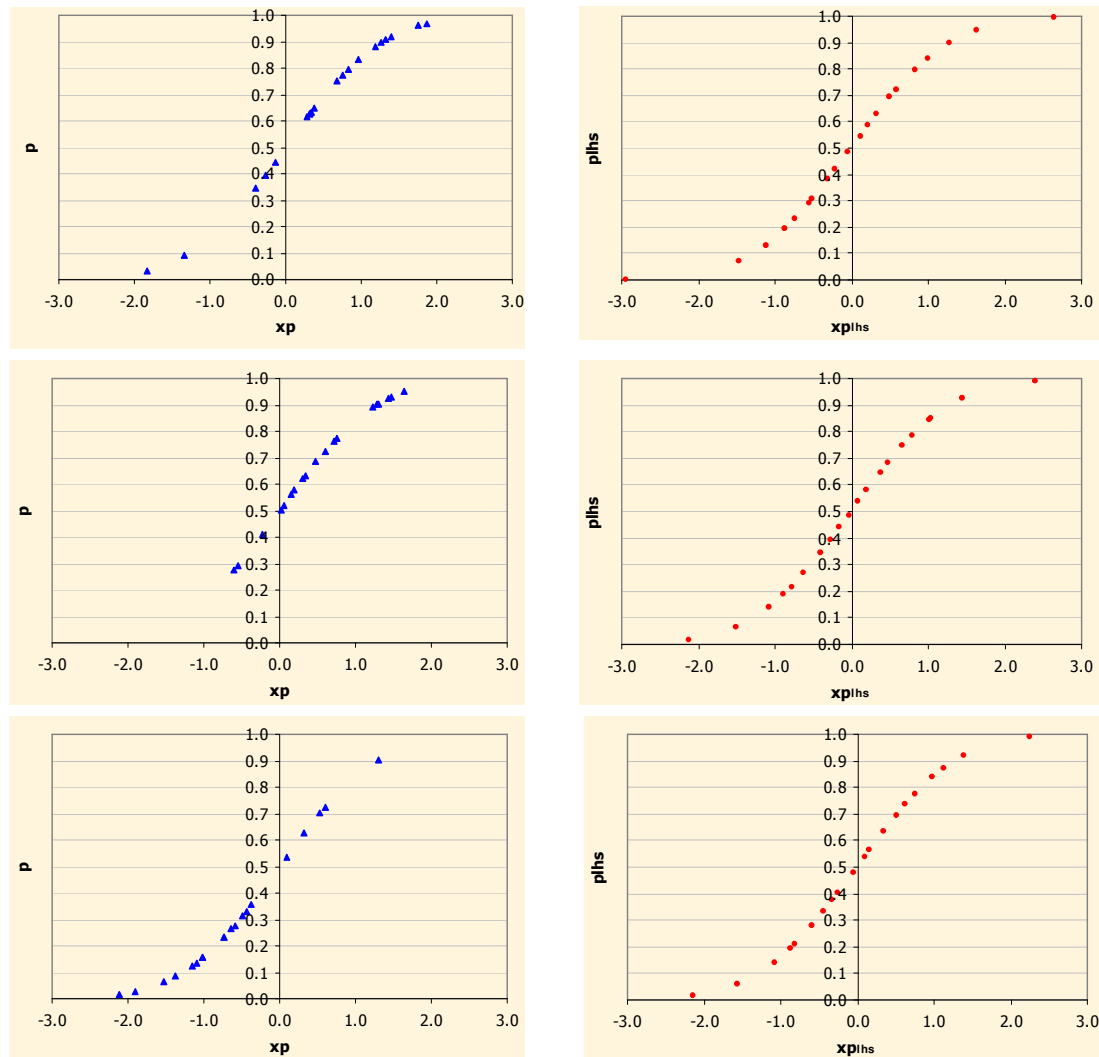


Figura 5.15: Distribuições experimentais de probabilidade Gaussiana, a partir de processos de simulação distintos de 20 realizações, de acordo com o método de Monte Carlo (esquerda) e LHS (direita).

Nessa figura, é visível que os três processos de simulação geram resultados bastante diferentes quando o método de Monte Carlo é aplicado. Em contrapartida, os resultados obtidos a partir da técnica LHS são muito similares. Isso explica porque, tradicionalmente, se recomenda

executar 100 realizações, quando se utiliza métodos de simulação condicional seqüencial. Com a técnica LHS, o espaço de incerteza pode ser suficientemente bem coberto a partir de um número menor de realizações.

Portanto, a técnica LHS apresenta mais uma característica vantajosa em relação ao método de Monte Carlo. A técnica LHS, além proporcionar tiragens que melhor representam a dp padrão $G(y)$, mostra vantagem quanto à “reprodutibilidade” dos resultados. Os resultados não são reproduzidos, mas são similares, pois uma tiragem é realizada de cada classe.

5.6. Comentários

Esse capítulo apresentou a técnica LHS (*Latin Hypercube Sampling*) formalmente, mostrando como ela pode ser útil quando combinada nos métodos seqüenciais de simulação estocástica. O capítulo iniciou com breve histórico da técnica, relacionando os métodos similares e aqueles que a deram origem. Em seguida, foi introduzido o conceito de Quadrado Latino, uma das bases teóricas do método. O método LHS foi apresentado em seu formato original, como um método redutor de variância, destinado à VA independentes. Sua versão destinada à VA dependentes também foi apresentada. Na seqüência, foi mostrado como o método de Monte Carlo atua nas rotinas de simulação condicional seqüencial e como o método LHS pode substituí-lo. Três variações foram apresentadas:

- i. LHS cíclico;
- ii. LHS descritivo;
- iii. LHS com caminho descritivo.

Foi explicado como o algoritmo pode ser implementado em rotinas de simulação geoestatística que são escritas em linguagem de programação Fortran. Finalmente, foram apresentados vários testes de verificação teórica da rotina implementada. Todos os testes foram satisfatórios. O final do capítulo contou com comparativos entre a rotina original e a rotina modificada. Esses comparativos indicaram vantagens para a técnica LHS, comprovando que a mesma pode ser implementada com sucesso nos algoritmos de simulação geoestatística.

O Capítulo 6 abordará um estudo de caso, no qual a rotina modificada (LHS) será aplicada, mostrando as diferenças e impactos provocados.

Capítulo 6

Estudo de caso: LHS aplicado

O Capítulo 6 apresenta um estudo de caso, no qual o algoritmo proposto, assim como suas variações, são aplicadas. São realizadas comparações entre as performances do algoritmo tradicional, que utiliza a técnica de Monte Carlo, e dos algoritmos que envolvem a técnica LHS. Essas comparações mostram, principalmente, o impacto provocado em relação à média das realizações, com enfoque sobre exatidão e precisão. Também são realizadas comparações de caráter prático, aplicando funções de transferência, como por exemplo, curvas de parametrização.

6.1. Banco de dados

O Capítulo 6 é desenvolvido a partir do banco de dados de domínio público Walker Lake. Nesse banco de dados, três variáveis são disponíveis, duas variáveis contínuas (V e U) e uma categórica (T). Entretanto, para esse estudo, foi utilizado um subconjunto de amostras da variável V, medida em ppm. Esse subconjunto é constituído por 195 amostras, regularmente espaçadas em uma área de 6,88 ha, como mostra a Figura 6.1. O espaçamento entre as amostras é de cerca de 20 m.

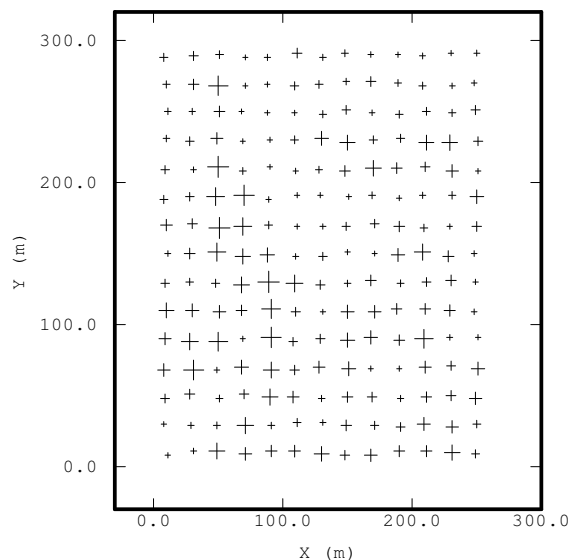


Figura 6.1: Mapa de localização dos furos de sondagem, com símbolos proporcionais aos teores (ppm) associados a variável V.

A Figura 6.1 mostra a localização dos furos de amostragem, por meio de símbolos proporcionais aos teores associados, proporcionando uma primeira impressão sobre a variabilidade espacial inerente aos dados.

6.2. Análise exploratória dos dados

A análise exploratória das amostras, provenientes da sondagem, revelou os parâmetros estatísticos apresentados na Figura 6.2. Essa figura, também, ilustra o histograma da distribuição, evidenciando sua assimetria positiva. Esse comportamento é característico de distribuições de metais como Au, nas quais o valor calculado para média é superior ao da mediana (Q50). Cada classe da Figura 6.2 apresenta amplitude de 25 ppm. Isso indica que 17,44 % das amostras (34 dados) assumem valores da classe $[0,0; 25,0[$. Enquanto que apenas 1 dado (0,005 %) se insere na classe $[975,0; 1000,0]$.

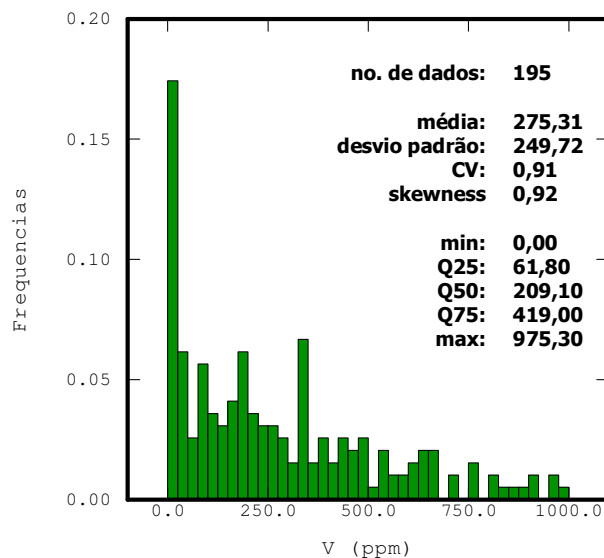


Figura 6.2: Histograma e parâmetros estatísticos da distribuição da variável V (ppm).

A Figura 6.3 mostra um mapa de localização das amostras no qual os dados foram classificados de acordo com os quartis da distribuição. Essa figura complementa a Figura 6.1, mostrando com clareza a distribuição espacial dos dados. Nela, é possível verificar uma concentração de altos teores no centro-oeste da área (correspondente aos dados da classe extrema superior do histograma da Figura 6.2). Como os furos de sondagem são regularmente espaçados, não há necessidade de aplicação de algoritmos de desagrupamento.

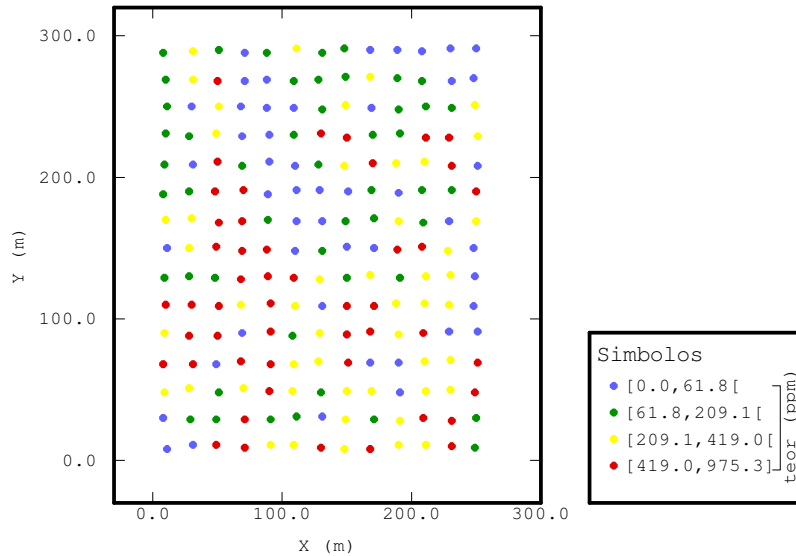


Figura 6.3: Mapa de localização dos furos de amostragem, de acordo com os quartis da distribuição da variável V (ppm).

6.2.1. Modelo de variabilidade espacial

Após a análise do variograma onidirecional, para calibração do efeito pepita e para uma primeira idéia sobre os alcances médios, os variogramas experimentais foram calculados segundo quatro direções. Esse procedimento mostrou a presença de anisotropia geométrica. Os variogramas experimentais direcionais foram modelados com o auxílio de três estruturas: uma de efeito pepita e duas esféricas. A primeira estrutura é isotrópica, ao contrário das outras duas. A Figura 6.4 ilustra os variogramas experimentais e teóricos segundo quatro direções.

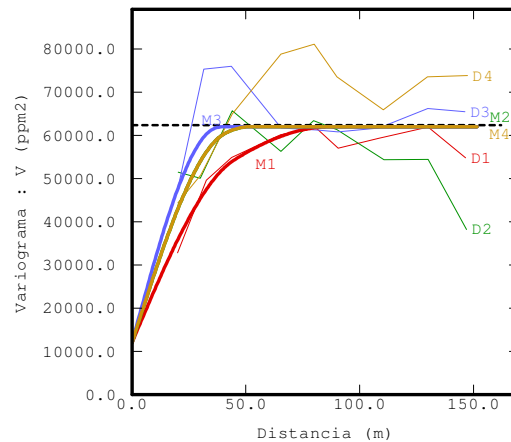


Figura 6.4: Variogramas experimentais (D) e modelos (M) para quatro direções. D1 e M1: azimute 158°; D2 e M2: azimute 113°; D3 e M3: azimute 68° e D4 e M4: azimute 23°.

O modelo de variabilidade espacial respeita a seguinte equação:

$$\gamma(h) = 12000 \cdot \text{efeito pepita} + \left\{ \begin{array}{l} 24000 \cdot \text{sph} \left(\frac{\text{azimute } 158^\circ}{45 \text{ m}}, \frac{\text{azimute } 68^\circ}{37 \text{ m}} \right) \\ + \\ 26000 \cdot \text{sph} \left(\frac{\text{azimute } 158^\circ}{87 \text{ m}}, \frac{\text{azimute } 68^\circ}{40 \text{ m}} \right) \end{array} \right\} \quad (6.1)$$

6.2.2. Malha de simulação

A simulação será realizada no suporte das amostras, em um malha quadrada pontual, cobrindo toda a área de amostragem (6,88 ha). A malha é constituída de 260 nós na direção leste-oeste e 300 nós na direção norte-sul, resultando em um total de 78000 nós. Cada nó é separado por um quadrado de lado unitário (um metro) do nó adjacente, configurando uma malha de dimensão 1 x 1 m. Considerando que cada amostra apresenta o suporte de um nó, para os processos de simulação decorrentes, não haverá a necessidade de mudança de suporte.

6.3. Anamorfose Gaussiana

Como descrito no Capítulo 3, a aplicação do método de ssG exige que VA original seja transformada em uma VA Gaussiana. Assim, a distribuição experimental original foi transformada em uma distribuição Gaussiana, por meio de uma função de anamorfose $\varphi(\cdot)$ (Equação 3.21, p. 32). Na prática, a função de anamorfose que relaciona a VA original e a VA Gaussiana, é, na verdade, uma função de anamorfose empírica $\hat{\varphi}(\cdot)$ (Figura 3.6, p. 36). A Figura 6.5 mostra a função de anamorfose empírica, assim como o modelo de anamorfose. A função de anamorfose empírica foi modelada com o auxílio dos polinômios de Hermite, em um polinômio de grau 30 (30 monômios). Foi decidido não extrapolar os trechos extremos da distribuição original, mantendo os valores extremos inalterados, configurando uma atitude conservadora.

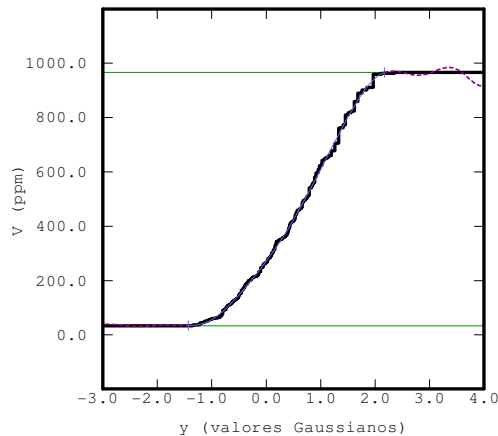


Figura 6.5: Anamorfose Gaussiana, empírica (linha espessa preta) e modelada (linha fina azul).

A Figura 6.6 ilustra o histograma e os parâmetros estatísticos da distribuição transformada (Gaussiana). A forma do histograma e suas características estatísticas são como as esperadas. Nesse histograma, as classes apresentam amplitude de 0,2. Na prática, são essas características de 1ª ordem que devem ser reproduzidas pelas realizações provenientes dos processos de simulação.

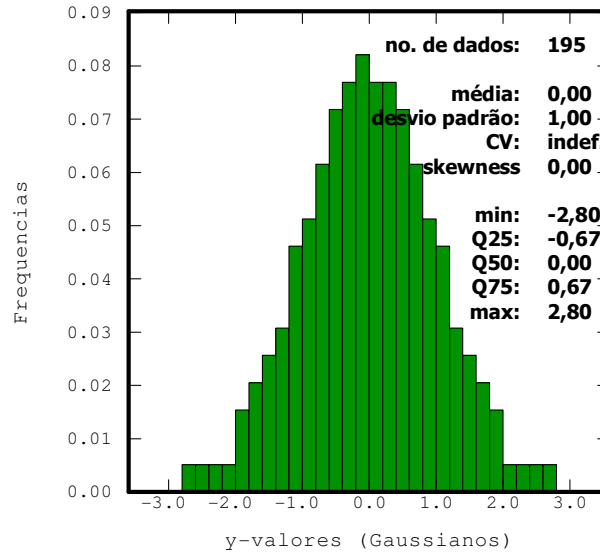


Figura 6.6: Histograma e parâmetros estatísticos da distribuição da variável Gaussiana.

6.3.1. Verificação de biGaussianidade

Para a verificação da biGaussianidade, exigida pelo método, foram realizados dois testes:

- i. relação madograma/semi-variograma e
- ii. diferença entre os valores de pares separados pelo vetor h onidirecional.

O teste da relação madograma/semi-variograma, apresentado no Capítulo 3 (p. 35), foi realizado a partir dessas funções, para vetores h onidirecionais. A Figura 6.7 apresenta um gráfico no qual é possível verificar o resultado da relação para diferentes dimensões dos vetores h (passo igual a 10 m, com tolerância de mais ou menos cinco metros).

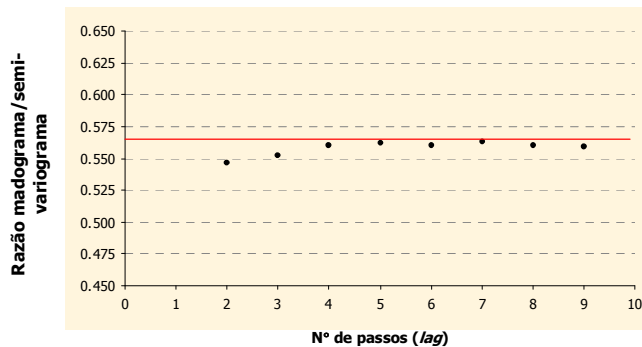


Figura 6.7: Teste de binormalidade da razão madograma/semi-variograma para os madogramas/semi-variogramas onidirecionais. A linha vermelha representa o valor ideal para a relação (Equação 3.25).

O segundo teste de binormalidade, consiste em verificar se a distribuição resultante da diferença entre pares de y -valores, separados pelo vetor h , caracteriza uma distribuição Gaussiana. Ou seja, se a distribuição resultante de $y(u) - y(u + h)$, $\forall u, \forall h$, também apresenta as características de uma distribuição normal. Para isso, foram considerados os pares de dados separados pelo vetor h onidirecional. Foram testadas quatro dimensões para o vetor h : 20, 40, 60 e 80 m. Esses vetores contaram com uma tolerância de meio metro para mais ou para menos. Portanto, foram considerados os pares de dados separados por:

- i. 19,5 a 20,5 m;
- ii. 39,5 a 40,5 m;
- iii. 59,5 a 60,5 m e
- iv. 79,5 a 80,5 m.

A Figura 6.8 apresenta os histogramas cumulativos e os parâmetros estatísticos das distribuições resultantes.

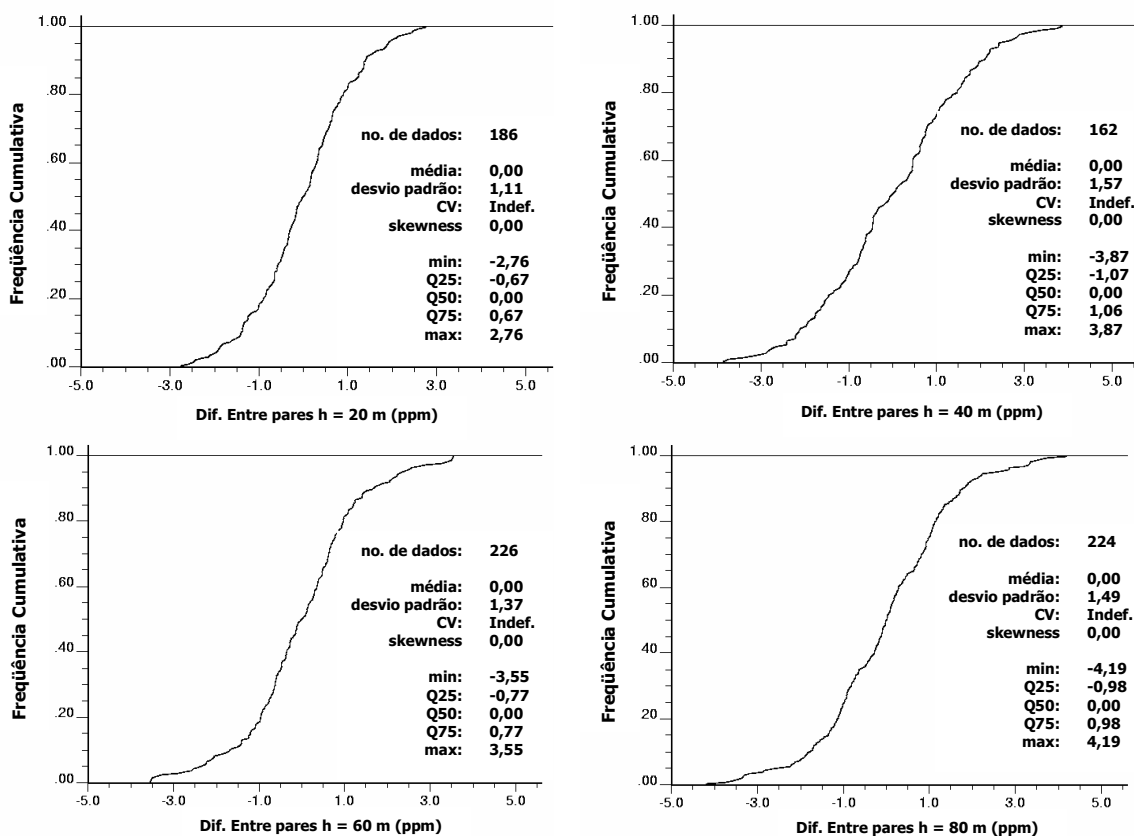


Figura 6.8: Histogramas de frequência cumulativa e parâmetros estatísticos para as distribuições da diferença de pares de dados, separados pelo vetor h onidirecional.

Os resultados foram satisfatórios para as quatro distribuições testadas. O valor assumido pela média das diferenças foi nulo para as quatro distribuições, como o esperado. Os desvios padrão das distribuições, embora não exatamente Gaussianos, assumiram valores muito próximos à unidade. Os

quatro histogramas também apresentaram simetria Gaussiana, denotado pelo grau de assimetria (skewness) nulo.

Portanto, os resultados desses testes validam a condição de binormalidade da distribuição Gaussiana resultante.

6.3.2. Modelo de variabilidade espacial Gaussiano

Com a distribuição Gaussiana validada, na sua condição bivariada, foi possível estudar a variabilidade espacial da VA Gaussiana $Y(u)$ resultante. Os semi-variogramas experimentais direcionais, revelaram praticamente a mesma anisotropia geométrica observada para a distribuição original. Entretanto, nesse caso, o semi-variograma foi modelado com o auxílio de duas estruturas de base: uma de efeito pepita e outra esférica. A Figura 6.9 ilustra os semi-variogramas direcionais experimentais e os respectivos modelos, segundo quatro direções: azimute 158°, azimute 113°, azimute 68° e azimute 23°.

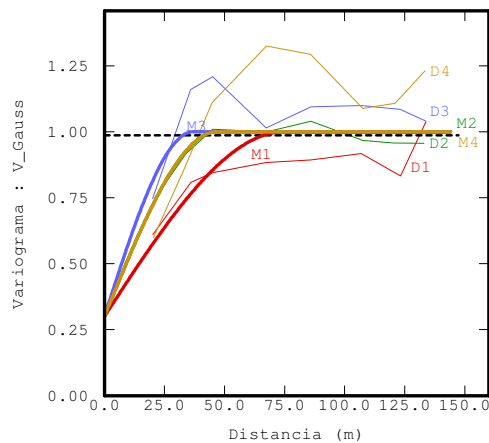


Figura 6.9: Variogramas experimentais (D) e modelos (M) da variável Gaussiana, para quatro direções. D1 e M1: azimute 158°; D2 e M2: azimute 113°; D3 e M3: azimute 68° e D4 e M4: azimute 23°.

A Equação 6.2 representa o modelo de variabilidade espacial da distribuição da VA $V(u)$, após a anamorfose Gaussiana. Para o caso Gaussiano, foi observada uma pequena redução na dimensão dos eixos da elipse. Entretanto, a razão de anisotropia é praticamente a mesma (~ 2).

$$\gamma(h) = 0,3 \cdot \text{efeito pepita} + 0,7 \cdot \text{sph} \left(\frac{\text{azimute } 158^\circ}{75 \text{ m}}, \frac{\text{azimute } 68^\circ}{37 \text{ m}} \right) \quad (6.2)$$

O modelo de variabilidade espacial Gaussiano (Equação 6.2) e o histograma, juntamente com os parâmetros estatísticos, da Figura 6.6, constituem os momentos de 1ª e 2ª ordem condicionantes dos processos de simulação decorrentes. Contudo, após a transformação inversa, os histogramas e semi-variogramas das realizações também devem reproduzir, em média, os momentos de 1ª e 2ª ordem da distribuição original, representados na Equação 6.1 e na Figura 6.2.

6.3.3. Validação do modelo de variabilidade espacial e estratégia de busca

Antes que o modelo de variabilidade espacial Gaussiano fosse tomado como condicionante, o mesmo foi validado, por meio do procedimento de validação cruzada. Esse recurso também foi utilizado para a definição da estratégia de busca, a ser realizada para a construção das matrizes de krigagem.

Foram testadas várias composições de estratégia de busca, variando o número mínimo e máximo de dados, a dimensão da elipse e o tipo de busca (por quadrantes ou não). Os resultados mais satisfatórios foram obtidos a partir da estratégia apresentada na Tabela 6.1.

Tabela 6.1: Parâmetros da estratégia de busca.

Elipse		Nº de quadrantes	Nº mínimo de dados	Nº máximo de dados
Orientação	Dimensão (eixos)			
Azimute 158°	90 x 50 m	1	4	16

No processo de validação cruzada, é interessante observar se os erros normalizados (pelos desvios padrão das estimativas) se inserem dentro de um determinado intervalo de confiança de uma distribuição Gaussiana. Geralmente, os erros normalizados que se inserem dentro desses intervalos são chamados de robustos. De forma padrão, o intervalo $[-2,5; 2,5]$ define os erros normalizados dentro do limite de confiança de 99% de uma distribuição Gaussiana. A Tabela 6.2 apresenta a porcentagem dos erros classificados como robustos, assim como um comparativo entre os erros robustos e brutos, utilizando o intervalo $[-2,5; 2,5]$. Os resultados do procedimento de validação cruzada são apresentados na Figura 6.10.

Tabela 6.2: Comparação estatística entre os erros robustos e brutos, após o processo de validação cruzada.

Estimativas robustas (%)	Erro robusto		Erro bruto	
	Média	Variância	Média	Variância
97,95	-0,0394	0,5873	0,0127	0,7052

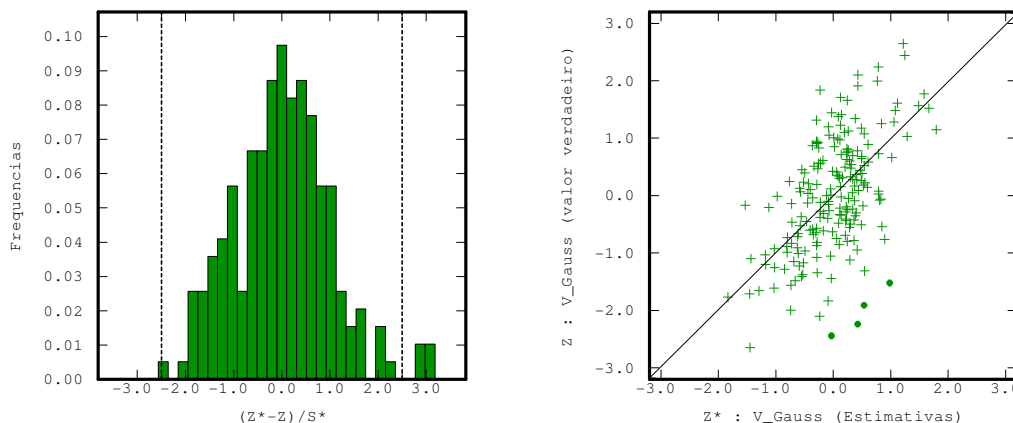


Figura 6.10: Histograma da distribuição dos erros normalizados sobre os desvios padrão das estimativas (esquerda) e diagrama de dispersão entre os valores verdadeiros e estimados (direita).

A Figura 6.10 mostra o histograma dos erros normalizados e um gráfico de dispersão entre os valores reais e estimados. No histograma, as linhas tracejadas indicam os valores limite do intervalo $[-2,5; 2,5]$. Os símbolos redondos, no gráfico de dispersão, representam os dados nos quais os erros das estimativas foram classificados como não robustos. A Figura 6.11 ilustra um mapa que indica o erro bruto, de maneira proporcional (o tamanho do símbolo é proporcional ao valor do erro bruto). Nesse mapa, é possível observar as localizações onde os dados foram subestimados e superestimados.

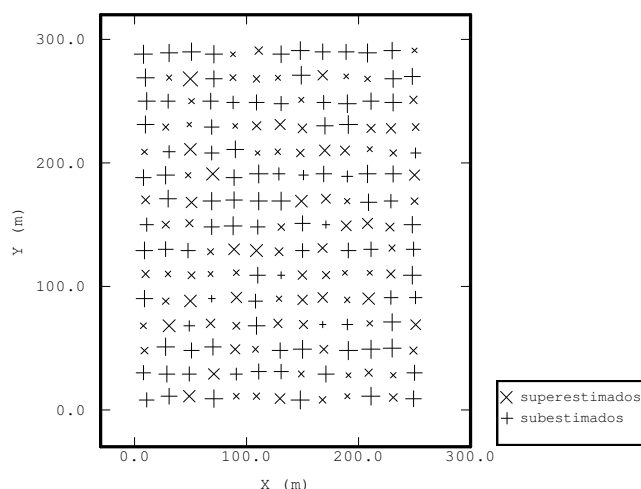


Figura 6.11: Mapa proporcional do erro bruto, indicando as localizações onde os dados foram subestimados e superestimados, segundo o processo de validação cruzada.

Os resultados do processo de validação cruzada, além de indicar a estratégia de busca mais adequada a ser utilizada, validaram o modelo de variabilidade espacial da VA Gaussiana $Y(u)$.

6.4. Simulação seqüencial condicional

O modelo de variabilidade espacial da VA $Y(u)$ e a estratégia de busca, validado e definida na seção anterior (classificada como adequada), foram utilizados nos processos de simulação seqüencial condicional apresentados a seguir. Entretanto, a estratégia de busca utilizada nos processos de simulação, envolve um parâmetro não mencionado anteriormente: o número de dados previamente simulados. Após alguns experimentos, esse parâmetro foi definido como sendo oito. Portanto, em cada nó da malha, a matriz de KS é construída a partir de no máximo 16 dados originais, além de oito previamente simulados. Essa captura de dados é realizada a partir da elipse de busca, centrada no nó da malha a ser simulado. A Figura 6.12 ilustra o esquema da estratégia de busca utilizada nos processos de simulação. Nessa figura, a malha não é representada, pois é muito densa (1 x 1 m).

Os processos de simulação foram realizados segundo os algoritmos de ssG original (MC) e modificado (LHS). Evidentemente, as variações no algoritmo, envolvendo a técnica LHS (apresentadas

no Capítulo 5), também foram aplicadas. Todos os processos de simulação, apresentados a seguir, foram realizados a partir da mesma estratégia de busca. Como mencionado anteriormente, esses processos foram realizados por meio da rotina *sgsim* da GSLIB (Deutsch e Journel, 1998, pp. 170 - 174). Os algoritmos envolvendo a técnica LHS também foram implementados sobre essa rotina. A Figura 6.13 apresenta o arquivo de parâmetros de simulação, necessário à execução da rotina.

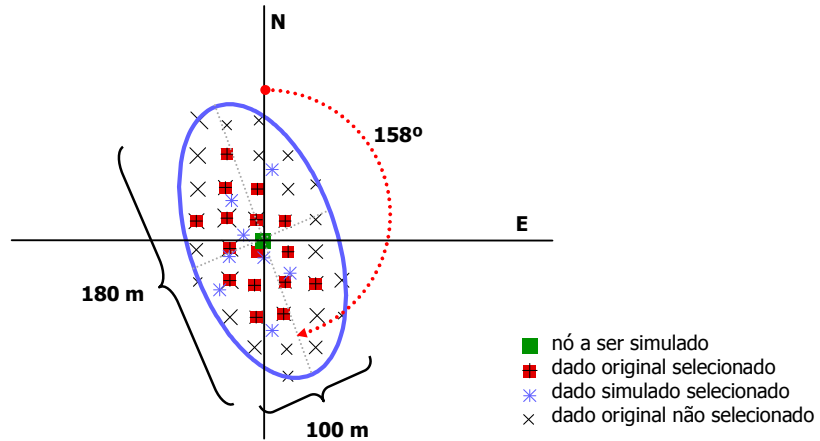


Figura 6.12: Esquema da estratégia de busca utilizada nos processos de simulação.

```

WalkerC1_normal.out      \ file with data
2 3 0 5 0 0             \ columns for X,Y,Z,vr,wt,sec.var.
1.0e21 1.0e21           \ trimming limits
0                        \ transform the data (0=no, 1=yes)
walker_nscore#1.trn     \ file for output trans table
0                        \ consider ref. dist (0=no, 1=yes)
histsmth.out            \ file with ref. dist distribution
1 2                     \ columns for vr and wt
0.0 975.3               \ zmin,zmax(tail extrapolation)
2 2.5                   \ lower tail option, parameter
4 2.0                   \ upper tail option, parameter
0                        \ debugging level: 0,1,2,3
sgsim_walker.dbg        \ file for debugging output
sgsim_walker.out        \ file for simulation output
8                        \ number of realizations to generate
260 0.0 1.0             \ nx,xmn,xsiz
300 0.0 1.0             \ ny,ymn,ysiz
1 0.0 1.0               \ nz,zmn,zsiz
73073                   \ random number seed
4 0.5                   \ number of classes and proportion
4 16                    \ min and max original data for sim
8                        \ number of simulated nodes to use
0                        \ assign data to nodes (0=no, 1=yes)
1 3                      \ multiple grid search (0=no, 1=yes),num
0                        \ maximum data per octant (0=not used)
90.0 50.0 1.0           \ maximum search radii (hmax,hmin,vert)
158.0 00.0 0.0          \ angles for search ellipsoid
0 0.0                   \ ktype:
                        0=SK,1=OK,2=LVM,3=EXDR,4=COLC
../data/ydata.dat       \ file with LVM, EXDR, or COLC variable
0                        \ column for secondary variable
1 0.3                   \ nst, nugget effect
1 0.7 158.0 00.0 00.0   \ it,cc,ang1,ang2,ang3
                        75.0 37.0 00.0
                        \ a_hmax, a_hmin, a_vert
    
```

Figura 6.13: Arquivo de parâmetros dos processos de simulação (padrão GSLIB).

6.4.1. Algoritmo original

Para caracterizar o espaço de incerteza da VA V , foram executadas 200 realizações por meio do algoritmo de ssG original, isto é, aquele que utiliza a técnica de Monte Carlo (MC). A Figura 6.14 mostra as médias e os desvios padrão das realizações, no espaço normal. Nessa figura, a linha espessa vermelha representa os valores condicionantes. Uma análise sobre essa figura indica que o processo de simulação foi realizado com sucesso, pois os parâmetros de média e desvio padrão das realizações oscilaram simetricamente em torno dos valores condicionantes. Esse comportamento configura uma característica conhecida como flutuações ergódicas.

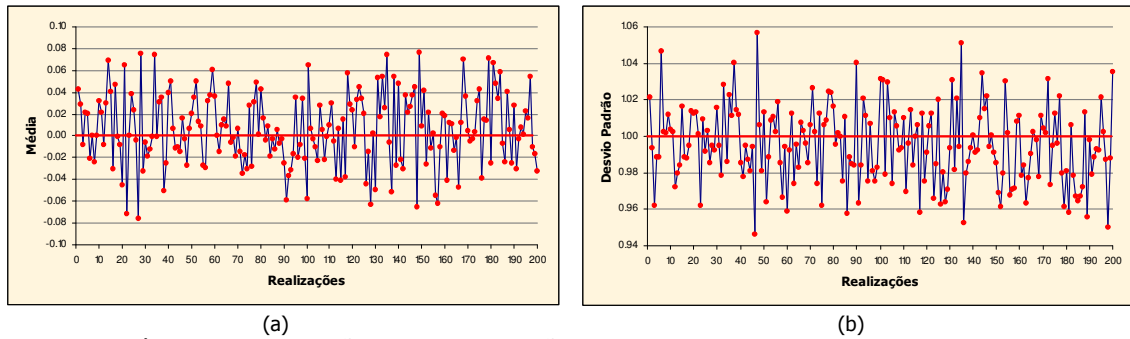


Figura 6.14: Médias (a) e desvio padrão (b) das 200 realizações obtidas por meio do algoritmo original – no espaço normal.

No gráfico relativo à média, 109 realizações mostraram valores médios superiores a zero. Em relação ao desvio padrão, 85 realizações obtiveram valores superiores a um. Além da oscilação simétrica, o desvio padrão desses parâmetros se mostrou reduzido, como recomendado. A Figura 6.15 apresenta o histograma (classes com 0,05 de amplitude) e os parâmetros estatísticos das médias e dos desvios padrão das 200 realizações. Nessa figura, pode ser verificado que, em média, os momentos de 1ª e 2ª ordem da distribuição normalizada condicionante foram reproduzidos. Ainda, os histogramas revelam simetria, comprovando o comportamento oscilatório homogêneo das realizações.

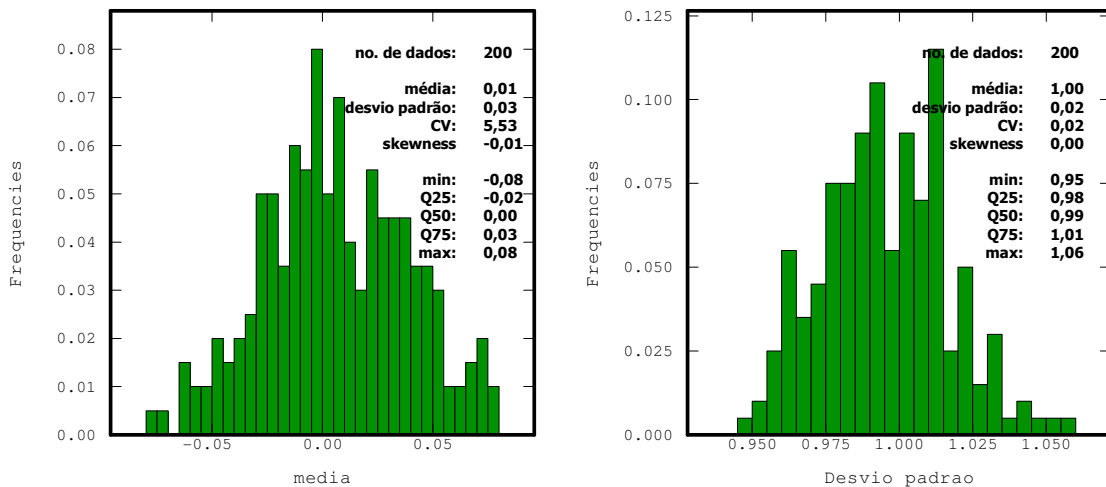


Figura 6.15: Histogramas das médias (esquerda) e dos desvios padrão (direita) das 200 realizações obtidas por meio do algoritmo original – no espaço normal.

Após a transformação inversa, os resultados do processo de simulação também se mostraram satisfatórios. A Figura 6.16 apresenta os desvios padrão e os valores médios assumidos pelas 200 realizações, no espaço original. Como no espaço normal, também foram verificadas oscilações ergódicas, praticamente homogêneas, em torno dos momentos condicionantes (linha espessa vermelha).

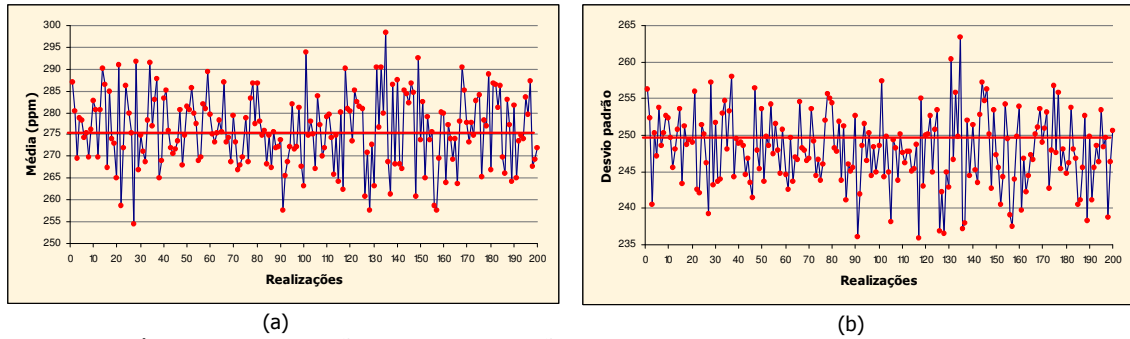


Figura 6.16: Médias (a) e desvio padrão (b) das 200 realizações obtidas por meio do algoritmo original – no espaço original.

Dentre as 200 realizações, 103 apresentaram média superior à média condicionante (275,31). Em relação ao desvio padrão, 73 realizações obtiveram valores superiores ao condicionante (249,72). Esse comportamento é evidenciado no grau de simetria dos histogramas da Figura 6.17. Nesses histogramas, as classes têm 1 ppm de amplitude. Essa figura, também, apresenta os parâmetros estatísticos dessas distribuições. Pode ser observado que, no espaço original, as flutuações ergódicas também não se mostraram excessivas. Essa característica, juntamente com a reprodução dos momentos condicionantes, validam o processo de simulação no espaço original.

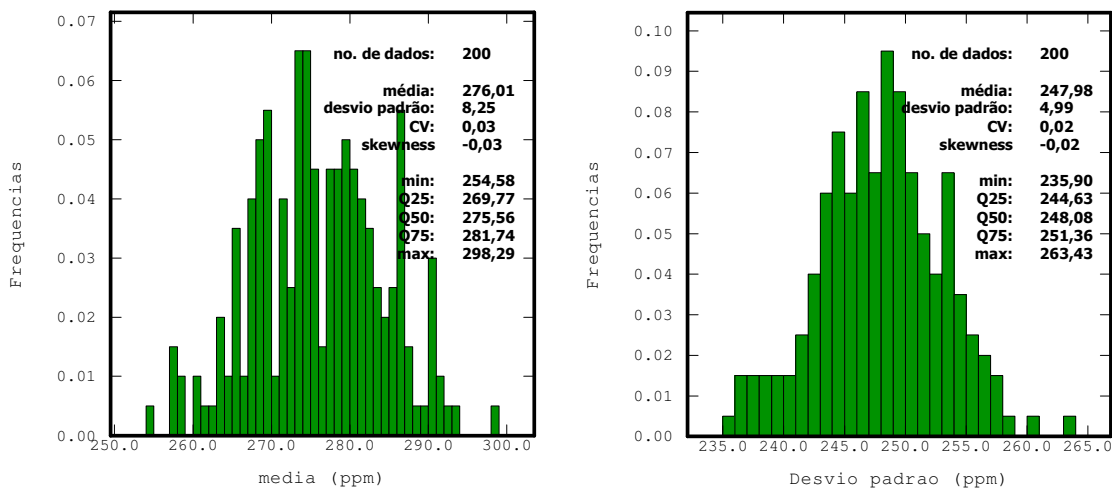


Figura 6.17: Histogramas das médias (esquerda) e dos desvios padrão (direita) das 200 realizações obtidas por meio do algoritmo original – no espaço original.

Após o processo de simulação, foi realizada uma análise de variabilidade entre os 200 cenários equiprováveis gerados. Para essa análise, foram verificados a média e o coeficiente de

variação (CV) dos conjuntos de médias das realizações, de acordo com as Equações (6.3) e (6.4), respectivamente. O passo utilizado foi de 10 realizações. Isto é, primeiramente foram analisadas as 10 primeiras realizações, depois as 20 primeiras e assim por diante. Por exemplo, para o conjunto c das 20 primeiras realizações (c = 20), foram analisados a média das médias (E[m]₂₀) e o CV das médias CV(m)₂₀ desse conjunto. A Tabela 6.3 apresenta os resultados desses parâmetros, além do desvio padrão [σ(m)_c], para os 20 conjuntos de realizações resultantes.

$$E[m]_c = 1/c \cdot \sum_{l=1}^c m_l \quad c = 10, 20, 30, \dots, 200 \quad (6.3)$$

Nessa equação, m_l representa a média de cada realização l (l = 1, ..., L).

$$CV(m)_c = 100 * (\sigma(m)_c/E[m]_c) \quad c = 10, 20, 30, \dots, 200 \quad (6.4)$$

Tabela 6.3: Média, desvio padrão e CV dos conjuntos de médias das realizações (c).

Conjunto de realizações (c)	E[m] _c	σ(m) _c	CV(m) _c
10	277,2930	5,1328	1,8510
20	277,2603	6,7872	2,4479
30	276,5521	9,0826	3,2842
40	276,7904	8,9203	3,2228
50	276,5138	8,3112	3,0057
60	277,0146	8,0450	2,9042
70	276,8748	7,6483	2,7624
80	276,7635	7,6160	2,7518
90	276,3851	7,3368	2,6545
100	275,7722	7,5424	2,7350
110	275,8985	7,5219	2,7263
120	275,8531	7,5958	2,7536
130	275,6264	7,7775	2,8218
140	275,9953	8,1888	2,9670
150	276,1697	8,3353	3,0182
160	275,9247	8,4071	3,0469
170	275,8921	8,3860	3,0396
180	275,9495	8,3163	3,0137
190	276,0747	8,3224	3,0145
200	276,0061	8,2531	2,9902

A Figura 6.18 (a) apresenta um gráfico mostrando a relação entre as médias das médias dos 20 conjuntos de realizações. Essa figura mostra que a média entre os valores médios, obtidos nas realizações, adquire um valor estável a partir da 160ª realização. Na Figura 6.18 (b) são apresentados os CV das médias para os mesmos conjuntos de realizações. De maneira similar ao caso da média, o CV da distribuição das médias assume valores estáveis a partir da 150ª realização. Os valores de média e de CV, considerados estáveis são, 275,92 e 3,05, respectivamente. Embora a diferença entre a média dos valores médios em relação aos conjuntos de 100 e de 160 realizações seja mínimo, o mesmo não ocorre para o CV das médias. Após assumir falsos valores estáveis a partir da 70ª

realização, a diferença entre esses valores e aqueles assumidos a partir da 150ª realização é mais significativa. Esse comportamento nem sempre seria percebido. Geralmente, assume-se $L = 100$ para garantir que todo o espaço de incerteza da VA $V(u)$ seja representado, independente das características estatísticas e de variabilidade espacial dos dados. Entretanto, esse caso alerta para o fato de que às vezes são necessárias mais do 100 realizações para que o espaço de incerteza da VA $V(u)$ seja caracterizado.

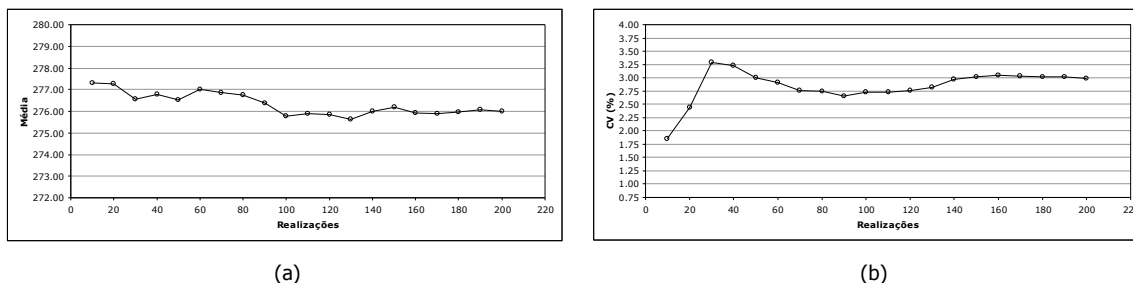


Figura 6.18: Média (a) e CV (b) da distribuição das médias das realizações, de acordo com a rotina *sgsim* original.

6.4.2. Algoritmo LHS clássico

O desempenho do algoritmo modificado, com a técnica LHS clássica, foi avaliado a partir dos resultados, obtidos segundo vários graus de discretização da distribuição G (números de classes M). Como nesse algoritmo $L = M$, foram realizados vários processos de simulação diferentes. Primeiramente, foram realizados 10 processos de simulação, utilizando 10 tipos de discretização: $M = 10, 20, 30, 40, 50, 60, 70, 80, 90$ e 100 classes. Os resultados apresentados na Tabela 6.4 validam os processos de simulação no espaço normal. Essa tabela, apresenta os valores médios das médias e dos desvios padrão das realizações, conforme os 10 processos de simulação. Em todos os processos, os valores médios foram muito similares aos respectivos valores condicionantes.

Tabela 6.4: Média das médias e dos desvios padrão para 10 processos de simulação, conforme o número de classes M – espaço normal.

Número de classes (M)	Média	
	Média	Desvio padrão
10	0,0008	0,9962
20	0,0039	0,9950
30	0,0054	0,9998
40	0,0057	0,9972
50	0,0064	0,9981
60	0,0047	0,9987
70	0,0055	0,9985
80	0,0036	0,9969
90	0,0053	0,9964
100	0,0062	0,9969

A Figura 6.19 ilustra os resultados de média e desvio padrão de dois processos de simulação: $L = M = 100$ e $L = M = 50$. Nessa figura, é observado que os parâmetros medidos oscilaram simetricamente em torno dos valores condicionantes, indicados pela linha espessa vermelha. Para $M = 100$, 53 % das realizações mostraram valores médios superiores a zero. Em relação ao desvio padrão, 42 % das realizações obtiveram valores superiores ao condicionante. Na Figura 6.19 (b) (para $M = 50$), 58 % das realizações apresentaram média superior à média condicionante. Considerando o desvio padrão, 48 % das realizações obtiveram valores superiores a um.

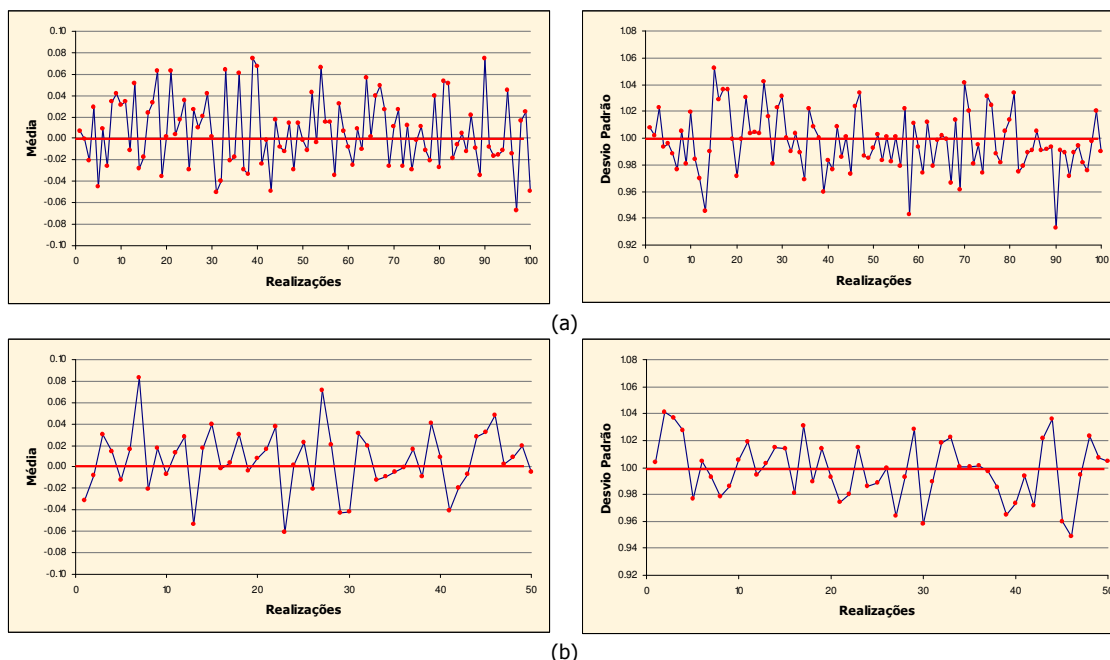


Figura 6.19: Médias (esquerda) e desvios padrão (direita) das realizações, resultantes dos processos de simulação a partir de 100 classes de discretização (a) e de 50 classes de discretização (b) – no espaço normal.

As características mostradas pela Figura 6.19 são compartilhadas por todos os processos de simulação testados. Esse comportamento indica que, utilizando o algoritmo modificado, os processos de simulação foram realizados com sucesso, independente da discretização da distribuição de probabilidade Gaussiana padrão (G) empregada.

Os processos de simulação, também, foram validados no espaço original. A Tabela 6.5 apresenta os valores médios das médias e dos desvios padrão das realizações, conforme os 10 processos de simulação. Nesses processos de simulação, ambos os parâmetros obtiveram valores próximos aos condicionantes. Entretanto, como já verificado no espaço normal (Tabela 6.4), em geral, as médias assumiram valores ligeiramente superiores à média condicionante. Ao contrário, os desvios padrão assumiram valores levemente inferiores. A Figura 6.20 ilustra as médias e os desvios padrão, em função das realizações, de dois processos de simulação: um com 100 e outro com 50 realizações. Essa figura retrata os resultados da Figura 6.19 no espaço original. Na Figura 6.20 (a), dentre as 100

realizações, 54 apresentaram médias superiores à média condicionante (275,31). Em relação ao desvio padrão, 38 % das realizações obtiveram valores superiores ao condicionante (249,72). Na Figura 6.20 (b), dentre as 50 realizações, 29 apresentaram médias superiores à média condicionante e 32 % obtiveram valores de desvio padrão superiores ao condicionante.

Tabela 6.5: Média das médias e dos desvios padrão para 10 processos de simulação, conforme o número de classes M – espaço original.

Número de classes (M)	Média	
	Média	Desvio padrão
10	274,9139	247,7289
20	275,5644	248,0152
30	276,3814	248,7383
40	276,1949	248,3474
50	276,4786	248,6770
60	276,0902	248,5538
70	276,2498	248,5282
80	275,7853	247,9378
90	276,0552	248,0414
100	276,3172	248,2249

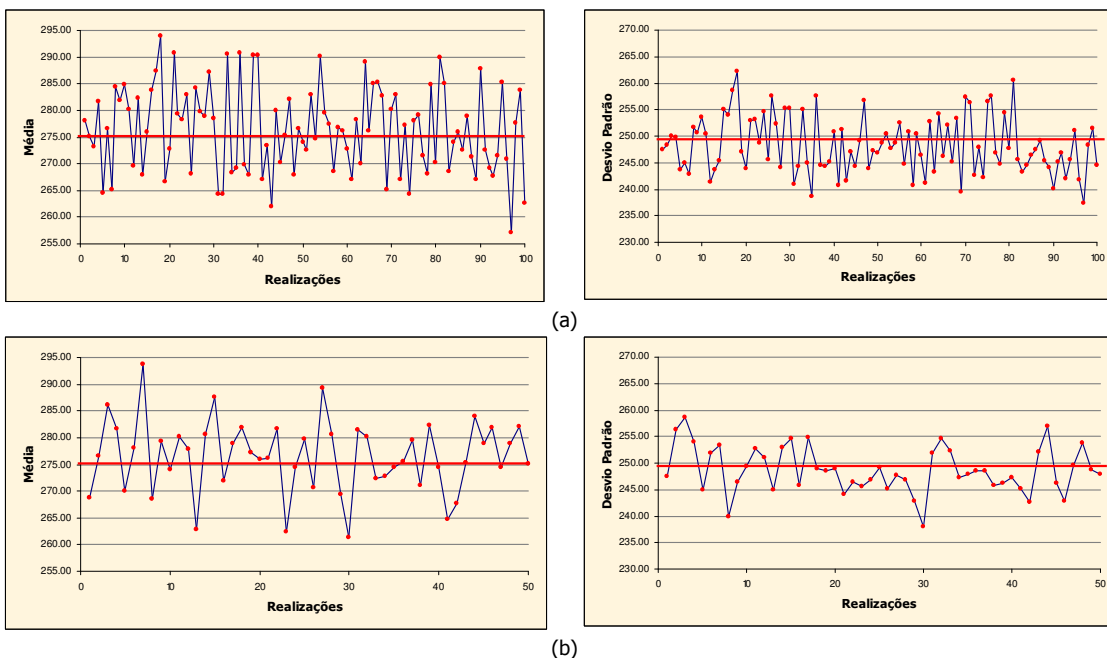


Figura 6.20: Médias (esquerda) e desvios padrão (direita) das realizações, resultantes dos processos de simulação a partir de 100 classes de discretização (a) e de 50 classes de discretização (b) – no espaço original.

Para avaliar o desempenho do algoritmo LHS clássico em relação ao algoritmo original (MC), foi realizada uma análise de variabilidade entre os vários cenários equiprováveis, gerados a partir dos 10 processos de simulação, no espaço original. Para cada processo de simulação (composto por números de realizações diferentes), foram calculados a média e o CV dos conjuntos constituídos pelas

médias das realizações. A Tabela 6.6 apresenta os resultados desses parâmetros, além do desvio padrão. A Figura 6.21, ilustra dois gráficos mostrando a relação entre a média e o CV em função dos processos de simulação.

Tabela 6.6: Média, desvio padrão e CV dos conjuntos de médias das realizações, para 10 processos de simulação.

Número de classes (M)	Média	Desvio padrão	CV (%)
10	274,9140	7,1692	2,6078
20	275,5644	7,1985	2,6123
30	276,3813	8,4646	3,0627
40	276,1699	8,0075	2,8995
50	276,4786	6,6922	2,4205
60	276,0902	8,3808	3,0355
70	276,2498	7,4922	2,7121
80	275,7852	8,8180	3,1974
90	276,0553	7,8998	2,8617
100	276,3173	8,0428	2,9107

Na Figura 6.21, pode ser observado que esses valores são próximos àqueles considerados estáveis, determinados a partir das 200 realizações executadas por meio do algoritmo original e representados pela linha vermelha. Nessa figura, o comportamento não tão contínuo, principalmente em relação ao CV das médias, é explicado pela característica descontínua do cálculo desses parâmetros. Diferente do caso original, cada ponto dos gráficos são oriundos de processos de simulação diferentes¹.

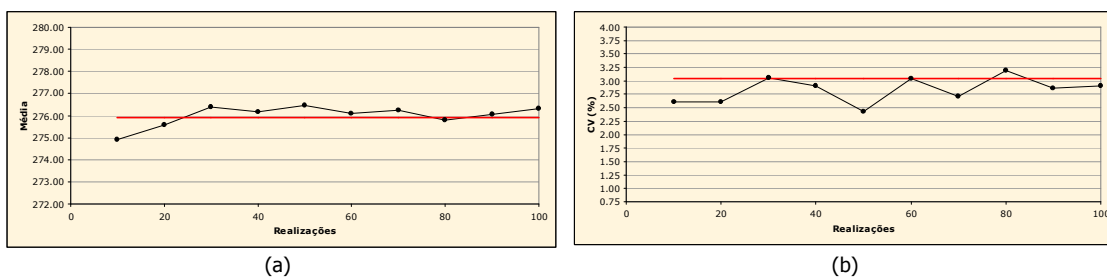


Figura 6.21: Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS clássico; a linha vermelha representa os valores considerados estáveis.

Em seguida, foi verificada a eficiência do algoritmo para um número menor de realizações. Para isso, o algoritmo foi testado utilizando números de classes M entre quatro e 20. Portanto, foram realizados 15 processos de simulação adicionais, resultantes da discretização de 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18 e 19 classes. A Tabela 6.7 apresenta os valores médios das médias e dos desvios padrão das realizações, conforme esses 15 processos de simulação, nos espaços normal e original.

¹ No caso original, foi executado um processo de simulação de 200 realizações. Nesse caso, foram executados 10 processos de simulação de 10, 20, 30, 40, 50, 60, 70, 80, 90 e 100 realizações (totalizando 550 realizações).

Tabela 6.7: Média das médias e dos desvios padrão para 15 processos de simulação, conforme o número de classes M, no espaço normal e no original. Valores condicionantes entre parênteses.

Número de classes (M)	Média			
	Espaço normal		Espaço original	
	Média (0,00)	Desvio padrão (1,00)	Média (275,31)	Desvio padrão (249,72)
4	0,0091	1,0009	277,0664	249,4481
5	0,0123	0,9996	277,7351	249,7080
6	0,0061	0,9991	276,3269	249,0333
7	0,0091	0,9992	277,1455	248,8780
8	0,0031	0,9984	275,7751	248,2563
9	0,0078	1,0013	277,0367	249,5217
11	0,0070	0,9933	276,1785	247,7190
12	0,0088	0,9981	277,0920	248,7812
13	0,0087	0,9977	276,9378	248,7056
14	0,0054	0,9978	276,0066	248,4089
15	0,0044	1,0032	276,2787	249,4540
16	0,0043	0,9987	276,1263	248,5183
17	0,0071	0,9949	276,3579	248,0974
18	0,0040	0,9944	275,5571	247,8482
19	0,0078	0,9955	276,6521	248,0225

Embora os resultados apresentados nessa tabela configurem as mesmas características observadas nas Tabelas 6.4 e 6.5, os conjuntos de realizações mostram flutuações ergódicas similares aos casos anteriores. As Figuras 6.22 e 6.23, ilustram essas flutuações, para o caso em que $M = 15$, nos espaços normal e original, respectivamente. Nessas figuras, pode ser observado que as realizações oscilam simetricamente em torno dos valores condicionantes. Na Figura 6.22, em relação à média, dentre as 15 realizações, nove apresentaram médias superiores à média condicionante. Em relação ao desvio padrão, sete realizações apresentaram desvios padrão superiores ao valor condicionante. No espaço original (Figura 6.23), 53 % das realizações mostraram médias superiores à média condicionante, enquanto que 40 % delas apresentaram desvios padrão superiores ao condicionante.

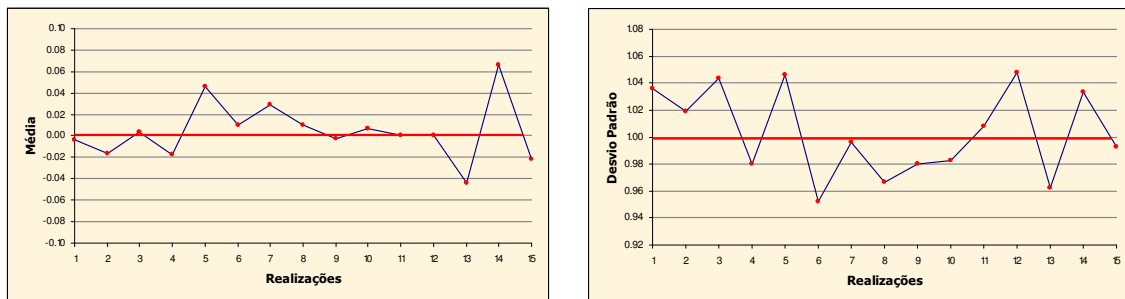


Figura 6.22: Médias (esquerda) e desvios padrão (direita) das realizações, resultantes dos processos de simulação a partir de 15 classes de discretização, no espaço normal.

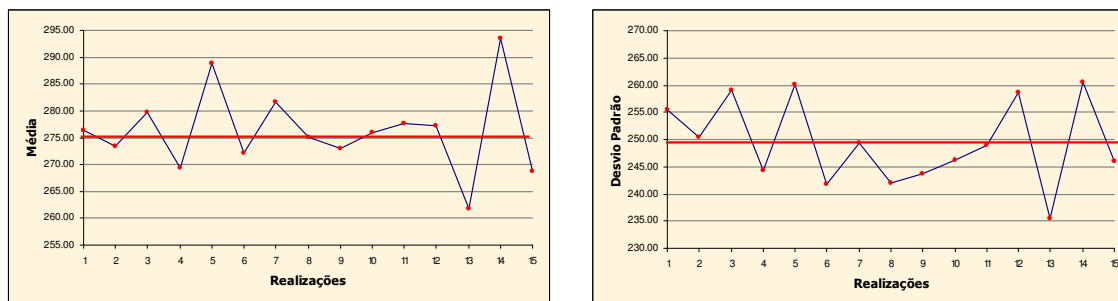


Figura 6.23: Médias (esquerda) e desvios padrão (direita) das realizações, resultantes dos processos de simulação a partir de 15 classes de discretização, no espaço original.

Na seqüência, como anteriormente, foi analisada a variabilidade em torno da média das realizações, segundo os 15 processos de simulação adicionais. A Tabela 6.8 apresenta os resultados de média, desvio padrão e CV, para 17 processos de simulação (incluindo os processos para os quais $M = 10$ e $M = 20$). A Figura 6.24 apresenta dois gráficos que ilustram a relação da média e do CV da distribuição das médias, em função dos processos de simulação.

Tabela 6.8: Média, desvio padrão e CV dos conjuntos de médias das realizações, para 17 processos de simulação.

Número de classes (M)	Média	Desvio padrão	CV (%)
4	277,0664	4,5970	1,6592
5	277,7351	2,5124	0,9046
6	276,3269	6,9816	2,5266
7	277,1455	9,6348	3,4764
8	275,7751	10,4219	3,7791
9	277,0367	7,8233	2,8239
10	274,9140	7,1692	2,6078
11	276,1785	6,5540	2,3731
12	277,0920	8,2818	2,9888
13	276,9378	6,1169	2,2088
14	276,0066	8,3748	3,0343
15	276,2787	7,5264	2,7242
16	276,1263	7,2817	2,6371
17	276,3578	7,3219	2,6494
18	275,5572	7,3665	2,6733
19	276,6521	9,9088	3,5817
20	275,5644	7,1985	2,6123

Na Figura 6.24, em geral, os parâmetros avaliados mostraram valores próximos aos considerados estáveis, representados pela linha vermelha. Os resultados são mais satisfatórios, principalmente, quando a distribuição G é dividida em mais do que 13 estratos. A partir desse grau de estratificação, em relação à média, os resultados são muito satisfatórios, oscilando levemente em torno do valor considerado estável (275,92 ppm). O CV da distribuição das médias se mostra um pouco mais errático, mas tolerável para um número reduzido de realizações. É importante enfatizar

que os gráficos da Figura 6.24 não podem ser avaliados dentro de um contexto de variáveis contínuas, pois cada ponto dos gráficos foi calculado a partir de processos de simulação diferentes.

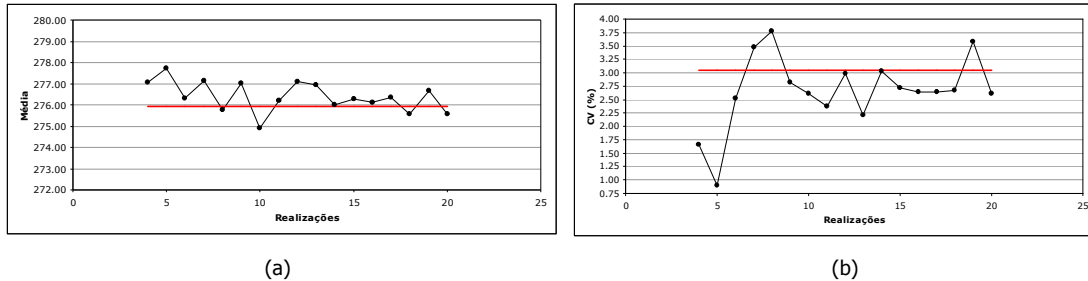


Figura 6.24: Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS clássico. A linha vermelha representa os valores considerados estáveis.

6.4.2.1. Validação prática do algoritmo LHS clássico

No Capítulo 5, na seção 5.5.3, foram apresentadas verificações teóricas do algoritmo proposto. Nessa seção, será mostrado como o algoritmo funciona na prática. Por exemplo, a Tabela 6.9 apresenta a solução e os parâmetros necessários à resolução da Equação 5.20, na situação em que $M = 15$, para a localização u_1 ($j = 1$). Essa equação fornece os valores simulados, de acordo com a técnica LHS. Na Tabela 6.9, na segunda e na terceira colunas, são apresentadas as classes de amostragem e as respectivas probabilidades de amostragem, de acordo com a realização. As três seguintes colunas apresentam o valor Gaussiano $x_{p_{lhs}}$ e os parâmetros calculados por meio de KS (a própria estimativa e o desvio padrão de krigagem). Finalmente, a última coluna mostra os valores simulados, no espaço normal, segundo as realizações.

Tabela 6.9: Parâmetros de simulação para a localização $j = 1$ da malha ($L = M = 15$).

Realização (i)	Classe $[\pi_i(u_1)]$	p_{lhs}	$x_{p_{lhs}}^{(i)}(u_1)$	$\gamma_{KS}^*(u_1)$	$\sigma_{KS}(u_1)$	$\gamma^{(i)}(u_1)$
1	12	0,7522	0,6813	-0,9573	0,6314	-0,5271
2	11	0,6937	0,5062	-0,5159	0,6295	-0,1973
3	9	0,5624	0,1570	-0,6228	0,6405	-0,5223
4	5	0,3282	-0,4450	-1,8071	0,7521	-2,1418
5	1	0,0400	-1,7508	-1,0839	0,6215	-2,1720
6	13	0,8077	0,8696	-1,3119	0,6620	-0,7363
7	15	0,9637	1,7948	-0,9065	0,6203	0,2067
8	6	0,3942	-0,2683	-0,6482	0,7494	-0,8493
9	3	0,1819	-0,9082	-1,2737	0,6829	-1,8940
10	10	0,6413	0,3619	-0,6512	0,7648	-0,3744
11	12	0,1236	-1,1573	-0,8143	0,7612	-1,6952
12	4	0,2389	-0,7100	-1,4791	0,6225	-1,9210
13	7	0,4224	-0,1957	-0,4011	0,6711	-0,5324
14	8	0,4686	-0,0787	-1,4097	0,6292	-1,4593
15	14	0,9047	1,3087	-0,5447	0,7192	0,3966

Os dados apresentados na Tabela 6.9, validam, na prática, o algoritmo. Em termos de probabilidade, quando $M = 15$, as classes têm amplitude de 6,67 %. Portanto, os resultados práticos confirmam que cada probabilidade plhs foi sacada de classes distintas, sem substituição.

No Capítulo 5, foi enfatizado que, em uma mesma realização, podem ser tiradas probabilidades plhs associadas às mesmas classes, desde que sejam de localizações diferentes (Capítulo 5, p. 118, primeiro parágrafo). A Figura 6.25 ilustra, na prática, essa situação. Nessa figura, um histograma mostra a freqüência de tiragens de classes, ao longo de uma realização. Nesse caso, foram utilizadas 100 classes ($M = 100$). A tiragem de classes é realizada de maneira aleatória, mesmo assim, nessa realização, 16 % das localizações tiveram probabilidades plhs associadas à classe número 73. Em contrapartida, as classes menos solicitadas, nessa realização, foram as de número 8, 35, 78 e 89, com 3 % de freqüência. Para as classes restantes, a média de tiragem foi de cerca de 8 %.

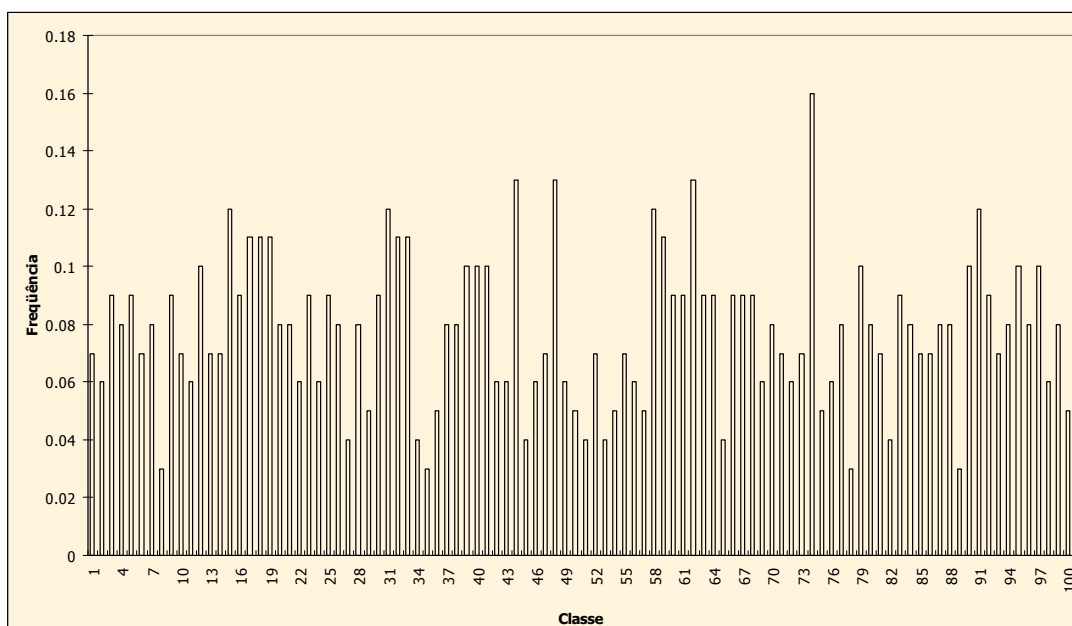


Figura 6.25: Histograma de freqüência de tiragens de classes ao longo de uma realização.

Quanto ao procedimento prático de tiragens de classes, a Tabela 6.10 apresenta, principalmente, o número aleatório que determina a classe amostrada. De acordo com o algoritmo, descrito na seção 5.5.2.4 (p. 113), em cada localização, a classe amostrada é determinada por meio de um número aleatório (*random*, representado por $\pi_i(u_j)$) nas Equações 5.18 e 5.19). Esse número aleatório é multiplicado por 100 e adicionado uma unidade. Finalmente, a classe amostrada é determinada pela parte inteira dessa operação. A Tabela 6.10 apresenta esse número aleatório e a classe resultante, além do caminho aleatório seguido em uma determinada realização. Nessa tabela, são apresentadas as cinco localizações extremas superiores e as cinco extremas inferiores, de acordo com o caminho aleatório. A primeira coluna indica, simplesmente, o número seqüencial de simulação.

Tabela 6.10: Procedimento prático de tiragens de classes, segundo as 10 localizações extremas do caminho aleatório de uma realização.

i	j (caminho aleatório)	$\pi_i(u_j)$ (número aleatório)	Classe
1	62164	0,0000	1
2	49828	0,0850	9
3	6028	0,6014	61
4	37204	0,8916	90
5	43468	0,9680	97
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮
77996	16429	0,2060	21
77997	42858	0,7599	76
77998	55542	0,5274	53
77999	47100	0,2176	22
78000	4733	0,8452	85

6.4.3. Algoritmo LHS cíclico clássico

Para avaliar o desempenho do algoritmo LHS cíclico clássico, foram testados diversos números de discretização M . Para cada grau de discretização da distribuição G , foram realizados processos de simulação diferentes, com números de ciclos diferentes. De acordo com o número M , foram executados tantos ciclos quanto necessários para alcançar, no máximo, 100 realizações. Por exemplo, quando a distribuição G foi discretizada em quatro classes ($M = 4$), foram realizados 25 ciclos, com cada ciclo sendo constituído por quatro realizações. No total, foram avaliados 17 modelos de discretização da distribuição G . A Tabela 6.11 apresenta esses 17 modelos de discretização, assim como o número de ciclos e o número total de realizações.

Tabela 6.11: Número de ciclos e número total de realizações de acordo a discretização da distribuição G .

Número de classes (M)	Número de ciclos	Número total de realizações
4	25	100
5	20	100
6	16	96
7	14	98
8	12	96
9	11	99
10	10	100
11	9	99
12	8	96
13	7	91
14	7	98
15	6	90
16	6	96
17	5	85
18	5	90
19	5	95
20	5	100

Nas seções seguintes, serão apresentados os resultados obtidos segundo cada modelo de discretização da distribuição de probabilidade Gaussiana padrão G.

6.4.3.1. M = 4, 25 ciclos

Primeiramente, a distribuição de probabilidade Gaussiana padrão G foi discretizada em quatro classes, o que permitiu a execução de 25 ciclos de quatro realizações. Os gráficos apresentados na Figura 6.26 validam o processo de simulação. Esses gráficos mostram a média e o desvio das 100 realizações executadas, nos espaços normal e original. A linha vermelha, representando os valores condicionantes, ajuda a verificar a magnitude das flutuações ergódicas. No espaço normal, dentre as 100 realizações, 61 apresentaram média superior à zero. Em relação ao desvio padrão, 40 realizações apresentaram valor superior ao condicionante. No espaço original, 57 % das realizações mostraram média superior à média condicionante, e 40 % delas obtiveram o desvio padrão maior do que o condicionante.

Nos gráficos da Figura 6.26, as linhas verticais separam os 25 ciclos. É possível verificar que, mesmo dentro dos ciclos, as realizações flutuaram de maneira simétrica em torno dos valores condicionantes.

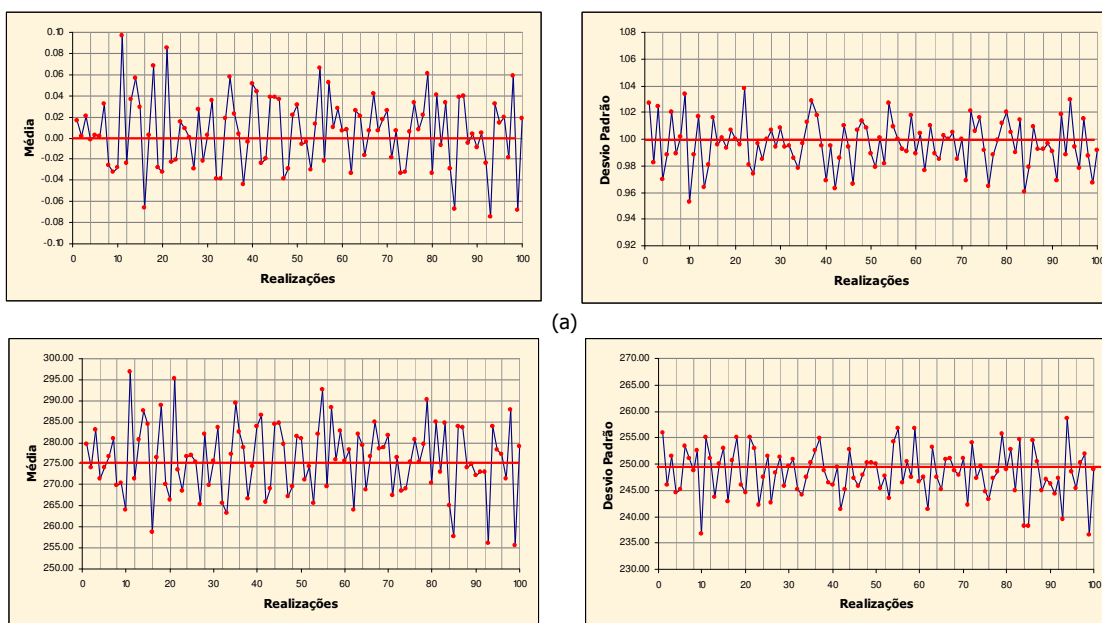


Figura 6.26: Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de 25 ciclos de quatro realizações, no espaço normal (a) e espaço original (b).

Como nos casos anteriores, foi avaliada a variabilidade entre as médias das realizações. O passo utilizado foi o de um ciclo (quatro realizações). Isto é, ao final de cada ciclo foi calculado a média e o CV da distribuição das médias de todas as realizações executadas previamente. A Tabela

6.12 apresenta os valores de média e CV, além do desvio padrão, de acordo com os conjuntos de realizações.

Tabela 6.12: Média, desvio padrão e CV dos conjuntos de médias das realizações, para 25 ciclos de quatro realizações.

N. de realizações	Média	Desvio padrão	CV (%)
4	277,0664	21,1323	1,6592
8	276,2242	19,5570	1,6010
12	276,0252	65,4007	2,9298
16	276,4845	81,5690	3,2666
20	276,2801	80,0711	3,2388
24	276,6519	84,1556	3,3159
28	276,4047	77,7628	3,1904
32	276,0600	74,6608	3,1300
36	276,2936	77,0430	3,1768
40	276,2596	73,4487	3,1022
44	276,2762	74,3109	3,1202
48	276,1962	72,4960	3,0828
52	276,2564	68,4794	2,9955
56	276,3417	71,7553	3,0654
60	276,6308	69,9874	3,0242
64	276,5872	68,6953	2,9966
68	276,6226	66,6146	2,9505
72	276,5929	64,5081	2,9038
76	276,4254	62,8955	2,8690
80	276,5466	62,6940	2,8632
84	276,5641	63,0780	2,8717
88	276,4839	65,4576	2,9262
92	276,3441	63,0781	2,8740
96	276,2423	65,3469	2,9263
100	276,1319	68,7003	3,0017

A Figura 6.27 ilustra dois gráficos que relacionam a média e o CV das médias das realizações, de acordo com os conjuntos de realizações. Ambos os parâmetros medidos mostraram valores próximos àqueles considerados estáveis, indicados pela linha vermelha nos gráficos da Figura 6.27.

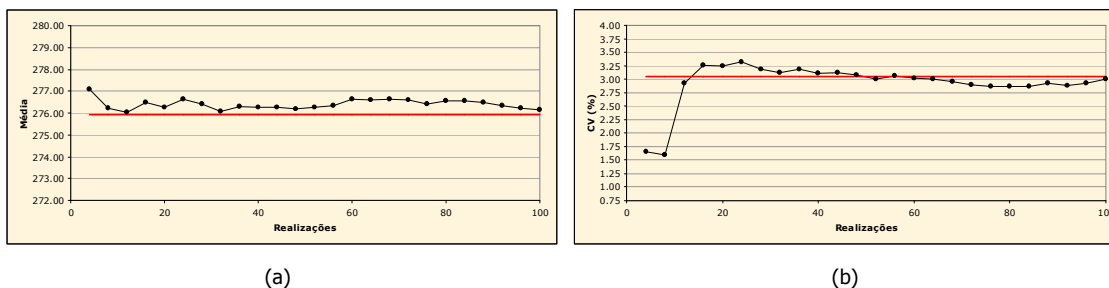


Figura 6.27: Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (25 ciclos de quatro realizações). A linha vermelha representa os valores considerados estáveis.

6.4.3.2. M = 5, 20 ciclos

A partir da discretização da distribuição padrão G em cinco classes, foram executadas 100 realizações, por meio de 20 ciclos. O processo de simulação foi validado nos espaços normal e original, como mostra a Figura 6.28. A intensidade de variação das médias e dos desvios padrão das realizações é como a recomendada (de fraca magnitude). Nos gráficos dessa figura, as linhas verticais indicam os limites de ciclo.

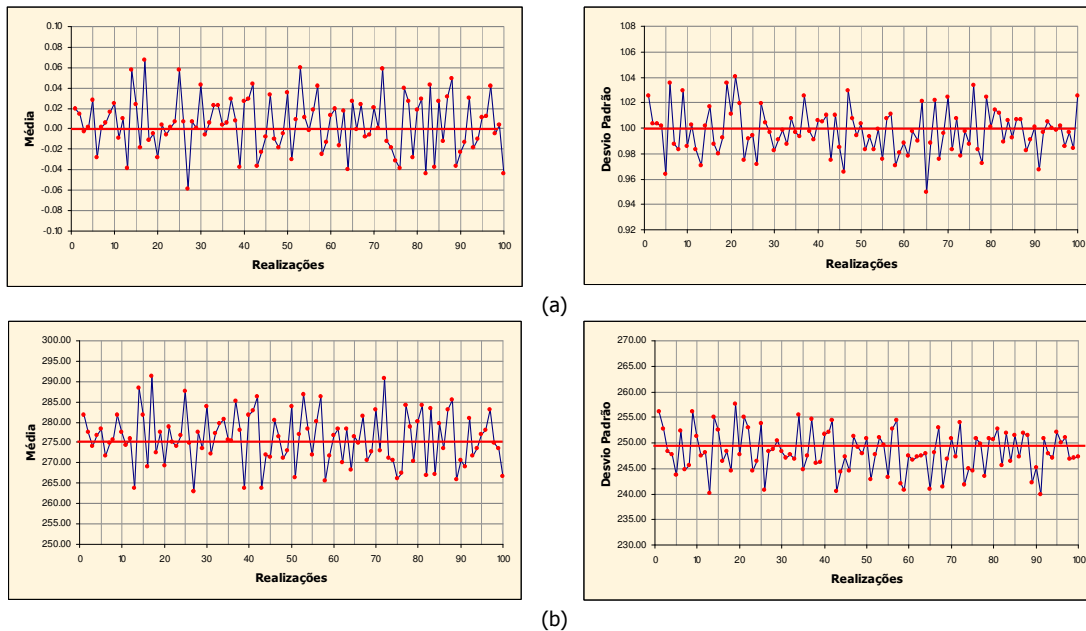


Figura 6.28: Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de 20 ciclos de cinco realizações, no espaço normal (a) e espaço original (b).

A Figura 6.29 ilustra os gráficos que mostram a relação da média e dos desvios padrão das médias das realizações em função dos conjuntos de realizações. Para esse grau de discretização, em relação aos valores considerados estáveis (indicados, nos gráficos, pela linha vermelha), o desempenho da média foi superior ao do CV. Nesse caso, o CV das médias das realizações foi sempre inferior ao estável, indicando que o espaço de incerteza foi subestimado. A Tabela 6.13 apresenta esses parâmetros, além do desvio padrão.

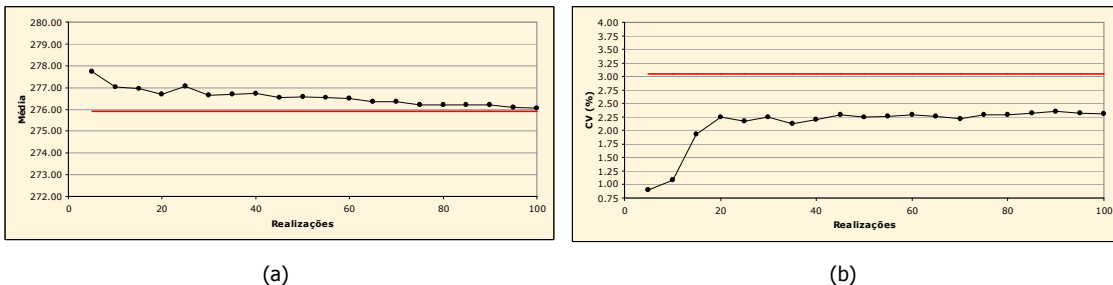


Figura 6.29: Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (20 ciclos de cinco realizações). A linha vermelha representa os valores considerados estáveis.

Tabela 6.13: Média, desvio padrão e CV dos conjuntos de médias das realizações, para 20 ciclos de cinco realizações.

N. de realizações	Média	Desvio padrão	CV (%)
5	277,7351	6,3124	0,9046
10	277,0259	9,0547	1,0862
15	276,9481	28,3988	1,9242
20	276,6984	38,5711	2,2445
25	277,0641	36,0660	2,1675
30	276,6383	38,6899	2,2485
35	276,7002	34,4546	2,1214
40	276,7196	36,8958	2,1951
45	276,5528	40,4058	2,2985
50	276,5912	38,5913	2,2460
55	276,5446	39,3014	2,2669
60	276,5113	40,1823	2,2925
65	276,3411	38,8488	2,2555
70	276,3563	37,7247	2,2225
75	276,2202	40,3045	2,2984
80	276,2169	40,2511	2,2969
85	276,2193	41,2825	2,3261
90	276,1933	42,1004	2,3493
95	276,1017	40,9604	2,3180
100	276,0582	40,3648	2,3014

6.4.3.3. M = 6, 16 ciclos

Com a utilização de seis classes de discretização, foram executadas 96 realizações, divididas em 16 ciclos. A validação do processo de simulação é resumida na Figura 6.30. Essa figura mostra o comportamento da média e do desvio padrão das realizações, no espaço Gaussiano e no espaço original.

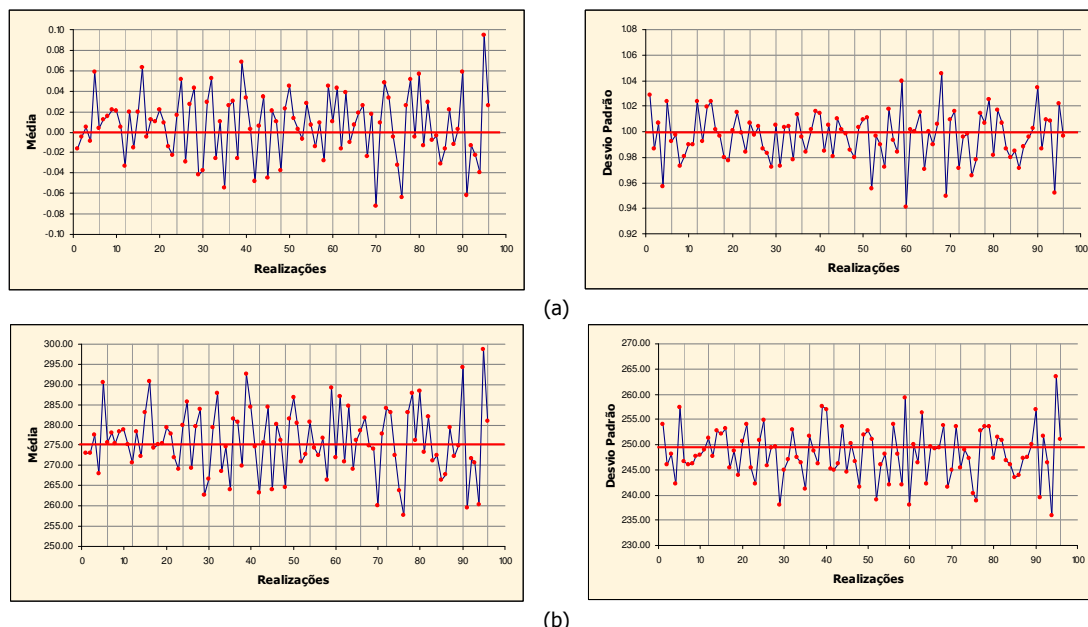
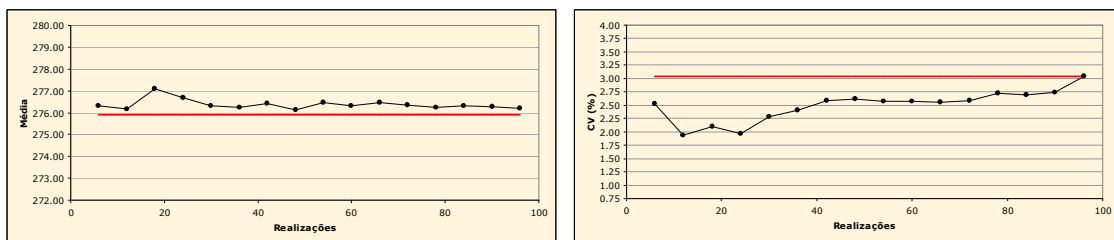


Figura 6.30: Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de 16 ciclos de seis realizações, no espaço normal (a) e espaço original (b).

A Tabela 6.14 apresenta a média, o desvio padrão e o CV das distribuições das médias das realizações, obtidos ao longo dos 16 ciclos. A Figura 6.31 ilustra os gráficos que relacionam os parâmetros de média e CV (dados na tabela) em função dos conjuntos de realizações. Os resultados observados nesses gráficos são favoráveis em relação à média. Em relação ao CV, novamente, é verificado que a variabilidade da distribuição das médias é subestimada. Isso indica que, para esse grau de discretização, o espaço de incerteza da VA V foi caracterizado inadequadamente.

Tabela 6.14: Média, desvio padrão e CV dos conjuntos de médias das realizações, para 16 ciclos de seis realizações.

N. de realizações	Média	Desvio padrão	CV (%)
6	276,3269	48,7422	2,5266
12	276,1762	28,4819	1,9324
18	277,0938	33,9890	2,1040
24	276,7035	29,7740	1,9720
30	276,3086	39,9950	2,2888
36	276,2498	44,0303	2,4020
42	276,4300	51,0881	2,5857
48	276,1483	52,4696	2,6231
54	276,4470	50,6257	2,5738
60	276,3257	50,5906	2,5740
66	276,4537	50,1012	2,5604
72	276,3681	50,9583	2,5830
78	276,2366	56,4905	2,7209
84	276,3123	55,2081	2,6891
90	276,2783	57,2461	2,7386
96	276,2200	70,9431	3,0493



(a)

(b)

Figura 6.31: Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (16 ciclos de seis realizações). A linha vermelha representa os valores considerados estáveis.

6.4.3.4. M = 7, 14 ciclos

Para o caso no qual a distribuição de probabilidade Gaussiana padrão G foi dividida em sete classes, foram executadas 98 realizações, constituídas por 14 ciclos de sete realizações. O processo de simulação foi realizado com sucesso, como pode ser observado pelo comportamento da média e do desvio padrão das realizações, nos espaços normal e original (Figura 6.32). Entretanto, com relação à média, a amplitude das flutuações ergódicas, nos espaços normal e original, foi maior em relação aos casos anteriores. A média da realização número 66 extrapolou os valores de 0,10 e 300, nos espaços

normal e original, respectivamente. Para contrabalançar, a média da realização número 96 foi inferior a -0,10 e 255, nos espaços normal e original, respectivamente. Nos gráficos da Figura 6.32, as linhas verticais limitam os ciclos, de sete realizações. A linha horizontal vermelha indica os valores condicionantes.

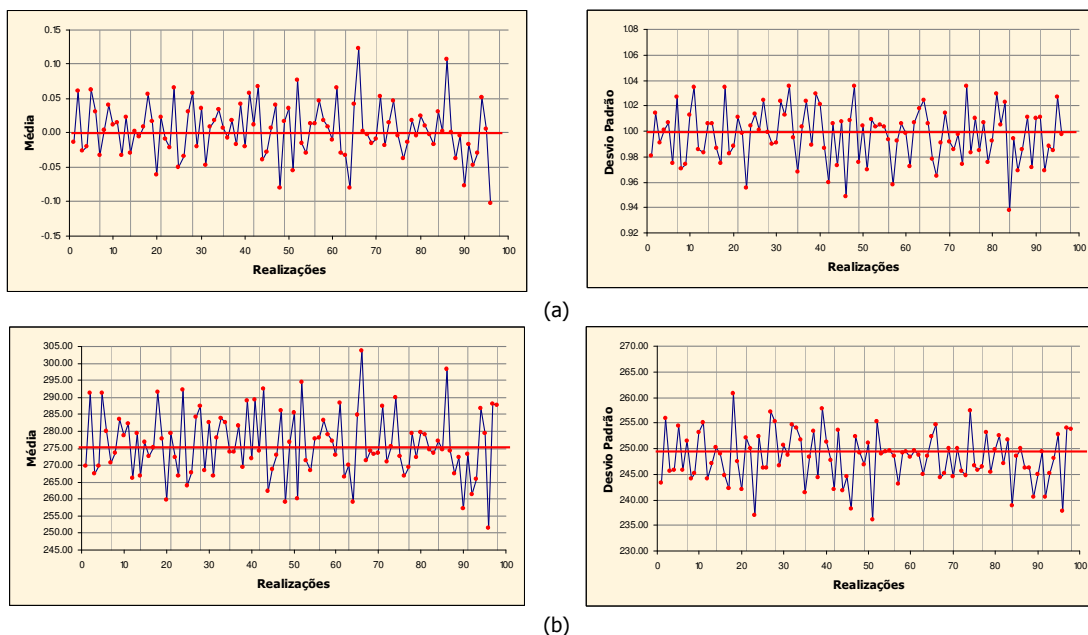


Figura 6.32: Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de 14 ciclos de sete realizações, no espaço normal (a) e espaço original (b).

A Tabela 6.15 apresenta os parâmetros de média, desvio padrão e CV das médias dos conjuntos de realizações, utilizados para avaliar a variabilidade em torno das médias das realizações. A Figura 6.33 ilustra, graficamente, o comportamento dos parâmetros de média e CV para os ciclos cumulativos de realizações.

Tabela 6.15: Média, desvio padrão e CV dos conjuntos de médias das realizações, para 14 ciclos de sete realizações.

N. de realizações	Média	Desvio padrão	CV (%)
7	277,1455	93,5134	3,4892
14	276,4528	68,5952	2,9959
21	276,3618	71,0287	3,0496
28	276,3632	80,9269	3,2551
35	276,4134	73,3285	3,0980
42	276,7559	71,3088	3,0512
49	276,3681	79,7200	3,2307
56	276,3893	83,4804	3,3058
63	276,4314	79,7236	3,2300
70	276,5070	88,5648	3,4035
77	276,4677	86,8404	3,3707
84	276,4688	80,2848	3,2409
91	276,2676	84,4679	3,3267
98	276,1304	91,8788	3,4713

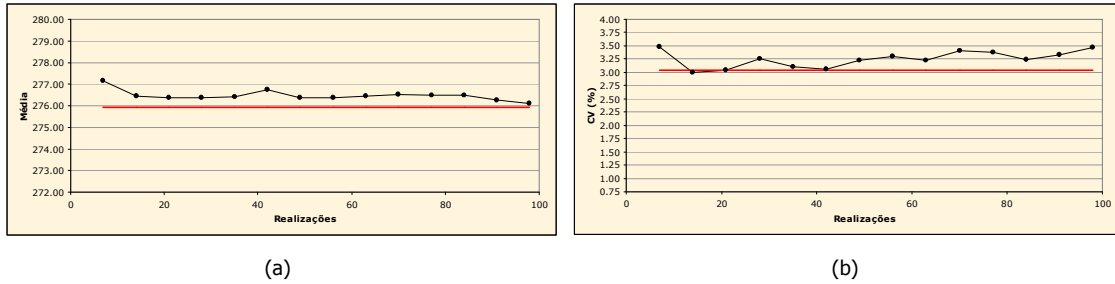


Figura 6.33: Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (14 ciclos de sete realizações). A linha vermelha representa os valores considerados estáveis.

Os parâmetros medidos e representados na forma de gráficos (Figura 6.33), mostram pouca variação em relação àqueles considerados estáveis (representados pela linha vermelha). Entretanto, esses valores mostraram-se levemente superiores, indicando uma pequena superestimativa do espaço de incerteza da VA V.

6.4.3.5. M = 8, 12 ciclos

Para oito classes de discretização, 12 ciclos foram necessários para executar 96 realizações. As médias e os desvios padrão das realizações validam o processo de simulação (Figura 6.34). Tanto no espaço Gaussiano, quanto no espaço original, esses parâmetros oscilaram em torno dos valores condicionantes (Figura 6.34 – linhas horizontais vermelhas).

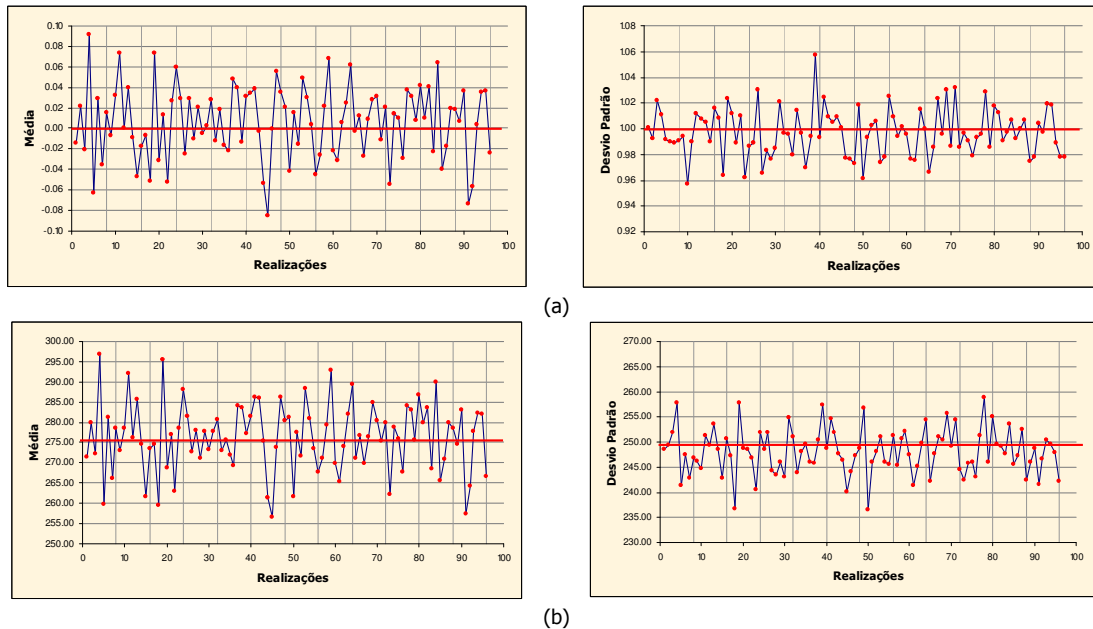


Figura 6.34: Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de 12 ciclos de oito realizações, no espaço normal (a) e espaço original (b).

Com relação à variabilidade apresentada pelas médias das realizações, a Tabela 6.16 lista as médias e os CV, além do desvio padrão, dos conjuntos de médias. Essa tabela é representada graficamente na Figura 6.35. Essa figura mostra que, a partir do 5º ciclo, ambos os parâmetros atingem valores muito próximos àqueles considerados estáveis, obtidos por meio do algoritmo original, a partir de 200 realizações.

Tabela 6.16: Média, desvio padrão e CV dos conjuntos de médias das realizações, para 12 ciclos de oito realizações.

N. de realizações	Média	Desvio padrão	CV (%)
8	275,7751	108,6164	3,7791
16	276,3524	90,7343	3,4469
24	276,1112	103,2456	3,6800
32	276,2353	80,7139	3,2523
40	276,4071	70,2154	3,0316
48	276,3023	77,8982	3,1943
56	276,1619	75,8555	3,1538
64	276,3910	77,0899	3,1767
72	276,4462	70,9055	3,0460
80	276,4789	69,9544	3,0251
88	276,5380	68,9097	3,0018
96	276,2868	70,5849	3,0409

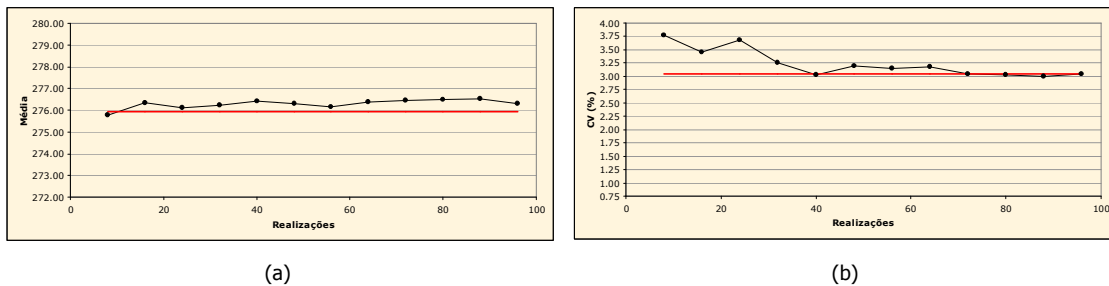


Figura 6.35: Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (12 ciclos de oito realizações). A linha vermelha representa os valores considerados estáveis.

6.4.3.6. M = 9, 11 ciclos

Para uma discretização de nove classes, foi necessário executar 11 ciclos para se atingir 99 realizações. Os gráficos da Figura 6.36 validam o processo de simulação no espaço normal.

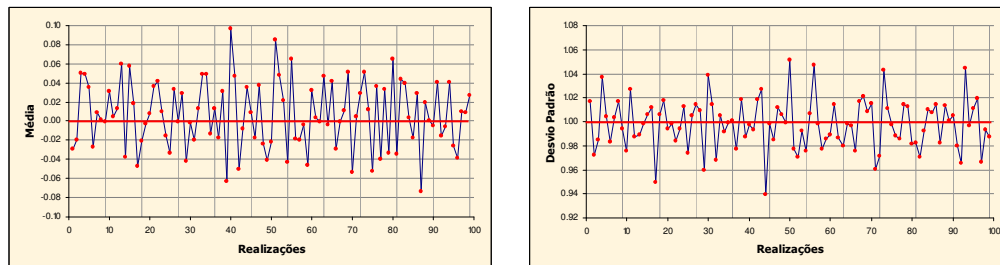


Figura 6.36: Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de 11 ciclos de nove realizações, no espaço normal.

O processo também foi validado no espaço original, como mostram os gráficos da Figura 6.37. Em ambos os espaços, a média e o desvio padrão das realizações flutuaram moderadamente em torno dos valores condicionantes, dados pela linha horizontal vermelha.

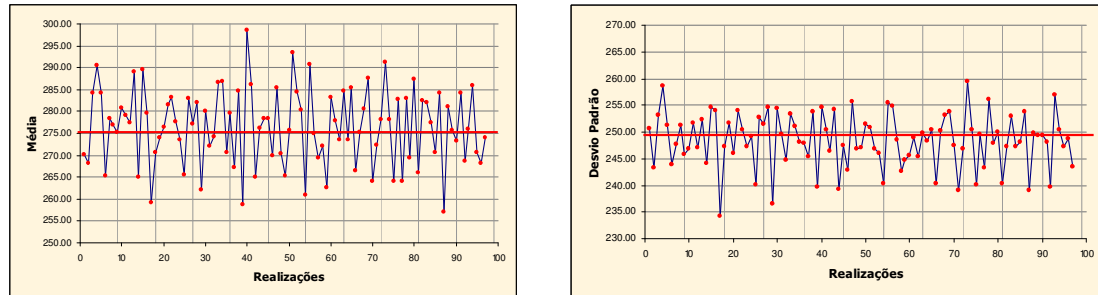
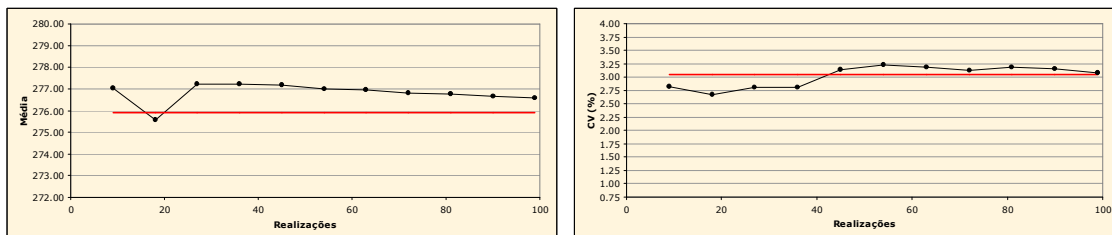


Figura 6.37: Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de 11 ciclos de nove realizações, no espaço original.

A Tabela 6.17 apresenta os parâmetros de medida de variabilidade entre as realizações, medidos a partir da distribuição cumulativa das médias. Para essa distribuição, a Tabela 6.17 descreve a média, o desvio padrão e o CV, de acordo com os conjuntos de realizações (ciclos cumulativos). Os gráficos da Figura 6.38 descrevem o comportamento da média e do CV. Esses gráficos mostram que, principalmente, a média foi superestimada ao longo do processo de simulação.

Tabela 6.17: Média, desvio padrão e CV dos conjuntos de médias das realizações, para 11 ciclos de nove realizações.

N. de realizações	Média	Desvio padrão	CV (%)
9	277,0367	61,2036	2,8239
18	275,5572	54,2647	2,6733
27	277,2323	60,7990	2,8126
36	277,2063	60,2590	2,8003
45	277,1689	75,4204	3,1333
54	276,9988	79,7590	3,2241
63	276,9448	77,7379	3,1836
72	276,8147	75,0128	3,1288
81	276,7514	77,6617	3,1843
90	276,6703	76,2793	3,1568
99	276,5962	72,6296	3,0811



(a)

(b)

Figura 6.38: Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (11 ciclos de nove realizações). A linha vermelha representa os valores considerados estáveis.

6.4.3.7. M = 10, 10 ciclos

Com o emprego de 10 classes de discretização, foram executadas 100 realizações, divididas em 10 ciclos. A Figura 6.39 apresenta o comportamento da média e do desvio padrão das realizações, nos espaços normal e original. É verificado que ambos os parâmetros assumiram valores similares aos condicionantes (linha horizontal vermelha), apresentando oscilações de baixa magnitude. Esse comportamento valida o processo de simulação. Nos gráficos da Figura 6.39, as linhas verticais ajudam a identificar os ciclos. É possível verificar que dentro dos ciclos, em média, os parâmetros condicionantes são reproduzidos.

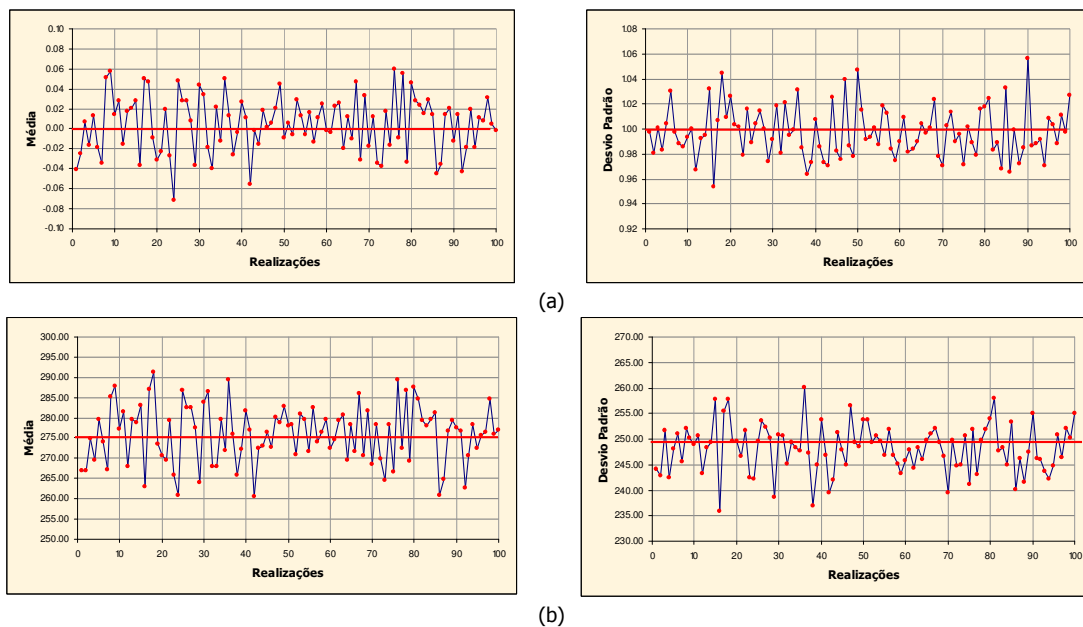


Figura 6.39: Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de 10 ciclos de 10 realizações, no espaço normal (a) e espaço original (b).

Os parâmetros de variabilidade entre as realizações são apresentados na Tabela 6.18. Nessa tabela, são descritos os parâmetros de média, desvio padrão e CV dos conjuntos de médias das realizações.

Tabela 6.18: Média, desvio padrão e CV dos conjuntos de médias das realizações, para 10 ciclos de 10 realizações.

N. de realizações	Média	Desvio padrão	CV (%)
10	274,9140	51,3975	2,6078
20	276,2800	62,3707	2,8585
30	275,9511	68,1304	2,9912
40	275,9439	65,9951	2,9440
50	275,7935	59,6876	2,8013
60	275,9383	52,4325	2,6241
70	275,9637	49,5066	2,5496
80	276,0036	52,7634	2,6318
90	276,0236	52,4676	2,6242
100	275,9257	50,2513	2,5691

Os gráficos da Figura 6.40 ajudam a analisar os parâmetros descritos na Tabela 6.18. Esses gráficos mostram a relação da média e do CV em função dos conjuntos de médias. Com relação à média, os resultados são muito satisfatórios. A partir do 2º ciclo, as médias assumem valores quase que idênticos ao valor considerado estável, representado pela linha horizontal vermelha. Em contrapartida, em relação ao CV, os resultados são subestimados. No 3º e 4º ciclo, esse parâmetro se mostra muito próximo ao considerado estável, porém, ainda, levemente subestimado.

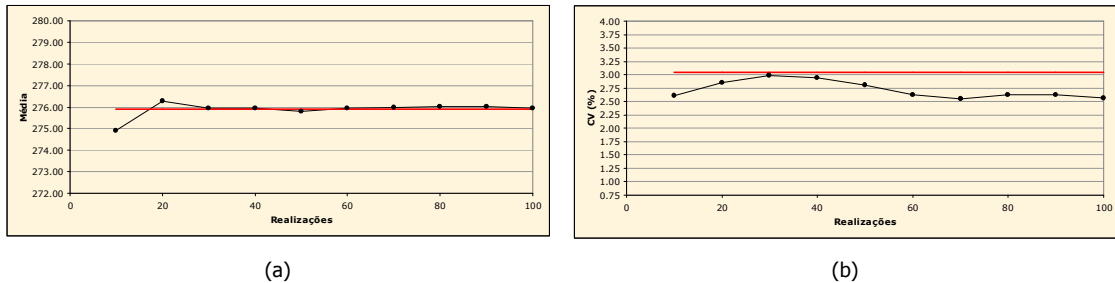


Figura 6.40: Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (10 ciclos de 10 realizações). A linha vermelha representa os valores considerados estáveis.

6.4.3.8. M = 11, 9 ciclos

Com a utilização de 11 classes de discretização, foram executadas 99 realizações, divididas em 9 ciclos. Os valores assumidos pelas médias e desvios padrão das realizações (Figura 6.41) validam o processo de simulação.

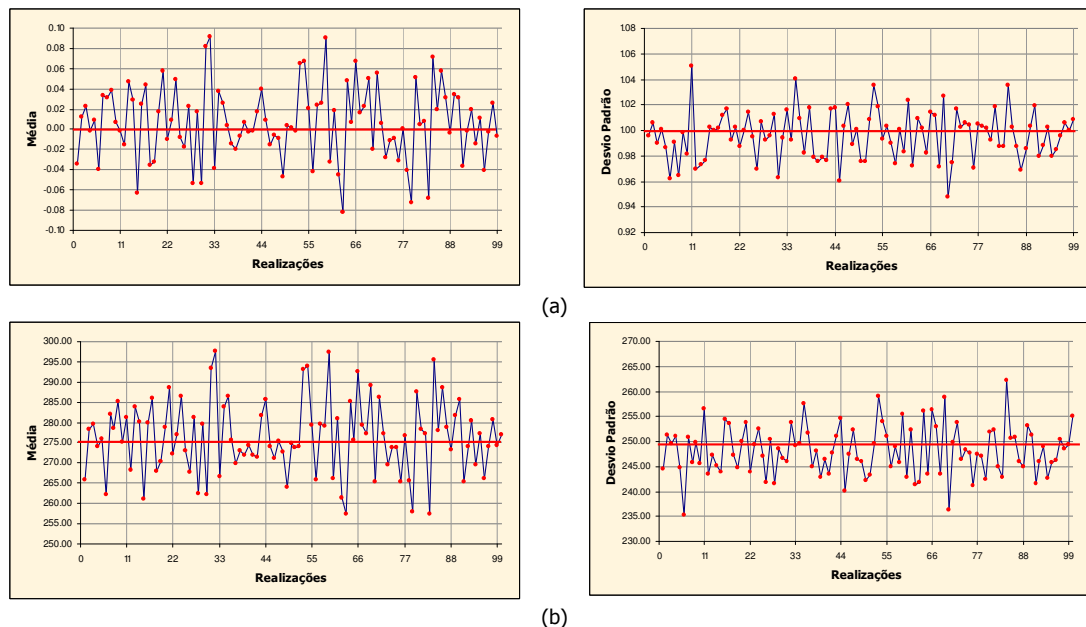


Figura 6.41: Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de nove ciclos de 11 realizações, no espaço normal (a) e espaço original (b).

Nos gráficos da Figura 6.41 é verificado que, em média, os parâmetros avaliados reproduzem os valores condicionantes, representados pelas linhas horizontais vermelhas. Nesse gráficos, as linhas verticais limitam os ciclos.

A Tabela 6.19 lista os parâmetros utilizados para avaliar a variabilidade entre as realizações de ciclos cumulativos (média, desvio padrão e CV). Esses parâmetros foram calculados a partir da distribuição das médias das 99 realizações executadas. A Figura 6.42 ilustra dois gráficos, que descrevem o comportamento da média e do CV dos conjuntos de médias das realizações. Uma análise sobre a Tabela 6.19 e a Figura 6.42, permite afirmar que, em geral, para esse modelo de discretização, o espaço de incerteza da VA V foi superestimado. Os valores de CV obtidos para o 4º e 5º ciclos são satisfatórios, entretanto, as médias (para os mesmos ciclos) são levemente superestimadas.

Tabela 6.19: Média, desvio padrão e CV dos conjuntos de médias das realizações, para nove ciclos de 11 realizações.

N. de realizações	Média	Desvio padrão	CV (%)
11	276,1785	42,9545	2,3731
22	276,1503	55,2864	2,6925
33	276,4514	80,8693	3,2529
44	276,5717	69,6455	3,0174
55	276,6439	70,5452	3,0361
66	276,6163	83,3347	3,3002
77	276,5009	78,9217	3,2129
88	276,4634	85,7857	3,3502
99	276,3424	80,7049	3,2509

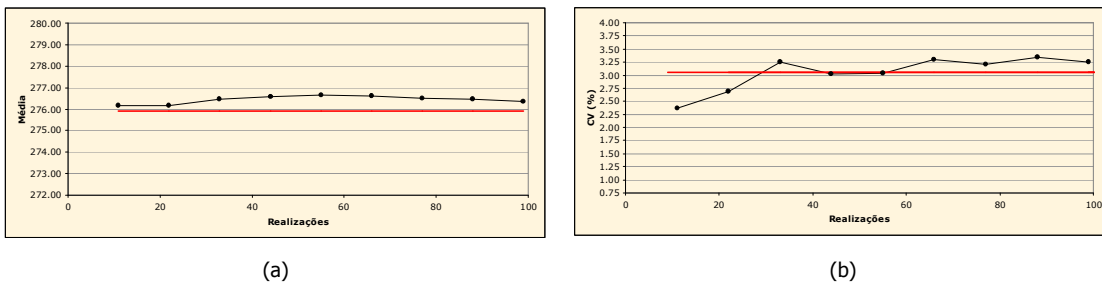


Figura 6.42: Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (nove ciclos de 11 realizações). A linha vermelha representa os valores considerados estáveis.

6.4.3.9. M = 12, 8 ciclos

Para um modelo de discretização de 12 classes, foram executadas 96 realizações, em oito ciclos. Esses ciclos, compostos por 12 realizações, foram validados nos espaços Gaussiano e original, como mostra a Figura 6.43. Essa figura exibe dois gráficos que descrevem as variações assumidas pelas médias e desvios padrão das realizações. Nesses gráficos, a linha horizontal vermelha indica os respectivos valores condicionantes, enquanto que as linhas verticais identificam os ciclos. As flutuações ergódicas de pouca intensidade e a reprodução, em média, dos valores condicionantes, validam o processo de simulação.

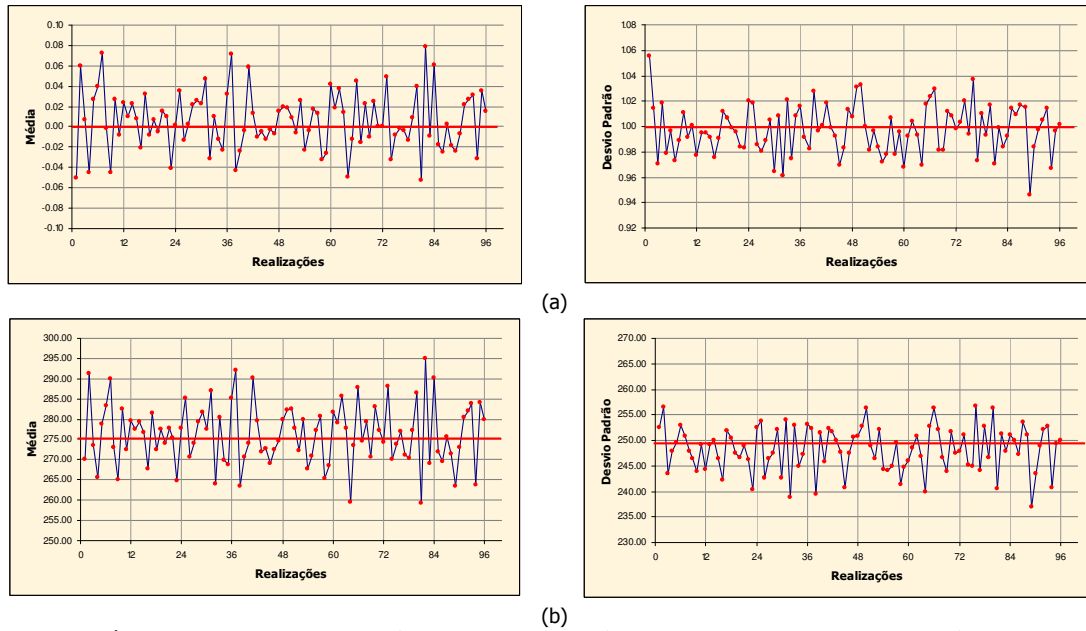


Figura 6.43: Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de oito ciclos de 12 realizações, no espaço normal (a) e espaço original (b).

A Tabela 6.20 lista os parâmetros de média, desvio padrão e CV dos conjuntos de médias das realizações. A Figura 6.44 ilustra, graficamente, o comportamento da média e do CV em função dos conjuntos de realizações.

Tabela 6.20: Média, desvio padrão e CV dos conjuntos de médias das realizações, para oito ciclos de 12 realizações.

N. de realizações	Média	Desvio padrão	CV (%)
12	277,0920	68,5877	2,9888
24	276,1451	45,7617	2,4497
36	276,4181	47,5712	2,4952
48	276,2842	51,7295	2,6032
60	276,1482	48,7152	2,5275
72	276,2761	49,1975	2,5388
84	276,4218	56,9332	2,7297
96	276,2364	56,1381	2,7124

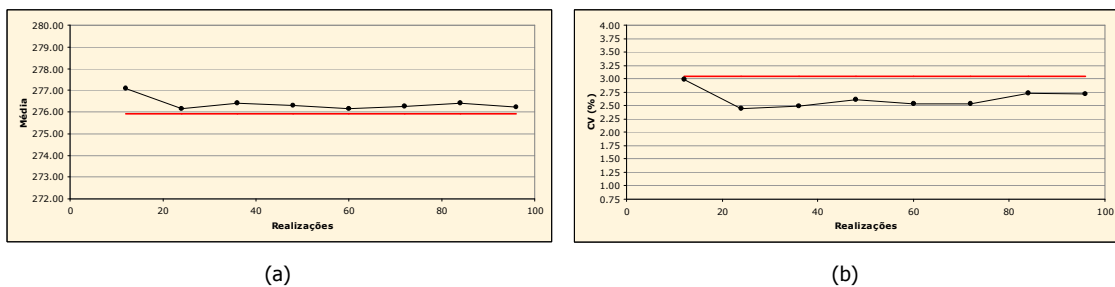


Figura 6.44: Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (oito ciclos de 12 realizações). A linha vermelha representa os valores considerados estáveis.

Na Figura 6.44, embora os valores de média sejam muito similares à média considerada estável (representada pela linha horizontal vermelha), os valores de CV são muito inferiores ao valor considerado estável. Esse comportamento indica que esse modelo de discretização não é adequado para caracterizar o espaço de incerteza da VA V, pois o mesmo o subestima.

6.4.3.10. M = 13, 7 ciclos

A discretização da distribuição de probabilidade Gaussiana G em 13 classes, resultou na execução de sete ciclos, totalizando 91 realizações. O processo de simulação foi validado nos espaços normal e original, por meio da análise da Figura 6.45. Os gráficos dessa figura indicam que, em média, os valores condicionantes (representados pela linha horizontal vermelha) foram reproduzidos. Essa afirmação é corroborada pela não tendenciosidade (simetria) das flutuações ergódicas, mostradas pelos parâmetros de média e desvio padrão das realizações.

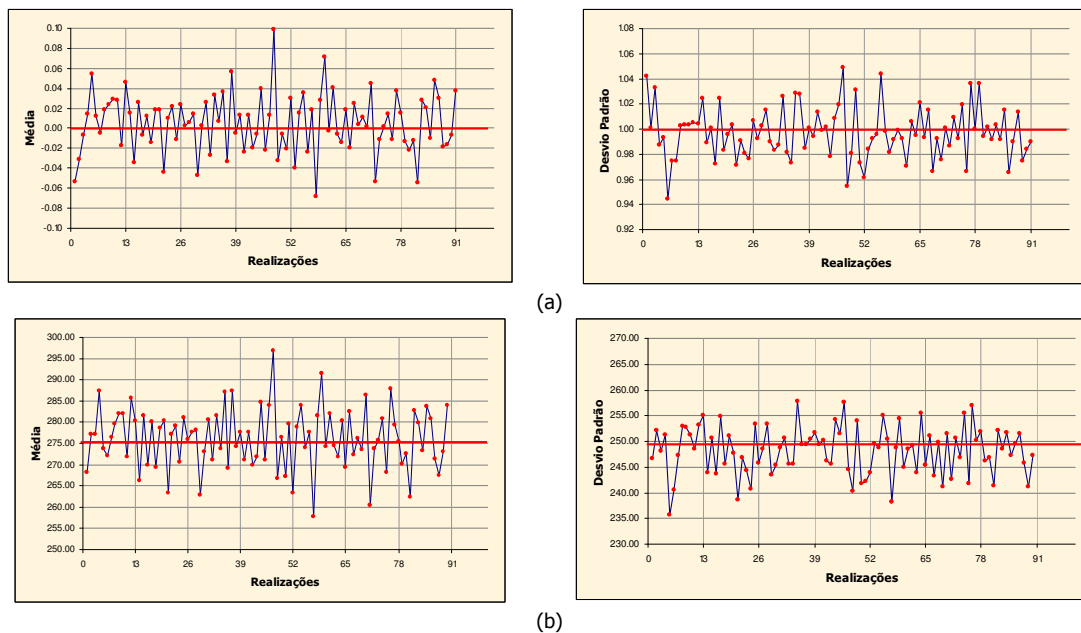


Figura 6.45: Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de sete ciclos de 13 realizações, no espaço normal (a) e espaço original (b).

A Tabela 6.21 exhibe os parâmetros de avaliação da variabilidade entre as realizações.

Tabela 6.21: Média, desvio padrão e CV dos conjuntos de médias das realizações, para sete ciclos de 13 realizações.

N. de realizações	Média	Desvio padrão	CV (%)
13	276,9378	37,4165	2,2088
26	276,0629	38,1683	2,2379
39	276,1744	40,1771	2,2951
52	276,2628	46,5386	2,4694
65	276,2600	51,4636	2,5968
78	276,2015	51,7612	2,6048
91	276,0511	50,4018	2,5718

A Figura 6.46 ilustra o comportamento de dois dos parâmetros apresentados na Tabela 6.21: a média e o CV dos conjuntos de médias das realizações. Nessa figura, é evidenciado que o CV da distribuição das médias é sistematicamente inferior ao CV considerado estável (representado pela linha horizontal vermelha). Exceto pelo 1º ciclo, a média exibe valores muito satisfatórios. Contudo, em um contexto geral, esse modelo de discretização não se mostra adequado para avaliar o espaço de incerteza da VA V.

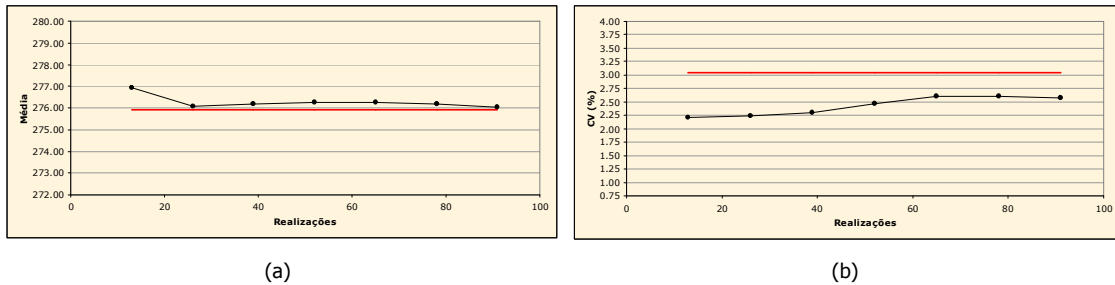


Figura 6.46: Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (sete ciclos de 13 realizações). A linha vermelha representa os valores considerados estáveis.

6.4.3.11. M = 14, 7 ciclos

Com o emprego de 14 classes de discretização, 98 realizações foram executadas, em um processo de 7 ciclos. As médias e os desvios padrão das realizações, apresentados na Figura 6.47, nos espaços Gaussiano e original, validam o processo de simulação.

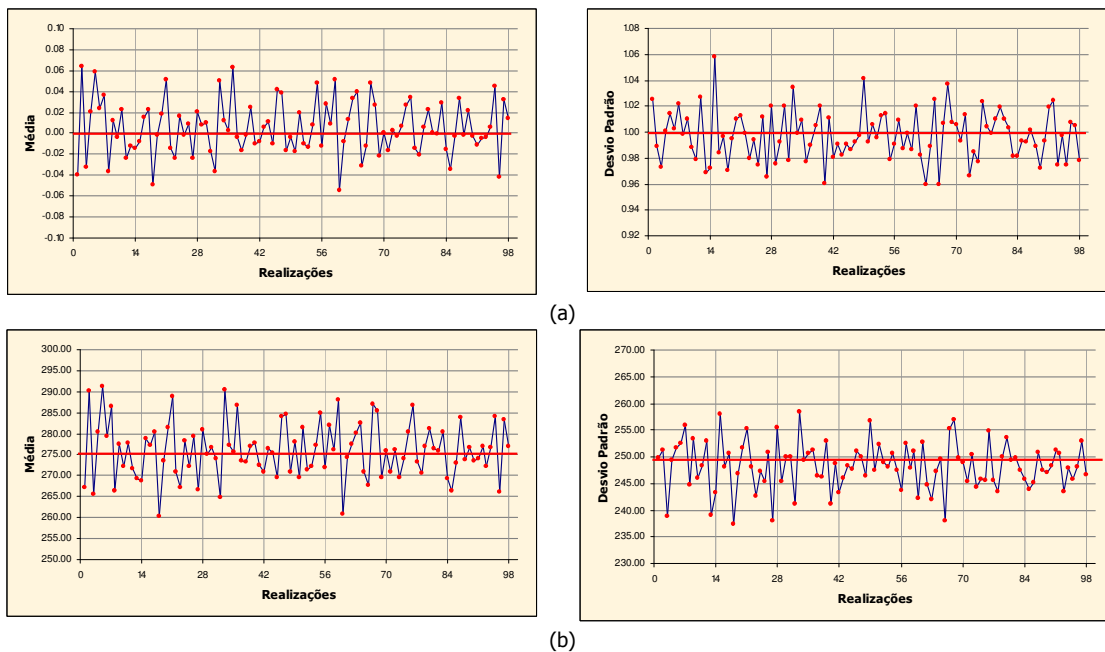
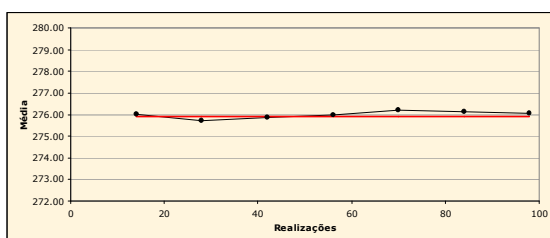


Figura 6.47: Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de sete ciclos de 14 realizações, no espaço normal (a) e espaço original (b).

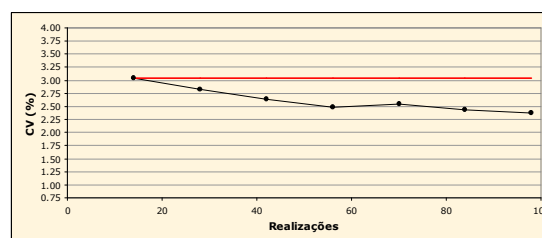
A Tabela 6.22 apresenta os parâmetros medidos para avaliar a variabilidade entre os conjuntos de médias das realizações. Essa tabela exhibe a média, o desvio padrão e o CV de acordo com os ciclos cumulativos de realizações. A representação gráfica dessa tabela, para os parâmetros de média e CV, é dada na Figura 6.48. No gráfico referente à média, todos os ciclos mostram valores muito similares à média considerada estável (representada pela linha horizontal vermelha). Com relação ao CV, somente o 1º ciclo se mostra satisfatório. Nos demais ciclos, o CV é sistematicamente subestimado.

Tabela 6.22: Média, desvio padrão e CV dos conjuntos de médias das realizações, para sete ciclos de 14 realizações.

N. de realizações	Média	Desvio padrão	CV (%)
14	276,0066	70,1374	3,0343
28	275,7234	60,9466	2,8314
42	275,8702	53,0039	2,6391
56	275,9716	47,0901	2,4866
70	276,1887	49,5201	2,5479
84	276,1317	45,2725	2,4367
98	276,0499	42,9920	2,3752



(a)



(b)

Figura 6.48: Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (sete ciclos de 14 realizações). A linha vermelha representa os valores considerados estáveis.

6.4.3.12. M = 15, 6 ciclos

A partir de uma discretização de 15 classes, foram executadas 90 realizações, por meio de seis ciclos. A Figura 6.49 ilustra os gráficos que representam a média e o desvio padrão das realizações no espaço normal. A Figura 6.50 apresenta os mesmos gráficos, no espaço original.

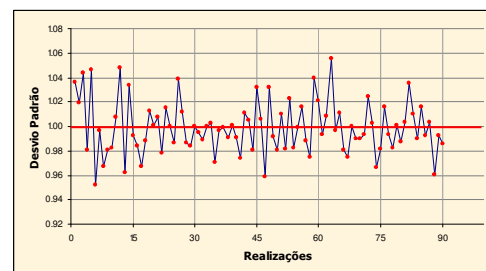
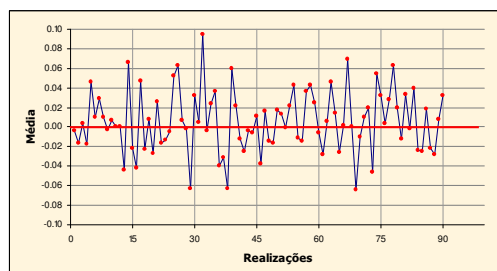


Figura 6.49: Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de seis ciclos de 15 realizações, no espaço normal.

Os gráficos das Figuras 6.49 e 6.50 validam o processo de simulação em ambos os espaços (normal e original). Nessas figuras, os respectivos parâmetros oscilaram moderadamente em torno dos valores condicionantes, representados pela linha horizontal vermelha. Em média, esses valores são reproduzidos. Nesses gráficos, as linhas verticais identificam os ciclos.

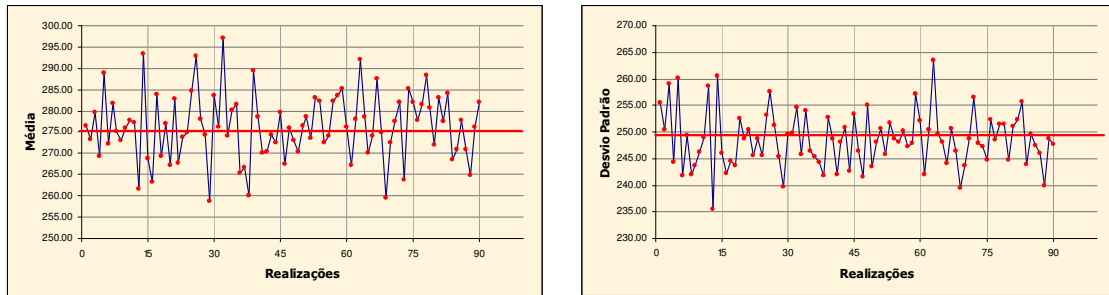
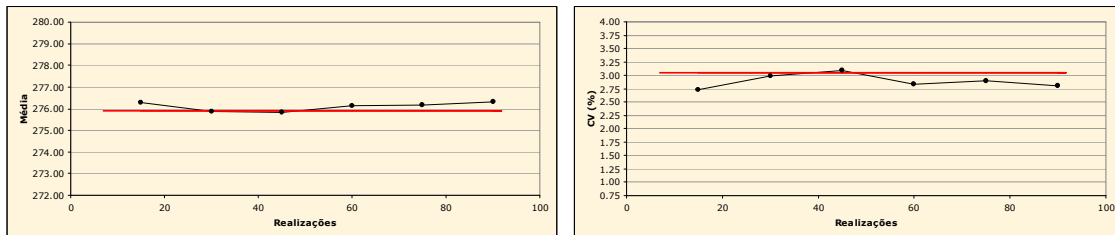


Figura 6.50: Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de seis ciclos de 15 realizações, no espaço original.

O valor dos parâmetros empregados para avaliar a variabilidade entre a médias das realizações, são apresentados na Tabela 6.23. Essa tabela mostra o valor da média, do desvio padrão e do CV referente aos conjuntos de médias de realizações, considerando os seis ciclos realizados. A Figura 6.51, exibe as representações gráficas dos parâmetros de média e CV em função dos conjuntos de realizações, dados pelos ciclos cumulativos. Os gráficos dessa figura mostram que, para esse modelo de discretização, a média e o CV das médias de realizações são praticamente similares aos valores considerados estáveis (representados pela linha horizontal vermelha). Principalmente, o 2º e o 3º ciclos se mostram muito satisfatórios.

Tabela 6.23: Média, desvio padrão e CV dos conjuntos de médias das realizações, para seis ciclos de 15 realizações.

N. de realizações	Média	Desvio padrão	CV (%)
15	276,2787	56,6461	2,7242
30	275,8808	67,7314	2,9831
45	275,8295	72,5731	3,0885
60	276,1188	61,4228	2,8384
75	276,1679	63,9400	2,8954
90	276,3161	60,0903	2,8054



(a)

(b)

Figura 6.51: Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (seis ciclos de 15 realizações). A linha vermelha representa os valores considerados estáveis.

6.4.3.13. M = 16, 6 ciclos

A utilização de 16 classes de discretização levou à geração de seis ciclos, totalizando 96 realizações. O processo de simulação foi validado nos espaços normal e original, como mostra a Figura 6.52. Essa figura mostra a oscilação das médias e dos desvios padrão das realizações, em torno dos respectivos valores condicionantes (representados pela linha horizontal vermelha). Nessa figura, os ciclos podem ser identificados por meio das linhas verticais, que limitam os mesmos.

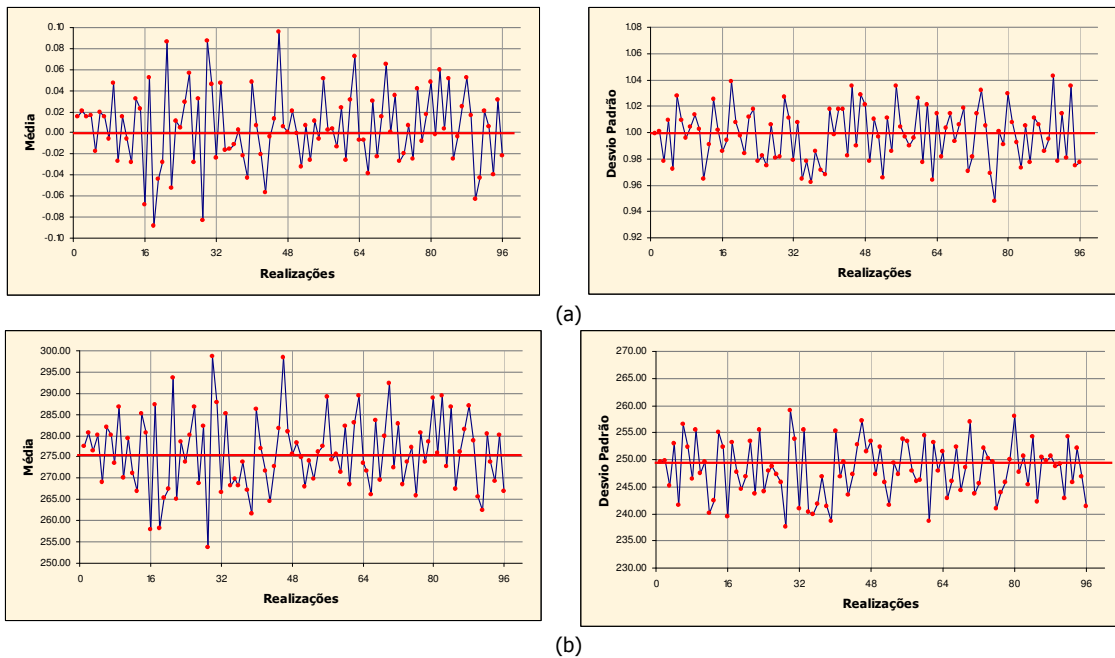


Figura 6.52: Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de seis ciclos de 16 realizações, no espaço normal (a) e espaço original (b).

A Tabela 6.24 exibe as médias, desvios padrão e CV dos conjuntos de médias das realizações. A Figura 6.53 apresenta os gráficos da média e do CV, cujos valores preenchem a Tabela 6.24, em função dos conjuntos de realizações. As médias dos conjuntos de médias das realizações são muito similares à média considerada estável, como pode ser verificado na Tabela 6.24 e no gráfico da esquerda da Figura 6.53. Nesse gráfico, os valores de média se inserem quase que sobre a linha que representa a média considerada estável.

Tabela 6.24: Média, desvio padrão e CV dos conjuntos de médias das realizações, para seis ciclos de 16 realizações.

N. de realizações	Média	Desvio padrão	CV (%)
16	276,1263	53,0239	2,6371
32	276,0009	105,5122	3,7217
48	275,7354	98,4726	3,5989
64	275,9577	83,9747	3,3207
80	276,0901	78,6900	3,2130
96	276,0576	76,0225	3,1584

Na Figura 6.53, em relação ao CV, os valores foram sistematicamente superestimados, exceto no 1º ciclo. Em um contexto geral, esse modelo de discretização não se mostra adequado para caracterizar o espaço de incerteza da VA V.

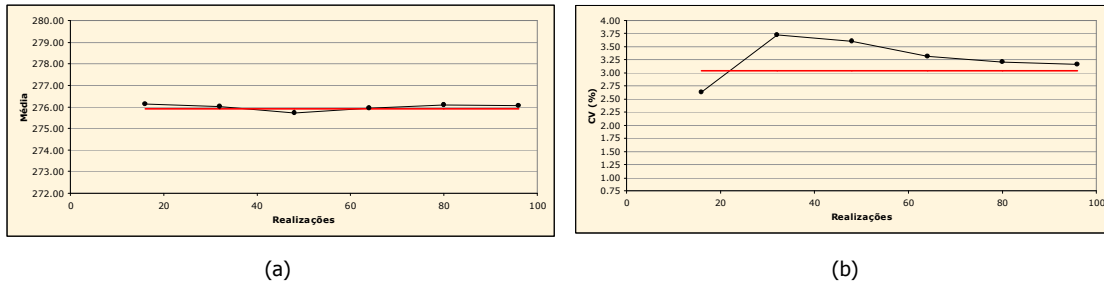


Figura 6.53: Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (seis ciclos de 16 realizações). A linha vermelha representa os valores considerados estáveis.

6.4.3.14. M = 17, 5 ciclos

Com a utilização de 17 classes de discretização, 85 realizações foram executadas, divididas em cinco ciclos. O processo de simulação foi realizado com sucesso, como verificado na Figura 6.54. Nessa figura, as médias e os desvios padrão das realizações, nos espaços normal e original, oscilam em torno dos respectivos valores condicionantes (representados, nos gráficos, pelas linhas horizontais vermelhas). As flutuações ergódicas, representadas na figura, se desenvolvem de maneira simétrica, e, em média, os valores condicionantes são reproduzidos.

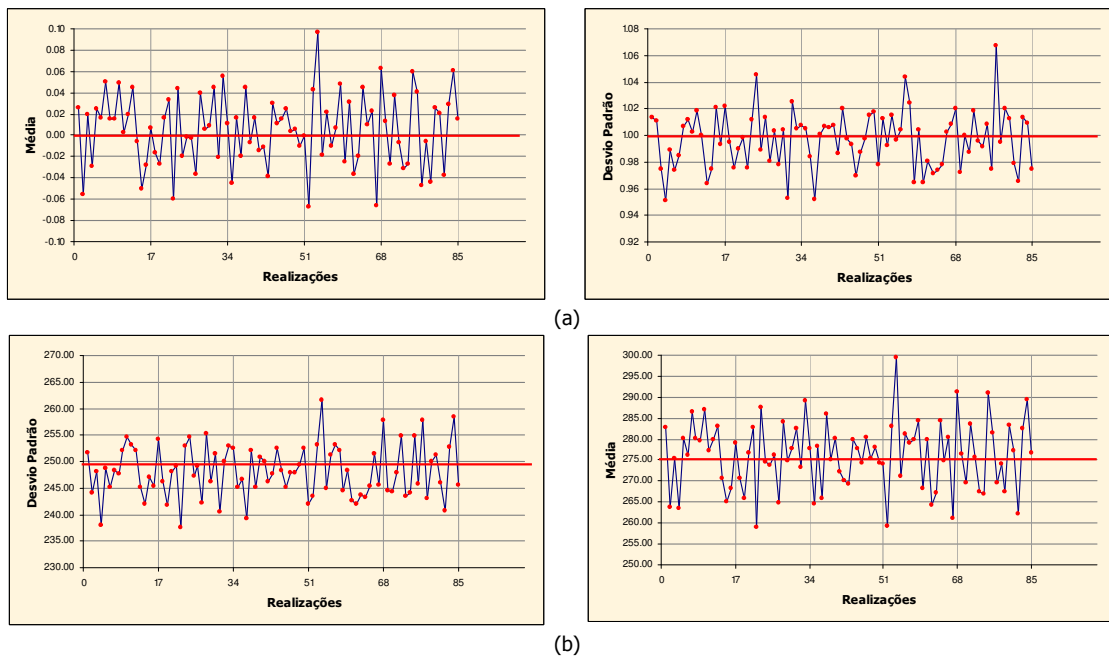


Figura 6.54: Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de cinco ciclos de 17 realizações, no espaço normal (a) e espaço original (b).

A Figura 6.55 exibe dois gráficos que relacionam a média e o CV, dos conjuntos de médias das realizações, em função dos ciclos cumulativos de realizações. Os respectivos valores representados nos gráficos são apresentados na Tabela 6.25. Nessa tabela, também são listados os desvios padrão, de acordo com os conjuntos de realizações. Uma análise sobre essa tabela e sobre os gráficos da Figura 6.55, indica que esse modelo de discretização não foi capaz de caracterizar adequadamente o espaço de incerteza da VA V. Embora os resultados obtidos em relação à média tenham sido satisfatórios, o mesmo não pode ser afirmado em relação ao CV. Esse parâmetro foi sistematicamente subestimado ao longo dos ciclos de realizações.

Tabela 6.25: Média, desvio padrão e CV dos conjuntos de médias das realizações, para cinco ciclos de 17 realizações.

N. de realizações	Média	Desvio padrão	CV (%)
17	276,3578	53,6107	2,6494
34	276,1710	57,2743	2,7403
51	275,7952	48,0779	2,5141
68	276,1059	63,2959	2,8815
85	276,1172	63,5863	2,8879

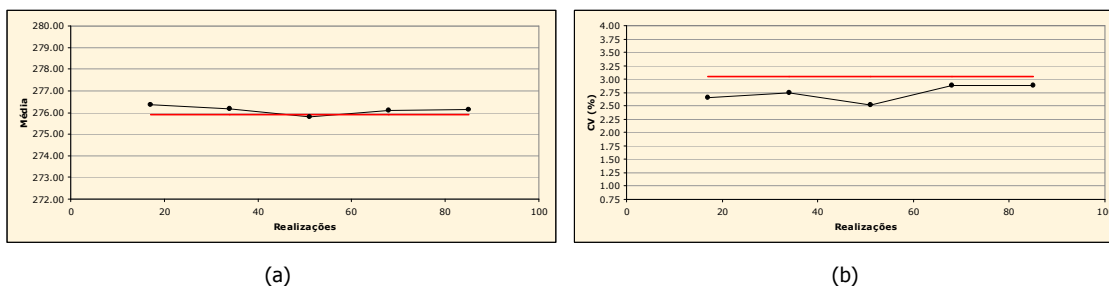


Figura 6.55: Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (cinco ciclos de 17 realizações). A linha vermelha representa os valores considerados estáveis.

6.4.3.15. M = 18, 5 ciclos

Para uma discretização de 18 classes, foram necessários cinco ciclos para se atingir 90 realizações. O processo de simulação foi validado nos espaços normal e original, como mostram os gráficos das Figuras 6.56 e 6.57.

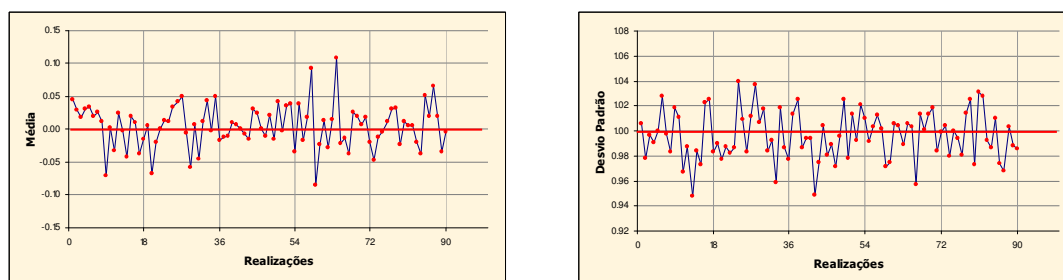


Figura 6.56: Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de cinco ciclos de 18 realizações, no espaço normal.

Nos gráficos de média das Figuras 6.56 e 6.57, foi verificado uma ampliação do domínio de variação em relação aos processos de simulação anteriores. A realização número 64 foi a responsável por este comportamento. Entretanto, isso não invalida o processo de simulação. Isso porque, em média, os parâmetros condicionantes, representados nesses gráficos pelas linhas horizontais vermelhas, foram reproduzidos.

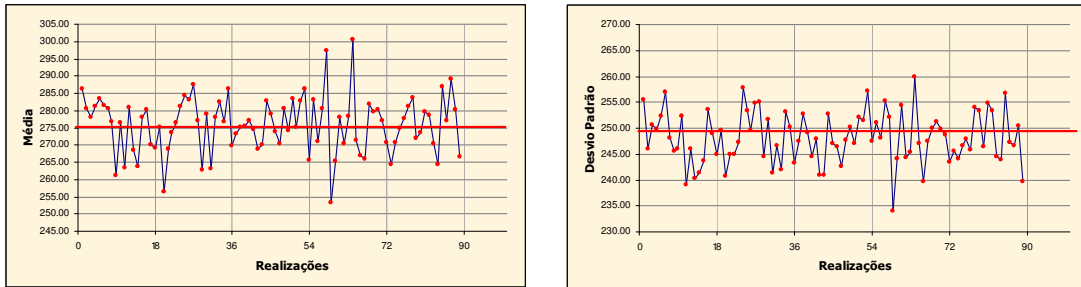
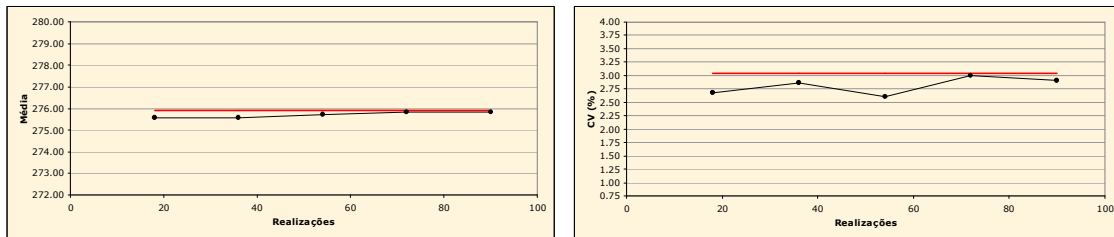


Figura 6.57: Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de cinco ciclos de 18 realizações, no espaço original.

A Tabela 6.26 exibe as médias e o CV, além do desvio padrão, das distribuições das médias dos conjuntos de realizações. A representação gráfica da média e do CV é dada na Figura 6.58. Pode ser verificado que, ambos os parâmetros medidos, foram subestimados ao longo dos ciclos de realizações. Somente quando foram incluídos os últimos dois ciclos esses parâmetros convergiram para valores toleráveis.

Tabela 6.26: Média, desvio padrão e CV dos conjuntos de médias das realizações, para cinco ciclos de 18 realizações.

N. de realizações	Média	Desvio padrão	CV (%)
18	275,5572	54,2647	2,6733
36	275,5863	62,3720	2,8657
54	275,7306	51,4073	2,6003
72	275,8475	68,2218	2,9943
90	275,8288	64,3007	2,9072



(a)

(b)

Figura 6.58: Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (cinco ciclos de 18 realizações). A linha vermelha representa os valores considerados estáveis.

6.4.3.16. M = 19, 5 ciclos

Com o emprego de 19 classes de discretização, cinco ciclos foram executados, totalizando 95 realizações. As médias e os CV das realizações, nos espaços normal e original, validam o processo de

simulação (Figura 6.59). Esses parâmetros oscilaram moderadamente em torno dos respectivos valores condicionantes, representados pelas linhas horizontais vermelhas. Em média, esses valores foram reproduzidos. Nesses gráficos, as linhas verticais delimitam os cinco ciclos realizados.

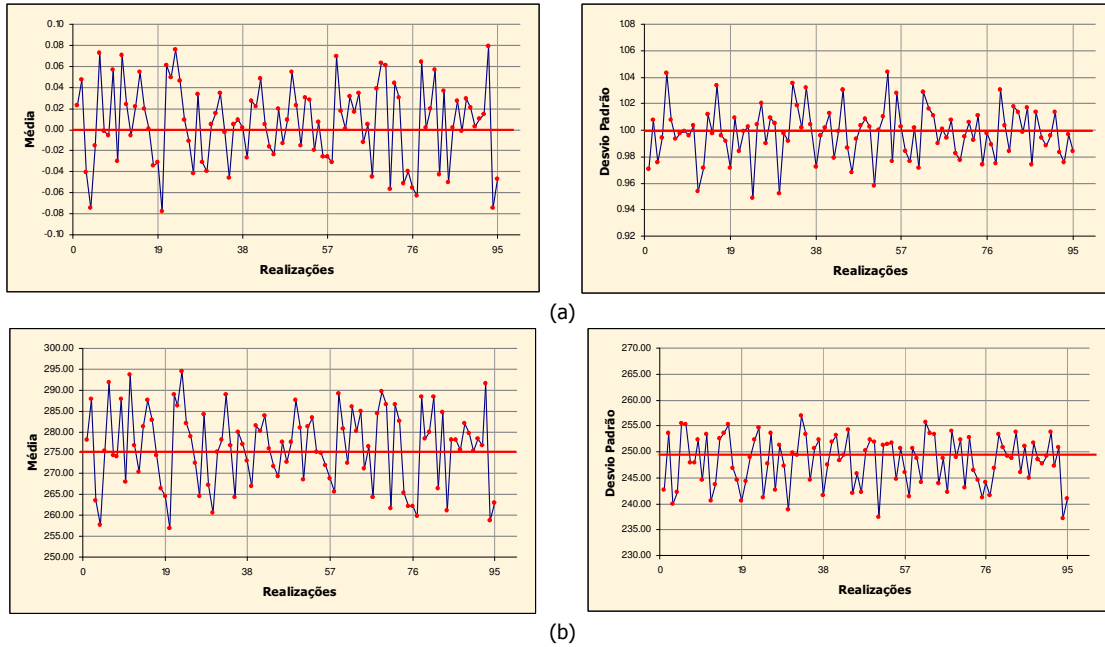


Figura 6.59: Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de cinco ciclos de 19 realizações, no espaço normal (a) e espaço original (b).

A Figura 6.60 ilustra graficamente dois dos parâmetros apresentados na Tabela 6.27: a média e o CV dos conjuntos de médias das realizações. Essa tabela também exibe os respectivos desvios padrão, de acordo com os ciclos cumulativos de realizações.

Tabela 6.27: Média, desvio padrão e CV dos conjuntos de médias das realizações, para cinco ciclos de 19 realizações.

N. de realizações	Média	Desvio padrão	CV (%)
19	276,6521	98,1841	3,5817
38	276,4592	98,3928	3,5880
57	276,4000	76,6920	3,1684
76	276,4073	82,4493	3,2851
95	276,3258	84,3538	3,3238

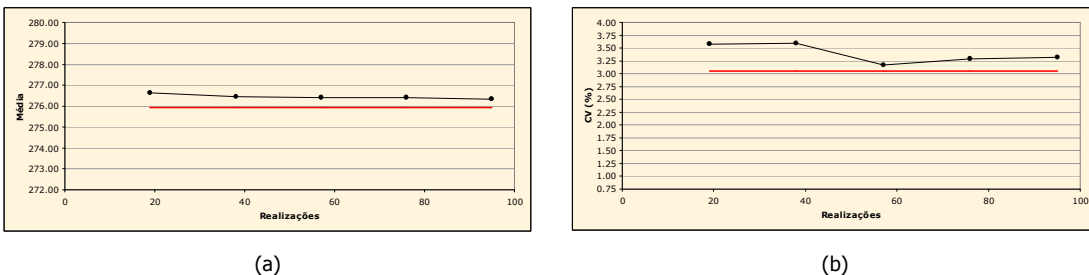


Figura 6.60: Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (cinco ciclos de 19 realizações). A linha vermelha representa os valores considerados estáveis.

A Figura 6.60 mostra que o desempenho desse modelo de discretização não foi satisfatório. Ambos os parâmetros medidos foram superestimados ao longo dos ciclos de realizações.

6.4.3.17. M = 20, 5 ciclos

A discretização da distribuição de probabilidade Gaussiana G em 20 classes, levou à execução de 100 realizações, divididas em cinco ciclos. O processo de simulação foi validado nos espaços normal e original, conforme mostra a Figura 6.61. Essa figura mostra quatro gráficos que descrevem a relação da média e do desvio padrão em função das realizações, em ambos os espaços. Nesses gráficos, os valores condicionantes são representados pelas linhas horizontais vermelhas. As flutuações ergódicas não tendenciosas, em torno desses valores, indicam que o processo de simulação foi realizado com sucesso. As linhas verticais, representam os limites de cada ciclo.

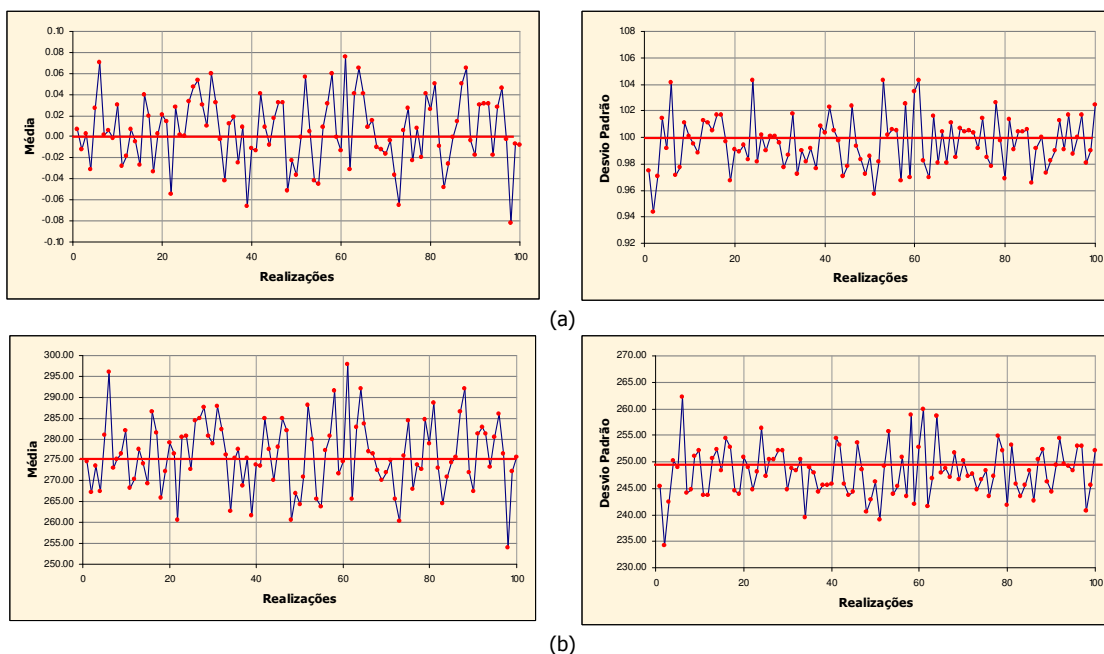


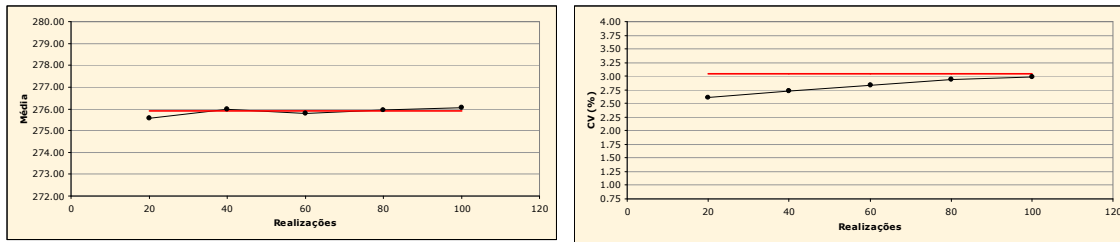
Figura 6.61: Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação a partir de cinco ciclos de 20 realizações, no espaço normal (a) e espaço original (b).

Os parâmetros de média e CV da distribuição das médias das realizações, apresentadas na Tabela 6.28 (além do CV) e representadas na Figura 6.62, indicam que o espaço de incerteza da VA V foi subestimado.

Tabela 6.28: Média, desvio padrão e CV dos conjuntos de médias das realizações, para cinco ciclos de 20 realizações.

N. de realizações	Média	Desvio padrão	CV (%)
20	275,5644	51,8181	2,6123
40	275,9893	56,7151	2,7287
60	275,7832	61,4970	2,8435
80	275,9551	66,0424	2,9449
100	276,0452	67,9956	2,9872

Na Figura 6.62, os resultados em relação à média são satisfatórios, variando levemente em torno da média considerada estável (representada pela linha horizontal vermelha). Contudo, em relação ao CV, os resultados são sistematicamente inferiores ao CV considerado estável. Somente nos últimos dois ciclos o CV adquire valores toleráveis.



(a)

(b)

Figura 6.62: Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico clássico (cinco ciclos de 20 realizações). A linha vermelha representa os valores considerados estáveis.

6.4.3.18. Validação prática do algoritmo LHS cíclico clássico

Na seção 5.5.3 do Capítulo 5 (pp. 116 - 118), o algoritmo LHS cíclico clássico foi testado teoricamente. Nessa seção, o algoritmo é verificado na prática, com os dados do estudo de caso.

A verificação do algoritmo LHS cíclico clássico foi realizada na localização u_1 ($j = 1$) da malha, em um caso no qual 10 classes discretizavam a distribuição de probabilidade Gaussiana G. Foram executados três ciclos, totalizando 30 realizações. A Tabela 6.29 mostra as classes e as probabilidades (plhs), assim como os valores Gaussianos x_{plhs} associados, segundo os três ciclos executados.

Tabela 6.29: Classes e probabilidades (plhs), tiradas na localização u_1 , para a obtenção dos valores Gaussianos x_{plhs} , de acordo com três ciclos de 10 realizações ($M = 10$).

Seqüência no ciclo	Ciclo 1			Ciclo 2			Ciclo 3		
	classe	plhs	x_{plhs}	classe	plhs	x_{plhs}	classe	plhs	x_{plhs}
1	10	0,9283	1,4632	8	0,7854	0,7906	9	0,8749	1,1499
2	3	0,2405	-0,7047	1	0,0583	-1,5692	5	0,4360	-0,1611
3	7	0,6436	0,3681	7	0,6336	0,3414	7	0,6908	0,4981
4	8	0,7922	0,8141	5	0,4029	-0,2458	6	0,5557	0,1401
5	5	0,4600	-0,1004	9	0,8570	1,0669	4	0,3270	-0,4482
6	1	0,0116	-2,2701	4	0,3087	-0,4995	10	0,9349	1,5133
7	4	0,3455	-0,3975	3	0,2252	-0,7547	1	0,0696	-1,4788
8	2	0,1914	-0,8727	10	0,9911	2,3698	3	0,2065	-0,8186
9	9	0,8728	1,1397	2	0,1601	-0,9940	2	0,1717	-0,9475
10	6	0,5619	0,1558	6	0,5114	0,0286	8	0,7507	0,6767

A Figura 6.63 representa, graficamente, os resultados exibidos na Tabela 6.29. Nessa figura, as tiragens de cada ciclo são identificadas por símbolos de cores diferentes. É verificado que exatamente três tiragens, de ciclos diferentes, são realizadas entre as linhas horizontais que delimitam

o domínio dos ciclos. Isso indica que, em cada ciclo, uma tiragem foi realizada de cada classe, o que comprova a eficiência de construção do algoritmo.

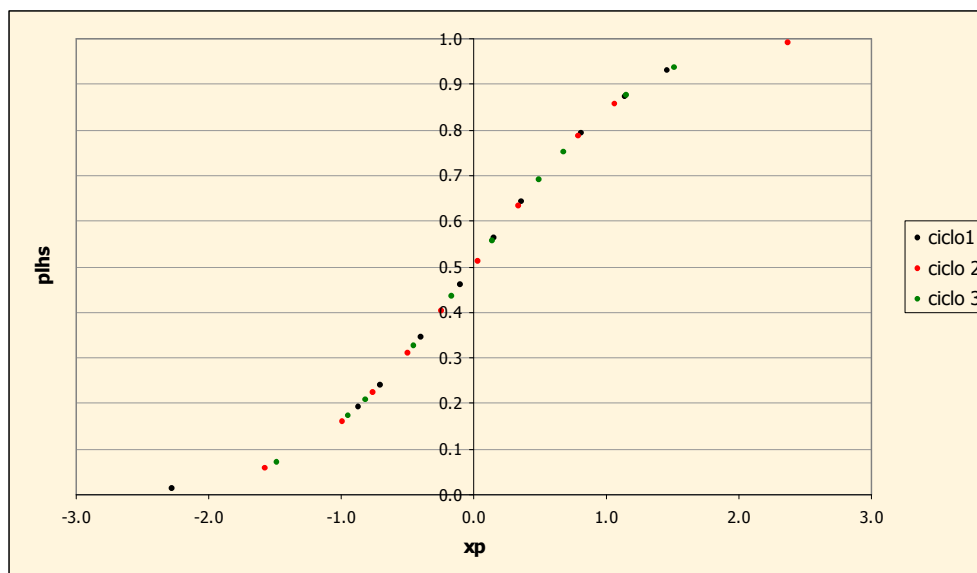


Figura 6.63: Distribuição experimental de probabilidade Gaussiana, associada à localização u_i , para o caso no qual $M = 10$, resultante da aplicação do algoritmo LHS cíclico clássico, em três ciclos, totalizando 30 tiragens.

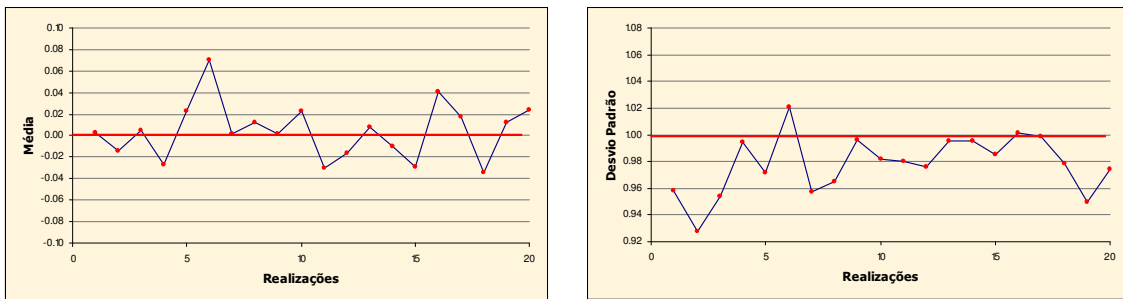
6.4.4. Algoritmo LHS determinístico

Para avaliar a eficiência do algoritmo LHS determinístico, foram testados vários modelos de discretização da distribuição de probabilidade Gaussiana G . Em todos os modelos, a mediana das classes foi amostrada ($p = 0,5$).

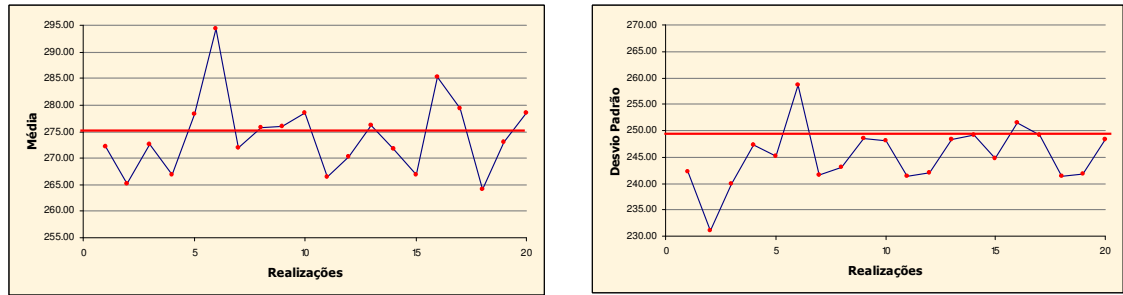
A Tabela 6.30 mostra a média e o desvio padrão das realizações, nos espaços normal e original, de acordo com os modelos de discretização. Nessa tabela, os valores condicionantes, referentes aos respectivos espaços, são dados entre parênteses. É verificado que são necessárias no mínimo 50 realizações ($M = 50$) para que o processo de simulação seja validado. A utilização de modelos de discretização com menos de 50 classes, resulta em processos de simulação nos quais não há reprodução dos valores condicionantes (erros inseridos domínio [1,50; 3,07 %]). Nesses casos, principalmente o desvio padrão é subestimado. As Figuras 6.64 e 6.65 representam graficamente a média e o desvio padrão das realizações, nos espaços normal e original, segundo dois modelos de discretização: $M = 20$ e $M = 60$. Nos gráficos em relação ao modelo de 20 classes, é observado que o desvio padrão é sistematicamente subestimado. Em ambos os espaços, somente duas realizações apresentaram desvio padrão superior ao condicionante (representado pela linha horizontal vermelha). Em relação ao modelo de 60 classes, no espaço original, o desvio padrão é levemente subestimado, apresentando, em média, erro de 0,83 %.

Tabela 6.30: Média das médias e dos desvios padrão para 19 processos de simulação, conforme o número de classes M, no espaço normal e no original. Valores condicionantes entre parênteses.

Número de classes (M)	Média			
	Espaço normal		Espaço original	
	Média (0,00)	Desvio padrão (1,00)	Média (275,31)	Desvio padrão (249,72)
10	0,0029	0,9623	272,563	241,7491
11	0,0063	0,9633	273,4611	242,2664
12	0,0081	0,9682	274,3590	243,3537
13	0,0077	0,9701	274,4535	243,6557
14	0,0045	0,9743	273,8932	244,2628
15	0,0035	0,9792	274,0972	245,2136
16	0,0045	0,9765	274,3426	244,6811
17	0,0058	0,9752	274,4371	244,6442
18	0,0043	0,9753	274,0858	244,5499
19	0,0061	0,9777	274,7862	244,9828
20	0,0040	0,9780	274,1775	245,1214
30	0,0052	0,9876	275,4092	246,7567
40	0,0055	0,9881	275,4337	246,8739
50	0,0064	0,9910	275,9496	247,5720
60	0,0046	0,9929	275,6502	247,6645
70	0,0055	0,9932	275,8938	247,7701
80	0,0037	0,9925	275,5041	247,2939
90	0,0056	0,9924	275,8499	247,5005
100	0,0065	0,9933	276,1256	247,7393



(a)



(b)

Figura 6.64: Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação (LHS determinístico) a partir de 20 realizações (M = 20), no espaço normal (a) e espaço original (b).

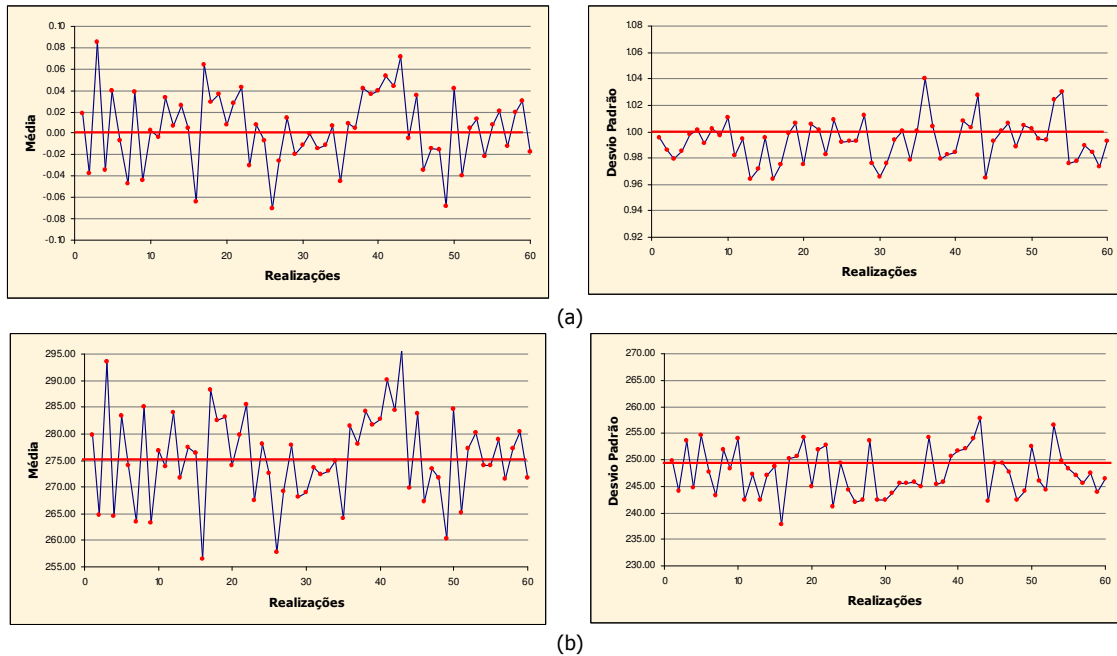


Figura 6.65: Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação (LHS determinístico) a partir de 60 realizações ($M = 60$), no espaço normal (a) e espaço original (b).

A variabilidade entre os conjuntos de médias das realizações, foi avaliado de maneira análoga aos casos anteriores. A Tabela 6.31 exibe a média, o desvio padrão e o CV desses conjuntos, de acordo o modelo de discretização.

Tabela 6.31: Média, desvio padrão e CV dos conjuntos de médias das realizações, para 19 processos de simulação.

Número de classes (M)	Média	Desvio padrão	CV (%)
10	272,5630	45,3330	2,4702
11	273,4611	42,3658	2,3802
12	274,3590	62,9263	2,8913
13	274,4536	32,7404	2,0848
14	273,8932	68,0358	3,0115
15	274,0972	58,2002	2,7833
16	274,3426	44,8688	2,4416
17	274,4371	47,3392	2,5071
18	274,0858	54,4424	2,6920
19	274,7862	86,9670	3,3938
20	274,1775	50,3931	2,5891
30	275,4092	64,5792	2,9179
40	275,4337	63,0592	2,8831
50	275,9496	45,3461	2,4403
60	275,6502	69,5223	3,0249
70	275,8939	56,0167	2,7128
80	275,5041	75,6254	3,1565
90	275,8499	62,0372	2,8553
100	276,1427	63,6134	2,8883

Os parâmetros apresentados na Tabela 6.31 são representados graficamente na Figura 6.66. Os gráficos dessa figura mostram que, a partir de 30 realizações, ambos parâmetros começam a apresentar valores próximos àqueles considerados estáveis, indicados pela linha horizontal vermelha. Também, é observado que, a partir de 30 realizações, esse algoritmo se torna muito similar ao algoritmo LHS clássico. Isso é evidenciado pelos resultados. Os gráficos das Figuras 6.21 e 6.66 mostram um padrão muito similar, a partir de 30 realizações. Isso acontece porque o domínio de variação das classes (em termos de probabilidades) se torna mais restrito com o aumento do número de classes.

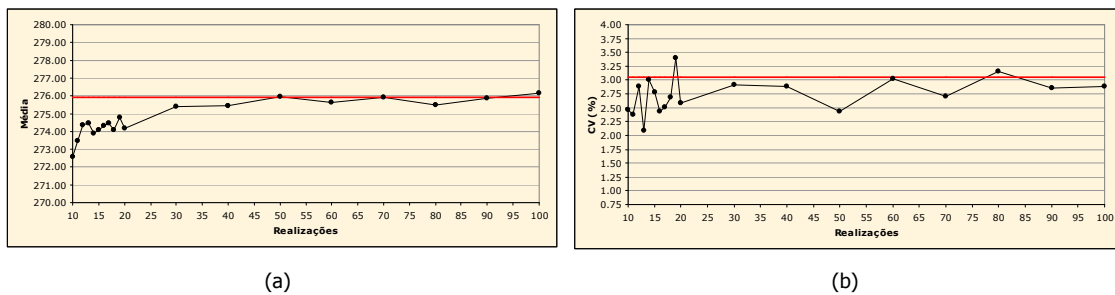


Figura 6.66: Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS determinístico. A linha vermelha representa os valores considerados estáveis.

6.4.4.1. Validação prática do algoritmo LHS determinístico

Na seção 5.5.3 do Capítulo 5 (pp. 116 - 118), o algoritmo LHS determinístico foi validado teoricamente. Nessa seção, esse algoritmo é verificado na prática, com os dados do estudo de caso.

A verificação do algoritmo LHS determinístico foi realizada na localização u_1 ($j = 1$) da malha, para o caso no qual a distribuição de probabilidade Gaussiana G é discretizada em 10 classes. Em cada classe, uma amostragem foi realizada na sua mediana (proporção, $p = 0,5$). A Tabela 6.32 mostra as classes e as probabilidades ($plhs_{ad}$), assim como os valores Gaussianos x_{plhs} associados. O valor simulado, no espaço normal, também é apresentado.

Tabela 6.32: Parâmetros de simulação para a localização $j = 1$ da malha ($L = M = 10$).

Realização (i)	Classe [$\pi_i(u_i)$]	$plhs_{ad}$	$x_{plhs}^{(i)}(u_i)$	$y^{(i)}(u_i)$
1	10	0,95	1,6449	-0,2649
2	3	0,25	-0,6745	-1,4966
3	7	0,65	0,3853	-0,9907
4	8	0,75	0,6745	-0,2256
5	5	0,45	-0,1257	-1,1477
6	1	0,05	-1,6449	-2,2998
7	4	0,35	-0,3853	0,3903
8	2	0,15	-1,0364	-2,3182
9	9	0,85	1,0364	-0,4455
10	6	0,55	0,1257	-1,258

A Figura 6.67 exibe um gráfico no qual relaciona as probabilidades ($p_{lhs_{ad}}$) e os valores Gaussianos ($x_{p_{lhs}}$) associados. Esse gráfico configura a distribuição experimental de probabilidade Gaussiana, associada à localização u_1 ($j = 1$). Nesse gráfico, as linhas horizontais tracejadas indicam a mediana das classes, enquanto que a linha horizontal contínua delimita o domínio das classes (em termos de probabilidades). É verificado que exatamente uma tiragem é realizada na mediana de cada classe. Esse comportamento valida, na prática, a construção do algoritmo.

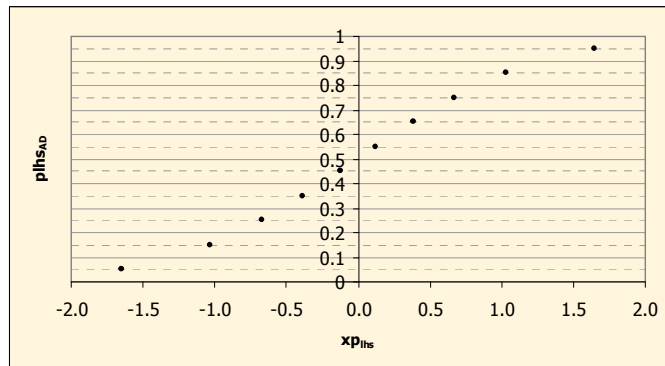


Figura 6.67: Distribuição experimental de probabilidade Gaussiana, associada à localização u_1 , para o caso no qual $M = 10$, resultante da aplicação do algoritmo LHS determinístico.

6.4.5. Algoritmo LHS cíclico determinístico

Para avaliar o desempenho em relação à característica cíclica do algoritmo determinístico, foi realizado um procedimento análogo ao executado na seção 6.4.3. Foram testados vários modelos de discretização, executando processos de simulação diferentes, com números de ciclos diferentes. De acordo com o número de classes M , foram executados tantos ciclos quanto necessários para alcançar, no máximo, 100 realizações. A Tabela 6.33 apresenta a média e o desvio padrão das realizações, nos espaços normal e original, de acordo com os modelos de discretização.

Tabela 6.33: Média das médias e dos desvios padrão para oito processos de simulação, conforme o número de classes M , no espaço normal e no original. Valores condicionantes entre parênteses.

Número de classes (M)	Média			
	Espaço normal		Espaço original	
	Média (0,00)	Desvio padrão (1,00)	Média (275,31)	Desvio padrão (249,72)
4	0,0069	0,9125	268,3910	232,0317
5	0,0049	0,9288	269,8638	235,3980
6	0,0059	0,9395	271,0322	237,6771
8	0,0056	0,9538	272,4504	240,3538
10	0,0046	0,9628	272,9017	242,0009
12	0,0057	0,9680	273,7342	243,2237
15	0,0051	0,9747	274,2098	244,4577
20	0,0053	0,9797	274,7202	245,3138

Na tabela 6.33 é percebido que, em média, tanto no espaço normal quanto no original, as realizações não reproduzem os valores condicionantes (dados entre parênteses). Os maiores erros são observados no espaço original, superando 7,5 %, em relação ao desvio padrão. Esse comportamento já era esperado, pois os testes de aplicação do algoritmo LHS determinístico (seção 6.4.4) mostraram que são necessárias no mínimo 50 realizações ($M = 50$) para que o processo de simulação seja validado.

Os gráficos da Figura 6.68 mostram a média e o desvio padrão em função das realizações, para o caso no qual $M = 10$ (100 realizações, 10 ciclos de 10 realizações). Principalmente, em relação ao desvio padrão, as realizações mostram valores inferiores ao condicionante (indicado pela linha horizontal vermelha). No espaço normal, a variabilidade de 97 % das realizações foi subestimada. No espaço original: 96 %. A representação gráfica da média e do CV das distribuições das médias dos conjuntos de realizações, é dada na Figura 6.69.

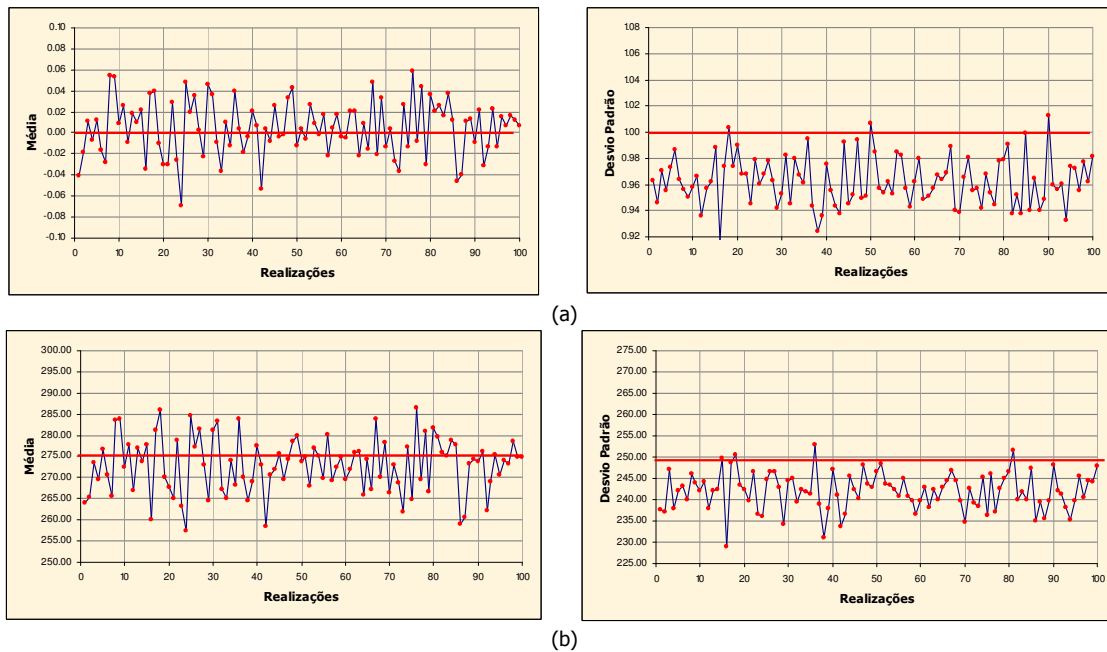


Figura 6.68: Médias (esquerda) e desvios padrão (direita) das realizações, resultantes do processo de simulação (LHS cíclico determinístico) a partir de 10 ciclos de 10 realizações, no espaço normal (a) e espaço original (b).

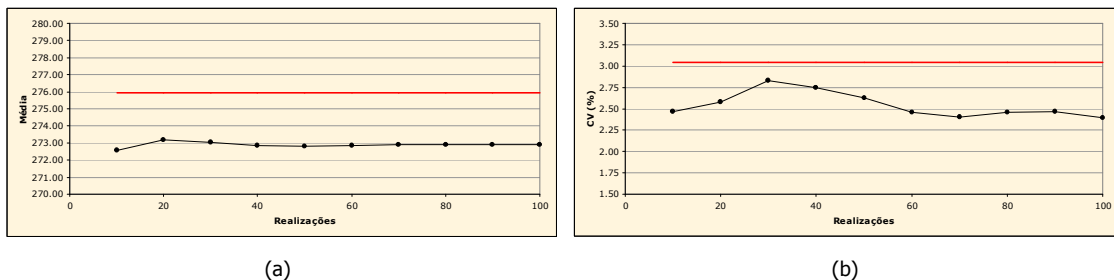


Figura 6.69: Média (a) e CV (b) da distribuição das médias das realizações, de acordo com o algoritmo LHS cíclico determinístico (10 ciclos de 10 realizações). A linha vermelha representa os valores considerados estáveis.

A partir dos resultados apresentados acima (Figuras 6.68 e 6.69), pode-se afirmar que o caráter cíclico não pode ser aplicado com sucesso no algoritmo LHS determinístico. A execução de dois ciclos de 50 realizações não se justifica, pois o resultado obtido com a utilização de 50 classes ($M = 50$), por meio do algoritmo LHS determinístico, não foi satisfatório.

6.4.6. Estratificação da dclp

Antes de aplicar a terceira variação do algoritmo LHS (com caminho determinístico), é preciso verificar o comportamento das distribuições condicionais locais de probabilidade (dclp) com relação à sua estratificação. Para realizar essa tarefa, foram selecionadas três localizações, inseridas em zonas de diferentes concentrações da VA V. A verificação foi realizada em localizações de:

- i. baixo teor (média de 25,00 ppm);
- ii. médio teor (média de 280 ppm) e
- iii. alto teor (média de 500 ppm).

Nessas localizações, as dclp, obtidas por meio do algoritmo que envolve a técnica LHS (clássico), foram comparadas com aquelas obtidas por meio do algoritmo original, que utiliza a técnica de Monte Carlo (MC). Essa comparação foi realizada nos espaços normal e original, a partir de processos de simulação de 20 realizações.

A Figura 6.70 mostra as dclp, segundo os dois algoritmos, na localização selecionada inserida na zona de baixo teor.

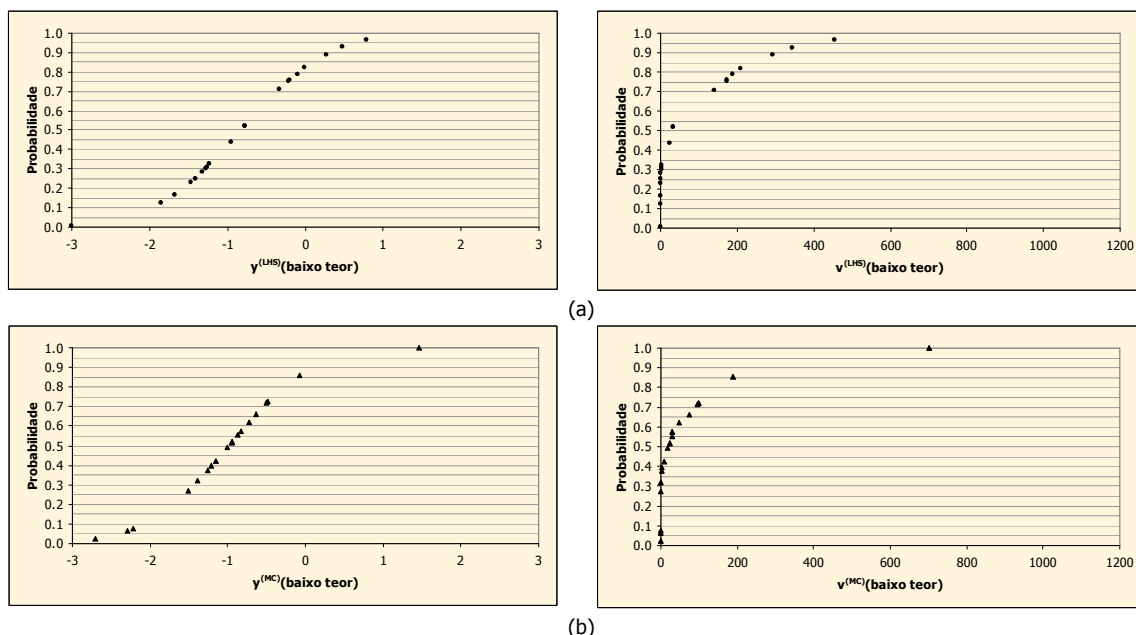


Figura 6.70: Modelos de incerteza (dclp) associados à localização inserida em uma zona de baixo teor, de acordo com os algoritmos LHS clássico (a) e com o algoritmo original (b), nos espaços normal (esquerda) e original (direita).

Na Figura 6.70, é verificado que a qualidade de definição das dclp, obtidas por meio do algoritmo original, é inferior em relação à obtida por meio do algoritmo modificado. Esse comportamento é evidente em ambos os espaços. Entretanto, nenhuma dclp da Figura 6.70 é perfeitamente estratificada (as linhas horizontais auxiliam na identificação dos estratos). As dclp provenientes dos dois algoritmos mostram 14 classes amostradas. Contudo, as seis classes não amostradas pelo algoritmo LHS clássico estão mais distribuídas no domínio de probabilidade. Isso proporciona uma maior definição dos modelos de incerteza. A Tabela 6.34 mostra as classes não amostradas (em termos de probabilidade), segundo os dois algoritmos.

Tabela 6.34: Classes não amostradas da Figura 6.70, em termos de probabilidade, segundo os dois algoritmos (LHS e MC).

Classes não amostradas	
LHS	MC
[0,05; 0,10)	[0,10; 0,15)
[0,35; 0,40)	[0,15; 0,20)
[0,45; 0,50)	[0,20; 0,25)
[0,55; 0,60)	[0,75; 0,80)
[0,60; 0,65)	[0,80; 0,85)
[0,65; 0,70)	[0,90; 0,95)

Nas dclp do espaço normal da Figura 6.70, é natural observar a ocorrência de uma freqüência maior de valores negativos, visto que essa localização se encontra em uma zona de baixos teores. A curva oriunda do algoritmo LHS clássico está centrada em -0,8148, enquanto que aquela proveniente do algoritmo original, está centrada em -0,9885.

A Figura 6.71 mostra as distribuições experimentais de probabilidade Gaussiana, oriundas dos dois algoritmos. Essas distribuições foram utilizadas para determinar os valores simulados do espaço normal, apresentados na Figura 6.70. A distribuição de probabilidade relativa ao algoritmo LHS clássico é perfeitamente estratificada, como esperado. Contudo, essa característica não foi suficiente para resultar em dclp estratificadas.

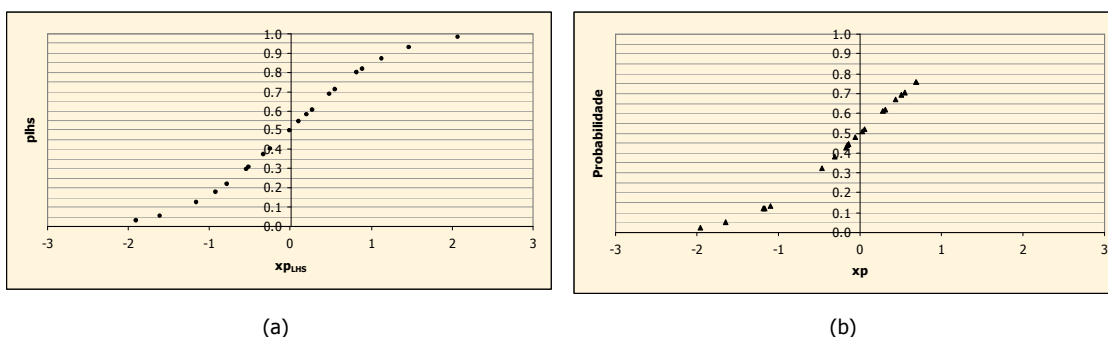


Figura 6.71: Distribuições experimentais de probabilidade Gaussiana, de acordo com o algoritmo LHS clássico (a) e original (b), utilizadas para a determinação dos valores simulados da Figura 6.70.

Na Figura 6.72, são apresentados os modelos de incerteza (dclp) associados à localização selecionada na zona de médio teor. Essas dclp foram determinadas de acordo com os algoritmos LHS clássico e com o original, nos espaços normal e original. Novamente, a qualidade de definição das curvas oriundas do algoritmo modificado é superior, embora não sejam estratificadas. É visível que, os modelos de incerteza, obtidos por meio do algoritmo original, apresentam mais lacunas de amostragem. Essa observação é corroborada pelo número de classes não amostradas. Enquanto que as curvas da Figura 6.72 (a) mostram cinco classes não amostradas, as da Figura 6.72 (b) apresentam sete. A Tabela 6.35 exibe essas classes, de acordo com os algoritmos.

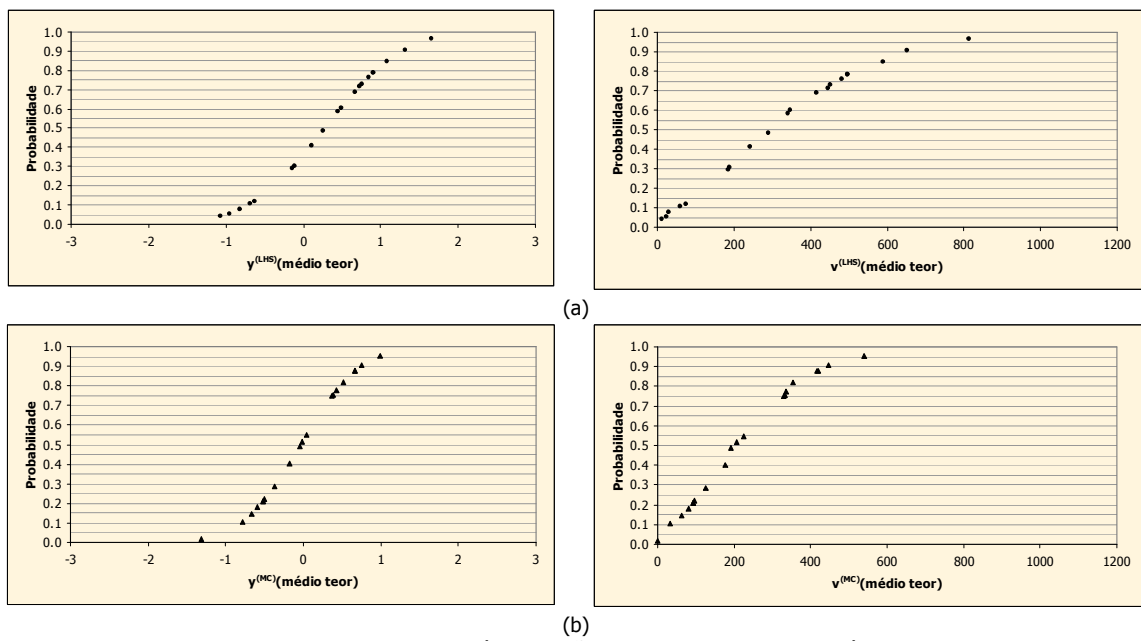


Figura 6.72: Modelos de incerteza (dclp) associados à localização inserida em uma zona de médio teor, de acordo com os algoritmos LHS clássico (a) e com o algoritmo original (b), nos espaços normal (esquerda) e original (direita).

Tabela 6.35: Classes não amostradas da Figura 6.72, em termos de probabilidade, segundo os dois algoritmos (LHS e MC).

Classes não amostradas	
LHS	MC
[0,15; 0,20)	[0,05; 0,10)
[0,20; 0,25)	[0,30; 0,35)
[0,35; 0,40)	[0,35; 0,40)
[0,50; 0,55)	[0,55; 0,60)
[0,85; 0,90)	[0,60; 0,65)
	[0,65; 0,70)
	[0,70; 0,75)

As dclp, do espaço normal, da Figura 6.72 mostram-se praticamente centradas. Esse comportamento já era esperado, pois esses modelos de incerteza estão localizados em uma região de

médios teores. Nesse caso, a dclp obtida por meio do algoritmo LHS clássico está centrada em 0,2897, enquanto que aquela oriunda do algoritmo original, está centrada em -0,0348.

A Figura 6.73 ilustra as distribuições experimentais de probabilidade Gaussiana, obtidas por meio dos dois algoritmos. Novamente, a estratificação da distribuição de probabilidade Gaussiana, oriunda do algoritmo LHS clássico, não foi suficiente para proporcionar modelos de incerteza estratificados.

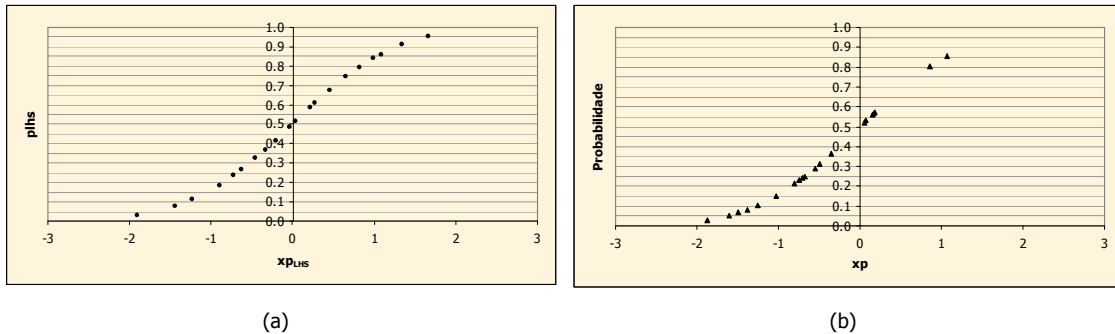


Figura 6.73: Distribuições experimentais de probabilidade Gaussiana, de acordo com o algoritmo LHS clássico (a) e original (b), utilizadas para a determinação dos valores simulados da Figura 6.72.

A Figura 6.74 ilustra os modelos de incerteza associados à localização selecionada na zona de alto teor, de acordo com os dois algoritmos testados. Como nas figuras anteriores, a Figura 6.74 também apresenta as dclp nos espaços normal e original.

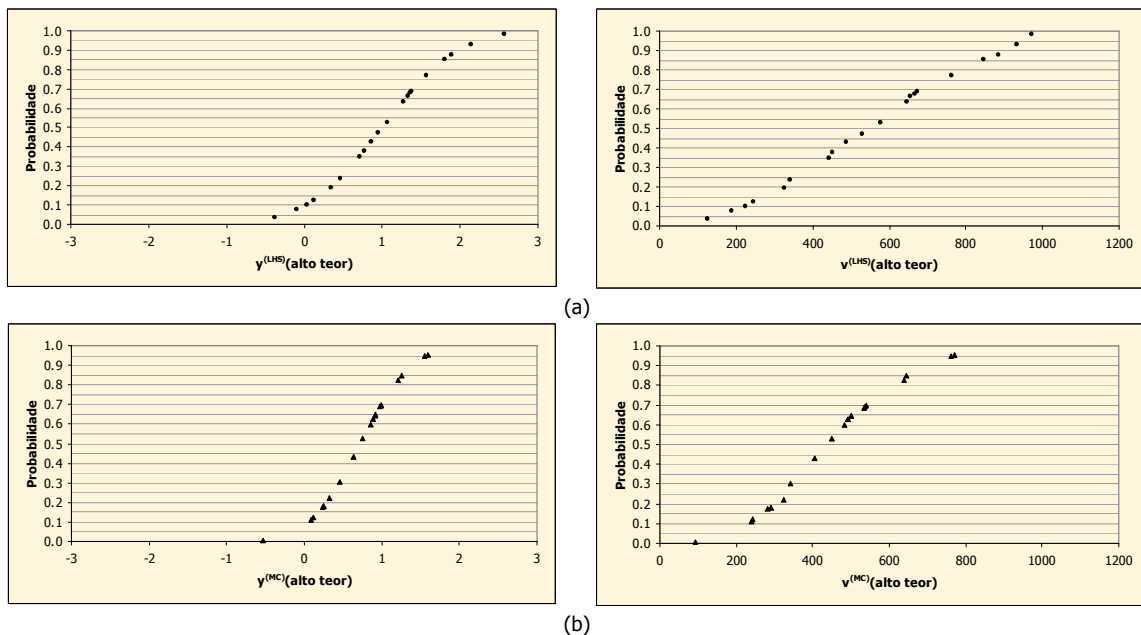


Figura 6.74: Modelos de incerteza (dclp) associados à localização inserida em uma zona de alto teor, de acordo com os algoritmos LHS clássico (a) e com o algoritmo original (b), nos espaços normal (esquerda) e original (direita).

Na Figura 6.74, as curvas obtidas por meio do algoritmo LHS clássico apresentam maior definição, em relação àquelas derivadas por meio do algoritmo original. Essa característica está associada à maior estratificação das dclp determinadas por meio desse algoritmo (as linhas horizontais auxiliam na identificação dos estratos). Os modelos de incerteza obtidos por meio do algoritmo LHS clássico apresentam 80 % de estratificação. Isso significa que, dentre as 20 classes, 16 foram amostradas. Em relação ao algoritmo original, 14 classes foram amostradas (70 %). A Tabela 6.36 apresenta as classes que deixaram de ser amostradas, segundo os dois algoritmos.

Tabela 6.36: Classes não amostradas da Figura 6.74, em termos de probabilidade, segundo os dois algoritmos (LHS e MC).

Classes não amostradas	
LHS	MC
[0,25; 0,30)	[0,05; 0,10)
[0,55; 0,60)	[0,25; 0,30)
[0,70; 0,75)	[0,35; 0,40)
[0,80; 0,85)	[0,45; 0,50)
	[0,70; 0,75)
	[0,75; 0,80)

Nas curvas da Figura 6.74 (a), as classes [0,10; 0,15) e [0,85; 0,90) foram duplamente amostradas, enquanto que a classe [0,65; 0,70) foi triplamente amostrada. As dclp da Figura 6.74 (b), apresentaram duas classes duplamente amostradas: [0,10; 0,15) e [0,15; 0,20); e duas triplamente amostradas: [0,60; 0,65) e [0,65; 0,70).

As dclp do espaço normal, da Figura 6.74, estão deslocadas para a direita, em relação ao zero. Esse comportamento é justificado pelos dados que contribuíram para a construção desses modelos de incerteza. A localização simulada está inserida em uma área de altos teores. Por isso, nesse caso, a dclp obtida por meio do algoritmo LHS clássico está centrada em 1,0100, enquanto que aquela proveniente do algoritmo original, está centrada em 0,7268.

A Figura 6.75 exibe as distribuições experimentais de probabilidade Gaussiana, utilizadas na determinação dos valores simulados (do espaço normal) das dclp da Figura 6.74.

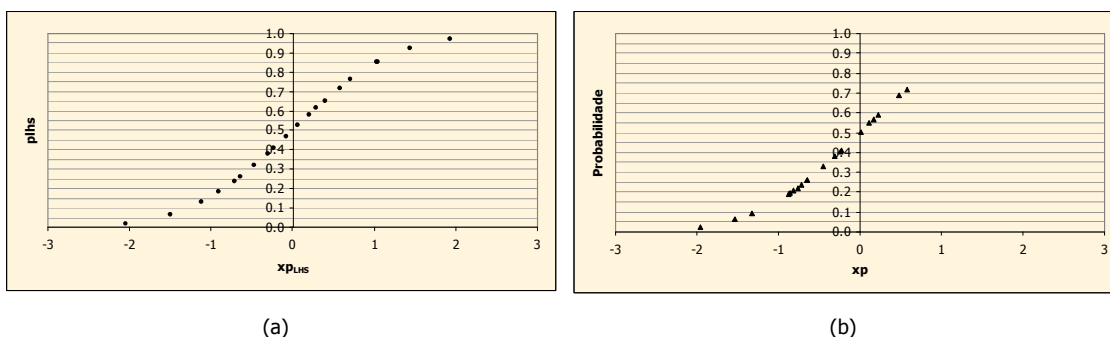


Figura 6.75: Distribuições experimentais de probabilidade Gaussiana, de acordo com o algoritmo LHS clássico (a) e original (b), utilizadas para a determinação dos valores simulados da Figura 6.74.

As figuras anteriores compararam as dclp obtidas por meio dos algoritmos LHS clássico e original. A Figura 6.76 mostra as dclp determinadas por meio dos algoritmos LHS cíclico clássico e LHS determinístico, associadas à localização selecionada na zona de médio teor.

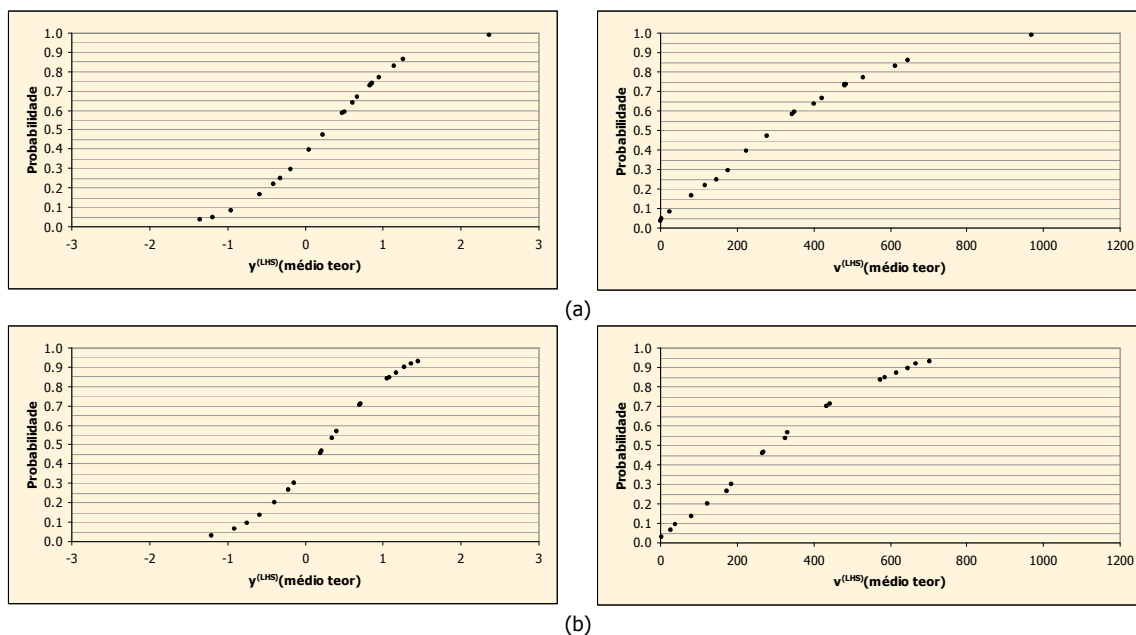


Figura 6.76: Modelos de incerteza (dclp) associados à localização inserida em uma zona de médio teor, de acordo com os algoritmos LHS cíclico clássico (a) e LHS determinístico (b), nos espaços normal (esquerda) e original (direita).

Como mostra a Figura 6.76, as dclp resultantes da aplicação dos algoritmos LHS cíclico clássico e LHS determinístico não são estratificadas. Nos modelos de incerteza relativos ao algoritmo LHS cíclico clássico, cinco classes deixaram de ser amostradas. Nas dclp referentes ao algoritmo LHS determinístico, sete classes não foram amostradas. A Tabela 6.37 apresenta essas classes.

Tabela 6.37: Classes não amostradas da Figura 6.76, em termos de probabilidade, segundo os algoritmos LHS cíclico clássico (20 realizações, $M = 4$) e LHS determinístico ($M = 20$).

Classes não amostradas	
LHS cíclico clássico	LHS determinístico
[0,10; 0,15)	[0,20; 0,25)
[0,30; 0,35)	[0,35; 0,40)
[0,40; 0,45)	[0,40; 0,45)
[0,50; 0,55)	[0,60; 0,65)
[0,90; 0,95)	[0,65; 0,70)
	[0,75; 0,80)
	[0,95; 1,00]

6.4.7. Algoritmo LHS com caminho determinístico

Como descrito na seção 5.5.1.3 (p. 111), a aplicação do algoritmo LHS determinístico pode gerar dclp mais estratificadas. Isso porque, nesse algoritmo, todas as realizações seguem a mesma seqüência de estimativa (caminho). Assim, os valores assumidos por σ_{KS} e y^*_{KS} nas localizações u_j ($j = 1, \dots, N$) têm mais chance de serem aproximadamente constantes, em cada simulação. Dessa forma, o peso dos valores Gaussianos x_{plhs} , estratificados, será maior. Conseqüentemente, a chance de que as dclp resultantes sejam estratificadas aumenta.

A Tabela 6.38 apresenta os parâmetros de simulação e os valores simulados (no espaço normal), resultantes da aplicação do algoritmo LHS clássico com caminho determinístico. Essa tabela retrata a situação na qual $M = 20$, associada à localização selecionada de baixo teor [u_1 ($j = 1$)].

Tabela 6.38: Parâmetros de simulação para a localização $j = 1$ da malha ($L = M = 20$), por meio da aplicação do algoritmo LHS clássico com caminho determinístico.

Realização (i)	Classe [$\pi_i(u_1)$]	plhs	$x_{plhs}^{(i)}(u_1)$	$y^*_{KS}(u_1)$	$\sigma_{KS}(u_1)$	$y^{(i)}(u_1)$
1	17	0,8141	0,8932	-0,7738	0,6314	-0,2098
2	13	0,6328	0,3394	-1,3595	0,6314	-1,1452
3	7	0,3402	-0,4118	-0,7961	0,6314	-1,0561
4	20	0,9881	2,2610	-0,8598	0,6314	0,5678
5	14	0,6728	0,4478	-0,5743	0,6314	-0,2916
6	18	0,8720	1,1358	-0,0879	0,6314	0,6292
7	11	0,5465	0,1169	0,2574	0,6314	0,3312
8	8	0,3549	-0,3722	-0,4654	0,6314	-0,7004
9	2	0,0722	-1,4595	-0,2038	0,6314	-1,1253
10	1	0,0106	-2,3048	-2,2224	0,6314	-3,6777
11	4	0,1656	-0,9718	-2,3717	0,6314	-2,9853
12	9	0,4192	-0,2040	-1,3941	0,6314	-1,5229
13	16	0,7946	0,8224	-1,7864	0,6314	-1,2671
14	12	0,5968	0,2450	-0,9221	0,6314	-0,7674
15	6	0,2531	-0,6647	-1,1883	0,6314	-1,608
16	5	0,2222	-0,7646	-1,674	0,6314	-2,1568
17	19	0,9253	1,4413	-0,421	0,6314	0,489
18	10	0,4886	-0,0286	-0,2379	0,6314	-0,256
19	3	0,1219	-1,1657	-0,1496	0,6314	-0,8856
20	15	0,7038	0,5353	-2,4609	0,6314	-2,1229

A Tabela 6.38 mostra que, ao longo das 20 realizações, os valores associados a σ_{KS} permanecem constantes. Isso comprova que a posição espacial dos dados na vizinhança de u_1 , utilizados para a obtenção dos pesos de KS, é a mesma nas 20 realizações (portanto, mesmo caminho). Entretanto, o valor das estimativas (y^*_{KS}) varia bastante ao longo das realizações ($\sigma = 0,78$). Isso é justificado pelo uso de oito dados previamente simulados nas matrizes de KS. Esses dados, são os únicos valores variáveis nas matrizes de KS.

A Figura 6.77 mostra que as dclp resultantes da aplicação do algoritmo LHS clássico com caminho determinístico não são estratificadas. Em relação à estratificação, os resultados são piores do que aqueles obtidos por meio do algoritmo LHS clássico. Nesse caso, sete classes não foram amostradas, quatro classes foram duplamente amostradas e uma classe foi triplamente amostrada. A Tabela 6.39 apresenta essas classes.

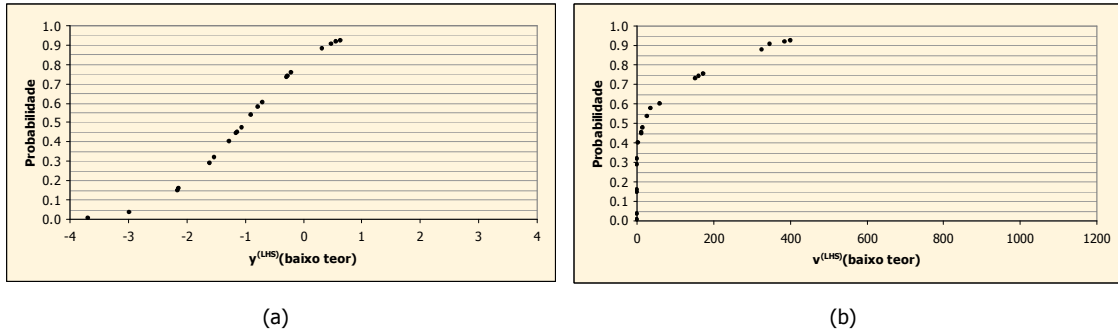


Figura 6.77: Modelos de incerteza (dclp) associados à localização u_1 , de acordo com o algoritmo LHS clássico com caminho determinístico, nos espaços normal (a) e original (b).

Tabela 6.39: Classes não amostradas, duplamente amostradas e triplamente amostradas da Figura 6.77, em termos de probabilidade, segundo o algoritmo LHS clássico com caminho determinístico.

Classes não amostradas	Classes duplamente amostradas	Classe triplamente amostrada
[0,05; 0,10)	[0,15; 0,20)	[0,90; 0,95)
[0,10; 0,15)	[0,40; 0,45)	
[0,20; 0,25)	[0,45; 0,50)	
[0,35; 0,40)	[0,70; 0,75)	
[0,65; 0,70)		
[0,80; 0,85)		
[0,95; 1,00]		

A diminuição do uso de dados previamente simulados não apresentou melhorias em relação à estratificação das dclp. Em um dos testes, foram utilizados apenas dois dados previamente simulados e as dclp continuaram não estratificadas, como as da Figura 6.77. Os modelos de incerteza só serão completamente estratificados em uma única situação: quando não forem utilizados dados previamente simulados. A Figura 6.78 ilustra essa situação. Os parâmetros de construção dessas dclp são mesmos da Figura 6.77, exceto o número de dados previamente simulados.

Portanto, o algoritmo LHS com caminho determinístico não apresentou desempenho satisfatório, em relação à estratificação das dclp, a partir da mesma estratégia de busca utilizada nos testes anteriores. Assim, não foram realizados testes de caracterização do espaço de incerteza da VA V , como aqueles realizados com as outras variações do algoritmo LHS. Os algoritmos (MC, LHS e suas variações) foram comparados a partir de suas aplicações com os mesmos parâmetros de entrada.

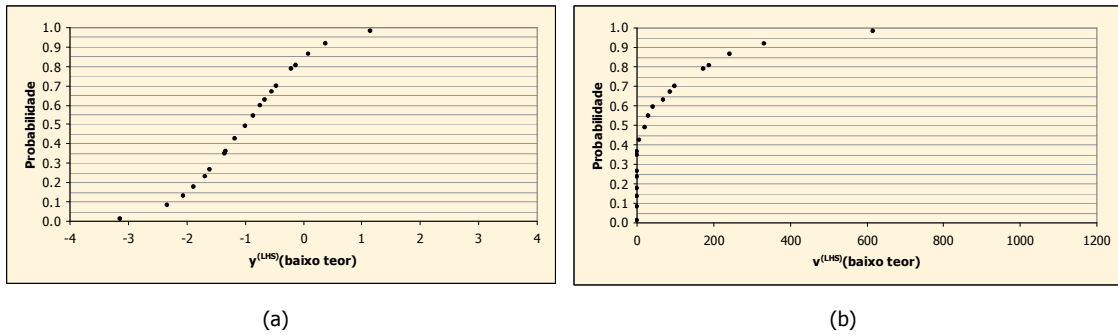


Figura 6.78: Modelos de incerteza (dclp) associados à localização u_1 , de acordo com o algoritmo LHS clássico com caminho determinístico, nos espaços normal (a) e original (b).

6.4.8. Comparativos entre os algoritmos de simulação (MC x LHS)

Nessa seção, os resultados obtidos por meio dos algoritmos de simulação envolvendo a técnica LHS, são comparados com aqueles obtidos pela aplicação do algoritmo original (MC). Esses algoritmos foram aplicados de acordo com os mesmos parâmetros de entrada, inclusive a semente aleatória. Esse parâmetro é especialmente importante, pois determina a ordem seqüencial de simulação, ao longo das realizações. A Tabela 6.40 mostra o caminho aleatório de simulação, seguido por uma realização qualquer, de acordo com os algoritmos original, LHS clássico e LHS determinístico. Nessa tabela, *index* representa as localizações u_j ($j = 1, \dots, 78000$), aleatoriamente ordenadas, representando o caminho de simulação. Essa tabela comprova que os algoritmos foram executados respeitando as mesmas seqüências de simulação.

Tabela 6.40: Ordem seqüencial de simulação (caminho) de acordo com os algoritmos original, LHS clássico e LHS determinístico.

Seqüência	MC	LHS Clássico	LHS determinístico
	Index	Index	Index
1	62164	62164	62164
2	49828	49828	49828
3	6028	6028	6028
4	37204	37204	37204
5	43468	43468	43468
.	.	.	.
.	.	.	.
.	.	.	.
77995	65800	65800	65800
77996	16429	16429	16429
77997	42858	42858	42858
77998	55542	55542	55542
77999	47100	47100	47100
78000	4733	4733	4733

As figuras seguintes (Figura 6.79 e 6.80) comparam o desempenho dos algoritmos, em relação à caracterização do espaço de incerteza da VA V. Essas figuras ilustram gráficos que mostram a relação entre a média e o CV, dos conjuntos constituídos pelas médias das realizações, em função das realizações. Entretanto, é conveniente lembrar que, nos gráficos referentes ao algoritmo LHS, os pontos são provenientes de processos de simulação diferentes.

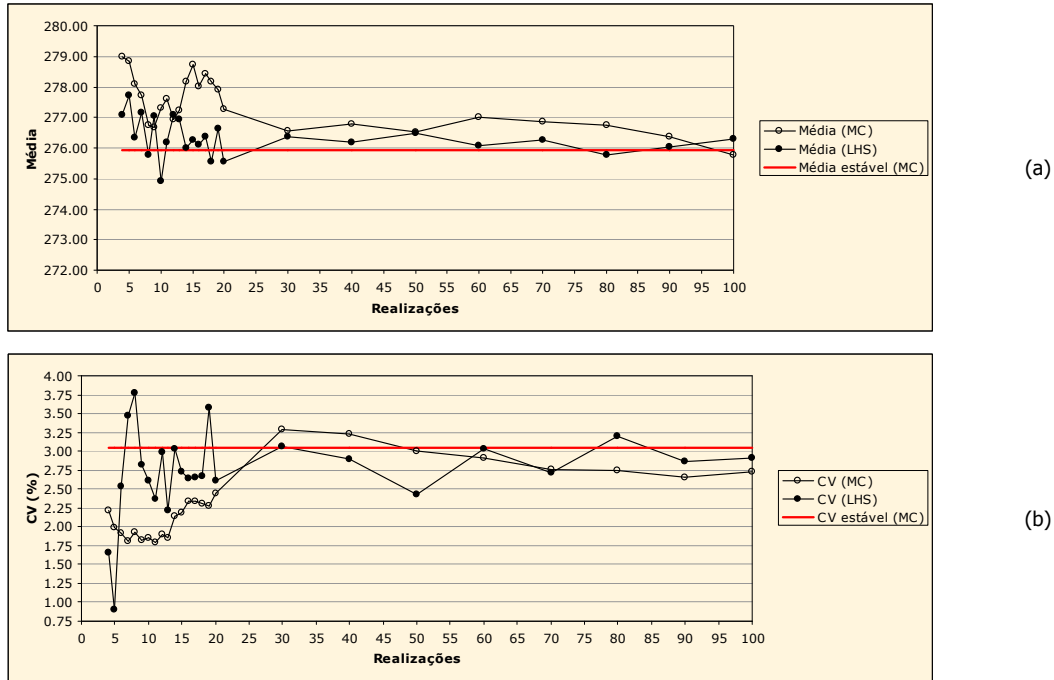


Figura 6.79: Média (a) e CV (b) da distribuição das médias das realizações, de acordo com os algoritmos original e LHS clássico. A linha horizontal vermelha indica os respectivos valores considerados estáveis.

Na Figura 6.79, é verificado que o desempenho do algoritmo LHS clássico, em relação à média [Figura 6.79 (a)], é superior ao apresentado pelo algoritmo original. Isso se observa, principalmente, a partir da utilização de 14 classes de discretização ($M = 14$). Em relação ao CV, o desempenho do algoritmo LHS clássico é superior em duas situações:

- i. com a utilização de nove a 30 classes; e
- ii. com a utilização de mais do que 60 classes.

A partir dessa análise, considerando os dois parâmetros de avaliação, pode-se afirmar que o algoritmo LHS clássico leva vantagem, em relação ao algoritmo original, nessas duas situações. Essa vantagem é principalmente destacada, quando são executadas poucas realizações ($9 < L < 20$). Nesses casos, o desempenho apresentado pelo algoritmo LHS clássico não é o ideal, mas é nitidamente superior ao apresentado pelo algoritmo original.

Os gráficos da Figura 6.80 mostram que, para menos do que 30 realizações, a aplicação do algoritmo LHS determinístico não proporciona resultados melhores do que os apresentados pelo

algoritmo original. Entretanto, é preciso lembrar que os processos de simulação, com a utilização de menos do que 50 classes de discretização, não foram validados. Com isso, verifica-se que o desempenho desse algoritmo é superior em relação ao algoritmo original, quando se utiliza 60 ou mais classes de discretização.

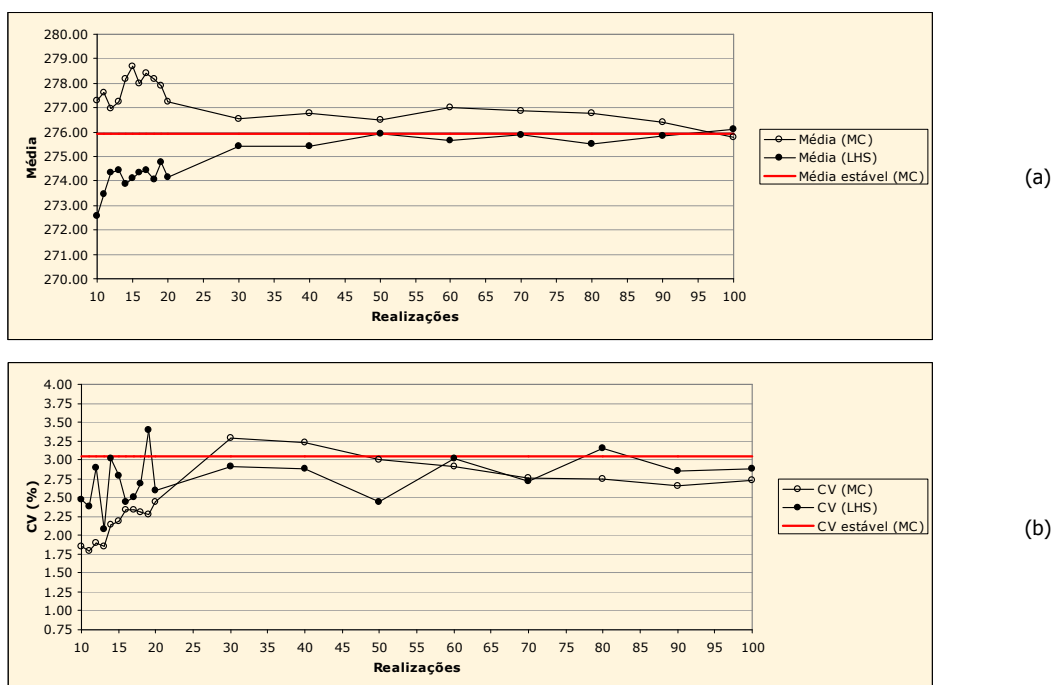


Figura 6.80: Média (a) e CV (b) da distribuição das médias das realizações, de acordo com os algoritmos original e LHS determinístico. A linha horizontal vermelha indica os respectivos valores considerados estáveis.

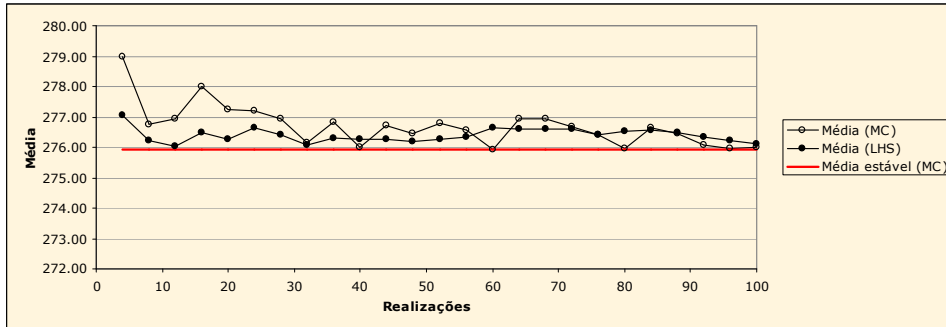
A seguir, o desempenho do algoritmo LHS cíclico clássico é comparado com o do algoritmo original. Para isso, foram selecionados os processos de simulação que proporcionaram os resultados mais próximos daqueles considerados estáveis. Foram selecionados quatro modelos de discretização da distribuição de probabilidade Gaussiana (G):

- i. $M = 4$;
- ii. $M = 8$;
- iii. $M = 10$; e
- iv. $M = 15$

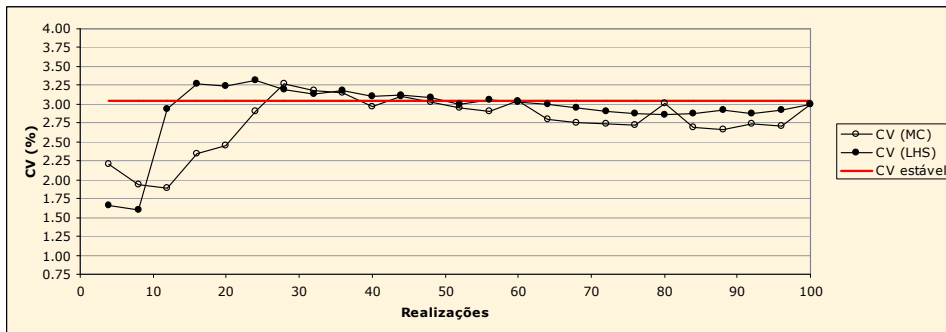
O procedimento de comparação foi realizado da mesma maneira da qual foram comparados os desempenhos dos algoritmos LHS clássico e LHS determinístico, em relação ao algoritmo original. As Figuras 6.81, 6.82, 6.83, 6.84 e 6.85 exibem os gráficos da média e do CV (das distribuições das médias das realizações) em função das realizações.

A Figura 6.81 mostra os gráficos referentes ao caso no qual foram utilizados quatro classes de discretização. Nessa figura, é observado que, em geral, os resultados obtidos por meio do algoritmo

LHS cíclico clássico são melhores do que aqueles apresentados por meio do algoritmo original. Esse comportamento é evidenciado, principalmente, em relação ao CV. A partir do sétimo ciclo, os valores assumidos por esse parâmetro variam muito pouco em relação ao CV considerado estável. Considerando os dois parâmetros avaliados, somente nos dois primeiros ciclos o desempenho do algoritmo LHS cíclico clássico (com $M = 4$) é inferior ao do algoritmo original.



(a)



(b)

Figura 6.81: Média (a) e CV (b) da distribuição das médias das realizações, de acordo com os algoritmos original e LHS cíclico clássico (25 ciclos de quatro realizações). A linha horizontal vermelha indica os respectivos valores considerados estáveis.

As Figuras 6.82 e 6.83 mostram os gráficos resultantes da aplicação do algoritmo LHS cíclico clássico, com a utilização de oito classes discretização. Na Figura 6.82 é verificado que o desempenho do algoritmo LHS cíclico clássico é superior ao apresentado pelo algoritmo original.

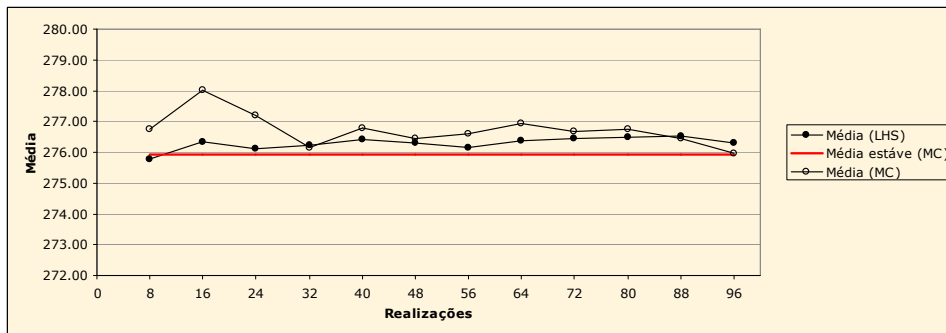


Figura 6.82: Média da distribuição das médias das realizações, de acordo com os algoritmos original e LHS cíclico clássico (12 ciclos de oito realizações). A linha horizontal vermelha indica os respectivos valores considerados estáveis.

Em relação ao CV (Figura 6.83), somente nos três primeiros ciclos, o desempenho do algoritmo LHS cíclico clássico não é satisfatório. Nos demais ciclos (exceto pelo quatro), os resultados apresentados pelo algoritmo LHS cíclico clássico, não só são melhores do que os apresentados pelo algoritmo original, com também são quase idênticos àqueles considerados estáveis. Em um contexto geral (considerando a média e o CV), pode-se afirmar que o desempenho do algoritmo LHS cíclico clássico, com esse grau de discretização ($M = 8$), é muito satisfatório a partir do quinto ciclo.

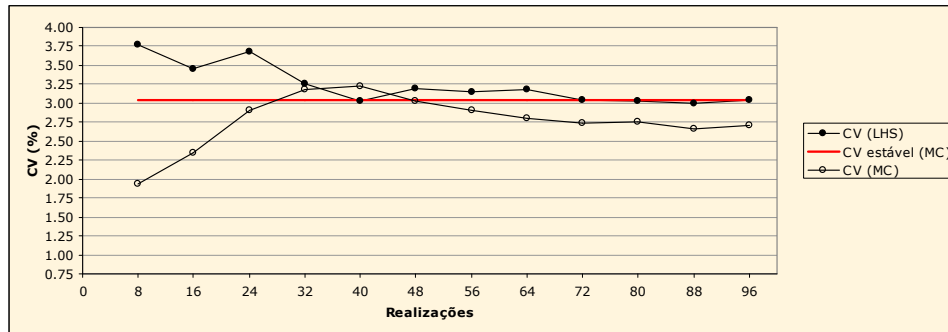
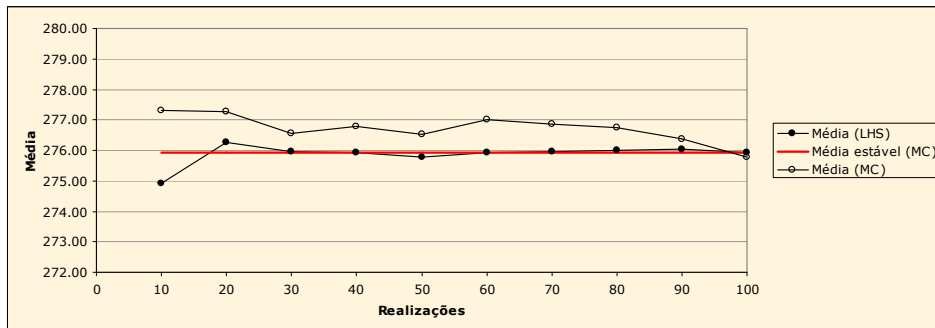
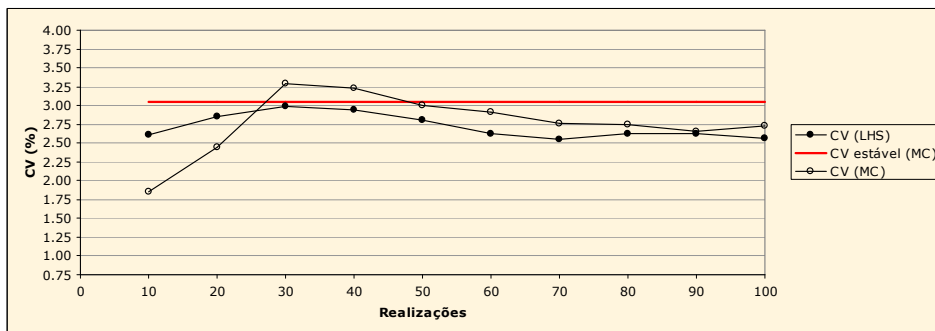


Figura 6.83: CV da distribuição das médias das realizações, de acordo com os algoritmos original e LHS cíclico clássico (12 ciclos de oito realizações). A linha horizontal vermelha indica os respectivos valores considerados estáveis.

A Figura 6.84 apresenta os gráficos referentes ao caso no qual foram empregadas de 10 classes de discretização.



(a)



(b)

Figura 6.84: Média (a) e CV (b) da distribuição das médias das realizações, de acordo com os algoritmos original e LHS cíclico clássico (10 ciclos de 10 realizações). A linha horizontal vermelha indica os respectivos valores considerados estáveis.

Na Figura 6.84 (a), é observado que, em todos os ciclos, os resultados obtidos por meio do algoritmo LHS cíclico clássico são melhores do que os obtidos por meio do algoritmo original. Inclusive, a partir do segundo ciclo, os resultados são muito satisfatórios, pois a média assume valores quase idênticos ao valor considerado estável. Em relação ao CV, os resultados não são tão bons quanto aos relativos à média. Somente nos quatro primeiros ciclos, o desempenho do algoritmo LHS cíclico clássico é superior ao apresentado pelo algoritmo original.

A Figura 6.85 apresenta os resultados da aplicação dos algoritmos LHS cíclico clássico, com $M = 15$, e original (MC), em função das realizações ou processos de simulação. Os resultados de ambos os parâmetros avaliados revelam a superioridade do algoritmo LHS cíclico clássico. Em relação à média, todos os ciclos apresentaram desempenhos superiores. Em relação ao CV, somente o quarto ciclo apresenta desempenho levemente inferior ao mostrado pelo CV das médias das 60 realizações, obtidas por meio do algoritmo original. Considerando os dois parâmetros avaliados, conclui-se que, para esse estudo de caso, esse modelo de discretização pode ser utilizado com sucesso.

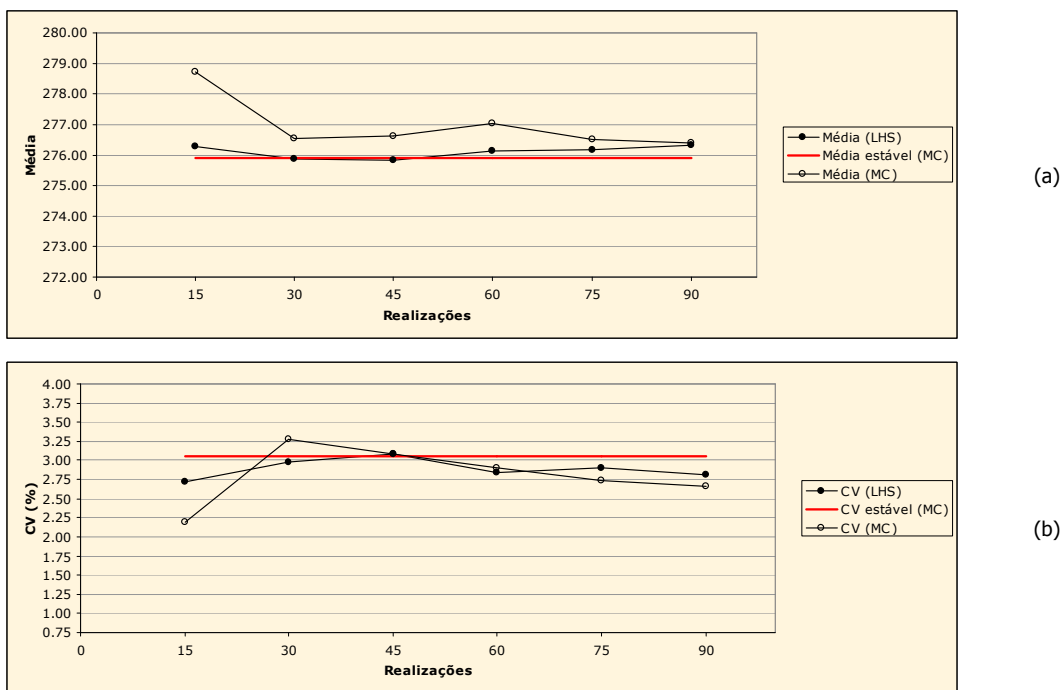
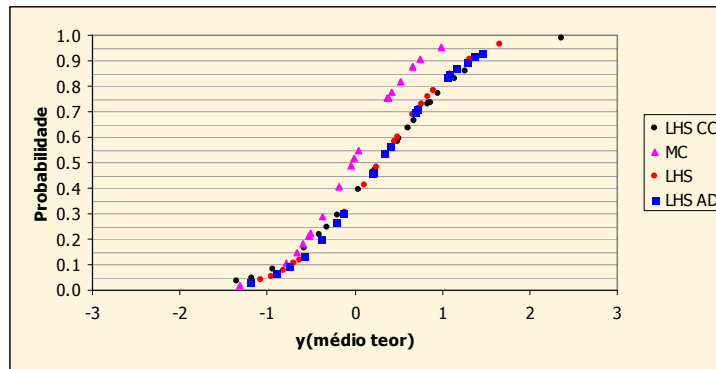


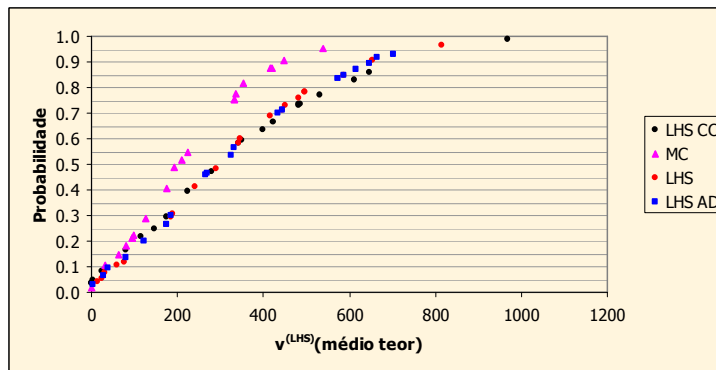
Figura 6.85: Média (a) e CV (b) da distribuição das médias das realizações, de acordo com os algoritmos original e LHS cíclico clássico (seis ciclos de 15 realizações). A linha horizontal vermelha indica os respectivos valores considerados estáveis.

As figuras anteriores compararam os algoritmos, original e LHS, com enfoque na caracterização da incerteza global. A Figura 6.86 compara esses algoritmos em relação à incerteza local. O desempenho dos algoritmos é comparado na localização selecionada de médio teor, quando são executadas 20 realizações. Essa figura mostra como o algoritmo original subestima a incerteza local. O uso do algoritmo LHS clássico e suas variações (LHS cíclico clássico e LHS determinístico)

resulta em dclp locais mais espalhadas (maior inclinação das curvas). Esse comportamento caracteriza uma maior incerteza local. Por exemplo, as dclp da Figura 6.86 (b) mostram que, com o uso do algoritmo original, existe cerca de 40 % de chance de o teor, nessa localização, ser inferior ou igual a 176 ppm. As dclp, derivadas por meio do algoritmo LHS (e suas variações), mostram que existe 30 % de chance. As diferenças são ainda maiores quando se considera o trecho superior das dclp. Por exemplo, as dclp mostram 10 % (MC) contra 27 % (LHS) de chance de a concentração ser superior a 450 ppm.



(a)



(b)

Figura 6.86: Modelos de incerteza (dclp) associados à localização inserida em uma zona de médio teor, de acordo com os algoritmos LHS clássico, LHS cíclico clássico, LHS determinístico e original (MC), nos espaços normal (a) e original (b).

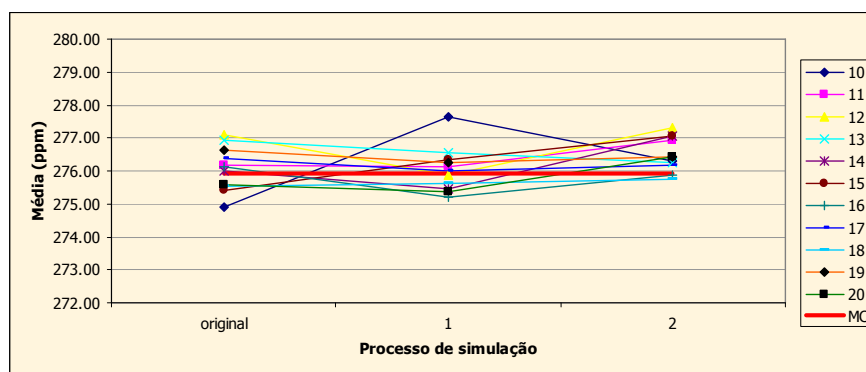
6.4.8.1. Reprodutibilidade dos resultados

Na seção 6.4.8, foi descrito que o uso do algoritmo LHS clássico é vantajoso, em relação ao algoritmo original, principalmente, quando são executadas poucas realizações ($9 < L < 20$). Isto é, o emprego desse algoritmo é, especialmente, vantajoso nos casos nos quais a distribuição de probabilidade Gaussiana (G) é dividida em até 20 estratos. Nesses casos, como a amplitude das classes (em termos de probabilidade) é relativamente grande (5 %, para $M = 20$), é conveniente verificar a reprodutibilidade dos resultados obtidos para esses modelos de discretização. Para isso, foram executados dois processos de simulação adicionais (com os modelos de discretização de 10 a 20 classes), com a utilização de sementes aleatórias diferentes. A Tabela 6.41 apresenta as médias e

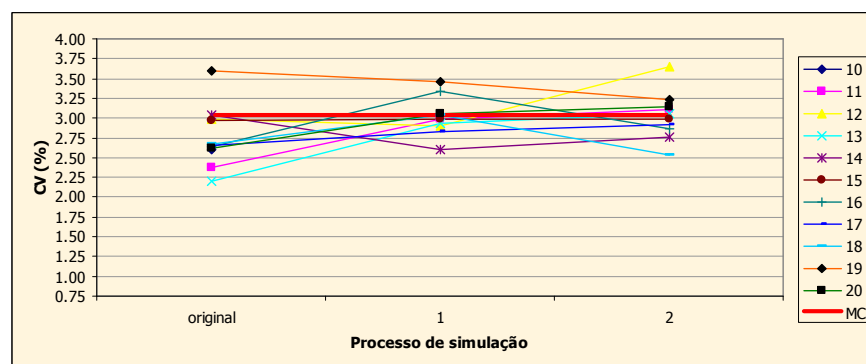
os CV (das distribuições das médias), segundo os modelos de discretização, para os dois processos de simulação adicionais, além do original (apresentado nas seções 6.4.2 e 6.4.8). Nessa tabela, também são apresentados os valores médios (das médias e dos CV) dos três processos de simulação. A Figura 6.87 representa graficamente a Tabela 6.41.

Tabela 6.41: Média e CV, da distribuição das médias das realizações, segundo os processos de simulação (LHS clássico) e de acordo os modelos de discretização.

Modelo de discretização (M)	Média				CV			
	processo de simulação				processo de simulação			
	original	1	2	média	original	1	2	média
10	274,9139	277,6459	276,2823	276,281	2,6078	3,0050	3,1444	2,919
11	276,1785	276,1221	276,9324	276,411	2,3731	2,9780	3,1071	2,819
12	277,0920	275,8827	277,2957	276,757	2,9888	2,8995	3,6503	3,180
13	276,9378	276,5278	276,249	276,572	2,2088	2,9274	3,0537	2,730
14	276,0066	275,4403	277,0574	276,168	3,0343	2,6078	2,7558	2,799
15	275,4085	276,3348	277,0632	276,269	2,9741	2,9826	2,9797	2,979
16	276,1263	275,2045	275,8947	275,742	2,6371	3,3297	2,8690	2,945
17	276,3579	275,9842	276,162	276,168	2,6494	2,8269	2,9137	2,797
18	275,5571	275,6155	275,7543	275,642	2,6733	3,0459	2,5283	2,749
19	276,6521	276,2663	276,4288	276,449	3,6000	3,4497	3,2328	3,428
20	275,5644	275,371	276,4054	275,780	2,6123	3,0557	3,1466	2,938



(a)



(b)

Figura 6.87: Média (a) e CV (b), da distribuição das médias das realizações, segundo os processos de simulação (LHS clássico) e de acordo os modelos de discretização.

Uma análise sobre a Tabela 6.41 e a Figura 6.87, revela que, em relação à média, o modelo de discretização de 20 classes mostra valores mais próximos do considerado estável, ao longo dos três processos de simulação. Em relação ao CV, esse comportamento é verificado no modelo de discretização de 15 classes. A Figura 6.88 auxilia na realização dessa análise. Essa figura mostra os erros médios, em relação à média e ao CV, de acordo com os modelos de discretização.

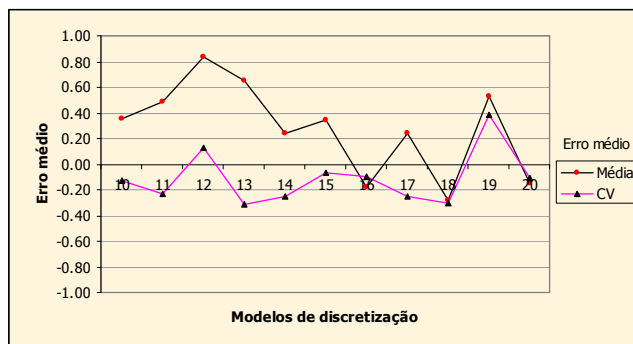


Figura 6.88: Erro médio apresentado pela média e CV, da distribuição das médias das realizações, segundo os modelos de discretização.

Considerando a média e o CV concomitantemente, o modelo de discretização de 20 classes leva vantagem em relação ao de 15 classes. Isso pode ser evidenciado na Figura 6.89. Essa figura ilustra um gráfico que mostra a soma dos erros médios absolutos da média e do CV, em função dos modelos de discretização. Entretanto, deve ser ressaltado que esses resultados retratam a média dos três processos de simulação. Considerando a regularidade ao longo dos três processos de simulação, o desempenho do modelo de discretização de 15 classes é melhor, como mostra a Figura 6.87.

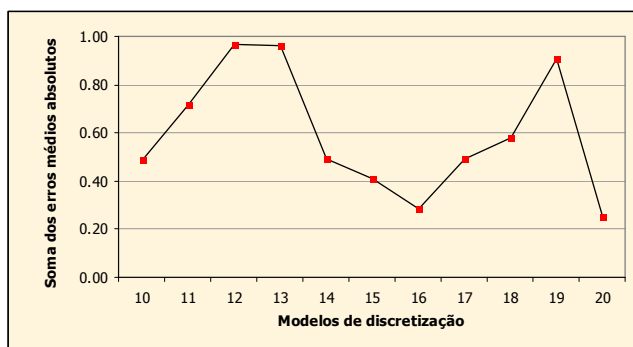


Figura 6.89: Soma dos erros médios absolutos da média e do CV, da distribuição das médias das realizações, segundo os modelos de discretização.

6.4.8.2. Caracterização da incerteza x curvas de parametrização – Walker Lake

Nas seções anteriores foi mostrado que o uso do algoritmo de simulação original (MC), a partir de um número reduzido de realizações, pode subestimar o espaço de incerteza da VA estudada. A utilização do algoritmo modificado, com a técnica LHS, pode minimizar esse efeito.

Na prática, o espaço de variabilidade de uma VA, caracterizado por meio da aplicação desses algoritmos, pode ser comparado por meio de uma análise das curvas de parametrização do depósito mineral. A Figura 6.90 ilustra a curva tonelagem x *cutoff* para 14 cenários equiprováveis, gerados por meio dos dois algoritmos.

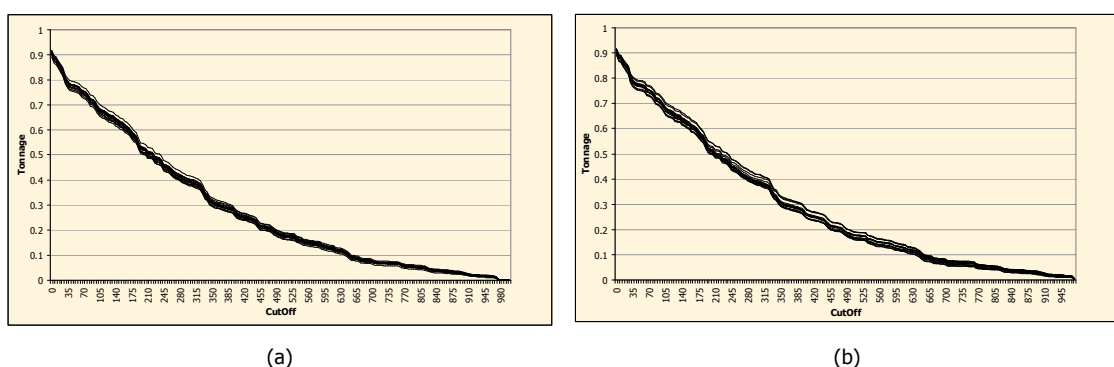


Figura 6.90: Curvas de parametrização (t x *cutoff*) produzidas a partir de 14 realizações geradas por meio dos algoritmos original (a) e LHS clássico (b).

Embora os dois conjuntos de curvas pareçam muito similares, a amplitude de variação das curvas produzidas a partir das 14 realizações geradas por meio do algoritmo LHS clássico é maior, como mostra a Figura 6.91. Essa figura apresenta os envelopes das 14 curvas, segundo os dois algoritmos. Esse comportamento evidencia que o uso do algoritmo LHS clássico faz com que as análises de sensibilidade decorrentes sejam realizadas de maneira mais realista.

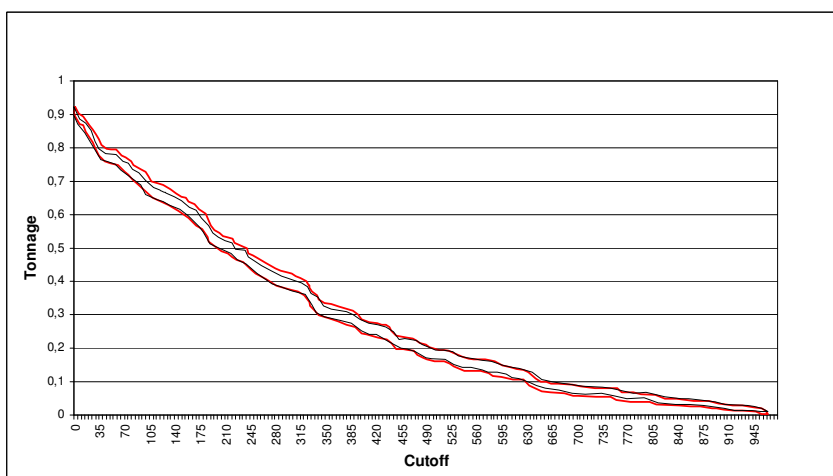


Figura 6.91: Amplitude das curvas de parametrização (t x *cutoff*), segundo os algoritmos original (linha preta) e LHS clássico (linha vermelha).

Como complemento da Figura 6.91, a Figura 6.92 apresenta o CV dos 14 resultados de tonelage, segundo cada *cutoff*, para cada um dos algoritmos. A análise dessa figura reflete as conclusões anteriores, evidenciando que a variabilidade em torno dos possíveis valores de tonelage é maior quando se utiliza o algoritmo modificado.

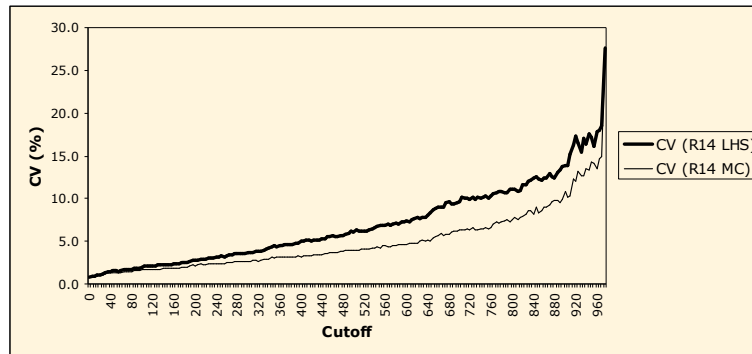


Figura 6.92: CV da distribuição dos 14 valores de tonelage, de acordo com os algoritmos original e LHS clássico.

As Figuras 6.93, 6.94 e 6.95 descrevem o comportamento do CV, tal como na Figura 6.92, para 15, 20 e 60 realizações, respectivamente.

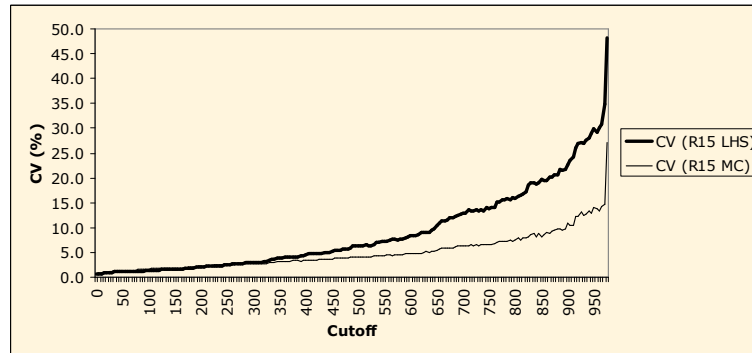


Figura 6.93: CV da distribuição de 15 valores de tonelage, de acordo com os algoritmos original e LHS clássico.

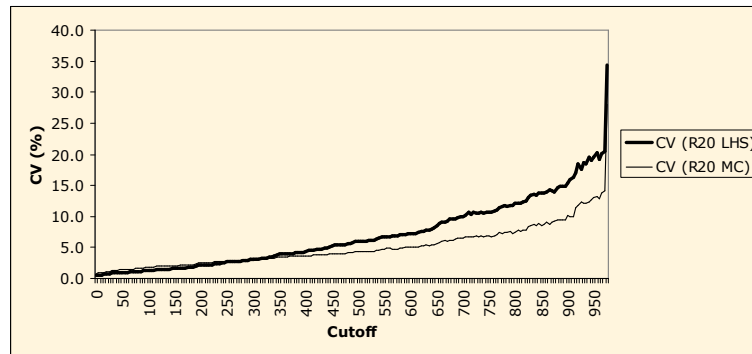


Figura 6.94: CV da distribuição de 20 valores de tonelage, de acordo com os algoritmos original e LHS clássico.

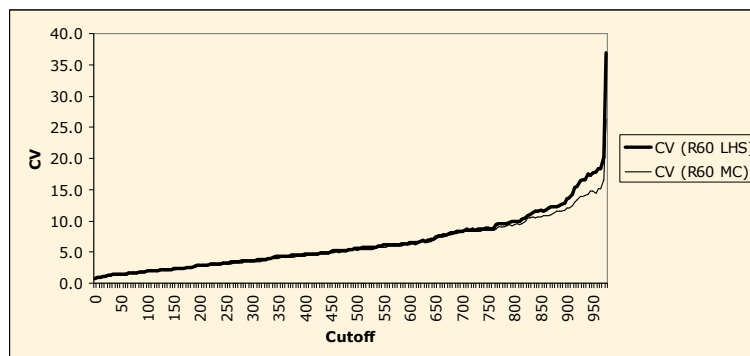


Figura 6.95: CV da distribuição de 60 valores de tonelagem, de acordo com os algoritmos original e LHS clássico.

A Figura 6.95 mostra que, para 60 realizações, o espaço de variabilidade caracterizado por ambos algoritmos é muito similar. Somente para *cutoffs* mais elevados (acima de 750 ppm), o espaço de incerteza estimado por meio do algoritmo LHS clássico é maior. Entretanto, quando é executado um número reduzido de realizações, a diferença entre os algoritmos é destacada. As Figuras 6.92, 6.93 e 6.94 mostram que a diferença de amplitude do espaço de incerteza é significativa para *cutoffs* acima de 100 ppm.

O comportamento do CV da distribuição dos 200 possíveis valores de tonelagem, obtidos pelo algoritmo original, é quase idêntico ao CV (LHS) apresentado na Figura 6.92. Isso porque, o espaço de variabilidade caracterizado por meio de 14 realizações, obtidas pelo algoritmo LHS clássico, é muito similar ao caracterizado por meio de 200 realizações obtidas pelo algoritmo original.

6.4.8.3. Caracterização da incerteza x curvas de parametrização – Fe

Embora esse capítulo seja destinado, preferencialmente, ao banco de dados Walker Lake, essa seção apresenta os resultados obtidos pela aplicação dos algoritmos MC e LHS clássico para um estudo de caso real. Foi considerado um corpo de minério de Fe (itabirito), sob concessão da Companhia Vale do Rio Doce (CVRD) (cujo uso foi gentilmente autorizado). Principalmente, esse estudo de caso se distingue do anterior (Walker Lake) por ser real, e em três dimensões (3 D).

O uso de simulação geoestatística, em minério de Fe, é muito recente. Um dos primeiros estudos, foi realizado em agosto 2004, pela CVRD, com o auxílio do Laboratório de Pesquisa Mineral e Planejamento Mineiro (LPM) do departamento de Engenharia de Minas da UFRGS. Nessa ocasião, foram simulados diversos corpos de minério, e por motivos operacionais, foram executadas, no máximo, 20 realizações. O estudo de caso, apresentado nessa seção, é referente a apenas um desses corpos de minério.

Os algoritmos MC e LHS foram utilizados para simular 54.910 blocos, de dimensão 25 x 25 x 10 m (6.250 m³), definidos em um domínio irregular determinado pela equipe de geólogos da CVRD. De acordo com os dois algoritmos, foram executados três processos de simulação; com 20, 14 e 10

realizações. Portanto, para a aplicação do algoritmo LHS clássico, foram utilizados três modelos de discretização: $M = 20$, $M = 14$ e $M = 10$.

A Figura 6.96 ilustra as curvas de parametrização, tonelagem x *cutoff*, resultantes da aplicação dos dois algoritmos, para o caso no qual foram executadas 20 realizações. A amplitude de variação dessas 20 curvas parece muito similar, entretanto, o gráfico da Figura 6.97 revela as diferenças. Essa figura apresenta o CV dos valores de tonelagem, segundo cada *cutoff*. É verificado que o espaço de incerteza, determinado por meio do algoritmo LHS clássico, é mais amplo. Esse comportamento se evidencia a partir do *cutoff* de 34 %.

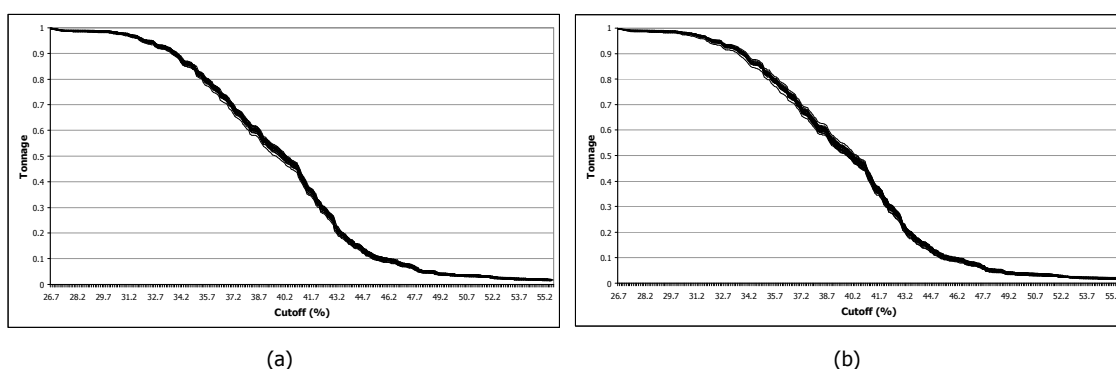


Figura 6.96: Curvas de parametrização (t x *cutoff*) produzidas a partir de 20 realizações geradas por meio dos algoritmos original (a) e LHS clássico (b) – estudo de caso: Fe.

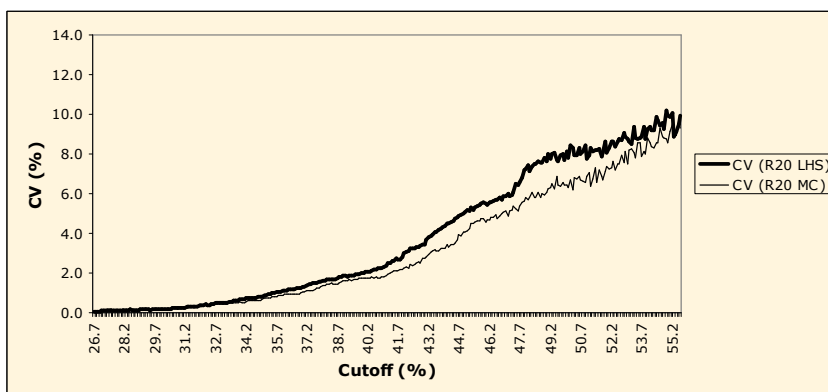


Figura 6.97: CV da distribuição dos 20 valores de tonelagem, de acordo com os algoritmos original e LHS clássico - estudo de caso: Fe.

A diferença entre os dois algoritmos se torna mais evidente quando são executadas menos do que 20 realizações. As Figuras 6.98 e 6.99 mostram os CV dos valores de tonelagem para os casos nos quais foram executadas 14 e 10 realizações, respectivamente, segundo cada algoritmo. Em ambos processos de simulação, o espaço de incerteza, caracterizado por meio do algoritmo LHS clássico, é mais amplo. A diferença de amplitude ocorre a partir do *cutoff* de 34 %, aumentando progressivamente com o *cutoff*.

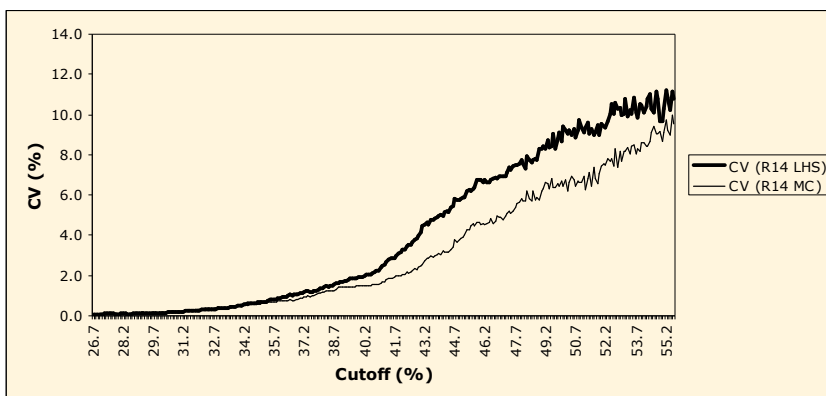


Figura 6.98: CV da distribuição dos 14 valores de tonelagem, de acordo com os algoritmos original e LHS clássico - estudo de caso: Fe.

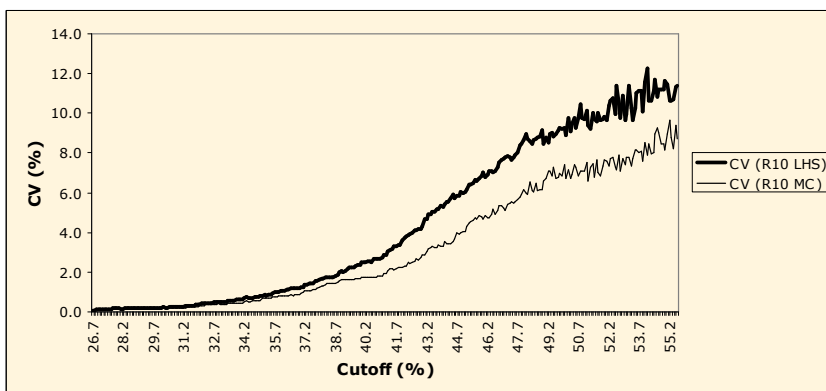


Figura 6.99: CV da distribuição dos 10 valores de tonelagem, de acordo com os algoritmos original e LHS clássico - estudo de caso: Fe.

Na prática, essa diferença na amplitude de variação das curvas de parametrização, pode resultar em análises de risco otimistas. Isso porque, a cada *cutoff*, são considerados subdomínios, contidos nos domínios reais de variação de tonelagem. Por exemplo, a Tabela 6.42 apresenta os domínios de variação de tonelagem, além da diferença entre eles, para os três processos de simulação, segundo os dois algoritmos (MC e LHS). Para isso, a densidade média do itabirito foi considerada como sendo 3,31 t/m³. Assim, a tonelagem de cada bloco foi considerada constante: 20.687,50 t. O *cutoff* utilizado foi o mesmo empregado pela CVRD para esse minério: 48 %.

Tabela 6.42: Domínio de variação de tonelagem e diferença de amplitude dos domínios, para os três processos de simulação, segundo os algoritmos MC e LHS clássico, de acordo com o *cutoff* de 48 %.

Realizações	MC	LHS clássico	Diferença de amplitude (M t)
	Tonelagem (domínio de variação)	Tonelagem (domínio de variação)	
10	[53.389.679,38; 62.477.284,38]	[47.709.926,25; 65.885.136,25]	9,088
14	[53.389.679,38; 62.477.284,38]	[53.276.084,31; 69.292.988,13]	6,929
20	[52.253.728,75; 62.477.284,38]	[49.981.827,50; 64.749.185,63]	4,544

A partir dos resultados exibidos na Tabela 6.42, é possível realizar uma análise quantitativa das Figuras 6.97, 6.98 e 6.99 para o *cutoff* de 48 %. Ou seja, é verificado que, para esse *cutoff*, a amplitude de variação é progressivamente maior com a diminuição do número de realizações. Ainda, que as diferenças, embora pequenas, podem ser significativas. Por exemplo, para 14 realizações, uma diferença de 1,84 %, no CV dos valores de tonelagem (para o *cutoff* de 48 %), resulta em uma diferença de quase 7 M t. Nesse caso, utilizando o algoritmo tradicional (MC), para cobrir essa diferença, seria necessário considerar mais 334,95 blocos. Ao limite superior do domínio de variação, seria necessário adicionar 329,46 blocos. Ao limite inferior, seria necessário subtrair a tonelagem equivalente a 5,49 blocos.

As Figuras 6.100, 6.101 e 6.102 apresentam os histogramas cumulativos, dos valores de tonelagem, a partir do *cutoff* de 48 %, para os três processos de simulação, segundo os dois algoritmos. Nessas figuras, é possível verificar graficamente a amplitude de variação de tonelagem, segundo cada algoritmo. Também é evidente a diferença de frequência dos resultados. Por exemplo, na Figura 6.100, enquanto que os resultados obtidos por meio do algoritmo LHS clássico mostram que 40 % dos valores de tonelagem são inferiores ou iguais a 57,50 M t, os resultados obtidos por meio do algoritmo MC mostram que 35 % dos valores são inferiores ou iguais a 57,50 M t.

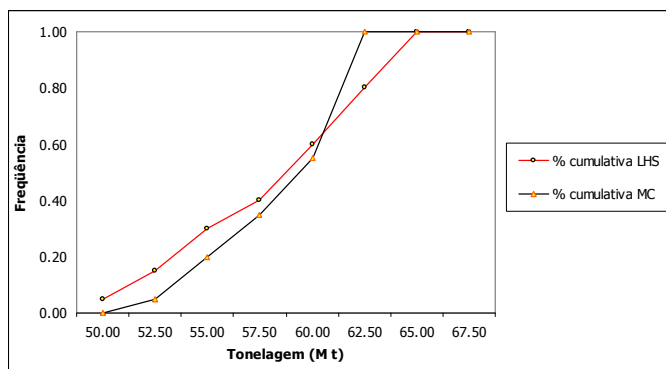


Figura 6.100: Histograma cumulativo dos 20 valores de tonelagem, segundo os algoritmos MC e LHS clássico, de acordo com o *cutoff* de 48 % - estudo de caso: Fe.

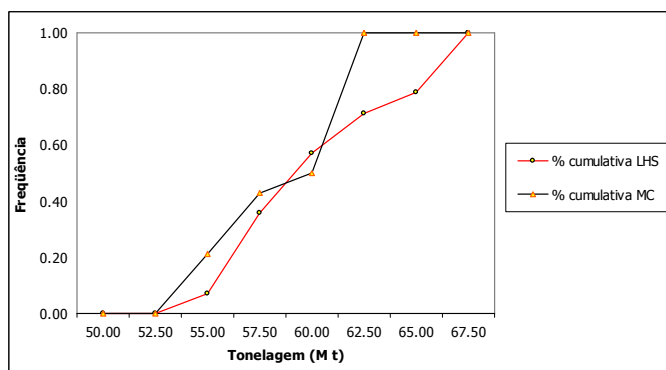


Figura 6.101: Histograma cumulativo dos 14 valores de tonelagem, segundo os algoritmos MC e LHS clássico, de acordo com o *cutoff* de 48 % - estudo de caso: Fe.

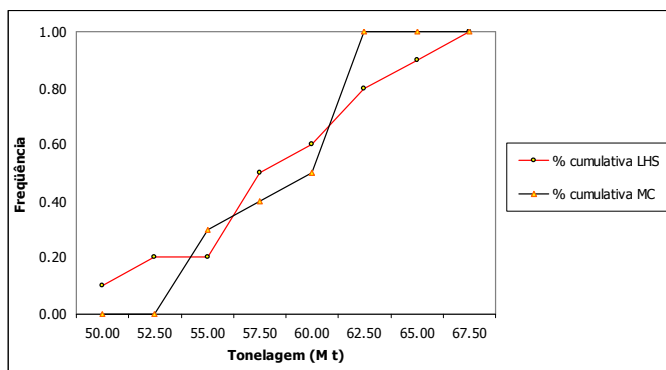


Figura 6.102: Histograma cumulativo dos 10 valores de tonelagem, segundo os algoritmos MC e LHS clássico, de acordo com o *cutoff* de 48 % - estudo de caso: Fe.

6.5. Comentários

Esse capítulo mostrou, na prática, a aplicabilidade e a funcionalidade do algoritmo proposto. Foram apresentadas as vantagens do uso do algoritmo LHS (e suas variações), em relação ao algoritmo original (MC); mostrando que o mesmo pode caracterizar adequadamente o espaço de variabilidade da VA estudada em menos realizações. Essa tarefa foi realizada por meio de um estudo de caso, a partir de uma porção de um banco de dados conhecido: Walker Lake.

O capítulo iniciou com a apresentação do banco de dados, seguido da execução de 200 cenários equiprováveis das concentrações da VA V no espaço. Em seguida, o algoritmo LHS clássico e suas variações foram aplicados, mostrando, em cada etapa, suas validações práticas e funcionais. Em cada etapa, também, foram destacados seus respectivos desempenhos, em função dos modelos de discretização, sendo nem sempre satisfatórios. Os resultados obtidos por meio dos algoritmos LHS foram comparados àqueles obtidos por meio do algoritmo original. Também, foi destacado o comportamento dos algoritmos LHS em relação à estratificação dos modelos de locais de incerteza. Verificações de reprodutibilidade dos resultados também foram realizadas. O capítulo foi encerrado mostrando impacto do uso dos algoritmos propostos sobre as curvas $t \times \textit{cutoff}$, segundo os modelos de discretização. Finalmente, foi apresentado as curvas $t \times \textit{cutoff}$, referentes a um caso real (minério de Fe), obtidas por meio dos algoritmos MC e LHS clássico. Esse exemplo mostrou, quantitativamente, a aplicabilidade e as vantagens do uso do algoritmo modificado.

O Capítulo 7 apresenta as conclusões e as recomendações para trabalhos futuros.

Capítulo 7

Conclusões e recomendações

Esse último capítulo expõe as conclusões, em relação ao algoritmo de simulação proposto, formuladas com base nos seis capítulos precedentes. Inicialmente, é realizado um breve resumo dos capítulos anteriores. São relacionados, principalmente, os resultados teóricos e práticos, apresentados nos Capítulos 5 e 6, com a meta e os objetivos expostos no Capítulo 1. Essa Tese é encerrada com a apresentação de algumas recomendações para futuras pesquisas.

7.1. Resumo dos primeiros capítulos

Embora os resultados efetivos dessa Tese tenham sido apresentados nos Capítulos 5 e 6, o desenvolvimento dos quatro primeiros capítulos também foi muito importante. Nesses capítulos, foi apresentada uma revisão bibliográfica consistente, e pertinente, para que o objetivo fosse alcançado.

O Capítulo 1 apresentou o problema abordado nessa Tese, bem como as metas e os objetivos propostos. Foi realizada uma revisão do estado da arte da geoestatística, enfocando, principalmente, os desenvolvimentos em torno dos métodos de simulação estocástica, enfatizando as últimas pesquisas nesse tema. O uso de simulação geoestatística na mineração também foi abordado, mostrando porque o emprego desses métodos tem se tornado cada vez mais freqüente e popular.

No Capítulo 2, foi realizada uma revisão dos conceitos básicos de geoestatística linear, essenciais para o entendimento e o desenvolvimento matemático dos seus métodos associados. Foram revisados os conceitos relacionados às variáveis regionalizadas e às funções aleatórias, bem como aqueles, envolvendo as hipóteses intrínsecas e estacionárias. Nesse capítulo, também foram revisadas as propriedades elementares da função de covariância espacial e algumas propriedades estatísticas de combinações lineares. Principalmente, foram apresentadas a definição do conceito de combinações lineares autorizadas (CLA) e suas relações com as hipóteses intrínsecas e estacionárias, fundamentais para a aplicação das equações de (co)krigagem.

O Capítulo 3 foi o responsável pela revisão dos conceitos de simulação e condicionamento, apresentando o algoritmo seqüencial de simulação condicional. Foi revisado, também, o conceito do método de Monte Carlo, empregado nesse algoritmo. O Capítulo 3 apresentou dois métodos,

paramétrico e não-paramétrico, para a construção dos modelos de incerteza, ou distribuições condicionais locais de probabilidade (dclp), que constituem o alicerce dos métodos condicionais seqüenciais.

O Capítulo 4 apresentou as utilidades dos modelos de incerteza (dclp). Foi mostrado como medir a magnitude de incerteza associada diretamente aos atributos medidos ou a partir de funções de transferência, de acordo com a finalidade desejada (incerteza local x incerteza espacial). Para isso, foi apresentado o conceito de função de transferência, mostrando como a incerteza pode se propagar por meio dessas funções. Principalmente, esse capítulo apresentou uma ferramenta muito eficaz para a construção dos modelos de incerteza das variáveis resposta das FTs, e também, para a análise da propagação da incerteza. Essa ferramenta é o método *Latin Hypercube Sampling*, LHS. Seu estudo despertou a curiosidade e motivação para a concretização dessa Tese.

7.2. Meta e objetivo x resultados teóricos e práticos

A meta dessa Tese era desenvolver um algoritmo de simulação geoestatística eficiente, capaz de atingir os objetivos propostos com agilidade, exigindo menos realizações (e como consequência, tempo de processamento) e espaço virtual do que os algoritmos tradicionais. Para alcançar essa meta, foi proposta uma modificação na estratégia de tiragens aleatórias, empregada nos algoritmos seqüenciais de simulação estocástica. Essa modificação envolveu a substituição do método de Monte Carlo (MC) pelo de *Latin Hypercube Sampling* (LHS). Os resultados teóricos, apresentados no Capítulo 5, e os práticos, apresentados no Capítulo 6, indicam que o objetivo foi alcançado com sucesso.

No Capítulo 1, foram identificadas cinco etapas a serem desenvolvidas, para que o objetivo proposto fosse alcançado.

- i. Desenvolver o algoritmo de tiragens aleatórias, partindo-se da idéia e do embasamento teórico da técnica LHS.
- ii. Verificar a funcionalidade teórica do algoritmo desenvolvido na etapa (i).
- iii. Inserir o algoritmo desenvolvido na etapa (i) nos algoritmos de simulação seqüencial estocástica.
- iv. Verificar a eficiência e funcionalidade prática dos algoritmos modificados.
- v. Quantificar o impacto provocado pela alteração dos algoritmos modificados em problemas reais de mineração.

As três primeiras etapas foram desenvolvidas no Capítulo 5. Nesse capítulo, além de apresentar o método LHS formalmente, o mesmo foi adaptado às VAs espaciais. Essa adaptação resultou na elaboração de algoritmos de tiragens aleatórias, baseados na técnica LHS, capazes de substituir o método de Monte Carlo, tradicionalmente utilizado nos métodos seqüenciais de simulação estocástica. Além da versão baseada na técnica LHS clássica, foram criados mais três algoritmos: LHS

determinístico, LHS cíclico e LHS com caminho determinístico. Sendo que, a característica a cíclica, pode ser aplicada nos algoritmos LHS clássico, LHS determinístico e LHS com caminho determinístico. Foi mostrado, em detalhes, como o método LHS e suas variações podem ser implementados nas rotinas de simulação seqüencial Gaussiana escritas em Fortran.

No Capítulo 5, também foram realizadas verificações teóricas, que comprovaram a eficiência de construção dos algoritmos. Essas verificações, também mostraram as vantagens teóricas do uso da técnica LHS, em substituição ao método de Monte Carlo. Nesses testes, a qualidade de amostragem da distribuição de probabilidade Gaussiana, por meio da técnica LHS, foi visivelmente superior àquela proporcionada pelo método de Monte Carlo. Ainda, os resultados obtidos por meio da técnica LHS, se mostraram mais robustos quanto à reprodutibilidade dos resultados.

As duas últimas etapas foram desenvolvidas no Capítulo 6. Para isso, foram utilizados dois bancos de dados: um em 2 D, exaustivamente amostrado e de domínio público; e outro em 3 D, de propriedade da CVRD. Entretanto, o Capítulo 6 foi desenvolvido, principalmente, sobre o banco de dados em 2 D (Walker Lake). Os algoritmos, tradicional e modificados, foram aplicados nesse banco de dados. Quanto aos resultados, as conclusões são as seguintes:

- i. Considerando o algoritmo LHS clássico:
 - Conclui-se que, o mesmo pode ser utilizado com sucesso para caracterizar o espaço de incerteza de VAs, em menos realizações do que seriam necessárias por meio do algoritmo original. Além de produzir processos de simulação válidos (nas quais os momentos condicionantes são, em média, reproduzidos), o uso do algoritmo LHS clássico sempre proporcionou resultados melhores do aqueles obtidos por meio do algoritmo original (Figura 6.79).
 - O uso desse algoritmo é principalmente recomendado quando, por motivos operacionais, for necessário executar um número reduzido de realizações. Nesses casos, a aplicação do algoritmo LHS clássico produz resultados (em termos de caracterização da incerteza) muito superiores àqueles produzidos por meio do algoritmo original. Para essas situações, recomenda-se o uso de, no mínimo, nove classes de discretização.
 - A característica cíclica pode ser utilizada com sucesso; porém, não em todos modelos de discretização. Os resultados foram satisfatórios quando foram empregados os seguintes modelos de discretização (Figuras 6.81, 6.82, 6.83, 6.84 e 6.85):
 - $M = 4$;
 - $M = 8$;
 - $M = 10$; e
 - $M = 15$.

-
- Além de caracterizar a incerteza global adequadamente e em menos realizações, do que seriam necessárias com o uso do algoritmo original, localmente, o uso do algoritmo LHS clássico produziu modelos de incerteza mais amplos (Figura 6.86).
 - Para poucas realizações, utilizando modelos de discretização de até 20 classes, os resultados são, em geral, reproduzidos. Isso indica que os processos de simulação podem ser adequadamente reproduzidos, a partir da utilização de diferentes sementes aleatórias. Os modelos de discretização de 15 e 20 classes são os que apresentaram melhores desempenhos.
 - A função de transferência utilizada, representada pelas curvas $t \times cutoff$, indica que, para um mesmo número de realizações, o desempenho do algoritmo LHS clássico é superior àquele apresentado pelo algoritmo original. Essa constatação foi corroborada pelos gráficos das Figuras 6.92, 6.93, 6.94 e 6.95. Essas figuras mostraram que o espaço de incerteza, caracterizado por meio do algoritmo original, sempre é subestimado.
- ii. Considerando o algoritmo LHS determinístico:
- Conclui-se que os resultados foram satisfatórios a partir do uso de um modelo de discretização de 60 ou mais classes ($M \geq 60$); quando esse algoritmo se torna muito similar ao LHS clássico. Embora os resultados, em relação à média (média da distribuição das médias das realizações), tenham sido satisfatórios a partir da utilização de um modelo de discretização de 30 classes ($M = 30$), em relação ao CV, os resultados mostraram uma oscilação significativa. Esse comportamento ocorreu quando foi empregado o modelo de discretização de 50 classes ($M = 50$) (Figuras 6.66 e 6.80). De qualquer forma, os processos de simulação, executados com a utilização de modelos de discretização com menos do que 50 classes, não foram validados. Nesses casos, os momentos de 1ª e 2ª ordem das realizações não reproduziram os respectivos momentos condicionantes. Por isso, o uso desse algoritmo só é recomendado com o emprego de 60, ou mais classes, de discretização.
 - A característica cíclica não pode ser utilizada com sucesso nesse algoritmo. Isso é determinado por duas razões: (i) os processos de simulação, executados com o uso de menos do que 50 classes, não foram validados; (ii) os resultados sempre foram subestimados, tanto em relação à média, quanto em relação ao CV. Isso indica que o espaço de incerteza, determinado por meio do algoritmo LHS cíclico determinístico, sempre é subestimado.
- iii. Considerando o algoritmo LHS com caminho determinístico:
- Esse algoritmo foi criado com o intuito de aumentar a estratificação dos modelos de incerteza locais (dclp) resultantes. Contudo, a aplicação desse algoritmo não

proporcionou resultados melhores do que aqueles obtidos por meio do algoritmo LHS clássico (Figura 6.77).

- Esse comportamento é devido ao número de dados previamente simulados, utilizados na construção das matrizes de krigagem. Esses dados são os únicos valores variáveis ao longo das realizações. A estratégia de busca empregada contava com a utilização de oito dados dessa natureza. A utilização de menos dados previamente simulados não proporcionou melhores resultados.
- Como mostrou a Figura 6.78, as dclp só serão perfeitamente estratificadas quando não forem utilizados dados previamente simulados. Entretanto, o uso desses dados é extremamente importante para a validação dos processos de simulação. Como visto no Capítulo 3 (seção 3.3.1, p. 24), esses dados são os responsáveis pelo condicionamento das realizações. Portanto, a não utilização de dados previamente simulados gera realizações que não reproduzem os momentos condicionantes.
- Para que as dclp sejam completamente estratificadas, utilizando dados previamente simulados, a ordem de tiragem das classes também deveria ser determinística. Isto é, para cada localização de uma mesma realização, os valores Gaussianos $x_{p_{LHS}}$ deveriam ser tirados das mesmas classes. Contudo, é preciso verificar se esse procedimento não agregaria uma tendenciosidade no sistema, e, se haveria reprodução dos momentos condicionantes.

Para o estudo de caso real, utilizando o banco de dados de minério de Fe (itabirito, de propriedade da CVRD), foram comparados os resultados obtidos por meio da aplicação do algoritmo tradicional (MC) e do algoritmo envolvendo a técnica LHS clássica. Em relação a esse estudo de caso, as conclusões foram as seguintes:

- Embora não apresentado, ambos algoritmos produziram realizações que, em média, reproduziram os momentos condicionantes.
- Para o mesmo número de realizações, o espaço de incerteza, caracterizado por meio do algoritmo LHS clássico, foi mais amplo do que aquele caracterizado por meio do algoritmo original (MC). Essa constatação é evidente na análise do Tabela 6.42 e dos gráficos das Figuras 6.97, 6.98 e 6.99. Considerando o teor de corte aplicado na CVRD, essa diferença na caracterização do espaço de incerteza, embora pequena, representa vários milhões de toneladas.

7.3. Recomendações

Para trabalhos futuros, que podem vir a dar seqüência à essa pesquisa, destacam-se as seguintes recomendações e sugestões:

-
- i. Realização de um estudo de caso completo, utilizando outro banco de dados de caráter real. A aplicação do algoritmo LHS, e de suas variações, a partir de outra situação mineira real, validaria por completo o emprego dos algoritmos envolvendo a técnica LHS.
 - ii. Utilização de outras funções de transferência, como por exemplo o VPL (valor presente líquido). Ao relacionar um parâmetro de projeto mineiro de caráter financeiro, o impacto provocado pela substituição do algoritmo original, realçaria as vantagens do uso do algoritmo LHS (e de suas variações).
 - iii. Pesquisar o modelo de discretização mais adequado para cada situação. A escolha do número M de classes [que discretiza a distribuição $G(y)$] ainda é um assunto a ser estudado. Nessa Tese, o modelo de discretização mais adequado foi determinado de maneira empírica.
 - iv. Pesquisar variações no algoritmo LHS, visando aumentar o grau de discretização dos modelos de incerteza resultantes. O algoritmo LHS com caminho determinístico, com tiragem de classes determinística, descrita na seção anterior, pode proporcionar esse efeito. Resta saber se os processos de simulação resultantes serão validados.
 - v. Estudar a possibilidade de, a partir do histograma original, privilegiar a tiragem de certas classes, consideradas como primordiais para as análises decorrentes.
 - vi. Alimentar as funções de transferência com as realizações P10, P50 e P90; visando posteriores análises de risco, a partir do algoritmo modificado.
 - vii. Aplicar a técnica LHS no algoritmo de simulação seqüencial dos indicadores (ssi).
 - viii. Quantificar o impacto provocado pela implantação da técnica LHS no algoritmo ssi.

Referências Bibliográficas

- Alabert, F. G., 1987a. The practice of fast conditional simulations through the LU decomposition of the covariance matrix. *Mathematical Geology*, vol. 19, no. 5, pp. 369-386.
- Alabert, F. G., 1987b. *Stochastic Imaging of Spatial Distributions Using Hard and Soft Information*. M.Sc. thesis, Stanford University, Stanford, California, 197p.
- Anderson, T. W., 1958. *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, New York, 374p.
- Anderson, T. W., 1984. *Multivariate Analysis*. Wiley. (Wiley series in probability and statistics: applied probability and statistics section), New York, 675p.
- Anscombe, F. J., 1959. Quick analysis methods for random balance design. *Technometrics*, vol. 1, pp. 195-209.
- Armstrong, M.; Carignan, J., 1997. *Géostatistique Linéaire: Application au Domaine Minier*. Les presses de l'École des Mines, Paris, 115p.
- Armstrong, M.; Galli, A.; Le Loc'h, G.; Geoffroy, F.; Eschard, R., 2003. *Plurigaussian Simulations in Geosciences*. Springer-Verlag, New York, 160p.
- Arpat, B. G.; Caers, J., 2004. A Multiple-scale, Pattern-based Approach to Sequential Simulation. In: *7th International Geostatistics Congress*, September 26th - October 1st, 2004 Banff, Alberta, Canada, ref. no. 102.
- Bahadur, R. R., 1961. A representation of the joint distribution of responses to n dichotomous items. In: *Solomon, H., Studies in item analysis and prediction*, Stanford University Press, Stanford, California, 310p.
- Barnes, R.; Johnson, T., 1984. Positive kriging. In: Verly, G.; David, M.; Journel, A. G.; Marechal, A., (Eds), *Geostatistics for natural resources characterization*, Riedel, Dordrecht, pp. 231-244.

-
- Bartlett, M., S., 1966. *An Introduction to Stochastic Processes, With Special Reference to Methods and Applications*. Cambridge University Press, London, 362p.
- Bonato, F. S.; Costa, J. F. C. L.; Koppe, J. C., 2000a. Planejamento de lavra baseado na variabilidade de atributos geológicos. *REM – Revista da Escola de Minas de Ouro Preto*, Ouro Preto, vol. 53, no. 4, pp. 275-284.
- Bonato, F. S.; Costa, J. F. C. L.; Koppe, J. C., 2000b. Geostatistical simulation: a framework to forecast fluctuations during mining In: *Mine planning and equipment selection 2000, 2000, Athens. Proceedings of the 9th International Symposium on mine planning and equipment selection*. Rotterdam: Balkema, vol.1. pp. 167-172.
- Bonato, F. S., 2000c. *Planejamento de Lavra com Auxílio de Simulação Geoestatística*. Dissertação de Mestrado. Programa de Pós-Graduação em Engenharia de Minas, Metalúrgica e Materiais – PPGEM, Universidade Federal do Rio Grande do Sul – UFRGS, 87p.
- Brus, D. J.; Gruijter, J. J. de, 1993. Does kriging really give unbiased and minimum variance predictions of spatial means? *J. Soil Sci*, vol. 44, pp. 99-111.
- Budne, T. A., 1959. The application of random balance designs. *Technometrics*, vol. 1, pp. 139-155.
- Caers, J., 2000a. Adding local accuracy to direct sequential simulation. *Mathematical Geology*, vol. 32, no. 1, pp. 815-850.
- Caers, J., 2000b. Direct sequential indicator simulation. In: Kleingeld, W. J.; Krige, D. G. (Eds.), *Proceedings of the Geostatistics 2000*. Cape Town, South Africa, pp. 39-48.
- Chien, Y.-Ju., 1998. *A Geostatistical Approach to Sampling Design for Contaminated Site Characterization*. M. Sc. thesis, Stanford University, Stanford, California.
- Chien, Yi-Ju; Journel, A. G., 1998. An algorithm for determination of additional sampling locations or how to maximize potential additional information. In: *Proceedings of the NACOG 1998 Meeting (North American Council on Geostatistics)*.

-
- Chilès, J.-P., 1984. Simulation of a nickel deposit: Problems encountered and practical solutions. In: Verly, G., David, M., Journel, A. and Marechal, A. (Eds.) *Geostatistics for Natural Resources Characterization*. Reidel Publishing Company, Dordrecht, pp. 1015-1030.
- Chilès, J.-P., 1995. Quelques méthodes de simulation de fonctions aléatoires intrinsèques. *Cahiers de Géostatistique Ecole des Mines de Paris*, pp. 97-112.
- Chilès, J.-P.; Delfiner P., 1999. *Geostatistics :Modeling Spatial Uncertainty*. Wiley (Wiley series in probability and statistics: applied probability and statistics section), New York, 695p.
- Clark, I.; White, B., 1976. Geostatistical modelling of an orebody as an aid to mine planning. In: *14th APCOM Symposium*, Littleton, Colorado, pp 1004-1012.
- Costa, J. F., 1997. *Developments in Recoverable Reserves Estimation and Ore Body Modeling*. Ph.D. thesis, The University of Queensland, Australia, 333p.
- Costa, J. F., Koppe, J. C.; Zingano, A. C., 1997. Uncertainty Analysis of Stripping Ratio and Enhanced Coal Mine Planning. In: *Proceedings of the MineIT'97 - First International Conference on Information Technologies in the Minerals Industry (Internet)*, Athens, Greece.
- Dagbert, M., 1981. The simulation of space-dependent data in geology. In: Craig, R.G.; Labovitz, M. L. (Eds.) *Future Trends in Geomathematics*. Pion Limited, London, pp. 29-47.
- Damsleth, E.; Tjolsen, C. B.; Omre, H.; Haldorsen, H. H., 1992. A 2-stage stochastic model applied to a North-Sea reservoir. *Journal of Petroleum Technology*, vol. 44, no. 4, pp. 402-408.
- David, M., 1973. Tools for planning: Variances and conditional simulations. In: *11th APCOM Symposium*, University of Arizona, pp. D10-D23.
- David, M., 1988. *Handbook of Applied Advanced Geostatistical Ore Reserve Estimation*. Elsevier Scientific Publisher, Developments in Geomathematics 6, Amsterdam, 216p.
- David, M.; Dowd, P.; Korobov, S. 1974. Forecasting Departure from Planning in Open Pit design and Grade Control. In: *12th APCOM Symposium*, Colorado School of Mines, pp. F131-F153.
- Davis, B. M.; Greenes, K. A., 1983. Estimation using spatially distributed multivariate data: An example with coal quality. *Mathematical Geology*, vol. 15, no. 2, pp. 287-300.

-
- Davis, M., 1987. Production of conditional simulations via the LU decomposition of the covariance matrix. *Mathematical Geology*, vol. 19, no. 2, pp. 91-98.
- Dempster, A. P., 1960. Random allocation designs I: On general classes of estimation methods. *Annals of Mathematical Statistics*, vol. 31, pp. 885-905.
- Dempster, A. P., 1961. Random allocation designs II: Approximate theory for simple random allocation. *Annals of Mathematical Statistics*, vol. 32, pp. 387-405.
- Deraisme, J., De Fouquet, C. ; Fraise, H., 1984. Geostatistical orebody model for computer optimization of profits from different underground mining methods. In: *18th APCOM Symposium*, London, pp 583-590.
- Deutsch, C. V., 1992. *Annealing Techniques Applied to Reservoir Modeling and Integration of Geological and Engineering (Well Test) Data*. PhD thesis, Stanford University, Stanford, USA, 304p.
- Deutsch, C. V., 1997. Direct assessment of local accuracy and precision. In: Baafi, E. Y., Schofield, N. A. (Eds.), *Geostatistics Wollongong '96*. Kluwer Academic Publishing, Dordrecht, pp. 115-125.
- Deutsch, C. V., 2002. *Geostatistical Reservoir Modeling*. Oxford University Press, New York, 376p.
- Deutsch, C. V.; Journel, A.G., 1998. *GSLIB: Geostatistical Software Library and User's Guide*. Oxford University Press, New York, 2nd ed., 369p.
- Dimitrakopoulos, R., 1990. Conditional simulation of intrinsic random functions of order K. *Journal of International Association for Mathematical Geology*, vol. 22, pp. 361-380.
- Dimitrakopoulos, R., 1997. Conditional Simulations: Tools for Modelling Uncertainty in Open Pit Optimisation. In: *Proceedings of Optimising with Whittle 97*, Perth, 8-9 April., pp. 32-49.
- Dimitrakopoulos, R. 1998. Conditional simulation algorithms for modelling orebody uncertainty in open pit optimization. *International Journal of Surface Mining, Reclamation and Environment*, no. 12, pp. 173-179.

-
- Dimitrakopoulos, R., Farrelly, C. T.; Godoy, M. 2002. Moving forward from traditional optimisation: Grade uncertainty and risk effects in open pit mine design, *Transcript of the Institute of Mining and Metallurgy*, vol. Section A: Minerals Industry, no. 111, pp. A82-A89.
- Dimitrakopoulos, R., Luo, X. 2004. Generalized sequential Gaussian simulation on group size v and screen-effect approximations for large field simulations. *Mathematical Geology*, vol. 36, no. 5, pp. 567-591.
- Dowd, P. A., 1992. A review of recent developments in geostatistics. *Computer & Geosciences*, Vol. 17, no. 10, pp. 1481-1500.
- Dowd, P. A., 1994. Risk assessment in reserve estimation and open pit planning. *Transcript of the Institute of Mining and Metallurgy*, vol. Section A: Minerals Industry, no. 103, pp. A148-A154.
- Ehrenfeld, S.; Zacks, S., 1951. Randomization and factorial experiments. *Annals of Mathematical Statistics*, vol. 32, pp. 270-297.
- Ehrenfeld, S.; Zacks, S., 1967. Testing hypotheses in randomized factorial experiments. *Annals of Mathematical Statistics*, vol. 38, pp. 1494-1507.
- Euler, L., 1782. Recherches Sur une Espèce de Carrés Magiques. *Verh. Zeeuw. Gen. Weten. Vlissingen*, vol. 9, pp. 85-239.
- Fisher, R. A., 1925. *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh. Fourteenth edition also published by Hafner, New York (1973).
- Fisher, R. A., 1926. The arrangement of field experiments. *Journal of the Ministry of Agriculture*, vol. 33, pp. 503-513.
- Galli, A. G., Beucher, H., Le Loc'h, G., Doligez, B., 1994. The pros and cons of the truncated Gaussian method. In: Armstrong, M.; Dowd, P. A., (eds.) *Geostatistical simulations*, Kluwer, Dordrecht, pp. 217-233.
- Gandin, L. S., 1963. *Objective Analysis of Meteorological Fields*. Gidrometeorologicheskoe Izdatel'stvo (GIMIZ), Leningrad (translated by Israel Program for Scientific Translations, Jerusalem, 1965), 238p.

-
- Garcia, M.; Froidevaux, R., 1997. Application of geostatistics to 3-D modelling of contaminated sites: a case-study. In: Soares, A. (Ed.), *geoENV I – Geostatistics for Environmental Applications*. Kluwer Academic Publishing, Dordrecht, pp. 309–325.
- Glacken, I. M., 1996. *Change of Support by Direct Conditional Block Simulation*. M.Sc. thesis, Stanford University, Stanford, 124p.
- Godoy, M. C., 2002. *The Effective Management of Geological Risk in Long-Term Production Scheduling of Open Pit Mines*. PhD. thesis, University of Queensland, Australia, 256p.
- Godoy, M., Dimitrakopoulos, R.; Costa, J. F., 2001. Economic functions and geostatistical simulation applied to grade control: in Edwards, A.C. (Ed.) *Mineral Resource and Ore Reserve Estimation - The AusIMM Guide to Good Practice*. AusIMM., Melbourne, pp. 591-599.
- Goovaerts, P., 1997. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York, 464p.
- Goovaerts, P., 1999. Geostatistics in soil science: state-of-the-art and perspectives. *Geoderma*, vol. 89, pp. 1-45.
- Goovaerts, P., 2001. Geostatistical modelling of uncertainty in soil science. *Geoderma*, vol. 103, pp. 3-26.
- Goovaerts, P.; Van Meirvenne, M., 2001. Accounting for measurement and interpolation errors in soil contaminant mapping and decision-making. In: *7th Annual Conference of the International Association for Mathematical Geology*, Cancun, Mexico, Sept. 6-12, 2001.
- Goria, S.; Armstrong, M.; Galli, A., 2001. Quantifying the impact of additional drilling on an openpit gold project. In: *7th Annual Conference of the International Association for Mathematical Geology*, Cancun, Mexico, Sept. 6-12, 2001.
- Gotway, C.A. and Rutherford, B.M., 1994. stochastic simulation for imaging spatial uncertainty: comparison and evaluation of available algorithms In: Armstrong, M.; Dowd, P. A., (eds.) *Geostatistical simulations*, Kluwer, Dordrecht, pp. 1-21.

- Gross, H.; Caers, J., 2004. A Direct Sequential Simulation Approach to Streamline-Based History Matching. In: *7th International Geostatistics Congress*, September 26th - October 1st, 2004 Banff, Alberta, Canada, ref. no. 302.
- Halton, J. H., 1970. A retrospective and prospective survey of the Monte Carlo method. In: Zemanian, A.H. (Ed.) *SIAM Review. Society for Industrial and Applied Mathematics*, pp. 1-63.
- Hammersley, J. M.; Handscomb, D. C., 1964. *Monte Carlo Methods*, Chapman and Hall, New York.
- Heuvelink, G. B. M.; Pebesma, E. J., 1999. Spatial aggregation and soil processes modelling. *Geoderma*, vol. 89, pp. 47-65.
- Hewett, T.A., 1986. Fractal Distributions of Reservoir Heterogeneity and their Influence on FluidTransport. In: *Society of Petroleum Engineers (SPE) paper no. 15386, 61st Annual Technical Conference of SPE*, New Orleans, USA.
- Huijbregts, C. J.; Matheron, G., 1970. Universal Kriging - an Optimal Approach to Trend Surface Analysis. In: *Decision Making in Mineral in Mineral Industry*, Special vol. no. 12, CIMM, Montreal, pp. 159-169.
- Iman, R., L.; Conover, W., J., 1982. A Distribution Free Approach to Inducing Rank Correlations Among Input Variables. *Communications in Statistics, Part B - Simulation and Computation*, vol. 11, pp. 311-334.
- Isaaks, E., H., 1990. *The Application of Monte Carlo Methods to the Analysis of Spatially Correlated Data*, Ph.D. thesis, Stanford University, Stanford, 213p.
- Isaaks, E. H.; Srivastava M. R., 1989. *An Introduction to Applied Geostatistics*. Oxford University Press, New York, 561p.
- Ito, K., 1954. Stationary Random Distributions. In: *University of Kyoto Memoirs, Series A, XXVIII, 3*, pp. 209-223.
- Jenssen, R. J., 1975. Square and cubic lattice sampling. *Biometrics*, vol. 31, pp. 449-471.
- Johnson, M., 1987. *Multivariate Statistical Simulation*. Wiley, New York, 240p.

-
- Journal, A. G., 1974. Geostatistics for conditional simulation of ore bodies. *Economic Geology*, vol. 69, no. 5, pp. 673-687.
- Journal, A. G., 1980. The Lognormal approach to predicting local distributions of selective mining unit grades. *Mathematical Geology*, vol. 12, no. 4, pp. 285-303.
- Journal, A. G., 1983. Non Parametric Estimation of Spatial Distributions. *Journal of International Association for Mathematical Geology*, vol. 15, no. 3, pp. 445-468.
- Journal, A. G., 1986a. Geostatistics: models and tools for the earth sciences. *Mathematical Geology*, vol. 18, no. 1, pp. 119-140.
- Journal, A. G., 1989. *Fundamentals of Geostatistics in Five Lessons*. Short Course in Geology. American Geophysical Union, vol. 8, Washington, DC, 40p.
- Journal, A. G. 1992. Computer imaging in the mineral industry - Beyond mere aesthetics. In: *23rd APCOM Symposium*, Littleton, Colorado, pp. 3-13.
- Journal, A. G., 1994. Modelling uncertainty: some conceptual thoughts. In: Dimitrakopoulos R., ed., *Geostatistics for the next century*. Kluwer Academic Publishers, Dordrecht, Netherlands, pp. 30-43.
- Journal, A. G., 2004. Beyond Covariance: Training Image as Explicit Models of Prior Connection Information. In: *7th International Geostatistics Congress*, September 26th - October 1st, 2004 Banff, Alberta, Canada, ref. no. 113.
- Journal, A. G.; Alabert, F., 1989. Non Gaussian data expansion in the earth sciences. *Terra Nova*, vol. 1, no. 2, pp. 123-134.
- Journal A. G.; Gomez-Hernandez, J., J., 1989. Stochastic imaging of the Wilmington classic sequence. In: *Society of Petroleum Engineers (SPE) paper no. 19857, 62nd Annual Technical Conference of SPE*, New Orleans, USA.
- Journal, A. G.; Huijbregts, C. J., 1978. *Mining Geostatistics*. Academic Press, London, 600p.
- Journal, A. G.; Posa, D., 1990. Characteristic behavior and order relations for indicator variograms. *Mathematical Geology*, vol. 22, no. 8, pp. 1011-1025.

-
- Journel, A. G.; Xu, W., 1994. Posterior identification of histograms conditional to local data. *Mathematical Geology*, vol. 22, no. 3, pp. 323-359.
- Kalos, M. H.; Whitlock, P. A., 1986. *Monte Carlo Methods, Volume I: Basics*. John Wiley & Sons, New York, 186p.
- Kolmogorov, A. N., 1941. Interpolation and Extrapolation of Stationary Random Sequences. *Izvestiia Akademii Nauk SSR, Serii Matematicheskaja*, vol. 5, pp. 3-14 (translation Memo RM. 3090-PR, Rand Corporation, Santa Monica, California, 1962, 14p.).
- Kolmogorov, A. N., 1950. *Foundations of the theory of probability*. Chelsea, New York.
- Koppe, V. C.; Gambin, F.; Costa, J. F.; Koppe, J. C.; Fallon, G.; Davies, N., 2004. Análise de Incerteza Associada à Determinação da Velocidade de Onda Sônica em Depósitos de Carvão Obtida por Perfilagem Geofísica – Newland Coal Mine – Austrália. In: *III Congresso Brasileiro de Mina A Céu Aberto & III Congresso Brasileiro de Mina Subterrânea*, Belo Horizonte, MG.
- Krige, D. G., 1951. *A Statistical Approach to Some Mine Valuations and Allied Problems at the Witwatersrand*. MSc. thesis, University of Witwatersrand, Johannesburg, South Africa.
- Lantuéjoul, C., 1994. *Cours de Sélectivité*. Cours C-140, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau, 72p.
- Lantuéjoul, C., 1997. Iterative algorithms for conditional simulations. In: Baafi, E. Y., Schofield, N. A. (Eds.), *Geostatistics Wollongong '96*. Kluwer Academic Publishing, Dordrecht, pp. 27-40.
- Lantuéjoul, C., 2002. *Geostatistical Simulation: Models and Algorithms*. Springer, New York, 256p.
- Lassonde, P., 1990. *The Gold book - The complete investment guide to precious metals*. Financial Times of Canada Series. Penguin, Montreal.
- Leuangthong, O., 2004. The Promises and Pitfalls of Direct Simulation. In: *7th International Geostatistics Congress*, September 26th - October 1st, 2004 Banff, Alberta, Canada, ref. no. 136.
- Luenberger, D., G., 1969. *Optimization by Vector Space*. John Wiley & Sons, New York, 344p.

-
- Luster, G., R., 1985. *Raw Materials for Portland Cement: Applications of Conditional Simulation of Correogionalization*, Ph.D. thesis, Stanford University, Stanford, 532p.
- Manchuk, J., Leuangthong, O.; Deutsch, C. V., 2004. Direct Geostatistical Simulation on Unstructured Grids. In: *7th International Geostatistics Congress*, September 26th - October 1st, 2004 Banff, Alberta, Canada, ref. no. 116.
- Manteufel, R. D., 2000. Parameter Importance Coupled with Latin Hypercube Sampling. In: *Proceedings, 8th ASCE Specialty Conference on Probabilistic Mechanics and Structural Reliability*, University of Notre Dame, Notre Dame, Indiana, July 24-26, 2000. CD-ROM Proceedings (Kareem, Haldar, Spencer, and Johnson, eds.), paper PMC2000-109, 6 pages.
- Marcotte, D.; David, M., 1985. The Bigaussian Approach: A Simple Method for Recovery Estimation. *Journal of International Association for Mathematical Geology*, vol. 20, no. 63, pp 625-644.
- Marechal, A., 1970. *Cokrigage et regression en correlation intrinsèque*. Publication N-205, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau.
- Matern, B. 1960. Spatial variation - Stochastic models and their applications to some problems in forest surveys and other sampling investigations. *Meddelanden fraen Statens Skogsforskningsinstitut*, Vol. 49, no. 5, pp. 1-44.
- Matheron, G. 1963. Principles of geostatistics. *Economic Geology*, Volume 58, pp. 1246-1266.
- Matheron, G., 1965. *Les variables régionalisées et leur estimation: une application de la théorie des fonctions aléatoires aux sciences de la nature*. Masson, Paris, 305p.
- Matheron, G., 1970. *La théorie des variables régionalisées et ses applications*. Les Cahiers du Centre de Morphologie Mathématique, Fascicule 5, Ecole des Mines de Paris, Fontainebleau, 212p.
- Matheron, G., 1971. *La théorie des fonctions aléatoires intrinsèques généralisées*. Publication N-252, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau.
- Matheron, G., 1972. *Quelques aspects de la montée*. Publication N-271, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau.

- Matheron, G., 1973. The intrinsic random functions and their applications. *Advances in Applied Probability*, no.5, pp. 439-468.
- Matheron, G., 1976. A Simple Substitute for Conditional Expectation: The Disjunctive Kriging. In: *Proceedings of Advanced Geostatistics in the Mining Industry*, (Guarascio, M.; David, M. and Huijbregts, eds.), NATO Advanced Study Institute, Rome, Italy, 13-25 October 1975, D. Reidel Publishing Co., Dordrecht, Holland, pp. 221-236.
- Matheron, G., 1982. *Pour une analyse krigéante des données régionalisées*. Rapport N-732, Centre de Géostatistique, Fontainebleau, França.
- Matheron, G., Beucher, H., de Fouquet, C., Galli, A., Guerillot, D.; Ravenne, C., 1987. Conditional Simulation of the Geometric of Fluvio-Deltaic Reservoir. *Society of Petroleum Engineers (SPE)*, paper no. 16753.
- McKay, M. D.; Beckman, R. J.; Conover, W. J., 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, vol. 21, no. 2, pp. 239-245.
- Metropolis, N.; Ulam, S., 1949. The Monte Carlo method. *Journal of American Statistical Association* no. 44, pp. 335-341.
- Meyer, M. A., 1956. *Symposium on Monte Carlo Methods*. John Wiley & Sons, New York.
- Myers, D. E., 1982. Matrix formulation of cokriging. *Mathematical Geology*, vol. 14, no. 3, pp. 249-257.
- Myers, J. C., 1996. *Geostatistical Error Management: Quantifying Uncertainty for Environmental Sampling and Mapping*. Van Nostrand Reinhold, New York, 545p.
- Pachepsky, Y.; Acock, B., 1998. Stochastic imaging of soil parameters to assess variability and uncertainty of crop yield estimates. *Geoderma*, vol. 85, pp. 213-229.
- Oz, B.; Deutsch, C. V.; Tran, T. T. ; Xie, Y., 2003. DSSIM-HR : A Fortran 90 program for direct sequential simulation with histogram reproduction. *Computers & Geoscience*, vol. 29, pp. 39-51.

- Patterson, H. D., 1954. The errors of lattice sampling. *Journal of the Royal Statistical Society, Series B*, vol. 16, pp. 140–149.
- Pebesma, E., J.; Heuvelink, G., B., M., 1999. Latin Hypercube Sampling of Gaussian Random Fields. *Technometrics*, vol. 41, no. 4, pp. 303-312.
- Pilger, G. G., 2000. *Crerios para Locaço Amostrai Baseados em Simulaço Estocstica*. Dissertaço de Mestrado, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil, 127p.
- Pilger, G. G.; Costa, J. F. C. L.; Koppe, J. C., 2000. Projeto de amostragem: uma proposiço baseada em simulaço estocstica. *REM – Revista da Escola de Minas de Ouro Preto*, Ouro Preto, vol. 53, no. 4, pp. 265-274.
- Pilger, G. G.; Costa, J. F. C. L.; Koppe, J. C., 2001. Additional Samples: Where they Should be Located?. *Natural Resource Research*, vol. 10, no. 3, pp. 197-207
- Pilger, G. G.; Costa, J. F. C. L.; Koppe, J. C., 2002. Optimizing the value of a sample. In: *Application of Computers and Operations Research in the Mineral Industry, 2002, Phoenix. In: Proceedings of the 30th International Symposium*. Littleton: Society for Mining, Metallurgy and Exploration, Inc. (SME), 2002. vol. 1, pp. 85-94.
- Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Wetterling, W. T., 1992. *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge University Press.
- Raj, D., 1968. *Sampling Theory*. McGraw-Hill, New York, 302p.
- Ravenscroft, P. J., 1992. Risk analysis for mine scheduling by conditional simulation. *Transcripts of Institution of Mining and Metallurgy*, no. 101, pp. A104-A108.
- Remacre, A. Z.; Feitosa V. N.; Pires, J. S., 1992. Conditional Simulation as a Tool for Mine Planning. In: *Proceedings of the 23rd APCOM (International Symposium on the Application of Computers and Mathematics in the Mineral Industry*, Tucson, USA, SME-AIME, Littleton, CO, USA, pp. 259-266.
- Richmond, A., J., 2003. Financially efficient ore selections incorporating grade uncertainty. *Mathematical Geology*, vol. 35, no. 2, pp. 195-215.

-
- Richmond, A., J., 2004. *Financially Efficient Mining Decisions Incorporating Grade Uncertainty*. PhD. thesis, Imperial College, London, 208p.
- Riordan, J., 1958. *An Introduction to Combinatorial Analysis*. Wiley, New York, 210p.
- Rivoirard, J., 1991. *Introduction au Krigeage Disjonctif et à la Geostatistique Non Linéaire*. Cours C-139, 2^{ème} édition, Centre de Geostatistique, EMP, 99p.
- Rivoirard, J., 2001. Which models for collocated cokriging? *Mathematical Geology*, vol. 33, no. 2, pp. 117-131.
- Rivoirard, J., 2002. *Cours de Géostatistique Multivariable*. Cours C-172, Centre de Géostatistique, EMP. 67p.
- Rossi, M. E.; Parker, H. M., 1994. Estimating recoverable reserves: Is it hopeless? *Geostatistics for the Next Century*, Montreal, Canada, Kluwer Academic Publishers, pp. 259-276.
- Rossi, M.E.; van Brunt, B. H., 1997. Optimizing conditionally simulated orebodies with Whittle 4D. In: *Proceedings of Optimizing with Whittle 97*, Perth, 8-9 April., pp. 102-118.
- Rubinstein, R. Y., 1981. *Simulation and the Monte Carlo Method*. John Wiley & Sons, New York, 304p.
- Saliby, E., 1980. *A Reappraisal of Some Simulation Fundamentals*. Ph.D. thesis, University of Lancaster.
- Saliby, E., 1990. Descriptive Sampling: A better Approach to Monte Carlo Simulation. *Journal of the Operational Research Society*, vol. 41, pp. 1133-1142.
- Saliby, E., 1997. Descriptive sampling: an improvement over latin hypercube sampling. In: *Proceedings of the 1997 Winter Simulation Conference (WSC' 97)*, Atlanta, December 1997, ed. S. Andradóttir, K. J. Healy, D. H. Withers, and B. L. Nelson, pp. 230-233.
- Santos, S. F. dos, 2002. *Redução da Incerteza Associada a Modelos Estocásticos de Fácies Através do Condicionamento a Dados de Produção*. Dissertação de Mestrado, Universidade Federal do Espírito Santo, Vitória, 120p.
- Satterthwaite, F. E., 1959. Random balance experimentation. *Technometrics*, vol. 1, pp. 111-137.

-
- Soares, A., 2001. Direct sequential simulation and cosimulation. *Mathematical Geology*, vol. 33, no. 8, pp. 911-926.
- Srivastava, R. M., 1992. Reservoir characterization with probability field simulation. *Society of Petroleum Engineers*, paper no. 24753.
- Srivastava, R. M., 1994. An overview of stochastic methods for reservoir characterization. In: Yarus, J. M.; Chambers, R. L. (eds), *Stochastic modeling and geostatistics, The American Association of Petroleum Geologists*, Tulsa, pp. 3-16.
- Srivastava, R. M.; Frykman, P.; Jensen, M., 2004. Geostatistical Simulation of Discrete Fracture Network Models. In: *7th International Geostatistics Congress*, September 26th - October 1st, 2004 Banff, Alberta, Canada, ref. no. 154.
- Srivastava, R. M., Hartzell, D .R. and Davis, B. M., 1992. Enhanced Metal Recovery Through Improved Grade Control. In: *Proceedings of the 23rd APCOM (International Symposium on the Application of Computers and Mathematics in the Mineral Industry)*, (Kim, Y.C., ed.), SME-AIME, Littleton, CO, USA, pp. 243-249.
- Stein, M., 1987. Large sample properties of simulations using Latin Hypercube Sampling. *Technometrics*, vol. 29, no. 2, pp. 143-151.
- Steinberg, H. A., 1963. Generalized quota sampling. *Nuclear Science and Engineering*, vol. 15, pp. 142-145.
- Strebelle, S. B., 2001. Conditional simulation of complex geological structures using multiple-point statistics. *Mathematical Geology*, vol. 34, no. 1, pp. 1-21.
- Strebelle, S. B.; Zhang, T., 2004. Non-Stationary Multiple-Point Geostatistical Models. In: *7th International Geostatistics Congress*, September 26th - October 1st, 2004 Banff, Alberta, Canada, ref. no. 142.
- Sullivan, J., 1984. Conditional recovery estimation through probability kriging – Theory and practice. In: *Geostatistics for Natural Resources Characterization*, G. Verly, M. David, A. G. Journel, and A. Maréchal, eds. Reidel, Dordrecht, Holland, Part 1, pp. 365-384.

-
- Suro-Péres, V.; Journel, A. G., 1991. Indicator principal component kriging. *Mathematical Geology*, vol. 23, no. 5, pp. 759-788.
- Tran, T. T.; Deutsch, C. V.; Xie, Y., 2001. Direct geostatistical simulation with multiscale well, seismic, and production data. In: *Society of Petroleum Engineers (SPE) paper no. 71323, 62nd Annual Technical Conference of SPE*, New Orleans, USA.
- Ulam, S. M., 1987. *Stanislaw Ulam, 1909–1984*, Los Alamos National Laboratory, Los Alamos, New Mexico. A special issue of Los Alamos Science, no. 15.
- Vallee, M., 1999. Resource/reserve inventories: What are the objectives? *CIM Bulletin*, vol. 92, no. 1031, pp. 151-155.
- van Groenigen, J. W., 1999. *Constrained Optimization of Spatial Sampling – A Geostatistical Approach*. PhD thesis, Wageningen Agricultural University and ITC Enschede, 148p.
- van Groenigen, J., W., 2000. The influence of variogram parameters on optimal sampling schemes for mapping by kriging. *Geoderma*, vol. 97, pp. 223-236.
- van Groenigen, J. W.; Siderius, W.; Stein, A., 1999. Constrained optimization of soil sampling for minimisation of the kriging variance. *Geoderma*, vol. 87, pp. 239-259.
- van Meirvenne, M.; Goovaerts, P., 2001. Evaluating the probability of exceeding a site-specific soil cadmium contamination threshold. *Geoderma*, vol. 102, pp. 75-100.
- Veach, E., 1997. *Robust Monte Carlo Methods for Light Transport Simulation*. Ph.D. thesis, Stanford University, Stanford, 432p.
- Verly, G., 1983. The multiGaussian approach and its applications in the estimation of local recoveries. *Mathematical Geology*, vol. 15, no. 2, pp. 263-290.
- Verly, G., 1984. *Estimation of Spatial Point and Block Distributions: The MultiGaussian Model*. PhD thesis, Stanford University, Stanford, Stanford, 416p.
- Viscarra Rossel, R. A.; Goovaerts, P.; McBratney, A. B., 2001. Assessment of the production and economic risks of the site-specific liming using geostatistical uncertainty modeling. *Environmetrics*, vol. 12, no. 5, pp. 112-125.

-
- Wackernagel, H., 1993. *Cours de Géostatistique Multivariable*, Cours C-146, Centre de Géostatistique, EMP. 80p.
- Wackernagel, H., 1994. Cokriging versus kriging in regionalized multivariate data analysis. *Geoderma*, vol. 62, pp. 83-92.
- Wackernagel, H., 2002. *Notas de aula – C.F.S.G. 2001/2002*, ENSMP, Fontainebleau, França.
- Whittle, P., 1963. Stochastic processes in several dimensions. *Bulletin International Statistical Institute*, vol. 40, pp. 974-994.
- Yaglom, A. M., 1962. *An Introduction to the Theory of Stationary Random Functions*. Prentice-Hall, Englewood Cliffs, New Jersey, 256p.
- Yates, F., 1953. *Sampling Methods for Censuses and Surveys*, second ed., Griffin, London. Fourth edition also published by Macmillan, New York (1981).
- Youden, W. J.; Kempthorne, O.; Tukey, J. M.; Box, G. E. P.; Hunter, J. S., 1959. Discussion of the papers of Messrs. Satterthwaite and Budne. *Technometrics*, vol. 1, pp. 157-193.
- Zacks, S., 1963. On a complete class of linear unbiased estimators for randomized factorial experiments. *Annals of Mathematical Statistics*, vol. 34, pp. 769-779.
- Zacks, S., 1964. Generalized least squares estimators for randomized fractional replication designs. *Annals of Mathematical Statistics*, vol. 34, pp. 696-704.
- Zhang, T.; Switzer, P.; Journel, A. G., 2004. Sequential Conditional Simulation by Identification of Training Patterns. In: *7th International Geostatistics Congress*, September 26th - October 1st, 2004 Banff, Alberta, Canada, ref. no. 127.

Apêndice A

Decomposição de matrizes e ACP

Neste apêndice são apresentados os conceitos e definições de valores e vetores próprios, assim como a suas relações com a decomposição de matrizes. Ainda, a partir das definições precedentes, a teoria da análise dos componentes principais (ACP) é apresentada, juntamente com sua relação com a correlação entre as VA e os fatores independentes. A apresentação de um procedimento de decomposição de matrizes mais genérico, o método de decomposição em valores singulares, encerra esse apêndice.

A.1. Valores próprios ou autovalores

Seja B uma matriz quadrada $n \times n$. A equação

$$\det(\lambda I - B) = 0 \quad (\text{A.1})$$

apresenta n soluções para λ (sendo que I é a matriz identidade de B). Geralmente λ apresenta soluções compostas de valores complexos, salvo se B for simétrica. Nesse caso, todos os valores de λ são reais. Os valores da matriz coluna λ são chamados de valores próprios de B . Os valores próprios λ apresentam as seguintes características:

- i. A soma dos valores próprios λ é igual ao traço da matriz B

$$\text{tr}(B) = \sum_{i=1}^n b_{ii} = \sum_{p=1}^n \lambda_p \quad (\text{A.2})$$

- ii. O produto dos valores próprios λ é igual ao determinante da matriz B .

$$\det(B) = \prod_{p=1}^n \lambda_p \quad (\text{A.3})$$

A.2. Vetores próprios ou autovetores

Se B é uma matriz quadrada $n \times n$ e λ são os valores próprios de B , existem vetores distintos q_C e q_L (não nulos) de modo que:

$$(\lambda \cdot I - B) q_C = 0 \quad \text{e} \quad (q_L)^t \cdot (\lambda \cdot I - B) = 0 \quad (\text{A.4})$$

ou seja

$$B \cdot q_C = \lambda \cdot q_C \quad \text{e} \quad (q_L)^t \cdot B = \lambda \cdot (q_L)^t \quad (\text{A.5})$$

Os vetores q_C são os vetores próprios das colunas de B , enquanto que q_L são os vetores próprios das linhas de B . Se B for simétrica, então $q_C = q_L$.

A.3. Decomposição em valores e vetores próprios

Seja B uma matriz simétrica. Se a relação

$$B \cdot Q = Q \cdot \Lambda \quad (\text{A.6})$$

onde Q é ortogonal ($Q \cdot Q^t = I$), for verdadeira, então Q é a matriz dos vetores próprios de B e Λ é a matriz diagonal dos valores próprios λ de B . Sabendo que $Q \cdot Q^t = I = Q \cdot Q^{-1}$, logo $Q^t = Q^{-1}$, então a Equação (A.6) pode ser reescrita como:

$$B = Q \cdot \Lambda \cdot Q^t \quad (\text{A.7})$$

A.4. Análise dos componentes principais

A análise dos componentes principais (ACP) consiste em transformar um conjunto de VA correlacionadas em um conjunto de VA sem correlação espacial ($\forall h$). Em um contexto Gaussiano, essas VA não correlacionadas podem ser interpretadas como fatores independentes (ortogonais) e intrínsecos do fenômeno.

Se Z é a matriz $N \times n$ das medidas disponíveis centradas ($z(u_\alpha) - E[Z(u)]$, $\alpha=1, \dots, n$), a matriz de variância-covariância V é definida como:

$$V = \frac{1}{n} Z^t \cdot Z \quad (\text{A.8})$$

Seja Y uma matriz $N \times n$ contendo em suas linhas os n fatores Y_α ($\alpha=1, \dots, n$) experimentais, independentes e de esperança matemática nula, correspondentes aos dados $z(u_\alpha)$. A matriz diagonal de variância-covariância Λ ($N \times N$) dos fatores Y_α é definida como:

$$\Lambda = \frac{1}{n} Y^t \cdot Y \quad (\text{A.9})$$

onde os elementos dessa matriz são as variâncias dos fatores Y_α .

O objetivo da ACP é buscar uma matriz X , ortogonal e de dimensão $N \times N$, que transforme linearmente os valores correlacionados medidos em fatores Y_α independentes, ou seja:

$$Y = Z \cdot X \quad \text{sendo que } X^t \cdot X = I \quad (\text{A.10})$$

Multiplicando os dois lados dessa equação por $1/n$ e por Y^t

$$\frac{1}{n} Y^t \cdot Y = \frac{1}{n} Y^t \cdot Z \cdot X \quad (\text{A.11})$$

e substituindo Y por Z . X

$$\frac{1}{n} (Z \cdot X)^t \cdot Z \cdot X = \frac{1}{n} Z^t \cdot X^t \cdot Z \cdot X = X^t \cdot V \cdot X \quad (\text{A.12})$$

Portanto

$$V \cdot X = X \cdot \Lambda \quad (\text{A.13})$$

Ora, pela equação acima e pelas características da matriz X, pode ser imediatamente verificado que a matriz X (que está sendo buscada) pode ser identificada com a matriz de vetores próprios Q. Ainda, os valores próprios λ podem ser identificados como as variâncias dos fatores Y, constituindo a matriz diagonal Λ .

Logo

$$V \cdot Q = Q \cdot \Lambda \quad \text{sendo que } Q^t \cdot Q = I \quad (\text{A.14})$$

e os fatores Y podem enfim ser definidos como:

$$Y = Z \cdot Q \quad (\text{A.15})$$

A.5. Correlação das variáveis com os fatores

Seja B uma matriz de correlação entre N VA $Z_i(u)$ ($i=1,\dots,N$), simétrica ($h=0$). A correlação entre essas VA e os fatores $\gamma_q^s(u)$ (onde s é o n°. de estruturas do semi-variograma e q é o n°. de fatores dentro de cada estrutura s) é definida como:

$$\text{Corr}(Z_i(u), \gamma_q^s(u)) = A = a_{iq}^s = \sqrt{\Lambda} \cdot Q \quad (\text{A.16})$$

Na equação acima, os elementos a_{iq}^s da matriz A representam as covariâncias regionalizadas entre as VA $Z_i(u)$ e os fatores $\gamma_q^s(u)$, enquanto que Q e Λ são as matrizes dos vetores e valores próprios de B, respectivamente.

Logo, se B pode ser decomposta pelos componentes principais, de acordo com

$$B = Q \cdot \Lambda \cdot Q^t \quad (\text{A.17})$$

Essa decomposição pode ser escrita em termos de A:

$$B = A \cdot A^t \quad (\text{A.18})$$

A.6. Decomposição em valores singulares

A decomposição de matrizes em valores e vetores próprios, apresentada em A.3, só é válida para matrizes simétricas, ou seja, aquelas que são quadradas ($n \times n$) e que são iguais às suas

transpostas ($B = B^t$). Entretanto, na prática, às vezes se torna necessário decompor matrizes de variância-covariância ($h \neq 0$) que, na verdade, não são simétricas (a não ser que se for assumido que $C_{ij}(h) = C_{ji}(h)$). Nesses casos, outro tipo de decomposição de matrizes deve ser aplicado.

A decomposição em valores singulares (DVS) permite decompor matrizes retangulares ($n \times m$), seguindo o mesmo raciocínio empregado na decomposição de matrizes simétricas em valores e vetores próprios.

Seja B uma matriz retangular, de dimensão $n \times m$ e rank r , a decomposição de B em r valores singulares μ_p é dada por:

$$B = Q_1 \cdot \Sigma \cdot Q_2^t \quad (\text{A.19})$$

onde Q_1 e Q_2 são matrizes ortogonais, de dimensões $n \times n$ e $m \times m$, respectivamente. A matriz Σ é do tipo retangular e conta com os valores singulares μ_p na sua diagonal; enquanto que os demais elementos de Σ ($i \neq j$) são nulos. Por exemplo, se $n > m$, $r = m$, então:

$$\Sigma = \begin{pmatrix} \mu_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \mu_r \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 \end{pmatrix}$$

Essa decomposição pode ser realizada a partir do cálculo dos valores e vetores próprios de $B \cdot B^t$ e de $B^t \cdot B$. As matrizes ortogonais Q_1 e Q_2 são os vetores próprios de $B \cdot B^t$ e $B^t \cdot B$, respectivamente. Os valores singulares μ_p são as raízes quadradas dos valores próprios λ_p de $B \cdot B^t$ ou de $B^t \cdot B$ (idênticos, sendo que $\mu_p \geq 0$):

$$\mu_p = \sqrt{\lambda_p} \quad \text{ou} \quad \Sigma = \sqrt{\Lambda} \quad (\text{A.20})$$

Portanto, esse procedimento transforma matrizes assimétricas em simétricas, a partir do produto da matriz pela sua transposta. Dessa forma, a matriz transformada pode ser decomposta em valores e vetores próprios, conforme o item A.3 – pois a matriz deve ser simétrica para garantir que os valores próprios sejam reais.

Geralmente, em geoestatística, a análise dos componentes principais (ACP) é realizada a partir da decomposição das matrizes de variância-covariância em valores singulares (DVS). Isso se explica pela relação direta entre os valores singulares e a correlação entre as VA $Z_i(u)$ e os fatores regionalizados $Y_q^s(u)$, conforme a Equação (A.16). Quando o efeito de retardo é desconsiderado, assumindo-se $C_{ij}(h) = C_{ji}(h)$, as matrizes de variância-covariância são simétricas e portanto, $Q_1 = Q_2$.

Apêndice B

Reprodução da função de covariância espacial

Esse apêndice tem o objetivo de validar o algoritmo de simulação seqüencial Gaussiana modificado em relação ao momento condicionante de segunda ordem. Ao longo do Capítulo 4 (em todos os testes realizados) foi verificado que, em média, os desvios padrão das realizações se identificavam com o desvio padrão condicionante. Essa verificação permitiu a validação dos processos de simulação, com respeito à reprodução do momento condicionante de segunda ordem – variância. Entretanto, o procedimento de verificação realizado não contemplou uma abordagem em relação à reprodução da função de covariância espacial (função de h). Esse apêndice preenche essa lacuna, mostrando que a função de covariância espacial é reproduzida pelas realizações obtidas a partir do algoritmo modificado – LHS clássico. Para essa análise, foram avaliadas cinco realizações (no espaço Gaussiano) tiradas de forma aleatória, de processos de simulação diferentes.

B.1. Algoritmo original (MC)

Primeiramente, foram avaliadas cinco realizações produzidas a partir do algoritmo original, isto é, aquele que utiliza o método de Monte Carlo (MC). Foram tiradas, aleatoriamente, cinco realizações dentre as 200 realizações originalmente produzidas por meio desse algoritmo. Essas realizações foram as número 22, 73, 98, 127 e 183. A Figura B.1 mostra os semi-variogramas experimentais dessas realizações, com respeito à direção de menor variabilidade espacial (azimute 158°), juntamente com o respectivo semi-variograma condicionante. A Figura B.2 ilustra os semi-variogramas, experimentais e condicionante, associados à direção de maior variabilidade espacial (azimute 68°). Na Figura B.1, é observado que os patamares dos semi-variogramas experimentais são sistematicamente subestimados. Nesses semi-variogramas experimentais, o patamar condicionante é atingido em distâncias superiores a 75 m (condicionante). Entretanto, a variância a “priori” sempre é

reproduzida pelas realizações, como descrito no Capítulo 4. Os semi-variogramas experimentais da Figura B.2 reproduzem adequadamente o semi-variograma condicionante.

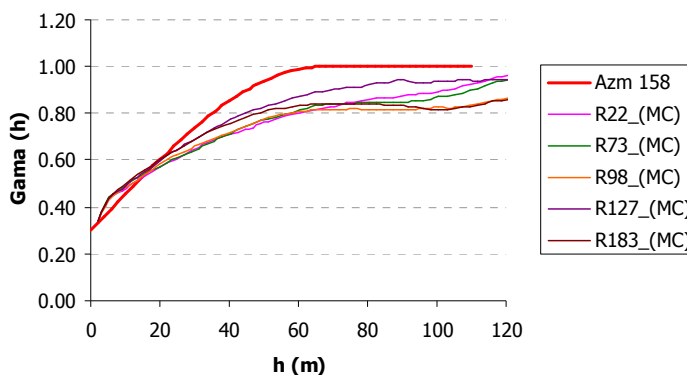


Figura B.1: Semi-variogramas experimentais, associados à direção de menor variabilidade espacial (azimute 158°), referentes a cinco realizações produzidas por meio do algoritmo original, superpostas ao respectivo semi-variograma condicionante.

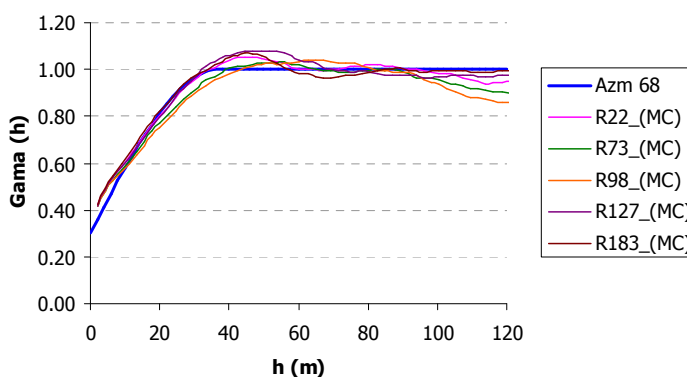


Figura B.2: Semi-variogramas experimentais, associados à direção de maior variabilidade espacial (azimute 68°), referentes a cinco realizações produzidas por meio do algoritmo original, superpostas ao respectivo semi-variograma condicionante.

B.2. Algoritmo modificado (LHS)

O desempenho do algoritmo modificado, em relação à reprodução da função de covariância espacial, foi avaliado a partir de cinco realizações oriundas de três processos de simulação diferentes. Os mesmos foram realizados a partir de diferentes modelos de discretização. Os três modelos foram os seguintes:

- i. $M = 10$, modelo com 10 classes de discretização;
- ii. $M = 14$, modelo com 14 classes de discretização e

iii. $M = 20$, modelo com 20 classes de discretização.

Dentre as 10 realizações, produzidas por meio do processo de simulação no qual $M = 10$, foram avaliadas as realizações de número dois, cinco, seis, oito e 10. A partir do processo de simulação no qual o modelo de discretização utilizado foi o de 14 classes ($M = 14$), foram avaliadas as realizações de número um, cinco, sete, nove e 12. Para o caso no qual M foi ajustado para 20 classes ($M = 20$), foram avaliadas as realizações de número quatro, nove, 10, 14 e 16. As Figuras B.3, B.4 e B.5 ilustram os semi-variogramas experimentais dessas realizações, assim como os condicionantes, segundo as direções de menor e maior variabilidade espacial. Essas figuras mostram que o comportamento dos semi-variogramas experimentais das realizações obtidas por meio do algoritmo modificado é o mesmo apresentado pelas realizações provenientes do algoritmo original.

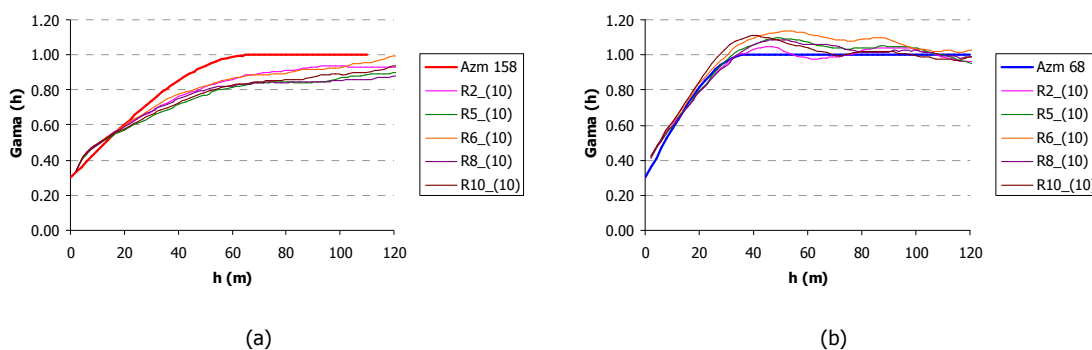


Figura B.3: Semi-variogramas experimentais, associados aos azimutes 158° (a) e 68° (b), referentes a cinco realizações produzidas por meio do algoritmo modificado (LHS), a partir de 10 classes de discretização ($M = 10$), superpostas aos respectivos semi-variogramas condicionantes.

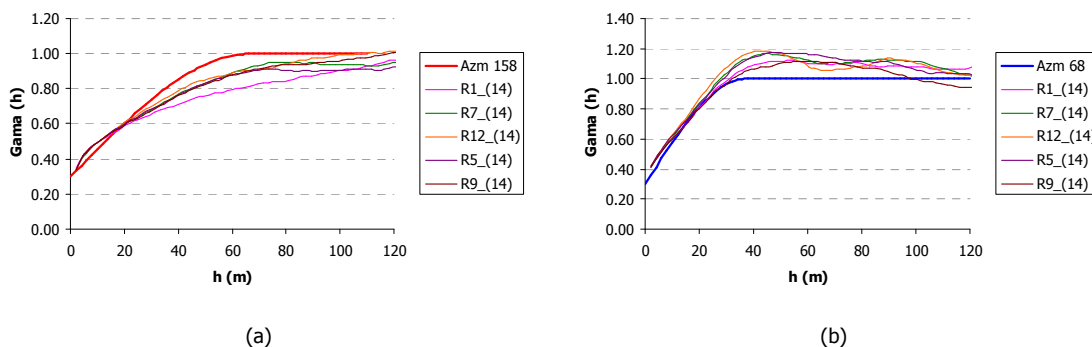


Figura B.4: Semi-variogramas experimentais, associados aos azimutes 158° (a) e 68° (b), referentes a cinco realizações produzidas por meio do algoritmo modificado (LHS) a partir de 14 classes de discretização ($M = 14$), superpostas aos respectivos semi-variogramas condicionantes.

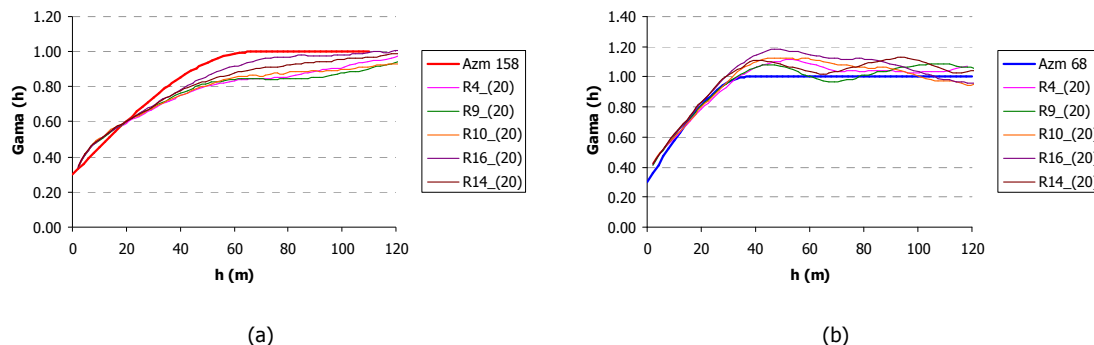


Figura B.5: Semi-variogramas experimentais, associados aos azimutes 158° (a) e 68° (b), referentes a cinco realizações produzidas por meio do algoritmo modificado (LHS) a partir de 20 classes de discretização ($M = 20$), superpostas aos respectivos semi-variogramas condicionantes.

Portanto, o comportamento dos semi-variogramas experimentais das Figuras B.3, B.4 e B.5 indicam que o algoritmo modificado não interfere na reprodução da função de covariância espacial. Esse comportamento já era esperado, pois a modificação do algoritmo só envolveu a etapa de tiragem de uma distribuição uniforme, para a obtenção do valor Gaussiano $x_{p_{lhs}}$ [Equação (5.17)]. A modificação não atinge o método de construção das N distribuições condicionais univariadas [Equação (3.1)]. Sendo assim, como provado por Journel (1994), em média, a função de covariância espacial das realizações tiradas a partir da fdp multivariada $f(u_1, \dots, u_N; z_1, \dots, z_N)$ se identifica com o modelo de covariância espacial condicionante.