

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
Programa de Pós-Graduação em Engenharia de Minas,
Metalúrgica e de Materiais (PPGE3M)

**OTIMIZAÇÃO DE PARÂMETROS DE KRIGAGEM
BASEADA NA MINIMIZAÇÃO DO ERRO ABSOLUTO E O ERRO QUADRÁTICO**

Ricardo José Hundelshausen Rubio

Tese para a obtenção do título de
Doutor em Engenharia

Porto Alegre, RS

2018

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
Programa de Pós-Graduação em Engenharia de Minas,
Metalúrgica e de Materiais (PPGEM)

**OTIMIZAÇÃO DE PARÂMETROS DE KRIGAGEM
BASEADA NA MINIMIZAÇÃO DO ERRO ABSOLUTO E O ERRO QUADRÁTICO**

Ricardo José Hundelshausen Rubio

Engenheiro Industrial, MsC.

Tese realizada no Laboratório de Pesquisa Mineral e Planejamento Mineiro da Escola de Engenharia da UFRGS, dentro do Programa de Pós-Graduação em Engenharia de Minas, Metalúrgica e de 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

2018

Esta tese foi julgada adequada para a obtenção do título de Doutor em Engenharia, área de concentração Metalurgia Extrativa / Tecnologia Mineral, e aprovada em sua forma final pelo Orientador e pela Banca Examinadora do Curso de Pós-Graduação.

Orientador:

Prof. Dr. João Felipe Coimbra Leite Costa

Co-orientador:

Prof. Dr. Diego Machado Marques

Banca Examinadora:

Prof. Dr. Marcelo Cheuiche Godoy

Prof. Dr. Ramón Giraldo Henao

Prof.^a Dr.^a Vanessa Cerqueira Koppe

Prof. Dr. Carlos Pérez Bergmann
Coordenador do PPGEM

*“Quem não assume um risco
nunca ganhará um jogo.”*

(G.M. Paul Keres)

A minha esposa Vanessa Szupszynski,
meus familiares e amigos.

AGRADECIMENTOS

Ao Prof. Dr. João Felipe Coimbra Leite Costa, por sua orientação, confiança, conselhos, por sua dedicação para tornar este projeto realidade e, sobretudo, pela amizade incondicional.

Ao Prof. Dr. Diego Machado Marques pela coorientação e pelas valorosas contribuições para o desenvolvimento deste trabalho.

Aos Professores do Departamento de Engenharia de Minas (DEMIN) e do Programa de Pós-Graduação em Engenharia de Minas, Metalúrgica e de Materiais (PPGE3M).

Aos colegas e amigos do LPM, que contribuíram para este trabalho com diferentes perspectivas, em especial, Diego, Marcel, Anuar, Pablo, Vanessa, Roberto, George, Ricardo, Samanta, Henrique, Péricles, Cristina, Jonas, Ryu, Augusto, David.

À empresa de Mineração Newmont, pela disponibilização dos dados para este estudo.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela bolsa concedida.

A meus pais Alejandra e Joaquín, meus irmãos Laura Andrea, Maria Alejandra e Joaquín Fernando, meus tios, primos e demais familiares.

A meus sogros Paulina e Miguel, meus cunhados Guilherme e Priscila, meu sobrinho Vinícius e Dona Nadzieja.

A meus amigos “colombo-gaúchos” Javier, Rosane, Victoria, Dayana, Fabrício, Fausto, Simalila, pelas alegrias compartilhadas.

A meus padrinhos de casamento Marcel Bassani e Waleska Campos, pela confiança e amizade incondicional nestes últimos anos.

A minha esposa Vanessa Szupczynski, por me orientar nas decisões fundamentais deste projeto, por sua dedicação na construção e execução dessa etapa da vida e pelo apoio, carinho e motivação sempre constantes, bem como a nosso cachorro Jake, por nos ter proporcionado felicidade todos os dias dessa jornada.

SUMÁRIO

LISTA DE FIGURAS.....	X
LISTA DE TABELAS.....	XIV
RESUMO.....	XV
ABSTRACT.....	XVI
CAPÍTULO 1 - INTRODUÇÃO.....	1
1.1. Introdução.....	1
1.2. Motivação.....	4
1.3. Meta.....	7
1.4. Objetivos.....	7
1.5. Metodologia.....	7
1.5.1. LKPO utilizando validação cruzada.....	7
1.5.2. LKPO utilizando simulações.....	10
1.6. Contribuição da Tese.....	13
1.7. Estrutura da Tese.....	13
CAPÍTULO 2 - CONCEITOS.....	15
2.1. Conceitos Gerais.....	15
2.1.1. Estacionariedade.....	15
2.1.2. Viés condicional.....	16
2.1.3. Inclinação da reta de regressão para estimativas de depósitos minerais.....	18
2.1.4. Peso da média da krigagem simples.....	18
2.1.5. Pesos negativos na krigagem.....	19
2.1.6. Variância da krigagem.....	20
2.1.7. Eficiência de krigagem.....	21
2.1.8. Validação cruzada.....	21
2.1.9. Erro das estimativas.....	22
2.1.10. Simulação geoestatística.....	23
2.1.11. gráficos de acurácia.....	26
2.2. Conceitos Específicos Para a Metodologia Proposta.....	27
2.2.1. Agrupamento de dados.....	27
2.2.2. K-means.....	27
2.2.3. Determinando o número de clusters.....	28
2.2.4. Medidas de distâncias.....	30

2.2.4.1. <i>Distância Euclidiana</i>	30
2.2.4.2. <i>Distância Manhattan</i>	30
2.2.4.3. <i>Distância Minkowski</i>	30
2.2.5. Criação da área/volume de influência.....	31
2.2.6. Parâmetros de krigagem.....	32
2.2.6.1. <i>Alcance da busca</i>	32
2.2.6.2. <i>Número de amostras</i>	33
2.2.6.3. <i>Número de setores angulares</i>	34
2.2.6.4. <i>Número de amostras por furo</i>	35
CAPÍTULO 3 - ESTUDOS DE CASO 2D E 3D	36
3.1. Banco de Dados 2D.....	36
3.1.1. Estatística básica.....	37
3.1.2. Escolhendo o número de domínios.....	37
3.1.2.1. <i>CASO I: Visual</i>	37
3.1.2.2. <i>CASO II: Visual</i>	38
3.1.2.3. <i>CASO III: Teor de corte pela mediana</i>	39
3.1.2.3. <i>CASO IV: Técnicas de agrupamentos de dados</i>	39
3.1.3. Modelagem da continuidade espacial.....	41
3.1.4. Definição das estratégias de busca.....	42
3.1.5. Definição da área de influência de cada amostra.....	43
3.1.6. Minimização do erro absoluto.....	43
3.1.7. Estimativas dentro de cada área de influência.....	44
3.2. Banco de Dados 3D.....	44
3.2.1. Domínios.....	46
3.2.2. Estatística básica.....	47
3.2.3. Modelagem da continuidade espacial.....	47
3.2.4. Definição do número de estratégias de busca.....	48
3.2.5. Definição da área de influência de cada amostra.....	49
CAPÍTULO 4 – RESULTADOS E ANÁLISES DA LKPO UTILIZANDO VALIDAÇÃO CRUZADA	50
4.1. Banco de Dados 2D.....	50
4.1.1. Resumo estatístico.....	51
4.1.2. Curvas teor-tonelagem.....	53
4.1.3. Variações do erro na curva teor-tonelagem.....	56

4.1.4. Análise de suavização da krigagem	58
4.1.5. Análise das médias locais (análise de deriva)	59
4.1.6. Análise visual	60
4.2. Banco de Dados 3D	61
4.2.1. Curvas teor-tonelagem por cada domínio	63
4.2.2. Variações do erro na curva teor-tonelagem	65
4.2.3. Unindo os domínios 3 e 4.....	66
4.2.4. Resumo estatístico.....	67
4.2.5. Análise das médias locais.....	67
4.2.6. Análise de suavização da krigagem	70
4.2.7. Análise da performance do algoritmo da LKPO.....	71
CAPÍTULO 5 – RESULTADOS E ANÁLISES DA LKPO UTILIZANDO SIMULAÇÕES	73
5.1. Verificação dos Modelos Simulados	73
5.1.1. Reprodução dos histogramas	73
5.1.2. Reprodução dos variogramas	73
5.1.3. Gráfico de acurácia.....	73
5.1.4. Estatística básica	77
5.1.5. Curvas teor-tonelagem	77
5.1.6. Análise das médias locais.....	78
5.1.7. Análise de suavização da krigagem	80
CAPÍTULO 6 – CONSIDERAÇÕES FINAIS.....	81
6.1. Comparação entre os Modelos LKPO/LKNA	81
6.2. Comparação entre os Modelos LKPO e Krigagem Ordinária Tradicional.....	83
CAPÍTULO 7 - CONCLUSÕES	86
7.1. Conclusões	86
7.2. Trabalhos Futuros	87
7.3. Recomendações Finais.....	87
REFERÊNCIAS BIBLIOGRÁFICAS.....	89
APÊNDICE A - CONTINUIDADE ESPACIAL 3D	96
APÊNDICE B - PLATAFORMA DO ALGORITMO DA LKPO.....	97
APÊNDICE C - CONTINUIDADE ESPACIAL 2D	99

LISTA DE FIGURAS

Figura 1.1. Mapa de localização das amostras.....	6
Figura 1.2. Fluxograma da metodologia da LKPO.....	8
Figura 1.3. Etapas do processo de LKPO utilizando simulações.	10
Figura 1.4. Simulações geoestatísticas (a) suporte de pontos (b) suporte blocos (re-blocagem). 11	
Figura 1.5. Minimização do erro quadrático em relação às incertezas de cada estratégia para o (a) bloco #1 (b) bloco #2.	12
Figura 2.1. Efeito do viés condicional entre o valor real e o valor estimado. Figura modificada de Vann et al. (2003), originalmente proposta por Journel & Huijbregts (1978).....	17
Figura 2.2. Histograma do erro obtido por validação cruzada (a) viés positivo (b) viés negativo (c) sem viés. Figura modificada de Isaaks & Srivastava (1989).....	22
Figura 2.3. Flutuações ergódicas nos variogramas. Imagem modificada do (Deutsch & Journel, 1998).	23
Figura 2.4. Princípio das bandas rotativas em 2D (imagem extraída de Chilès & Delfiner,1999).	24
Figura 2.5. Gráfico de acurácia (imagem extraída de Pyrcz & Deutsch 2014).	26
Figura 2.6. Agrupamento de dados usando o algoritmo <i>k-means</i> (a) <i>clusters</i> iniciais (b) atualização dos <i>clusters</i> (c) <i>clusters</i> finais. Figura modificada de Han et al. (2012).	28
Figura 2.7. Identificação do número de <i>clusters</i> através do ponto <i>Elbow</i> (a) Homogeneidade (b) Heterogeneidade. Figura modificada de Kodinariya & Makwana (2013).....	29
Figura 2.8. Identificação do número de <i>clusters</i> através índice Calinski-Harabasz.	29
Figura 2.9. Área de influência de cada amostra pelo (a) método poligonal e (b) método do vizinho mais próximo. Figura modificada de Sinclair & Blackwell (2002).....	31
Figura 2.10. Tamanho de busca para amostras dentro (em vermelho) e fora (em azul) da região de busca, representado por uma elipse.	32
Figura 2.11. Efeito do elipsoide de busca e da escolha do número mínimo e máximo de amostras para a estimativa de um bloco.....	33
Figura 2.12. Efeito sobre a escolha do número de setores angulares para a estimativa de um bloco.	34
Figura 2.13. Efeito sobre a escolha do número de amostras por furo para a estimativa de um bloco (a) limitar amostras através dos setores angulares (b) limitar as amostras através do número de amostras por furo.	35
Figura 3.1. Mapa de localização das amostras do banco de dados 2D.	36
Figura 3.2. Estatística básica (a) Histograma de frequência (b) Histograma acumulado.	37
Figura 3.3. Domínios gerados no CASO I (a) zonas de alto teor (b) zona de baixo teor.	38

Figura 3.4. Domínios gerados no CASO II (a) zonas de alto teor (b) zona de baixo teor.....	38
Figura 3.5. Domínios gerados no CASO III (a) zonas de alto teor (b) zona de baixo teor.	39
Figura 3.6. Identificação do ponto <i>Elbow</i> (a) Heterogeneidade (b) Homogeneidade.	40
Figura 3.7. Índice Calinski-Harabasz.	40
Figura 3.8. (a) Mapa de localização dos teores (b) Mapa de localização dos domínios.....	41
Figura 3.9. Área de influência das amostras pelo método do vizinho mais próximo.....	43
Figura 3.10. Exemplo ilustrativo da minimização do erro absoluto para uma amostra, utilizando várias estratégias de krigagem.	44
Figura 3.11. Mapa de localização das amostras do banco de dados em 3D.	45
Figura 3.12. Histograma das amostras por domínio.	46
Figura 3.13. Mapa de localização das amostras (a) domínio 3 (b) domínio 4.....	46
Figura 3.14. Histograma de frequência das amostras (a) domínio 3 (b) domínio 4.	47
Figura 3.15. Área de influência das amostras pelo método do vizinho mais próximo (a) Domínio 3 (b) Domínio 4.....	49
Figura 4.1. Mapa de localização das amostras do Walker Lake Exaustivo (a) blocos de 1x1m (b) blocos de 5x5m.	50
Figura 4.2. CASO I: Curvas teor-tonelagem. Modela LKPO (linha vermelha), Modela LKNA (linha azul) e Modelo de referência (linha verde).....	54
Figura 4.3. CASO II: Curvas teor-tonelagem. Modela LKPO (linha vermelha), Modela LKNA (linha azul) e Modelo de referência (linha verde).....	54
Figura 4.4. CASO III: Curvas teor-tonelagem. Modela LKPO (linha vermelha), Modela LKNA (linha azul) e Modelo de referência (linha verde).....	55
Figura 4.5. CASO IV: Curvas teor-tonelagem. Modelo LKPO (linha vermelha), Modelo LKNA (linha azul) e Modelo de referência (linha verde).....	55
Figura 4.6. Caso I: (a) variações do erro na tonelagem (b) variações do erro no teor médio. Modelo da LKPO (linha vermelha), modelo da LKNA (linha azul) e modelo de referência (eixo X).....	56
Figura 4.7. Caso II: (a) variações do erro na tonelagem (b) variações do erro no teor médio. Modelo da LKPO (linha vermelha), modelo da LKNA (linha azul) e modelo de referência (eixo X).	57
Figura 4.8. Caso III: (a) variações do erro na tonelagem (b) variações do erro no teor médio. Modelo da LKPO (linha vermelha), modelo da LKNA (linha azul) e modelo de referência (eixo X).	57
Figura 4.9. Caso IV: (a) variações do erro na tonelagem (b) variações do erro no teor médio. Modelo da LKPO (linha vermelha), modelo da LKNA (linha azul) e modelo de referência (eixo X).	57

Figura 4.10. Curvas teor-tonelagem. Modelo LKPO (linha vermelha), Modelo LKNA (linha azul) e Modelos de referência obtida pela mudança de suporte (linhas verdes).....	58
Figura 4.11. Análise de deriva ao longo das principais direções (a) plano X (b) plano Y.	59
Figura 4.13. Análise visual (a) Dados Exaustivos (b) Modelo LKPO (c) Modelo LKNA.	60
Figura 4.13. Mapa de localização das amostras de <i>blasthole</i>	61
Figura 4.14. Modelo de referência do domínio 3, utilizando dados de <i>blasthole</i> (a) Mapa de localização (b) Seção x-x' (c) Histograma de frequências.	62
Figura 4.15. Modelo de referência do domínio 4, utilizando dados de <i>blasthole</i> (a) Mapa de localização (b) Seção x-x' (c) Histograma de frequências.	63
Figura 4.16. Curvas teor-tonelagem (a) domínio 3 (b) domínio 4.....	64
Figura 4.17. Variações do erro do domínio 3 (a) Tonelagem (b) Teor médio.....	65
Figura 4.18. Variações do erro do domínio 4 (a) Tonelagem (b) Teor médio.....	65
Figura 4.19. Domínios 3 e 4 unificados (a) Curvas teor-tonelagem (b) Variações do erro na tonelagem (c) Variações do erro no teor médio.....	66
Figura 4.20. Análise de deriva do domínio 3 ao longo das principais direções (a) plano X (b) plano Y (c) plano Z.....	68
Figura 4.21. Análise de deriva do domínio 4 ao longo das principais direções (a) plano X (b) plano Y (c) plano Z.....	69
Figura 4.22. Curvas teor-tonelagem. Modelo LKPO (linha vermelha), Modelo LKNA (linha azul) e Modelos de referência obtida pela mudança de suporte (linhas verdes).....	70
Figura 4.23. Tempos obtidos pelo algoritmo da LKPO (a) domínio 3 (b) domínio 4.	72
Figura 5.1. Validação dos modelos simulados (linhas verdes) através dos histogramas acumulados, usando como referência os dados desagrupados (linha vermelha) para o caso IV (a) domínio 1 (b) domínio 2 (c) domínio 3 (d) domínio 4 (e) domínio 5.....	74
Figura 5.2. Validação dos modelos simulados (linhas verdes) na direção de maior continuidade através de variogramas, usando como referência o modelo dos dados originais (linha vermelha) para o caso IV (a) domínio 1 na N0° (b) domínio 2 na N157° (c) domínio 3 isotrópico (d) domínio 4 isotrópico (e) domínio 5 na N157°.....	75
Figura 5.3. Validação dos modelos simulados através dos gráficos de acurácia para o caso IV (a) domínio 1 (b) domínio 2 (c) domínio 3 (d) domínio 4 (e) domínio 5.....	76
Figura 5.4. Caso IV utilizando simulações (a) Curvas teor-tonelagem (b) Variações do erro na tonelagem (c) Variações do erro no teor médio.....	78
Figura 5.5. Análise de deriva ao longo das principais direções (a) plano X (b) plano Y.	79
Figura 5.6. Curvas teor-tonelagem. Modelo LKPO utilizando simulações (linha vermelha), Modelo LKNA (linha azul) e Modelos de referência obtida pela mudança de suporte (linhas verdes).....	80

Figura 6.1. Análise visual (a) Dados Exaustivos (b) Modelo LKPO- <i>Cross-Validation</i> (c) Modelo LKPO- <i>Simulations</i>	81
Figura 6.2. Gráfico de dispersão entre o LKPO- <i>Simulations</i> e o LKPO- <i>Cross-Validation</i>	82
Figura 6.3. Curvas teor-tonelagem entre os modelos LKPO- <i>simulations</i> (linha azul) e LKPO- <i>Cross-Validations</i> (linha vermelha) comparadas com o modelo exaustivo (linha verde).....	82
Figura 6.4. Gráfico de dispersão do erro absoluto entre os modelos LKPO (<i>Cross-Validations</i>) e LKPO (<i>Simulations</i>).....	83
Figura 6.5. Curvas teor-tonelagem entre o modelo tradicional (linha azul) e LKPO- <i>Cross-Validations</i> (linha vermelha) comparadas com o modelo exaustivo (linha verde).....	84
Figura 6.6. Gráfico de dispersão entre o modelo tradicional e o modelo da LKNA.....	84
Figura 6.7. Gráfico de dispersão do erro absoluto entre os modelos (a) LKPO/Tradicional (b) LKPO/LKNA.....	85
Figura A-1. Correlogramas modelados nas três principais direções do domínio 3 (a) Maior: N112 (b) Intermediária: N22 (c) Menor: D90.....	96
Figura A-2. Correlogramas modelados nas três principais direções do domínio 4 (a) Maior: N45 (b) Intermediária: N135 (c) Menor: D90.....	97
Figura B-1. Aba geral do SGeMS.....	97
Figura B-2. Aba geral do algoritmo da LKPO implementado dentro do SGeMS.....	98
Figura C-1. CASO I: Modelos de continuidade espacial (a) domínio 1: N157 (b) domínio 1: N167 (c) domínio 2: omnidirecional.....	99
Figura C-2. CASO II: Modelos de continuidade espacial (a) domínio 1: N157 (b) domínio 1: N167 (c) domínio 2: omnidirecional.....	100
Figura C-3. CASO III: Modelos de continuidade espacial (a) domínio 1: N157 (b) domínio 1: N167 (c) domínio 2: omnidirecional.....	100
Figura C-4 parte I. CASO IV: Modelos de continuidade espacial (a) domínio 1: N0 (b) domínio 1: N90 (c) domínio 2: N157 (b) domínio 2: N67.....	100
Figura C-4 parte II. CASO IV: Modelos de continuidade espacial (a) domínio 3: omnidirecional (b) domínio 4: omnidirecional (c) domínio 5: N157 (b) domínio 5: N67.....	100

LISTA DE TABELAS

Tabela 1. Parâmetros estruturais dos modelos de continuidade espacial dos domínios para cada caso.	42
Tabela 2. Combinações das estratégias de busca.	42
Tabela 3. Parâmetros dos modelos de continuidade espacial para os domínios 3 e 4.	48
Tabela 4. Combinações das estratégias de busca no banco de dados 3D.	48
Tabela 5. Resumo estatístico por cada domínio no banco de dados 2D.	53
Tabela 6. Resumo estatístico por cada domínio no banco de dados 3D.	67
Tabela 7. Resumo das configurações básicas dos computadores.	71
Tabela 8. Número de amostras é número de blocos por domínio.	71
Tabela 9. Resumo estatístico para o banco de dados 2D.	77

RESUMO

A definição da estratégia de busca na krigagem, feita pelo usuário, pode ter impactos significativos nos resultados das estimativas. Domínios estacionários são geralmente estimados utilizando-se uma única estratégia de busca para todo o domínio. No entanto, utilizar uma estratégia de busca global, ou seja, interpolar todos os blocos dentro do domínio estacionário usando-se a mesma estratégia de busca, ignora completamente as variações locais dentro de cada domínio. Nesta tese, é proposta a otimização de parâmetros de krigagem localizada (LKPO, em inglês) como uma metodologia alternativa que considera as melhores configurações de parâmetros de krigagem para estimativas locais (bloco a bloco) baseada na minimização do erro absoluto de cada amostra.

Atualmente, a prática utilizada em estimativa de recursos e reservas minerais consiste, primeiramente, em dividir o depósito mineral inteiro em domínios que refletem zonas homogêneas e, em seguida, realizar as modelagens geoestatísticas e as etapas de estimativa dentro de cada um deles. Nessa primeira etapa, o *geomodeller* geralmente utiliza uma ampla variedade de informações geológicas (e.g., zonas de oxidação, litologias, limites estruturais, distribuição espacial de teores, etc.), visando caracterizar adequadamente o domínio geológico. Entretanto, quando são escassas essas informações, a distribuição dos teores é a única informação que pode ser utilizada para caracterizar aqueles domínios. Nesse sentido, foram testadas 4 alternativas diferentes para criar os domínios geológicos em um estudo de caso 2D, começando por uma simples análise visual até a utilização de métodos de agrupamento para análises de dados através do algoritmo *k-means++* (algoritmo usado para agrupar dados em padrões similares), cujo objetivo é particionar um conjunto de dados em um número determinado de *clusters* (domínios), utilizando somente a informação das amostras.

Para ilustrar a metodologia proposta, dois bancos de dados foram utilizados neste trabalho. O primeiro, é um depósito mineral sintético (2D), na qual foram utilizados como referência os dados exaustivos e métodos clássicos de mudança de suporte para realizar as respectivas comparações. O segundo, é um depósito mineral de ouro (3D), na qual foram utilizados como referência os dados provenientes de *blasthole* e os métodos clássicos de mudança de suporte. Uma ampla variedade de métodos de validação, tais como estatística básica, análise de deriva, curvas teor-tonelagem e análise de suavização da krigagem, demonstram que o uso de parâmetros de krigagem locais melhora significativamente o teor estimado, obtendo-se resultados mais precisos e acurados que as metodologias atualmente disponíveis na literatura geoestatística.

Por outra parte, no processo de otimização, tanto o número de combinações nas estratégias de busca como o fato de as estimativas serem locais (bloco a bloco) demandam grandes operações matemáticas, que, por sua vez, requerem um grande esforço computacional. Em razão disso, criou-se um algoritmo dentro da plataforma do SGeMS para facilitar ao usuário a implementação dessa metodologia. O desempenho em diferentes hardwares desse algoritmo desenvolvido evidenciou

que o aumento no número de amostras e a capacidade computacional têm um grande impacto nos tempos de execução do algoritmo da LKPO.

ABSTRACT

The definition of the search neighborhood (defined by the user) in kriging can have a significant impact on the resultant estimates. Stationary domains are usually estimated using a unique search strategy for the entire block model. However, the use of a global search neighborhood ignores the local variations within each domain, i.e., all blocks within a stationary domain are interpolated using a unique search strategy. Localized Kriging Parameter Optimization (LKPO) is proposed in this paper as an alternative methodology that considers the best “local estimation parameter settings” block by block.

Currently, the practice used to estimate mineral resources and reserves consists first of all, in dividing the entire mineral deposit into domains that reflecting homogeneous zones, and then performing the geostatistical modeling and estimation within each of them. In this first step, the geomodeller generally uses a wide variety of geological information (e.g., oxidation zones, lithologies, structural boundaries, grade spatial distribution, etc.), in order to adequately characterize the geological domain. However, when this information is limited, the grade distribution is the unique information that can be used to characterize those domains. In this regard, four different alternatives were tested to create the geological domains in a 2D case study, that are developed both by a simple visual analysis and by the use of advanced clustering methods for data analysis through the k-means ++ algorithm (algorithm used to group data into similar standards), which purpose is to partition a dataset into clusters using only the sample information.

To illustrate the proposed methodology, two datasets are presented in this work. The first is a synthetic mineral deposit (2D), using as reference the exhaustive data and change of support models for future comparisons. The second is a gold deposit (3D) using as reference the blasthole and change of support models. A wide variety of validation checks (e.g., swath plots, grade–tonnage curves, summary statistics and kriging smoothing analysis) show that the use of local kriging parameters improves significantly the grade estimation, obtaining more precise and accurate results than the methodologies currently available in the geostatistical literature.

On the other hand, in the optimization process the number of combinations in the search strategies and the estimates are located (block by block) require large mathematical operations, which in turn, require a great computational effort. because of this, an algorithm was created within the SGeMS platform to facilitate the user to implement this methodology. The performance in different hardware of this developed algorithm evidenced that the increase in the number of samples and the computational capacity have a great impact in the execution times of the LKPO algorithm.

CAPÍTULO 1 - INTRODUÇÃO

1.1. INTRODUÇÃO

O conceito de estacionariedade (definido formalmente por Matheron (1962-1963)) tem sido objeto de discussão por alguns autores ao longo do tempo. Para Philip & Watson (1986), significa que a média e a variância dos valores das amostras não dependem da localização. Eles questionam sobre como pode um depósito mineral ser assim quando nele existem diferentes zonas de mineralização (e.g., zonas de alto e baixo teor). No entanto, Myers (1989) adverte que os referidos autores interpretaram incorretamente o conceito de estacionariedade e basearam suas conclusões sobre essa definição incorreta. Segundo o autor, a estacionariedade refere-se à função aleatória e não aos dados. Por outro lado, muitos autores concordam que o conceito de estacionariedade está intimamente ligado ao termo de homogeneidade (termo usado pelos geólogos para caracterizar domínios com características geológicas semelhantes, e.g., tipos de mineralização). Para Rossi & Deutsch (2014), o conceito de “populações estatisticamente homogêneas” é denominado estacionariedade, ao passo que Sinclair & Blackwell (2002) usam o conceito de quase-estacionariedade (i.e., estacionariedade local) com o objetivo de separar estatística e espacialmente as zonas de alto e baixo teor dentro de um depósito mineral. Segundo Isaaks & Srivastava (1989), a estacionariedade local assumida por todos os métodos de estimativa é, muitas vezes, um pressuposto viável, mesmo em conjuntos de dados para os quais a estacionariedade global é claramente inadequada.

A prática atual em estimativa de recursos e reservas minerais consiste, primeiramente, em dividir o depósito mineral inteiro em domínios que refletem zonas homogêneas e, em seguida, realizar as modelagens geoestatísticas e etapas de estimativa dentro de cada um deles (Rossi & Deutsch (2014), Glacken & Trueman (2014), Duke & Hanna (2014)). Cada domínio tem os seus próprios parâmetros estruturais (contribuições do sill, alcance, anisotropia, etc.) e as suas estratégias de busca (número de amostras utilizadas dentro do elipsoide, divisão dos setores angulares, número de amostras por furo, etc.). No entanto, de acordo com Boisvert & Deutsch (2011), essa prática é razoável, se a direção de continuidade espacial muda suavemente sobre o domínio e a mudança ocorre em grandes distâncias. Por outro lado, Boisvert et al. (2009) comentam que, nos casos em que o comportamento espacial do atributo de interesse é curvilíneo (e.g., depósitos dobrados¹), uma única direção de anisotropia por domínio não é suficiente, ou seja, a hipótese de estacionariedade de segunda ordem (que assume que a função de covariância de uma variável aleatória é invariante em translação) é questionável, pois não existe uma única direção de continuidade que possa ser aplicada nesse tipo de domínios não-estacionários. Em razão disso, o uso da anisotropia local despertou muito interesse entre alguns autores ao longo do tempo.

¹ Nesses casos, pode ser utilizado a técnica de *unfolding* (McArthur, 1998; Deutsch, 2002)

Haas (1990) propôs um método para gerar um campo anisotrópico baseado na automatização do variograma e calculado a partir dos dados que estão dentro de uma janela móvel. A vantagem desse procedimento é a sua simplicidade. No entanto, ele reduz o número de amostras para cada cálculo variográfico e pode afetar a qualidade dos variogramas.

Isaaks (2004) desenvolveu um algoritmo para identificar as orientações das tendências estruturais controladas localmente em mineralização de alto teor. Essas orientações locais são realizadas bloco a bloco, contribuindo para uma seleção adequada dos dados no processo de interpolação. O algoritmo é conhecido como krigagem com Anisotropia Local (LAK, em inglês). O autor comenta que o LAK é baseado na diferença de vetores entre pares de amostras e sua implementação precisa de informações abundantes para obter resultados plausíveis. Diferente do autor anterior, Stroet & Snepvangers (2005) apresentaram um novo método de LAK. De acordo com esses autores, este método utiliza um algoritmo gradiente da análise de imagem (Van Munster et al., 1995) para estimar a anisotropia local (determinada apenas a partir dos dados), seguida de uma estimativa por krigagem. Depois, para cada bloco, o algoritmo determina as anisotropias usando uma área de busca em torno do mesmo e o processo de krigagem é realizado com essas anisotropias. Esse processo é realizado iterativamente.

Boisvert et al. (2009) implementaram um algoritmo conhecido como Anisotropia Variável Local (LVA). O algoritmo usa o campo de anisotropia local para inferir o variograma e modelar os teores usando distâncias não-euclidianas. Segundo os autores, o LVA permite determinar o caminho ideal entre os pontos que proporcionam a máxima covariância possível na presença de LVA. Uma vez que o melhor caminho foi verificado, os resultados da covariância podem ser usados no processo de interpolação ou simulação (e.g., krigagem ou Simulação Sequencial Gaussiana). Uma variação de LVA pode ser visitada em Machuca-Mory & Deutsch (2013). Magneron et al. (2010) propuseram uma nova metodologia chamada Geoestatística Móvel (M-GS, em inglês) com base na otimização local de parâmetros envolvidos em modelos baseados na variografia. De acordo com os autores, a M-GS considera que os parâmetros estruturais e a estratégia de busca são um conjunto de parâmetros interdependentes a serem otimizados localmente. Este processo é realizado durante uma fase de análise M-estrutural e conduz a um conjunto de parâmetros locais para cada modelo variográfico.

Outro aspecto importante a ser mencionado é a constante busca por modelos cada vez mais precisos e acurados nas estimativas de recursos minerais. A Análise da Vizinhança da Krigagem (KNA, em inglês) é uma das metodologias que buscam otimizar a estratégia de busca minimizando o erro entre o valor real e o valor estimado através de uma análise quantitativa da eficiência de krigagem (KE, em inglês) e/ou a inclinação da regressão (SR, em inglês). Entretanto, a maioria das pesquisas realizadas até o momento são baseadas na otimização da estratégia de busca global para cada domínio geológico, ou seja, todos os blocos dentro de um domínio estacionário são interpolados usando-se uma única estratégia de busca (mesmo que seja otimizada). Isaaks & Srivastava (1989) afirmaram que, embora seja uma prática comum definir uma única estratégia de busca para todos os blocos dentro do domínio estacionário, nem sempre esta é uma boa abordagem.

O que funciona em determinadas áreas de um determinado conjunto de dados pode não funcionar em outras. Por exemplo, na parte central de um conjunto de dados, pode ser possível escolher amostras relevantes simplesmente tomando todas as amostras que estão dentro do elipsoide de busca. Contudo, perto dos limites do domínio (do mesmo conjunto de dados), essa mesma estratégia pode ser inadequada.

Vann et al. (2003) argumentaram que a seleção de parâmetros de krigagem pode ter um impacto significativo na qualidade das estimativas. Uma busca com raio muito grande (alta continuidade espacial) pode gerar muitos pesos negativos (aumentando a possibilidade de gerar estimativas negativas) e/ou suavização excessiva. De outro modo, uma busca muito pequena (curta continuidade espacial) pode resultar em um viés condicional muito alto. A metodologia proposta pelos autores, conhecida como Análise da Vizinhança de Krigagem Quantificada (QKNA, em inglês), envolve alguns testes estatísticos consolidados na literatura geoestatística. O objetivo do QKNA é determinar a melhor combinação entre a estratégia de busca e o tamanho do bloco, minimizando o viés condicional. Os critérios considerados pelos autores para a avaliação da estratégia de busca são: SR, peso dado à média usando krigagem simples em cada bloco, variância da krigagem e distribuição dos pesos de krigagem (principalmente os pesos negativos).

Coombes (2008) resumiu, de forma estratégica, os estudos realizados sobre otimização de estratégias de busca. De acordo com a autora, a ideia principal é começar com a configuração máxima de todos os parâmetros envolvidos na estratégia de busca e reduzi-los para um único parâmetro, de modo que haja uma perda mínima na qualidade das estimativas. O objetivo desta abordagem é identificar um conjunto de parâmetros que resultem no maior SR e KE e garantam estimativas mais confiáveis.

Segundo Boyle (2010), a estratégia de busca é um parâmetro importante a ser considerado no processo de estimativa e pode ser determinada a partir do peso atribuído à média global ao usar krigagem simples na estimativa de um dado bloco. Ou seja, uma SR perto de um (para minimizar o viés condicional) e o peso da média deve ser próximo de zero (para definir uma busca mínima). O autor comentou que os resultados de uma KNA podem superestimar ou subestimar os benefícios entre uma estratégia de busca e outra, concluindo que há aspectos mais importantes a serem considerados do que simplesmente uma KNA, por exemplo, domínios das amostras, pressupostos de estacionariedade e acurácia na amostragem.

Abedini et al. (2012) propuseram uma krigagem ordinária baseada em algoritmos genéticos (GA, em inglês) com técnicas de vizinhança móvel para escolher as estratégias de busca apropriadas a serem incluídas no processo de estimativa. Os autores concluíram que o GA foi capaz de capturar a dinâmica do depósito estudado de maneira eficiente e argumentaram que a definição dos parâmetros de estratégia de busca baseada nas coordenadas geográficas e nos teores melhora significativamente as estatísticas de validação cruzada quando comparadas com aquelas baseadas apenas nas coordenadas geográficas.

Coombes & Boamah (2015) argumentaram que valores mais elevados de KE e SR implicam maior precisão no processo de estimativa com um nível adequado de suavização para o tamanho do bloco selecionado, bem como o número de amostras e a configuração delas entorno do bloco. Segundo os autores, a KE e SR fornecem indicadores úteis sobre a confiabilidade nas estimativas de cada bloco individualmente, propondo uma nova metodologia conhecida como Análise da Vizinhança da Krigagem Localizada (LKNA, em inglês) que funciona como uma extensão do KNA. A LKNA permite realizar a otimização local dos parâmetros das estimativas bloco a bloco. O fundamento desta metodologia é que cada bloco dentro do modelo de blocos é independente de todos os outros blocos do mesmo modelo, permitindo escolher as estratégias de busca localmente. Logo, cada bloco é livre para ser estimado com base no seu próprio conjunto de parâmetros otimizados. O processo de otimização consiste simplesmente em selecionar a estimativa associada à melhor combinação de parâmetros com maior KE e SR.

Jeannée et al. (2013) apresentaram a Geoestatística Local (LGS, em inglês) como uma possível solução para calcular modelos variográficos e/ou vizinhanças locais em um determinado conjunto de dados usando a técnica de validação cruzada. A ideia principal da otimização da LGS é aplicar, para cada bloco, as validações cruzadas de todas as amostras que estão situadas dentro de uma janela móvel (ou sua dilatação), realizando todas as variações respectivas entre o modelo variográfico e os parâmetros de krigagem. No final, o conjunto de parâmetros com o menor erro absoluto é armazenado para cada bloco e posteriormente utilizado no processo de estimativa. Ao repetir este processo iterativamente, todo o modelo de blocos é preenchido com os parâmetros locais que proporcionem a melhor estratégia local com base ao menor erro médio absoluto.

As principais desvantagens nos métodos KNA/QKNA é que o processo de otimização é baseado em uma única estratégia de busca global para todo o modelo de bloco (dentro do domínio estacionário), ignorando as variações locais (bloco a bloco) nela existentes. Todos os blocos são interpolados usando a mesma configuração de parâmetros na estratégia de busca. Por outro lado, a LKNA usa métricas ineficientes com base na otimização local, como, por exemplo, a KE e SR que são métricas que dependem da variância da krigagem e ignoram a variabilidade local dos teores. Além disso, uma alta eficiência na LKNA significa que a variância de krigagem é baixa, ou seja, muitas amostras devem ser consideradas, aumentando o risco de aparecerem pesos negativos que possam influenciar nas estimativas. Finalmente, a desvantagem na LGS é que o algoritmo é limitado ao número de amostras que estão dentro da janela móvel escolhida pelo usuário, permitindo que outros usuários possam ter escolhas diferentes desse número e, por conseguinte, outros resultados de otimização.

1.2. MOTIVAÇÃO

A técnica de estimativa conhecida como krigagem, considerado como o melhor estimador linear sem viés (BLUE, em inglês), é um termo genérico aplicado a uma série de metodologias de

estimativa (pontual ou em bloco) que minimizam a variância do erro (Sinclair & Blackwell, 2002) através da otimização dos pesos dados a cada amostra condicionante. Para garantir o melhor resultado da estimativa, é necessário escolher um conjunto de parâmetros (e.g. número mínimo e máximo de amostras, número de setores angulares, distância de busca, entre outros) adequados à técnica que está sendo aplicada. Se forem utilizados parâmetros diferentes, um novo modelo será gerado, com características locais possivelmente diferentes. Desse modo, várias questões podem ser levantadas:

- i. Quais seriam os critérios utilizados pelo usuário para escolher, dentre vários modelos estimados com distintos parâmetros, o melhor, se todos modelos seriam *a priori* aceitáveis?
- ii. E se existisse um terceiro modelo de estimativa que não tivesse sido testado pelo usuário e ele for mais preciso e acurado que os dois anteriores?
- iii. E se o modelo de estimativa escolhido não fosse criado apenas por uma estratégia “a melhor de todas”, mas sim por um conjunto de estratégias de krigagem (e.g. otimização local bloco a bloco) que seja mais preciso e acurado que este último escolhido pelo usuário?

Para resolver a primeira questão, podemos mencionar que existem várias alternativas para validar um modelo estimado e ele ser considerado como aceitável dentro dos critérios de cada usuário, entre elas estão:

- i. Reprodução da média global;
- ii. Reprodução das médias locais ou análise de deriva;
- iii. Presença de excessiva quantidade de pesos negativos na krigagem;
- iv. Validação visual;
- v. Verificação da suavização da krigagem;
- vi. Coeficiente de regressão;
- vii. Coeficiente de correlação entre o modelo estimado e o real usando validação cruzada;
- viii. Erro absoluto médio (MAE, em inglês) ou quadrático (MSE, em inglês), entre outros.

As métricas mencionadas anteriormente tendem a ser subjetivas e variam de um usuário para outro. Por exemplo, algumas pessoas aceitam um modelo estimado com uma tolerância de desvio da média global de até $\pm 5\%$ em relação aos dados desagrupados; outras, de até $\pm 3\%$. Em geral, não existe uma regra quantitativa que estabeleça os limites de permissibilidade para aceitar ou rejeitar cada uma das técnicas. O esperado seria, que o conjunto das técnicas utilizadas nas validações, leve a um modelo o mais preciso e acurado possível.

A segunda questão levantada é em relação ao número de tentativas de parâmetros diferentes usadas pelo usuário nas estimativas (que geralmente são poucas). Isso acontece, em alguns casos, pela própria complexidade do modelo a ser estimado, pelo tempo de entrega do modelo para as próximas etapas de planejamento de lavra que usam esse modelo ou até pelo próprio esforço computacional excessivo para que se processe muitos modelos alternativos. ***Seja qual for o motivo, não podemos descartar da ideia de que existe a possibilidade de encontrar modelos mais precisos e acurados utilizando parâmetros de estimativa diferentes aos que não foram testados pelo usuário.*** O interessante da última questão é que o modelo gerado como “o melhor de todos”

corresponde a uma única estratégia de estimativa para todo o modelo de blocos dentro de cada domínio estacionário. Não seria interessante se esse modelo não fosse gerado apenas por essa única estratégia, mas sim pelo conjunto de todas as estratégias utilizadas, gerando um novo modelo estimado bloco a bloco, cuja precisão e acuracidade fosse melhor que a escolhida inicialmente?

Imagine-se que a figura 1.1 representa um depósito mineral amostrado para um elemento qualquer (adaptado de Isaaks & Srivastava 1989), onde as amostras com baixo e alto teor são representadas pelas cores azuis e vermelhas-amarelas, respectivamente. Note, que existe um alto adensamento da malha amostral nas zonas ricas (alto teor) e pouco adensamento nas zonas pobres (baixo teor). A prática comum na estimativa de recursos minerais recomenda que primeiramente deve-se particionar todo o depósito em zonas com características ou propriedades homogêneas similares (domínios geológicos) e posteriormente realizar todo o processo de modelagem e estimativas dentro de cada domínio. No entanto, quando realizamos o processo de estimativa, é usual que, em cada domínio, se use uma única estratégia de busca para todo o modelo de blocos, ignorando as variações locais presentes em cada um deles.

Utilizar os mesmos parâmetros de estimativa dentro de toda a área que abrange cada domínio pode gerar estimativas tendenciosas. Por exemplo, muita informação (malha mais adensada) pode ser redundante e pode-se correr o risco de gerar estimativas negativas devido a distribuição dos pesos da krigagem que tendem a gerar pesos negativos. Por outro lado, pouca informação (malha menos adensada) pode gerar muita suavização e extrapolar teores, correndo o risco de estimar valores influenciados por dados extremos e superestimar o depósito. Esta tese foi inspirada precisamente na necessidade de criar uma otimização local (bloco a bloco) com todos os possíveis parâmetros de krigagem que ajude a minimizar o erro da estimativa (teor real vs. teor estimado) em cada bloco individualmente.

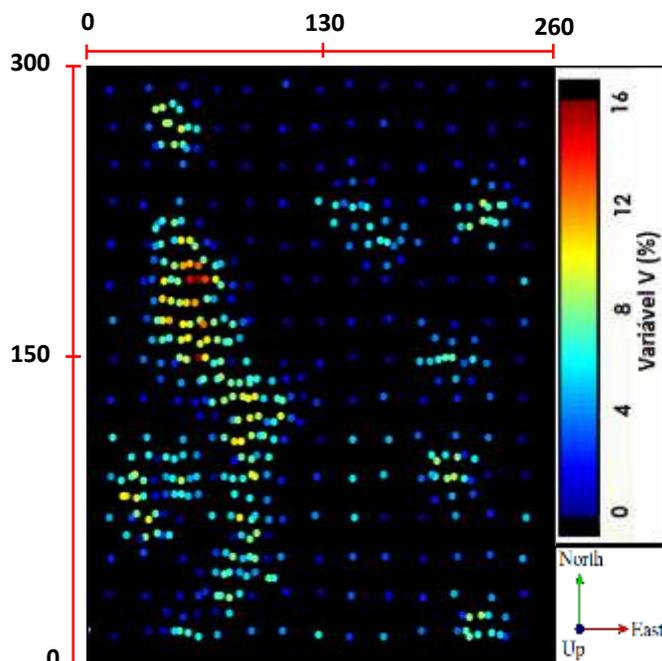


Figura 1.1. Mapa de localização das amostras.

1.3. META

Investigar se é possível uma metodologia que permita utilizar múltiplas estratégias de estimativas diferentes para capturar as variações locais (i.e., bloco a bloco) dentro de cada domínio estacionário, visando melhorar a precisão e a acuracidade dos modelos estimados.

1.4. OBJETIVOS

- i. Desenvolver uma técnica de otimização de parâmetros de krigagem que permita obter o menor erro local possível (i.e., bloco a bloco);
- ii. Investigar o desempenho desta metodologia em banco de dados sintéticos e em depósitos minerais reais;
- iii. Desenvolver um algoritmo que permita facilitar ao usuário o uso desta técnica;
- iv. Comparar os resultados obtidos pelos métodos tradicionais com a metodologia proposta, usando como referência o modelo exaustivo dos dados (quando dispor deles) ou modelos teóricos;
- v. Avaliar os resultados por meio dos métodos tradicionais de validações utilizados na geoestatística.

1.5. METODOLOGIA

A Otimização de Parâmetros de Krigagem Localizada (LKPO, em inglês) é apresentada neste trabalho como uma variação da LKNA, onde parâmetros locais são estabelecidos através da minimização de uma função (processo de otimização). Diferente da LKNA, que utiliza como métricas a eficiência de krigagem e o *slope regression* no processo de otimização, este trabalho propõe duas alternativas diferentes a essas métricas, uma é através da minimização do erro absoluto obtido a partir da distribuição do erro gerado na validação cruzada e, outra, através da minimização do erro quadrático obtido a partir da distribuição do erro gerado nas simulações.

1.5.1. LKPO UTILIZANDO VALIDAÇÃO CRUZADA

A LKPO utilizando validação cruzada é apresentada neste trabalho como uma variação da LKNA, onde cada bloco (dentro da área/volume de influência de cada amostra) é interpolado usando a estratégia de busca que forneça o menor erro absoluto na validação cruzada. Assumindo a estacionariedade intrínseca (um semi-variograma válido por domínio), o objetivo é investigar os benefícios do uso de vizinhanças locais para avaliar a precisão e a acuracidade bloco a bloco, considerando as melhores configurações de parâmetros locais de cada estimativa.

A construção de cada domínio estacionário será obtida através do algoritmo *k-means++* (algoritmo usado para agrupar dados em padrões similares) e métodos visuais. Obviamente,

existem outras alternativas para modelagem dos domínios, mas nesse estudo se desejam compartimentar o depósito em zonas (domínios) de uma forma rápida e expedita. Finalmente, para ilustrar a metodologia proposta, será utilizado um conjunto de dados sintéticos e um depósito mineral real, cujas estimativas são comparadas com o modelo de referência (e.g., dados exaustivos) e com o modelo obtido pela LKNA. A LKPO é definida, neste trabalho, através das seguintes etapas:

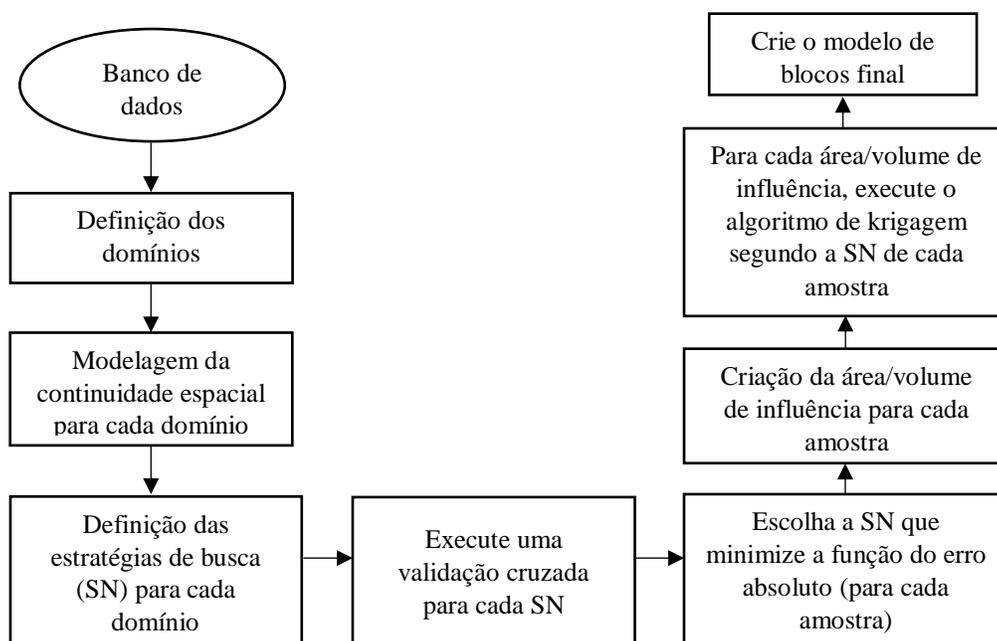


Figura 1.2. Fluxograma da metodologia da LKPO.

- i. *Definição dos domínios geológicos dentro do conjunto de dados, incluindo o modelo de blocos e o modelo de continuidade espacial para cada domínio.* Determinar domínios geológicos para uma análise comum é um passo importante na modelagem geoestatística e, talvez, a decisão mais crítica na estimativa de recursos e reservas minerais (Wilde & Deutsch, 2012). O processo de definição de domínios de estimativa envolve a combinação de um número significativo de parâmetros (e.g., zonas de oxidação, litologias, alterações, limites estruturais, distribuição espacial de teores, etc.) que devem ser considerados para caracterizar o domínio geológico (Rossi & Deutsch, 2014; Emery & Ortiz, 2005; Romary et al., 2012; Chanderman et al., 2017, Abzalov & Humphreys, 2002, Abzalov, 2016). Nesta primeira etapa, o geomodeller é livre para escolher as melhores técnicas para delinear os domínios geológicos que serão utilizados no processo de otimização. Seguidamente, realize o cálculo e modelamento da continuidade espacial para cada domínio estacionário.
- ii. *Definição do conjunto de estratégias de busca e execute uma validação cruzada para cada uma delas.* A Vizinhança de Busca (SN, em inglês) define quantas e quais amostras serão

incluídas dentro do processo de estimativa. A SN é representada graficamente por uma elipse (2D) ou elipsoide (3D) centrada no ponto que está sendo estimado e, a orientação/alcance é determinada pela anisotropia ou isotropia do modelo de continuidade espacial. Todos esses parâmetros são imputados pelo usuário. Logo, é fundamental definir, em primeiro lugar, o conjunto de estratégias de busca que serão utilizadas no processo de otimização. Dentro desses parâmetros, podemos mencionar, entre outros:

- número mínimo de amostras incluídas dentro do elipsoide;
- número máximo de amostras incluídas dentro do elipsoide;
- número de setores angulares em que o elipsoide é dividido;
- número máximo de amostras utilizadas por furo de sondagem;
- limites nos teores máximos (*capping*), entre outros.

Finalmente, após definir cada SN, execute uma validação cruzada para cada uma delas. Lembre-se, que o número de estimativas é proporcional ao número de SN definidas pelo usuário.

- iii. *Criação da área/volume de influência ao redor de cada amostra.* Conforme mencionado neste trabalho, cada bloco dentro do modelo de bloco é interpolado considerando as melhores configurações de parâmetros locais de estimativa. Para isso, precisamos definir quais blocos serão atribuídos a cada amostra. Utilizar-se-á o conceito de área de influência (2D) ou volume (3D) para definir o conjunto de blocos que pertencem a cada amostra utilizando o método do vizinho mais próximo (definido no Capítulo 2).
- iv. *Escolha dos parâmetros das estratégias de busca (para cada amostra) que forneçam o menor erro absoluto.* Na etapa 2, várias estimativas pontuais foram realizadas através da validação cruzada. Portanto, cada amostra possui uma distribuição de valores estimados associados a cada SN definido pelo usuário. Em primeiro lugar, é necessário calcular a distribuição do erro absoluto para cada amostra através da diferença absoluta entre o valor real e o valor estimado. Posteriormente, deve-se encontrar o valor mínimo desta distribuição e, finalmente, escolher os parâmetros da SN associados ao valor mínimo daquela distribuição. Neste trabalho, esse processo é designado de otimização de parâmetros de estimativa e é obtido através da minimização do erro absoluto:

$$SN_i \rightarrow mAE_i = \min(|Z_i - Z_1^*|, |Z_i - Z_2^*|, \dots, |Z_i - Z_n^*|) \quad i = 1, 2, \dots, n \text{ (equação 1)}$$

onde,

- Z_n^* são os valores estimados associados a cada SN para cada amostra;
- Z_i é o valor real associado à i -ésima amostra;
- mAE_i é o menor erro absoluto entre Z_i e Z_n^* ; e
- SN_i são os parâmetros de SN que minimizam a função mAE_i para cada amostra.

- v. *Para cada área/volume de influência é executado o algoritmo de krigagem de acordo com os parâmetros de SN de cada amostra obtidos no passo anterior. Conforme calculado nas etapas (iii) e (iv), cada amostra tem a sua própria área/volume de influência e seus próprios parâmetros de SN otimizados. Assim, as estimativas serão executadas individualmente dentro de cada área/volume usando os parâmetros de SN associados a cada uma delas. Foi utilizado o algoritmo de krigagem ordinária (Matheron, 1963).*
- vi. *Finalmente, para cada domínio, faz-se a criação de um modelo de estimativa global a partir de estimativas locais otimizadas no passo (v). Com base em todos os blocos otimizados localmente (bloco a bloco), será construído um modelo de blocos inteiro que representará cada domínio definido na etapa (i). A construção desse modelo é dada pelo conjunto das estimativas locais calculadas na etapa (v).*

1.5.2. LKPO UTILIZANDO SIMULAÇÕES

Nessa parte, será utilizado o conceito de simulação geoestatística como métrica para realizar todo o processo de otimização de parâmetros de krigagem. Diferente do caso anterior, que utiliza a validação cruzada como métrica para escolher os melhores parâmetros de krigagem bloco a bloco, a LKPO utiliza, neste caso, as simulações como métrica para criar, em primeiro lugar, o espaço de incerteza (bloco a bloco) e, posteriormente, escolher os melhores parâmetros cujas estratégias proporcionem, em média, o menor erro quadrático possível em relação às incertezas de cada uma das estimativas. As etapas que conformam a LKPO utilizando simulações, dado um conjunto de dados, são as seguintes:

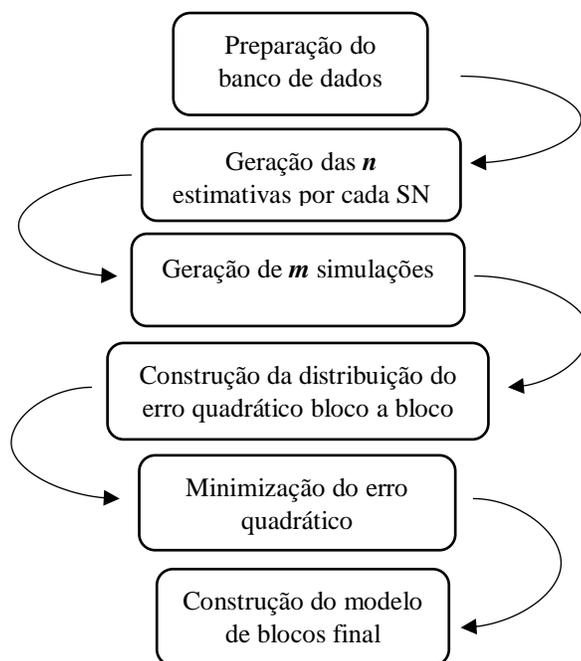


Figura 1.3. Etapas do processo de LKPO utilizando simulações.

- i. *Preparação do banco de dados.* Nessa primeira etapa, serão definidos todos os domínios geológicos dentro do conjunto de dados, incluindo o modelo de blocos e o modelo de continuidade espacial por cada domínio (similar à primeira etapa do capítulo 1.5.1).
- ii. *Geração das n estimativas por cada estratégia de busca.* Nessa segunda etapa, será realizada uma estimativa por cada estratégia de busca definida pelo usuário (diferente da segunda etapa do capítulo 1.5.1 que executa uma validação cruzada por cada estratégia de busca). O número de estimativas (n) é proporcional ao número de estratégias utilizadas no processo de otimização.
- iii. *Geração das m simulações.* Crie um conjunto de possíveis modelos equiprováveis (m) que permita capturar o espaço de incerteza existente nos nós do modelo. Para isso, será utilizada a simulação geoestatística, que permite gerar possíveis cenários e obter a distribuição dos valores da variável de interesse. Como cada simulação é gerada no suporte de pontos (figura 1.4a), é necessário realizar posteriormente um post-processamento de re-blocagem (*upscale*) para cada uma das simulações no mesmo suporte das estimativas geradas na etapa ii (figura 1.4b). *Upscale* é simplesmente a média aritmética dos pontos simulados que se situam dentro do suporte desejado.

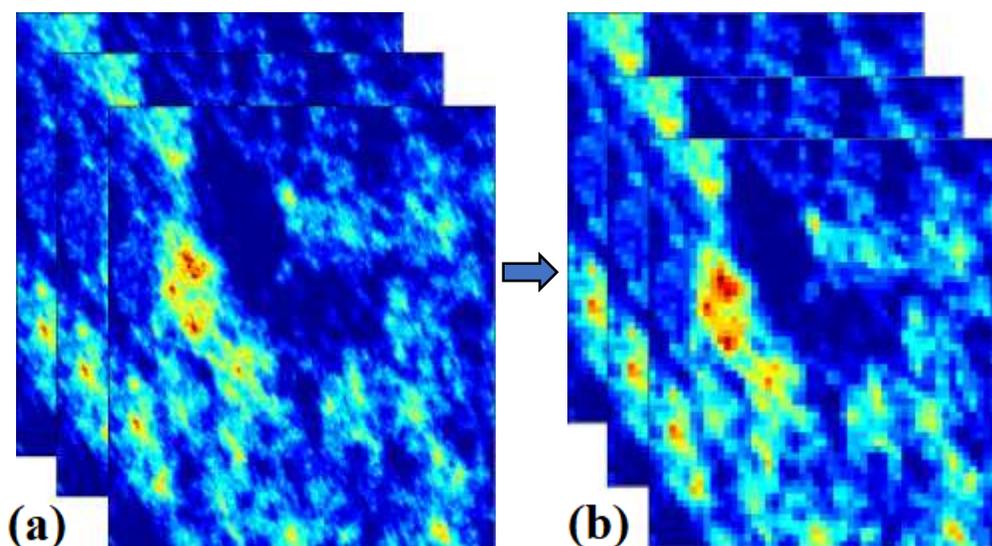


Figura 1.4. Simulações geoestatísticas (a) suporte de pontos (b) suporte blocos (re-blocagem).

- iv. *Construção da distribuição do erro quadrático bloco a bloco.* Para cada bloco dentro do modelo de blocos, é construído uma distribuição do erro quadrático (*QED*, em inglês) utilizando o teor estimado do bloco e todas as simulações pertencentes ao mesmo bloco (para cada estratégia de busca). Através da seguinte equação podemos obter a *QED* para cada bloco:

$$QED = ((Z_1^* - f(sim))^2, (Z_2^* - f(sim))^2, \dots, (Z_n^* - f(sim))^2) \quad i = 1, 2, \dots, n \text{ (equação 2)}$$

$$f(sim) = Z_{sim_1}^*, Z_{sim_2}^*, \dots, Z_{sim_m}^* \quad j = 1, 2, \dots, m \text{ (equação 3)}$$

onde:

- Z_i^* é o valor estimado do bloco associado à i -ésima estratégia de busca;
 - $f(sim)$ é distribuição dos modelos simulados;
 - $Z_{sim_j}^*$ é o valor simulado associado à j -ésima simulação, no mesmo suporte do Z_i^* .
- v. *Minimização do erro quadrático.* Uma vez obtida a distribuição do erro quadrático (bloco a bloco) para todas as estratégias de krigagem, em relação a todas as simulações, a escolha pelo teor estimado do bloco será o valor estimado associado aquela estratégia de busca que proporcione, em média, o menor erro quadrático possível, considerando às incertezas de cada uma das estimativas.

As figuras 1.5 a e b mostram um exemplo ilustrativo da distribuição do erro para dois blocos utilizando n estratégias de krigagem. Note que, para o bloco (#1), a estratégia que proporcionou a menor incerteza em relação a todas as simulações foi a número 4 (círculo em vermelho), já que a distribuição dos teores simulados proporcionou, em média, o menor erro quadrático. Portanto, o teor que minimiza o erro quadrático para o bloco (#1) é o valor correspondente à estratégia de krigagem (#4). De maneira similar, acontece para o bloco (#2), só que para este caso é a estratégia (#2) que fornece o menor erro quadrático médio, retendo novamente o valor estimado para esse bloco em particular. O mesmo procedimento é repetido para todos os blocos do modelo.

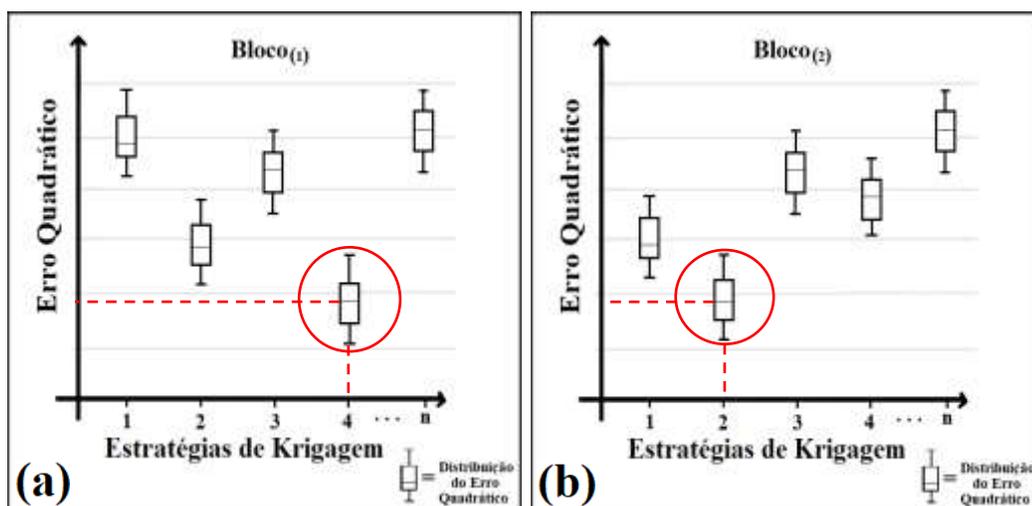


Figura 1.5. Minimização do erro quadrático em relação às incertezas de cada estratégia para o (a) bloco #1 (b) bloco #2.

- vi. *Finalmente, para cada domínio, faz-se a criação de um modelo a partir de estimativas locais otimizadas no passo (v). Com base em todos os blocos otimizados localmente (bloco a bloco), será construído um modelo de blocos completo que representará cada domínio definido na etapa (i). A construção desse modelo é dada pelo conjunto das estimativas locais calculadas na etapa (v).*

1.6. CONTRIBUIÇÃO DA TESE

A indústria de mineração sempre busca obtenção de maior precisão e acuracidade na geração dos modelos de teores nos depósitos minerais, devido principalmente à alta demanda no controle de qualidade dos teores. A maioria desses modelos são atualizados periodicamente tornando-se uma atividade laboral cíclica que pode gerar atrasos nas próximas etapas de planejamento.

A procura constante na construção de novos modelos de estimativa, cada vez mais precisos e acurados, tem sido objeto de pesquisa ao longo do tempo. Alguns desses modelos, baseados na análise da vizinhança da krigagem, buscam otimizar os parâmetros utilizados para interpolação por krigagem (ou estratégias de busca) que ajudem a minimizar o erro das estimativas (diferença entre o valor real e o valor estimado). Contudo, a grande maioria das pesquisas realizadas até o momento baseia-se na otimização dos parâmetros de krigagem globalmente, ou seja, uma única estratégia de krigagem usada para todos os blocos do modelo dentro de cada domínio estacionário. A LKPO, proposta neste trabalho, visa a otimizar esses parâmetros localmente (bloco a bloco) utilizando diferentes estratégias de busca que ajudem a minimizar o erro das estimativas.

Por outra parte, sabe-se que a geoestatística é uma ciência composta por um conjunto de técnicas apropriadas para prover uma estimativa desses modelos (krigagem ou cokrigagem) e, inclusive, sua incerteza (simulação). Portanto, um dos objetivos nos avanços da geoestatística é desenvolver metodologias que ajudem a melhorar o realismo dos modelos gerados. Esses avanços ajudariam a construir estimativas mais precisas e acuradas que permitiriam obter uma melhor previsão do modelo de recursos, uma melhor avaliação dos riscos e uma melhor informação para a tomada de decisões.

1.7. ESTRUTURA DA TESE

O primeiro capítulo introduz o conceito de estacionariedade e menciona como ele é importante no âmbito geoestatístico. Além disso, mostra como alguns autores ao longo do tempo despertaram interesse pelo impacto que produz, nos modelos estimados, as variações dos parâmetros de estimativa e definição de domínios estacionários. Por último, é definida a metodologia da LKPO através de uma série de passos e, ao final, é apresentado um fluxograma para facilitar ao leitor o desenvolvimento de cada um deles.

No capítulo 2, são definidos alguns conceitos gerais próprios do estado da arte e algumas métricas clássicas utilizadas na geoestatística para medir o desempenho da krigagem. Outros conceitos como agrupamento de dados utilizando técnicas de aprendizagem de máquinas e algoritmos mais comumente utilizados associados a essas técnicas, foram definidos neste capítulo com o intuito de mostrar ao leitor a forma como foram definidos os domínios estacionários nos casos de não ter nenhuma informação geológica. Finalmente, alguns tópicos relevantes sobre os parâmetros de krigagem com maior impacto nas estimativas, foram abordados nesse capítulo.

No capítulo 3 são descritos dois conjuntos de dados diferentes. O primeiro é um banco de dados (2D) consolidado na literatura; e, o segundo, um banco de dados (3D) provenientes de um depósito mineral real.

No capítulo 4, são apresentados os resultados e discussões desses dois estudos de casos na LKPO utilizando validação cruzada, com uma análise detalhada para cada conjunto de dados.

Analogamente, no capítulo 5, são apresentados os resultados e discussões do estudo de caso 2D na LKPO utilizando simulações.

O capítulo 6 apresenta as conclusões e as recomendações para trabalhos futuros.

CAPÍTULO 2 - CONCEITOS

Na primeira parte deste capítulo, definiremos alguns conceitos gerais da geoestatística e algumas das métricas mais utilizadas para quantificar o desempenho da krigagem. Entre elas: o conceito de estacionariedade; o viés condicional; coeficiente de regressão (*slope regression*); peso da média da krigagem simples; distribuição dos pesos negativos; variância da krigagem; eficiência da krigagem (*kriging efficiency*) e a precisão e viés local baseada na validação cruzada. Na segunda parte, serão abordados alguns conceitos utilizados no processo de otimização de parâmetros de krigagem, por exemplo, a definição dos parâmetros de krigagem que tem maior impacto nas estimativas, como são o número de dados condicionantes (número mínimo e máximo de amostras utilizadas no processo de interpolação) e o número de setores angulares em que é dividido o elipsoide de busca. Finalmente, serão definidos os conceitos básicos sobre a criação de domínios estacionários através da análise de agrupamento de dados utilizando o algoritmo de aprendizagem de máquinas (*k-means++*) e a criação de áreas/volumes de influência das amostras.

2.1. CONCEITOS GERAIS

2.1.1. ESTACIONARIEDADE

A geoestatística é baseada na teoria das variáveis regionalizadas (Matheron, 1963). Nessa teoria, são consideradas as características estruturais e aleatórias de variáveis espacialmente distribuídas. Segundo Journel & Huijbregts (1978), uma variável aleatória (RV, em inglês), a qual pode ser contínua ou categórica, é uma variável numérica que apresenta uma distribuição de probabilidade. No caso de variáveis contínuas (e.g., teor mineral), uma função de densidade de probabilidade (PDF, em inglês) pode ser utilizada para representar a probabilidade de todos os resultados possíveis da RV contínua (Boisvert, 2010). Finalmente, um conjunto de RVs podem definir uma função aleatória (RF, em inglês), ou seja, uma RF pode ser considerada como uma descrição probabilística da distribuição espacial de uma variável.

O conceito de estacionariedade pode ser difícil de entender, mas aqui está compreendido como uma relação estreita com o termo homogeneidade, usado pelos geólogos para caracterizar domínios de características geológicas semelhantes, tais como tipos de mineralização (Sinclair & Blackwell, 2002). Segundo Isaaks & Srivastava (1989), a estacionariedade é uma propriedade do modelo da RF e não é uma característica intrínseca da variável que está sendo analisada. Assumir que um domínio D é **estacionário de primeira ordem** implica que a média (m) de uma RF, $Z(u)$, dentro do domínio é constante em todas as localizações u , ou seja:

$$E\{Z(u)\} = m \quad \forall u \in D \quad (\text{equação 4})$$

A **estacionariedade de segunda ordem** assume que a função covariância, $C(h)$, de uma RV é invariante em translação, ou seja, $C(h)$ é constante dentro do domínio estacionário. A covariância entre dois pontos separados por uma distância h é dada pela seguinte equação:

$$E\{Z(u+h) \cdot Z(u)\} - m^2 = C(h) \quad \forall u \in D \quad (\text{equação 5})$$

A **hipóteses de estacionariedade intrínseca** de uma RF assume que o valor esperado existe (estacionariedade de primeira ordem) e, para todos os vetores h o incremento $Z(u+h) - Z(u)$ tem uma variância finita que não depende de u , ou seja:

$$\text{Var}\{Z(u+h) - Z(u)\} = E\{[Z(u+h) - Z(u)]^2\} = 2\gamma(h) \quad \forall u \in D \quad (\text{equação 6})$$

onde $2\gamma(h)$ representa a função do variograma, definida como a variância do incremento $[Z(u+h) - Z(u)]$.

Finalmente, de acordo com Journel & Huijbregts (1978), a função estrutural, a função covariância ou a função do variograma, são usadas apenas para distâncias de separação h limitadas ($|h| \leq b$), onde b representa, por exemplo, o limite do diâmetro da vizinhança de busca (geralmente determinado pelo alcance da continuidade espacial). Nos casos em que $|h| > b$, onde b pode ser descrito como a extensão de uma zona homogênea, $Z(u)$ e $Z(u+h)$ não podem ser mais considerados como vindo da mesma mineralização homogênea se $|h| > b$. Em tais casos, as funções citadas anteriormente, não são mais do que localmente estacionária (para toda distância $|h|$ menores ou iguais que o limite da vizinhança b). Esta limitação é conhecida na geoestatística como **hipóteses quase-estacionaria ou estacionariedade local**.

2.1.2. VIÉS CONDICIONAL

Segundo Vann et al. (2003), a definição da estratégia de busca na krigagem pode trazer impactos significativos sobre os resultados nas estimativas, gerando principalmente sérios problemas de viés condicional (Krige, 1994, 1996a, 1996b). O viés condicional² está quase sempre presente nas estimativas, devido ao efeito de suavização gerado pela krigagem (ou qualquer outro interpolador que utilizem pesos para as amostras) e o espaçamento largo entre as amostras. McLennan & Deutsch (2002), baseados nas equações de Journel & Huijbregts (1978), comentam que o viés condicional ocorre quando o valor esperado do teor real (Z_v) condicionado ao teor estimado ($Z_v^* = z$) não é igual ao valor do teor estimado (Z_v^*), ou seja:

² Uma análise do viés condicional e seu impacto na estimativa de recursos minerais é apresentada em Isaaks (2005).

$$E\{Z_v | Z_v^* = z\} \neq z \quad (\text{equação 7})$$

onde v representa algum volume de estimativa, por exemplo, a unidade seletiva de lavra (SMU, em inglês) e E representa a esperança matemática. A falta de informação em uma malha pouco adensada, a relação volume-variância e o efeito suporte implicam que a correlação entre as estimativas e o valor real seja imperfeita, superestimando os valores dos teores altos e sobrestimando os valores dos teores baixos. A figura 2.1 mostra a regressão linear entre Z_v^* e Z_v , originalmente proposta por Journel & Huijbregts (1978). Note, que a correlação entre Z_v^* e Z_v sempre é imperfeita em qualquer situação prática (i.e., $Z_v^* \neq Z_v$) e, geralmente a linha da regressão (Z_v/Z_v^* , linha contínua) será menos inclinada que a linha da regressão ($Z_v^* = Z_v$, linha tracejada).

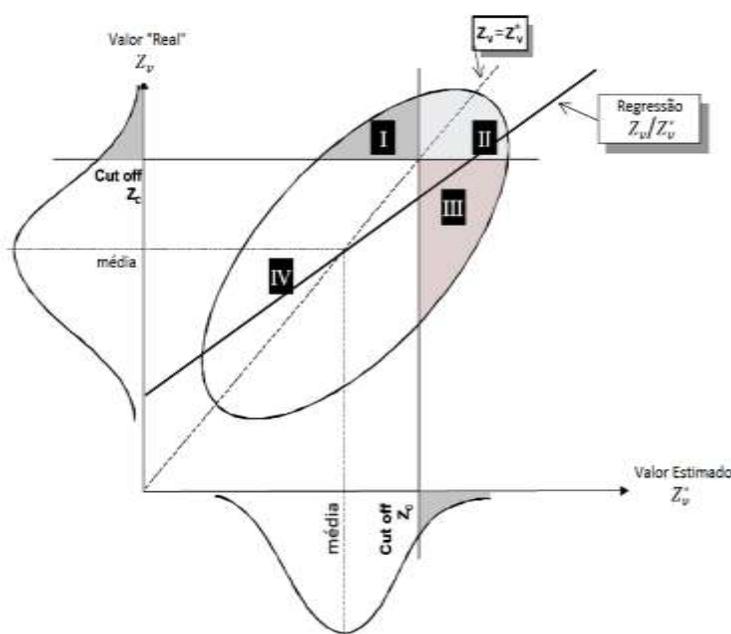


Figura 2.1. Efeito do viés condicional entre o valor real e o valor estimado. Figura modificada de Vann et al. (2003), originalmente proposta por Journel & Huijbregts (1978).

Vann et al. (2003) comentam que, embora na prática nunca conheçamos o teor real dos blocos, pode-se inferir a relação entre os valores reais dos blocos e os valores estimados sob certas premissas. Uma delas é que o variograma tenha sido modelado de forma confiável representando, de forma adequada, a continuidade espacial no domínio de interesse, ou seja, a hipótese de estacionariedade intrínseca é aceitável. Outra das premissas é que uma simples regressão linear pode capturar a essência da relação entre os teores reais dos blocos e os teores dos blocos estimados. A hipótese de estacionariedade intrínseca permite o cálculo de covariâncias/correlações entre quaisquer suportes específico, por exemplo, entre valores reais e estimados.

2.1.3. INCLINAÇÃO DA RETA DE REGRESSÃO PARA ESTIMATIVAS DE DEPÓSITOS MINERAIS

Com base nas premissas previamente citadas (variograma válido e o modelo linear de ajuste para a regressão), é possível calcular os principais parâmetros da reta de ajuste da regressão entre o teor real do bloco (Z_v) e o teor estimado (Z_v^*). A equação 8 representa o cálculo da inclinação da reta de regressão em termos de variância e covariância entre os blocos estimados e as amostras utilizadas.

$$a = \frac{Cov(Z_v, Z_v^*)}{Var(Z_v^*)} \quad (\text{equação 8})$$

onde:

- a = é a inclinação da reta de regressão;
- $Cov(Z_v, Z_v^*)$ = é a covariância entre o teor real do bloco (Z_v) e o teor estimado do bloco (Z_v^*);
- $Var(Z_v^*)$ = é a variância dos blocos estimados.

A variância e a covariância podem ser calculadas da seguinte maneira:

$$Var(Z_v^*) = \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j C(x_i, x_j) \quad (\text{equação 9})$$

$$Cov(Z_v, Z_v^*) = \sum_{i=1}^N \lambda_i \bar{C}(x_i, V) \quad (\text{equação 10})$$

onde $C(x_i, x_j)$ representa o valor da função covariância entre a amostra localizada na posição x_i e a amostra localizada na posição x_j e, λ_i e λ_j são os pesos da krigagem associados a essas amostras respectivamente. por outra parte, $\bar{C}(x_i, V)$ representa a média dos valores da função covariância entre as amostras localizadas nas posições x_i e o bloco a ser estimado V . Uma regressão linear perfeita deve ter como inclinação da reta de regressão um (1.0), ou seja $Z_v^* = Z_v$. Isto implica, que as nossas estimativas não apresentam viés condicional. A inclinação da reta de regressão e a sua interpretação são discutidas amplamente por Krige (1994; 1996a), Rivoirard (1987) e Deutsch (2007; 2014).

2.1.4. PESO DA MÉDIA DA KRIGAGEM SIMPLES

Na krigagem simples (SK, em inglês), uma parte dos pesos são atribuídos às amostras locais (amostras dentro da vizinhança de busca) e a outra parte para a média global sob a hipótese

de estacionariedade da média (assume-se média conhecida em todo o domínio). O peso que é atribuído à média global na SK mostra o quanto a krigagem é dependente das amostras locais. Se o peso da média estiver alto, significa que se tem poucas amostras ou amostras distantes do ponto estimado. Por outro lado, se o peso da média estiver baixo, significa que se tem muitas amostras ou amostras perto do ponto estimado. Segundo Boyle (2010), conforme a vizinhança de busca aumenta, a tendência do peso da média da SK é de diminuir (caso encontre novas amostras). Contudo, o mesmo autor recomenda que o incremento da busca seja até que a adição de novas amostras não impacte significativamente no peso da média da SK (decrementemente) ou até que haja presença de alguns pesos negativos nas estimativas. Quando os pesos negativos aparecerem, a tendência é que o peso sobre a média da SK aumente, deteriorando consigo a precisão e acurácia das estimativas, apesar das novas informações adicionadas.

O peso da média da SK proporciona uma ideia da qualidade da estratégia de busca usada em uma dada krigagem em relação à malha de amostragem disponível, já que ele pode ser visto como uma medida de fraqueza do efeito de filtragem dos dados mais próximos. As equações da SK foram definidas por Matheron (1963).

2.1.5. PESOS NEGATIVOS NA KRIGAGEM

Pesos negativos na krigagem surgem quando algumas amostras mais próximas do local a ser estimado filtram a influência de outras. Esta filtragem é conhecida na literatura como efeito eclipse (*screen effect*). Dependendo do variograma e da quantidade de amostras eclipsadas, os pesos negativos podem ter um impacto significativo nas estimativas. Sinclair & Blackwell (2002) citam vários exemplos sobre a influência e os problemas que podem causar os pesos negativos na krigagem:

- **Exemplo 1:** imaginemos que um bloco será estimado por 5 amostras e, uma delas, representa 1/3 do teor médio das outras, cujo peso atribuído para ela é de -0.1, portanto, a soma de todos os pesos da krigagem será de 1.1. Assumindo teores de 1 g/t e 3 g/t respectivamente, a média do teor estimado será de $(-0.1 * 1) + (1.1 * 3) = 3.2 \text{ g/t}$, um valor maior que as amostras utilizadas na estimativa. Portanto, o peso negativo atribuído a um teor baixo, neste caso, levou a obter um valor acima do maior valor entre os dados na vizinhança (superestimativa).
- **Exemplo 2:** consideremos um caso similar do exemplo 1, onde o peso negativo da krigagem de -0.1 é atribuído para a amostra cujo teor representa 3 vezes mais que o teor médio das outras, ou seja, 3 g/t e 1 g/t respectivamente. A soma dos pesos será de 1.1 e a média do teor estimado será de $(-0.1 * 3) + (1.1 * 1) = 0.8 \text{ g/t}$, um valor menor que de qualquer uma das amostras utilizadas na estimativa. Portanto, o peso negativo atribuído a um teor alto, neste caso, levou a uma subestimativa.
- **Exemplo 3:** consideremos um mesmo caso do exemplo 2, exceto que o peso negativo é aplicado a um valor extremo de 75 g/t. O teor médio da estimativa será de $(-0.1 * 75) +$

$(1.1 * 1) = -6.4 \text{ g/t}$, um valor fisicamente impossível de ser obtido, porém resultante matematicamente em uma estimativa.

- **Exemplo 4:** consideremos um mesmo caso do exemplo 3, exceto que o peso negativo aplicado ao mesmo valor extremo é muito pequeno (-0.01). O teor médio da estimativa será de $(-0.01 * 75) + (1.01 * 1) = -0.26 \text{ g/t}$, um valor fisicamente impossível em uma estimativa. Este resultado, se fosse ligeiramente positivo, poderia classificar minério como estéril.

A presença de pesos negativos na krigagem é matematicamente correta, mas resultados negativos não tem nenhum sentido físico, principalmente quando causado por um peso negativo atribuído a uma determinada amostra com um alto teor, já que esses pesos podem gerar estimativas negativas. Deutsch (1996) propõe um algoritmo para redefinir os pesos negativos da krigagem e corrigi-los para zero. Por outra parte, Van et al. (2003) comentam que os pesos negativos não são problemáticos, desde que eles representem uma pequena proporção do peso total (e.g., menos de 5%).

2.1.6. VARIÂNCIA DA KRIGAGEM

Isaaks & Srivastava (1989) definem a krigagem ordinária como o melhor estimador linear não-tendencioso (BLUE, em inglês). Melhor, porque minimiza a variância do erro da estimativa; linear, porque suas estimativas são combinações lineares ponderadas dos dados existentes e não tendencioso, porque a média do erro da estimativa (diferença entre o valor real e o valor estimado) é zero. Segundo Vann *et. al.* (2003), baseado nas equações de Matheron (1963), a expressão que minimiza a variância da estimativa, em termos de variograma, é calculada pela seguinte expressão:

$$\sigma_{OK}^2 = \sum_{i=1}^n \lambda_i \bar{\gamma}(x_i, V) - \bar{\gamma}(V, V) + \mu \quad (\text{equação 11})$$

onde σ_{OK}^2 é a variância de krigagem; λ_i são os pesos da krigagem; $\bar{\gamma}(x_i, V)$ é a média dos valores do variograma entre as amostras e o bloco que está sendo estimado; $\bar{\gamma}(V, V)$ é a média dos valores de variograma dos dados que estão sendo usados na estimativa do bloco dentro do domínio estacionário e μ é o parâmetro lagrangeano. A variância de krigagem é vista como uma métrica deficiente para avaliar a qualidade das estimativas, já que ela não considera os valores dos teores das amostras em seu cálculo, mas ela permite ter uma clara ideia da densidade amostral e a qualidade da sua disposição espacial em relação ao local sendo estimado.

2.1.7. EFICIÊNCIA DE KRIGAGEM

Segundo Deutsch (2014), a eficiência da krigagem (KE, em inglês) foi introduzida por Krige (1997) como uma medida de eficiência da qualidade da estimativa dos blocos. A KE é expressa como a diferença entre a variância real dos blocos (σ^2) e a variância da krigagem (σ_K^2), normalizada pela variância dos blocos reais, medida em porcentagem (%), ou seja:

$$KE(\%) = \frac{\sigma^2 - \sigma_K^2}{\sigma^2} = \frac{\bar{C}(v, V) - \sigma_K^2}{\bar{C}(v, V)} \quad (\text{equação 12})$$

onde $\bar{C}(v, V)$ representa a covariância média dos pontos (v) dentro dos blocos³ (V). Deutsch (2014) comenta que, uma alta eficiência significa que a variância de krigagem é baixa, e a variância dos blocos estimados é aproximadamente igual à variância dos blocos reais. Por outro lado, uma baixa eficiência significa que a variância de krigagem é alta em relação à variância do bloco. A variância de krigagem varia de bloco a bloco, de modo que a eficiência de krigagem variará também.

2.1.8. VALIDAÇÃO CRUZADA

Segundo Sinclair & Blackwell (2002), a validação cruzada é um procedimento amplamente utilizado na geoestatística na qual os dados pontuais são extraídos individualmente do banco de dados e, posteriormente, para cada ponto retirado é realizado uma estimativa pontual pelo grupo de amostras que se encontram na sua vizinhança. Desse modo, uma distribuição do erro é gerada a partir do cálculo do erro da estimativa. A técnica de validação cruzada permite, através da comparação entre valores reais e estimados, escolher entre diferentes procedimentos de estimativa, diferentes estratégias de busca e/ou diferentes modelos variográficos. Os resultados são usados para comparar as distribuições do erro de estimativa dos diferentes procedimentos de estimativa selecionados (Isaaks & Srivastava, 1989).

O histograma do erro é um dos gráficos mais importantes que podem ser obtidos através desta técnica e ele é formado pela distribuição dos valores do erro entre o valor estimado e o valor real, cujo enviesamento pode ser avaliado através da média aritmética dos valores dessa distribuição⁴. Por exemplo, se a média do histograma do erro é maior ou menor que zero, podemos dizer que as estimativas estão enviesadas positiva ou negativamente (figuras 2.2 a e b). Por outra

³ $\bar{C}(v, V)$ pode ser calculada a partir de fórmulas tradicionais da relação de Krige (Journel & Huijbregts, 1978, pp. 61-67).

⁴ Outros tipos de distribuições do erro de estimativas são analisados mais amplamente por Isaaks & Srivastava (1989), pag. 261-264.

parte, uma estimativa não enviesada, normalmente apresenta-se quando a média do erro fica próxima de zero (ver Figura 4c).

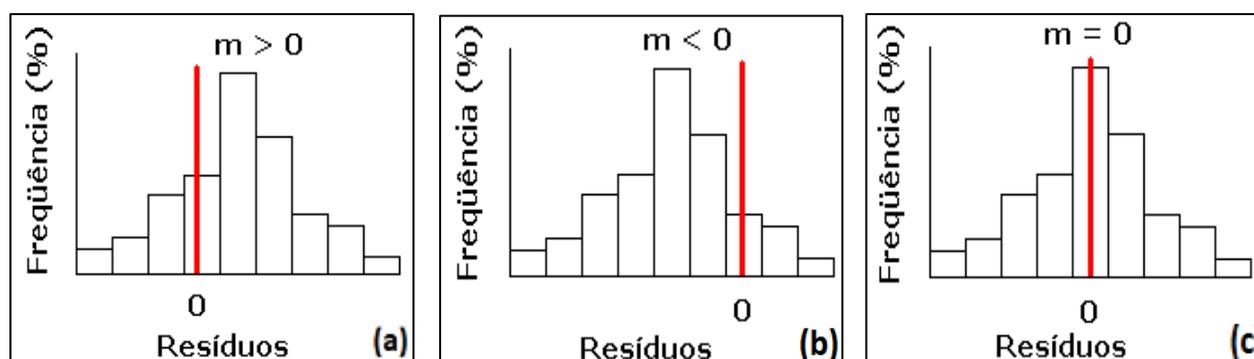


Figura 2.2. Histograma do erro obtido por validação cruzada (a) viés positivo (b) viés negativo (c) sem viés. Figura modificada de Isaaks & Srivastava (1989).

2.1.9. ERRO DAS ESTIMATIVAS

Isaaks & Srivastava (1989) definem o erro de cada ponto estimado como a diferença entre o valor estimado e o valor real:

$$Erro = r = Z_v^* - Z_v \quad (\text{equação 13})$$

onde Z_v representa o valor real e Z_v^* representa o valor estimado. Note, que o erro (r) pode apresentar-se em três formas diferentes:

- i. erro positivo ($Z_v^* > Z_v$), o que indica que houve uma superestimativa do valor estimado em relação ao valor real;
- ii. erro negativo ($Z_v^* < Z_v$), o que indica que houve uma subestimativa do valor de estimado em relação ao valor real;
- iii. sem erro ($Z_v^* = Z_v$), o que indica que tanto o valor real como o valor estimado são iguais.

Como já mencionamos anteriormente, a ideia principal da LKPO é a minimização do erro e, sabendo que erros podem ser negativos em alguns casos, utilizaremos a função do erro absoluto (AE, em inglês) que permite que a distribuição do erro esteja sempre definida positivamente ou igual a zero, ou seja:

$$AE = |Z_v^* - Z_v| \quad (\text{equação 14})$$

onde:

$$AE \geq 0 \quad (\text{equação 15})$$

2.1.10. SIMULAÇÃO GEOESTATÍSTICA

Segundo Sinclair & Blackwell (2002), uma simulação não é uma estimativa, ela é considerada um conjunto de valores (em duas ou três dimensões) com a mesma natureza estatística e espacial dos dados originais. Essas estatísticas podem ser representadas pelos histogramas e variogramas que variam ligeiramente em cada simulação.

A simulação seqüencial em cada local (u) está condicionada basicamente por duas características principais: primeiro, aos valores dos dados originais (amostras) e segundo, aos valores previamente simulados. Segundo comentado por Koppe (2009), a simulação seqüencial funciona da seguinte maneira: no primeiro local a ser simulado, é construída uma distribuição acumulada para valores do atributo em estudo. Essa distribuição é construída condicionada aos valores dos dados originais. Sorteia-se, então, um valor dessa distribuição, que corresponde ao valor simulado para esse primeiro local. No segundo local a ser simulado, é construída, novamente, uma distribuição acumulada para valores do atributo. Essa distribuição é construída condicionada aos valores dos dados originais e ao valor simulado anteriormente. Sorteia-se, então, um valor dessa distribuição, que corresponde ao valor simulado nesse segundo local. A simulação dos próximos locais ocorre da mesma maneira em cada novo local a ser simulado.

Um conjunto de realizações fornece uma medida de incerteza, resultante do nosso conhecimento imperfeito sobre a distribuição no espaço do parâmetro ou variável de trabalho. Essa distribuição de possíveis cenários é chamada de espaço de incerteza. As discrepâncias entre as realizações e o modelo adotado são conhecidas como flutuações ergódicas. A figura 2.3 mostra as flutuações ergódicas (variogramas) correspondentes a 20 realizações (linhas azuis) em relação ao modelo espacial de referência (linha preta).

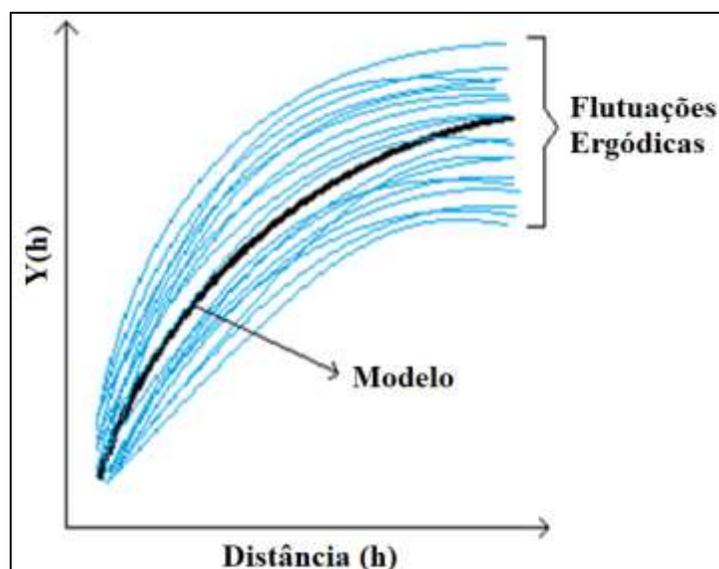


Figura 2.3. Flutuações ergódicas nos variogramas. Imagem modificada do (Deutsch & Journel, 1998).

Sinclair & Blackwell (2002) comentam, que as flutuações locais numa distribuição espacial dos dados são suavizadas pelos métodos tradicionais de estimativas, de tal forma que essas estimativas não refletem as variações locais dos dados. Portanto, em um processo de interpolação, estamos preocupados pela minimização da variância do erro da estimativa; enquanto, nas simulações nos preocupamos pela reprodução da dispersão da variância dos dados originais. Existem vários algoritmos de simulação geoestatística na literatura, entre eles estão a Simulação Sequencial Gaussiana (Deutsch & Journel, 1998), a Simulação Annealing (Deutsch, 1992), a Simulação por Caminhos Aleatórios (Ribeiro et al., 2012), a Simulação por Bandas Rotativas (Chilès & Delfiner, 1999), entre outros. Neste trabalho, será utilizado a Simulação por Bandas Rotativas.

O algoritmo de simulação por Bandas Rotativas foi desenvolvido por Matheron (1973). Este método, usado pela primeira vez por Chentsov (1957) no caso especial para funções aleatórias brownianas, permite gerar um grande número de simulações independentes definidas por linhas (bandas) varrendo o plano (ou espaço). Segundo Chilès & Delfiner (1999), o valor da simulação no ponto x do plano é a soma dos valores assumidos nas projeções de x nas diferentes linhas através das simulações unidimensionais correspondentes. Considerando um sistema de n_D linhas varrendo o plano regularmente a partir da origem no espaço (figura 2.4), o ângulo entre duas linhas adjacentes é π/n_D .

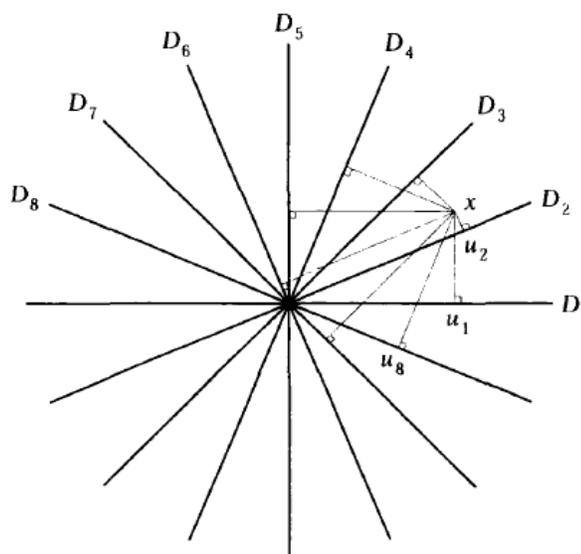


Figura 2.4. Princípio das bandas rotativas em 2D (imagem extraída de Chilès & Delfiner,1999).

onde:

- $\theta_t \in [0, \pi]$ é o ângulo da linha D_t com o eixo x .
- u_t é o vetor unidade de D_t com os componentes $\cos\theta_t$ e $\sin\theta_t$
- s_t é a abscissa em D_t centrado na origem

As simulações não condicionais independentes $S_t(s_t)$ com média zero e com covariância $C_1(h)$ são associadas às linhas D_t . Considerando um ponto $x = (x, y)$ no plano, a sua projeção em D_t é um ponto com abscissa:

$$s_t = \langle x, u_t \rangle = x \cos\theta_t + y \sin\theta_t \quad (\text{equação 16})$$

Portanto, a simulação no ponto x é definida da seguinte maneira:

$$Z^*(x) = \frac{1}{\sqrt{n_D}} \sum_{i=1}^{n_D} S_t(s_t) \quad (\text{equação 17})$$

onde:

- $Z^*(x)$ = Valor simulado no ponto x
- n_D = Número de linhas que dividem o plano
- S_t = Valor simulado no local s_t
- s_t = Local da projeção s_t na n -ésima linha que divide o plano.

Com as simulações elementares sendo independentes e com a mesma covariância $C_1(h)$, a covariância generalizada de $Z^*(x)$ é dada por:

$$C_2(x) = \frac{1}{n_D} \sum_{i=1}^{n_D} C_1(\langle h, u_t \rangle) \quad (\text{equação 18})$$

O método de simulação por bandas rotativas pode ser realizado de maneira não-condicional (reproduzindo o modelo de covariância) e de maneira condicional (reproduzindo os dados amostrais nas suas localizações). Para a realização da simulação condicional, são necessárias duas etapas: primeiro, os valores obtidos em cada local reproduzem o modelo de covariância dos dados amostrados e segundo, esses valores obtidos são então condicionados aos dados (Marques, 2010). Segundo Journel (1974), o condicionamento dos valores simulados pode ser realizado da seguinte maneira:

$$Y_c^s(u_1) = Y_k(u_1) + [Y_k^s(u_1) - Y^s(u_1)] \quad (\text{equação 19})$$

onde:

- $Y_c^s(u_1)$ = simulação de Y no local u_1 , condicionada aos dados (normais);
- $Y_k(u_1)$ = estimativa de krigagem de Y no local u_1 , condicionada aos dados;
- $Y_k^s(u_1)$ = estimativa de krigagem de Y no local u_1 , condicionada aos valores simulados (não-condicionais) nos locais dos dados;
- $Y^s(u_1)$ = simulação não-condicional de Y no local u_1 .

2.1.11. GRÁFICOS DE ACURÁCIA

A validação cruzada é uma técnica tradicional utilizada para testar parâmetros de estimativas (vizinhança/variograma) na maioria dos casos usando krigagem, mas também pode ser usada para checar o espaço de incerteza das simulações. Caxeta et al., 2015 comentam que, a técnica consiste em remover um ou um conjunto de amostras e re-estimar (ou re-simular) o local (u) usando os dados remanescentes. Em seguida, o conjunto de valores obtidos é comparado com os valores reais. Segundo os autores, uma boa maneira de analisar o conjunto de valores simulados em um local (u) é através de gráficos de acurácia (Deutsch, 1996; Pyrcz & Deutsch 2014). Em termos de simulação, ela é expressa por um conjunto de intervalos de probabilidade p na abscissa, e, a proporção de valores verdadeiros dentro de cada intervalo de probabilidade $\xi(p)$ no eixo Y. Se os pontos caírem acima da linha de 45° indica, que os modelos simulados são acurados (mas imprecisos). Por outra parte, se os pontos caírem abaixo desta mesma linha indica, que esses mesmos modelos são pouco acurados e imprecisos. Se os pontos caírem na linha de 45° indica, que os modelos simulados são, simultaneamente, precisos e acurados. A figura 2.5 mostra um exemplo do gráfico de acurácia.

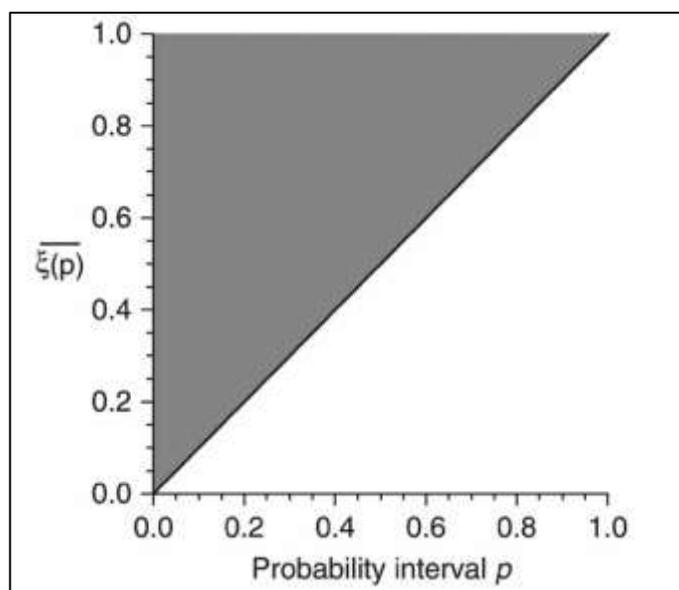


Figura 2.5. Gráfico de acurácia (imagem extraída de Pyrcz & Deutsch 2014).

As notações de intervalos de probabilidade são como de costume. O $p = 0.9$ representa o intervalo entre os quantis 0.05 e 0.95; o $p = 0.5$ é o intervalo entre os quantis 0.25 e 0.75 e assim por diante. Elas podem ser definidas pela média e variância gaussianas (somente para SGS) ou pela construção de uma distribuição de frequências acumulada (cdf) a partir dos valores simulados. A verificação é realizada para cada intervalo de probabilidade por meio de indicadores, definindo 1 para o ponto se o valor verdadeiro estiver dentro desse intervalo ou 0 se não.

Então, a média dos indicadores é plotada em relação ao seu intervalo de probabilidade no gráfico de acurácia. Deutsch (1996) propôs algumas medidas para quantificar a exatidão e precisão dos modelos simulados. O usado aqui é o *Goodness* (equação 20), que é uma medida da saída dos pontos da linha de 45°, representando uma melhor ou pior medida se os pontos forem precisos ou imprecisos, respectivamente. Ela varia de [0,1], sendo 1 a máxima acurácia.

$$G = 1 - \left[\int_0^1 a(p)[\bar{\xi}(p) - p]dp + 2 \int_0^1 [1 - a(p)][p - \bar{\xi}(p)]dp \right] \quad (\text{equação 20})$$

onde p representa os intervalos de probabilidade, $a(p)$ a proporção de valores acurados dentro do intervalo p e $\bar{\xi}(p)$ é a proporção de valores verdadeiros dentro do intervalo p .

2.2. CONCEITOS ESPECÍFICOS PARA A METODOLOGIA PROPOSTA

2.2.1. AGRUPAMENTO DE DADOS

Agrupamento (*clustering*) é um método de aprendizagem de máquina não supervisionado para análise de dados utilizado em vários campos do conhecimento, incluindo mineração de dados, reconhecimento de padrões, análise de imagem, recuperação de informações e bioinformática (Hastie, 2009). Existem vários métodos de agrupamento na literatura (Han *et al.* 2012; Aggarwal & Reddy 2013; Witten *et al.* 2016). No entanto, de acordo com Fouedjio *et al.* (2017), os métodos clássicos de agrupamento desconsideram as informações espaciais e tendem a produzir grupos de amostras espacialmente dispersos, o que certamente não é desejável do ponto de vista da modelagem geológica. Para tratar este problema, foram desenvolvidas diversas técnicas que utilizam o agrupamento aplicado a dados geoestatísticos (Romary *et al.* 2012, 2015; Fouedjio 2016a,b,c, 2017a,b; Fouedjio *et al.* 2017).

2.2.2. K-MEANS

O algoritmo de *k-means* é um dos métodos de agrupamento amplamente utilizado sendo ele um método iterativo simples usado para particionar um conjunto de dados em um número determinado de *k-clusters* especificado pelo usuário. O algoritmo começa escolhendo um número representativo de k -pontos como centroides iniciais (ponto central) e, posteriormente, cada amostra é associada ao centroide mais próximo com base numa medida de proximidade (e.g., distância euclidiana). Uma vez que os *clusters* são formados, o centroide de cada *cluster* é atualizado. O algoritmo repete iterativamente este procedimento até que os centroides não mudem ou qualquer outro critério de convergência seja cumprido (Aggarwal & Reddy, 2013). O objetivo do *k-means*

é minimizar as distâncias entre os pontos pertencentes ao mesmo *cluster* e maximizar as distâncias entre *clusters* diferentes (figura 2.6).

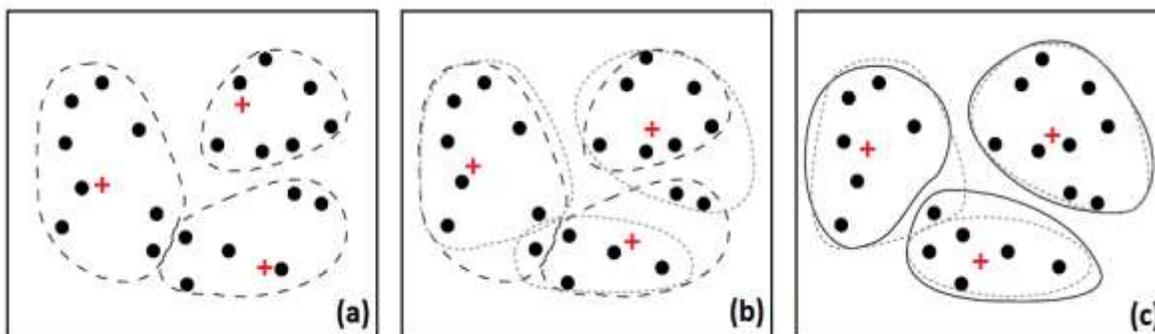


Figura 2.6. Agrupamento de dados usando o algoritmo *k-means* (a) *clusters* iniciais (b) atualização dos *clusters* (c) *clusters* finais. Figura modificada de Han et al. (2012).

No entanto, de acordo com Han *et al.* (2012) a convergência para o ótimo global neste método não é garantida e, muitas vezes, termina em um ótimo local. Os resultados dependem da seleção aleatória inicial dos centros dos *clusters* e, para obter bons resultados na prática, é comum executar o algoritmo *k-means* várias vezes com diferentes centros de *clusters* iniciais. Na literatura, existem alguns esforços para melhorar o algoritmo tradicional de *k-means* (Pelleg & Moore, 1999; Kanungo *et al.* 2002). Aggarwal & Reddy (2013) mencionaram outros métodos de inicialização do *k-means* que foram usados com sucesso para melhorar o desempenho de cada cluster. Um desses métodos é o algoritmo de agrupamento *k-means++* (Arthur & Vassilvitskii, 2007) que seleciona cuidadosamente os centroides iniciais do agrupamento *k-means*. Este algoritmo segue uma abordagem baseada em probabilidade em que o primeiro centroide é selecionado através de uma função aleatória e, a seleção do próximo, é dada pela maximização da distância do centroide atualmente selecionado. Essa seleção é determinada através de um índice de probabilidade ponderada. Neste trabalho, utilizaremos o algoritmo *k-means++* para determinar os domínios estacionários.

2.2.3. DETERMINANDO O NÚMERO DE CLUSTERS

O problema de estimar corretamente o número de *clusters* (domínios) é um dos principais desafios dos *k-means++*. Existem vários métodos na literatura para tratar esse problema (Aggarwal & Reddy, 2013; Kodinariya & Makwana, 2013), um deles é o método *Elbow* (cotovelo). *Elbow* é um método de interpretação e validação da coerência dentro da análise de agrupamento, projetada para ajudar a encontrar o número apropriado de *clusters* dentro de um conjunto de dados. Este método permite avaliar como a homogeneidade ou heterogeneidade dentro dos *clusters* varia para o valor de cada cluster. Para representar a homogeneidade (figura 2.7a), é utilizada a inércia (soma dos quadrados das distâncias dentro dos *clusters*) e para representar a heterogeneidade (figura 2.7b) é utilizada a variância (soma dos quadrados das distâncias entre os *clusters*). No ponto *Elbow*

(homogeneidade e heterogeneidade não tem ganhos significativos), é possível escolher um número adequado de *clusters*.

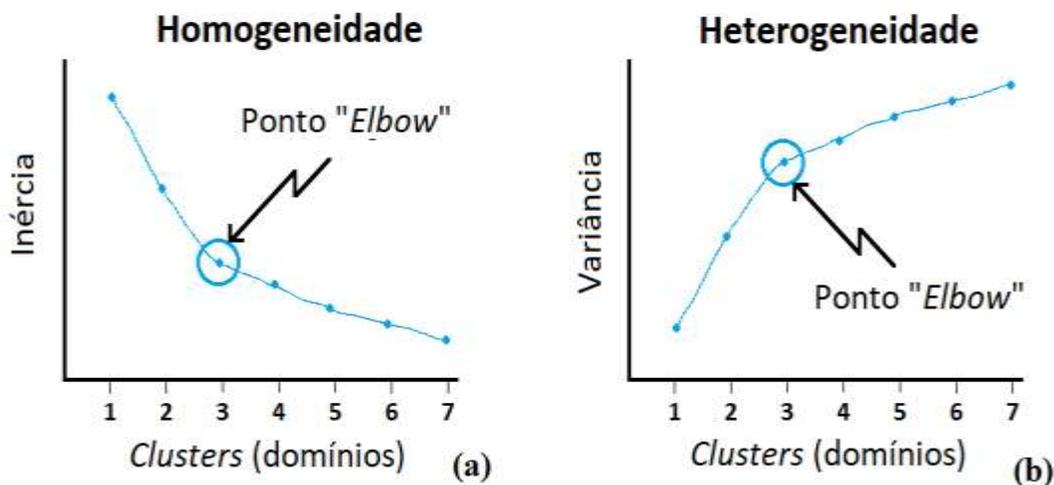


Figura 2.7. Identificação do número de *clusters* através do ponto *Elbow* (a) Homogeneidade (b) Heterogeneidade. Figura modificada de Kodinariya & Makwana (2013).

Outro método para escolher o número ótimo de *clusters* é o índice de Calinski-Harabasz (Calinski and Harabasz 1974), que é a relação das variações entre e dentro dos *clusters*. Quanto melhor o agrupamento, maior a variação entre *cluster* (heterogeneidade) e menor variação dentro do *cluster* (homogeneidade). Assim, o número ótimo de *clusters* é o que maximiza esse índice (figura 2.8).

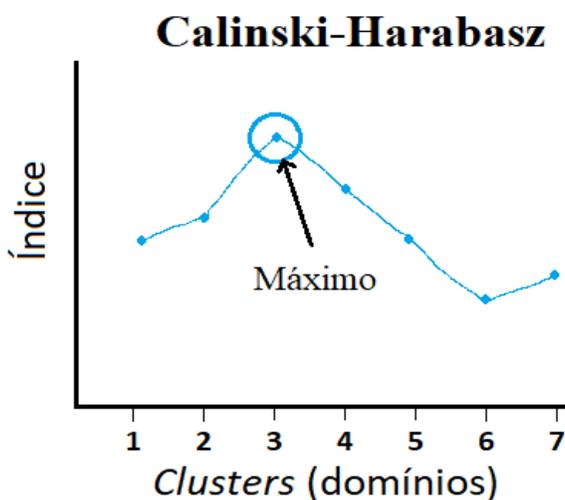


Figura 2.8. Identificação do número de *clusters* através índice Calinski-Harabasz.

Para calcular o índice de Calinski-Harabasz, é utilizada a seguinte expressão:

$$CH(K) = \frac{B(K)}{W(K)} * \frac{(N - K)}{(K - 1)} \quad (\text{equação 21})$$

onde N representa o número de dados (amostras), B(K) e W(K) são a soma dos quadrados das distâncias entre e dentro dos *clusters* respectivamente e K é o número de *clusters*.

2.2.4. MEDIDAS DE DISTÂNCIAS

O *k-means++* é um método de agrupamento baseado no cálculo das distâncias entre elementos dos vetores, dentre elas, as distâncias mais comuns são: Euclidiana, Manhattan e Minkowski (Aggarwal & Reddy 2013).

2.2.4.1. DISTÂNCIA EUCLIDIANA

A distância Euclidiana é simplesmente a distância geométrica no espaço M-dimensional de vetores, definida pela seguinte equação:

$$dist_2(d_1, d_2) = \sqrt{\sum_{i=1}^M (d_1^i - d_2^i)^2} \quad (\text{equação 22})$$

onde d_1^i e d_2^i são os i-ésimos elementos dos vetores d_1 e d_2 , respectivamente. No caso do *k-means++*, é possível representar o d_1^i como o vetor de elementos M-dimensional do centroide do *cluster* e d_2^i o vetor de elementos M-dimensional dos dados.

2.2.4.2. DISTÂNCIA MANHATTAN

A distância Manhattan é a soma das diferenças absolutas dos elementos dos vetores, definida da seguinte forma:

$$dist_1(d_1, d_2) = \sum_{i=1}^M |d_1^i - d_2^i| \quad (\text{equação 23})$$

2.2.4.3. DISTÂNCIA MINKOWSKI

A distância Minkowski é uma generalização da distância Euclidiana e Manhattan, definida através da seguinte equação:

$$dist_h(d_1, d_2) = \sqrt[h]{\sum_{i=1}^M |d_1^i - d_2^i|^h} \quad (\text{equação 24})$$

onde h é um número real tal que $h \geq 1$.

2.2.5. CRIAÇÃO DA ÁREA/VOLUME DE INFLUÊNCIA

De acordo com Sinclair & Blackwell (2002), o vizinho mais próximo (NN, em inglês) é uma aproximação numérica do método da poligonal para geração de grids, onde um valor de um nó é obtido pelo valor da amostra mais próxima a esse nó. Os polígonos de influência de uma amostra são definidos por uma série de bissetrizes perpendiculares de linhas que juntam as diversas amostras. Cada polígono contém uma única amostra e qualquer outro ponto no polígono está mais próximo do dado contido do que qualquer dado externo (figura 2.9a). No entanto, as estimativas são geralmente realizadas em blocos regulares, fazendo que NN funcione como uma aproximação ao método da poligonal. A figura 2.9b mostra um exemplo ilustrativo usando NN. Note, que a área de influência da amostra 1 corresponde aos quatro blocos em azul (amostra mais próxima do centro dos blocos do que as outras), para a amostra 2 corresponde aos seis blocos em verde e para amostra 3 os dois blocos em amarelo. Esse processo é executado em todo o conjunto de dados para todos os blocos dentro do mesmo domínio.

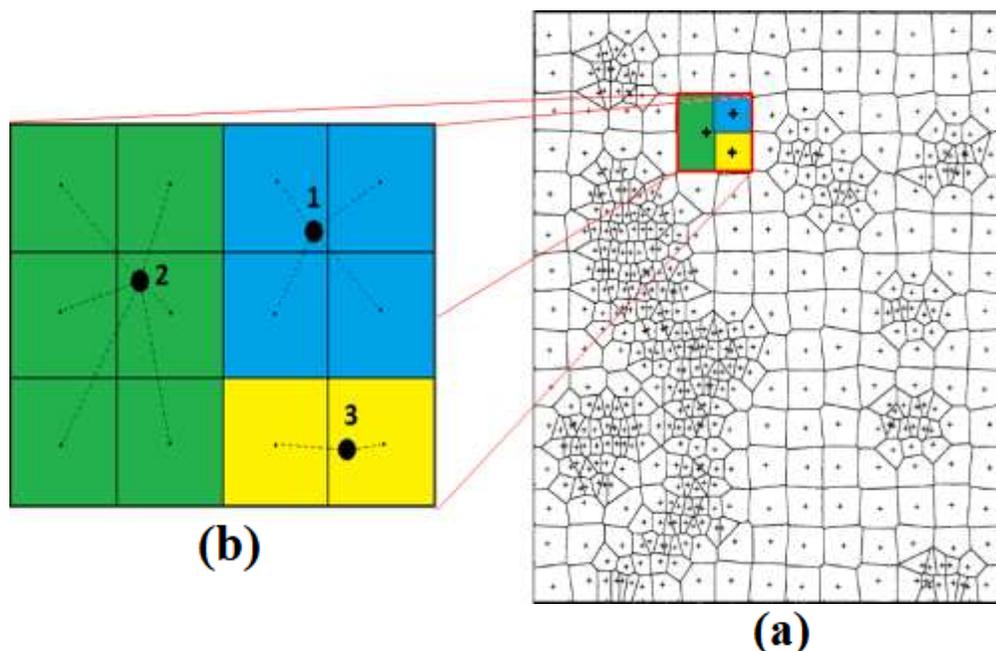


Figura 2.9. Área de influência de cada amostra pelo (a) método poligonal e (b) método do vizinho mais próximo. Figura modificada de Sinclair & Blackwell (2002).

2.2.6. PARÂMETROS DE KRIGAGEM

Os algoritmos de interpolação (ou simulação condicional) utilizados na geoestatística necessitam de um conjunto de parâmetros de vizinhança de busca adequados. Esses parâmetros afetam diretamente a qualidade das estimativas, já que definem quantas e quais amostras serão incluídas dentro do processo de interpolação. O objetivo da escolha da estratégia ou vizinhança de busca “ótima” é selecionar as amostras que garantam a estimativa com maior precisão e acuracidade. Para quantificar quais parâmetros mais impactam nas estimativas foram considerados os seguintes aspectos:

2.2.6.1. ALCANCE DA BUSCA

Na krigagem, o resultado de cada bloco estimado é produto da média ponderada das amostras que estão dentro do raio (ou alcance) de busca (normalmente determinado pelo alcance do variograma). O tamanho da busca é representado por uma elipse (2D) ou elipsoide (3D). A figura 2.10 mostra um exemplo do alcance da elipse (em vermelho) para a estimativa de um bloco (cinza escuro). Note, que para este caso hipotético, somente as amostras em vermelho (dentro do alcance) serão consideradas para estimar o bloco. Já as amostras azuis (fora do alcance) serão desconsideradas. A função principal desta busca é limitar a distância em que serão procuradas amostras para a estimativa do nó do grid. A busca é limitada até determinada distância para garantir maior precisão e acuracidade local e evitar suavização excessiva do modelo (o que poderia ocorrer se fossem usadas todas as amostras disponíveis simultaneamente).

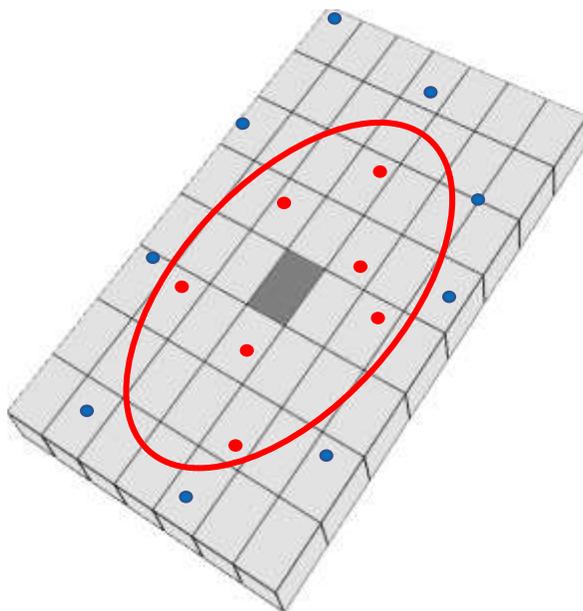


Figura 2.10. Tamanho de busca para amostras dentro (em vermelho) e fora (em azul) da região de busca, representado por uma elipse.

2.2.6.2. NÚMERO DE AMOSTRAS

A estimativa de um bloco também é fortemente influenciada pela quantidade de amostras utilizadas no processo de interpolação. Utilizar muitas ou poucas amostras depende de vários fatores, entre eles podemos mencionar: o tipo de mineralização, a distribuição de probabilidade dos dados (simétrica ou assimétrica), a malha de amostragem ou o adensamento amostral em algumas regiões mais que em outras. Utilizar o maior número de amostras na estimativa de um bloco nem sempre é melhor opção, já que o uso de muita informação pode gerar suavização excessiva na krigagem e, o efeito de eclipsar algumas amostras, pode gerar pesos negativos afetando diretamente o teor do bloco. Por outra parte, utilizar poucas amostras implica que a estimativa seja, muitas vezes, pouco suavizada (e.g., vizinho mais próximo). Imaginemos que, na figura 2.11, desejamos estimar um bloco usando o raio de busca representado pela linha vermelha contínua. Algumas situações podem ser destacadas:

- i. Se o usuário decide escolher no máximo 12 amostras para estimar o bloco, todas as amostras que se encontram dentro do elipsoide de busca serão utilizadas para estimar o bloco, já que há exatamente 12 amostras dentro do raio (amostras vermelhas).
- ii. Se o usuário decide agora escolher no máximo 6 amostras para estimar o bloco, somente algumas amostras dentro do elipsoide de busca serão utilizadas, neste caso, as 6 amostras mais próximas ao centroide do bloco (linha vermelha tracejada), mesmo que dentro do elipsoide de busca (linha contínua) existam 12 amostras. O elipsoide de busca pode ser reduzido conseqüentemente, devido às restrições de número de amostra máximo definidas pelo usuário.
- iii. Se o usuário decide escolher no máximo 24 amostras para estimar o bloco, somente serão utilizadas as amostras que estão dentro do elipsoide de busca, neste caso, 12 amostras.
- iv. Se o usuário decide escolher no mínimo 14 amostras para estimar o bloco, nenhuma estimativa será feita, já que dentro do elipsoide de busca somente existem 12 amostras.

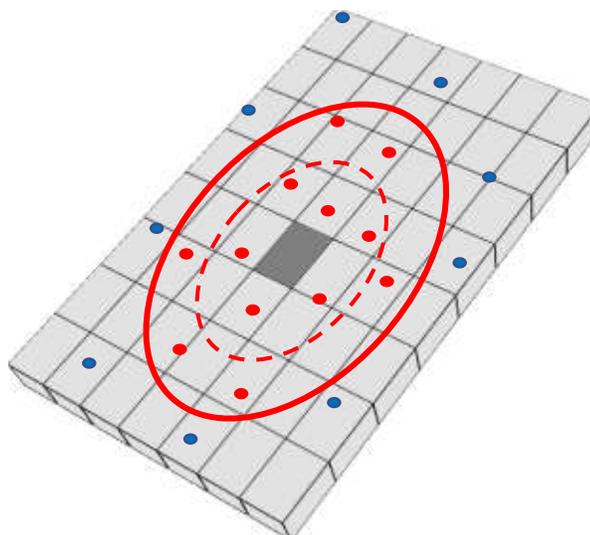


Figura 2.11. Efeito do elipsoide de busca e da escolha do número mínimo e máximo de amostras para a estimativa de um bloco.

A medida que o elipsoide de busca aumenta, a tendência é encontrar mais amostras, aumentando consigo a complexidade do número de combinações entre o número mínimo e máximo das amostras. A escolha pela quantidade de amostras utilizadas para um bloco determinado é muito importante e, talvez, a de maior impacto nas estimativas, desde que o objetivo seja obter o menor erro possível em cada bloco.

2.2.6.3. NÚMERO DE SETORES ANGULARES

Muitas vezes, a concentração de amostras em uma determinada região é mais evidente do que em outras, trazendo como consequência que algumas dessas amostras sejam redundantes para a estimativa de um bloco. Isto deve-se a que certas amostras acabam filtrando a influência de outras devido ao chamado efeito eclipse. Para atenuar esse efeito, é recomendado utilizar uma subdivisão do elipsoide de busca em setores angulares. A figura 2.12 mostra o elipsoide de busca dividido em 4 setores. Note, que nos setores I e III existem amostras que acabam sendo redundantes (eclipsadas), e não agregam muita informação relevante para a estimativa do bloco que não possa ser proporcionada pelas amostras mais próximas. Portanto, limitar algumas vezes a SN para uma ou duas amostras por setor, por exemplo, pode resultar ser mais preciso e acurado que utilizar todas as amostras mais próxima independente da sua localização. A escolha do número de setores angulares dentro do elipsoide de busca é outro dos parâmetros com grande impacto nas estimativas, já que permite filtrar a influência de algumas amostras que acabam sendo redundantes dentro do elipsoide de busca. Um outro ponto importante é que ocorra sempre interpolação e evite-se extrapolação.

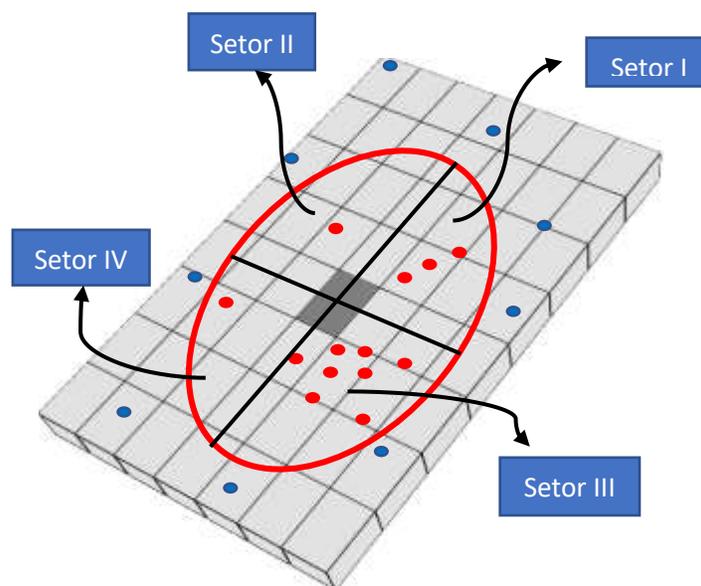


Figura 2.12. Efeito sobre a escolha do número de setores angulares para a estimativa de um bloco.

2.2.6.4. NÚMERO DE AMOSTRAS POR FURO

Em banco de dados 3D, muitas vezes o número de amostras por setor angular não é suficiente para caracterizar uma escolha adequada das amostras que estão ao redor do bloco a ser estimado. Imaginemos que, na figura 2.13a, o elipsoide de busca do bloco foi dividido em 4 setores angulares e desejamos estimar o bloco utilizando 4 amostras por cada setor. Note, que dentro do setor I existem dois furos: F-04 (que está mais próximo do centroide do bloco) e F-05 (mais longe) com 4 amostras cada um (2 de baixo teor em azul e 2 de alto teor de amarelo-vermelho). Nesta situação em particular, se a nossa escolha fosse utilizar 4 amostras por setor para estimar o bloco, provavelmente somente as amostras do F-04 iriam ser utilizadas no processo de interpolação, já que elas estão espacialmente mais próximas do centroide do bloco que as do F-05, misturando os teores de alto e baixo teor.

Uma escolha adequada neste caso seria, além de utilizar as 4 amostras por setor angular, limitar o número de amostras por furo, ou seja, poderíamos utilizar somente duas amostras por cada furo (figura 2.13b linha tracejada). Desse modo, evita-se utilizar informações contíguas de um mesmo furo, as quais são em muitos casos redundantes devido a sua proximidade espacial.

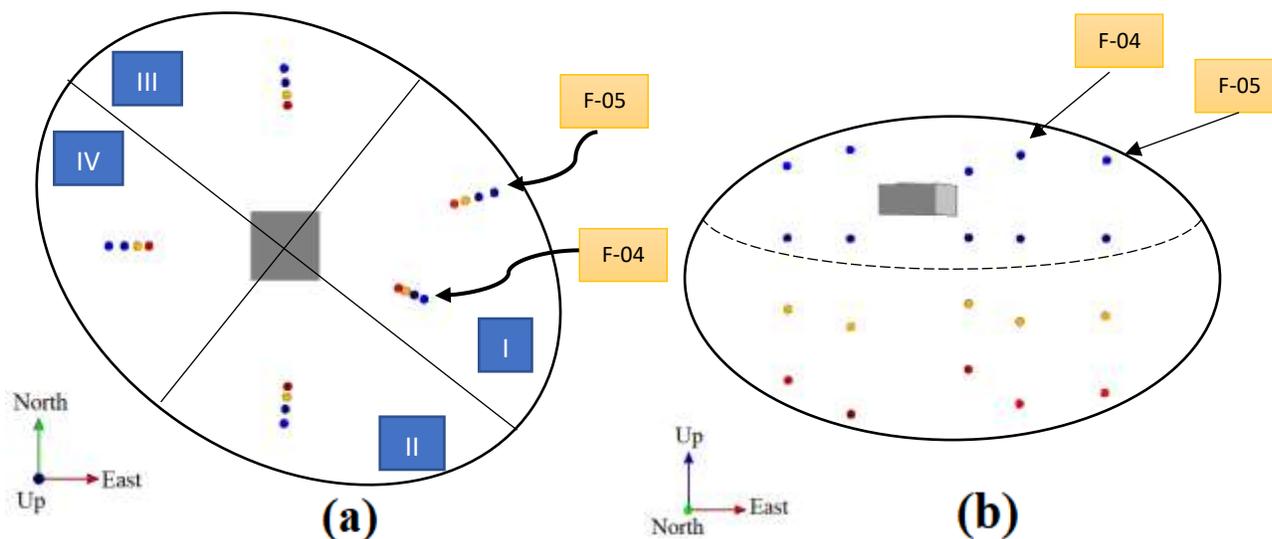


Figura 2.13. Efeito sobre a escolha do número de amostras por furo para a estimativa de um bloco (a) limitar amostras através dos setores angulares (b) limitar as amostras através do número de amostras por furo.

CAPÍTULO 3 - ESTUDOS DE CASO 2D E 3D

Neste capítulo, são analisados dois conjuntos de dados diferentes. O primeiro é um banco de dados de elevações de um terreno (2D) disponível na literatura e, o segundo, um banco de dados (3D) provenientes de um depósito mineral real. A estatística básica, mapa de localização das amostras, definição de domínios, definição das estratégias de busca e análise da continuidade espacial são alguns dos tópicos que serão definidos neste capítulo. O software utilizado para a análise dos dados e geração de gráficos é o SGeMS⁵, o mesmo software que foi utilizado para criar o algoritmo de otimização de parâmetros de krigagem.

3.1. BANCO DE DADOS 2D

O primeiro banco de dados corresponde aos dados provenientes do Walker Lake (Isaaks & Srivastava, 1989), cuja área abrange aproximadamente 260x300 m no plano XY. Foi utilizada a variável V como atributo de interesse, passando por uma pequena transformação nos valores (dividimos por 100) para reproduzir as características de um depósito mineral real, por exemplo, concentrações de algum material em percentagem (%). Duas campanhas de amostragem podem ser observadas, a primeira com uma malha de aproximadamente 20x20 m. cobrindo toda a área e uma segunda campanha de aproximadamente 5x5 m. nas zonas de alto teor (pontos vermelhos e amarelos). O banco de dados é bidimensional (nos eixos X e Y) e possui uma alta densidade amostral principalmente nas zonas de alto teor (figura 3.1).

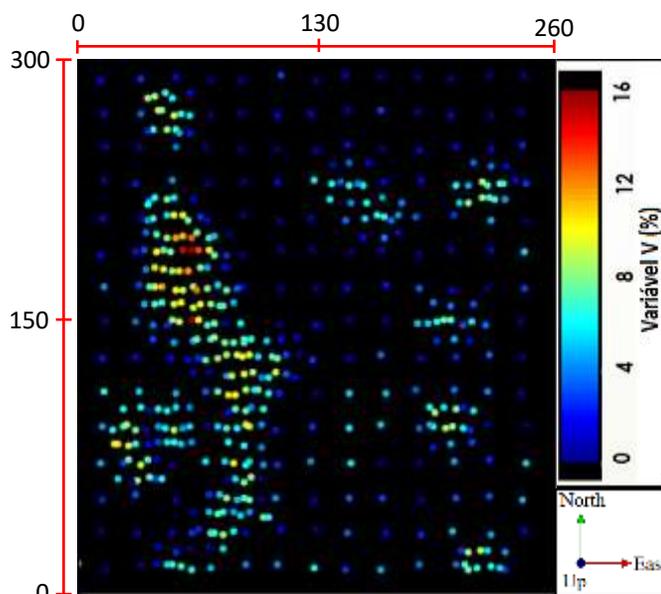


Figura 3.1. Mapa de localização das amostras do banco de dados 2D.

⁵ Stanford Geostatistical Modelling Software (SGeMS)

3.1.1. ESTATÍSTICA BÁSICA

Na figura 3.2, é apresentado um resumo estatístico da variável V (%) através do histograma de frequências e o histograma de frequências cumulativo. Note, que a distribuição dos teores (figura 3.2a) apresenta uma leve assimetria positiva (0.46) devido à presença de alguns valores extremos. Esses valores representam aproximadamente 5% do conjunto de dados, ou seja, valores acima de 9% (figura 3.2b). O subconjunto de 470 amostras foi extraído dos dados exaustivos do banco de dados Walker Lake, cuja média e desvio padrão é de 4.35% e 3%, respectivamente. A distribuição dos teores varia entre 0% e 15.28% .

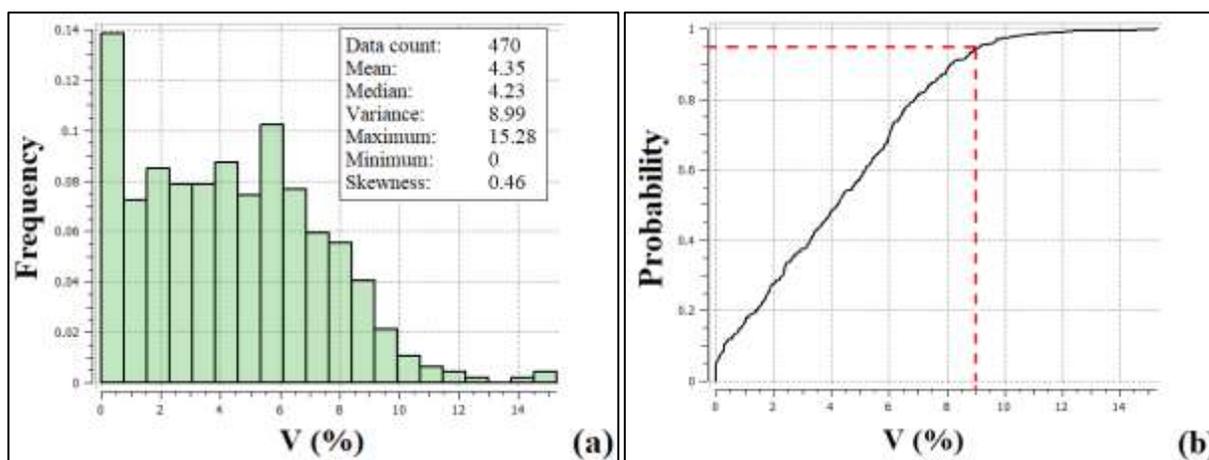


Figura 3.2. Estatística básica (a) Histograma de frequência (b) Histograma acumulado.

3.1.2. ESCOLHENDO O NÚMERO DE DOMÍNIOS

Como foi mencionado anteriormente, a escolha pelo número de domínios em um depósito mineral nem sempre é uma tarefa fácil, cabe ao especialista na área de modelagem escolher as melhores técnicas para delinear os domínios geológicos nele existentes. Neste estudo de caso, será utilizado a LKPO dentro de um mesmo domínio estacionário, portanto, foram criadas 4 versões diferentes sobre o delineamento dos domínios geológicos, entre elas estão: duas visuais, uma utilizando um teor de corte pela mediana e uma utilizando técnicas de agrupamento de dados. Os três primeiros casos foram escolhidos ao meu entendimento geológico sobre o depósito mineral em estudo e, no último caso, foi utilizado um algoritmo matemático (*k-means++*) baseado na criação de grupos (domínios) através de técnicas de agrupamento de dados recentemente utilizados na literatura.

3.1.2.1. CASO I: VISUAL

Para o primeiro caso, foram determinados dois domínios geológicos, identificando principalmente as zonas de alto e baixo teor. O primeiro, encontra-se localizado na zona principal

de alto teor e maior adensamento amostral, com algumas amostras de baixo teor ao redor do domínio para garantir, que o processo de transição entre uma zona e outra, não seja fortemente afetada (figura 3.3a). O segundo domínio corresponde à zona de baixo teor, cujas amostras estão mais espaçadas com algumas lentes de alto teor (figura 3.3b).

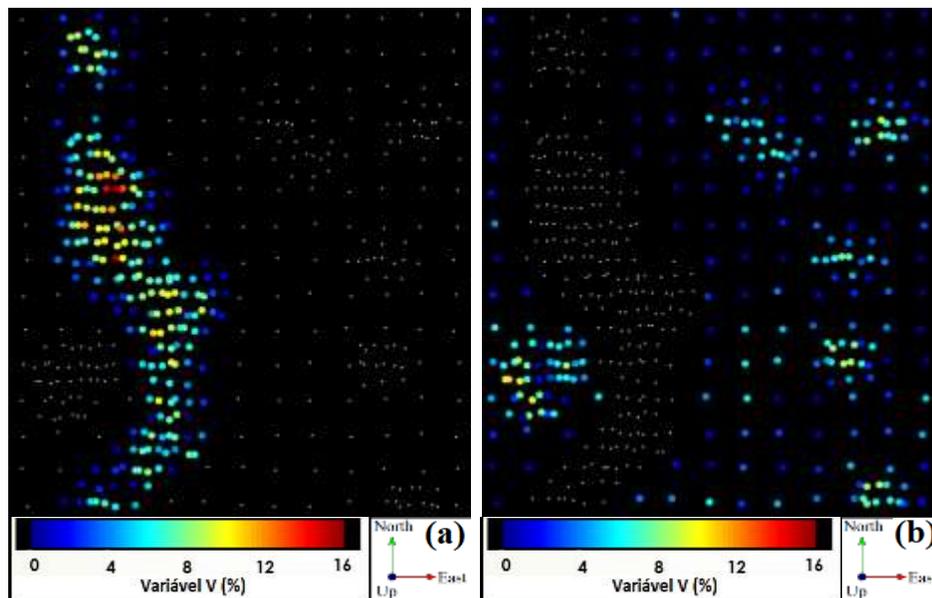


Figura 3.3. Domínios gerados no CASO I (a) zonas de alto teor (b) zona de baixo teor.

3.1.2.2. CASO II: VISUAL

Para o segundo caso, foram determinados dois domínios geológicos, dessa vez, a identificação dos domínios foi baseada na zona de alto teor localizada no sudoeste e noroeste do depósito (figura 3.4a) e na zona de baixo teor localizada no nordeste e sudeste do mesmo depósito (figura 3.4b).

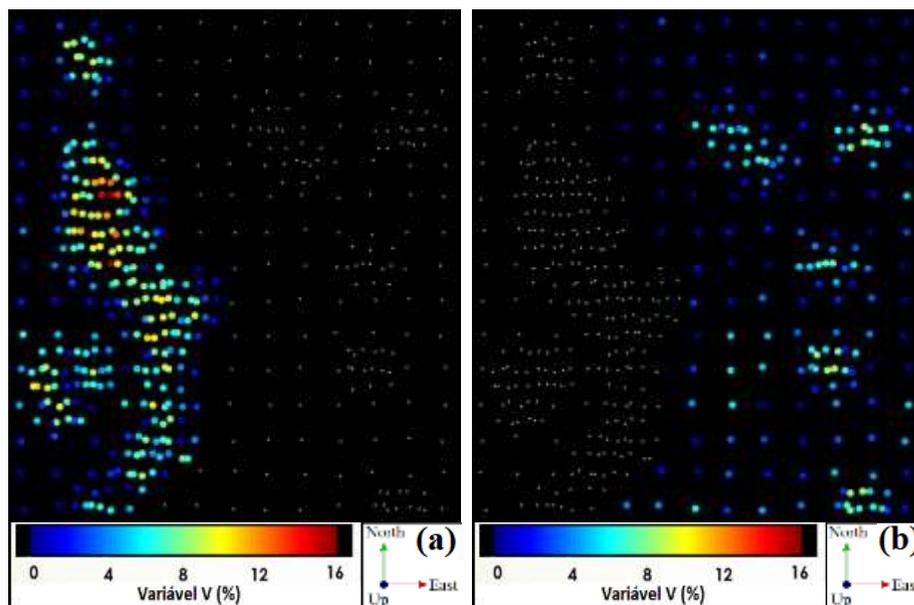


Figura 3.4. Domínios gerados no CASO II (a) zonas de alto teor (b) zona de baixo teor.

3.1.2.3. CASO III: TEOR DE CORTE PELA MEDIANA

Para o terceiro caso, foram determinados dois domínios geológicos baseados no teor de corte pela mediana do banco de dados (4.23%). Para o primeiro domínio, foram agrupados os teores acima do teor de corte, com algumas amostras ao redor para garantir o processo de transição entre um domínio e outro (figura 3.5a), e, para o segundo domínio, foram agrupadas as amostras que estão abaixo do teor de corte (figura 3.5b).

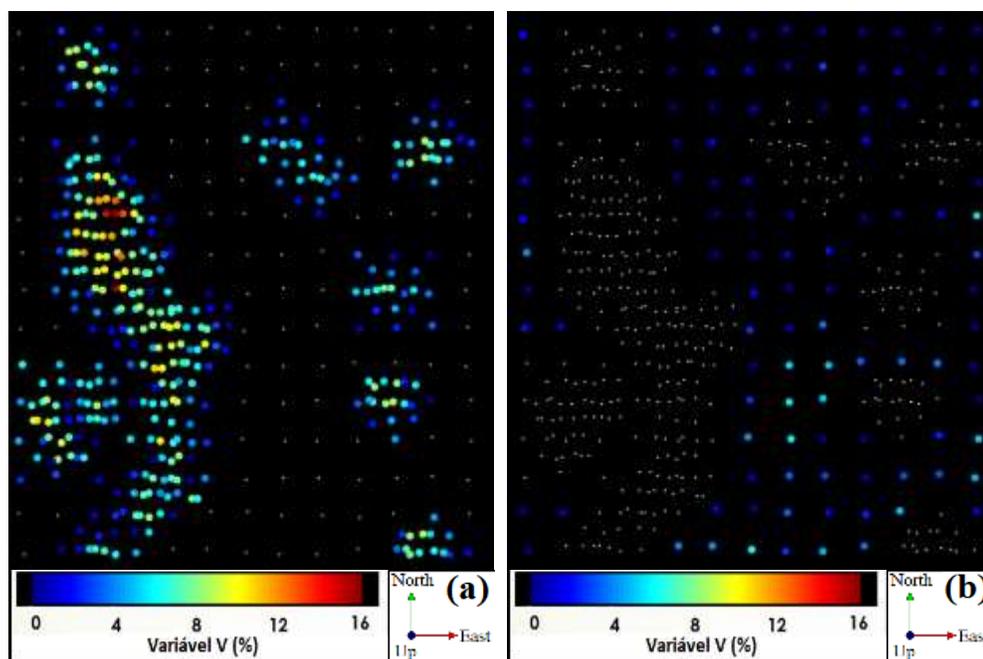


Figura 3.5. Domínios gerados no CASO III (a) zonas de alto teor (b) zona de baixo teor.

3.1.2.3. CASO IV: TÉCNICAS DE AGRUPAMENTOS DE DADOS

Para o quarto caso, utilizaremos o conceito de análise de agrupamento (*clustering*) para definir os domínios estacionários. A ideia principal é agrupar amostras dentro de grupos (*clusters*) com características semelhantes. Assim, as amostras que pertencem a um mesmo grupo são mais parecidas entre si que aquelas que pertencem a grupos diferentes (Romary et al., 2012, Fouedjio et al., 2018).

Em primeiro lugar, foi determinado o número de domínios (*clusters*) baseados na análise de agrupamento das amostras. Para isso, foi selecionado o algoritmo *k-means++* implementado dentro da biblioteca do *scikit-learn*⁶. Neste estudo, foram utilizadas tanto as coordenadas espaciais associadas aos valores dos atributos como a variável de interesse. Abedini *et al.* (2008) argumentaram que é mais eficaz utilizar agrupamento de dados com base nas coordenadas

⁶ É uma biblioteca de aprendizagem de máquina, de domínio público para uso da linguagem de programação Python (Pedregosa et al., 2011).

geográficas e os valores das amostras do que com base apenas nas coordenadas geográficas, além disso, utilizar coordenadas espaciais evita que os *clusters* não sejam desconexos, o que seria indesejável para qualquer análise de estacionariedade.

Como foi mencionado no capítulo anterior, existem vários métodos na literatura para escolher o número adequado de *clusters*; no entanto, neste trabalho foi utilizado o método *Elbow* e o índice de Calinski-Harabasz. Para método *Elbow*, dois gráficos são apresentados na figura 3.6, mostrando como a homogeneidade e heterogeneidade variam para cada número de cluster. O cálculo da homogeneidade foi obtido através da inercia (soma dos quadrados das distâncias dentro dos *clusters*) e o cálculo da heterogeneidade através da variância (soma dos quadrados das distâncias entre os *clusters*). Note, que o ponto *Elbow* (homogeneidade e heterogeneidade não tem ganhos significativos) estabiliza em 5 *cluster* (linha vermelha).

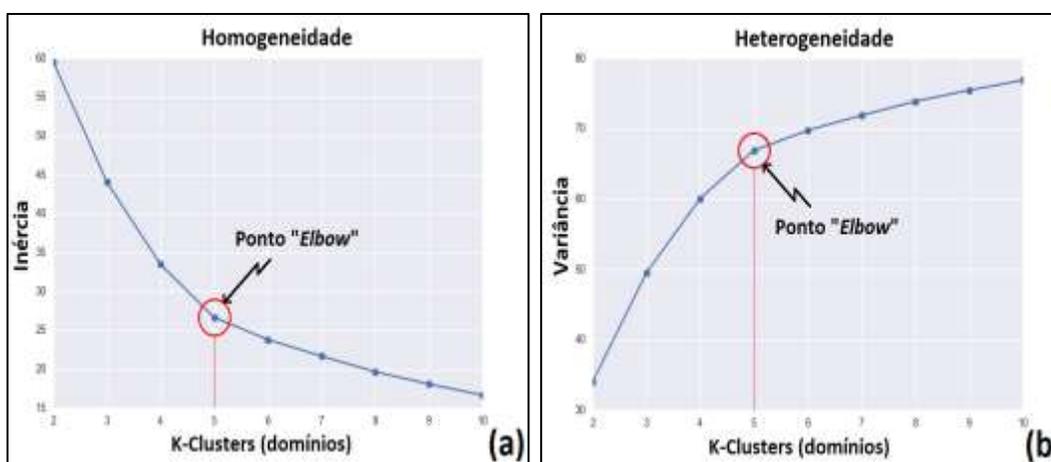


Figura 3.6. Identificação do ponto *Elbow* (a) Heterogeneidade (b) Homogeneidade.

O índice de Calinski-Harabasz mede a relação das variações entre e dentro dos *clusters*. O objetivo é maximizar o índice para encontrar o número de *clusters* adequado. A figura 3.7 mostra índice Calinski-Harabasz para cada número de cluster, obtido através da equação 20 definida no capítulo 2. Note, que a maximização deste índice ocorre quando utilizamos 5 *clusters* (linha vermelha).

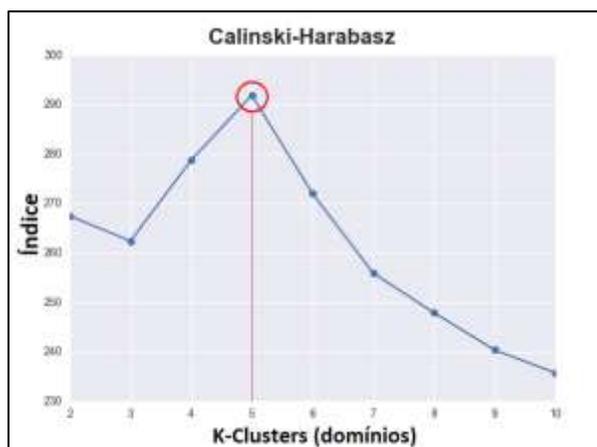


Figura 3.7. Índice Calinski-Harabasz.

Finalmente, foi considerado um método de interpretação visual nas zonas de alto e baixo teor para garantir que esses domínios fossem fisicamente (geologicamente) consistentes. A figura 3.8 mostra a distribuição espacial dos teores e dos 5 *clusters* para este conjunto de dados. Observe, que os domínios não podem ser muito pequenos para que não restem poucos dados em um domínio impossibilitando uma descrição estatística ou inferência confiável e, também não podem ser muito grandes de forma que os dados mostrem que podem ser subdivididos em conjuntos geológicos menores (Rossi & Deutsch, 2014).

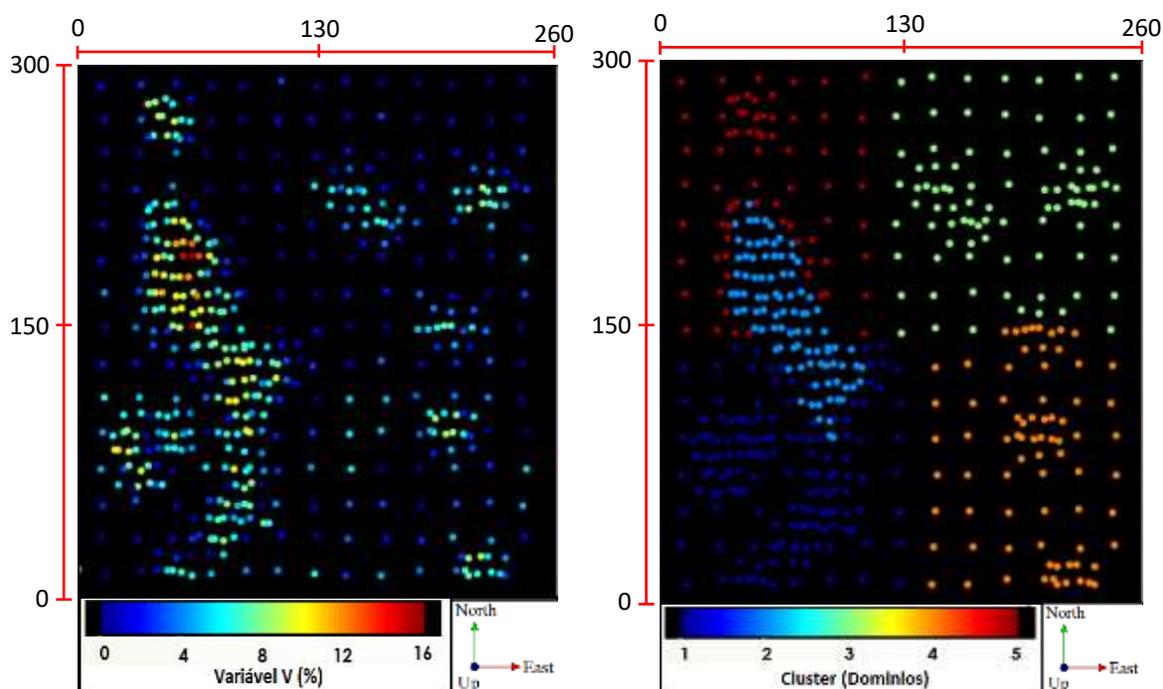


Figura 3.8. (a) Mapa de localização dos teores (b) Mapa de localização dos domínios.

3.1.3. MODELAGEM DA CONTINUIDADE ESPACIAL

Uma vez definido os domínios para cada caso, foi calculada e modelada a continuidade espacial para cada um deles através de variogramas e correlogramas experimentais direcionais (tabela 1). Note, que o domínio 1 (para os casos I, II e III) foi representado pela mesma direção preferencial de maior e menor continuidade (N157° e N67°). Já para o domínio 2, nos mesmos casos, não existe uma continuidade preferencial definida (isotrópico). Para o caso IV, onde foram criados 5 domínios através das técnicas de agrupamento de dados, os domínios 2 e 5 apesar de ter a mesma direção preferencial de maior e menor continuidade (N157° e N67°), o alcance no domínio 2 (maior densidade amostral) é menor do que o alcance no domínio 5 (menor densidade amostral). Isto mostra, que o efeito da malha e a distribuição dos teores pode ter influência nos alcances dos modelos de continuidade espacial. O efeito pepita foi obtido através do variograma

omnidirecional no plano XY e foram utilizadas duas estruturas de modelo esférico para modelar cada domínio. Todos os alcances estão em metros (m).

Tabela 1. Parâmetros estruturais dos modelos de continuidade espacial dos domínios para cada caso.

CASO	Domínio	Tipo	Direção		Estruturas								
					c0	Primeira Estrutura			Segunda Estrutura				
			D1	D2		Modelo	c1	a (D1)	a (D2)	Modelo	c2	a (D1)	a (D2)
I	1	Correlograma	N157°	N67°	0.1	Esférico	0.3	12	12	Esférico	0.6	52	23
	2	Variograma	Isotrópico	Isotrópico	2		1.4	11	11		2.49	45	45
II	1	Correlograma	N157°	N67°	0.1	Esférico	0.4	12	12	Esférico	0.5	75	25
	2	Variograma	Isotrópico	Isotrópico	1		1	13	13		3.16	42	42
III	1	Correlograma	N157°	N67°	0.1	Esférico	0.4	13	13	Esférico	0.5	75	30
	2	Variograma	Isotrópico	Isotrópico	0.5		1.08	16	16		1.06	143	143
IV	1	Correlograma	N0°	N90°	0.15	Esférico	0.45	15	10	Esférico	0.4	45	30
	2	Correlograma	N157°	N67°	0.15		0.84	24	14		0.01	25	15
	3	Variograma	Isotrópico	Isotrópico	0.5		4.39	34	34		0.01	35	35
	4	Variograma	Isotrópico	Isotrópico	0.5		2.08	10	10		2	25	25
	5	Correlograma	N157°	N67°	0.05		0.15	22	20		0.8	53	30

D1/D2 - Direção de maior e menor continuidade, respectivamente, **c0** - efeito pepita, **a** - alcance, **c1/c2** - Contribuições da primeira e segunda estrutura para a variância total, respectivamente.

3.1.4. DEFINIÇÃO DAS ESTRATÉGIAS DE BUSCA

Para cada domínio, foi gerado um grande número de alternativas diferentes de estratégias de busca variando tanto o número de amostras incluídas dentro do elipsoide de busca como o número de amostras utilizadas por cada setor angular (tabela 2). No total, foram geradas 175 estratégias de busca diferentes produto da combinação entre o número mínimo e máximo de amostras e o número máximo de amostras por cada setor angular. É importante mencionar, que os algoritmos de interpolação implementados dentro do software SGeMS utilizam 4 setores angulares (quadrantes) em banco de dados 2D e 8 setores angulares (optantes) em banco de dados 3D.

Tabela 2. Combinações das estratégias de busca.

Número Mínimo de Amostras	Número de Setores Angulares	Número Máximo de Amostras por Cada Setor Angular	Número Máximo de Amostras
4	4	1	4
		2	5
		3	6
		4	.
		5	.
		6	.
		7	28

3.1.5. DEFINIÇÃO DA ÁREA DE INFLUÊNCIA DE CADA AMOSTRA

A área de influência, neste trabalho, é definida como o conjunto de blocos associados a cada amostra do banco de dados. Para realizar esse processo, foi utilizado o método do vizinho mais próximo, onde cada bloco é interpolado (carimbado) com o teor da amostra mais próxima a seu centroide. Como o NN usa o teor de cada amostra e, sabemos que dentro de um conjunto de dados existe a possibilidade de ter várias amostras com o mesmo teor, foi necessário criar uma identificação (ID) diferente para cada uma delas, visando não ter blocos com a mesma identificação. Neste caso, foi realizado o carimbo de cada amostra proporcional ao número de amostras do conjunto de dados, ou seja, se o banco de dados possui 470 amostras, o ID começa desde a amostra número 1 até a amostra número 470.

Na figura 3.9, podemos observar o cálculo da área de influência para todo o modelo de blocos, onde cada cor do bloco representa uma área de influência para cada amostra. Por exemplo, para a amostra com ID = 300, nesse conjunto de dados, somente 7 blocos (amarelo) foram carimbados como pertencentes aquela amostra, dado a sua vizinhança mais próxima ao centroide dos blocos do que as outras. Ao final, cada amostra dentro do conjunto de dados terá a sua própria área de influência.

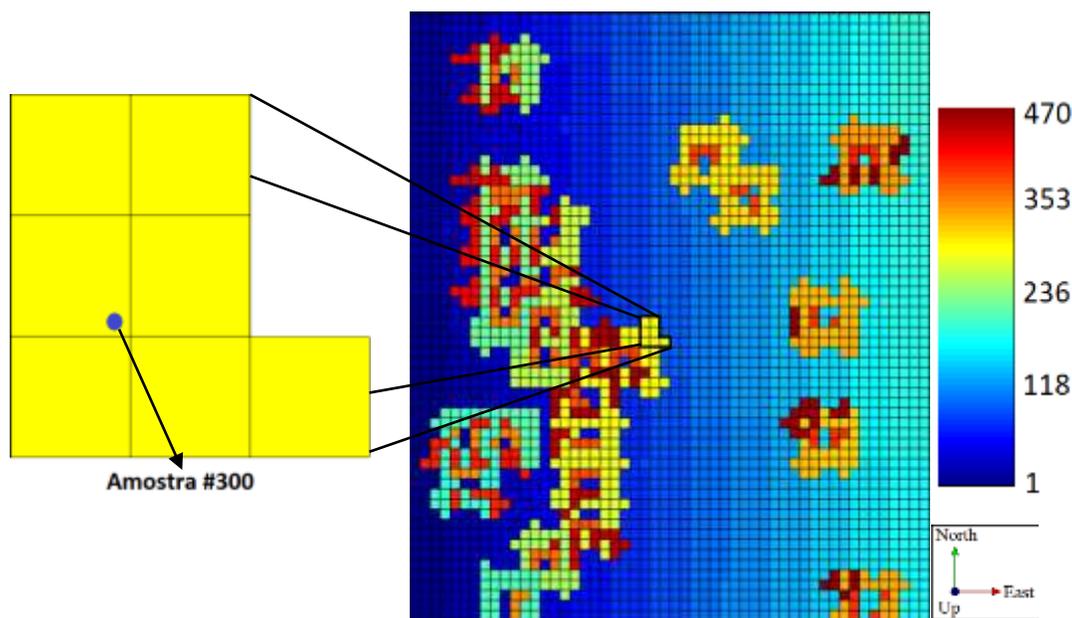


Figura 3.9. Área de influência das amostras pelo método do vizinho mais próximo.

3.1.6. MINIMIZAÇÃO DO ERRO ABSOLUTO

Todo o processo de otimização de parâmetros de krigagem é baseado na minimização do erro absoluto. Para isso, é realizado inicialmente uma validação cruzada para cada uma das estratégias definidas no item 3.1.4. Lembre-se, que a validação cruzada gera para cada amostra um

erro entre o valor real e o valor estimado. Portanto, cada amostra dentro do conjunto de dados terá uma distribuição do erro absoluto (figura 3.10), que será a base para a escolha dos parâmetros que levaram a obter o menor erro possível. Finalmente, dentro do conjunto de estratégias de busca, serão selecionados aqueles parâmetros que minimizam a função do erro absoluto para cada amostra.

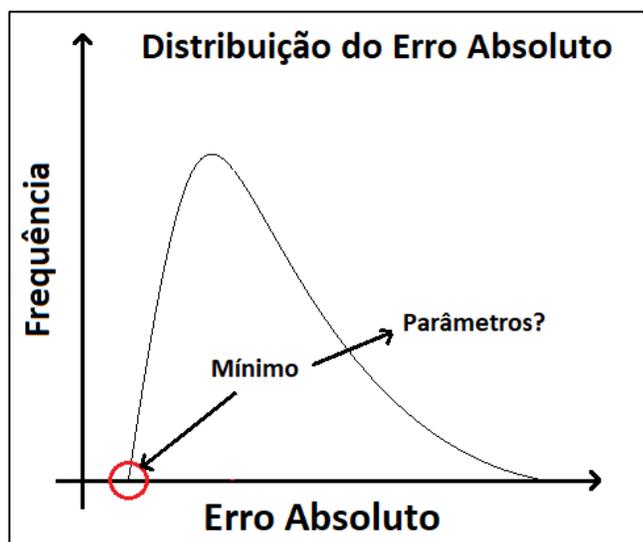


Figura 3.10. Exemplo ilustrativo da minimização do erro absoluto para uma amostra, utilizando várias estratégias de krigagem.

3.1.7. ESTIMATIVAS DENTRO DE CADA ÁREA DE INFLUÊNCIA

Finalmente, para cada área de influência (item 3.1.5) será realizado o processo de estimativa (e.g., krigagem ordinária) utilizando os parâmetros obtidos no item 3.1.6. O alcance e a orientação do elipsoide de busca é determinado pelo mesmo alcance/orientação do modelo de continuidade espacial (definidos na tabela 1). Este processo de estimativas torna-se uma tarefa cíclica e repetitiva para todos os blocos dentro de cada domínio, portanto, foi necessário a criação de um algoritmo dentro do software SGeMS.

3.2. BANCO DE DADOS 3D

O segundo banco de dados corresponde a um depósito mineral de ouro em 3D, onde os valores de ouro são expressos em (g/t) e cuja área total abrange aproximadamente 2.1 km² no plano XY e 730 m entre os níveis mínimo e máximo ao longo do eixo Z (figura 3.11)⁷. Duas campanhas

⁷ O sistema de coordenadas foi trocado para garantir a privacidade da empresa patrocinadora desta pesquisa.

de amostragem podem ser observadas, a primeira com uma malha de amostragem variável entre 45 e 150 m (no plano XY) e uma segunda com uma malha mais densa de aproximadamente 30x30 m na área central do depósito. A maioria dos furos são verticais com alguns inclinados. O tamanho das amostras é de aproximadamente 7.5 m.

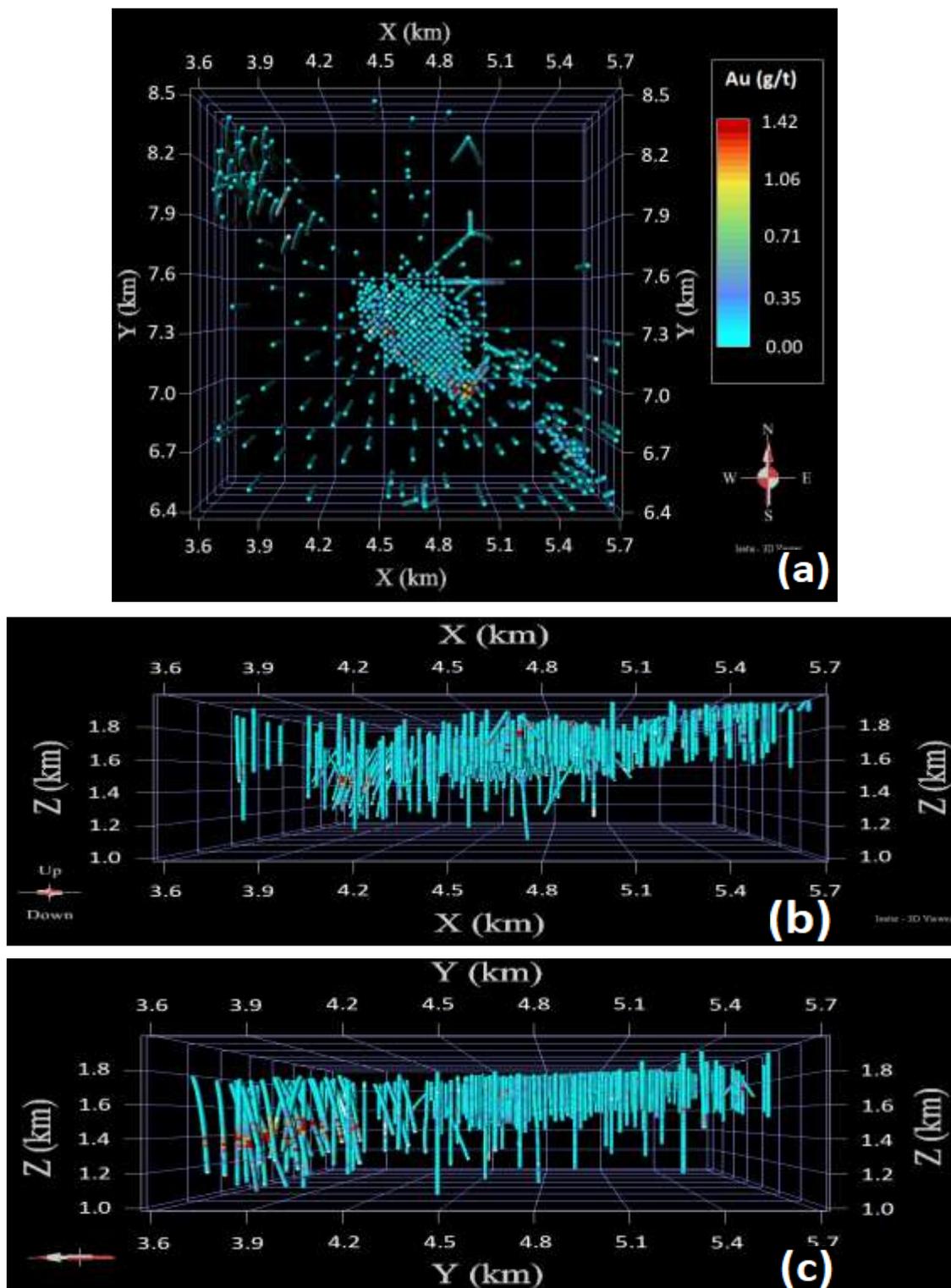


Figura 3.11. Mapa de localização das amostras do banco de dados em 3D.

3.2.1. DOMÍNIOS

Diferente do caso anterior, neste estudo os domínios já tinham sido definidos pela empresa de mineração que forneceu este conjunto de dados. Os domínios foram definidos a partir de informações litológicas. A figura 3.12 mostra o histograma de frequência das amostras para cada domínio (1 a 6). A maioria das amostras (mais de 80%) pertencem aos domínios 3 e 4; portanto, para os propósitos deste trabalho, somente os dados destes dois domínios serão usados. As figuras 3.13 a e b mostram o mapa de localização das amostras para cada domínio estudado. Em ambos os casos, os dados estão localizados na parte central do depósito, indicando ser a zona com maior concentração de minério.

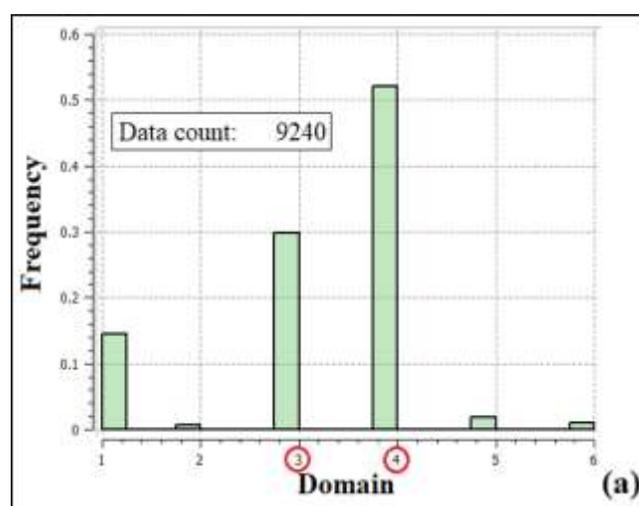


Figura 3.12. Histograma das amostras por domínio.

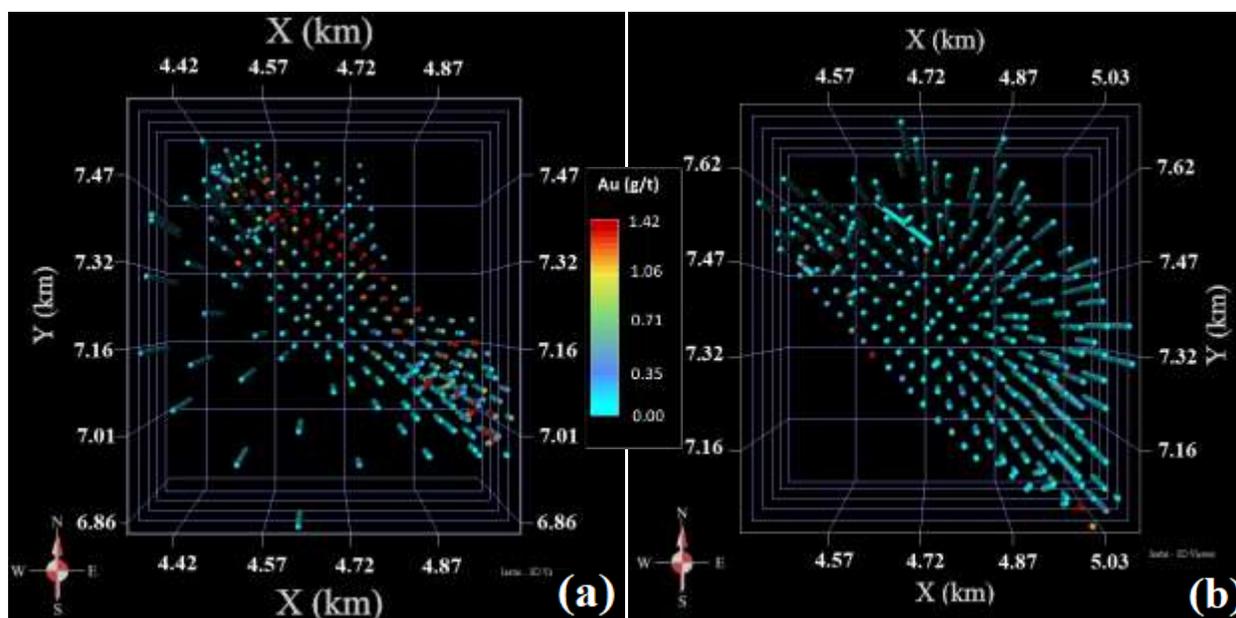


Figura 3.13. Mapa de localização das amostras (a) domínio 3 (b) domínio 4.

3.2.2. ESTATÍSTICA BÁSICA

Nas figuras 3.14 a e b podemos observar a distribuição de teores através do histograma de frequências para os domínios 3 e 4, respectivamente. Em ambos casos, as duas distribuições apresentam uma alta assimetria positiva (7.02 e 6.07) indicando a presença de alguns valores extremos que afetam o comportamento assimétrico dessas distribuições. Outra maneira de observar esse comportamento assimétrico é através do coeficiente de variação (CV), que é a razão entre o desvio padrão e a média. Os CVs para os domínios 3 e 4 são 2.26 e 2.69, respectivamente. Em estimativas de recursos minerais, um CV acima de 2 indica que a presença dos valores extremos pode afetar as estimativas locais. Esses valores extremos⁸ representam aproximadamente 5% da distribuição dos dados com teores acima de 1.42 g/t. O número de amostras para ambos domínios são 2701 e 4114, respectivamente, com uma média de aproximadamente 0.28 g/t (domínio 3) e 0.32 g/t (domínio 4).

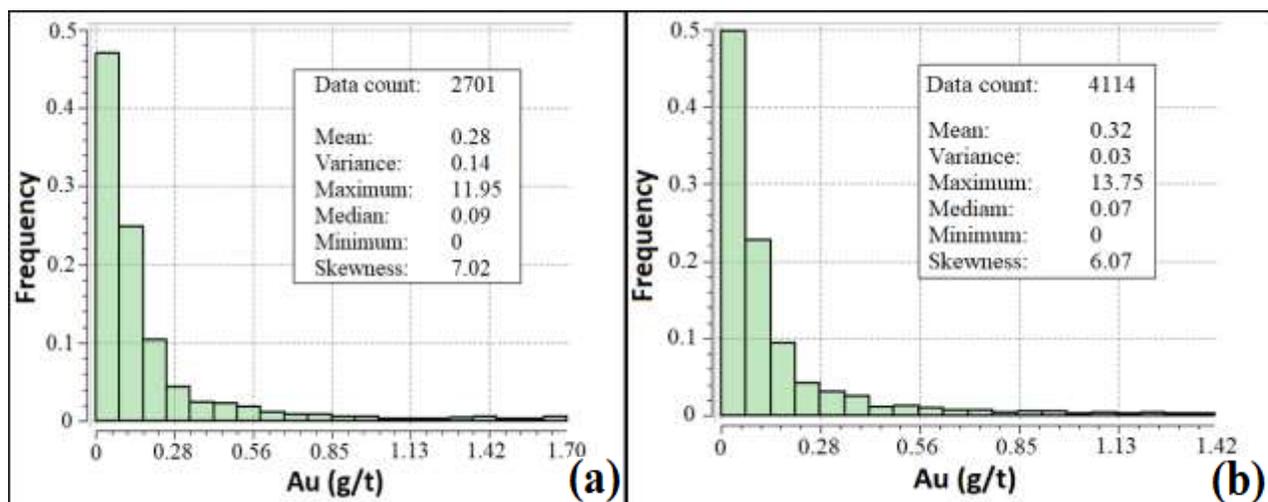


Figura 3.14. Histograma de frequência das amostras (a) domínio 3 (b) domínio 4.

3.2.3. MODELAGEM DA CONTINUIDADE ESPACIAL

A tabela 3 mostra um resumo dos parâmetros utilizados na modelagem da continuidade espacial para os domínios 3 e 4, respectivamente. Nessa modelagem, foram utilizados correlogramas experimentais para atenuar o efeito da presença de alguns valores extremos. Dado que a maioria dos furos são verticais e o tamanho da composta é menor nessa direção, o efeito

⁸ Sinclair & Blackwell (2002) comentam que uma distribuição de dados com mais de três (3) desvios padrão do valor médio podem ser considerados como valores extremos.

pepita foi obtido através do correlograma vertical (ao longo do furo). Foram utilizadas duas estruturas com modelos esféricos em ambos domínios. Todos os alcances estão em metros (m).

Tabela 3. Parâmetros dos modelos de continuidade espacial para os domínios 3 e 4.

Domínio	Tipo	Direção			Estruturas										
		D1	D2	D3	c0	Primeira Estrutura					Segunda Estrutura				
						Modelo	c1	a (D1)	a (D2)	a (D3)	Modelo	c2	a (D1)	a (D2)	a (D3)
3	Correlograma	N112°	N22°	D90°	0.1	Esférico	0.45	45	40	36	Esférico	0.45	190	60	55
4	Correlograma	N45°	N135°	D90°	0.1	Esférico	0.45	45	40	28	Esférico	0.45	245	150	45

D1/D2/D3 - Direção de maior, intermediária e menor continuidade respectivamente, **c0** – efeito pepita, **a** - alcance, **c1/c2/c3** – contribuições da primeira, segunda e terceira estrutura para a variância total, respectivamente.

3.2.4. DEFINIÇÃO DO NÚMERO DE ESTRATÉGIAS DE BUSCA

Finalmente, para cada domínio, foram realizadas 256 estimativas diferentes, produto das combinações entre o número mínimo e máximo de amostras e o número máximo de amostras por cada setor angular (Tabela 4). É importante mencionar que a performance do algoritmo de otimização está diretamente relacionada, principalmente, aos seguintes aspectos:

- i. Número de estimativas;
- ii. Tamanho do modelo de blocos;
- iii. Número de amostras;
- iv. Capacidade computacional.

O aumento dos três primeiros aspectos e a diminuição da capacidade computacional pode gerar demoras (ou atrasos) nos resultados do processo de otimização. No próximo capítulo, será realizada uma comparação dos tempos de resposta para cada processo utilizando diferentes especificações de hardware.

Tabela 4. Combinações das estratégias de busca no banco de dados 3D.

Número Mínimo de Amostras	Número de Setores Angulares	Número Máximo de Amostras por Cada Setor Angular	Número Máximo de Amostras
4	8	1	4
		2	5
		3	6
		4	.
		5	.
		6	.
		7	.
		8	32

3.2.5. DEFINIÇÃO DA ÁREA DE INFLUÊNCIA DE CADA AMOSTRA

Similar ao caso anterior, nesse banco de dados foi realizado o carimbo (ID) de cada amostra proporcional ao número total de amostras por cada domínio (2701 e 4114, respectivamente). Nas figuras 3.15 a e b, podemos observar o cálculo da área de influência para todo o modelo de blocos dentro dos domínios 3 e 4, respectivamente. Cada cor do bloco representa uma área de influência para cada amostra.

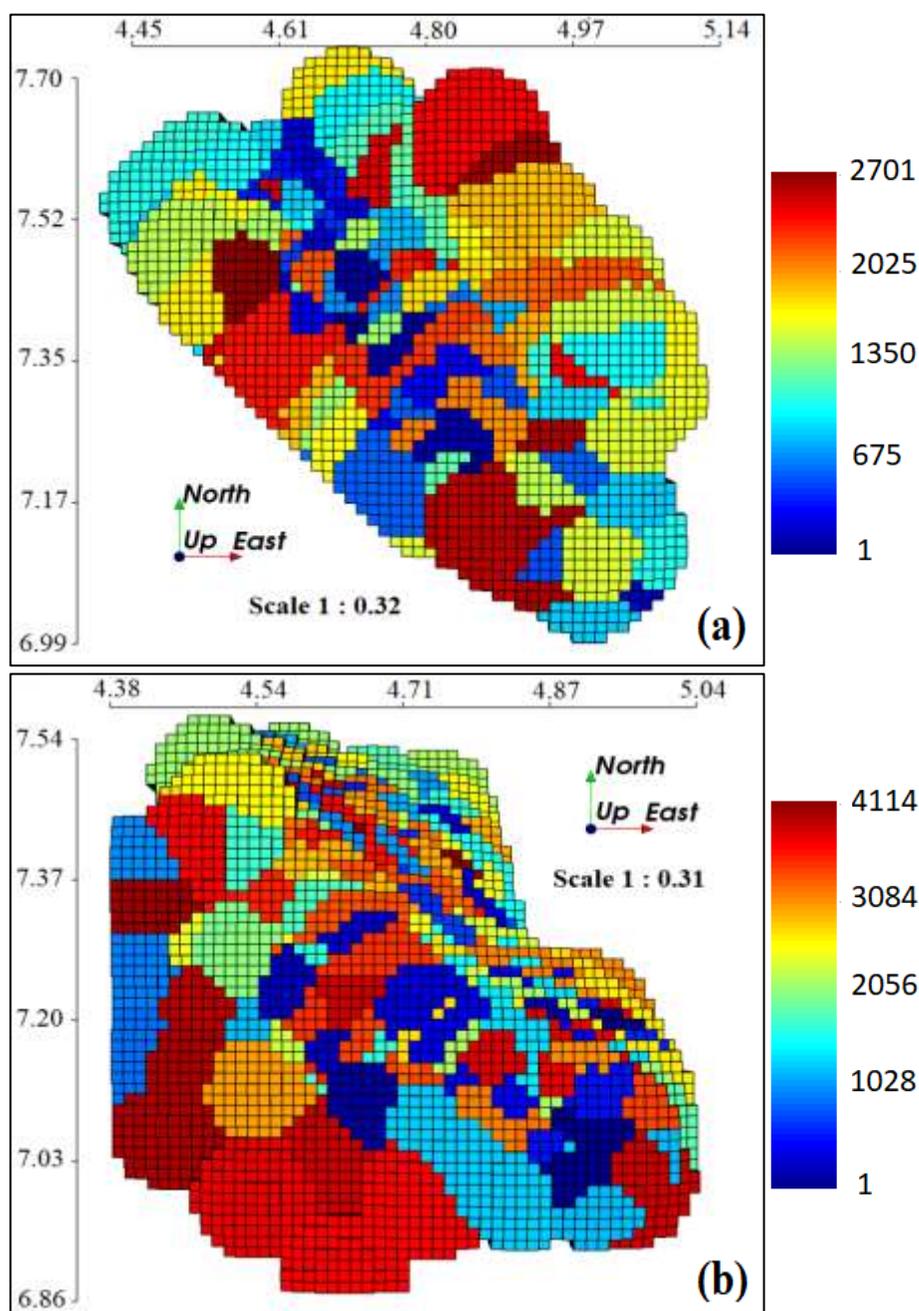


Figura 3.15. Área de influência das amostras pelo método do vizinho mais próximo (a) Domínio 3 (b) Domínio 4.

CAPÍTULO 4 – RESULTADOS E ANÁLISES DA LKPO UTILIZANDO VALIDAÇÃO CRUZADA

Neste capítulo, serão analisados todos os resultados obtidos nos estudos de caso 2D e 3D definidos anteriormente no capítulo 3. Cada resultado da LKPO será comparado com a metodologia atualmente disponível na literatura (LKNA⁹), o modelo de referência (quando disponível) ou modelos teóricos¹⁰. Finalmente, para avaliar a precisão e acurácia de cada modelo, serão usados os métodos de validações atualmente disponível na literatura, como reprodução de estatísticas básicas, de curvas teor-tonelagem, grau de suavização da krigagem e análise das médias locais (análise de deriva).

4.1. BANCO DE DADOS 2D

Para avaliar os resultados deste primeiro banco de dados, foi utilizado como modelo de referência o Walker Lake exaustivo (Isaaks & Srivastava, 1989). Esse modelo abrange toda a área amostrada em blocos de 1x1m (figura 4.1a). Para fins de comparação dos resultados no mesmo suporte das estimativas, foi realizado um “*upscaling*” (média aritmética dos teores de volumes menores dentro de um volume maior) para um tamanho de blocos de 5x5m (figura 4.1b).

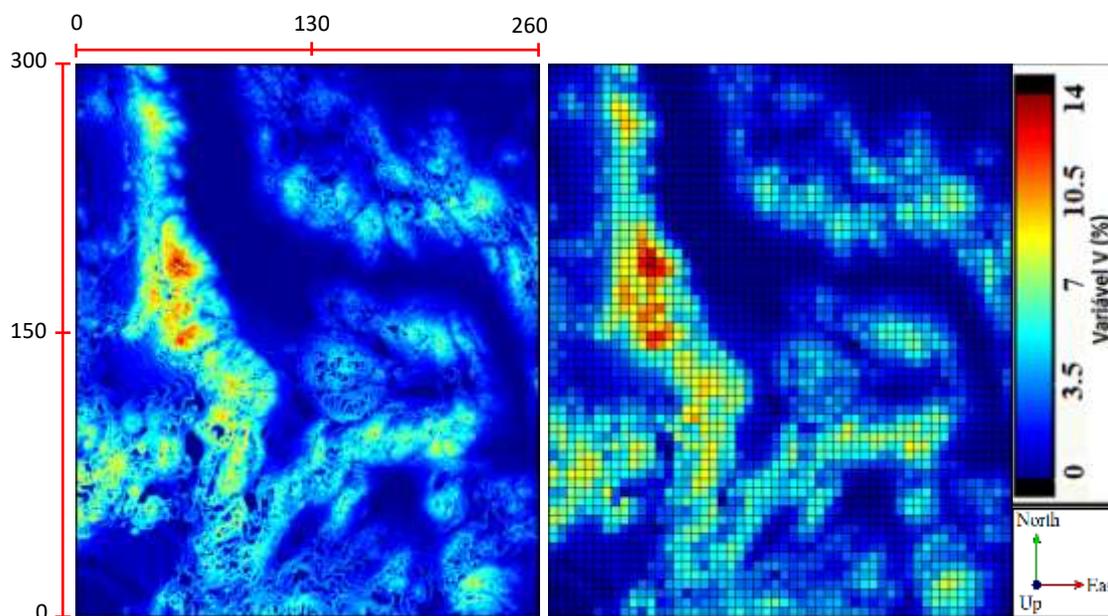


Figura 4.1. Mapa de localização das amostras do Walker Lake Exaustivo (a) blocos de 1x1m (b) blocos de 5x5m.

⁹ Apesar de a indústria utilizar métodos como a krigagem e a KNA, que utilizam uma única estratégia de busca para todo o domínio, optou-se, conforme os objetivos do presente trabalho, por comparar o método da LKPO com o da LKNA, uma vez que a LKNA busca, localmente (bloco a bloco), melhorar o processo da KNA.

¹⁰ Definidos no item 4.1.4 (Análise de suavização da krigagem).

4.1.1. RESUMO ESTATÍSTICO

Uma primeira análise concentra-se nas estatísticas básicas como média, mediana e variância dos três modelos estudados: referência (V_Exhaustive), metodologia proposta (V_LKPO) e a metodologia atualmente disponível na literatura (V_LKNA) proposta por Coombes & Boamah (2015). A partir dessas estatísticas, foram inferidas algumas medidas de erro como média, mediana e variância do erro relativo (em percentagem) da seguinte maneira:

$$\text{Erro Relativo (\%)} = \frac{\text{Valor Estimado} - \text{Valor Real}}{\text{Valor Real}} * 100 \quad (\text{Equação 25})$$

Note que, o erro relativo pode ser positivo (apontando que o valor estimado está superestimando o valor real); negativo (apontando que o valor estimado está subestimando o valor real) ou nulo (indicando que o valor real e o valor estimado são iguais). Por outra parte, para cada modelo foi calculado a distribuição do erro, o erro absoluto e o erro quadrático. Entenda por erro a diferença entre o valor real e o valor estimado em cada local u dentro de todo o modelo de blocos. Obtiveram-se os valores médios desses erros como acurácia (tendenciosidade) e sua variabilidade como medida de precisão, utilizando as seguintes equações:

$$\text{Erro Médio (ME)} = \frac{1}{n} \sum_{i=1}^n (Z_u^* - Z_u) \quad (\text{Equação 26})$$

$$\text{Erro Absoluto Médio (MAE)} = \frac{1}{n} \sum_{i=1}^n |Z_u^* - Z_u| \quad (\text{Equação 27})$$

$$\text{Erro Quadrático Médio (MSE)} = \frac{1}{n} \sum_{i=1}^n (Z_u^* - Z_u)^2 \quad (\text{Equação 28})$$

onde Z_u e Z_u^* representam o valor real e o valor estimado em cada local u , respectivamente. Por último, foi calculada o coeficiente de correlação de Pearson entre o modelo de referência e cada uma das metodologias citadas anteriormente através da seguinte equação:

$$\rho = \frac{\sum_{i=1}^n (Z_u^* - \bar{Z}^*)(Z_u - \bar{Z})}{\sqrt{(\sum_{i=1}^n (Z_u^* - \bar{Z}^*)^2) \cdot \sqrt{(\sum_{i=1}^n (Z_u - \bar{Z})^2)}} \quad (\text{Equação 29})$$

onde:

- $Z_1^*, Z_2^*, \dots, Z_n^*$ são os valores interpolados de cada bloco no lugar i para a variável estimada;

- Z_1, Z_2, \dots, Z_n são os valores reais de cada bloco no lugar i para a variável real;
- \bar{Z} e \bar{Z}^* são as médias aritméticas da variável real e estimada, respectivamente.

A Tabela 5 mostra um resumo estatístico dos modelos da LKPO, LKNA e a referência para cada domínio nos 4 casos estudados anteriormente. Dessa tabela, pode-se inferir as seguintes conclusões¹¹:

- O erro relativo da **média** em quase todos os domínios (exceto no domínio 2-caso III) sempre é menor quando utilizamos o modelo obtido com a LKPO do que o modelo obtido com a LKNA.
- No erro relativo da **mediana**, com exceções do domínio 1-caso I e do domínio 2-caso III, o modelo gerado pela LKPO sempre foi menor do que quando utilizamos o modelo gerado pela LKNA.
- O erro relativo da **variância** em todos os domínios e em todos os casos, sempre é menor quando utilizamos o modelo gerado pela LKPO do que o modelo obtido com a LKNA.
- As **correlações** em todos os domínios e para todos os casos, apresentaram leves diferenças significativas entre ambos os modelos¹² para caracterizar que um deles é melhor que o outro. Entretanto, com exceção do domínio 2-caso III, todas as correlações encontram-se acima de 0.72 indicando uma correlação forte (Hinkle *et al.* (2003)).
- O erro **médio** obtido a partir da distribuição do erro, em quase todos os domínios (exceto nos domínios 2 dos casos II e III) sempre é menor quando utilizamos o modelo da LKPO do que o modelo da LKNA.
- O erro **absoluto médio** obtido a partir da distribuição do erro absoluto, com exceções do domínio 2-caso II e do domínio 3-caso IV, o modelo gerado pela LKPO sempre foi menor do que quando utilizamos o modelo gerado pela LKNA.
- O erro **quadrático médio** obtido a partir da distribuição do erro quadrático, em quase todos os domínios (exceto no caso II para ambos os domínios, o domínio 1-caso I e o domínio 3-caso IV) o modelo gerado pela LKNA sempre foi maior do que o modelo gerado pela LKPO.

¹¹ As cores das barras (azul e vermelha) no erro relativo (%) indicam a presença de superestimativa ou subestimativa nas estatísticas básicas, respectivamente. Por outro lado, os tamanhos dessas barras indicam a proporção que elas representam, ou seja, entre maior é a barra, maior é o erro (ou vice-versa).

¹² O modelo gerado pela LKPO/REFERÊNCIA e o modelo gerado pela LKNA/REFERÊNCIA.

Tabela 5. Resumo estatístico por cada domínio no banco de dados 2D.

CASO	Variável	Domínio	Estatística Básica			Erro Relativo (%)			ρ	Distribuição do Erro		
			Média (%)	Mediana (%)	Variância (%) ²	Média	Mediana	Variância		ME (%)	MAE (%)	MSE (%) ²
I	V_Exhaustive		3.89	3.51	8.11							
	V_LKPO	1	3.96	3.4	7.22	1.80	-3.13	-10.97	0.92	-0.07	0.84	1.22
	V_LKNA	4	3.47	6.72	2.83	-1.14	-17.14	0.93	-0.11	0.85	1.17	
	V_Exhaustive		2.36	2.06	3.49							
	V_LKPO	2	2.43	2.21	2.48	2.97	7.28	-28.94	0.79	-0.08	0.87	1.31
	V_LKNA		2.49	2.34	2.02	5.51	13.59	-42.12	0.79	-0.13	0.92	1.35
II	V_Exhaustive		3.58	3.03	7.36							
	V_LKPO	1	3.67	3.17	6.26	2.51	4.62	-14.95	0.9	-0.08	0.91	1.44
	V_LKNA		3.73	3.29	5.7	4.19	8.58	-22.55	0.9	-0.14	0.92	1.43
	V_Exhaustive		2.24	1.99	3.07							
	V_LKPO	2	2.3	2.09	2.29	2.68	5.03	-25.41	0.79	-0.07	0.83	1.18
	V_LKNA		2.31	2.11	2.07	3.12	6.03	-32.57	0.79	-0.07	0.83	1.17
III	V_Exhaustive		3.56	3.18	6.27							
	V_LKPO	1	3.92	3.57	4.54	10.11	12.26	-27.59	0.86	-0.37	1.05	1.81
	V_LKNA	4	3.67	4.09	12.36	15.41	-34.77	0.85	-0.44	1.09	1.89	
	V_Exhaustive		1.84	1.61	2.38							
	V_LKPO	2	1.69	1.35	1.41	-8.15	-16.15	-40.76	0.69	0.16	0.83	1.29
	V_LKNA		1.7	1.38	1.03	-7.61	-14.29	-56.72	0.67	0.14	0.87	1.32
IV	V_Exhaustive		3.31	3.12	4.03							
	V_LKPO	1	3.23	3.11	2.91	-2.42	-0.32	-27.79	0.76	0.08	1.01	1.73
	V_LKNA		3.13	3	2.12	-5.44	-3.85	-47.39	0.72	0.18	1.08	1.99
	V_Exhaustive		7.45	7.43	4.81							
	V_LKPO	2	7.7	7.6	4.34	3.36	2.29	-9.77	0.85	-0.25	0.9	1.41
	V_LKNA		8.12	8.07	2.9	8.99	8.61	-39.71	0.84	-0.67	1.04	1.9
	V_Exhaustive		1.92	1.6	2.63							
	V_LKPO	3	1.82	1.41	2.31	-5.21	-11.88	-12.17	0.78	0.1	0.75	1.1
	V_LKNA		1.78	1.34	2.05	-7.29	-16.25	-22.05	0.79	0.13	0.72	1.03
	V_Exhaustive		2.75	2.54	3.26							
	V_LKPO	4	3.13	3.09	2.29	13.82	21.65	-29.75	0.74	-0.37	1.03	1.65
	V_LKNA		3.22	3.16	1.78	17.09	24.41	-45.40	0.74	-0.46	1.05	1.67
V_Exhaustive		1.94	1.49	3.75								
V_LKPO	5	1.81	1.41	3.05	-6.70	-5.37	-18.67	0.89	0.14	0.58	0.78	
V_LKNA		1.66	1.22	2.61	-14.43	-18.12	-30.40	0.87	0.28	0.65	1	

ME = Erro Médio; MAE = Erro Absoluto Médio; MSE = Erro Quadrático Médio; ρ = Coef. de correlação de Pearson

4.1.2. CURVAS TEOR-TONELAGEM

A análise dos próximos resultados foi feita sobre todo o depósito mineral. Para isso, foi necessário construir um modelo de blocos global a partir da união de todos os domínios gerados em cada caso. Portanto, a segunda análise concentra-se nas curvas teor-tonelagem, na qual, duas curvas são envolvidas: uma é a tonelagem (em percentagem) acima do teor de corte versus o teor médio (lado direito dos gráficos) e a outra é o teor médio associado a essa tonelagem (lado esquerdo dos gráficos). As figuras 4.2 a 4.5 mostram as curvas teor-tonelagem para todos os casos na ordem crescente (I, II, III e IV). Observe, que o comportamento das duas curvas obtidas pela

LKPO (linhas vermelhas) se aproxima mais ao modelo de referência (linhas verdes) do que as curvas geradas pelo modelo da LKNA (linhas azuis) em todos os 4 casos. Isso significa que, para qualquer teor de corte neste estudo, a metodologia da LKPO levou a modelos mais precisos e acurados do que o uso da metodologia da LKNA.

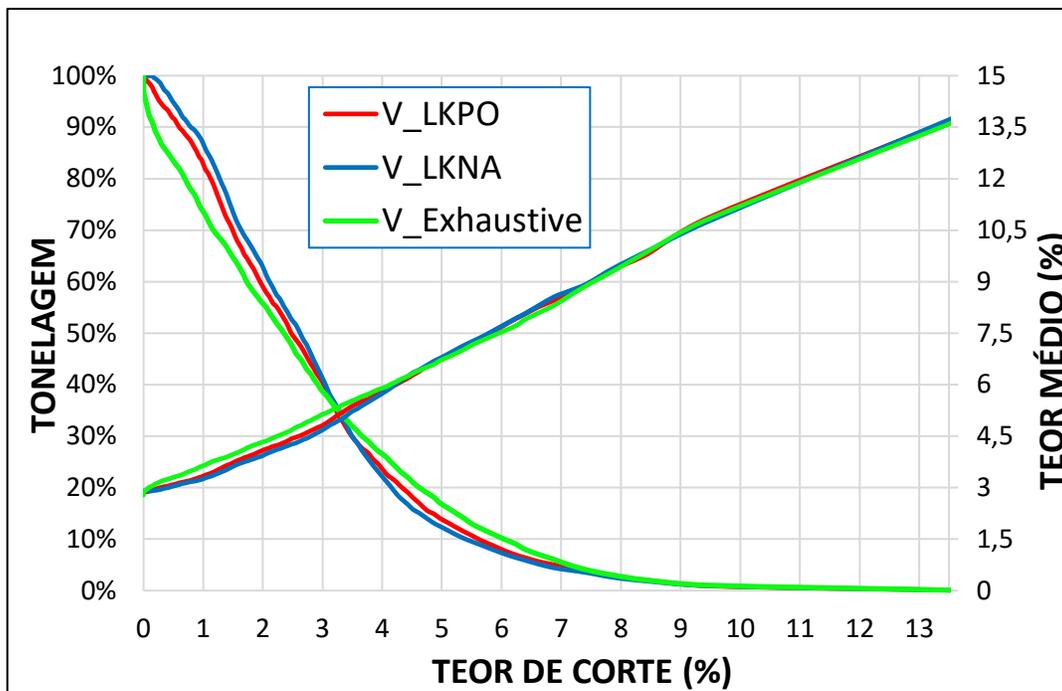


Figura 4.2. CASO I: Curvas teor-tonelagem. Modela LKPO (linha vermelha), Modela LKNA (linha azul) e Modelo de referência (linha verde).

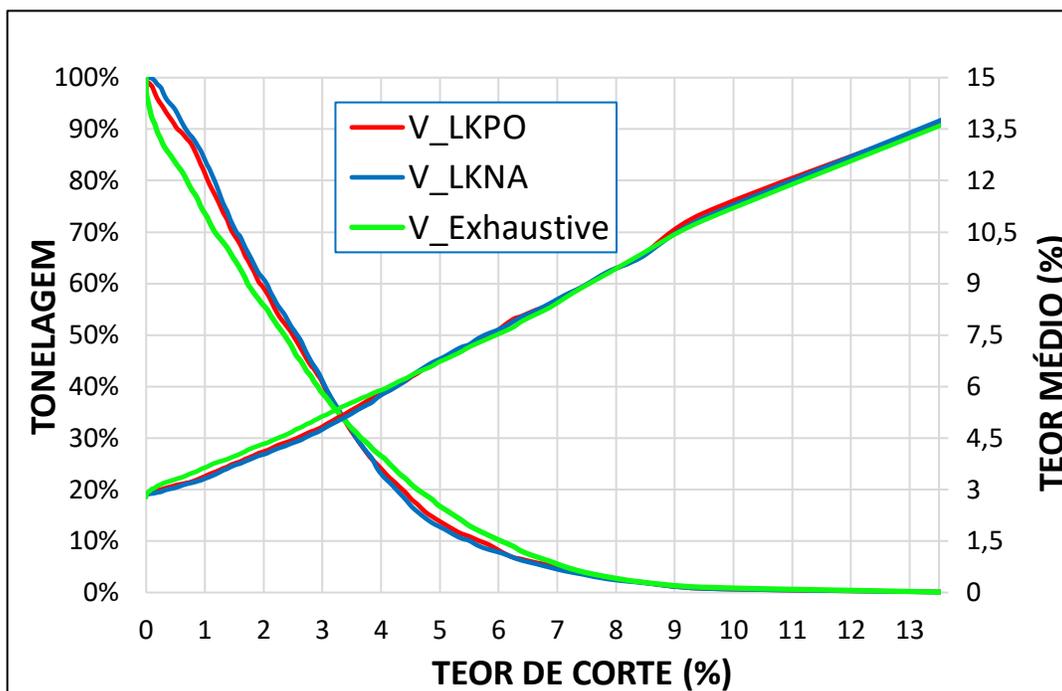


Figura 4.3. CASO II: Curvas teor-tonelagem. Modela LKPO (linha vermelha), Modela LKNA (linha azul) e Modelo de referência (linha verde).

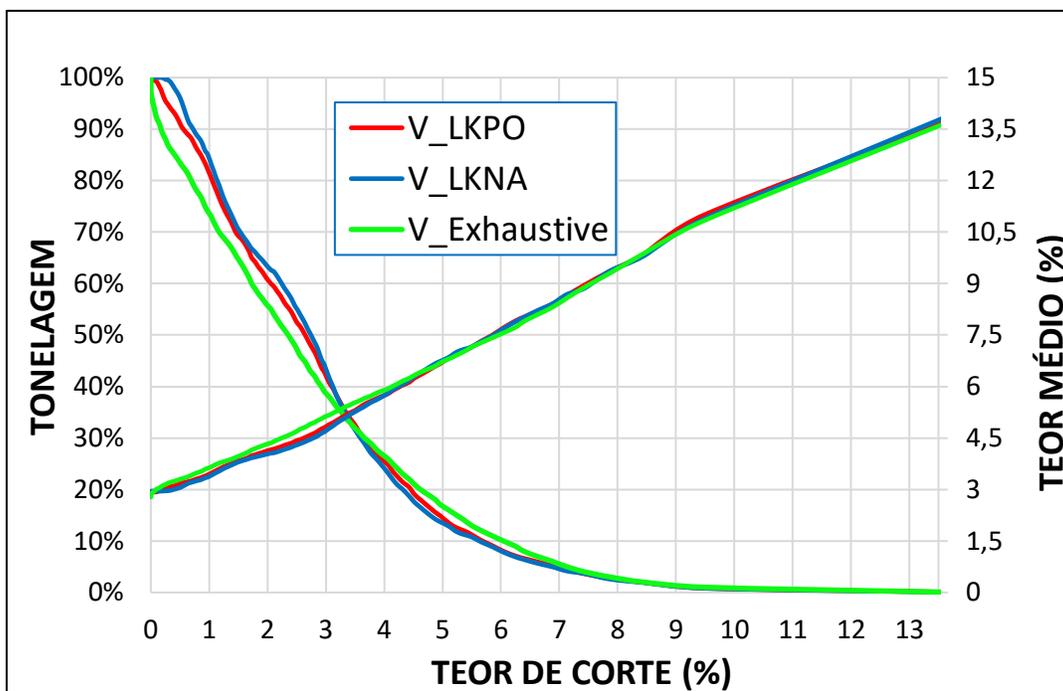


Figura 4.4. CASO III: Curvas teor-tonelagem. Modela LKPO (linha vermelha), Modela LKNA (linha azul) e Modelo de referência (linha verde).

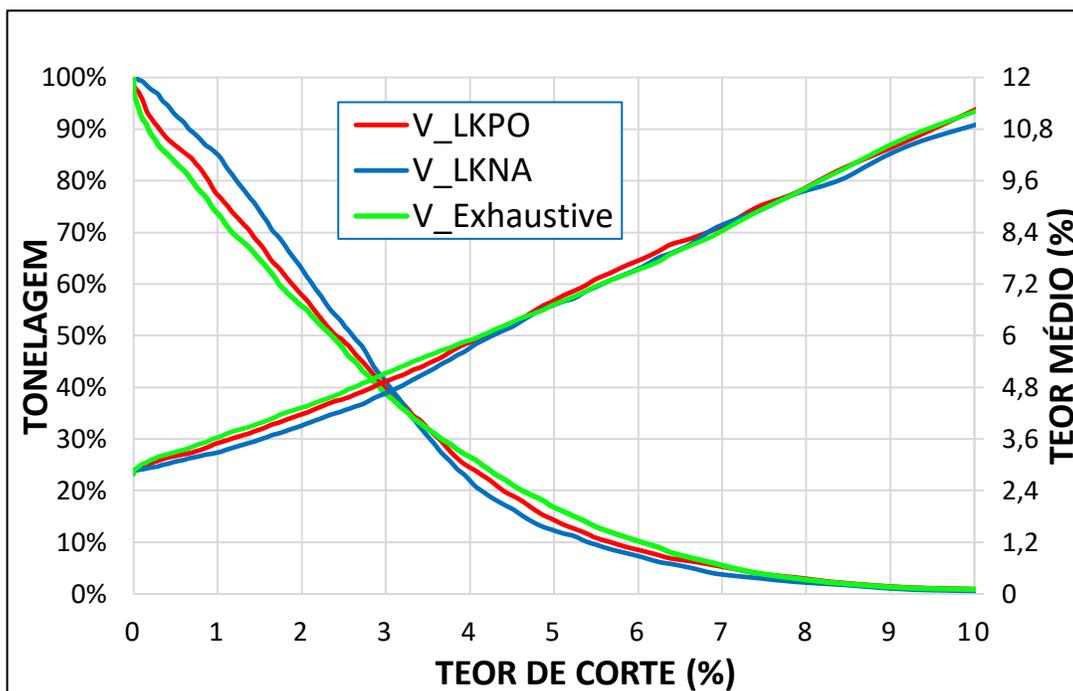


Figura 4.5. CASO IV: Curvas teor-tonelagem. Modelo LKPO (linha vermelha), Modelo LKNA (linha azul) e Modelo de referência (linha verde).

4.1.3. VARIAÇÕES DO ERRO NA CURVA TEOR-TONELAGEM

Para medir o comportamento do erro das curvas teor-tonelagem em relação ao modelo de referência, foi calculado o erro das duas curvas (em porcentagem) em relação a cada percentil da distribuição de teores. Usando como referência a linha preta no eixo X, as figuras 4.6 a 4.9 mostram, que as variações do erro (tanto na tonelagem como no teor médio) quase sempre são maiores utilizando o modelo da LKNA (linha azul) que quando utilizamos o modelo da LKPO (linha vermelha). Por exemplo, para um teor de corte de 2% (figura 33a) o erro na tonelagem do modelo LKNA é de aproximadamente 7% enquanto o erro no modelo LKPO, para o mesmo teor de corte, é de aproximadamente 3%. O mesmo pode ser observado nas variações do erro do teor médio.

De todos os quatro casos estudados, o modelo gerado pela LKPO teve melhor aderência às curvas teor-tonelagem e, por conseguinte, menores variações do erro entre essas duas curvas foi o caso IV (figura 4.9). Portanto, as seguintes análises estarão baseadas somente nesse caso, que utilizou o conceito de agrupamento de dados utilizando algoritmos de aprendizagem de máquinas para gerar cinco domínios geológicos.

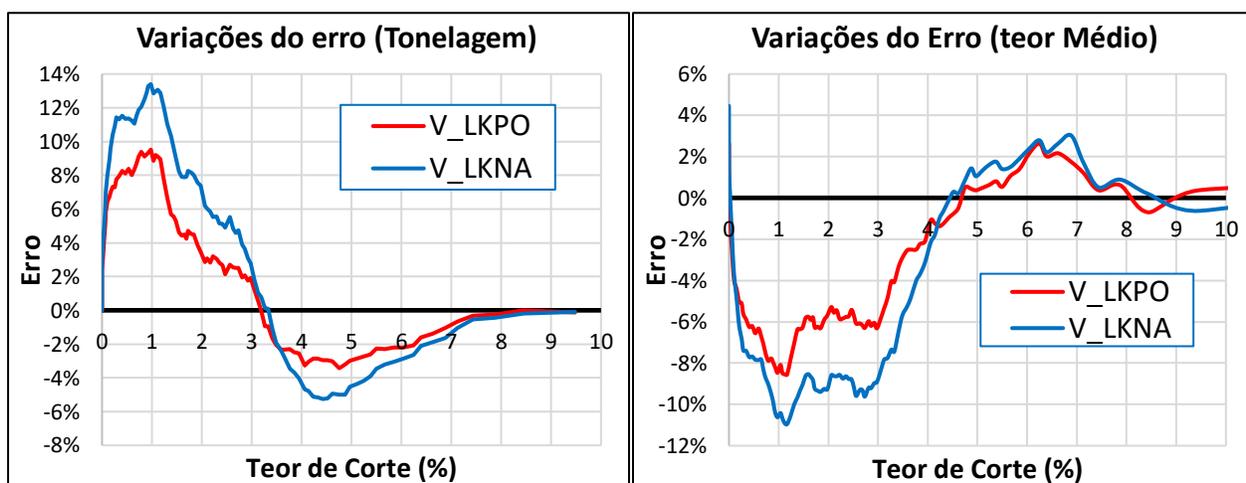


Figura 4.6. Caso I: (a) variações do erro na tonelagem (b) variações do erro no teor médio. Modelo da LKPO (linha vermelha), modelo da LKNA (linha azul) e modelo de referência (eixo X).

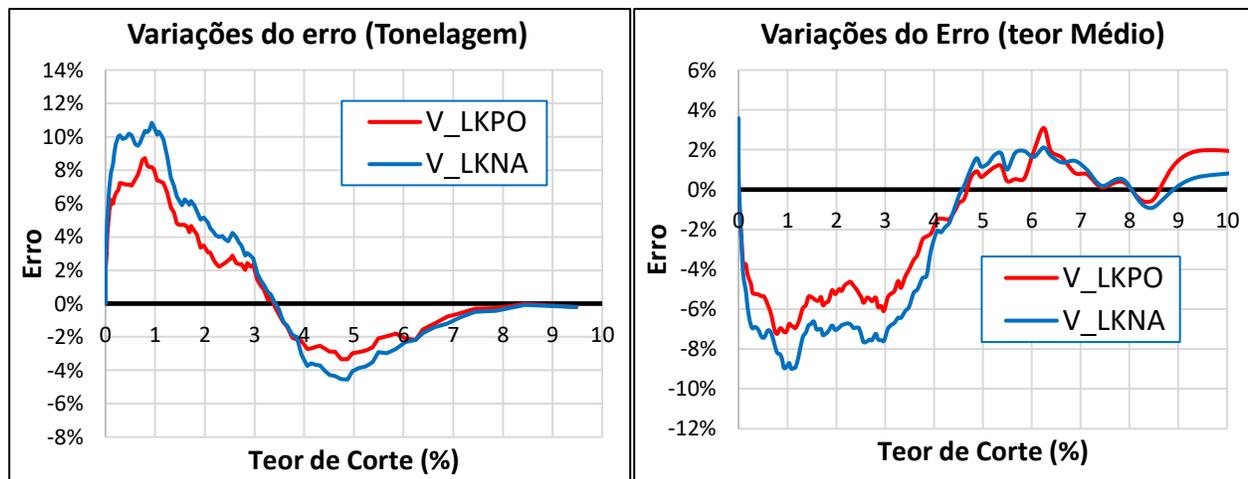


Figura 4.7. Caso II: (a) variações do erro na tonelagem (b) variações do erro no teor médio. Modelo da LKPO (linha vermelha), modelo da LKNA (linha azul) e modelo de referência (eixo X).

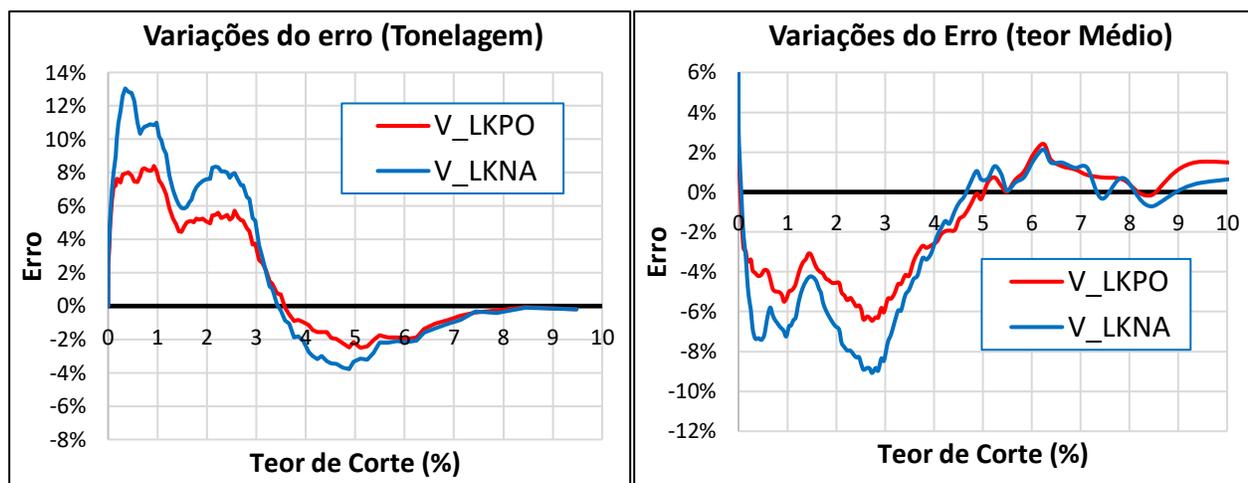


Figura 4.8. Caso III: (a) variações do erro na tonelagem (b) variações do erro no teor médio. Modelo da LKPO (linha vermelha), modelo da LKNA (linha azul) e modelo de referência (eixo X).

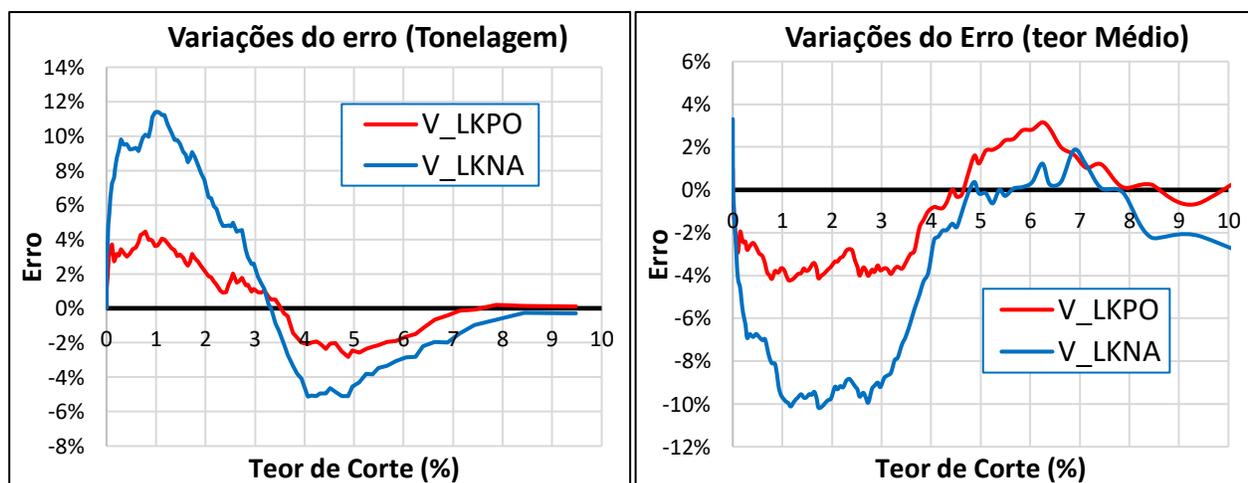


Figura 4.9. Caso IV: (a) variações do erro na tonelagem (b) variações do erro no teor médio. Modelo da LKPO (linha vermelha), modelo da LKNA (linha azul) e modelo de referência (eixo X).

4.1.4. ANÁLISE DE SUAUIZAÇÃO DA KRIGAGEM

A partir desse momento, a etapa de análise de suavização de krigagem (item 4.1.4) e a etapa de análise das médias locais (item 4.1.5) utilizarão como base os modelos que foram gerados no caso IV, já que esse modelo foi mais preciso e acurado que os mesmos modelos gerados nos casos I, II e III. As seguintes validações foram baseadas em uma análise de suavização da krigagem usando modelos de mudança de suporte. Alguns exemplos desses modelos podem ser encontrados na literatura geostatística, tais como: correção afim, lognormal indireta (IL) e gaussiana discreta (DG) (Isaaks e Srivastava 1989; Lajaunie 2000; Emery 2004, 2008; Matheron 1976, 1984, 1985). Para isso foi realizado, em primeiro lugar, um desagrupamento nos dados através do método de células móveis (Isaaks e Srivastava, 1989) para obter os pesos de cada amostra¹³. Posteriormente, foram utilizados esses pesos para realizar cada uma das correções de mudança de suporte citadas anteriormente. Semelhante ao comportamento das figuras das curvas teor-tonelagem vistas anteriormente, a figura 4.10 mostra que as duas curvas geradas pelos modelos da LKPO e as referências (afim, IL e DG) estão mais próximas entre si do que o modelo escolhido pelo método da LKNA. Claramente, isso demonstra que o efeito de suavização gerado no modelo da LKNA é maior do que o efeito de suavização gerado no modelo da LKPO.

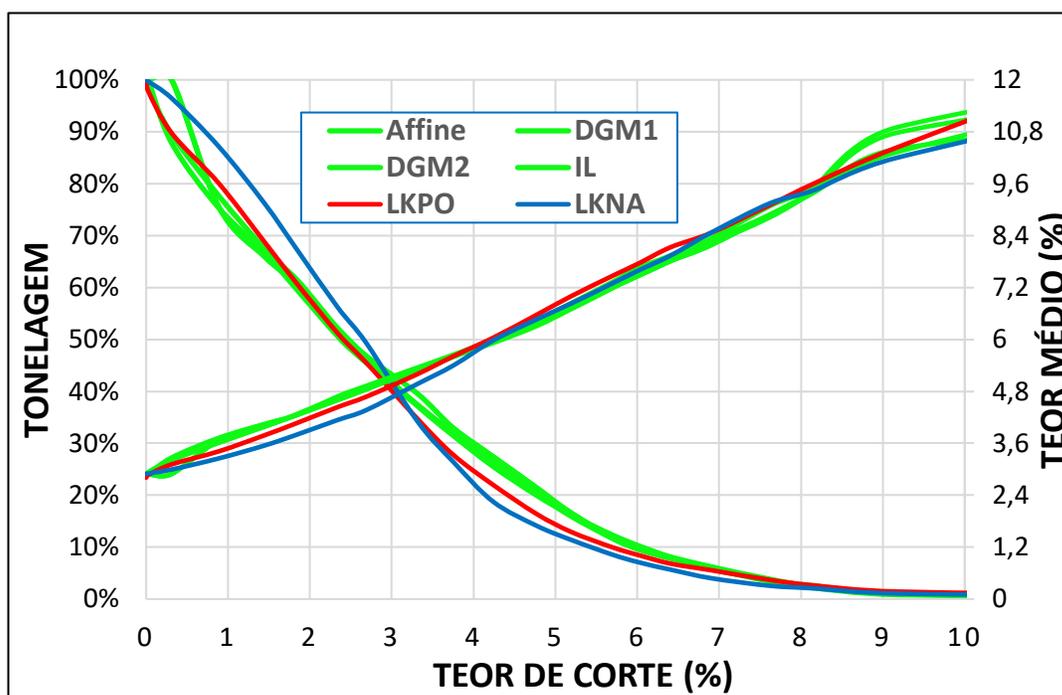


Figura 4.10. Curvas teor-tonelagem. Modelo LKPO (linha vermelha), Modelo LKNA (linha azul) e Modelos de referência obtida pela mudança de suporte (linhas verdes).

¹³ Segundo Isaaks e Srivastava (1989), no desagrupamento por células toda a área é dividida em regiões retangulares chamadas células. Cada amostra recebe um peso inversamente proporcional ao número de amostras que caem dentro da mesma célula. Amostras agrupadas geralmente receberão pesos menores com esse método, porque as células nas quais elas estão localizadas também conterão várias outras amostras.

4.1.5. ANÁLISE DAS MÉDIAS LOCAIS (ANÁLISE DE DERIVA)

Outras verificações de validações foram baseadas na análise de deriva, onde as médias locais foram plotadas em relação a cada modelo. Este método consiste basicamente em fazer comparações locais dos valores entre o modelo de blocos interpolado e o modelo de referência ao longo das bandas nas direções principais (X, Y e Z). A média é calculada para cada banda e o resultado é plotado em relação à sua localização ao longo das direções principais. Os gráficos mostram a coerência entre a média local ao longo de cada faixa nas duas estimativas.

As figuras 4.11 a e b mostram os gráficos de análise de deriva para os modelos LKPO, LKNA e de referência ao longo das duas direções principais, isto é, X e Y, respectivamente. Observe, que na maioria das partes do gráfico, as médias locais do modelo LKPO estão mais próximas dos valores do modelo de referência do que os valores derivados do modelo LKNA. As bandas usadas neste trabalho foram de 20 m de largura ao longo de X e 25 m ao longo de Y.

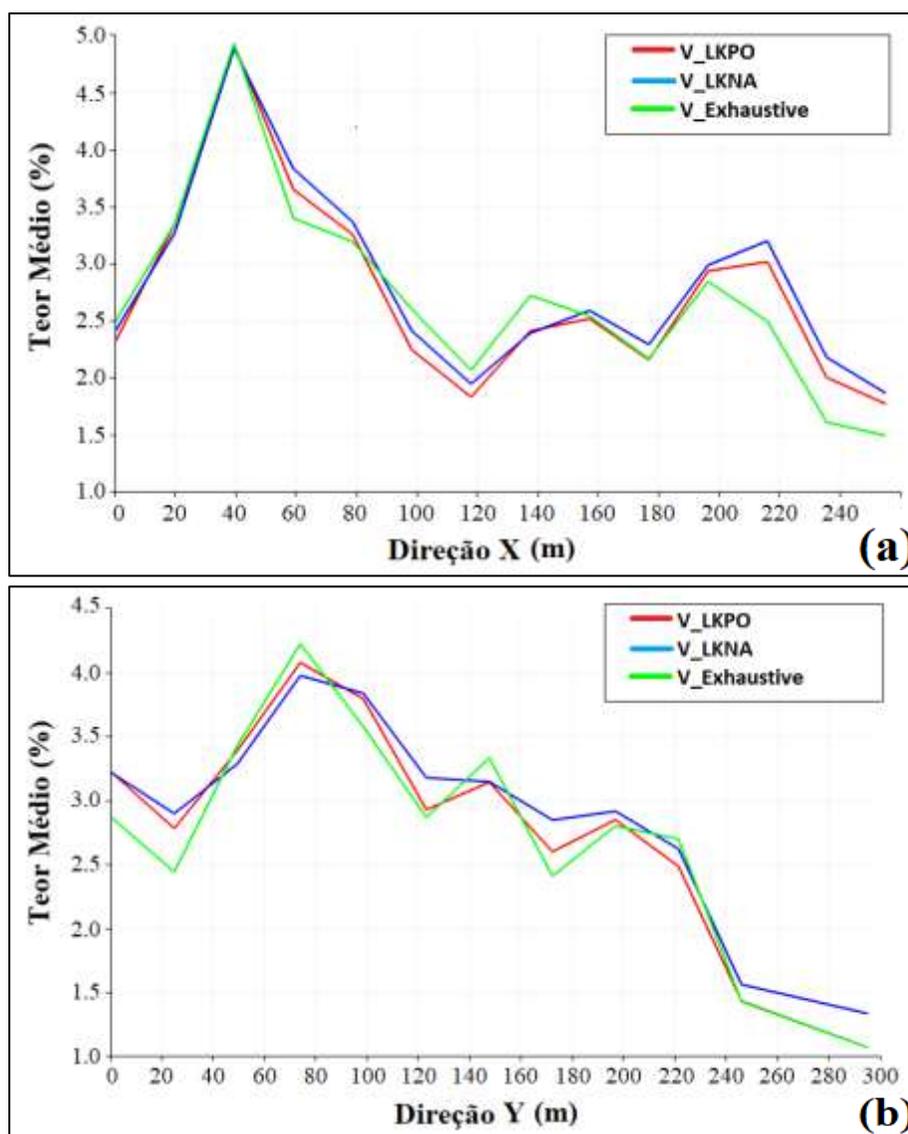


Figura 4.11. Análise de deriva ao longo das principais direções (a) plano X (b) plano Y.

4.1.6. ANÁLISE VISUAL

As últimas verificações de validação foram baseadas na análise visual. Este tipo de análise consiste em realizar comparações entre o modelo de bocós estimado e as amostras. Se o modelo estimado tiver resultados coerentes, as zonas de alto e baixo teor dos dados amostrais estarão correlacionadas com as zonas de alto e baixo teor do modelo estimado. Neste caso, como dispomos dos dados exaustivos do banco de dados, será realizada uma comparação mais precisa entre os modelos gerados na LKPO e LKNA com o modelo de referência (dados exaustivos). A figura 4.12 mostra o mapa de localização de cada um dos modelos analisados. Note que, apesar dos modelos LKPO (figura 4.12b) e LKNA (figura 4.12c) identificarem satisfatoriamente as zonas de alto e baixo teor em relação ao modelo de referência (figura 4.12a), o modelo LKNA é notoriamente mais suavizado que no modelo LKPO.

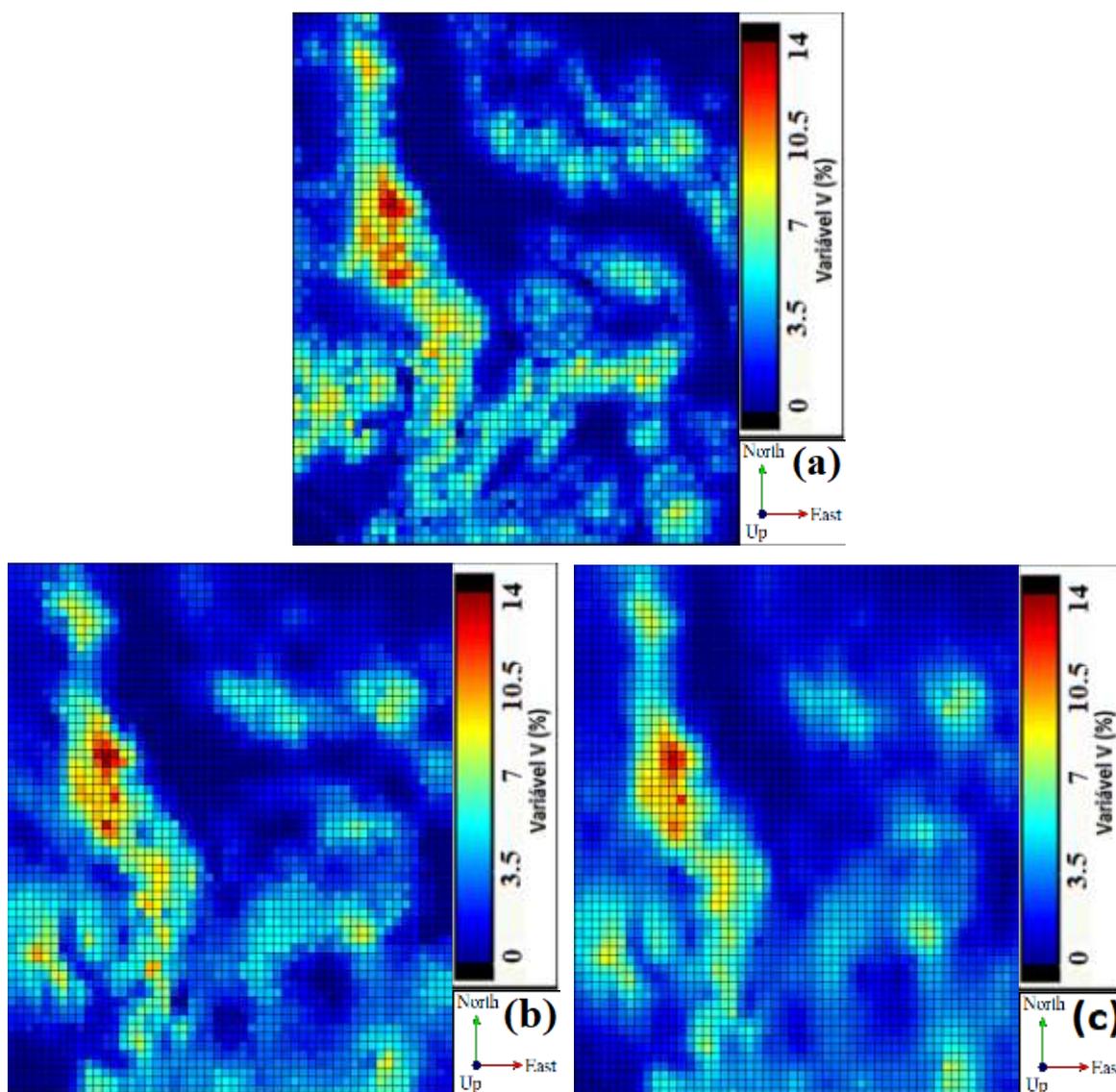


Figura 4.13. Análise visual (a) Dados Exaustivos (b) Modelo LKPO (c) Modelo LKNA.

4.2. BANCO DE DADOS 3D

Para avaliar os resultados do segundo estudo de caso, foi utilizado inicialmente como modelo de referência, os dados provenientes de *blasthole*¹⁴. As figuras 4.13 a-d mostram o mapa de localização dessas amostras com uma malha de amostragem de aproximadamente 7.5x7.5x7.5 m. A localização desses dados situa-se na parte central do depósito, que é a área aonde os domínios 3 e 4 possuem maior informação. O número de amostras para este conjunto de dados é de aproximadamente 160000.

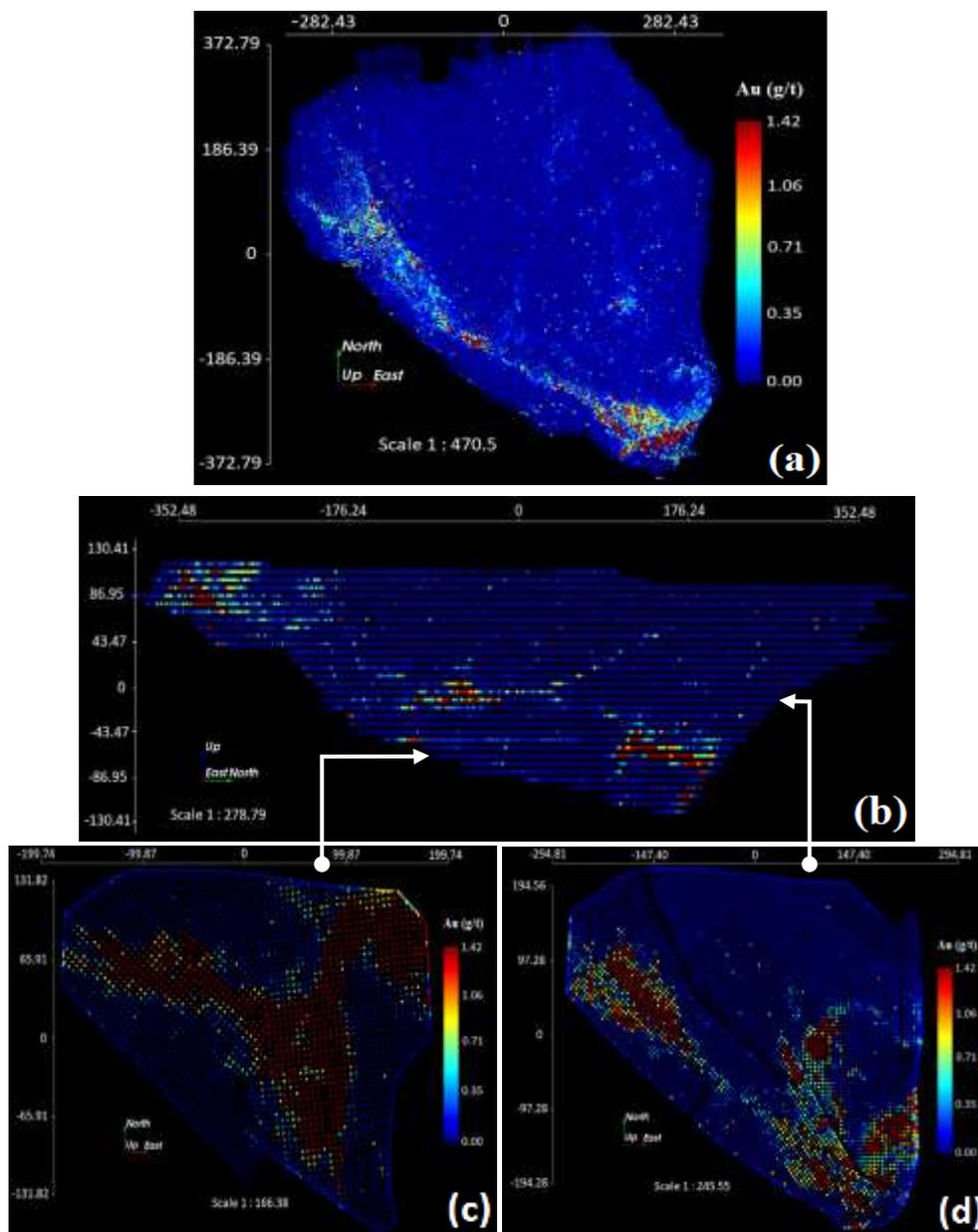


Figura 4.13. Mapa de localização das amostras de *blasthole*.

¹⁴ dados obtidos pela empresa de mineração apoiadora dessa tese

Como os dados de *blasthole* não possuíam nenhuma classificação por domínio, foi realizado o seguinte procedimento para obter cada modelo de referência:

- i. Para cada modelo de blocos dentro do domínio 3 e 4 (15x15x7.5 m), foi realizado uma sub-blocagem para um tamanho de blocos de 3x3x1.5 m.
- ii. Dentro de cada novo modelo, foi realizado uma estimativa pontual através do vizinho mais próximo utilizando os dados de *blasthole*.
- iii. Para cada estimativa no passo anterior, foi realizado uma re-blocagem para o tamanho de blocos original que corresponde ao mesmo suporte das estimativas (15x15x7.5 m).

É importante mencionar que esses dados abrangem apenas uma pequena parte da área amostrada dentro dos domínios 3 e 4, ou seja, os dados de *blasthole* não cobrem completamente a área total de cada domínio. As figuras 4.14 e 4.15 mostram os resultados da construção dos modelos de referência para o domínio 3 (BlastHole_d3) e o domínio 4 (BlastHole_d4), respectivamente. Note, que no primeiro modelo existem 9205 blocos, representando aproximadamente 16% de todo o modelo de blocos do domínio 3 (56153 blocos). Já para o segundo modelo, esse número aumenta para 10024 blocos, representando aproximadamente 17% de todo o modelo de blocos do domínio 4 (57375 blocos).

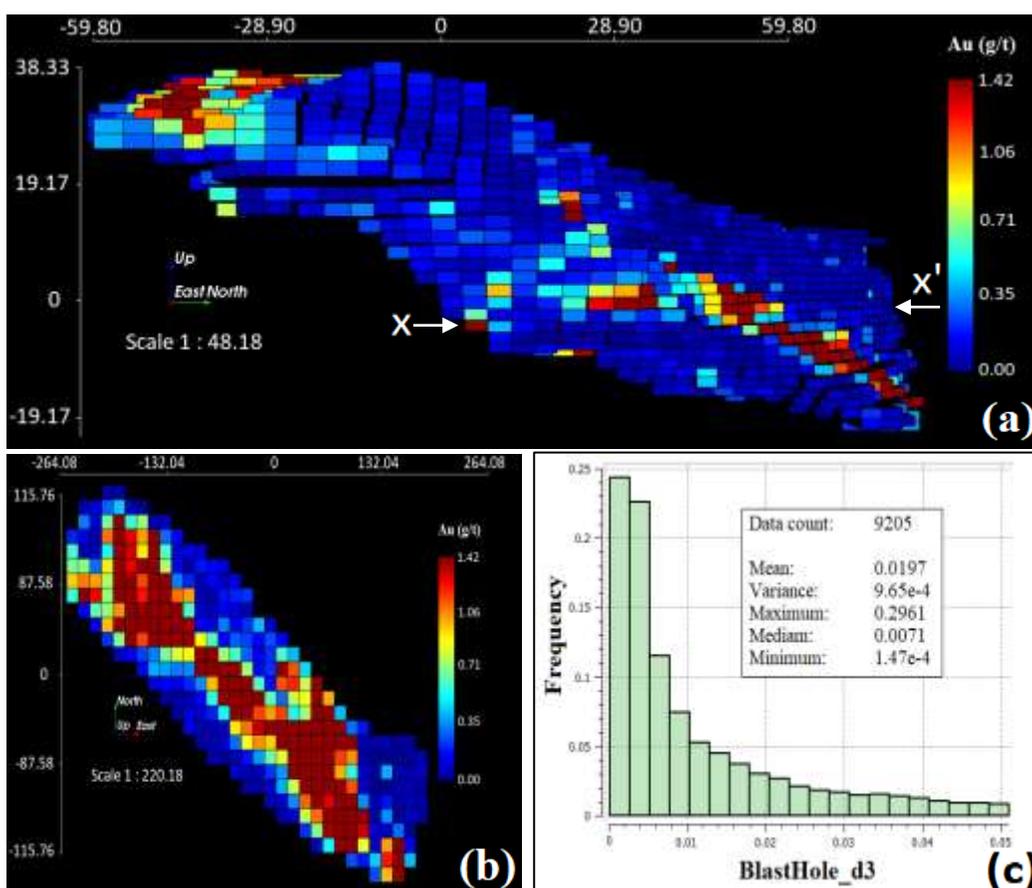


Figura 4.14. Modelo de referência do domínio 3, utilizando dados de *blasthole* (a) Mapa de localização (b) Seção x-x' (c) Histograma de frequências.

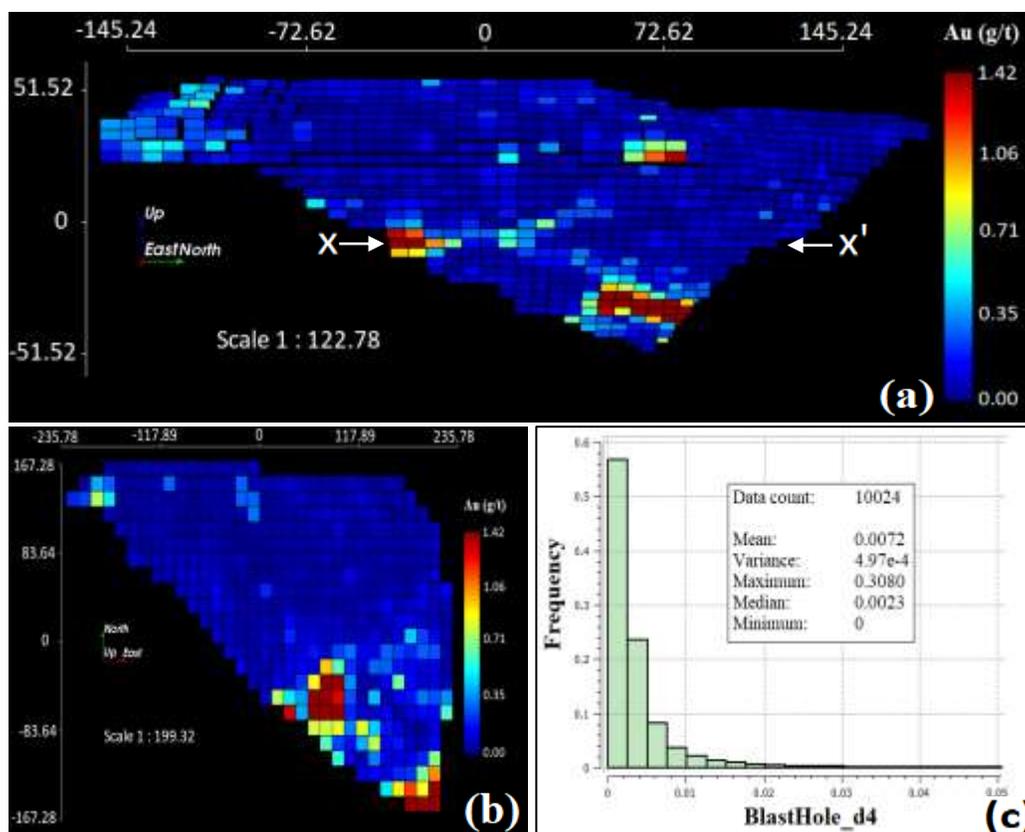


Figura 4.15. Modelo de referência do domínio 4, utilizando dados de *blasthole* (a) Mapa de localização (b) Seção x-x' (c) Histograma de frequências.

Para este caso específico, as primeiras análises de validação de modelos foram realizadas utilizando apenas os blocos que compartilham a mesma informação (mesmo que o modelo de blocos de cada domínio seja maior) tanto para a LKPO como para a LKNA. Na etapa seguinte de validação, será realizado uma mudança de suporte na qual serão aproveitados todos os blocos dentro de cada domínio.

4.2.1. CURVAS TEOR-TONELAGEM POR CADA DOMÍNIO

Uma primeira análise foi realizada utilizando as curvas teor-tonelagem já apresentadas anteriormente. Essas curvas mostram como o teor médio e a tonelagem variam ao longo de cada teor de corte. Neste estudo de caso, as curvas do modelo de referência (linha verde) são comparadas com as curvas do modelo da LKPO (linha vermelha) e o modelo da LKNA (linha azul) utilizando apenas os blocos que compartilham a mesma informação. Na figura 4.16a, podemos observar como as duas curvas do modelo da LKPO no domínio 3, na maioria das vezes, estão mais próximas ao modelo de referência do que as mesmas curvas do modelo da LKNA. Em relação a tonelagem (lado esquerdo), por exemplo, aproximadamente 80% do modelo da LKPO se aproxima mais ao modelo de referência (linhas verdes) do que as curvas geradas pelo modelo da LKNA (linhas

azuis). Nos últimos 20% da tonelagem, o modelo da LKNA apresenta leves melhorias em relação ao modelo da LKPO. Por outra parte, um comportamento similar pode ser observado no teor médio (lado direito), onde o modelo da LKPO acompanha claramente melhor essa tendência do que o modelo da LKNA ao longo de cada teor de corte. Para o domínio 4, uma análise similar pode ser observada nessas duas curvas (figura 4.16b).

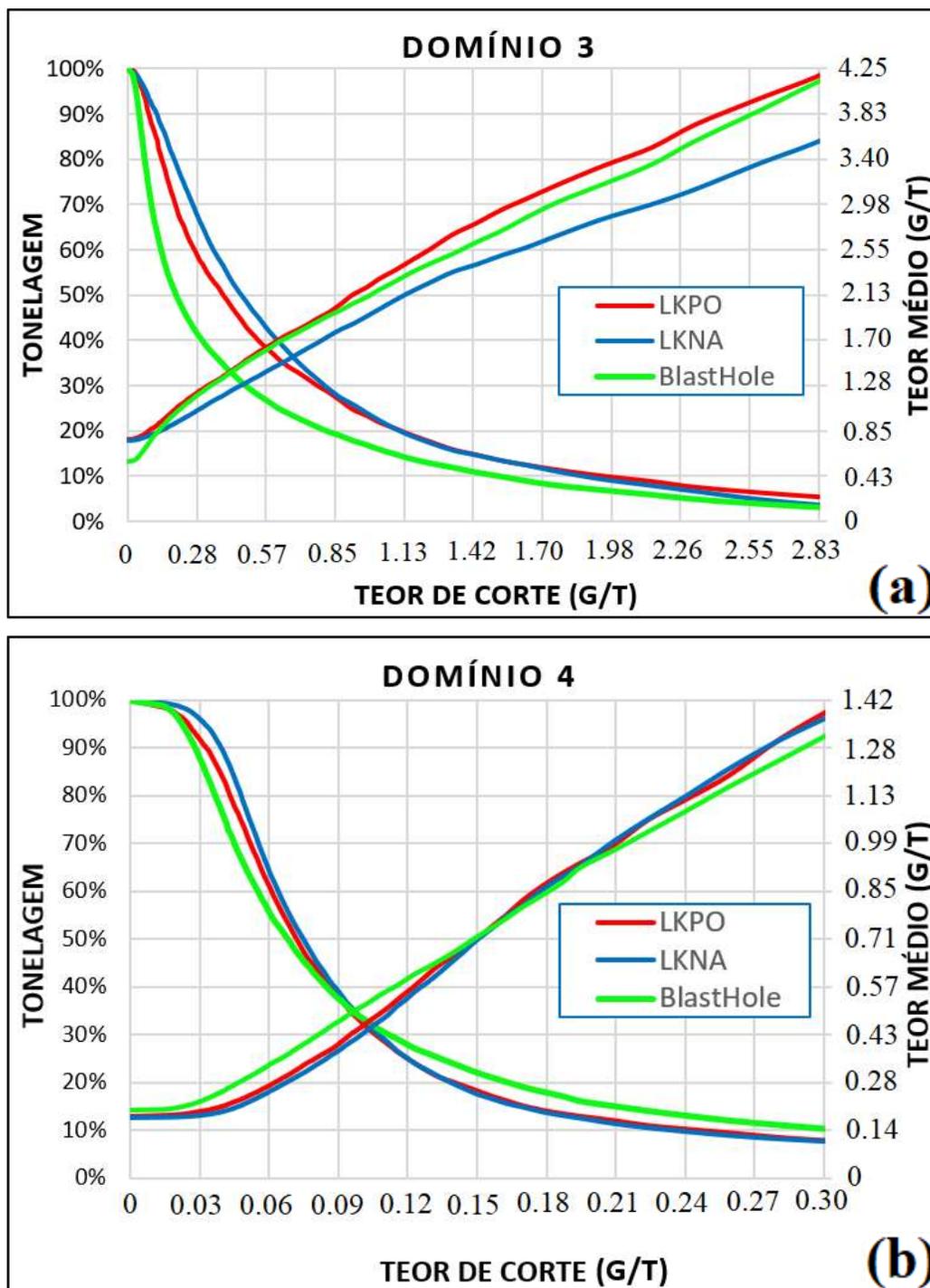


Figura 4.16. Curvas teor-tonelagem (a) domínio 3 (b) domínio 4.

4.2.2. VARIAÇÕES DO ERRO NA CURVA TEOR-TONELAGEM

Similar ao estudo de caso anterior, as figuras 4.17 e 4.18 mostram as variações do erro das curvas teor-tonelagem em relação ao modelo de referência para os domínios 3 e 4, respectivamente. O erro da curva tonelagem (em porcentagem) foi calculado como a diferença entre a tonelagem do modelo obtido com base nos dados de “blasthole” e a tonelagem de cada modelo de otimização (figuras 4.17a e 4.18a). Por outro lado, o erro da curva de teor médio (em porcentagem) foi calculado como a diferença entre o teor médio do modelo obtido com base nos dados de “blasthole” e teor médio de cada modelo de otimização (figuras 4.17b e 4.18b). O eixo X (linha preta) representa o modelo de referência em relação a cada percentil da distribuição dos teores. Na maior parte dos gráficos, essas variações são menores quando utilizamos o modelo da LKPO (linha vermelha) do que quando utilizamos o modelo da LKNA (linha azul).

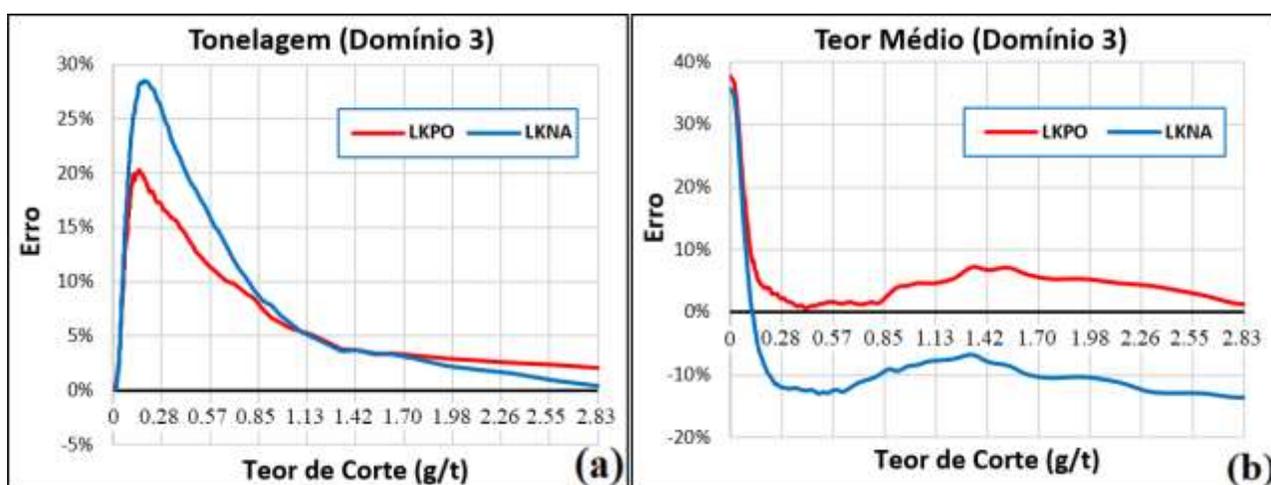


Figura 4.17. Variações do erro do domínio 3 (a) Tonelagem (b) Teor médio.

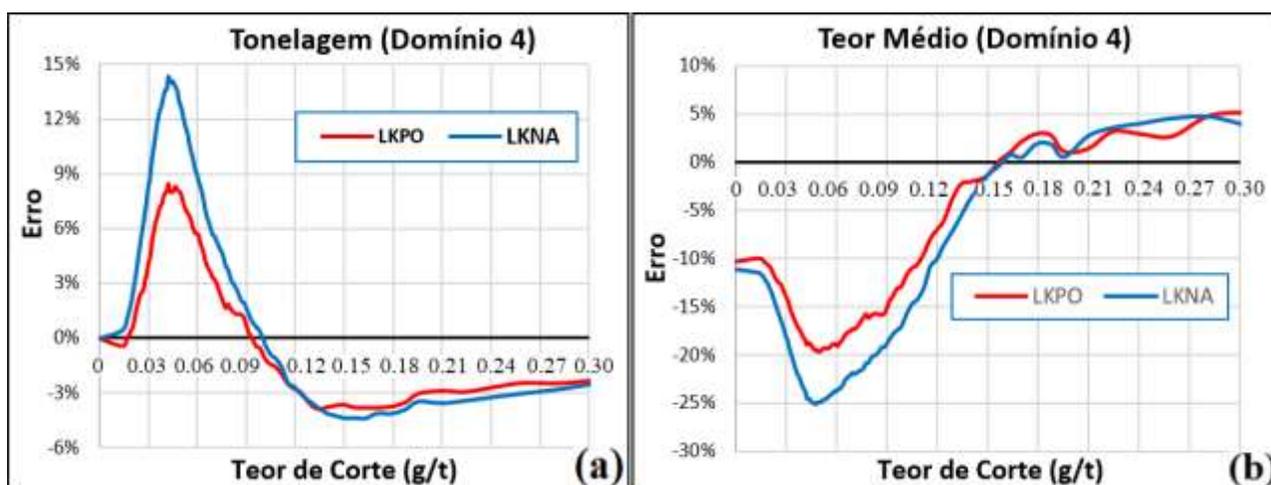


Figura 4.18. Variações do erro do domínio 4 (a) Tonelagem (b) Teor médio.

4.2.3. UNINDO OS DOMÍNIOS 3 E 4

Como foi mencionado anteriormente, os dados de referência para este estudo de caso representam aproximadamente 16% e 17% da área total dos domínios 3 e 4 respectivamente. A seguinte análise foi baseada a partir na unificação desses dois modelos, visando ter uma maior informação em relação ao número de blocos estimados. Esse mesmo procedimento foi realizado para os domínios 3 e 4 a fim de comparar blocos que compartilham a mesma informação. A figura 4.19a mostra as curvas teor-tonelagem para esse novo modelo. Note, que o comportamento das duas curvas é similar aos casos anteriores, onde a metodologia da LKPO levou a modelos mais precisos e acurados do que o uso da metodologia da LKNA. Essa precisão é medida através das variações do erro na curva da tonelagem (figura 4.19b) e a curva do teor médio (figura 4.19c).

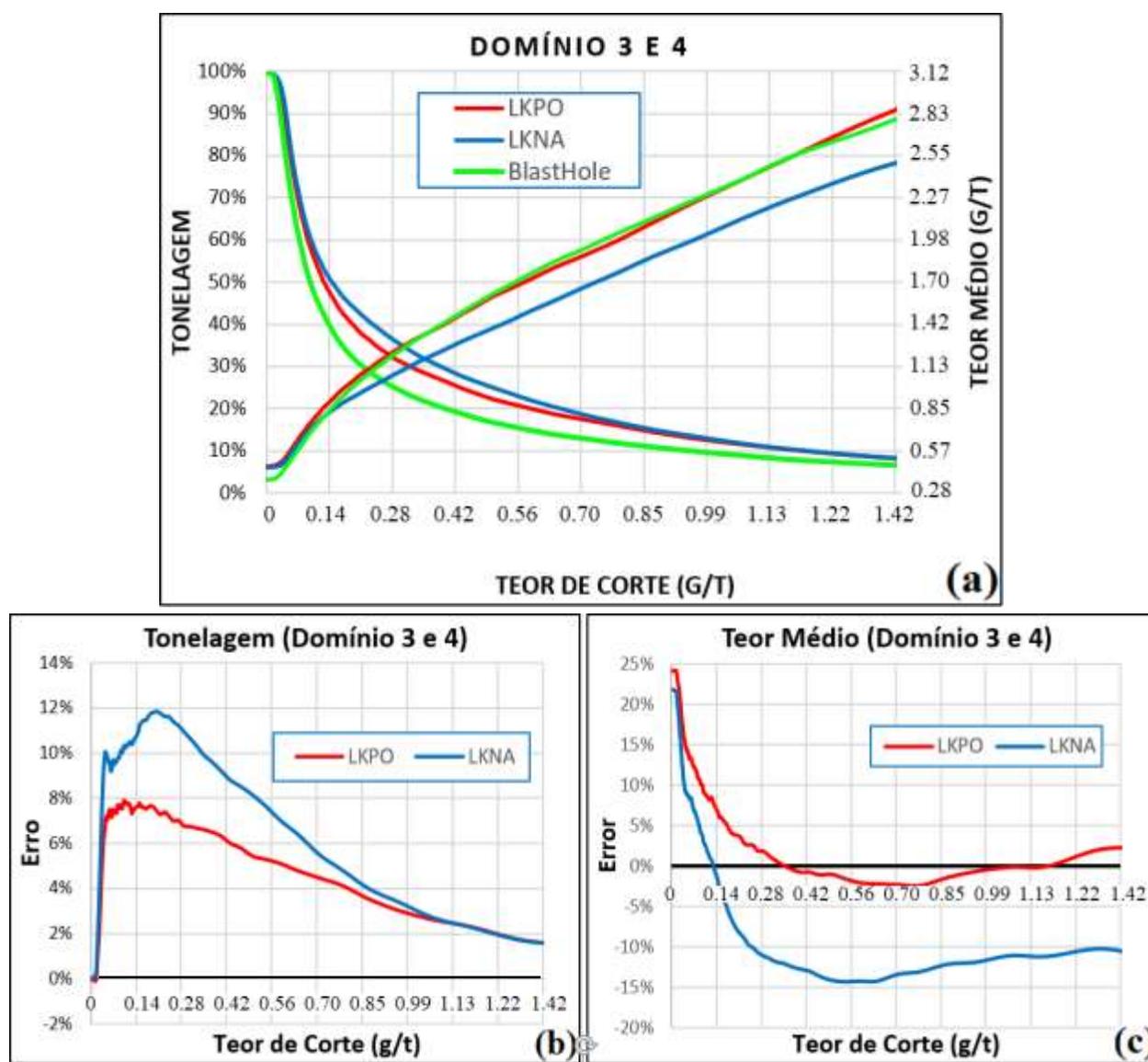


Figura 4.19. Domínios 3 e 4 unificados (a) Curvas teor-tonelagem (b) Variações do erro na tonelagem (c) Variações do erro no teor médio.

4.2.4. RESUMO ESTATÍSTICO

A fim de obter estatísticas representativas do depósito, foi realizado o desagrupamento dos dados através do vizinho mais próximo (NN) para cada domínio (dentro de cada modelo de blocos). A tabela 6 mostra as estatísticas básicas (média, mediana e variância) e os erros relativos calculados através da equação 25. Note, que em ambos domínios, o modelo LKPO apresenta algumas melhoras significativas em relação ao modelo LKNA, principalmente, na mediana (pouco afetada pela influência de valores extremos) e na variância (indicando o grau de suavização das estimativas). Entre mais próximos os erros relativos de zero (0), mais acuradas serão as estimativas em relação ao modelo de referência.

Tabela 6. Resumo estatístico por cada domínio no banco de dados 3D.

Variável	Domínio	Estatística Básica			Erro Relativo (%)		
		Média (g/t)	Mediana (g/t)	Variância (g/t) ²	Média	Mediana	Variância
NN		0.215	0.075	0.014			
LKPO	3	0.219	0.086	0.009	2.05	15.24	-33.09
LKNA		0.221	0.093	0.007	3.17	24.20	-52.06
NN		0.241	0.068	0.0159			
LKPO	4	0.237	0.084	0.0102	-1.99	23.24	-35.85
LKNA		0.235	0.093	0.0079	-2.82	37.21	-50.31

4.2.5. ANÁLISE DAS MÉDIAS LOCAIS

Aproveitando o modelo desagrupado do item anterior, foi utilizado o mesmo modelo de referência para calcular as médias locais em cada domínio ao longo das três direções principais, isto é, X, Y e Z. Nas figuras 4.20 e 4.21 são apresentados os gráficos (em cada direção) das médias locais para os domínios 3 e 4, respectivamente. Observe, que na maioria das partes do gráfico, as médias locais do modelo LKPO estão mais próximas dos valores do modelo de referência do que os valores derivados do modelo LKNA.

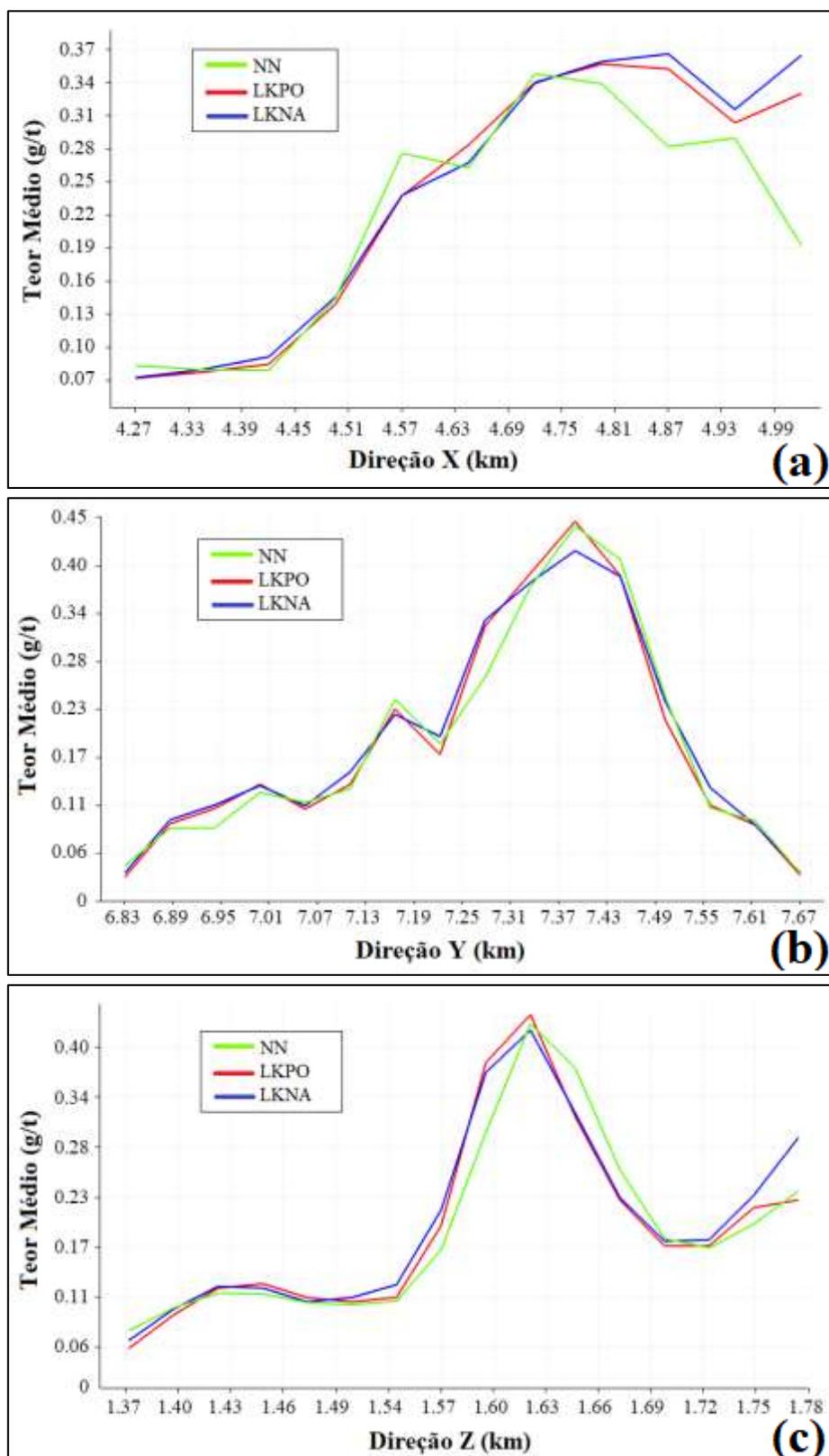


Figura 4.20. Análise de deriva do domínio 3 ao longo das principais direções (a) plano X (b) plano Y (c) plano Z.

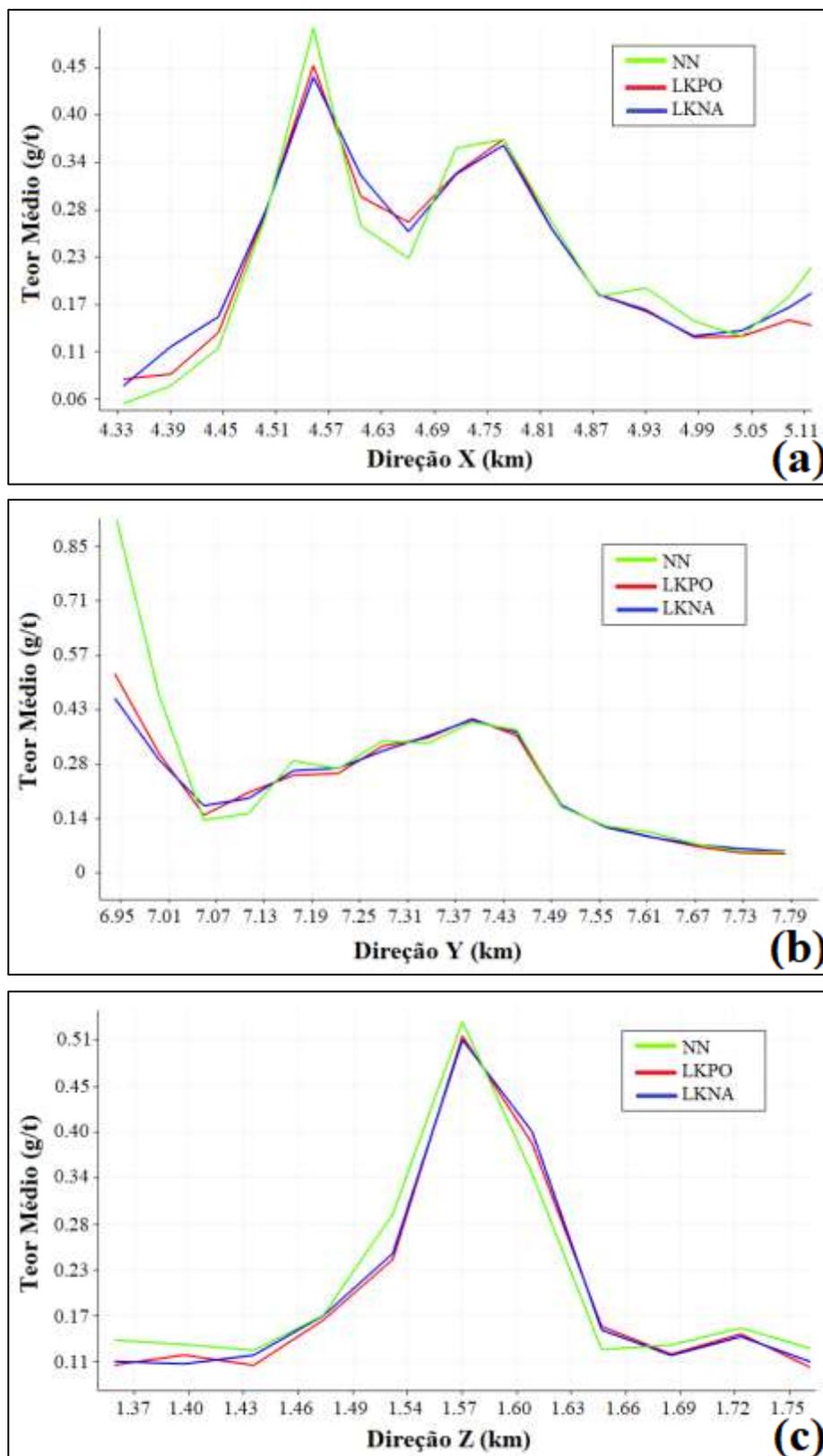


Figura 4.21. Análise de deriva do domínio 4 ao longo das principais direções (a) plano X (b) plano Y (c) plano Z.

4.2.6. ANÁLISE DE SUAUIZAÇÃO DA KRIGAGEM

Utilizamos o mesmo teste de suavização de krigagem apresentado anteriormente. Essas verificações de validação usam modelos de mudança de suporte para estimar a distribuição de teores do bloco a partir de uma distribuição conhecida de teores de amostras. A mudança de suporte foi realizada utilizando todas as amostras do conjunto de dados para todo o modelo de blocos dos domínios 3 e 4. Portanto, as comparações realizadas cobrem 100% da área de cada domínio. As figuras 4.22 a e b mostram as curvas de tonelagem-teor geradas pelos modelos LKPO, LKNA e as referências (afim, IL e DG) nos domínios 3 e 4, respectivamente. Note, que quase sempre, as curvas geradas pela LKPO (linha vermelha) estão mais próximas dos modelos de referência (linhas verdes) do que a LKNA (linha azul).

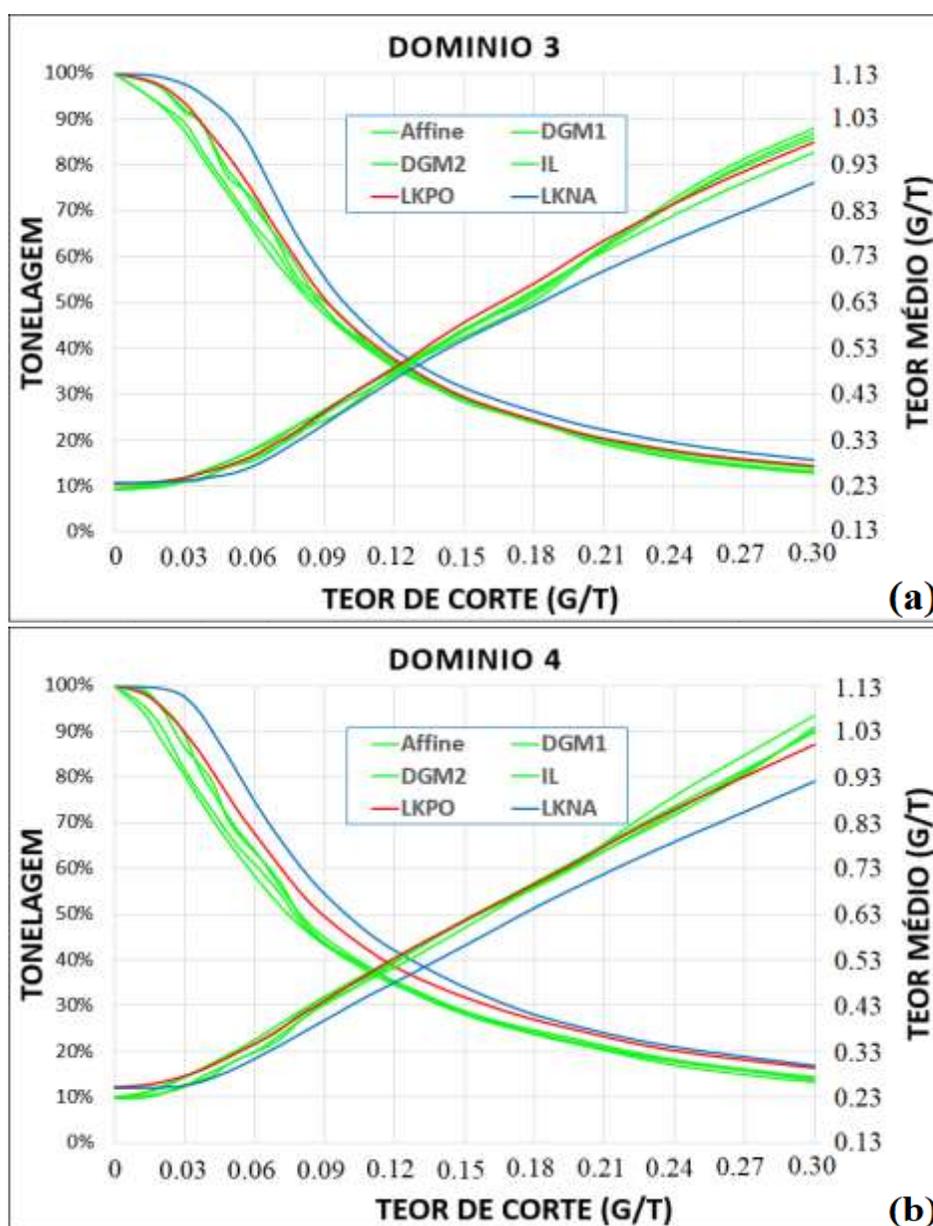


Figura 4.22. Curvas teor-tonelagem. Modelo LKPO (linha vermelha), Modelo LKNA (linha azul) e Modelos de referência obtida pela mudança de suporte (linhas verdes).

4.2.7. ANÁLISE DA PERFORMANCE DO ALGORITMO DA LKPO

Para ter-se uma ideia sobre o tempo de processamento do algoritmo da LKPO, foram utilizados dois computadores com as seguintes configurações¹⁵:

Tabela 7. Resumo das configurações básicas dos computadores.

Sistema	Computador #1	Computador #2
Processador	Intel(R) Core (TM) i5-6ª Geração CPU @ 2.3GHz	Intel(R) Core (TM) i7-4ª Geração CPU @ 2.4GHz
Memória RAM	8 GB	32 GB
Tipo de Processador	64-bit	64-bit

O número de amostras e o número de blocos dentro de cada domínio é apresentado na seguinte tabela:

Tabela 8. Número de amostras é número de blocos por domínio.

Domínio	Amostras	Blocos
3	2701	56153
4	4114	57375

A construção do algoritmo da LKPO foi realizada através de uma série de funções (processos) compatível com a sequência de passos proposta nesta metodologia (ver capítulo 1). Cada processo é realizado sequencialmente até obter o resultado final de otimização, são eles:

- i. *Cross-Validation*: Realiza todo o processo de interpolação através da técnica de validação cruzada para todas estratégias de busca definidas pelo usuário.
- ii. *Post-Processing*: Escolhe os parâmetros das estratégias de busca que minimizam a função do erro absoluto para cada amostra.
- iii. *Influence Area*: Cria a área/volume de influência (em blocos) ao redor de cada amostra.
- iv. *Kriging*: Realiza todo o processo de estimativa utilizando os parâmetros do processo *ii* dentro de cada área/volume de influência do processo *iii* (bloco a bloco).
- v. *Final Block Model*: Agrupa todos os blocos individuais dentro do modelo de blocos que está sendo estimado.
- vi. *Post-Processing*: Apaga ou elimina variáveis temporárias do passo *iii*.

¹⁵ Ressalta-se, que as configurações dos processadores aqui apresentados, servem somente para ter uma ideia aproximada do tempo de execução do algoritmo. Pequenas mudanças nessas configurações (e.g., passar de uma geração para outra) pode ter impactos significativos nos tempos de execução do algoritmo.

vii. *Post-Processing*: Apaga ou elimina variáveis temporárias do passo *iv*.

As figuras 4.23 a e b mostram o tempo acumulado (minutos) de cada computador durante a execução de cada processo nos domínios 3 e 4. Note, que para o caso do computador # 1 (linha azul), o tempo total é de aproximadamente 30 minutos para domínio 3 e 50 minutos para o domínio 4. Essa diferença (aproximadamente 40%) deve-se ao fato do domínio 4 ter mais amostras do que o domínio 3 (aproximadamente 34%), o que indica claramente que o número de amostras tem um grande impacto no desempenho do algoritmo. No caso do computador # 2 (linha laranja), essas diferenças entre os tempos totais diminuíram em 39% e 43% para ambos os domínios respectivamente, reduzindo de 30 para 18 minutos no domínio 3 e de 50 para 29 no domínio 4. Esta melhoria significativa é devida à capacidade computacional do computador # 2. É importante mencionar, que os maiores tempos de execução dos processos na LKPO são usados criação das áreas de influência (ou volumes) e nas estimativas (krigagem) de cada uma dessas áreas.

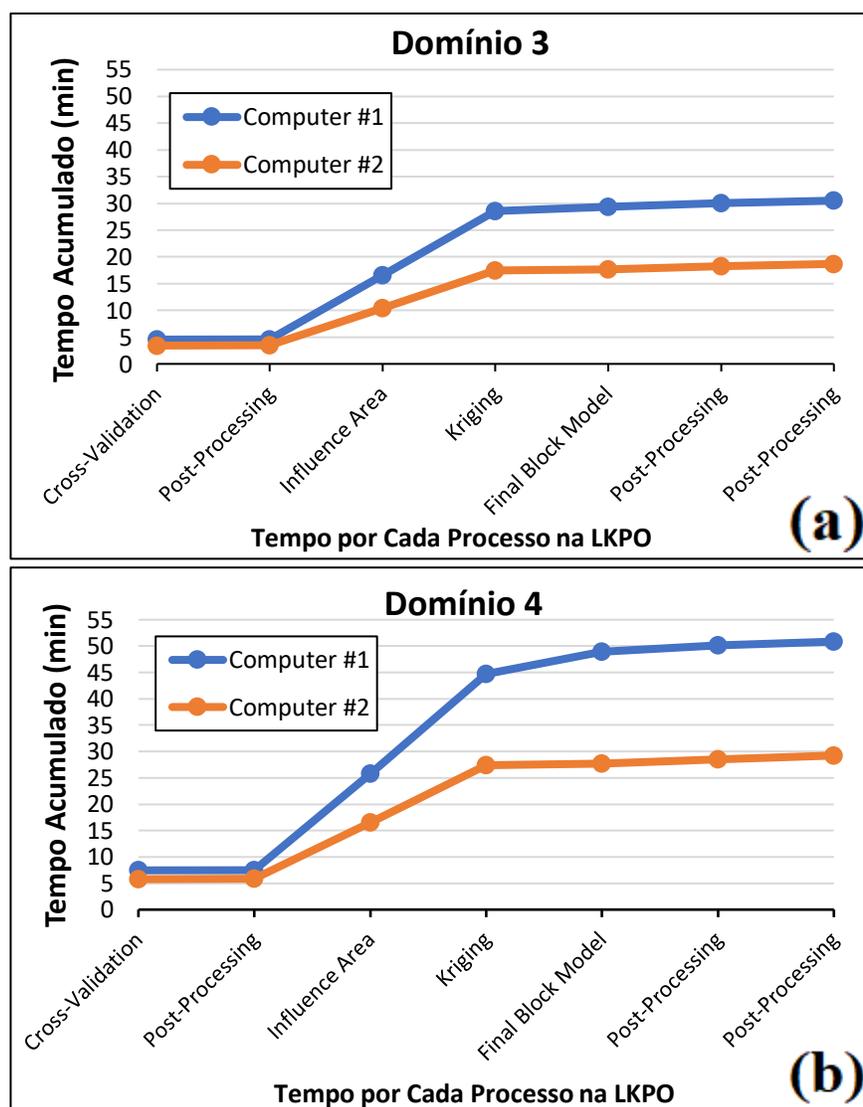


Figura 4.23. Tempos obtidos pelo algoritmo da LKPO (a) domínio 3 (b) domínio 4.

CAPÍTULO 5 – RESULTADOS E ANÁLISES DA LKPO UTILIZANDO SIMULAÇÕES

Para avaliar os resultados obtidos na LKPO utilizando simulações, foi utilizado o banco de dados 2D definido anteriormente no capítulo 3. Similar ao capítulo anterior, cada resultado será comparado com: o modelo LKNA, o modelo de referência e os modelos de mudança de suporte. Finalmente, serão utilizados os métodos de validações atualmente disponível na literatura, como reprodução de estatísticas básicas, de curvas teor-tonelagem, grau de suavização da krigagem e análise das médias locais (análise de deriva) para avaliar a precisão e acurácia de cada modelo.

5.1. VERIFICAÇÃO DOS MODELOS SIMULADOS

Nesse processo de LKPO, que utiliza como métrica as simulações na minimização do erro, é importante gerar o espaço de incerteza adequado para o conjunto de dados que está sendo analisado. Portanto, devemos ser criteriosos na hora de realizar o processo de simulação e verificar se elas representam, de fato, a variabilidade real do depósito e o espaço de incerteza adequado. Para isso, existem várias maneiras de verificar os modelos simulados, entre eles estão:

5.1.1. REPRODUÇÃO DOS HISTOGRAMAS

As figuras 5.1a até 5.1e mostram os histogramas acumulados dos modelos simulados (linhas verdes) e os dados desagrupados (linha vermelha) para cada um dos domínios definidos no caso IV. Note, que em todos os gráficos e para a maioria dos teores (exceto teores acima de 4% no domínio 4) as simulações reproduzem satisfatoriamente o histograma dos dados desagrupados (obtido através do vizinho mais próximo).

5.1.2. REPRODUÇÃO DOS VARIOGRAMAS

As figuras 5.2a até 5.2e mostram os variogramas dos modelos simulados (linhas verdes) e o modelo dos dados originais (linha vermelha) nas direções preferenciais para cada um dos domínios definidos no caso IV. Note, que o modelo dos dados originais, em todos os domínios, se adere às flutuações ergódicas dos modelos simulados.

5.1.3. GRÁFICO DE ACURÁCIA

A representação gráfica dos modelos simulados em gráficos de acurácia fornece uma ideia geral de quão precisas e acuradas são as simulações. As figuras 5.3a até 5.3e mostram os gráficos

de acurácia para cada um dos domínios estudados no caso IV. Observe, que todos os modelos simulados possuem uma boa acurácia. Isto significa, que dentro dos intervalos de confiança estabelecidos no gráfico, a probabilidade de acerto é sempre maior do que o intervalo de confiança.

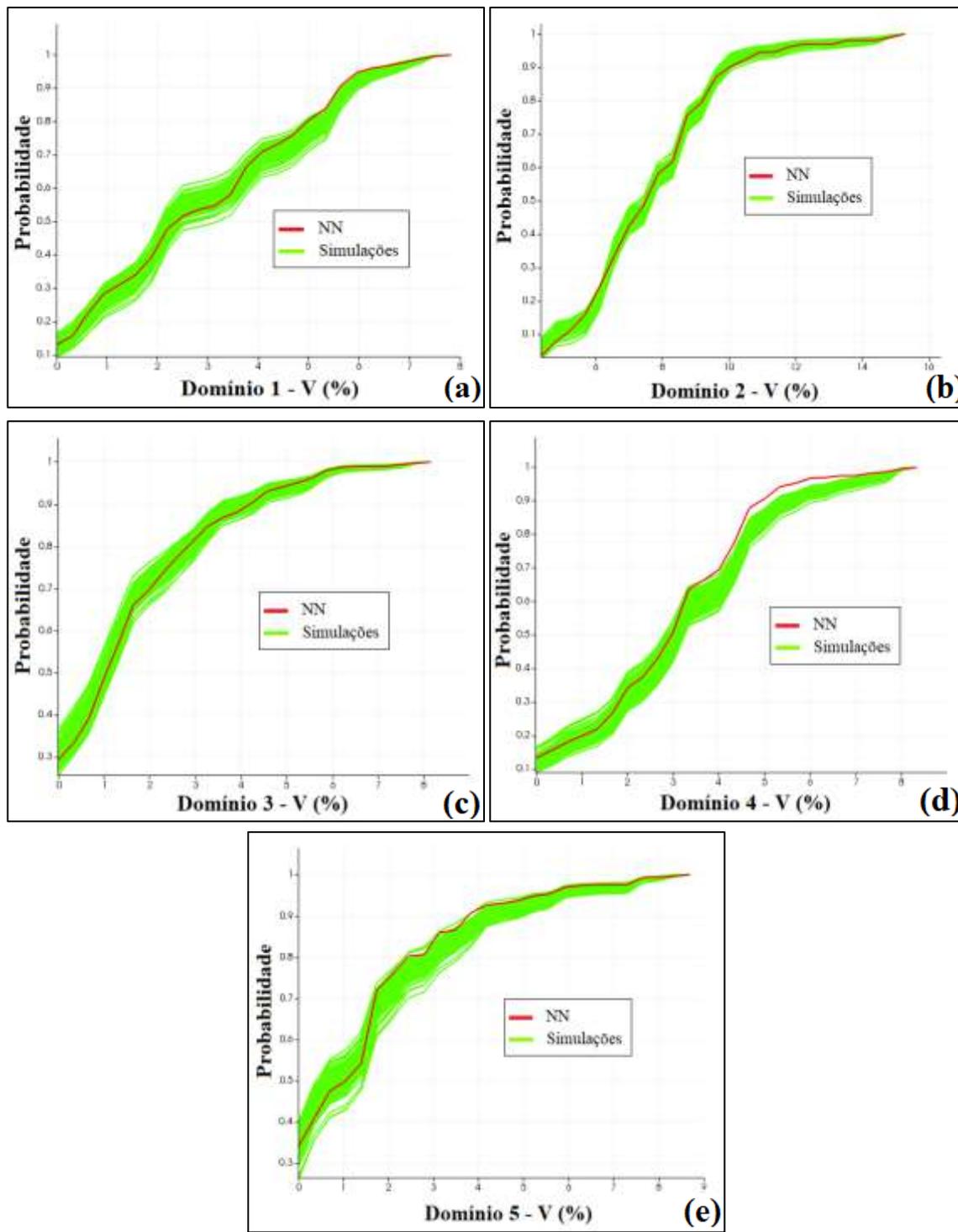


Figura 5.1. Validação dos modelos simulados (linhas verdes) através dos histogramas acumulados, usando como referência os dados desagrupados (linha vermelha) para o caso IV (a) domínio 1 (b) domínio 2 (c) domínio 3 (d) domínio 4 (e) domínio 5.

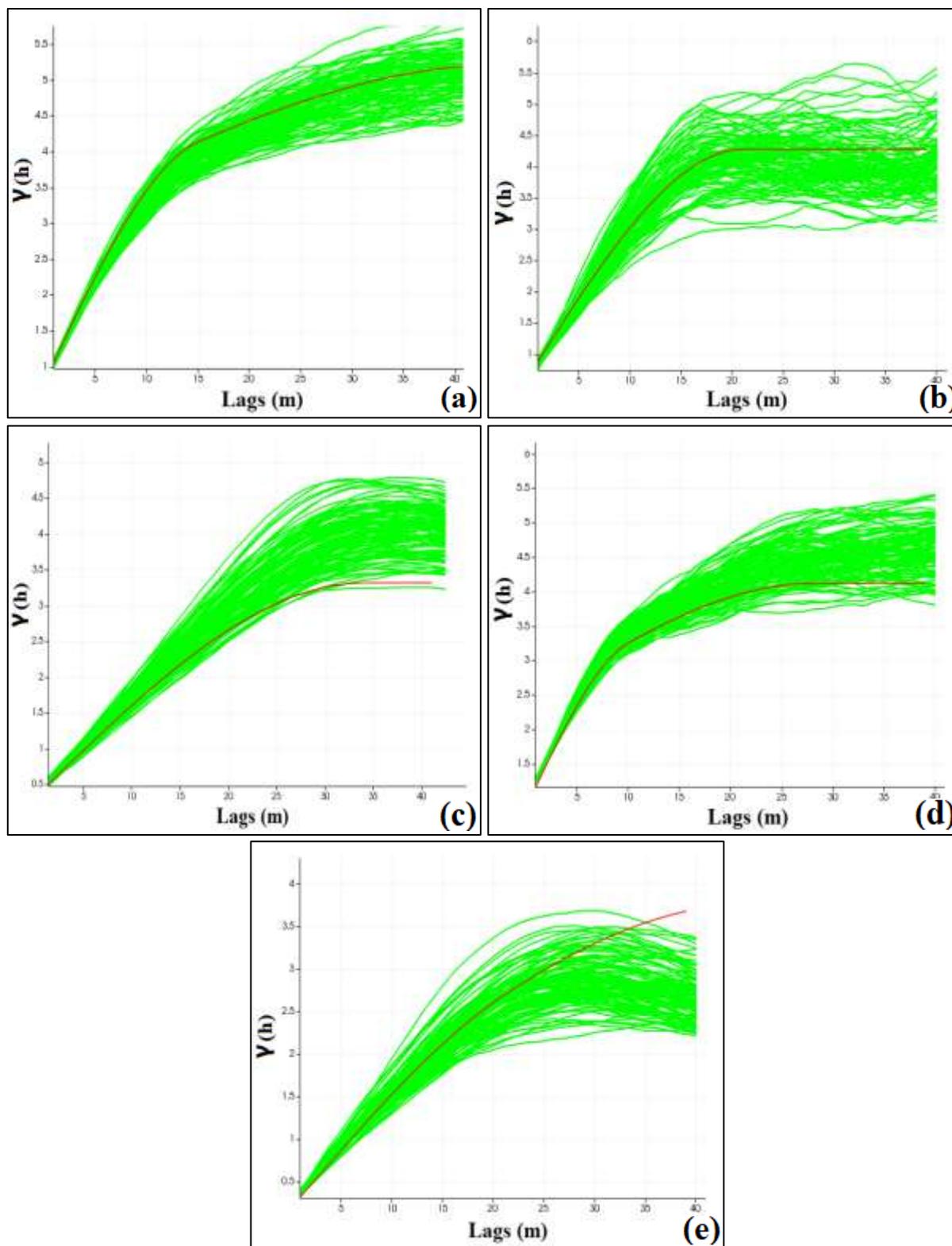


Figura 5.2. Validação dos modelos simulados (linhas verdes) na direção de maior continuidade através de variogramas, usando como referência o modelo dos dados originais (linha vermelha) para o caso IV (a) domínio 1 na $N0^\circ$ (b) domínio 2 na $N157^\circ$ (c) domínio 3 isotrópico (d) domínio 4 isotrópico (e) domínio 5 na $N157^\circ$.

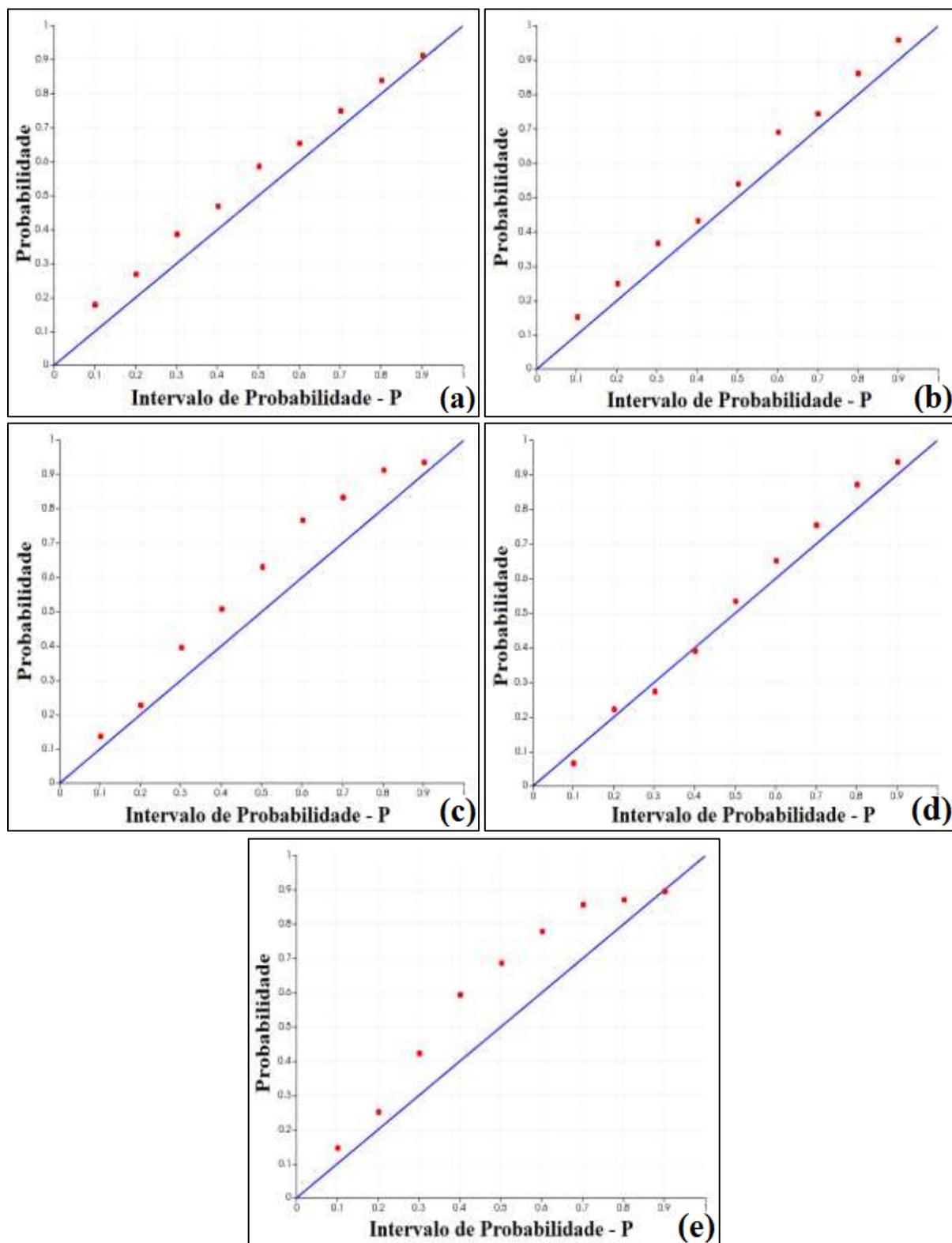


Figura 5.3. Validação dos modelos simulados através dos gráficos de acurácia para o caso IV (a) domínio 1 (b) domínio 2 (c) domínio 3 (d) domínio 4 (e) domínio 5.

5.1.4. ESTATÍSTICA BÁSICA

A Tabela 9 mostra um resumo estatístico dos modelos de LKPO, LKNA e a referência para o caso IV (ver tópico 3.1.2.3). Em primeiro lugar, note que o erro relativo da média, mediana e variância em todos os domínios sempre é menor quando utilizamos o modelo obtido com a LKPO do que o modelo da LKNA. Por outra parte, temos que o erro médio (ME), o erro absoluto médio (MAE) e o erro quadrático médio (MSE) na maioria dos domínios (exceto no domínio 3), o modelo gerado pela LKPO teve melhores resultados (próximos de zero) do que o modelo da LKNA. Por último, temos a medida de correlação ρ , embora ela seja muito parecida em ambos os casos, podemos notar que, com exceção ao domínio 3, o modelo gerado pela LKPO é ligeiramente maior do que o modelo gerado pela LKNA, enquanto no domínio 3 é a LKNA que está ligeiramente melhor.

Tabela 9. Resumo estatístico para o banco de dados 2D.

CASO	Variável	Domínio	Estatística Básica			Erro Relativo (%)			ρ	Distribuição do Erro		
			Média (%)	Mediana (%)	Variância (%) ²	Média	Mediana	Variância		ME (%)	MAE (%)	MSE (%) ²
IV	V_Exhaustive		3.31	3.12	4.03							
	V_LKPO	1	3.28	3.12	2.73	-0.91	0.00	-32.26	0.75	-0.03	1.03	1.8
	V_LKNA		3.13	3	2.12	-5.44	-3.85	-47.39	0.72	0.18	1.08	1.99
	V_Exhaustive		7.45	7.43	4.81							
	V_LKPO	2	7.66	7.49	3.62	2.82	0.81	-24.74	0.86	0.21	0.88	1.31
	V_LKNA		8.12	8.07	2.9	8.99	8.61	-39.71	0.84	-0.67	1.04	1.9
	V_Exhaustive		1.92	1.6	2.63							
	V_LKPO	3	1.88	1.4	2.1	-2.08	-12.50	-20.15	0.77	-0.03	0.79	1.09
	V_LKNA		1.78	1.34	2.05	-7.29	-16.25	-22.05	0.79	0.13	0.72	1.03
	V_Exhaustive		2.75	2.54	3.26							
	V_LKPO	4	3.1	3.1	1.95	12.73	22.05	-40.18	0.76	-0.34	0.99	1.49
	V_LKNA		3.22	3.16	1.78	17.09	24.41	-45.40	0.74	-0.46	1.05	1.67
	V_Exhaustive		1.94	1.49	3.75							
	V_LKPO	5	1.7	1.24	2.97	-12.37	-16.78	-20.80	0.9	-0.24	0.58	0.79
	V_LKNA		1.66	1.22	2.61	-14.43	-18.12	-30.40	0.87	0.28	0.65	1

5.1.5. CURVAS TEOR-TONELAGEM

A Figura 5.4a mostra as curvas teor-tonelagem para todo o depósito mineral (construído a partir da união dos domínios previamente definidos). Observe, que o comportamento das duas curvas obtidas pelo modelo da LKPO (linhas vermelhas) se aproxima mais ao modelo de referência (linhas verdes) do que as curvas geradas pelo modelo da LKNA (linhas azuis). Isso significa que, para a maior parte do gráfico, a metodologia da LKPO levou a modelos mais acurados do que o uso da metodologia da LKNA. Essa acurácia é medida através das variações do erro na curva da tonelagem (figura 5.4b) e a curva do teor médio (figura 5.4c).

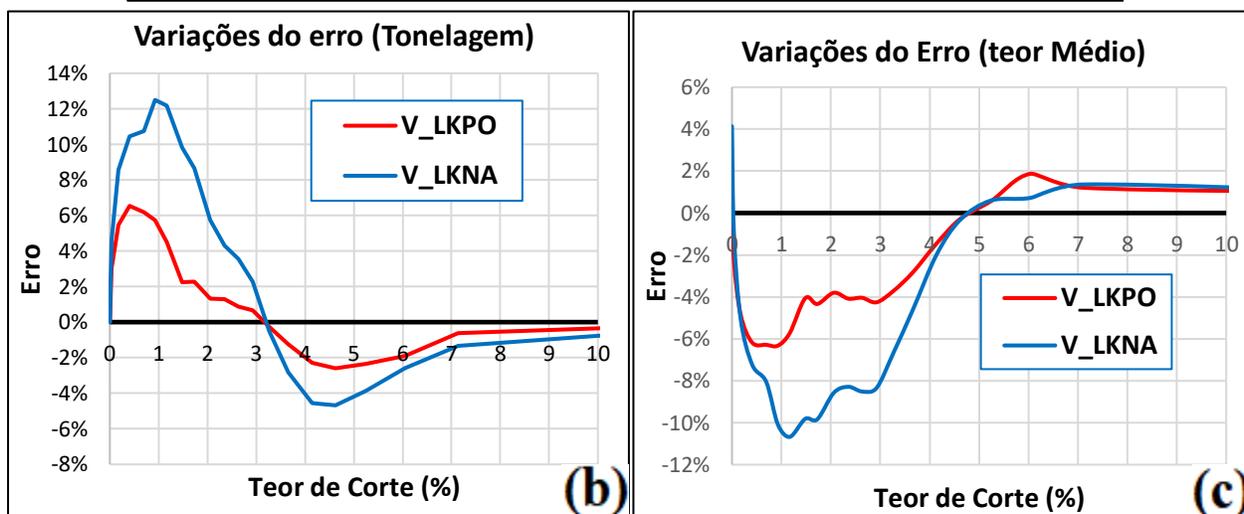
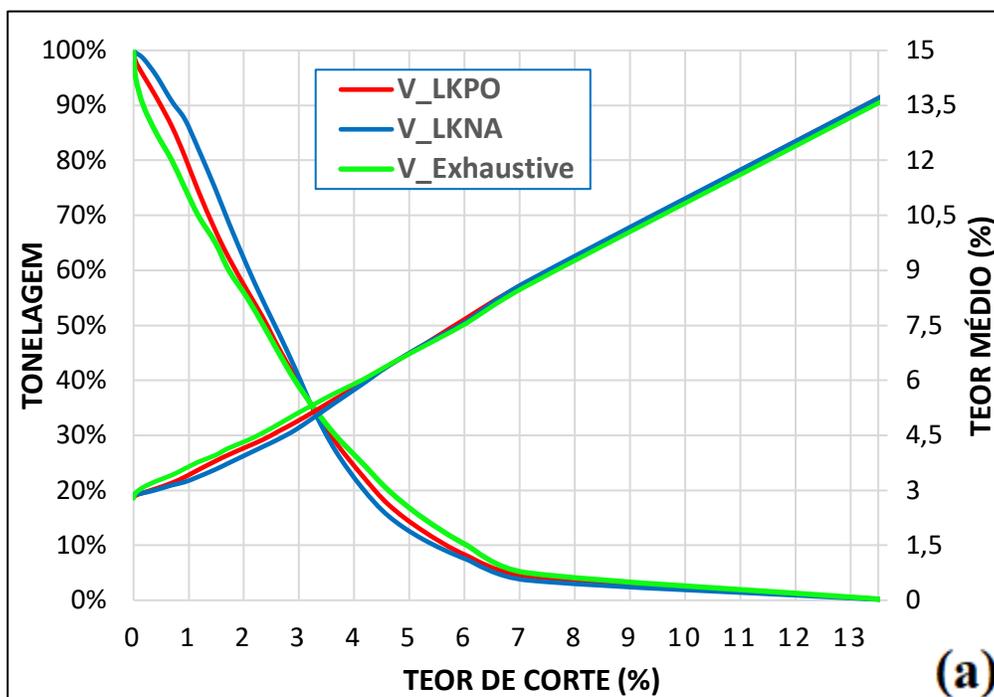


Figura 5.4. Caso IV utilizando simulações (a) Curvas teor-tonelagem (b) Variações do erro na tonelagem (c) Variações do erro no teor médio.

5.1.6. ANÁLISE DAS MÉDIAS LOCAIS

Como mencionado anteriormente, o método das médias locais realiza comparações locais dos valores entre o modelo de blocos interpolado e o modelo de referência ao longo das bandas nas direções principais (X, Y e Z). A média é calculada para cada banda e o resultado é plotado em relação à sua localização ao longo das direções principais. Os gráficos mostram a coerência entre a média local ao longo de cada faixa nas duas estimativas. As Figuras 5.5 a e b mostram os

gráficos de análise de deriva para os modelos LKPO, LKNA e de referência ao longo das duas direções principais, isto é, X e Y, respectivamente. Observe, que na maioria das partes do gráfico, as médias locais do modelo LKPO estão mais próximas dos valores do modelo de referência do que os valores derivados do modelo LKNA.

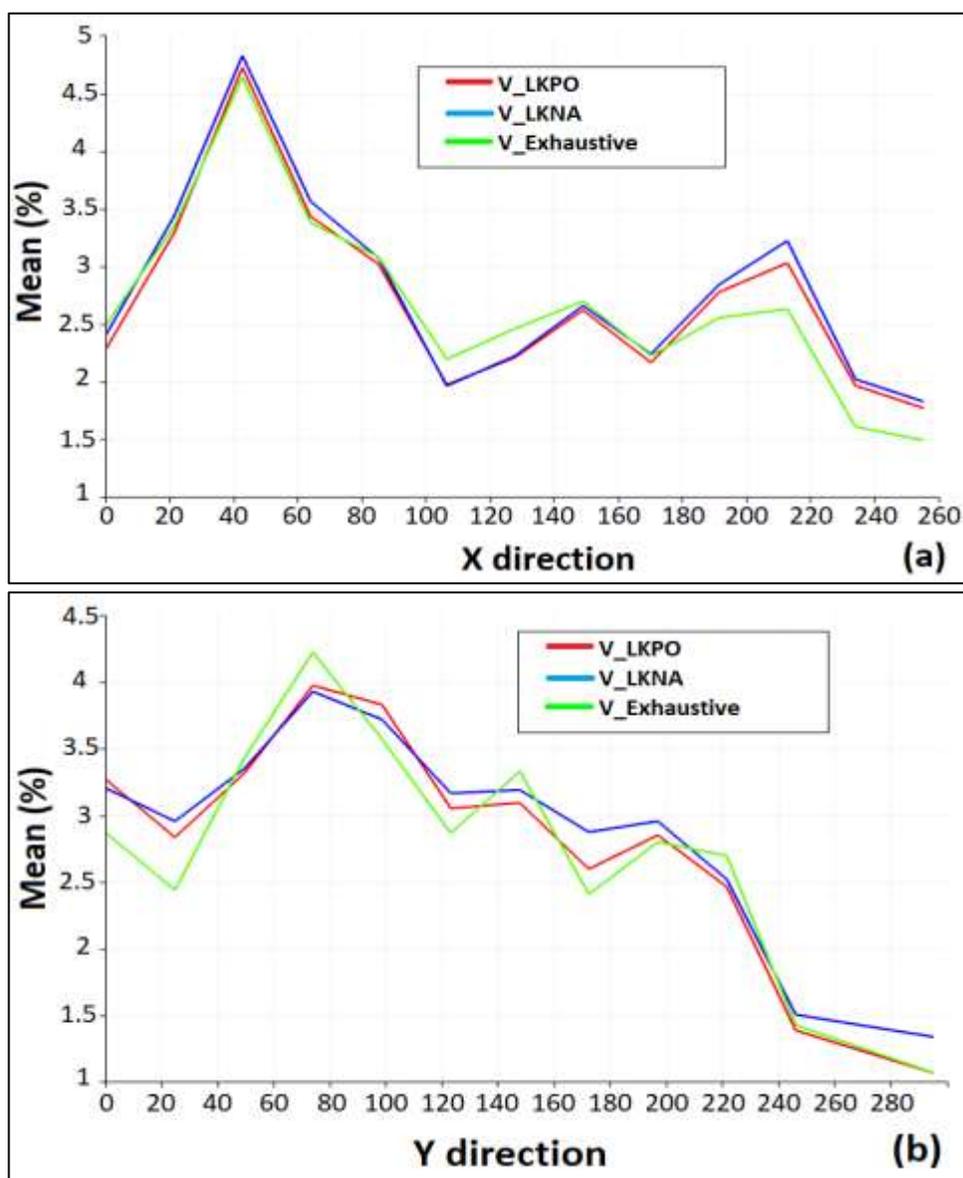


Figura 5.5. Análise de deriva ao longo das principais direções (a) plano X (b) plano Y.

5.1.7. ANÁLISE DE SUAUIZAÇÃO DA KRIGAGEM

Para realizar a análise de suavização da krigagem foi feito, em primeiro lugar, um desagrupamento nos dados através do método de desagrupamento por células (Isaaks e Srivastava, 1989) para obter os pesos de cada amostra e, posteriormente, foi utilizado esses pesos para realizar cada uma das correções de mudança de suporte citadas anteriormente. Semelhante ao comportamento das figuras das curvas teor-tonelagem vistas anteriormente, a figura 5.6 mostra que as duas curvas geradas pelos modelos da LKPO e as referências (afim, IL e DGs) estão mais próximas entre si do que o modelo LKNA. Claramente isso demonstra que o efeito de suavização gerado no modelo da LKNA é maior do que o efeito de suavização gerado no modelo da LKPO.

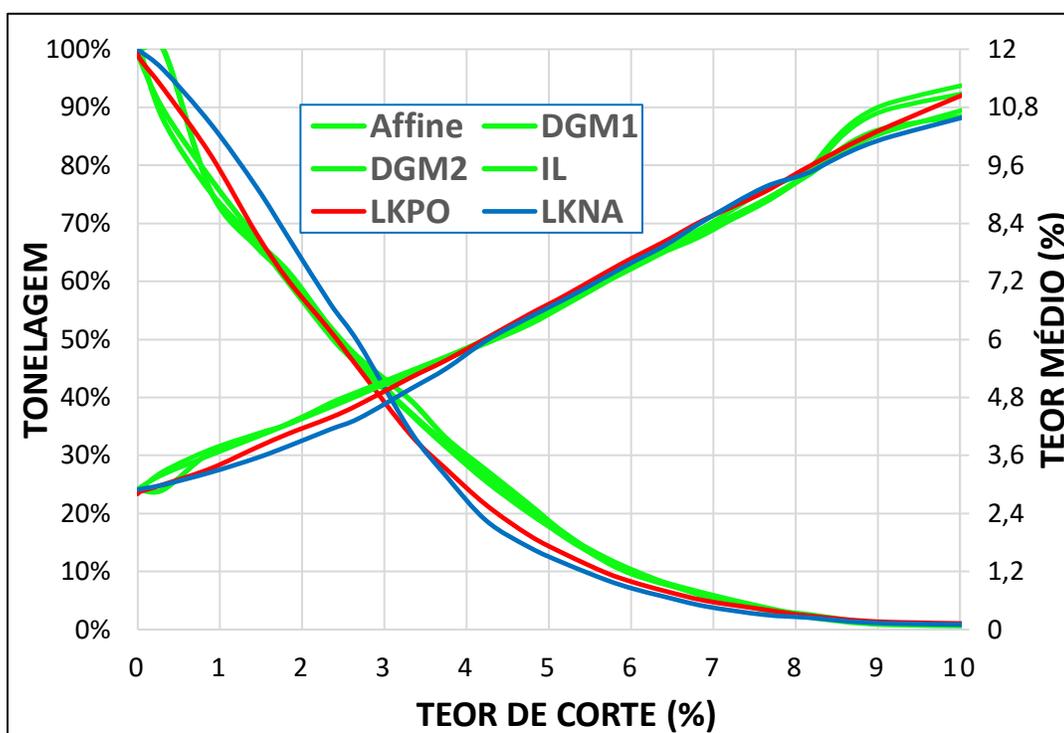


Figura 5.6. Curvas teor-tonelagem. Modelo LKPO utilizando simulações (linha vermelha), Modelo LKNA (linha azul) e Modelos de referência obtida pela mudança de suporte (linhas verdes).

CAPÍTULO 6 – CONSIDERAÇÕES FINAIS

6.1. COMPARAÇÃO ENTRE OS MODELOS LKPO/LKNA

As duas metodologias apresentadas neste trabalho visam otimizar os parâmetros de krigagem através da função de minimização do erro entre o valor estimado e o real. A principal diferença nestas metodologias é a métrica utilizada por cada uma delas para realizar o processo de otimização. Uma, utiliza a métrica de validação cruzada (LKPO *cross-validation*) e, a outra, o espaço de incerteza para realizar todo o processo de otimização ((LKPO *simulation*). Com o intuito de mostrar uma comparação direta entre essas duas metodologias e o modelo de referência, é apresentado alguns dos métodos de validação de estimativas apresentados anteriormente. Em primeiro lugar, a figura 6.1 mostra uma análise visual entre os dois modelos propostos (figuras b-c) e o modelo de referência (figura a). Note que, em ambos os casos, visualmente são muito parecidos entre eles, principalmente nas zonas de transição entre os teores altos e baixos. Além disso, nota-se a ausência de suavização entre ambos os modelos.

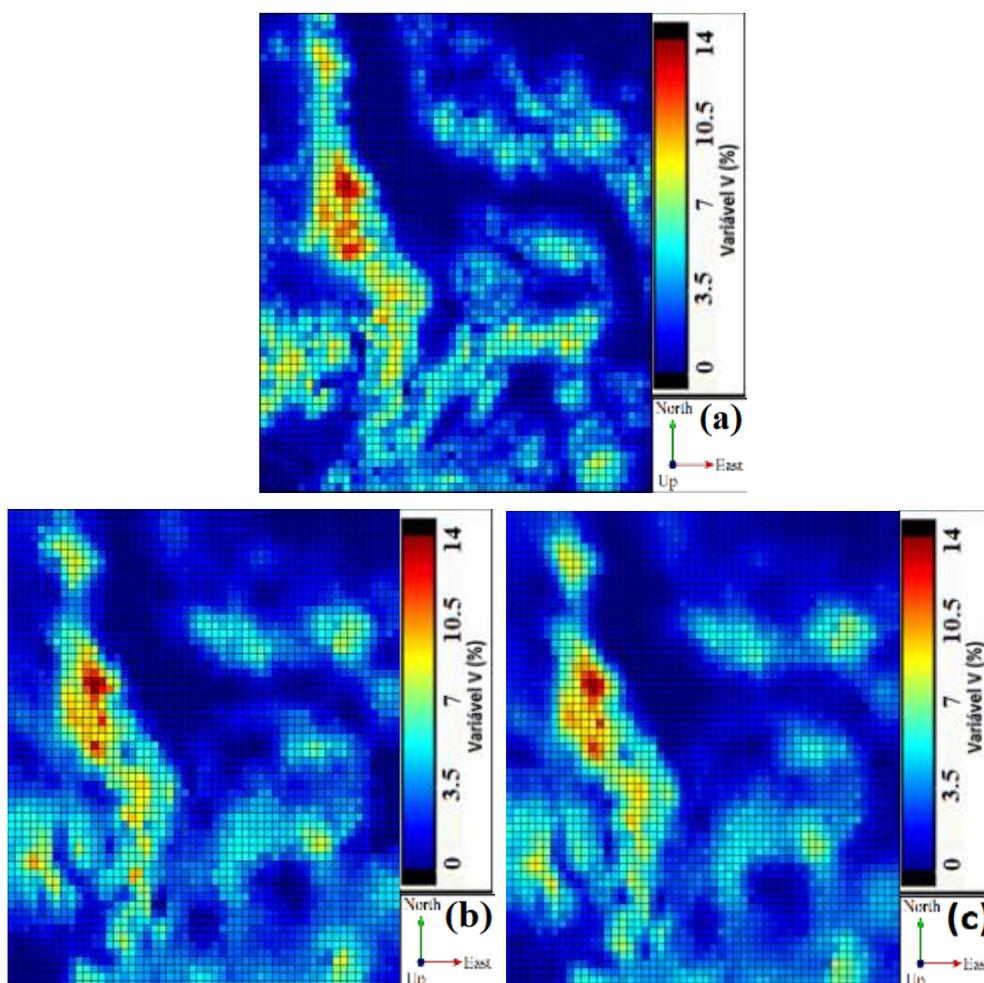


Figura 6.1. Análise visual (a) Dados Exaustivos (b) Modelo LKPO-Cross-Validation (c) Modelo LKPO-Simulations.

Em segundo lugar, foi realizado um diagrama de dispersão (*scatterplot*) que mede, quantitativamente, a relação entre duas variáveis, neste caso, os modelos obtidos através das duas metodologias propostas. Na figura 6.2 podemos observar como o coeficiente de correlação (*correlations*) é muito próximo de 1 (0,98), o que indica uma forte relação positiva entre as duas variáveis.

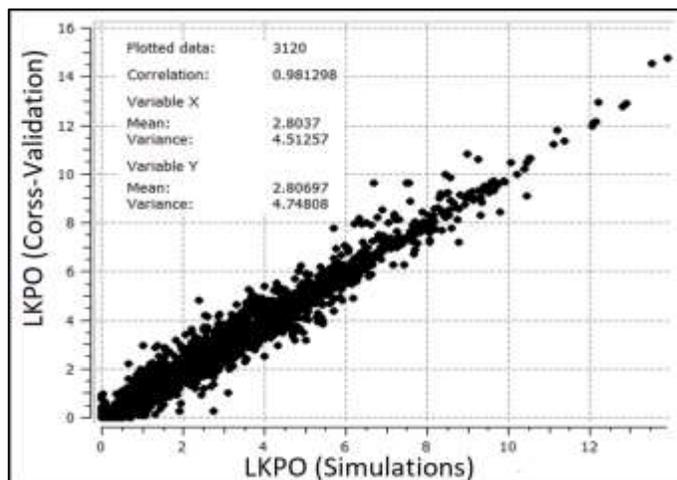


Figura 6.2. Gráfico de dispersão entre o LKPO-Simulations e o LKPO-Cross-Validation.

Em terceiro lugar, foi realizada uma comparação entre as curvas teor-tonelagem para ambas as metodologias, comparando com o modelo exaustivo. Na figura 6.3 podemos observar o modelo obtido por validação cruzada (linha vermelha), o modelo obtido por simulações (linha azul) e o modelo exaustivo (linha verde). Note que, os dois modelos do LKPO apresentam praticamente o mesmo comportamento tanto na tonelagem (exceto para 20% cujos teores são abaixo de 1%) como no teor médio.

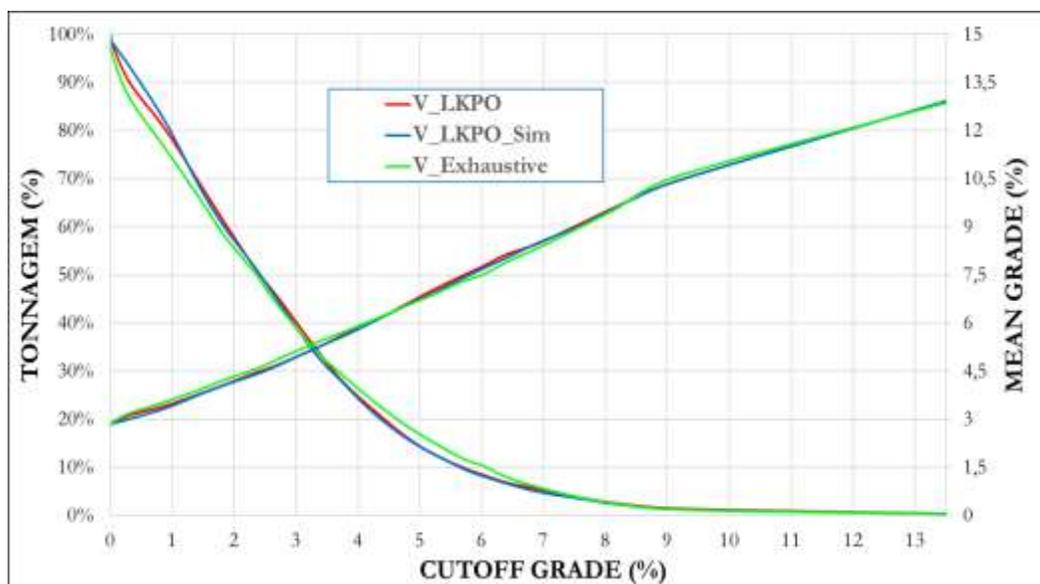


Figura 6.3. Curvas teor-tonelagem entre os modelos LKPO-simulations (linha azul) e LKPO-Cross-Validations (linha vermelha) comparadas com o modelo exaustivo (linha verde).

Finalmente, a figura 6.4 mostra o gráfico de dispersão do erro absoluto obtido entre o modelo da LKPO (*Cross-Validations*) e o modelo da LKPO (*simulations*). Esse erro é obtido através da diferença absoluta entre o modelo dos dados exaustivos e cada um dos modelos estudados. Note que, apesar do mapa de dispersão apresentar uma forte correlação na distribuição dos erros (0.88), existem alguns blocos onde esses erros são completamente diferentes e muito distantes um dos outros, como apresentado na figura.

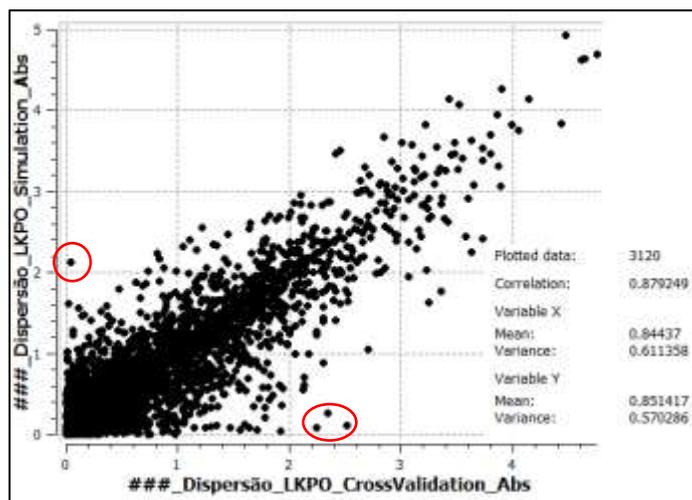


Figura 6.4. Gráfico de dispersão do erro absoluto entre os modelos LKPO (Cross-Validations) e LKPO (Simulations).

6.2. COMPARAÇÃO ENTRE OS MODELOS LKPO E KRIGAGEM ORDINÁRIA TRADICIONAL

Como mencionado anteriormente, atualmente o processo de estimativa costuma ser realizado utilizando uma única estratégia de busca para o domínio que está sendo estudado, ou seja, todos os blocos são interpolados utilizando os mesmos parâmetros de busca (parâmetros globais). Em alguns casos, esta premissa pode ocasionar algum viés condicional nas estimativas, principalmente porque dentro da maioria dos depósitos minerais existem zonas nas quais o mapeamento amostral é mais denso que em outras zonas (e.g., zonas de alto e baixo teor). Entretanto, este trabalho não foi omissivo a essa situação atual e preocupou-se por realizar as comparações entre o método tradicional utilizando parâmetros globais e o método da LKPO utilizando validação cruzada.

Em primeiro lugar foi realizado um gráfico das curvas teor-tonelagem, como pode ser observado na figura 6.5. Note que, o comportamento das curvas geradas no modelo da LKPO (linhas vermelhas) aproxima-se mais das curvas geradas no modelo exaustivo (linhas verdes) do que as curvas geradas no modelo tradicional (linhas azuis), indicando uma maior acurácia na metodologia proposta.

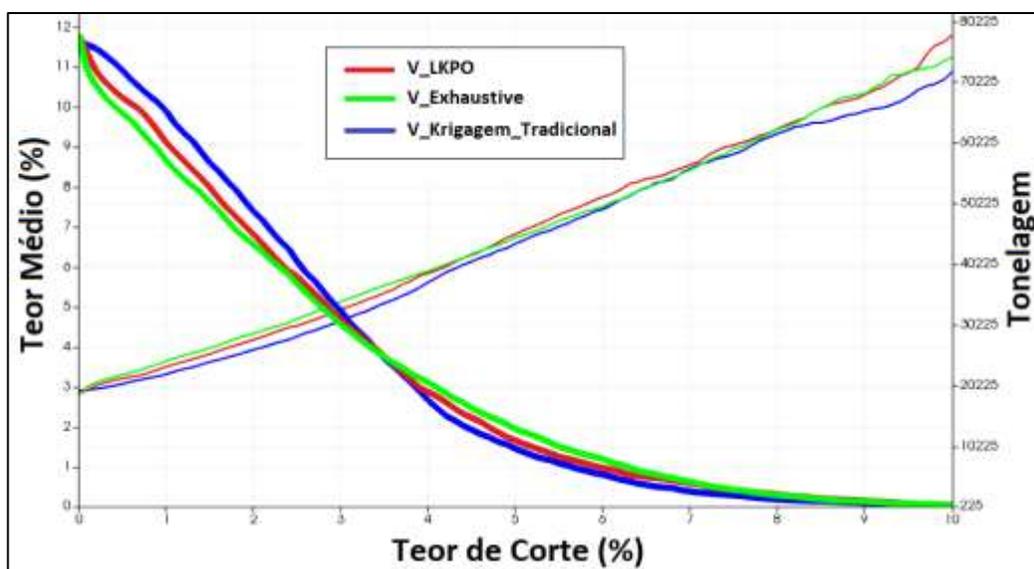


Figura 6.5. Curvas teor-tonelagem entre o modelo tradicional (linha azul) e LKPO-Cross-Validations (linha vermelha) comparadas com o modelo exaustivo (linha verde).

Outro ponto importante a ser ressaltado é a similitude dos resultados obtidos pelo método tradicional e o método da LKNA. Para isso foi realizado um diagrama de dispersão (figura 6.6) no qual podemos observar uma correlação quase perfeita (0,979). Isto poderia ser justificado pelo fato de o modelo tradicional utilizar, na maioria dos casos, uma grande quantidade de amostras na estratégia de busca e, como sabemos, uma das métricas utilizadas na LKNA é o *slope* da regressão (que depende da variância de krigagem). Porém, ao depender dessa variância, implica que ela seja minimizada e, para que isso aconteça, muitas amostras devem ser consideradas no processo de interpolação, pois como sabemos, a variância de krigagem depende do layout amostral e não do teor da amostra, ou seja, entre mais amostras utilizadas, menos variância de krigagem, consequentemente as duas metodologias acabam utilizando a maioria das amostras dentro de cada estratégia de busca. Outro fato consequente disso é a aparição de algumas estimativas negativas, produto da geração de pesos negativos que podem ser gerados pelo “efeito eclipse” ao utilizar uma grande quantidade de amostras.

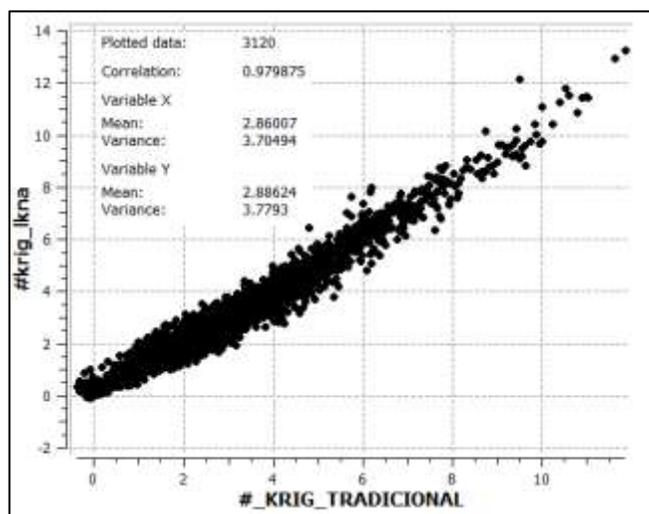


Figura 6.6. Gráfico de dispersão entre o modelo tradicional e o modelo da LKNA.

Finalmente, as figuras 6.7 (a, b) mostram os gráficos de dispersão do erro absoluto obtido entre os modelos da LKPO/Tradicional (figura 6.7a) e os modelos da LKPO/LKNA (figura 6.7b). Note que, apesar do mapa de dispersão apresentar uma forte correlação na distribuição dos erros em ambas as comparações (0.7 e 0.77), existem alguns blocos onde esses erros são completamente diferentes e muito distantes um dos outros, como apresentado na figura. Vale ressaltar que os modelos da LKPO/LKNA possuem uma ligeira melhora na distribuição do erro do que os modelos da LKPO/Tradicional, visto que este último utiliza uma estratégia única para todo o modelo de blocos.

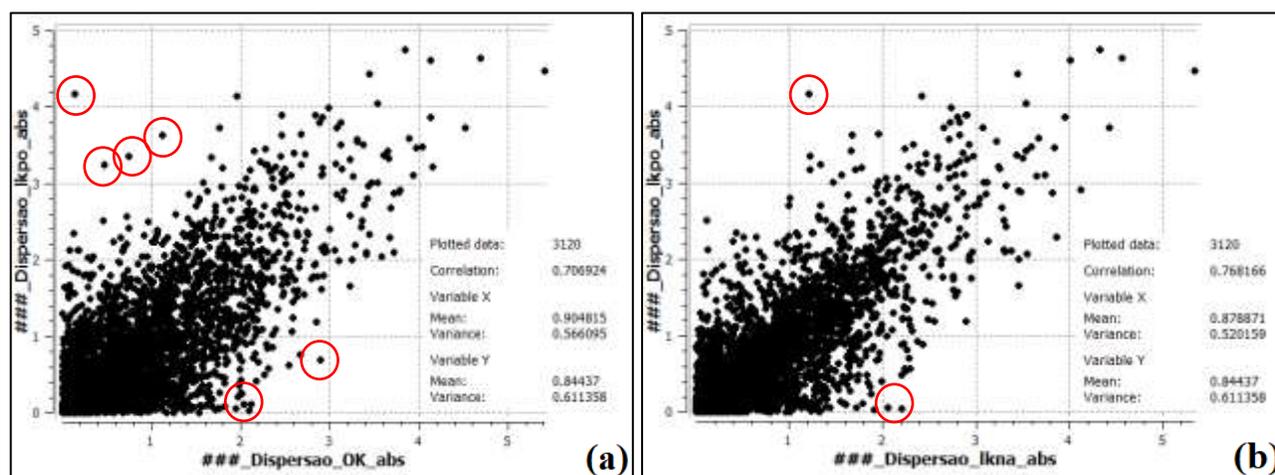


Figura 6.7. Gráfico de dispersão do erro absoluto entre os modelos (a) LKPO/Tradicional (b) LKPO/LKNA.

CAPÍTULO 7 - CONCLUSÕES

Nesse capítulo, serão apresentadas as conclusões finais, produto da implementação da metodologia proposta nesta tese. Uma breve síntese dos resultados obtidos nos estudos de caso 2D e 3D é discutida nessa seção. Ademais, são apresentadas algumas recomendações para trabalhos futuros como objeto de pesquisa nos avanços da geoestatística, cujo objetivo seja criar metodologias que ajudem a melhorar o realismo dos modelos gerados. Finalmente, serão prescritas algumas recomendações finais sobre a implementação desta metodologia.

7.1. CONCLUSÕES

Em muitas ocasiões, a decisão de utilizar determinados parâmetros de estimativa no processo de interpolação da krigagem é meramente subjetiva. Essa decisão baseia-se geralmente em alguns testes simples realizados antes do processo de interpolação (e.g., validação cruzada) ou após o processo de interpolação (e.g., média global, análise de deriva, soma dos pesos negativos da krigagem, dentre outros). Além disso, o processo de estimativa costuma ser realizado utilizando-se parâmetros globais (i.e., uma única estratégia de busca para todo o modelo de blocos). Por outra parte, o número de tentativas realizadas pelo usuário é relativamente pequeno em comparação com o grande número de parâmetros que podem afetar as estimativas (e.g., número mínimo e máximo de amostras, número de setores angulares, número de amostras por furo, dentre outros). Isso demonstra que há necessidade de se criar e implementar novas metodologias, considerando todos esses fatores, o que foi o objeto principal desta pesquisa.

Os estudos de caso 2D e 3D mostraram que o uso de parâmetros locais de krigagem melhoram significativamente a estimativa dos teores. Os benefícios foram evidentes ao usar o modelo da LKPO, reafirmando uma vez mais a utilidade de serem realizadas estimativas locais, nas quais cada bloco é interpolado com seu próprio parâmetro de estimativa. Foram obtidos modelos mais precisos e acurados ao se utilizar a metodologia da LKPO em comparação com as metodologias atualmente disponíveis na literatura (LKNA). Esses resultados foram medidos através de uma ampla variedade de métodos de validações (estatística básica, análise de deriva, suavização da krigagem, curvas teor-tonelagem, dentre outros).

É importante mencionar que um dos métodos de validação mais utilizados nesta tese foi a curva teor-tonelagem, que relaciona a quantidade de massa (tonelagem) e o teor médio para um determinado teor de corte. Em razão de essas curvas serem uma das ferramentas mais amplamente utilizadas na tomada de decisões, foi claramente identificado que utilizar estimativas locais através da LKPO melhora significativamente os comportamentos dessas duas curvas e, como consequência, os modelos se aderem melhor à realidade dos depósitos minerais.

Uma alternativa para agrupar dados nos quais não há nenhuma informação geológica disponível (e.g., zonas de oxidação, litologias, alterações, limites estruturais, distribuição espacial de teores, etc.) é utilizar as técnicas de agrupamento de análise de dados. Por meio do algoritmo *k-means++*, foi possível separar o conjunto de dados (2D) em vários domínios estacionários que permitiram representar as variações anisotrópicas locais dentro de cada domínio.

No processo de otimização, o número de combinações nas estratégias de busca e o fato de as estimativas serem locais (bloco a bloco) demandam grandes operações matemáticas, que, por sua vez, requerem um grande esforço computacional. O desempenho nesse processo evidenciou que o aumento do número de amostras e a capacidade computacional têm um grande impacto nos tempos de execução do algoritmo da LKPO. Essa análise foi realizada utilizando-se dois computadores com características computacionais diferentes.

7.2. TRABALHOS FUTUROS

Durante o desenvolvimento dessa metodologia, surgiram novas ideias que poderiam ser analisadas e implementadas em trabalhos futuros, entre elas estão:

- i. Paralelizar o algoritmo da LKPO com o objetivo de utilizar vários processadores ao mesmo tempo, o que reduziria consideravelmente o tempo de execução de cada processo, principalmente na criação da área de influência de cada amostra e das estimativas de cada uma delas.
- ii. Criar um algoritmo de validação cruzada dentro da plataforma do SGeMS, pois atualmente a LKPO faz uso do algoritmo de validação cruzada implementado dentro da plataforma do GSLib (*Geostatistics Software Library*).
- iii. Criar um sistema de desempate quando duas distribuições do erro quadrático (na LKPO usando simulações) possuam o mesmo teor médio otimizado, por exemplo, largura da banda da distribuição ou uma medida de distância entre a média e o desvio padrão.
- iv. Combinar anisotropias locais (e.g., LVA) com a LKPO.
- v. Adicionar no algoritmo a possibilidade de escolher o número de amostras que deseja retirar na validação cruzada entorno da amostra que foi extraída para a respectiva área de influência, visando não ter otimizações erradas por causa da influência de valores extremos.

7.3. RECOMENDAÇÕES FINAIS

É importante mencionar que o uso dessa metodologia requer que o usuário tenha um breve conhecimento em análises de estimativas de recursos minerais ou áreas afins. Resultados, esperados ou não, dependem do *input* de parâmetros e não da metodologia proposta. O desempenho da LKPO mostrou trabalhar adequadamente em dois conjuntos de dados diferentes, entretanto, não há evidências matemáticas (somente experimentais) de que funciona em qualquer

tipo de conjunto de dados. Portanto, recomenda-se ao leitor que esta metodologia seja testada em outros bancos de dados com amplo conhecimento das características intrínsecas do tipo de depósito objeto de estudo.

A maioria dos *softwares* geoestatísticos como o *SGeMS* e o *Isatis@*, por exemplo, proporcionam ao usuário a escolha no arquivo de saída várias informações que podem ser úteis além das estimativas, dentre elas estão: a variância de krigagem, soma dos pesos positivos, número de amostras utilizadas em cada bloco interpolado, distância média entre as amostras, a eficiência de krigagem (KE, em inglês), o *slope* da regressão (SR, em inglês), etc. Portanto, a implementação da LKNA ao longo deste trabalho foi realizada através da criação de um “*script*” desenvolvido pelo próprio autor, considerando principalmente os arquivos de saída do KE e o LR que foram gerados no processo de interpolação para cada um dos parâmetros utilizados. A ideia por trás desse “*script*” é escolher, para cada bloco, os parâmetros que proporcionem os maiores KE e SR ao longo de todo o processo de otimização.

REFERÊNCIAS BIBLIOGRÁFICAS

Abedini, M. J., Nasser, M., & Ansari, A. (2008). Cluster-Based Ordinary Kriging of Piezometric Head in West Texas/New Mexico — Testing of Hypothesis. *Journal of Hydrology*, 351(3–4), pp. 360–367, doi:10.1016/j.jhydrol.2007.12.030.

Abedini, M. J., Nasser, M., & Burn, D. H. (2012). The Use of a Genetic Algorithm-Based Search Strategy in Geostatistics: Application to a Set of Anisotropic Piezometric Head Data. *Computers & Geosciences*, 41, pp. 136–146, doi:10.1016/j.cageo.2011.08.024.

Abzalov, M. (2016). *Applied Mining Geology*. Cham: Springer. 448 p., doi:10.1007/978-3-319-39264-6.

Abzalov, M., & Humphreys, M. (2002). Geostatistically Assisted Domaining of Structurally Complex Mineralisation: Method and Case Studies. In *AusIMM Conference 2002 Victoria: AusIMM*, pp. 345–350.

Aggarwal, C. C. K., & Reddy, C. (2013). *Data Clustering: Algorithms and Applications*. Chapman & Hall/CRC, 616 p.

Arthur, D., & Vassilvitskii, S. (2007). K-Means++: The Advantages of Careful Seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete, Philadelphia: SIAM, Algorithms*, pp. 1027–1035.

Boisvert, J. B. (2010). *Geostatistics with Locally Varying Anisotropy*. PhD Thesis, University of Alberta, Edmonton, AB, 175 p., doi.org/10.7939/R31X5C.

Boisvert, J. B., Deutsch, C. V. (2011). Programs for Kriging and Sequential Gaussian Simulation with Locally Varying Anisotropy Using Non-Euclidean Distances. *Computers & Geosciences*, 37 (4), pp. 495–510.

Boisvert, J. B., Manchuk, J. G., & Deutsch, C. V. (2007). Calculating Distance in Presence of Locally Varying Anisotropy. *IAMG Geomathematics and GIS Analysis of Resources, Environment and Hazards*, pp. 82–85.

Boisvert, J. B., Manchuk, J. G., & Deutsch, C. V. (2009). Kriging in the Presence of Locally Varying Anisotropy Using Non-Euclidean Distances, *Mathematical Geosciences*, 41(5), pp. 585–601, doi:10.1007/s11004-009-9229-1.

Boyle, C. (2010). Kriging Neighbourhood Analysis by Slope of Regression and Weight of Mean — Evaluation with the Jura Data Set. *Mining Technology*, 119(2), pp. 49–58. doi:10.1179/037178410X12741755140804.

Calinski, T., & Harabasz, J. (1974). A Dendrite Method for Cluster Analysis. *Communications in Statistics - Theory and Methods*, 3(1), pp. 1–27. doi:10.1080/03610927408827101.

Caxeta, R. M., Costa, J. F., Machado, P. L., Ribeiro, D. T., & Pires, R. O. (2015). How Precise and Accurate is the Space of Uncertainty Derived from Simulations? In: *Proceedings of APCOM 2015*, Fairbanks, Alaska.

Chanderman, L., Dohm, C. E., & Minnitt, R. C. A. (2017). 3D Geological Modelling and Resource Estimation for a Gold Deposit in MALI. *Journal of the Southern African Institute of Mining and Metallurgy*, 117(2), pp. 189–197. doi:10.17159/2411-9717/2017/v117n2a10.

Chentsov, N. N. (1957). Levy Brownian Motion for Several Parameters and Generalized White Noise. *Theory of Probability and its Applications*, 2, pp. 265-266.

Chilès, J.P. & Delfiner, P. (1999). *Geostatistics: Modeling Spatial Uncertainty*. John Wiley & Sons, New York, 695 p.

Coombes, J. (2008). *The Art and Science of Resource Estimation. A practical Guide for Geologists and Engineers*. Subiaco, WA, Australia: Coombes Capability, 231 p.

Coombes, J., & Boamah, P. (2015). Local Kriging Neighbourhood Optimisation. In *Africa Australia Technical Mining Conference 2015*, Victoria: AusIMM, pp. 27–35.

Deutsch, C. V. (1992). *Annealing Techniques Applied to Reservoir Modeling and the Integration of the Geological and the Engineering (Well Test) Data*. PhD Thesis, Stanford University, USA, 306 p.

Deutsch, C. V. (1996). Correcting for Negative Weights in Ordinary Kriging. *Computers & Geosciences*, 22, pp. 765-773.

Deutsch, C. V. (2007). *The Slope of Regression for Kriging Estimators*. Report 09. Centre for Computational Geostatistics (CCG). University of Alberta, Canada, pp 314-4.

Deutsch, C. V., & Journel, A. G. (1998). *GSLIB: Geostatistical Software Library and User's Guide* (2nd ed.). New York: Oxford University Press, 369 p.

Deutsch, C. V. (2002). *Geostatistical Reservoir Modeling*. New York: Oxford University Press. 376 p.

Deutsch, C. V. (1996). Direct Assessment of Local Accuracy and Precision. In: *Geostatistics Wollongong '96*, Vol.- 1, Baafi and Schofield (eds.). Kluwer Academic Publishers, pp. 115–125.

Deutsch, J. L., Szymanski, J. & Deutsch, C. V. (2014). Checks and Measures of Performance for Kriging Estimates. *The Southern African Institute of Mining and Metallurgy*, Johannesburg, South Africa, 114, pp. 223-230.

Duke, J. H., & Hanna, P. J. (2014). Geological Interpretation for Resource Modelling and Estimation. In A. C. Edwards (Ed.), *Mineral Resource and Ore Reserve Estimation — The AusIMM Guide to Good Practice* (2nd ed.), Victoria: AusIMM, pp. 145–153.

Emery, X. (2004). On the Consistency of the Indirect Lognormal Correction. *Stochastic Environmental Research and Risk Assessment*, 18(4), pp. 258–264, doi:10.1007/s00477-004-0188-2.

Emery, X. (2008). Change of Support for Estimating Local Block Grade Distributions. *Mathematical Geosciences*, 40(6), pp. 671–688, doi:10.1007/s11004-008-9148-6.

Emery, X., & Ortiz, J. M. (2005). Estimation of Mineral Resources Using Grade Domains: Critical Analysis and a Suggested Methodology. *Journal of the South African Institute of Mining and Metallurgy*, 105, pp. 247–256.

Fouedjio, F. (2016a). A Hierarchical Clustering Method for Multivariate Geostatistical Data. *Spatial Statistics*, 18, pp. 333–351, doi:10.1016/j.spasta.2016.07.003.

Fouedjio, F. (2016b). Discovering Spatially Contiguous Clusters in Multivariate Geostatistical Data Through Spectral Clustering. In J. Li, X. Li, S. Wang, J. Li, & Q. Sheng (Eds.), *ADMA 2016: Advanced Data Mining and Applications. Lecture Notes in Computer Science 10086*, pp. 547–557. Cham: Springer. doi:10.1007/978-3-319-49586-6_38.

Fouedjio, F. (2016c). A Clustering Approach for Discovering Intrinsic Clusters in Multivariate Geostatistical Data. In P. Perner (Ed.), *Machine Learning and Data Mining in Pattern Recognition. Lecture Notes in Computer Science 9729*, pp. 491–500. Cham: Springer. doi:10.1007/978-3-319-41920-6_39.

Fouedjio, F. (2017a). A Spectral Clustering Approach for Multivariate Geostatistical Data. *International Journal of Data Science and Analytics*, 4(4), pp. 301–312, doi:10.1007/s41060-017-0069-7.

Fouedjio, F. (2017b). A Spectral Clustering Method for Large-Scale Geostatistical Datasets. In P. Perner (Ed.), *MLDM 2017: Machine Learning and Data Mining in Pattern Recognition. Lecture Notes in Computer Science 10358*, pp. 248–261. Cham: Springer, doi:10.1007/978-3-319-62416-7_18.

Fouedjio, F., Hill, E. J., & Laukamp, C. (2018). Geostatistical Clustering as an Aid for Ore Body Domaining: Case Study at the Rocklea Dome Channel Iron Ore Deposit, Western Australia. *Applied Earth Science*, 127(1), pp. 15–29, doi:10.1080/03717453.2017.1415114.

Glacken, I., & Trueman, A. (2014). Overview Mineral Resource Estimation. In A. C. Edwards (Ed.), *Mineral Resource and Ore Reserve Estimation — The AusIMM Guide to Good Practice* (2nd ed.), Victoria: AusIMM, pp. 263–276.

Haas, T. C. (1990). Kriging and Automated Variogram Modeling Within a Moving Window. *Atmospheric Environment. Part A. General Topics*, 24(7), pp. 1759–1769. doi:10.1016/0960-1686(90)90508-K.

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Waltham, MA: Morgan Kaufmann, 703 p.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning* (2nd ed.). New York: Springer-Verlag. doi:10.1007/978-0-387-98135-27, 745 p.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied Statistics for the Behavioral Sciences*. 5th ed. Boston: Houghton Mifflin, 756 p.

Isaaks, H. E. (2004). Local Anisotropy Kriging. <http://www.isaaks.com/software-lak.html>. Accessed 25 August 2017.

Isaaks, H. E. (2005). The Kriging Oxymoron: a Conditionally Unbiased and Accurate Predictor (2nd ed.). In: *Proceedings of geostatistics Banff 2004*. Springer, 2005, 1, pp. 363–374.

Isaaks, H. E., & Srivastava, M. R. (1989). *An Introduction to Applied Geostatistics*. Oxford: Oxford University Press. 547 p.

Jeannée, N., Bertoli, O. & Deraisme, J. (2013). Local Geostatistics, Concepts & Case Studies. *Geostats Rendezvous – Perth*, pp. 26-27.

Journel, A. G. & Huijbregts, C. J., (1978). *Mining Geostatistics*, Academic Press: London, 600 p.

Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., & Wu, A. Y. (2002). An Efficient K-Means Clustering Algorithm: Analysis and Implementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7), pp. 881–892. doi:10.1109/TPAMI.2002.1017616.

Kodinariya, T. M., & Makwana, P. R. (2013). Review on Determining Number of Cluster in K-Means Clustering. *International Journal of Advance Research in Computer Science and Management Studies*, 1(6), pp. 2321–7782.

Krige, D. G., (1994). An Analysis of Some Essential Basic Tenets of Geostatistics not Always Practised in Ore Valuations, in *Proceedings Regional APCOM: Computer Applications and Operations Research in the Minerals Industries, Slovenia*, pp. 15-18.

Krige, D. G., (1996a). A Basic Perspective on the Roles of Classical Statistics, Data Search Routines, Conditional Biases and Information and Smoothing Effects in Ore Block Valuations, in *Proceedings Conference on Mining Geostatistics, Kruger National Park, South Africa, Geostatistical Association of South Africa*, pp. 1-10

Krige, D. G., (1996b). A Practical Analysis of the Effects of Spatial Structure and Data Available and Accessed, on conditional biases in ordinary kriging, in Proceedings Fifth International Geostatistical Congress, Geostatistics Wollongong '96 (Eds: E Y Baafi and N A Schofield) pp. 799-810.

Koppe, V. C. (2009). Metodologia para Comparar a Eficiência de Alternativas para disposição de Amostras. Tese (Doutorado em Engenharia) – Faculdade de Engenharia de Minas, Metalúrgica e de Materiais, Universidade Federal Rio Grande do Sul, Porto Alegre, 236 p.

Lajaunie, C. (2000). Introduction to the Modelling of the Support Effect. In H. S. Pandalai & P. K. Saraswati (Eds.), *Geological Data Analysis: Statistical Methods*, New Delhi: Hindustan Publishing, pp. 155–171.

Machuca-Mory, D. F., & Deutsch, C. V. (2013). Non-Stationary Geostatistical Modeling Based on Distance Weighted Statistics and Distributions. *Mathematical Geosciences*, 45(1), pp. 31–48. doi:10.1007/s11004-012-9428-z.

Magneron, C., Jeanne, N., Le Moine, O., & Bourillet, J.-F. (2010). Integrating Prior Knowledge and Locally Varying Parameters with Moving-GeoStatistics: Methodology and Application to Bathymetric Mapping. In P. Atkinson & C. Lloyd (Eds.), *geoENV VII — Geostatistics for Environmental Applications. Quantitative Geology and Geostatistics*, Dordrecht: Springer, 16, pp. 405–415, doi:10.1007/978-90-481-2322-3_35.

Marques, D. M. (2010). Desenvolvimento de um Algoritmo para Simular a Variabilidade do Minério em Pilhas de Homogeneização. Dissertação (Mestrado em Engenharia) – Faculdade de Engenharia de Minas, Metalúrgica e de Materiais, Universidade Federal Rio Grande do Sul, Porto Alegre, 130p.

Matheron, G. (1962–1963). Tome I: *Traité de Géostatistique Appliquée*. Tome II: *Le krigeage*. I: *Mémoires du Bureau de Recherches Géologiques et Minières*, No. 14 (1962), Paris: Editions Technip. II: *Mémoires du Bureau de Recherches Géologiques et Minières*, No. 24 (1963), Paris: Editions B.R.G.M.

Matheron, G. (1963). Principles of Geoestistics. *Economic Geology*, No. 58, pp. 1246-1266.

Matheron, G. (1973). The Intrinsic Random Functions and Their Applications. *Advances in Applied Probability*, vol. 5, No. 3, pp. 439-468.

Matheron, G. (1976). Forecasting Block Grade Distributions: the Transfer Functions. In M. Guarascio, M. David, & C. J. Huijbregts (Eds.), *Advanced Geostatistics in the Mining Industry*, Dordrecht: Reidel, pp. 237–251.

Matheron, G. (1984). The Selectivity of the Distributions and “The Second Principle of Geostatistics.” In G. Verly, M. David, A. G. Journel, & A. Marechal (Eds.), *Geostatistics for*

Natural Resources Characterization, Dordrecht: Reidel, pp. 421–433, doi:10.1007/978-94-009-3699-7_24.

Matheron, G. (1985). Change of Support for Diffusion-Type Random Functions. *Journal of the International Association for Mathematical Geology*, 17(2), pp. 137–165. doi:10.1007/BF01033151.

McArthur, G. J. (1988). Using Geology to Control Geostatistics in the Hellyer Deposit. *Mathematical Geology*, v. 20, n. 4, pp. 343-366.

McLennan, J.A. & Deutsch, C. V. (2002). "Conditional Bias of Geostatistical Simulation for Estimation of Recoverable Reserves," CIM Vancouver 2002, April 28 to May 1.

Myers, D. E. (1989). To Be or Not to Be... Stationary? That is the Question. *Mathematical Geology*, 21(3), pp. 347–362, doi:10.1007/BF00893695.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, pp. 2825–2830.

Pelleg, D., & Moore, A. (1999). Accelerating Exact K-Means Algorithms with Geometric Reasoning. In *KDD '99: Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York: ACM, pp. 277–281, doi:10.1145/312129.312248

Philip, G. M., & Watson, D. F. (1986). Matheronian Geostatistics — Quo Vadis? *Mathematical Geology*, 18(1), pp. 93–117.

Pyrcz, M.J., and Deutsch, C.V. (2014). *Geostatistical Reservoir Modeling*. Oxford University Press, 2nd ed. 448 p.

Ribeiro, D.T., Cunha-Filho, E.M., Costa, J.F.C.L., Roldão, D.G., and Tamantini, T.B.F. 2012. Multidirectional Random Walk: An Alternative to Local Simulation of Gaussian Data. In: *Digital Proceedings of Ninth International Geostatistics Congress*, Oslo, Norway. v. 1. pp. 1–19.

Rivoirard, J. (1987). Two Key Parameters When Choosing the Kriging Neighbourhood, *J Mathematical Geology*, 19, pp. 851-856.

Romary, T., Ors, F., Rivoirard, J., & Deraisme, J. (2015). Unsupervised Classification of Multivariate Geostatistical Data: Two Algorithms. *Computers & Geosciences*, 85, pp. 96–103. doi:10.1016/j.cageo.2015.05.019.

Romary, T., Rivoirard, J., Deraisme, J., Quinones, C., & Freulon, X. (2012). Domaining by Clustering Multivariate Geostatistical Data. In P. Abrahamsen, R. Hauge, & O. Kolbjørnsen (Eds.), *Geostatistics Oslo 2012. Quantitative Geology and Geostatistics 17*, Dordrecht: Springer Netherlands, pp. 455–466, doi:10.1007/978-94-007-4153-9_37.

Rossi, M. E., & Deutsch, C. V. (2014). *Mineral Resource Estimation*. Dordrecht: Springer Netherlands. doi:10.1007/978-1-4020-5717-5, 332 p.

Sinclair, A. J. & Blackwell, G. H. (2002). *Applied Mineral Inventory Estimation*. Cambridge University Press, Cambridge - United Kingdom, 401p.

te Stroet, C. B. M., & Snepvangers, J. J. J. C. (2005). Mapping Curvilinear Structures with Local Anisotropy Kriging. *Mathematical Geology*, 37(6), pp. 635–649. doi:10.1007/s11004-005-7310-y.

Van Munster, R. J., Van Antwerpen, G., Meulblok, M., VanWestenbrugge, C. J., & Borst, J. P. (1995). DIGIPOL, Intelligent Interpolation of Depth Measurements. In *Joint European Conference and Exhibition on Geographical Information*, Basel: JEC-GI, 1, pp. 24–29.

Vann, J., Jackson, S., & Bertoli, O. (2003). Quantitative Kriging Neighbourhood Analysis for the Mining Geologist - A Description of the Method with Worked Case Examples. In *Fifth International Mining Geology Conference*, Victoria: AusIMM, pp. 17–19.

Wilde, B. J., & Deutsch, C. V. (2012). Kriging and Simulation in Presence of Stationary Domains: Developments in Boundary Modeling. In P. Abrahamsen, R. Hauge, & O. Kolbjørnsen (Eds.), *Geostatistics Oslo 2012. Quantitative Geology and Geostatistics 17*, Dordrecht: Springer, pp. 289–300, doi:10.1007/978-94-007-4153-9_23.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical Machine Learning Tools and Techniques* (4th ed.). Waltham, MA: Morgan Kaufmann, 654 p.

APÊNDICE A - CONTINUIDADE ESPACIAL 3D

Esta seção apresenta graficamente o modelamento dos correlogramas utilizados no estudo de caso 3D. As Figuras A-1 e A-2 apresentam os correlogramas experimentais (pontos vermelhos) e modelados (linha preta) nas três direções principais do domínio 3 e 4, respectivamente. Os parâmetros utilizados no cálculo de cada correlograma para ambos os domínios foram:

Maior / Intermediária

- Número de lags: 10
- Separação do lag: 100 m.
- Tolerância do lag: 50 m.

Menor

- Número de lags: 10
- Separação do lag: 25 m.
- Tolerância do lag: 12.5 m.

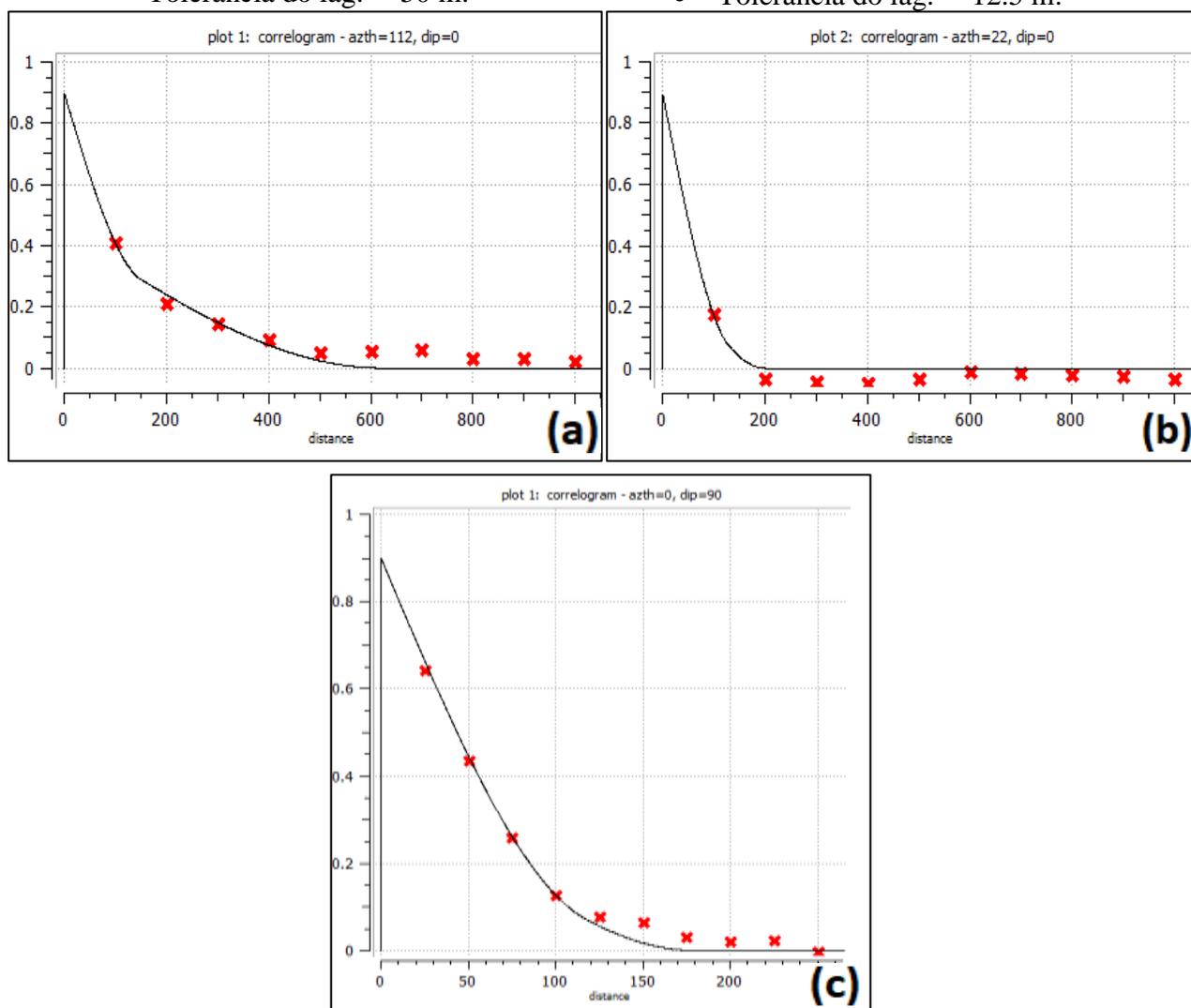


Figura A-1. Correlogramas modelados nas três principais direções do domínio 3 (a) Maior: N112 (b) Intermediária: N22 (c) Menor: D90.

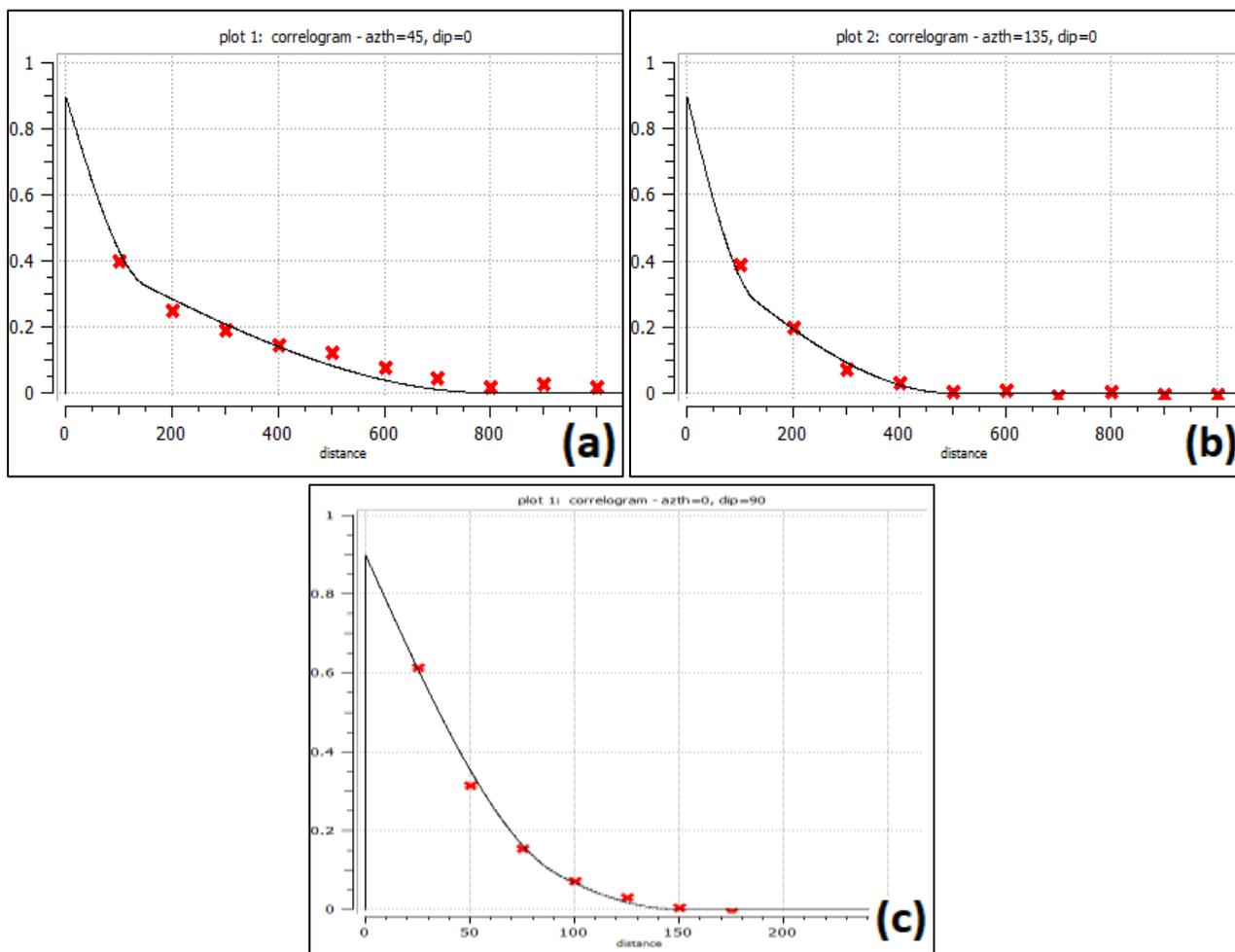


Figura A-2. Correlogramas modelados nas três principais direções do domínio 4 (a) Maior: N45 (b) Intermediária: N135 (c) Menor: D90.

APÊNDICE B - PLATAFORMA DO ALGORITMO DA LKPO

Esta seção apresenta uma descrição geral do algoritmo da LKPO implementado dentro da plataforma do SGeMS (Figuras B-1 e B-2).

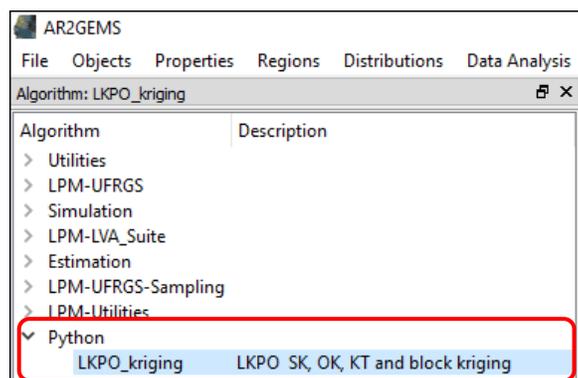


Figura B-1. Aba geral do SGeMS.

The image shows the 'General and Data' tab of the LKPO algorithm interface in SGeMS. The interface is divided into several sections with various input fields and controls. Red arrows point from descriptive labels to specific parts of the interface.

- General and Data Tab:** The top tab is labeled 'General and Data' and is highlighted with a red box.
- Kriging Grid:** A dropdown menu is set to '<- None ->'. A red arrow points to this dropdown with the label 'Grid'. Another red arrow points to a small icon next to the dropdown with the label 'Region (Grid)'. Below it is a text field for 'New Property Name' with a checkbox 're-use if existing'. A red arrow points to this text field with the label 'Property Name'.
- Hard Data:** A dropdown menu is set to '<- None ->'. A red arrow points to this dropdown with the label 'Dataset'. Below it is a folder selection icon. A red arrow points to this icon with the label 'Property'. Below the folder selection is a dropdown menu set to 'Simple Kriging (SK)'. A red arrow points to this dropdown with the label 'Kriging Type'. Below that is a text field for 'Mean'. A red arrow points to this text field with the label 'Mean for SK'.
- Block Kriging:** A checkbox labeled 'Block Kriging' is unchecked. Below it are three spinners for X, Y, and Z, all set to 5. A red arrow points to these spinners with the label 'Discretization of Block'.
- Search Ellipsoid:** A section containing 'Conditioning data' and 'Search Ellipsoid' parameters.
 - 'Conditioning data' includes 'Minimum: From 1 To 4' and 'Maximum: From 5 To 24'. A red bracket groups these with the label 'Variations in the Number of Samples'. Another red bracket groups the 'From' and 'To' values with the label 'From' and 'To'.
 - 'Search Ellipsoid' includes 'Samples by Sectors: From 1 To 4', 'Range 1: 100', 'Range 2: 100', 'Range 3: 10', 'Azimuth: 0.00', 'Dip: 0.00', and 'Strike: 0.00'. A red arrow points to the 'Range 1' spinner with the label 'Angles of Search Ellipsoid'. Another red arrow points to the 'Range 1', 'Range 2', and 'Range 3' spinners with the label 'Ranges of Search Ellipsoid'.
- Type of Optimization:** Three radio buttons are present: 'LKPO by Simulations' (unchecked), 'LKPO by Cross-Validation' (checked), and 'GSLIB Parameters' (unchecked). A red box highlights the 'LKPO by Cross-Validation' option with the label 'LKPO by Cross-Validation'.
- Executable Path:** A text field for 'Executable path:' is empty. A red arrow points to this field with the label 'Executable Path'. To the right, text says 'Create a folder and put inside:' followed by a list:
 - The kt3d executable (GSLib).
 - Dataset in GSLib format.
- Cross-Validation using:** Two radio buttons are present: 'Type: 1' (selected) and 'Type: 2' (unchecked). A red bracket groups these with the label 'Cross-Validation using: 1. Dataset. 2. Dataset + Drill Holes (DH)'. Below the radio buttons is a text field for 'Dataset Name:' with an example 'dataset.dat (.txt, .out, .ijnl)' and a note 'P.S.: DH is numeric (integer)'. A red arrow points to this text field with the label 'Dataset Name'.
- Columns of Dataset in GSLib Format:** A section titled 'Columns for:' with a table:

DH	X	Y	Z	Var
0	1	2	3	4

 A red arrow points to the 'Var' column with the label 'Variable'.
- Trimming Limits:** A section titled 'Trimming Limits:' with 'From: 0.00' and 'To: 1000.00'. A red arrow points to these fields with the label 'Trimming Limits'.
- Influence area (nearest neighbor):** A section titled 'Influence area (nearest neighbor):' with 'Range 3: 100', 'Range 2: 100', and 'Range 1: 10'. A red bracket groups these with the label 'Ranges of influence area for each sample obtained by nearest neighbor.'.
- Run Algorithm:** At the bottom, there are 'Load', 'Save', and 'Clear All' buttons, and a large 'Run Algorithm' button. A red arrow points to this button with the label 'Run Algorithm'.

Figura B-2. Aba geral do algoritmo da LKPO implementado dentro do SGeMS.

APÊNDICE C - CONTINUIDADE ESPACIAL 2D

Esta seção apresenta graficamente o modelamento dos correlogramas/variogramas utilizados em cada caso do estudo de caso 2D.

CASO I: Modelamento dos correlogramas/variogramas utilizados no estudo de caso 2D para o primeiro caso da geração dos domínios. Os parâmetros utilizados no cálculo de cada correlogramas/variogramas para ambos os domínios foram:

Maior / Menor

- Número de lags: 10
- Separação do lag: 10 m.
- Tolerância do lag: 5 m.

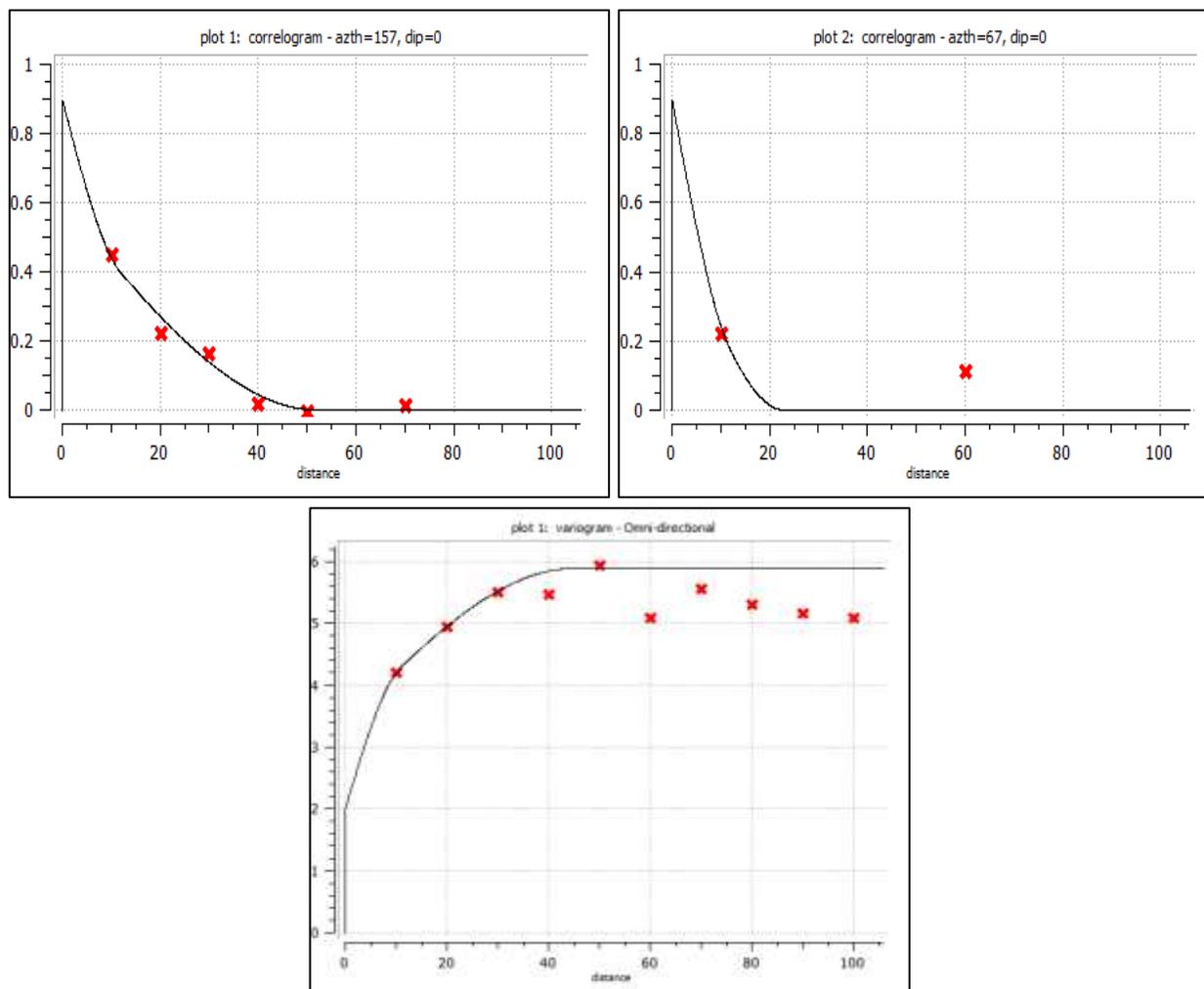


Figura C-1. CASO I: Modelos de continuidade espacial (a) domínio 1: N157 (b) domínio 1: N167 (c) domínio 2: omnidirecional.

CASO II: Modelamento dos correlogramas/variogramas utilizados no estudo de caso 2D para o segundo caso da geração dos domínios. Os parâmetros utilizados no cálculo de cada correlogramas/variogramas para ambos os domínios foram:

Maior / Menor

- Número de lags: 10
- Separação do lag: 10 m.
- Tolerância do lag: 5 m.

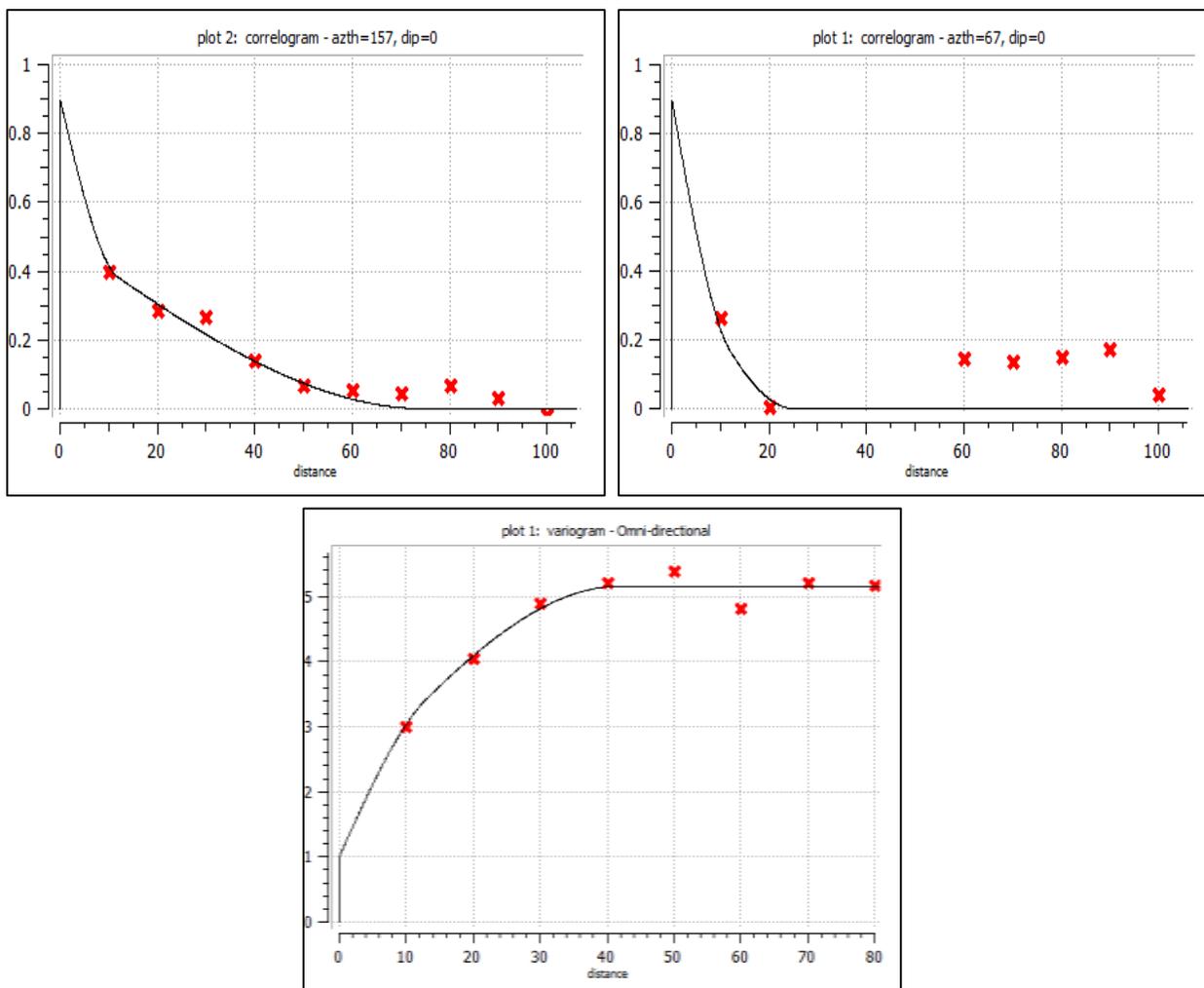


Figura C-2. CASO II: Modelos de continuidade espacial (a) domínio 1: N157 (b) domínio 1: N167 (c) domínio 2: omnidirecional.

CASO III: Modelamento dos correlogramas/variogramas utilizados no estudo de caso 2D para o quarto caso da geração dos domínios. Os parâmetros utilizados no cálculo de cada correlogramas/variogramas para ambos os domínios foram:

Maior / Menor

- Número de lags: 10
- Separação do lag: 15 m.
- Tolerância do lag: 7.5 m.

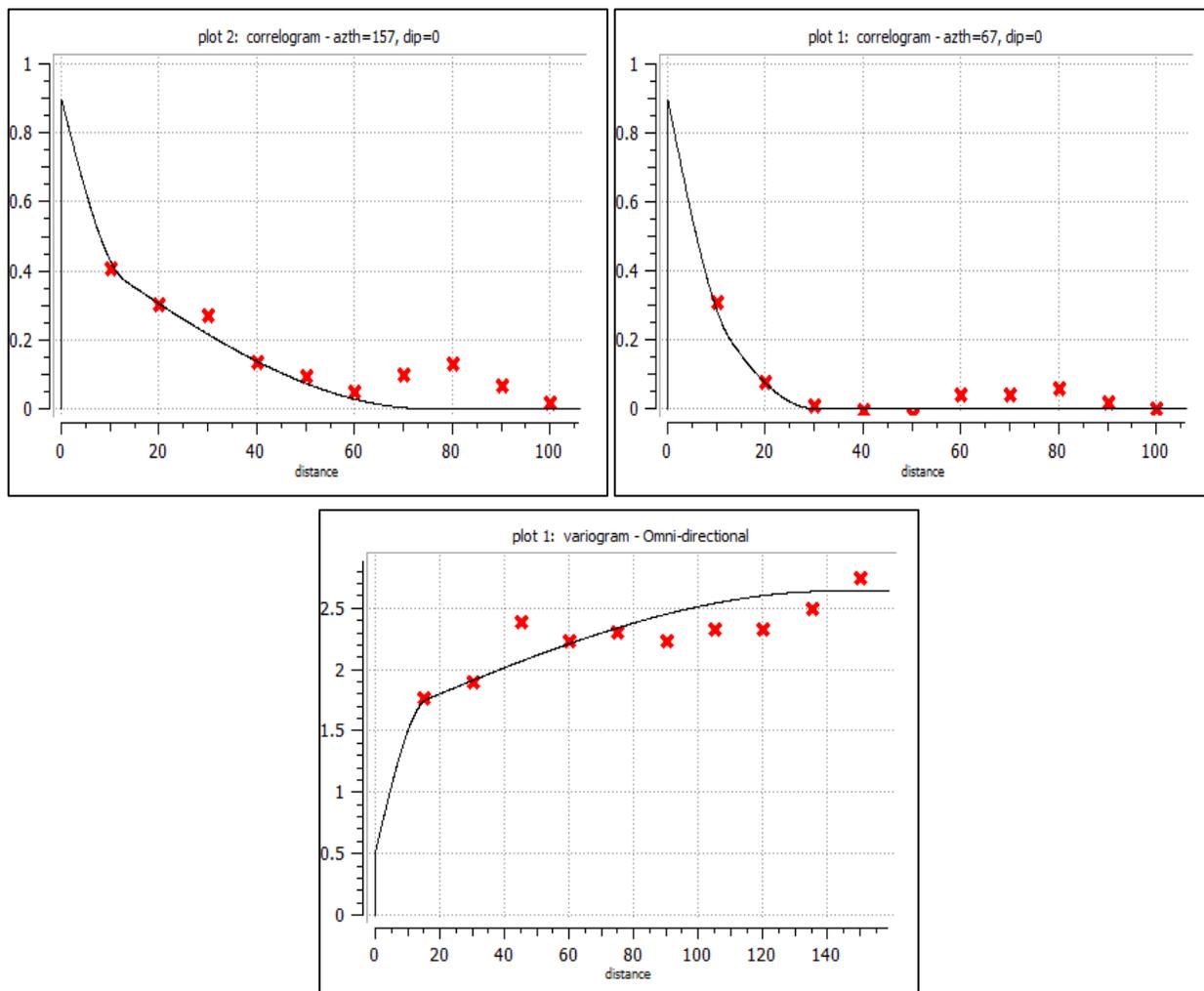


Figura C-3. CASO III: Modelos de continuidade espacial (a) domínio 1: N157 (b) domínio 1: N167 (c) domínio 2: omnidirecional.

CASO IV: Modelamento dos correlogramas/variogramas utilizados no estudo de caso 2D para o quarto caso da geração dos domínios. Os parâmetros utilizados no cálculo de cada correlogramas/variogramas para ambos os domínios foram:

Maior / Menor

- Número de lags: 10
- Separação do lag: 10 m.
- Tolerância do lag: 5 m.

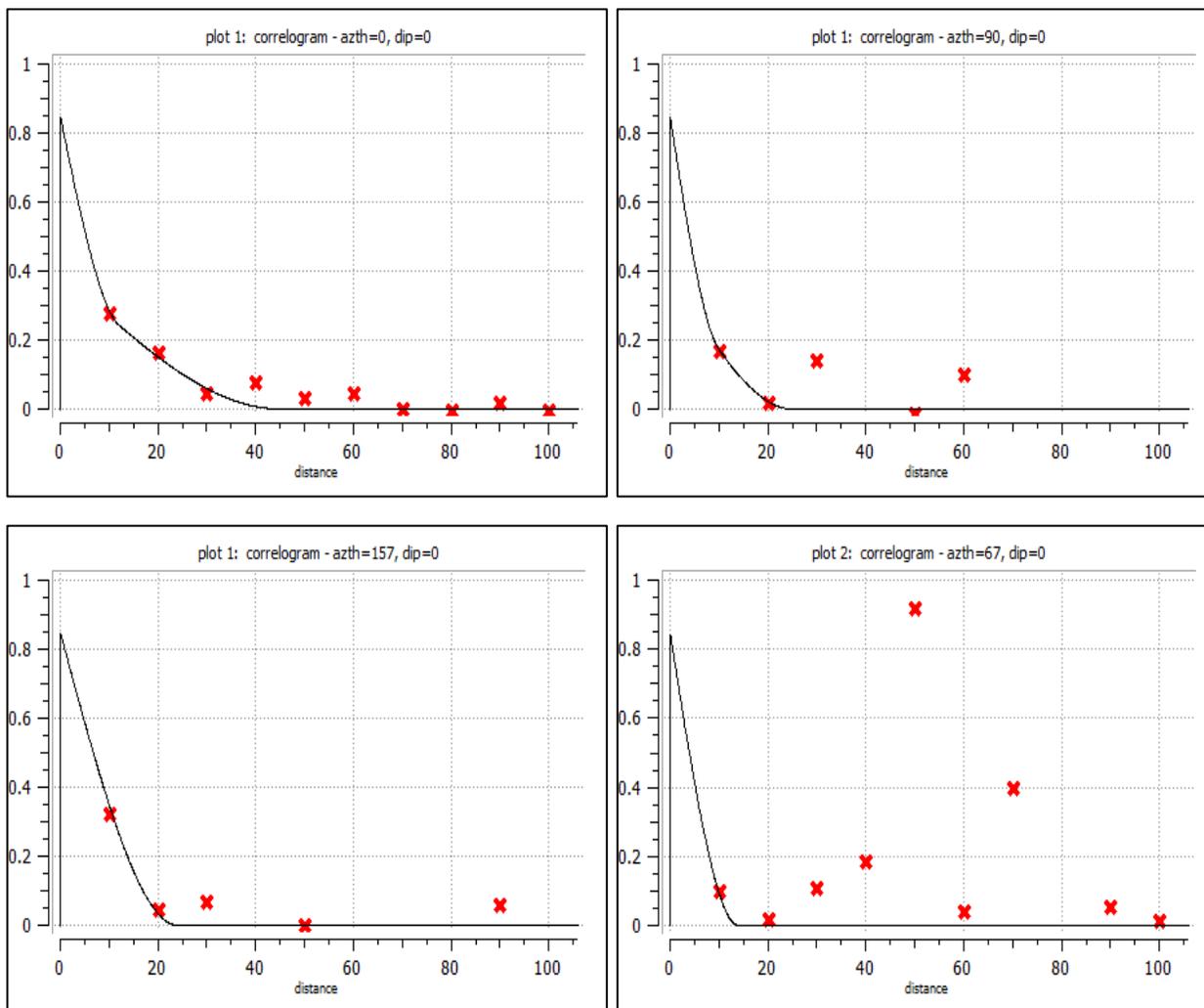


Figura C-4 parte I. CASO IV: Modelos de continuidade espacial (a) domínio 1: N0 (b) domínio 1: N90 (c) domínio 2: N157 (b) domínio 2: N67

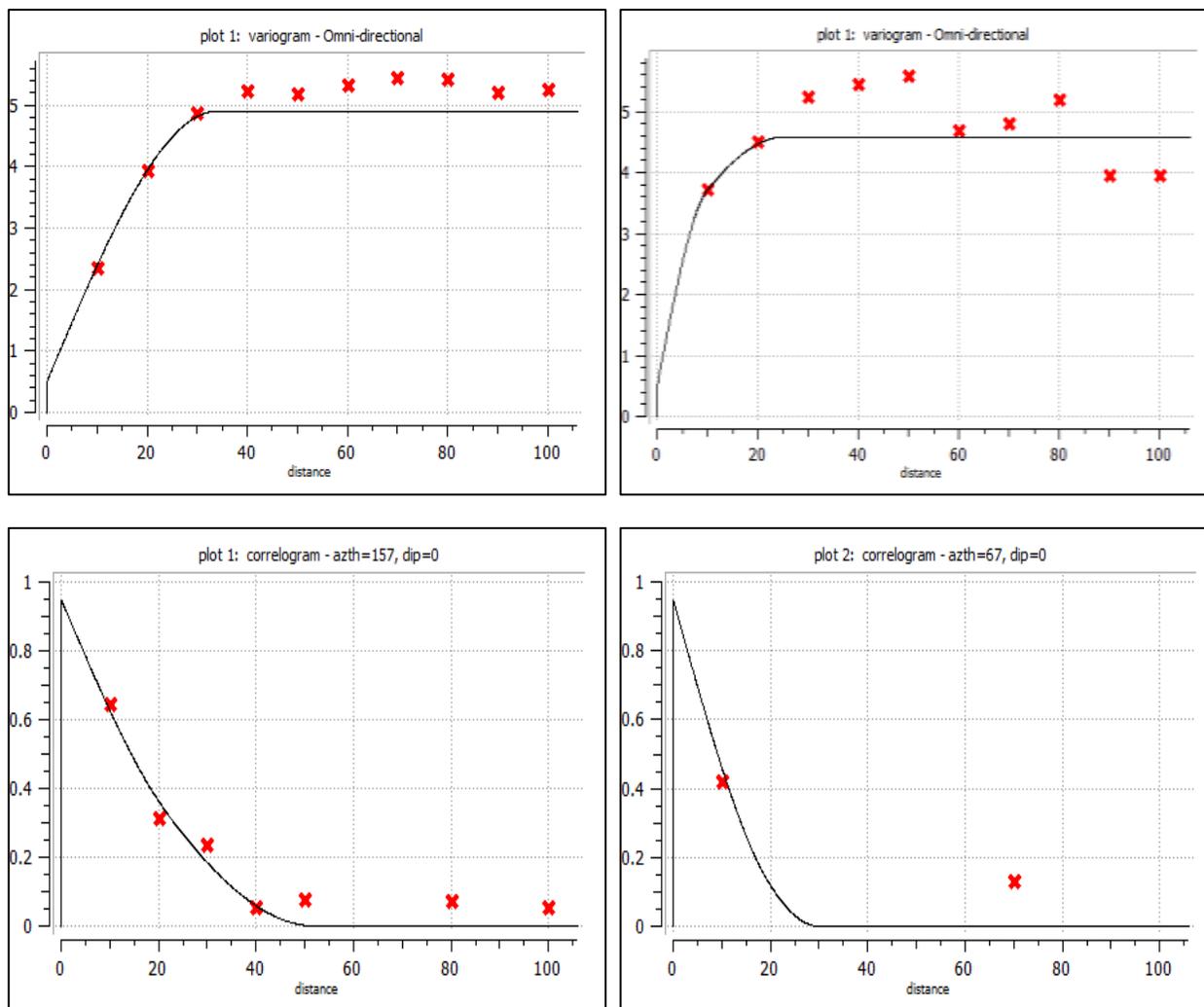


Figura C-4 parte II. CASO IV: Modelos de continuidade espacial (a) domínio 3: omnidirecional (b) domínio 4: omnidirecional (c) domínio 5: N157 (b) domínio 5: N67

