

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)

**CODA: UMA ALTERNATIVA PARA ESTIMATIVAS MULTIVARIADAS QUE
ENVOLVEM BALANÇOS DE MASSA GRANULOMÉTRICO E DAS
ESPÉCIES QUÍMICAS**

Ricardo José Hundelshaussen Rubio

Dissertação para a obtenção do título de
Mestre em Engenharia

Porto Alegre, RS

2014

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)

**CODA: UMA ALTERNATIVA PARA ESTIMATIVAS MULTIVARIADAS QUE
ENVOLVEM BALANÇOS DE MASSA GRANULOMÉTRICO E DAS
ESPÉCIES QUÍMICAS**

Ricardo José Hundelshausen Rubio

Engenheiro Industrial

Dissertação 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 Mestre em Engenharia.

Área de concentração: Metalurgia Extrativa e Tecnologia Mineral

Porto Alegre, RS

2014

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

Orientador:

Prof. Dr. João Felipe Coimbra Leite Costa

Banca Examinadora:

Dr. Jair Carlos Koppe

Dra. Maria Noel Morales Boezio

Dra. Vanessa Cerqueira Koppe

Prof. Dr. Telmo Roberto Strohaecker
Coordenador do PPGEM

“Quem não assume um risco, nunca ganhará um jogo”

G.M. Paul Keres

A meus pais Alejandra e Joaquín,
familiares e amigos pelo
apoio e dedicação.

AGRADECIMENTOS

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

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

À Dra. Maria Noel Morales Boezio pela orientação e valorosas contribuições para o desenvolvimento deste trabalho.

Aos colegas e amigos do LPM que contribuíram aportando um grão de areia neste trabalho, em especial Anuar, Diego, Pablo, George, Ricardo, Vanessa, Marcel, Roberto, Oluwafemi, Luis Gustavo, Samanta, Henrique, Péricles, Camila, Cristina, Luciana e Taís.

À empresa de Mineração Rio do Norte pela disponibilização dos dados para este estudo.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela bolsa concedida.

A meus pais Alejandra e Joaquín, meus irmãos Laura Andrea, Maria Alejandra e Joaquín Fernando, a todos meus tios, primos e demais familiares. A minha namorada Leidy pelo apoio e paciência.

A meus amigos Colombo-Gauchos Javier, Rosane, Karen, Grethel, Victoria, Dayana, Fabrício, Fausto, Silmalila, Carlos, Sonia, Lyda, Carolina, Felix e Renildes por compartilhar momentos agradáveis.

Sumário

Lista de Figuras.....	X
Lista de Tabelas	xii
Resumo.....	xv
Capitulo 1 – Introdução	1
1.1. Introdução	1
1.2. Metodologias.....	3
1.3. Aspectos gerais.....	5
1.4. Estado da arte.....	6
1.5. Problema	9
1.6. Meta da dissertação	10
1.7. Objetivos específicos da dissertação.....	10
1.8. Metodologia	11
1.9. Organização dessa dissertação.....	12
Capitulo 2 – Conceitos Gerais	13
2.1. Transformação de coordenadas.....	13
2.2. Acumulação	17
2.3. Continuidade espacial	19
2.3.1. Variograma:.....	20
2.3.2. Covariância:.....	20
2.3.3. Correlograma:	21
2.4. Estimativa	22
2.4.1. Krigagem ordinária (OK):.....	23
2.4.2. Cokrigagem Ordinária (CKO):	25
2.5. Simulação seqüencial	26
2.6. Simulação seqüencial gaussiana.....	27
2.7. Simulação por bandas rotativas	30
2.8. Dados composicionais	32
2.8.1. Conceitos básicos	32

2.8.2.	Estrutura do espaço vetorial no D-Simplex.....	33
2.8.3.	Transformações logarítmicas	34
2.8.4.	Metodologia proposta	36
2.8.5.	Presença de zeros.....	38
Capítulo 3 – Como Lidamos com Problemas Multivariados Quando as Somas das Estimativas Devem Fechar 100%?		39
3.1.	Localização.....	39
3.2.	Banco de dados	40
3.3.	Estatística básica.....	43
3.4.	Acumulação	45
3.5.	Desagrupamento	47
3.6.	Transformação de coordenadas.....	49
3.7.	Continuidade espacial	51
3.8.	Estimativa	56
3.9.	Análise parcial dos resultados	60
3.9.1.	Metodologia 1 (isotópico – individuais - krigagem ordinária).....	61
3.9.2.	Metodologia 2 (isotópico – único - krigagem ordinária)	63
3.9.3.	Metodologia 3 e 4 (heterotópico – único - individual - krigagem ordinária)	64
3.9.4.	Metodologia 5 (isotópico – único - individual - cokrigagem).....	67
Capítulo 4 – Transformações logarítmicas isométricas (ilr)		73
4.1.	Transformações logarítmicas isométricas (ilr)	73
4.2.	Continuidade espacial	80
4.3.	Estimativa	83
4.4.	Viés na média aritmética	85
4.5.	Reprodução da média geométrica.	88
4.6.	Retro-transformação (ilr-1)	89
4.7.	Análise preliminar dos resultados	92
4.7.1.	Resultados com metodologia 1 (ilr - isotópico – variogramas individuais - krigagem ordinária)	92

4.7.2. Resultados com metodologia 2 (ilr- isotópico – variograma único - krigagem ordinária).....	94
4.7.3. Metodologia 1 e 2 utilizando uma krigagem em blocos.....	96
4.8. Simulação.....	98
4.8.1. Aspectos gerais.....	99
4.8.2. Análise preliminar dos resultados.....	101
Capítulo 5 – Validações.....	105
5.1. Reprodução da média global.....	105
5.2. Reprodução da média local (análise de deriva)	106
5.3. Soma dos pesos positivos.....	108
5.4. Análise visual	109
5.5. E-type das simulações vs estimativas e desagrupamento	111
Capítulo 6 – Conclusões e Recomendações.....	113
5.1. Conclusões.....	113
5.1.1. Primeira fase	113
5.2.2. Segunda fase	115
5.2. Recomendações para trabalhos futuros	116
Referências Bibliográficas	117

Lista de Figuras

Figura 1.1. Dependência linear das variáveis em depósitos multi-elementais dados pelos balanços das espécies químicas e os balanços granulométricos. Figura modificada da tese doutoral de Boezio (2010).....	2
Figura 1.2. Metodologia de trabalho.	11
Figura 2.1. Interpretação de coordenadas entre diferentes amostras, (a) coordenadas cartesianas (b) coordenadas estratigráficas.....	14
Figura 2.2. Interpretações sobre eventos ou processos geológicos, imagem modificada de Deutsch (2002).	15
Figura 2.3. Correção estratigráfica de uma camada, imagem modificada de Deutsch (2002).....	16
Figura 2.4. Influência do suporte amostral na média.	17
Figura 2.5. Flutuações ergódicas nos variogramas. Imagem modificada do (Deutsch & Journel, 1998).	27
Figura 2.6. Princípio das bandas rotativas em 2D (imagem extraída de Chilès & Delfiner,1999).....	30
Figura 3.1. Mapa de localização da área de estudo, imagem tomada o dia 30-07-2013 às 13:00 do site da empresa MRN http://www.mrn.com.br/pt-BR/Sobre-MRN/Porto-Trombetas/MapDeLocalizacao/Paginas/default.aspx	39
Figura 3.2. Mapa de localização da variável FE na fração +14, 2D.	42
Figura 3.3. Histogramas de frequências (a) Variável Alap14 (b) Variável Sire14.....	44
Figura 3.4. (a) Histograma ALAP14 (Original), (b) Histograma ALAP14A (Acumulada). 46	
Figura 3.5. (a) Camada sem correção, (b) Camada corrigida pela Lapa. Imagem tomada da apresentação final do projeto MRN 2013.	50
Figura 3.6. (a) Distribuição do Erro Rec14 na CAPA, (b) Distribuição do Erro Rec14 na LAPA.....	50
Figura 3.7. Correlograma experimental Rec14 nas 8 direções principais e o modelo ajustado.	52
Figura 3.8. Correlograma experimental Rec-400 nas 8 direções principais e o modelo ajustado.	53
Figura 3.9. Correlograma não ergódico variável Rec14 (a) Maior, (b) Intermediária, (c) Menor, (d) Mapa variográfico no plano XY.	55
Figura 3.10. Anisotropia zonal da variável ALAP14A no plano XY.	59

Figura 3.11. Anisotropia zonal da variável Ti400A no plano vertical.	60
Figura 3.12. Histograma de fechamento metodologia 1, (a) faixa +14#, (b) faixa +400#, (c) frações mássicas.	62
Figura 3.13. Histograma de fechamento metodologia 2, (a) faixa +14#, (b) faixa +400#, (c) frações mássicas.	64
Figura 3.14. Histograma de fechamento metodologia 3, (a) faixa +14#, (b) faixa +400#, (c) frações mássicas.	65
Figura 3.15. Histograma de fechamento metodologia 4, (a) faixa +14#, (b) faixa +400#, (c) frações mássicas.	66
Figura 3.16. Diagramas de dispersão da krigagem ordinária comparada com cokrigagem utilizando o modelo da REC14, (a) Rec14, (b) Rec400, (c) Rec-400, (d) Histograma de fechamento Metodologia 5.	68
Figura 4.1. Processo de cálculo das transformações <i>ilr</i>	77
Figura 4.2. Histogramas das transformações <i>ilr</i> , (a) <i>ilr</i> _{14_1} , (b) <i>ilr</i> _{400_4}	78
Figura 4.3. Histogramas das transformações <i>ilr</i> acumuladas, (a) <i>ilr</i> _{14_1A} , (b) <i>ilr</i> _{400_4A}	79
Figura 4.4. Correlograma experimental <i>ilr</i> _{14_2A} nas 8 direções principais e o modelo ajustado.	80
Figura 4.5. Correlograma não ergódico da variável <i>ilr</i> _{14_2A} nas direções de (a) Maior, (b) Intermediária, (c) Menor continuidade, (d) Mapa variográfico no plano XY.	82
Figura 4.6. Anisotropia zonal da variável <i>ilr</i> _{400_2A} no plano vertical.	85
Figura 4.7. Processo para atenuar o viés na média aritmética utilizando transformações não lineares.	87
Figura 4.8. Processo de transformação logarítmica isométrica inversa (<i>ilr</i> -1).	91
Figura 4.9. Histograma de fechamento metodologia 1 usando uma krigagem pontual, (a) faixa +14#, (b) faixa +400#, (c) frações mássicas.	94
Figura 4.10. Histograma de fechamento metodologia 2 usando uma krigagem pontual, (a) faixa +14#, (b) faixa +400#, (c) frações mássicas.	95
Figura 4.11. Histograma de fechamento metodologia 1 usando uma krigagem em blocos, (a) faixa +14#, (b) faixa +400#, (c) frações mássicas.	97
Figura 4.12. Histograma de fechamento metodologia 2 usando uma krigagem em blocos, (a) faixa +14#, (b) faixa +400#, (c) frações mássicas.	98
Figura 4.13. Processo de simulação para as transformações logarítmicas isométricas (<i>ilr</i>).	99

Figura 4.14. Correlograma da variável <i>ilr14A</i> gaussiana, (a) Plano horizontal N0, (b) Plano vertical N90.....	100
Figura 4.15. Histogramas dos dados originais e simulações nas frações granulométricas (a) Rec14, (b) Rec14_Sim, (c) Rec400, (c) Rec400_Sim, (e) Rec-400, (e) Rec-400_Sim.....	102
Figura 4.16. Correlograma dos dados originais (cor vermelha) vs Correlograma das simulações (a) Rec14, (b) Rec400, (c) Rec-400.....	103
Figura 4.17. Histograma de fechamento nas simulações (a) Cenário 08, (b) Cenário 15, (c) Cenário 21, (d) Cenário 27.....	104
Figura 5.1. Análise de deriva da variável <i>Alap14</i> entre a krigagem ordinária (OK), krigagem das (<i>ilr</i>) e o desagrupamento por NN (a) Plano X, (b) Plano Y, (c) Plano Z.	107
Figura 5.2. Análise de deriva da variável <i>Rec14</i> entre a krigagem ordinária (OK), krigagem das (<i>ilr</i>) e o desagrupamento por NN (a) Plano X, (b) Plano Y, (c) Plano Z.	107
Figura 5.3. Soma dos pesos positivos da krigagem pontual (a) <i>Rec14</i> , (b) <i>Rec400</i>	108
Figura 5.4. Soma dos pesos positivos da krigagem pontual (a) <i>ilrRec_1</i> , (b) <i>ilrRec_2</i> .	109
Figura 5.5. Análise visual das estimativas vs as amostras (a) <i>Alap14</i> , (b) <i>Fe14</i> , (c) <i>Ti14</i> , (d) <i>Rec14</i>	110
Figura 5.6. Análise de deriva da variável <i>Rec14</i> entre o e-type das simulações, krigagem ordinária (OK), krigagem das (<i>ilr</i>) e o desagrupamento por NN (a) Plano X, (b) Plano Y, (c) Plano Z.	112
Figura 5.7. Análise de deriva da variável <i>Alap14</i> entre o e-type das simulações, krigagem ordinária (OK), krigagem das (<i>ilr</i>) e o desagrupamento por NN (a) Plano X, (b) Plano Y, (c) Plano Z.....	112

Lista de Tabelas

Tabela 3.1. Resumo das características principais do depósito de Bauxita.....	40
Tabela 3.2. Resumo das características das variáveis analisadas no depósito.....	41
Tabela 3.3. Resumo da estatística básica dos dados originais.....	43
Tabela 3.4. Coeficiente de assimetria das variáveis analisadas.....	45
Tabela 3.5. Resumo estatístico das variáveis acumuladas.....	47
Tabela 3.6. Resumo da média desagrupada ponderada e a média dos dados ponderados.....	48

Tabela 3.7. Parâmetros utilizados na Variografia Omnidirecional e Direcional.	53
Tabela 3.8. Resumo dos parâmetros encontrados na modelagem da variografia. ...	54
Tabela 3.9. Resumo das características metodológicas utilizadas na estimativa.	57
Tabela 3.10. Resumo dos parâmetros para os 4 testes da krigagem.	57
Tabela 3.11. Resumo dos parâmetros utilizados na busca da krigagem.	58
Tabela 3.12. Comparação entre os valores mínimos e máximos dos dados estimados e os dados originais (Metodologia 1).	61
Tabela 3.13. Comparação entre os valores mínimos e máximos dos dados estimados e os dados originais (Metodologia 2).	63
Tabela 3.14. Comparação entre os valores mínimos e máximos dos dados estimados e os dados originais (Metodologia 3).	65
Tabela 3.15. Comparação entre os valores mínimos e máximos dos dados estimados e os dados originais (Metodologia 4).	66
Tabela 3.16. Parâmetros das contribuições e o efeito pepita do modelo MCI nas recuperações (Metodologia 5).	69
Tabela 3.17. Coeficiente de Correlação (Metodologia 5) da Fração Granulométrica +14 Acumulada.	69
Tabela 3.18. Coeficiente de Correlação (Metodologia 5) da Fração Granulométrica +400 Acumulada.	70
Tabela 3.19. Comparação entre os valores mínimos e máximos dos dados estimados e os dados originais (Metodologia 5).	70
Tabela 3.20. Percentagem de blocos negativos obtidos pelos 6 testes da Cokrigagem.	72
Tabela 4.1. Partição sequencial binária para cada fração (+14#, +400#, -400#).	75
Tabela 4.2. Estatística básica das transformações logarítmicas isométricas (ilr).	78
Tabela 4.3. Resumo estatístico das variáveis transformadas (ilr) acumuladas.	79
Tabela 4.4. Resumo dos parâmetros encontrados na modelagem da variografia. ...	81
Tabela 4.5. Resumo dos parâmetros utilizados na busca da krigagem.	84
Tabela 4.6. Comparação da media aritmética usando transformações lineares e não lineares.	86
Tabela 4.7. Comparação da media geométrica usando transformações lineares e não lineares.	88
Tabela 4.8. Matriz ψ que representa as coordenadas para cada balanço.	90

Tabela 4.9. Valores mínimos e máximos dos dados estimados usando uma krigagem pontual e os dados originais (Metodologia 1).	93
Tabela 4.10. Valores mínimos e máximos dos dados estimados usando uma krigagem pontual e os dados originais (Metodologia 2).	95
Tabela 4.11. Valores mínimos e máximos dos dados estimados usando uma krigagem em blocos e os dados originais (Metodologia 1).	96
Tabela 4.12. Valores mínimos e máximos dos dados estimados usando uma krigagem em blocos e os dados originais (Metodologia 2).	97
Tabela 4.13. Valores mínimos e máximos dos dados simulados e os dados originais (Metodologia 1).	101
Tabela 5.1. Média desagrupada vs Média das Estimativas.	106

Resumo

Na maioria das minas, até poucos anos, só havia preocupação em estimar-se o teor de uma ou no máximo duas espécies químicas de interesse, por exemplo, teor de cobre, ouro ou ferro. Hoje, existem padrões de qualidade em produtos e concentrados que exigem um rigoroso controle de outros elementos além do metal/mineral minério. É comum ser necessário a estimativa de múltiplos elementos, possivelmente correlacionados e em algumas vezes com uma combinação de teores que deva fechar uma soma constante (por exemplo, 100%). Dados que somam uma constante são conhecidos como dados composicionais (CODA), carregando consigo informação relativa e não absoluta. Esta situação (soma de uma constante) condiciona a que as estimativas também devam fechar uma constante. As metodologias clássicas, como é o caso da krigagem e a cokrigagem ordinária, podem ser apropriadas para obter a melhor estimativa local do ponto de vista teórico. Mas, cada variável é estimada separadamente (no caso OK) com seus respectivos parâmetros de medidas de continuidade espacial, o que leva a obter pesos diferentes para cada atributo e em contraposto gera problemas para o fechamento final das estimativas. Novas técnicas estatísticas podem ser adaptadas ao estudo de CODA que permitam garantir, além de resultados coerentes, o fechamento após estimativas. Esta dissertação utilizou as transformações de razões logarítmicas isométricas (ilr) que permite projetar o espaço amostral dos dados originais (espaço euclidiano) ao espaço simplex de $D-1$ partes da composição (S^D). Os resultados mostraram-se satisfatórios, já que além de obter bons resultados, foi garantido o fechamento após estimativas. Esses resultados foram comparados com os resultados obtidos na krigagem tradicional utilizando os mesmos critérios de interpolação. Ambas as metodologias produziram resultados semelhantes em termos de qualidade de estimativa, mostrando-se ser uma metodologia alternativa para estimativa em depósitos multivariados. Uma desvantagem nesta transformação é o uso de transformações de tipo não lineares combinada com interpoladores lineares (OK), já que quando fazemos o processo de retro-transformação (ilr) estamos

expostos a um viés na média estimada. Para solucionar esse problema, foi utilizada a simulação das transformações (*ilr*) que não gera viés, já que o dado simulado é obtido a partir de uma tiragem randômica de função de probabilidade local. Comparou-se então, os resultados do e-type das simulações (*ilr*) com os resultados da krigagem (*ilr*) para observar a influência do viés da média aritmética localmente. Os resultados mostraram que, embora esse viés exista teoricamente, para este estudo de caso não foi representativo, ou seja, a média local entre o e-type das simulações (*ilr*) e a krigagem (*ilr*) não é maior que 5%.

ABSTRACT

In most of the mines, until a few years ago, there was only concern in estimating the content of one or at most two chemical species of interest, such as percentage of copper, gold or iron. Today, there are standards of quality in products and concentrates, which require careful control of other elements besides metal / ore minerals. It is common to estimate multiple elements, possibly correlated and sometimes with a combination of contents which must close a constant sum (e.g. 100%). Data that add up to a constant are known as compositional data (CODA), carrying information relative and not absolute. This (sum of a constant) determines that the estimates should also close a constant. The classical methods, such as kriging and ordinary cokriging, may be appropriate for the best local estimate. However, each variable is estimated separately (in case OK) with their specific parameters of spatial continuity, which leads to obtain different weights for each attribute and fails to obtain estimates that satisfy the constant sum constraint. New statistical techniques can be adapted to the study of CODA that guarantee consistent results after the closing estimates. This dissertation uses the isometric transformations of logarithmic ratios (*ilr*) that allows to transform the sample space of the original data (Euclidean space) into the simplex space with D-1 parts of the composition (S^D). The results were satisfactory providing closed sums after estimates. These results were compared with the results obtained using the traditional kriging interpolator. Both methods produced similar results in terms of quality of the estimation, proving to be an alternative methodology to estimate multivariate deposits. A disadvantage in this transformation is the use of nonlinear transformations combined with linear interpolation (OK), since when doing this process of retro-transformation (*ilr*) can lead to a bias in the estimated average. To solve this problem, it was used simulation of the transformation (*ilr*) and this solution do not generates bias, since the simulated data is obtained from randomly drawing of a local probability function. The average of all simulations wase compared with kriging to observe the impact of the arithmetic bias. The results showed that, although this bias exists theoretically, for this case study it

was not significant, i.e., the difference from the local average of the e-type simulations (ilr) and kriging (ilr) is not greater than 5%.

Capítulo 1 – Introdução

1.1. Introdução

Os depósitos minerais são caracterizados por ter, além do elemento principal (ouro, cobre, níquel, ferro, carvão, etc.) outras espécies químicas que afetam, em termos de viabilidade econômica, processos industriais e/ou planejamento de lavra. Além disso, há impactos ambientais e custos de penalização em relação ao nível de concentração de contaminantes presentes em cada minério. Isto levou as empresas de mineração a se interessarem pelo estudo apropriado de cada uma desses elementos e cumprir com as exigências estabelecidas pelos diversos controles de qualidade.

Uma característica importante nos depósitos como ferro, bauxita e manganês é a análise das espécies químicas em diversas faixas granulométricas. Cada uma dessas variáveis apresenta uma correlação em relação a elas e, também, entre as mesmas nas diversas faixas granulométricas. Boezio (2010) apresenta uma tabela onde explica claramente a dependência linear entre as variáveis analisadas dadas pelas relações de balanço entre as diferentes espécies químicas e, pelos balanços de massa entre as diferentes frações granulométricas (ver figura 1.1).

Outra característica importante neste tipo de depósitos multi-elementos é que teoricamente as somas de todas as espécies químicas devem fechar o balanço mássico entre as diferentes frações granulométricas e espécies químicas, neste caso, se estivéssemos falando em percentagem a soma seria 100%. Geralmente isto não acontece, pois existem muitos fatores que interatuam na qualidade da informação.

Espécie Química (EQ)	Faixa Granulométrica (FG)							BALANÇOS DE MASSA
	FG ₁	FG ₂	FG ₃	...	FG _i	...	FG _q	
EQ ₁	EQ ₁₁	EQ ₁₂	EQ ₁₃	...	EQ _{1i}	...	EQ _{1q}	BALANÇOS DE MASSA
EQ ₂	EQ ₂₁	EQ ₂₂	EQ ₂₃	...	EQ _{2i}	...	EQ _{2q}	
EQ ₃	EQ ₃₁	EQ ₃₂	EQ ₃₃	...	EQ _{3i}	...	EQ _{3q}	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
EQ _j	EQ _{j1}	EQ _{j2}	EQ _{j3}	...	EQ _{ji}	...	EQ _{jq}	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
EQ _n	EQ _{n1}	EQ _{n2}	EQ _{n3}	...	EQ _{ni}	...	EQ _{nq}	

BALANÇOS DAS ESPÉCIES QUÍMICAS

Figura 1.1. Dependência linear das variáveis em depósitos multi-elementares dados pelos balanços das espécies químicas e os balanços granulométricos. Figura modificada da tese doutorado de Boezio (2010).

Segundo Marat (2011), os erros mais comuns na má amostragem podem ser gerados em qualquer fase da análise amostral (extração, preparação ou ensaio químico das amostras). Segundo o mesmo autor, esses erros podem ser classificados como:

- *Erro fundamental de amostragem* (Gy, 1982), relacionada com as características intrínsecas do material amostrado, tais como mineralogia e textura da mineralização.
- *Erro de agrupamento e segregação* (Pitard, 1993), que é uma consequência da heterogeneidade da distribuição do material amostrado.
- *Erros relacionados com as práticas de amostragem*, que dependem de quão rigoroso, aplicado e seguido é o protocolo de amostragem, incluindo erros na extração, delimitação, preparação ou pesagem.
- *Erros instrumentais*, que ocorrem durante as operações analíticas, incluindo ensaios, análise de umidade, peso das amostras, análise de densidade, erros de precisão e viés causado pelo desempenho dos instrumentos analíticos.

Muitos desses erros simplesmente podem ser minimizados através de processos de melhorias nos protocolos de amostragem, outros sempre estarão presentes e não podem ser completamente eliminados. Portanto, é necessário

fazer-se um pós-processamento que leve ao fechamento dos dados em cada uma das frações analisadas.

Alguns dos procedimentos empíricos utilizados na indústria mineira quando o balanço granulométrico não fecha, por exemplo, em 100% são:

- Distribuir equitativamente uma pequena porção desse erro (chamaremos erro de fechamento à percentagem que falta para fechar uma fração numa constante, geralmente 100%) em cada uma das variáveis analisadas.
- Criar uma variável artificial que carregue todo o erro, ou seja, a nova variável terá a finalidade de fechar o balanço granulométrico.

No entanto, este tipo de variáveis conhecidos como *dados composicionais CODA* (Aitchison, 1981, 1982, 1983, 1984, 1986), estão condicionados ao fechamento numa constante dos dados analisados antes de começar com um processo de análise geoestatístico, para que o objeto de estudo seja além de ter uma boa estimativa, garantir a soma fechada dos dados composicionais no processo final após estimativas.

1.2. Metodologias

Os métodos tradicionais de estimativa foram desenvolvidos inicialmente por Matheron (1963) baseado nos estudos iniciais de Krige (1951) sobre estimativas de ouro numa mina na África do Sul. Matheron contribuiu com o estudo quantitativo dessas variáveis regionalizadas, levando em consideração a localização geográfica e dependência espacial das amostras.

Partindo do fato de que a melhor estimativa de uma variável é usando a sua própria informação, a krigagem ordinária (OK) surge como o primeiro método interpolador utilizado nesta dissertação. A vantagem principal é que cada variável é modelada com seu próprio variograma, o que garante uma adequada representatividade física da realidade do depósito. Sua desvantagem é que neste tipo de casos multivariados, cada variável, locada em um mesmo

ponto amostral, irá receber pesos diferentes em relação a seu modelo de continuidade, o que impede o fechamento químico e mássico em cada bloco estimado.

Uma segunda abordagem é definida na geoestatística multivariada e surge como necessidade de suprir as desvantagens na krigagem, ao determinar de forma simultânea, múltiplas variáveis regionalizadas (Matheron, 1965). O segundo método interpolador utilizado nesta dissertação é conhecido como cokrigagem ordinária (Marechal, 1970), a sua vantagem principal consiste em incorporar a correlação conjunta entre as variáveis analisadas, além de gerar estimativas não tendenciosas que minimizam a variância do erro (Wackernagel, 1995).

Boezio (2010) menciona uma serie de aspectos problemáticos na Cokrigagem:

- i. Estimativas fora do intervalo original das amostras e/ou negativas que precisam ser pós-processadas.
- ii. A não satisfação, por parte dos teores estimados, dos balanços de massa e das espécies químicas (somadas constantes), sendo necessária a distribuição do erro ou carregá-lo em uma variável que seja determinada a partir dessas relações.
- iii. Dificuldades na modelagem da corre regionalização que faz necessários softwares que consigam satisfazer as condições de definição positiva impostas pelo Modelo Linear de Corre regionalização (MLC), necessário na cokrigagem.
- iv. Modelos variográficos que não se ajustam adequadamente aos variogramas experimentais diretos e cruzados, mas satisfaçam o MLC.

Nesse sentido, foi proposto implementar uma metodologia que permita trabalhar com os dados composicionais (CODA) usando transformações de tipo logarítmicas nos dados originais antes de ser estimados, permitindo depois o uso da krigagem e a cokrigagem ordinária para fins de estimativas. Ao final, os

dados são retro-transformados voltando novamente ao espaço real, analisando o ganho na qualidade das estimativas e o fechamento granulométrico e das espécies químicas das variáveis estudadas. Essa metodologia é conhecida por Egozcue & Pawlowsky (2003) como transformações logarítmicas isométricas (ilr) definido no próximo capítulo.

Um novo problema surge ao utilizar transformações de tipo não lineares nas estimativas e é que a média aritmética é sensível a gerar viés na retro-transformação (ver exemplo capítulo 4). Esse viés é produzido pela mesma transformação não linear, já que os métodos clássicos como krigagem e a cokrigagem, baseiam as suas estimativas em interpolações lineares.

A simulação então surge como uma solução a esses problemas. O terceiro método utilizado nesta dissertação é a simulação seqüencial gaussiana e a simulação por bandas rotativas definidas no capítulo 2. A sua principal vantagem é produzir simulações não enviesadas e garantir os balanços químicos e mássicos utilizando as transformações logarítmicas isométricas.

1.3. Aspectos gerais

Os depósitos como bauxita são formados pela ação do intemperismo que forma rocha sedimentar, levando à formação de depósitos com camadas plano-paralelas ou em forma de bacias. Alguns eventos geológicos posteriores, como dobramentos ou a própria forma de bacia, afetam o modelamento geoestatístico tridimensional, pois pode acontecer que duas ou mais amostras estejam próximas uma da outra depositadas em épocas geológicas diferentes e não apresentar continuidade espacial.

Koppe (2005) diz que o modelamento geoestatístico pode ser aprimorado, se essas coordenadas cartesianas forem transformadas em coordenadas geológicas (ou estratigráficas), o que significa que camadas depositadas em uma mesma época geológica poderão ser mais bem representadas durante a etapa de análise de continuidade espacial realizada

na geoestatística. Deutsch (2002) sugere algumas aproximações que podem ser feitas para transformar coordenadas cartesianas em coordenadas estratigráficas.

1.4. Estado da arte

No espaço real, estamos acostumados a adicionar vetores, multiplicá-los pelo valor de uma constante ou escalar, observar propriedades como ortogonalidade, ou calcular a distância entre dois pontos. Tudo isto, e muito mais, é possível, porque o espaço real é um espaço linear vetorial com uma estrutura métrica euclidiana (Pawlowsky et al., 2010), mas esta geometria não é uma geometria adequada para os dados composicionais (CODA). O espaço ao qual pertencem esses dados composicionais é designado por Aitchison (1986) como *D-simplex*.

Um exemplo notável sucede quando temos duas composições, [5,65,30] e [10,60,30] cuja soma em cada composição é 100. Note as primeiras medidas nas duas amostras, 5 e 10, respectivamente e as segundas, 65 e 60. Intuitivamente, a distância euclidiana entre o primeira e a segunda componente é a mesma (5 unidades, i.e 5-10 ou 60-65) Mas, no primeiro caso, a proporção da primeira componente é dobrada (de 5 para 10), enquanto que para a segunda o aumento relativo é perto de 10% (de 60 para 65). O uso da diferença relativa parece ser mais adequado para descrever a variabilidade de composição (adaptado de Pawlowsky et al. 2010)

Assim, os dados composicionais (CODA) surgem do fato de que eles representam partes de um todo e, portanto, eles transmitem apenas informação relativa e não absoluta. Eles são sempre positivos e normalmente restritos a uma soma constante (Barceló-Vidal et al 2001, Eynatten et al 2002, Egozcue & Pawlowsky 2005, Thió-Henestrosa & Martín-Fernández 2005, Pawlowsky & Egozcue 2006, Egozcue & Pawlowsky 2006, Filzmoser & Hron 2008, Filzmoser et al 2009).

Diversas possibilidades de transformação de dados composicionais têm sido introduzidas na literatura, sendo mais amplamente utilizadas na família das transformações as Razões Logarítmicas Aditivas (alr) e as Razões Logarítmicas Centradas (clr) por Aitchison (1986). Cada uma delas possui vantagens e desvantagens que serão discutidas posteriormente na revisão bibliográfica. Recentemente, Egozcue & Pawlowsky (2003) propuseram uma nova transformação de dados com o objetivo de ter uma base ortonormal no espaço de composição (simplex), chamada razão logarítmica isométrica (ilr) que é o foco principal desta dissertação.

Boezio (2010) menciona que o princípio das transformações de razões logarítmicas está baseado no fato de que existe uma correspondência um a um entre os vetores composicionais e os vetores das razões logarítmicas associadas. A vantagem dessa transformação é que retira o problema de um espaço amostral restringido (soma de uma constante) para um espaço amostral não restringido (D-simplex), multivariado, o espaço real.

Bragulat et al. (2002) realizaram um estudo experimental em um depósito de Bauxita na Hungria usando como método interpolador a Krigagem. Eles fizeram uma comparação das estimativas entre quatro métodos diferentes: Krigagem Ordinária (OK), Krigagem das transformações de Razões Logarítmicas Aditivas (alr), Krigagem das transformações de Razões Logarítmicas Centradas (clr) e a Krigagem da Transformada Rápida de Fourier (FFT). Ao final, concluíram que os resultados obtidos em cada uma delas, em termos de qualidade das estimativas são muito similares e que a Krigagem (FFT) não leva em consideração o fato dos dados serem composicionais.

Bragulat & Sala (2003) fizeram outro estudo similar no mesmo depósito de Bauxita, agora comparando somente a Krigagem (alr), Krigagem (clr) e Krigagem das transformações de Razões Logarítmicas Isométricas (ilr) usando a Krigagem de matrizes de covariância calculadas na transformada (FFT). As conclusões obtidas por eles é que em termos de estimativas, a Krigagem (alr) apresenta uma estatística básica mais parecida com a estatística básica dos

dados originais, no entanto, para as demais os resultados dos demais métodos de estimativa foram muito próximos.

Boezio, Costa & Koppe (2012) fazem uma comparação entre a Cokrigagem direta e a Cokrigagem de Razões Logarítmicas Aditivas (*alr*) num depósito de ferro no Brasil. Eles usam a Cokrigagem (*alr*) como metodologia alternativa na estimativa de depósitos multivariados, onde se tem múltiplas variáveis correlacionadas com somas constantes. Os resultados obtidos na Cokrigagem (*alr*) se mostraram adequados, pois além de ter estimativas compatíveis com os dados originais, respeitou-se o fechamento químico e mássico das variáveis analisadas. Além disso, eles fazem ênfases à reprodução da média geométrica dos dados estimados em relação às amostras, mas, na indústria mineral isto carece de sentido, já que esta medida de tendência central não é apropriada para medir as características físicas do depósito e se a média aritmética.

Boezio, Abichequer & Costa (2012) partindo do fato da dificuldade na modelagem da Cokrigagem de múltiplas variáveis correlacionadas, fizeram uma combinação entre duas metodologias. Primeiro, utilizaram as transformações (*alr*) para os dados composicionais e depois descorrelacionaram cada uma delas pelo método conhecido como Min/Max Autocorrelation Factors (MAF) que descorrelaciona as múltiplas variáveis para poder modelar e estimar cada uma delas independentemente. Eles compararam os resultados obtidos na Cokrigagem Ordinária direta e concluíram que tanto a Cokrigagem (*alr*) como a Krigagem (MAF) tiveram melhores resultados enquanto à simplificação na modelagem da continuidade espacial e o fechamento dos balanços de massa (granulométricos e das espécies químicas), além de ter havido bons resultados nas estimativas.

Outros estudos feitos por Walvoort & Gruijter (2001) utilizam a Krigagem composicional como uma extensão direta da Krigagem Ordinária para avaliar qualidade das estimativas em dois estudos de caso (Wesepe e Walker Lake). Eles compararam os resultados obtidos com a Krigagem das transformações (*alr*) e concluíram que no primeiro estudo, a Krigagem Composicional produziu

previsões significativamente mais precisas do que a Krigagem (alr), mas no segundo caso essas estimativas foram muito similares.

Diferente do estudo desta dissertação, Bragulata et al. (2002 e 2003) não buscam o fechamento mássico granulométrico e das espécies químicas das variáveis estudadas, apenas analisam a qualidade das estimativas em cada uma das transformações utilizadas (alr, clr, FFT). Por sua parte, Boezio et al. (2012) utilizam a Cokrigagem (alr) como método interpolador e, embora seja claro a soma fechada dos dados estimados, o nosso trabalho utiliza a Krigagem das transformações (ilr) que preservam melhor a métrica (neste caso específico representa as mesmas distâncias do espaço euclidiano) do espaço amostral dos dados composicionais, além de não enfrentar aqueles problemas mencionados por Boezio (2010) em relação à Cokrigagem.

1.5. Problema

Trabalhar com depósitos multi-elementais, como é o caso em depósitos de bauxita, e, além disso, ter que lidar com dados composicionais, não necessariamente correlacionados fisicamente (correlação espúrias, Pearson, 1897), combinadas ao uso de métodos tradicionais podem levar ao não fechamento do balanço químico e mássico das múltiplas variáveis analisadas. Desse modo, um pós-processamento, normalmente, deve ser realizado para superar esta inconsistência. Por exemplo, com distribuição do erro de fechamento entre as diferentes faixas granulométricas para cada um dos elementos estimados.

As metodologias clássicas, como é o caso da krigagem e a cokrigagem ordinária, podem ser apropriadas para obter a melhor estimativa local do ponto de vista teórico. Mas, cada variável é estimada separadamente (no caso OK) com seus respectivos parâmetros de medidas de continuidade espacial, o que leva a obter pesos diferentes para cada atributo e em contraposto gera problemas para o fechamento final das estimativas.

Em relação a outros métodos mais robustos como a Cokrigagem, leva em consideração a correlação entre as variáveis, mas não garante o fechamento das espécies químicas. Ainda, apresenta o defeito de utilizar modelos de continuidade espacial não muito aderentes à realidade física do depósito, aumentando cada vez mais sua complexidade quando aumenta o número de variáveis estudadas.

1.6. Meta da dissertação

A meta dessa dissertação visa adaptar uma metodologia estatística aplicada em um depósito mineral de bauxita, fazendo uso das ferramentas geoestatísticas que permitem interpolar cada variável separadamente, garantindo o fechamento nos balanços químicos e mássicos após estimativas ou simulações.

1.7. Objetivos específicos da dissertação

Para atingir a meta proposta anteriormente, estabeleceram-se os seguintes objetivos:

- i. Estudar o uso das transformações logarítmicas isométricas (*ilr*) nos dados originais e, checar a reprodução dos balanços após o processo de estimativas.
- ii. Comparar os resultados obtidos pelos métodos tradicionais geoestatísticos e pelos métodos tradicionais utilizando as transformações (*ilr*).
- iii. Verificar a possível existência de um viés significativo na média aritmética das estimativas, comparando o E-type das simulações com a krigagem ordinária.
- iv. Investigar os resultados por meio de validações em cada uma das metodologias estudadas nesta dissertação.

1.8. Metodologia

Com base aos objetivos propostos anteriormente, foi criado um fluxograma para representar cada uma das etapas do processo a ser utilizado nesta dissertação (ver figura 1.2).

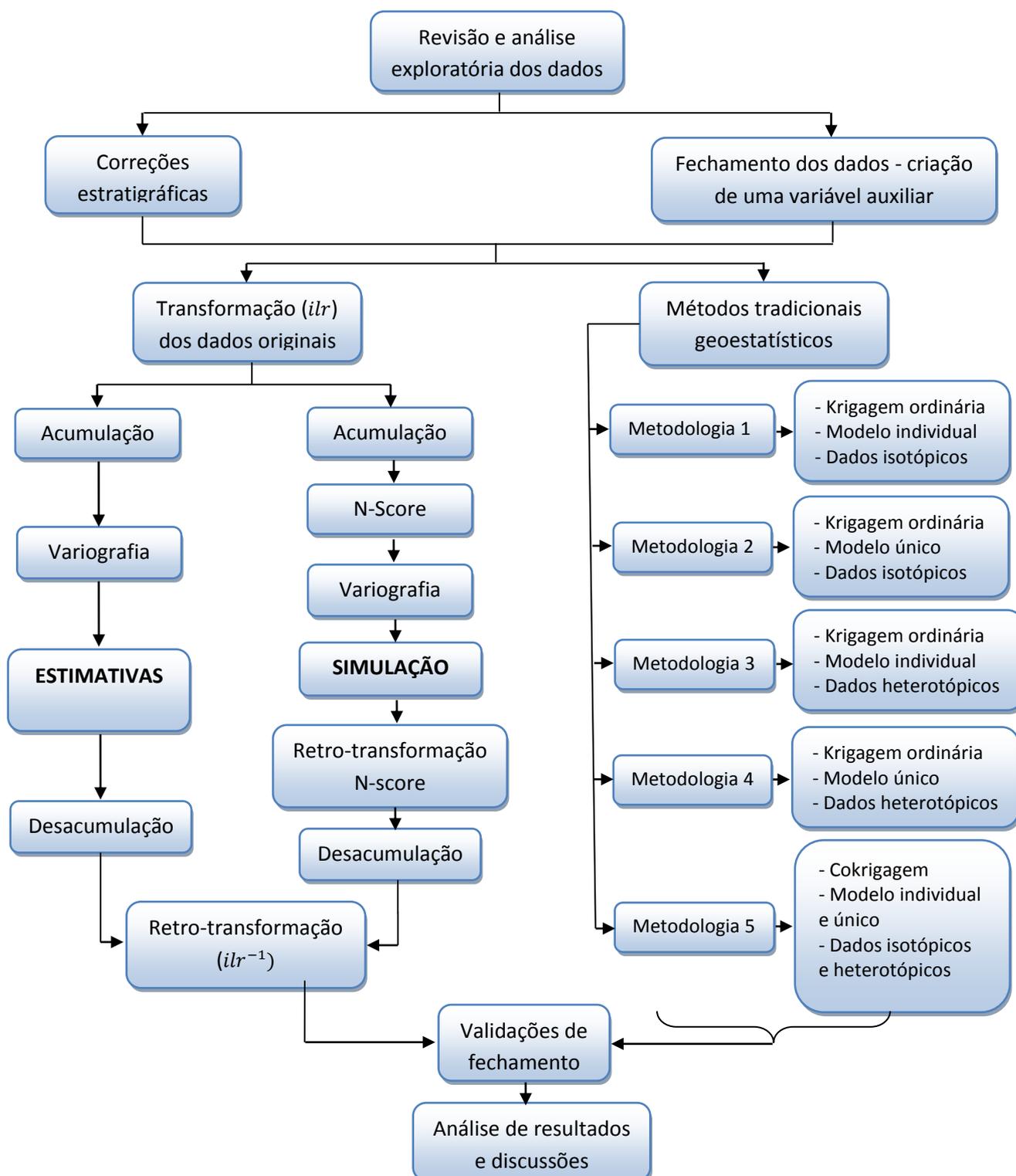


Figura 1.2. Metodologia de trabalho.

1.9. Organização dessa dissertação

A presente dissertação foi estruturada em relação a uma série de capítulos descritos da seguinte maneira:

O Capítulo 1 contém com uma breve introdução da dissertação, seguido das metodologias de interpolação geoestatísticas, estado da arte, o problema, a meta, os objetivos específicos e algumas considerações do tipo de depósito em estudo.

O Capítulo 2 introduz uma revisão bibliográfica das metodologias clássicas usadas na geoestatística, as medidas de continuidade espacial, os métodos clássicos de correções de coordenadas para depósitos estratiformes e a influencia do suporte amostral. Além disso, é apresentada a teoria básica dos dados composicionais proposta por Aitchison (1986) fazendo uso das transformações de razões logarítmicas isométricas (*ilr*) que é nossa metodologia proposta.

O Capítulo 3 apresenta as metodologias clássicas da geoestatística aplicada a um estudo de caso real em um depósito de Bauxita situada na cidade de Porto Trombetas - Pará da empresa MRN.

O Capítulo 4 corresponde ao desenvolvimento da metodologia proposta aplicado ao mesmo depósito. Além disso, se apresenta um análise previa dos resultados e uma breve comparação com a utilização dos métodos tradicionais utilizados no capítulo anterior.

O Capítulo 5 encerra com as conclusões obtidas ao longo do trabalho e as recomendações para trabalhos futuros.

Capítulo 2

Conceitos Gerais

Ao longo desta dissertação, utilizaremos uma série de conceitos que são necessários para o desenvolvimento do tipo de depósito em estudo, por exemplo, a utilização de coordenadas estratigráficas e a necessidade de fazer um processo de acumulação. Outros temas importantes vêm da teoria da geoestatística clássica e seus métodos interpoladores necessários para o processo de estimativa. Por último, serão definidos os aspectos teóricos da metodologia proposta referente aos dados composicionais (CODA) e a sua aplicabilidade ao longo deste trabalho.

2.1. Transformação de coordenadas

Uma das características principais dos depósitos sedimentares (carvão, alguns de fosfato, bauxita, etc.) é que esses depósitos são constituídos por camadas. Essas camadas podem passar por vários eventos geológicos posteriores que podem criar dobramentos, bacias ou erosões.

Um dos problemas mais comuns que ocorre neste tipo de depósitos é quando analisamos a continuidade espacial dos dados, já que se pode estar incorrendo em um erro combinando amostras de níveis estratigráficos diferentes. Por exemplo, duas amostras podem estar no mesmo nível topográfico ou cartesiano, mais em diferente nível estratigráfico ou geológico. Note que, na figura 2.1(a), as coordenadas cartesianas da amostra 1 e 4 se encontram no topo e a base respectivamente de uma camada, porém em uma mesma cota z - topográfica. No entanto, a análise da continuidade espacial e as estimativas deveriam ser feitas com amostras depositadas no mesmo horizonte geológico; portanto, a amostra 1 deveria ter mais conectividade espacial com a amostra 2 do que com a amostra 4 (ver figura 2.1b).

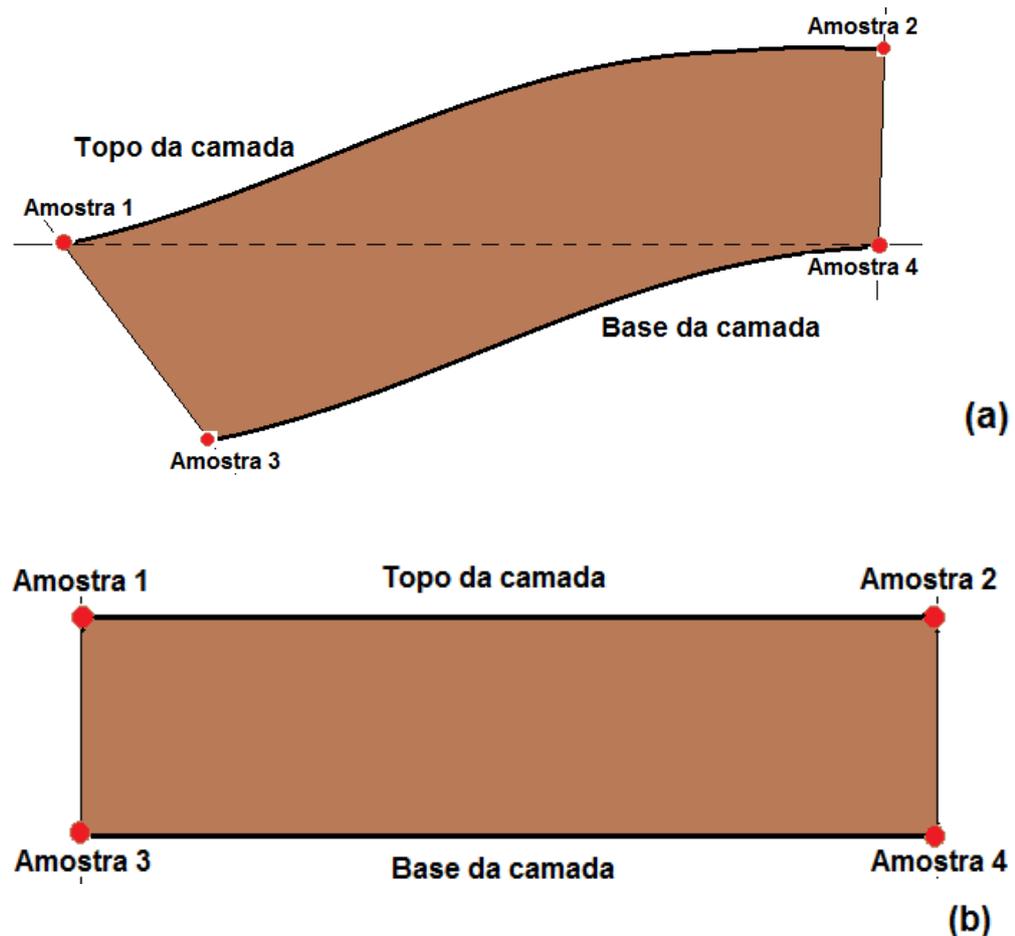


Figura 2.1. Interpretação de coordenadas entre diferentes amostras, (a) coordenadas cartesianas (b) coordenadas estratigráficas.

O modelamento geoestatístico está baseado na dependência espacial de amostras que foram depositadas no mesmo período de tempo e essa dependência (temporal/espacial) está ligada ao tipo de depósito que está sendo estudado (Koppe, 2005). Deutsch (2002) comenta sobre algumas aproximações que podem ser utilizadas para transformar coordenadas cartesianas (topográficas) em coordenadas geológicas (estratigráficas):

- Proporcional: A camada está conforme ao topo e a base existente. A camada pode variar em espessura devido a fatores geológicos, mas, a superfície da capa e a lapa são mantidas (ver figura 2.2).

- Truncada: A camada está conforme a base existente, mais, o topo foi erodido devido a processos geológicos (ver figura 2.2).
- Sobreposição: A camada está conforme ao topo existente, mais, a base foi depositada devido a processos geológicos (ver figura 2.2).
- Combinação: o topo e a base foram erodidos e depositados respectivamente devido a processos geológicos (ver figura 2.2).

As linhas contínuas representam o topo (capa) e a base (lapa) da camada e as linhas tracejadas são interpretações geológicas obtidas a partir das informações amostrais.

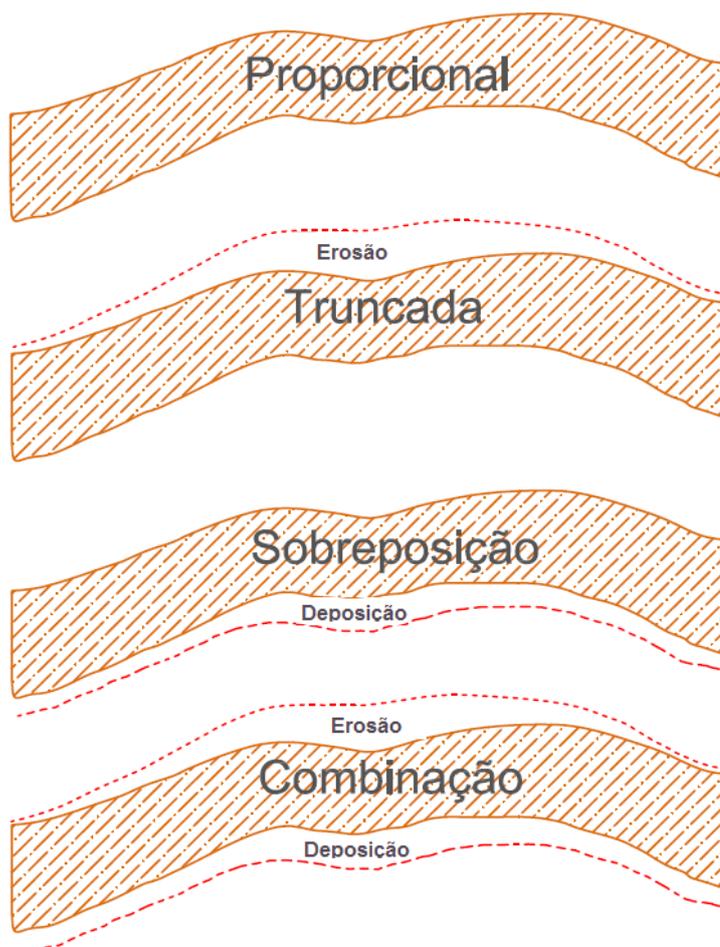


Figura 2.2. Interpretações sobre eventos ou processos geológicos, imagem modificada de Deutsch (2002).

A transformação das novas coordenadas estratigráficas pode ser calculada da seguinte maneira:

$$Z_{est} = \frac{z(i) - z(i)_a}{z(i)_b - z(i)_a} (E) \quad i = 1, \dots, n \quad (2.1)$$

Onde:

- Z_{est} = Nova coordenada vertical.
- $z(i)$ = Coordenada vertical original.
- $z(i)_a$ = Coordenada vertical da linha de erosão (topo).
- $z(i)_b$ = Coordenada vertical da linha de deposição (base).
- E = Espessura média da camada.
- n = Número de amostras.

Essa nova transformação irá gerar uma imagem plana, já que unicamente serão modificadas as coordenadas verticais e não as horizontais devido à dependência espacial conforme foi expresso anteriormente (ver figura 2.3).

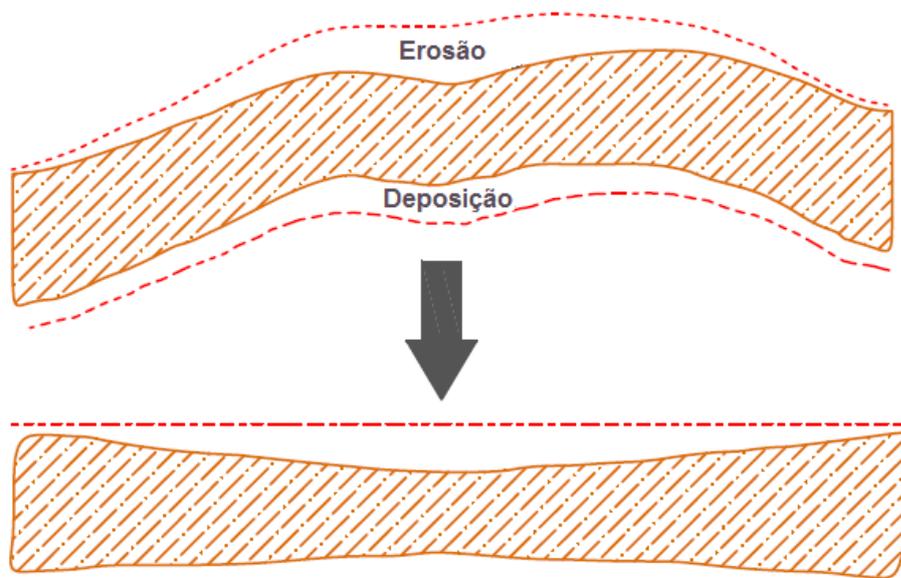


Figura 2.3. Correção estratigráfica de uma camada, imagem modificada de Deutsch (2002).

2.2. Acumulação

Outra característica importante para ser considerada na estimativa dos depósitos minerais é o suporte amostral na qual são coletadas as amostras. Geralmente, o banco de dados de uma campanha de sondagem traz consigo uma grande variedade de suportes (comprimento ou diâmetro em que são analisadas as amostras). Do ponto de vista estatístico, isto pode influenciar nos resultados já que os teores não carregam as mesmas informações (teores analisados em diferente suporte com possível viés e diferentes variâncias).

Na figura 2.4, podemos ver um exemplo da influência do suporte amostral. Imagine que foram analisadas três amostras ao longo de uma camada cujos teores são 12%, 13%, 18% cada uma com as espessuras 0.67m, 0.67m e 1.34m, respectivamente. Se calcularmos a média aritmética dos três teores sem considerar o seu suporte teremos uma média de 14.33%. Agora, se consideramos o suporte amostral obteríamos uma média ponderada de 15.25%, que nada mais é a somatória da multiplicação do teor vezes a espessura dividida pela somatória da espessura. A média ponderada é mais correta para ser utilizada neste tipo de casos, já que considera o suporte amostral. Por último, se comparamos o erro relativo entre as duas médias obteríamos um valor de aproximadamente 6%.

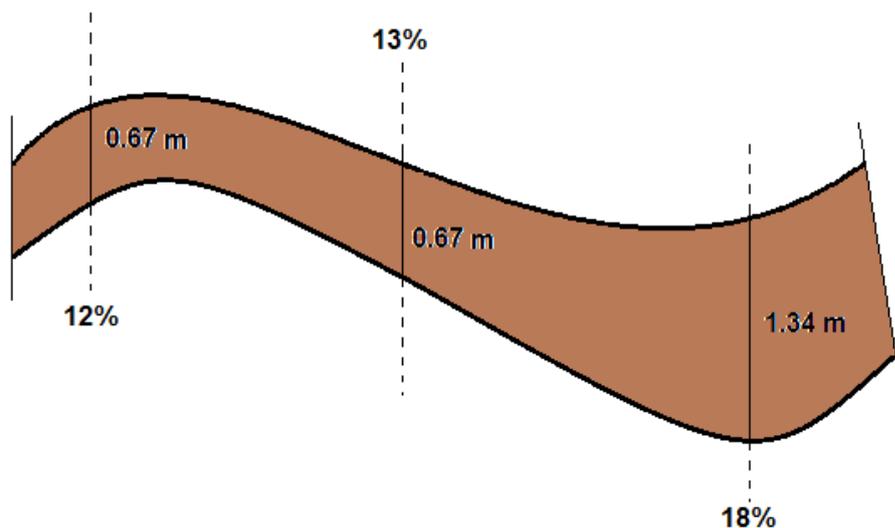


Figura 2.4. Influência do suporte amostral na média.

Acumulação é o processo de multiplicar o valor de um atributo por seu determinado suporte, antes de proceder com as estimativas, para não misturar dados que estão em suportes distintos. Por exemplo, quando se multiplica o teor de cinzas de depósitos de carvão pela espessura da camada correspondente naquele ponto amostral ou quando se multiplica o teor de ferro em determinada faixa granulométrica pela respectiva fração (massa) dessa faixa no ponto amostral. Ao final, após as estimativas dos atributos acumulados pelo seu suporte amostral, faz-se o processo inverso ou desacumulação. Consiste na obtenção do atributo de interesse através da razão da estimativa acumulada pela estimativa do suporte amostral no mesmo local (Vieira, 2011).

O principal problema no uso de variáveis acumuladas é a possibilidade de aparecer valores extremos fora do limite mínimo e máximo dos dados originais, produto do processo de desacumulação. Geralmente, esses valores indesejáveis aparecem quando as estimativas da variável acumulada possuem valores altos e, os valores estimados do acumulador são relativamente baixos ou vice-versa, já que o processo de desacumulação implica a divisão desses dois valores, produzindo assim um valor inconsistente.

Para solucionar ou diminuir esse problema, Dagbert (2001) sugere que ao fazer o processo de estimativa, utilize-se o mesmo variograma do acumulador nas variáveis acumuladas, outorgando assim os mesmos pesos para ambas as partes. Assim, espera-se ao ter-se uma estimativa alta na variável acumulada, provavelmente resultará também uma estimativa alta do acumulador no mesmo local devido à mesma distribuição dos pesos, que levaria a desacumulação não gerar muitos valores expúrios. O problema é utilizar-se o mesmo modelo variográfico para diferentes variáveis e que em alguns casos esse modelo não seja representativos das características espaciais de todas as variáveis (acumulação e ponderador).

Outra solução, utilizada por Boezio (2010) é fazer um pós-processamento dos blocos com valores expúrios, consistindo em substituir os blocos que estão fora dos intervalos válidos, por valores obtidos por outra metodologia de interpolação que forneça estimativas não tendenciosas ou

utilizem um modelo variográfico arbitrário, como é o caso da média local ou inverso ao quadrado da distância.

Em alguns casos especiais, geralmente quando o número de blocos fora do intervalo mínimo e máximo é bem pequeno e os valores não são muito além dos valores mínimos e máximos (menos de 1% do total de blocos do depósito) é feito um truncamento desses valores para o valor mínimo e máximo dos dados originais, já que eles não representam uma quantidade significativa do depósito.

2.3. Continuidade espacial

Um dos principais aspectos que distinguem os dados da ciência da terra da maioria dos outros dados é a disposição espacial na qual eles estão posicionados. Dois dados próximos são mais propensos a terem valores semelhantes que dois dados que estão distantes. Em um depósito mineral, os valores baixos tendem a estar perto de outros valores baixos e valores altos tendem a estar perto de outros valores mais altos (Isaaks & Srivastava 1989).

A continuidade espacial mede a correlação que existe entre valores ou dados de um ou vários atributos em função da distância espacial entre eles. As ferramentas que são utilizadas para descrever a relação linear entre duas variáveis são o coeficiente de correlação e covariância, e podem ser também utilizadas para descrever a relação entre o valor de uma variável e o valor da mesma em outro local separado por uma distância h . A melhor forma de observar isto é mediante os h -scatterplots que são utilizados para descrever a variabilidade da correlação em diversas distâncias e direções.

Existe uma grande variedade de ferramentas utilizadas na geoestatística para medir a continuidade espacial dos dados. Nessa dissertação, definiremos as mais comuns, sendo o correlograma tipo não ergódico (Srivastava, 1987) o escolhido nesse trabalho. Outras medidas de continuidade espacial podem ser visitadas no Isaaks & Srivastava (1989), Goovaerts (1997) & Deutsch & Journel (1998).

2.3.1. Variograma: O variograma ou semi-variograma $\gamma(h)$ é metade da diferença quadrática média entre os valores de dados pareados aproximadamente pelo vetor h . É a relação entre o momento de inércia de um h -scatterplot e h . O $\gamma(h)$ é calculado da seguinte maneira:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2 \quad (2.2)$$

Onde:

- $\gamma(h)$ = variograma de amostras separadas pelo vetor h ;
- $N(h)$ = número de pares separados aproximadamente pelo vetor h ;
- h = distância média entre amostras;
- x_i = é o valor da variável;
- $x_i + h$ = é o valor da variável na localização i mais o vetor h .

2.3.2. Covariância: Também conhecida como *covariância não ergódica* (Srivastava, 1987), é a relação entre a covariância de um h -scatterplot e os respectivos valores de h . A função covariância $C(h)$ é calculada da seguinte maneira:

$$C(h) = \frac{1}{N(h)} \sum_{i,j=1}^{N(h)} z(x_i) * z(x_i + h) - m_{-h} * m_{+h} \quad (2.3)$$

$$m_{-h} = \frac{1}{N(h)} \sum_{i=1}^{N(h)} z(x_i) \quad (2.4)$$

$$m_{+h} = \frac{1}{N(h)} \sum_{j=1}^{N(h)} z(x_j + h) \quad (2.5)$$

Onde:

- $C(h)$ = covariância de amostras separadas pelo vetor h ;
- $N(h)$ = número de pares separados aproximadamente pelo vetor h ;
- x_i = é o valor da variável;
- $x_i + h$ = é o valor da variável na localização i mais o vetor h ;
- m_{-h} = é a média de todos os valores dos dados cuja localização se encontra $-h$ de distância de outros dados;
- m_{+h} = é a média de todos os valores dos dados cuja localização se encontra $+h$ de distância de outros dados.

2.3.3. Correlograma: Também conhecido como *correlograma não ergódico* (Srivastava, 1987) é a função covariância estandardizada pelos respectivos desvios padrões. É a relação entre o coeficiente de correlação de um h -scatterplot e os respectivos valores de h , dado pela seguinte equação:

$$\rho(h) = \frac{C(h)}{\sigma_{-h} * \sigma_{+h}} \quad (2.6)$$

$$\sigma_{-h}^2 = \frac{1}{N(h)} \sum_{i=1}^{N(h)} z(x_i)^2 - m_{-h}^2 \quad (2.7)$$

$$\sigma_{+h}^2 = \frac{1}{N(h)} \sum_{j=1}^{N(h)} z(x_j + h)^2 - m_{+h}^2 \quad (2.8)$$

Onde:

- $\rho(h)$ = correlograma de amostras separadas pelo vetor h ;
- $N(h)$ = número de pares separados aproximadamente pelo vetor h ;
- x_i = é o valor da variável;
- $x_i + h$ = é o valor da variável no local i mais o vetor h ;
- σ_{-h} = é o desvio padrão de todos os valores dos dados cuja localização se encontra $-h$ de distância de outros dados;
- σ_{+h} = é o desvio padrão de todos os valores dos dados cuja localização se encontra $+h$ de distância de outros dados.

2.4. Estimativa

Os métodos geoestatísticos clássicos surgiram a partir dos anos sessenta com a idéia de prever valores em locais não amostrados baseados na conectividade espacial das amostras analisadas. Sichel (1947, 1949) observou a natureza assimétrica da distribuição dos dados em uma mina de ouro sul-africana semelhante a uma distribuição de probabilidade lognormal e, desenvolveu as fórmulas básicas para essa distribuição. Isto permitiu uma primeira estimativa das reservas, mas sobre o pressuposto de que as medidas eram independentes, em clara contradição com a experiência de que há zonas mais ricas do que outras.

Uma primeira abordagem para resolver esse problema foi dada pelo Eng. de minas Daniel Krige (1951) que propôs uma variante do método conhecido como média móvel, equivalente ao método da krigagem simples (SK), que é um dos métodos de estimativa lineares no espaço com mais elevada qualidade teórica. Uma formulação rigorosa e solução para o problema de previsão veio das mãos de Matheron (1962) na Escola de Minas de Paris. Nos anos subseqüentes, a teoria foi aperfeiçoada por alargar a sua gama de validade e reduzindo os pressupostos necessários. Técnicas desenvolvidas na mineração dentro da geoestatística foram disseminadas para outros campos, como hidrologia, física do solo, ciências da terra e mais recentemente, monitoramento ambiental e processamento de imagens de satélite.

A krigagem é um termo genérico utilizado em uma grande variedade de métodos de estimativa (pontual ou em bloco) que dependem da minimização do erro, geralmente através de um procedimento de mínimos quadrados. O termo foi incluído por G. Matheron e P. Carlier ao honrar ao D. Krige, cujo trabalho empírico sobre a estimativa de reserva em minas de ouro da África do Sul foi mais tarde desenvolvido pela teoria geoestatística desenvolvido por Matheron (Sinclair & Blackwell, 2002).

Existe uma grande variedade de métodos geoestatísticos associados à krigagem, entre eles podemos encontrar: krigagem simples (SK), krigagem ordinária (OK), krigagem dos indicadores (IK), krigagem universal (UK), krigagem probabilístico (PK), entre outros. Todas essas técnicas dependem de um conceito similar: a autocorrelação de uma variável regionalizada pode ser modelada por uma função matemática inferida de uma realização (dados) da variável regionalizada e pode ser utilizada para auxiliar a estimativa (Sinclair & Blackwell, 2002).

Freqüentemente, é necessário caracterizar as correlações espaciais de vários elementos de forma simultânea. Infelizmente, os métodos geoestatísticos univariados ignoram este tipo de informação, já que eles se caracterizam por produzir as melhores estimativas individualmente. Os métodos geoestatísticos multivariados surgem para captar essa dependência espacial entre vários elementos. Esta informação pode ser extraída a partir da matriz de variância-covariância ou matriz de correlação.

Krigagem é um termo utilizado para estimar valores desconhecidos de uma variável a partir dos valores conhecidos de ela mesma. Algumas vezes, desejamos estimar uma variável em função de outra variável, geralmente correlacionada. Essas variáveis, chamadas de variáveis secundárias, podem fornecer informações úteis para estimar a variável de interesse e essa informação pode ser levada em consideração para melhorar as estimativas das variáveis de interesse ou primária. Isto exige uma generalização multivariada da krigagem que é conhecida como cokrigagem (Chilès & Delfiner 1999).

2.4.1. Krigagem ordinária (OK): Isaaks e Srivastava (1989) definem a krigagem ordinária como o “melhor estimador linear não tendencioso” (BLUE, por suas siglas 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.

Uma característica fundamental da krigagem ordinária, que difere de outros tipos de krigagem, são as flutuações locais sobre a média, limitando a hipóteses de estacionaridade a uma vizinhança local ao longo da área do depósito. Assim, não necessita o conhecimento prévio da média global para realizar as estimativas locais.

A expressão que define o cálculo do valor estimado por krigagem ordinária é dada pela seguinte equação (Deutsch & Journel, 1998):

$$Z_{OK}^*(u) = \sum_{\alpha=1}^{n(u)} \lambda_{\alpha}^{OK}(u) Z(u_{\alpha}) \quad (2.9)$$

Onde:

- $Z_{OK}^*(u)$ = é o valor da estimativa;
- $Z(u_{\alpha})$ = é o valor da amostra α no local u ;
- $n(u)$ = número de amostras para estimativa no local u ;
- $\lambda_{\alpha}^{OK}(u)$ = pesos da krigagem ordinária calculados da seguinte maneira:

$$\sum_{\beta=1}^{n(u)} \lambda_{\beta}^{OK}(u) C(u_{\beta} - u_{\alpha}) + \mu_m^{OK}(u) = C(u - u_{\alpha}), \quad \alpha = 1, \dots, n(u) \quad (2.10)$$

$$\sum_{\beta=1}^{n(u)} \lambda_{\beta}^{OK}(u) = 1 \quad (2.11)$$

- $\mu_m^{OK}(u)$ = parâmetro lagrangeano.
- $C(u_{\alpha} - u_{\beta})$ = covariância entre amostras utilizadas na estimativa.
- $C(u - u_{\alpha})$ = covariância entre a amostra utilizada na estimativa e o ponto a ser estimado.

A variância de krigagem ordinária é dada por:

$$\sigma_{OK}^2(u) = C(0) - \sum_{\alpha=1}^{n(u)} \lambda_{\alpha}^{OK}(u) C(u_{\alpha} - u) - \mu_m^{OK}(u) \quad (2.12)$$

2.4.2. Cokrigagem Ordinária (CKO): A cokrigagem ordinária é uma metodologia que prove estimativas não tendenciosas minimizando a variância do erro e melhora as estimativas de uma variável chamada primária, ao incorporar informações secundárias provenientes de outras variáveis fortemente correlacionadas com a primária e, preferencialmente, estejam mais densamente amostradas, formando um banco de dados heterotópicos.

Neste caso, a média é considerada estacionária, somente numa vizinhança local e não na totalidade da área de estudo. O estimador de CKO para a variável de interesse Z_1 no local u é dado pela seguinte expressão:

$$Z_{CKO1}^*(u) = \sum_{\alpha1=1}^{n1} \lambda_{\alpha1}^{CKO} * Z_1(u_{\alpha1}) + \sum_{\alpha2=1}^{n2} \lambda_{\alpha2}^{CKO} * [Z_2(u_{\alpha2}) - m_2 + m_1] \quad (2.13)$$

Onde:

- $Z_{CKO1}^*(u)$ = valor estimado da variável primária;
- Z_1 = variável primária;
- Z_2 = variável secundária;
- m_1 e m_2 = médias desagrupadas de Z_1 e Z_2 respectivamente;
- Os pesos da cokrigagem ordinária são dados pelas seguintes equações:

$$\sum_{\alpha1=1}^{n1(u)} \lambda_{\alpha1}^{CKO} C_{11}(u_{\alpha1} - u_{\beta1}) + \sum_{\alpha2=1}^{n2(u)} \lambda_{\alpha2}^{CKO} C_{12}(u_{\alpha1} - u_{\alpha2}) + \mu(u) = C_{11}(u_{\alpha1} - u) \quad (2.14)$$

$$\sum_{\alpha1=1}^{n1(u)} \lambda_{\alpha1}^{CKO} C_{21}(u_{\alpha2} - u_{\alpha1}) + \sum_{\alpha2=1}^{n2(u)} \lambda_{\alpha2}^{CKO} C_{22}(u_{\alpha2} - u_{\beta2}) + \mu(u) = C_{21}(u_{\alpha2} - u) \quad (2.15)$$

$$\sum_{\alpha1=1}^{n1} \lambda_{\alpha1}^{CKO} = 1 \quad , \quad \sum_{\alpha2=1}^{n2} \lambda_{\alpha2}^{CKO} = 0 \quad (2.16)$$

A variância de cokrigagem ordinária é dada pela seguinte equação.

$$\sigma_{CKO}^2(u) = C_{11}(0) - \mu_1(u) - \sum_{\alpha 1=1}^{n1(u)} \lambda_{\alpha 1}^{CKO} * C_{11}(u_{\alpha 1} - u) - \sum_{\alpha 2=1}^{n2(u)} \lambda_{\alpha 2}^{CKO} * C_{21}(u_{\alpha 2} - u) \quad (2.17)$$

2.5. Simulação seqüencial

Conforme explica Blackwell & Sinclair (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 explica 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 2.5 mostra essas

flutuações (variogramas) correspondentes a 20 realizações (linhas azuis) em relação ao modelo espacial de referência (linha preta).

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. Em um processo de interpolação, estamos preocupados pela minimização da variância do erro; enquanto, nas simulações nos preocupamos pela reprodução da dispersão da variância dos dados originais (Blackwell & Sinclair, 2002).

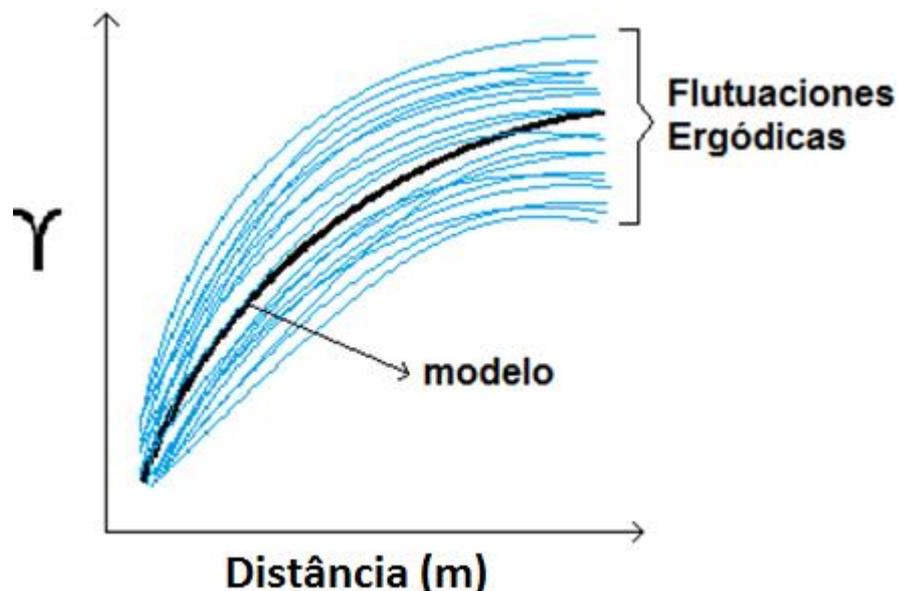


Figura 2.5. Flutuações ergódicas nos variogramas. Imagem modificada do (Deutsch & Journel, 1998).

2.6. Simulação seqüencial gaussiana

A simulação seqüencial gaussiana constrói basicamente uma função de distribuição Gaussiana acumulada para cada nó simulado. Essa distribuição está baseada em um modelo multiGaussiano próprio de cada conjunto de dados.

Conforme comenta Koppe (2009), o modelo multiGaussiano assume que para um atributo com distribuição de probabilidade normal padrão (média zero e desvio padrão um) qualquer combinação de valores desse atributo deve

resultar em novo valor com uma distribuição normal. A distribuição acumulada de possíveis valores do atributo em cada local u é normal e definida por parâmetros condicionados aos dados: esperança matemática e variância. Segundo Journel & Huijbregts (1978), esses parâmetros são iguais à estimativa por krigagem simples e a variância da krigagem simples respectivamente, considerando que os dados vizinhos são utilizados na interpolação.

Deutsch & Journel (1998) mencionam que o algoritmo da simulação seqüencial gaussiana utiliza, para cada local simulado, a krigagem simples como método de construção da distribuição normal acumulada, dada pela seguinte equação:

$$[Z_{SK}^*(u) - m(u)] = \sum_{\alpha=1}^{n(u)} \lambda_{\alpha}^{SK}(u) * [Z(u_{\alpha}) - m(u_{\alpha})] \quad (2.18)$$

Onde:

- $Z_{SK}^*(u)$ = é o valor da estimativa;
- $m(u)$ = média dos valores de Z no local u ;
- $Z(u_{\alpha})$ = valor da amostra α ;
- λ_{α}^{SK} = pesos da krigagem simples calculados da seguinte maneira:

$$\sum_{\beta=1}^{n(u)} \lambda_{\beta}^{SK}(u) C(u_{\alpha}, u_{\beta}) = C(u_{\alpha}, u_{\beta}), \quad \alpha = 1, \dots, n \quad (2.19)$$

A variância de krigagem simples é dada por:

$$\sigma_{SK}^2(u) = \sigma^2 - \sum_{\alpha=1}^{n(u)} \lambda_{\alpha}^{SK}(u) C(u) \quad (2.20)$$

Deutsch & Journel (1998) mencionam, além do algoritmo utilizado nesse trabalho (SGS), uma série de passos é seguida no processo de simulação seqüencial gaussiana, tal como:

1. Determinar a função de distribuição acumulada (*ccdf*) representativa da área de estudo, que não necessariamente corresponde aos dados originais, já que as amostras podem estar agrupadas;
2. Usar essa *ccdf* (originais ou desagrupados) para transformá-la em uma distribuição normal padrão (*n-score*) com média zero e desvio padrão um (ver Deutsch & Journel, 1998, p. 141);
3. Checar a hipótese de Bi-normalidade sobre os dados transformados (ver Deutsch & Journel, 1998, p. 142);
4. Se o modelo de referência multiGaussiano for considerado apropriado, a simulação seqüencial gaussiana é aplicada nos dados transformados, levando em consideração os seguintes parâmetros:
 - Definir um caminho aleatório, visitando cada ponto do *grid* a ser simulado, selecionando um conjunto de dados condicionantes ao redor de cada ponto a ser simulado, incluindo os dados originais $y(u')$ e os dados previamente simulados $y(u'')$;
 - Determinar em cada ponto (u') simulado os parâmetros (média e variância) da *ccdf* Gaussiana usando SK com o modelo variográfico ajustado aos dados normalizados;
 - Selecionar um valor simulado $y(u'')$ da *ccdf* obtida e adicioná-lo ao conjunto de dados condicionantes;
 - Proceder para o próximo ponto escolhido através de um processo aleatório e repetir os dois últimos passos citados;
 - Repetir o processo até que todos os N pontos sejam simulados;
5. Retro-transformar os dados simulados no espaço Gaussiano em valores simulados da variável original, aplicando a transformação inversa (ver Deutsch & Journel, 1998, p. 145);

Goovaerts (1997) menciona que os algoritmos de simulação seqüencial garantem a reprodução do modelo de continuidade espacial obtido a partir dos

dados originais, devido ao condicionamento dos dados originais e cada distribuição acumulada dada pelos dados previamente simulados.

2.7. Simulação por bandas rotativas

Outro algoritmo de simulação utilizado nesta dissertação foi Bandas Rotativas, 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 2.6), o ângulo entre duas linhas adjacentes é π/n_D .

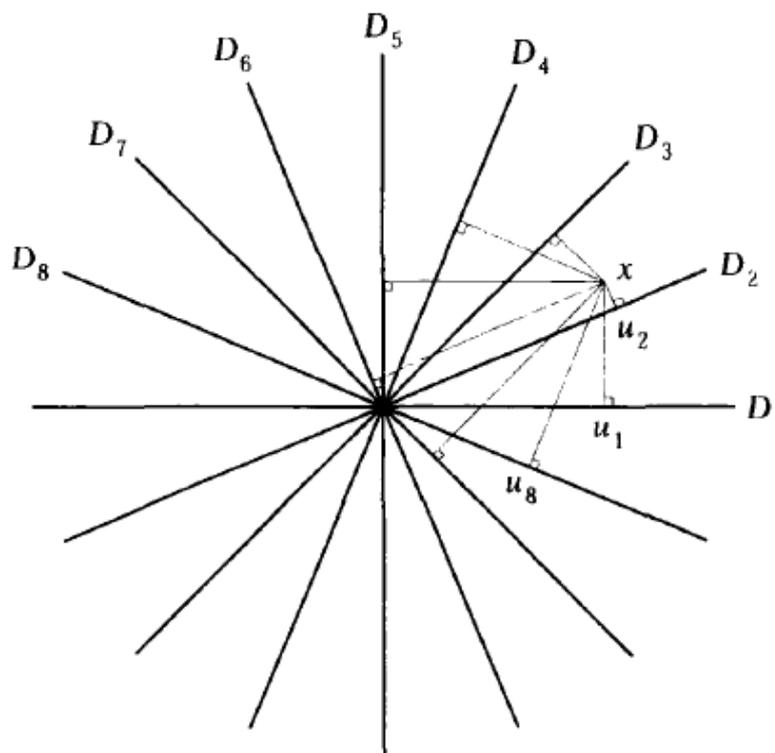


Figura 2.6. Princípio das bandas rotativas em 2D (imagem extraída de Chilès & Delfiner, 1999).

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:

$$s_t = \langle x, u_t \rangle = x \cos\theta_t + y \sin\theta_t \quad (2.21)$$

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

$$Z^*(x) = \frac{1}{\sqrt{n_D}} \sum_{i=1}^{n_D} S_t(s_t) \quad (2.22)$$

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:

$$C_2(x) = \frac{1}{n_D} \sum_{i=1}^{n_D} C_1(\langle h, u_t \rangle) \quad (2.23)$$

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).

2.8. Dados composicionais (CODA)

2.8.1. Conceitos básicos

Ao princípio dos anos 80's, Aitchison definiu os dados composicionais (CODA) como uma série de elementos ou componentes não negativos cuja soma é uma constante (geralmente 1 ou 100%). Uma característica importante neste tipo de dados é que eles carregam informação relativa e não absoluta, sobre os valores das componentes. Portanto, qualquer enunciado sobre uma composição pode ser expresso em termos de quocientes de componentes ou quocientes logarítmicos (Aitchison, 1981, 1982, 1983 e 1984).

Um vetor $x = [x_1, x_2, \dots, x_D]$ é definido como uma composição de D partes se todas as componentes são estritamente números positivos e carregam somente informação relativa. Essa informação está condicionada pela soma de uma constante e representam partes de um todo, por exemplo, unidade (1), percentagem (100%) ou partes por milhão (ppm).

O espaço amostral na qual são definidos os dados composicionais é conhecido como D -simplex, dado pela seguinte equação (Pawlowsky et al., 2010):

$$S^D = \left\{ x = [x_1, x_2, \dots, x_D] \mid x_i > 0, i = 1, 2, \dots, D; \sum_{i=1}^D x_i = k \right\} \quad (2.24)$$

onde as componentes de um vetor S^D são chamadas de *partes* de uma composição. A operação que define o fechamento de uma composição em uma constante k é dada pela seguinte equação (Pawlowsky et al., 2010):

$$C(Z) = \left[\frac{k * Z_1}{\sum_{i=1}^D Z_i}, \frac{k * Z_2}{\sum_{i=1}^D Z_i}, \dots, \frac{k * Z_D}{\sum_{i=1}^D Z_i} \right] \quad (2.25)$$

Onde:

- $C(Z)$ é a operação de fechamento;

- k é a constante de fechamento (geralmente 100%);
- Z_i é o valor da i -ésima amostra.

A restrição dos dados composicionais à soma de uma constante força que pelo menos uma covariância ou coeficiente de correlação entre um par de componentes da composição seja negativa. Se uma correlação tende a ser negativa, então nenhum dos coeficientes de correlação entre os elementos são livres de variar entre -1 e +1. Assim, correlações espúrias (Pearson, 1987) induzidas pelo fato de que a soma dos dados é uma constante (ou estão fechados) e há uma clara tendência para que a correlação seja negativa (Pawlowsky et al., 2006).

2.8.2. Estrutura do espaço vetorial no D-Simplex

As operações básicas de vetores que são utilizadas no espaço euclidiano, como soma de vetores, multiplicação por uma constante, distâncias entre pontos e princípio de ortogonalidade, são também utilizadas no espaço dos dados composicionais (D-simplex). Pawlowsky et al. (2010) define essas operações da seguinte maneira:

- A *perturbação* (análoga à soma no espaço euclidiano) de uma composição $x \in S^D$ vezes uma composição $y \in S^D$ é dada pela seguinte equação:

$$x \oplus y = C[x_1y_1, x_2y_2, \dots, x_Dy_D] \quad (2.26)$$

Onde C é a operação de fechamento definida na equação (2.25).

- A *potenciação* (análoga a potência no espaço euclidiano) de uma composição $x \in S^D$ por uma constante $\alpha \in R$ é dada pela seguinte equação:

$$\alpha \odot x = C[x_1^\alpha, x_2^\alpha, \dots, x_D^\alpha] \quad (2.27)$$

- O *produto interno*, *norma* e *distância* são dadas pelas seguintes equações respectivamente:

$$\langle x, y \rangle_a = \frac{1}{2D} \sum_{i=1}^D \sum_{j=1}^D \ln \frac{x_i}{x_j} \ln \frac{y_i}{y_j} \quad (2.28)$$

$$\|x\|_a = \sqrt{\frac{1}{2D} \sum_{i=1}^D \sum_{j=1}^D \left(\ln \frac{x_i}{x_j} \right)^2} \quad (2.29)$$

$$d_a \langle x, y \rangle = \|x \ominus y\|_a = \sqrt{\frac{1}{2D} \sum_{i=1}^D \sum_{j=1}^D \left(\ln \frac{x_i}{x_j} - \ln \frac{y_i}{y_j} \right)^2} \quad (2.30)$$

As propriedades utilizadas no espaço euclidiano, como comutativa, associativa, distributiva, elemento neutral, entre outros, também são válidas no espaço D-Simplex (ver Pawlowsky et al., 2010, p. 15).

2.8.3. Transformações logarítmicas

Aitchison (1986) definiu duas transformações logarítmicas baseado na teoria de que os dados composicionais carregam informação relativa e não absoluta. Essas transformações são conhecidas como Transformações Logarítmicas Aditivas (*alr*) e Transformações Logarítmicas centradas (*clr*) reproduzidas em Pawlowsky et al. (2010) da seguinte maneira:

1. *Transformações logarítmicas aditivas (alr)*: se x é definido como uma composição de D-partes no simplex S^D , então:

$$alr(x) = \ln \left(\frac{x_1}{x_D}, \frac{x_2}{x_D}, \dots, \frac{x_{D-1}}{x_D} \right) \quad (2.31)$$

onde (\ln) é o logaritmo natural aplicado componente a componente, portanto, para a primeira componente, a transformação alr é definida como $alr(x) = \ln(x_1/x_D)$. Esse novo espaço é conhecido por Aitchison (1986) como espaço *Real*. O processo inverso para trazer novamente o espaço real ao espaço simplex é dado pela seguinte equação:

$$x = C(\exp(alr_1(x), alr_2(x), \dots, alr_{D-1}(x), 0)) \quad (2.32)$$

As vantagens nestas transformações é que primeiramente reduzem o espaço simplex de uma composição de D partes para um espaço real de (D-1) partes. Segundo, preserva as propriedades de perturbação e potencia, ou seja:

$$alr((\alpha \odot x) \oplus (\beta \odot y)) = \alpha \cdot alr(x) + \beta \cdot alr(y) \quad (2.33)$$

para qualquer composição x, y e qualquer constante real α e β . No entanto, o principal inconveniente nas transformações alr é que não preservam as distâncias no espaço real, sendo um isomorfismo e não uma isometria.

2. *Transformações logarítmicas centradas (clr)*: se x é definido como uma composição de D-partes no simplex S^D , então:

$$clr(x) = \ln\left(\frac{x_1}{g_m(x)}, \frac{x_2}{g_m(x)}, \dots, \frac{x_D}{g_m(x)}\right) \quad (2.34)$$

onde $g_m(x)$ é a média geométrica calculada da seguinte maneira:

$$g_m(x) = \sqrt[D]{x_1 \cdot x_2 \cdot \dots \cdot x_D} \quad (2.35)$$

O processo inverso é dado pela seguinte equação:

$$x = C(\exp(\text{clr}_1(x), \text{clr}_2(x), \dots, \text{clr}_D(x))) \quad (2.36)$$

Diferentemente das *alr*, as transformações *clr* preservam as distâncias no espaço real, ou seja:

$$\langle x, y \rangle_a = \langle \text{clr}(x), \text{clr}(y) \rangle \quad (2.37)$$

$$\|x\|_a = \|\text{clr}(x)\|, \quad d_a(x, y) = d(\text{clr}(x), \text{clr}(y)) \quad (2.38)$$

Além disso, as propriedades de perturbação e potência são preservadas:

$$\text{clr}((\alpha \odot x) \oplus (\beta \odot y)) = \alpha \cdot \text{clr}(x) + \beta \cdot \text{clr}(y) \quad (2.39)$$

As desvantagem nesta transformação é que o espaço restrito das componentes das D-partes, ou seja, as dimensões do vetor D mudam para cada sub-composição por causa da média geométrica. A transformação *clr* é considerada, portanto, como uma isometria.

2.8.4. Metodologia proposta

Recentemente Egozcue & Pawlowsky (2003) definiram uma nova transformação que evita as desvantagem das transformações *alr* e *clr* apresentadas anteriormente. Essa transformação é conhecida como transformações logarítmicas isométricas *ilr* e é o foco principal de nossa dissertação. Começaremos definindo o conceito de base ortonormal essencial para este tipo de transformações.

Uma base ortonormal de S^D é uma série de composições e_1, e_2, \dots, e_{D-1} tal que $\langle e_i, e_j \rangle_a = 0$ para $i \neq j$ e $\|e_i\| = 1$. Para uma base fixa, as coordenadas de uma composição são obtidas usando a função:

$$x^* = \text{ilr}(x) = (\langle x, e_1 \rangle_a, \langle x, e_2 \rangle_a, \dots, \langle x, e_{D-1} \rangle_a) \quad (2.41)$$

A construção das coordenadas ortonormais é chamada de transformações *ilr* porque as coordenadas $x_j^* = ilr_j(x)$ são log-contraste e são isométricas (Pawlowsky et al., 2010):

$$ilr((\alpha \odot x) \oplus (\beta \odot y)) = \alpha \cdot ilr(x) + \beta \cdot ilr(y) \quad (2.42)$$

$$\langle x, y \rangle_a = \langle ilr(x), ilr(y) \rangle \quad (2.43)$$

$$\|x\|_a = \|ilr(x)\| \quad (2.44)$$

$$d_a(x, y) = d(ilr(x), ilr(y)) \quad (2.45)$$

Estas equações são análogas às propriedades dadas nas equações (2.38) e (2.39) das transformações *clr*. A única diferença é que o produto interno, a norma e a distância entre os vetores das coordenadas *ilr* correspondem à dimensão D-1 do espaço real, sendo um isomorfismo e uma isometria para S^D .

Como em qualquer espaço euclidiano, existe um número infinito de bases ortonormais no S^D que podem ser obtidas por vários métodos, por exemplo, o procedimento Gram-Schmidt definido por Egozcue et al. (2003) ou pelo procedimento Decomposição em Valores Singulares (SVD) comentado por Pawlowsky et al. (2010).

Pawlowsky (2005) propõe um novo método para obter uma base ortonormal, conhecida como partição seqüencial binária (SBP). Este procedimento e as equações que definem as transformações logarítmicas isométricas serão definidos passo a passo no capítulo 4 aplicado a um banco de dados real.

2.8.5. Presença de zeros

A análise dos dados composicionais está baseada nas razões logarítmicas, portanto, a presença de zeros (0) em uma composição traz consigo inconsistências na medida em que são feitas operações de tipo logarítmicas. Como consequência, qualquer análise de um vetor de componentes deve ser precedido por um tratamento desses zeros.

Os depósitos minerais são característicos pelo análise de múltiplas variáveis em conjunto. Em alguns casos, o banco de dados traz consigo variáveis analisadas cujo teor é zero (0). Muitos desses valores podem ser gerados a partir de erros de digitação na base de dados, arredondamento dos valores muito pequenos, variáveis não analisadas ou locais não amostrados (erros de digitação). No entanto, quando lidamos com este tipo de dados, se assume que as amostras analisadas como zero não podem ser descartadas nem eliminadas do banco de dados original, independentemente do significado que elas representem.

Pawlowsky et al., (2011) comentam uma serie de possíveis soluções de como lidar com este tipo de problemas. No entanto, para fins de nosso trabalho, as duas mais relevantes foram:

1. Substituir os zeros pelo limite de detecção de cada variável analisada.
2. Substituir os zeros por um valor muito pequeno, por exemplo, $1e^{-5}$.

Embora as duas soluções anteriores sejam empíricas, essa solução é dada em situações em que o teor de um elemento é exatamente zero, já que a metodologia de análise química utilizada não possui sensibilidade para detectar esse tipo de concentrações (0%) existente na amostra.

Capítulo 3

Como Lidamos com Problemas Multivariados Quando as Somas das Estimativas Devem Fechar 100%?

Neste capítulo, será apresentada uma série de metodologias usadas na geoestatística clássica quando se trabalha com depósitos multivariados (múltiplas variáveis) e, a soma dessas estimativas em cada bloco deva fechar uma constante (100% por exemplo). O estudo de caso é realizado em um depósito de Bauxita. Na primeira parte, faremos uma descrição geral das características principais do depósito e, posteriormente estudaremos cada uma das metodologias de estimativa usadas neste trabalho.

3.1. Localização

O estudo de caso corresponde aos dados provenientes de um depósito de Bauxita localizado na cidade de Porto Trombeta, estado de Pará-Brasil. O banco de dados foi disponibilizado pela empresa de Mineração Rio do Norte (MRN) situada aproximadamente a 880 km de Belém e 400 km de Manaus (ver figura 3.1).

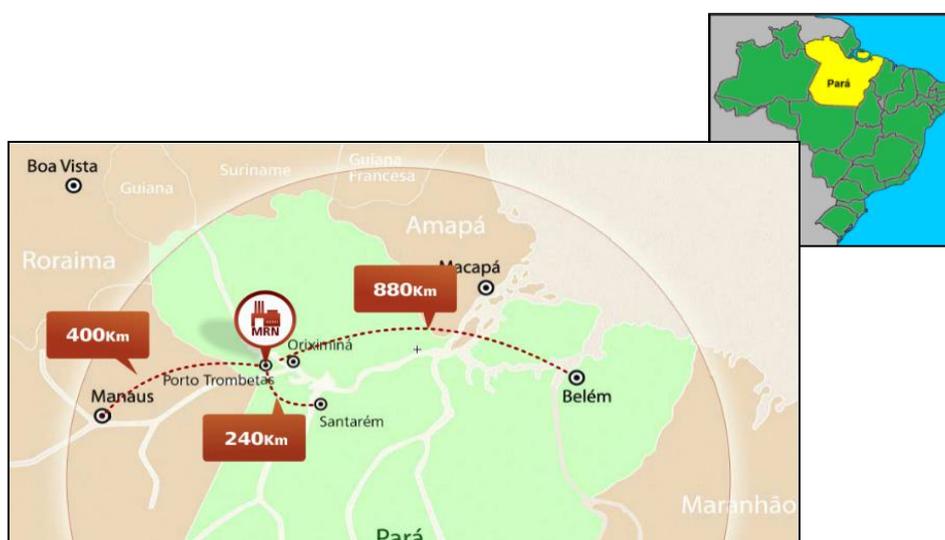


Figura 3.1. Mapa de localização da área de estudo, imagem tomada o dia 30-07-2013 às 13:00 do site da empresa MRN <http://www.mrn.com.br/pt-BR/Sobre-MRN/Porto-Trombetas/MapDeLocalizacao/Paginas/default.aspx>

3.2. Banco de dados

O depósito de Bauxita possui um domínio (Jamari) com grande extensão, que é o domínio que será estudado neste trabalho, contando com um grande número de amostras regularizadas em três dimensões e, uma peculiaridade importante da malha estar regularmente espaçada. Além disso, o banco de dados é isotópico, ou seja, todas as variáveis contêm informação em todos os lugares amostrados. Na Tabela 3.1, foi feito um resumo das características principais do depósito.

Tabela 3.1. Resumo das características principais do depósito de Bauxita.

No.	Descrição	Característica	Observações
1	Tipo de Depósito	Bauxita	-
2	Platô	Jamari	-
3	Domínio	Bauxita	-
4	Malha	200x200 m	Grid Regular
5	Número de Furos	626	Verticais (90°)
6	Número de Amostras	3735	Bauxita
7	Comprimento das Amostras	0.5 m	Variando de (0.3-0.8) metros
8	Dimensão da Área	13.5 km	Norte-Sul
9	Dimensão da Área	15 km	Leste-Oeste

As variáveis analisadas nesse estudo estão baseadas em dois tipos de balanços: espécies químicas e granulométrico, deduzidos da seguinte maneira:

1. O balanço granulométrico ou mássico é composto pelas frações +14#, +400#, -400#. Cada uma delas representa o material retido em cada fração, por exemplo, a fração granulométrica +14# se refere a todo material retido na peneira 14#, enquanto a fração granulométrica +400# diz respeito ao material passante na peneira 14# e retido na fração granulométrica 400#.
2. O balanço químico é composto pela soma das espécies químicas Fe₂O₃, TiO₂, Al₂O₃, SiO₂ e Resto. Cada uma delas representa a análise das variáveis químicas para cada fração granulométrica, por exemplo, o Ferro na fração 14# e 400# (Fe14 e Fe400) se referem às quantidades de Ferro (%) retidas nessas peneiras respectivamente.

Por último, na fração -400# não foram analisadas as espécies químicas, somente a massa.

O banco de dados não apresentava uma soma constante com as informações provenientes do laboratório, ou seja, a soma das espécies químicas e mássicas em cada fração era diferente de 100%. Para solucionar este problema, no caso das variáveis químicas, foi criada uma variável artificial chamada Resto14 e Resto400, que tem como objetivo carregar todo o teor faltante em cada faixa para fechar 100%. Por exemplo, se a soma das espécies químicas na faixa +14# é 90%, o 10% que falta para fechar o balanço é atribuído à variável Resto14.

No caso das frações granulométricas, o erro de fechamento foi distribuído proporcionalmente em cada uma das frações para poder fechar o balanço mássico em 100%. Por exemplo, se a soma das frações +14#, +400# e -400# numa análise é de 97%, o 3% restante é dividido entre as três frações, ou seja, é adicionado a cada fração uma pequena percentagem em relação a sua massa. Na Tabela 3.2, apresentam-se cada uma das variáveis analisadas detalhadamente, cuja nomenclatura é a mesma utilizada pela empresa. As cores verdes representam as frações granulométricas retidas em cada peneira (+14#, +400# e -400#) com suas respectivas espécies químicas (Fe_2O_3 , TiO_2 , Al_2O_3 , SiO_2 e Resto).

Tabela 3.2. Resumo das características das variáveis analisadas no depósito.

o.	DESCRIÇÃO	NOMENCLATURA	OBSERVAÇÕES
1	Fração granulométrica +14	Rec14	Recuperação na peneira +14
2	Ferro na fração +14	Fe14	Fe_2O_3
3	Titânio na fração +14	Ti14	TiO_2
4	Alumina Aproveitável na fração +14	Alap14	Al_2O_3
5	Sílica reativa na fração +14	Sire14	SiO_2
6	Variável artificial na fração +14	Resto14	Criada para fechar 100%
7	Fração granulométrica +400	Rec400	Recuperação na peneira +400
8	Ferro na fração +400	Fe400	Fe_2O_3
9	Titânio na fração +400	Ti400	TiO_2
10	Alumina Aproveitável na fração +400	Alap400	Al_2O_3
11	Sílica reativa na fração +400	Sire400	SiO_2
12	Variável artificial na fração +400	Resto400	Criada para fechar 100%
13	Fração granulométrica -400	Rec-400	Recuperação na peneira -400

No total, são 13 variáveis que correspondem a 5 espécies químicas em duas faixas granulométricas e 3 parâmetros físicos (massa retida em cada peneira). A variável principal é a alumina aproveitável (Al_2O_3) e o contaminante principal é a sílica reativa (SiO_2).

A Figura 3.2 mostra a disposição da malha amostral vista em planta, a fim de apresentar a área coberta pelos dados amostrados.

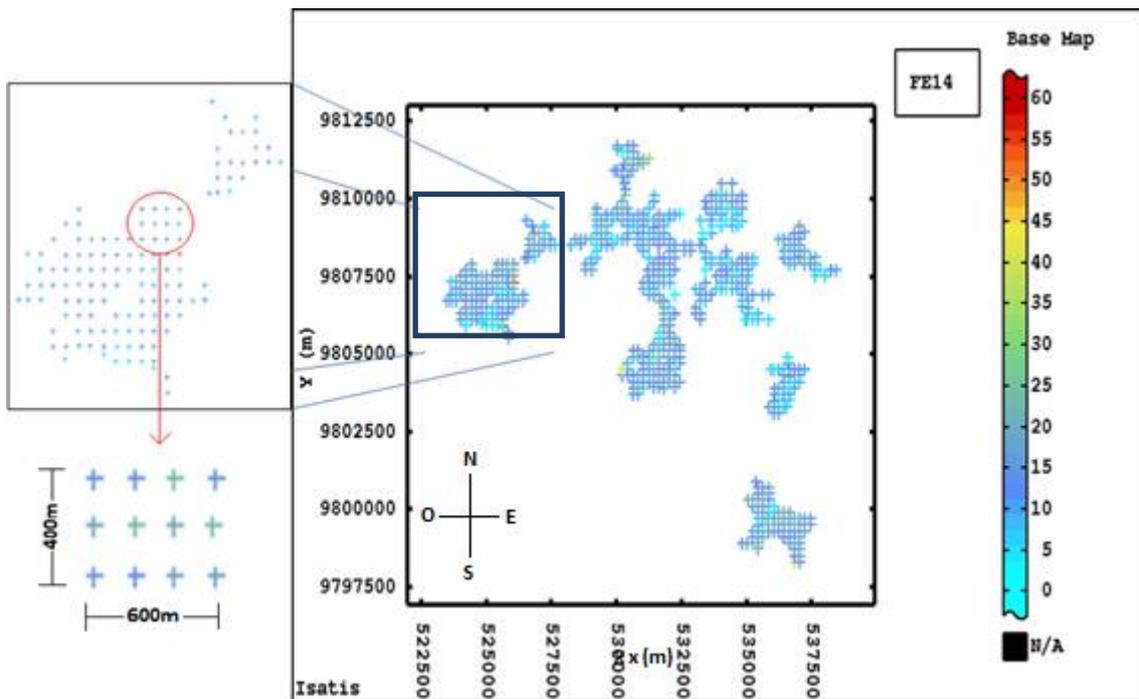


Figura 3.2. Mapa de localização da variável FE na fração +14.

Uma característica importante neste tipo de mineiro é a descontinuidade da camada laterita ferruginosa ao longo da área do depósito, encontrando lugares onde a camada desaparece devido a flutuações do lençol freático que dá origem a este depósito (processos lateríticos). A evolução do manto laterítico e o zoneamento regional provavelmente foram controlados tanto por variações climáticas no tempo e no espaço, como por fatores geomorfológicos, geotécnicos e sedimentológicos (Kotschoubey *et. al*, 2005).

3.3. Estatística básica

Na tabela 3.3, é apresentado um resumo estatístico das 13 variáveis analisadas (químicas e mássicas). No total, são 3735 amostras por cada variável, cujos valores em percentagem oscilam entre 0 e 100%, lembrando que estamos trabalhando com balanços químicos e granulométricos.

Tabela 3.3. Resumo da estatística básica dos dados originais.

Variável	Valores (%)								
	Dados	Mínimo	Máximo	Média Ponderada	Desvio Padrão	Variância	Q25	Q50	Q75
Alap14	3735	13.07	58.78	49.17	3.84	14.71	46.67	49.25	51.59
Fe14	3735	0.69	54.65	9.49	4.98	24.8	5.42	8.41	12.02
Sire14	3735	0.27	14.9	4.44	1.97	3.89	3.28	4.62	6.01
Ti14	3735	0.27	3.57	1.53	0.44	0.2	1.15	1.44	1.76
Resto14	3735	21.47	52.32	35.38	2.88	8.31	33.9	35.78	37.49
Alap400	3735	0.2	58.32	47.66	5.4	29.18	44.2	47.33	49.96
Fe400	3735	0.56	46.46	8.11	5.23	27.32	4.97	7.84	11.44
Sire400	3735	0.46	28.9	4.84	1.84	3.39	3.72	4.91	6
Ti400	3735	0.32	7.13	1.37	0.58	0.34	1.09	1.44	1.84
Resto400	3735	24.25	79.24	38.01	4.67	21.81	35.74	37.83	39.81
Rec14	3735	3.21	97.52	67.63	15.84	250.76	59.01	70.21	79.06
Rec400	3735	0.55	44.86	9.36	5.27	27.73	5.72	8.13	11.77
Rec-400	3735	1.04	94.14	23	12.89	166.11	13.27	21.17	30.46

Uma peculiaridade importante neste banco de dados é que as variáveis químicas foram analisadas em massas diferentes, ou seja, cada teor analisado é representativo da massa retida em cada fração granulométrica. Isso traz como consequência que cada teor tenha um peso (suporte) diferente. Portanto, para análise estatística não faremos uso da média aritmética e sim pela média ponderada que é calculada através do somatório das multiplicações entre os valores (teor) e os pesos (recuperação mássica) divididos pelo somatório dos pesos. O cálculo é feito da seguinte maneira:

$$\text{Média Ponderada} = \frac{\sum_{i=1}^n x_i * w_i}{\sum_{i=1}^n w_i} \quad (3.1)$$

Onde:

- x_i = Valor da i-ésima amostra.
- w_i = Valor do i-ésimo peso.

Como as medidas de dispersão (variância e desvio padrão) são sensíveis aos valores extremos, a identificação dos valores mínimos e máximos são essenciais para detectar este tipo de valores. Se considerarmos como valores extremos (altos ou baixos), por exemplo, aqueles dados que distanciam-se da média de 3 a 4 vezes o desvio padrão, podemos observar claramente que em todas as variáveis existem este tipo de valores extremos.

Uma das ferramentas mais amplamente utilizada para análise estatística é o histograma, pois ele permite descrever o comportamento da variável em relação a sua tendência central, forma e dispersão.

Considerando somente as variáveis de maior impacto econômico, a figura 3.3 mostra os histogramas de freqüências das variáveis Alap14 (a) e Sire14 (b) na fração granulométrica +14#, além da estatística básica e a distribuição de probabilidade dos teores.

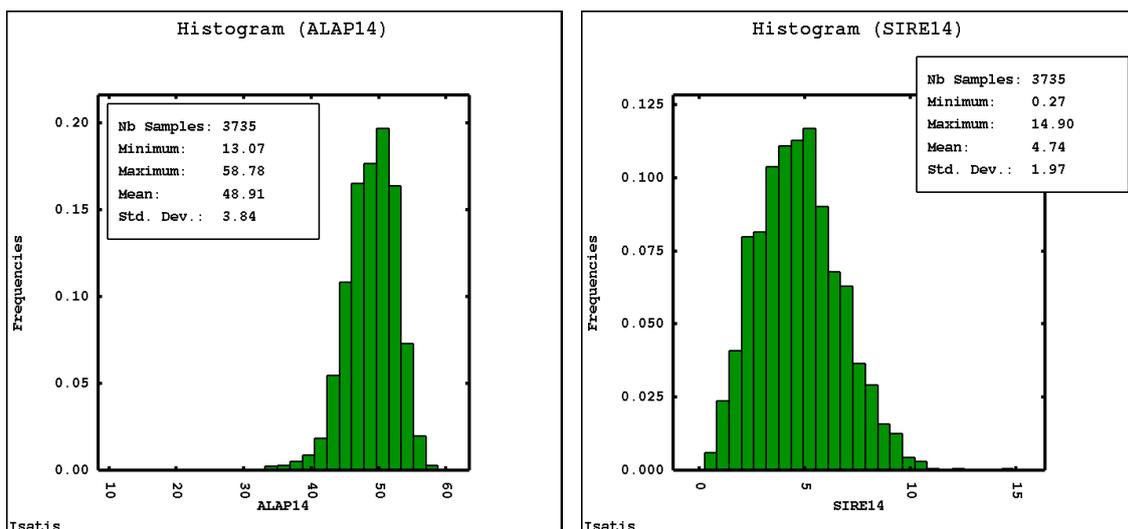


Figura 3.3. Histogramas de freqüências (a) Variável Alap14 (b) Variável Sire14.

Visualmente, a figura 3.3 (a) mostra uma assimetria negativa, isto devido a presença de alguns valores extremos na parte inferior do histograma. No entanto, a figura 3.2 (b) mostra somente uma leve assimetria positiva com poucos valores extremos na parte superior. Outro aspecto importante é que grande parte dos teores das amostras estão concentradas na faixa de (40-56)% da variável Alap14 e (2-10)% da variável Sire14.

A tabela 3.4 mostra os valores dos coeficientes de assimetria, indicando a forma assimétrica (positiva ou negativa) ou simétrica (aproximadamente zero) dos histogramas de cada variável.

Tabela 3.4. Coeficiente de assimetria das variáveis analisadas.

Variável	Coeficiente de assimetria
Alap14	-1.24
Fe14	1.4
Sire14	0.43
Ti14	0.5
Resto14	0.35
Alap400	-1.76
Fe400	1.42
Sire400	1.19
Ti400	1.26
Resto400	1.97
Rec14	0.9
Rec400	1.59
Rec-400	1.03

3.4. Acumulação

Para evitar a combinação de amostras com suportes diferentes, incorrendo em um erro em estimar teores com amostras de representatividade mássica distinta, foram utilizadas as variáveis acumuladas dos atributos de interesse nas faixas granulométricas analisadas.

As variáveis acumuladas são obtidas através da multiplicação do teor na faixa granulométrica pela respectiva porcentagem retida na fração granulométrica recuperada. Por exemplo, a variável Fe_2O_3 acumulada (%) na

fração granulométrica +14# corresponde ao teor de Fe_2O_3 na faixa granulométrica +14# multiplicado pela fração mássica recuperada da faixa granulométrica Rec14.

O processo de acumulação leva a obter uma variável com uma nova distribuição de probabilidades, considerando agora o fato das amostras estarem no mesmo suporte amostral. Na figura 3.4b, podemos observar como a nova distribuição dos dados da variável Alap14 se comporta ao considerar o mesmo suporte amostral. Note-se, que para esse caso, a nova distribuição dos dados acumulados apresenta uma leve assimetria em relação à distribuição dos dados originais cuja assimetria é bem definida, figura 3.4a.

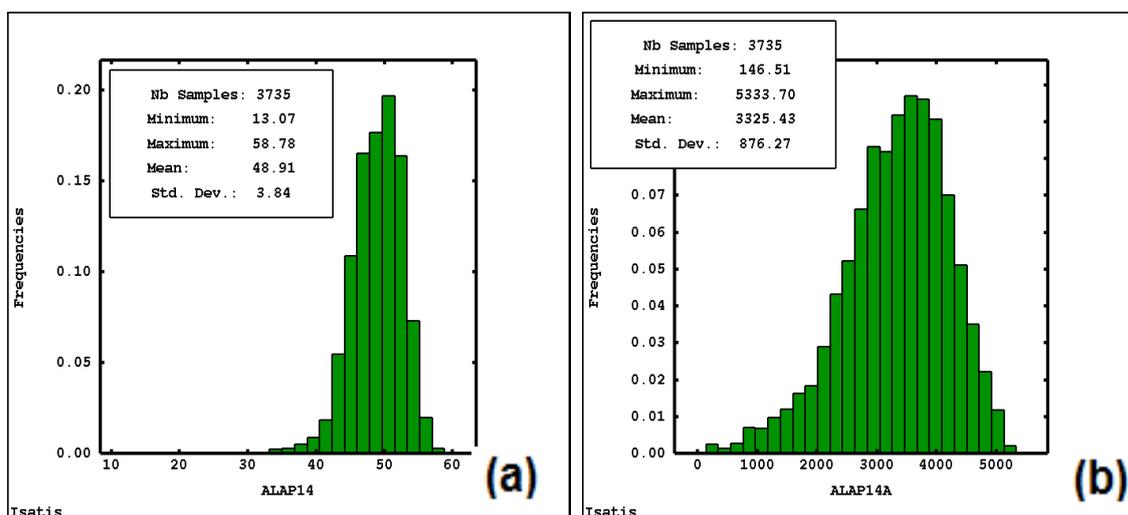


Figura 3.4. (a) Histograma ALAP14 (Original), (b) Histograma ALAP14A (Acumulada).

Esse processo é realizado para todas as espécies químicas nas suas respectivas frações granulométricas, portanto para a fração +14# utilizou-se como acumulador a Rec14 e para a fração +400# utilizou-se como acumulador a Rec400. A fração -400# não possui análise química, então não existe processo de acumulação.

No final, para obter novamente o teor de cada uma das variáveis (processo de desacumulação), cada teor acumulado estimado na sua fração granulométrica é dividido pela recuperação mássica estimada na fração correspondente. Por exemplo, a estimativa do Fe14A (Ferro acumulado na

fração +14#) é dividida pela estimativa da REC14 (Recuperação mássica na fração +14#).

A nova nomenclatura para essas variáveis acumuladas tem ao final a letra A que significa acumulação. Na tabela 3.5, foi feito um resumo estatístico das variáveis acumuladas.

Tabela 3.5. Resumo estatístico das variáveis acumuladas.

Variável	Dados	Mínimo	Máximo	Média	Variância	Desvio Padrão	Q25	Q50	Q75
ALAP14A	3735	146.51	5333.7	3325.43	767843.57	876.27	2798.21	3414.28	3960.81
FE14A	3735	4.85	4304.69	641.79	168149.32	410.06	333.55	564.6	874.41
SIRE14A	3735	5.71	1043.04	300.14	10736.88	103.62	229.49	301.22	367.79
TI14A	3735	0.93	331.15	103.17	2062.96	45.42	69.74	99.78	131.87
RESTO14A	3735	138.98	3727.41	2392.87	257365.69	507.31	2129.26	2458.97	2741.98
ALAP400A	3735	2.42	2466.85	446.26	78085.51	279.44	255.03	371.78	566.21
FE400A	3735	4.28	440.35	75.92	2929.42	54.12	36.72	62.01	100.56
SIRE400A	3735	2.12	349.11	45.35	1003.16	31.67	23.49	36.62	58.6
TI400A	3735	0.85	67.87	12.83	40.27	6.35	8.25	11.76	16.21
RESTO400A	3735	28.2	1772.42	355.9	40934.03	202.32	218.3	307.38	443.83

3.5. Desagrupamento

O interesse econômico da maioria das empresas de mineração é encontrar muito mais minério que estéril, e isto serve como referência para amostrar mais nos lugares onde se concentra grande parte do bem mineral. O principal problema que geralmente acontece é que a média das amostras é influenciada por esse agrupamento amostral e, muitas vezes pode gerar expectativas em subestimar ou superestimar a média de todo o depósito. Portanto, é necessário fazer um desagrupamento dessas amostras e obter uma média mais realista ao depósito.

O método escolhido nessa dissertação foi o desagrupamento poligonal ou polígono de Voronoi (Isaaks e Srivastava, 1989) onde cada amostra recebe um peso proporcional a área do polígono (2D) ou poliedro (3D) formado na sua vizinhança.

Lembrando que, como as espécies químicas possuíam suporte amostral diferente (teores em relação à recuperação em cada faixa) foi feita a média desagrupada ponderada, que consiste em outorgar um peso a cada amostra em relação a sua recuperação e logo dividir pelo somatório dos pesos. Para o caso das recuperações somente foi feita a média desagrupada, pois eles não têm o problema de suporte amostral. A tabela 3.6 mostra a média dos dados ponderados e a média desagrupada ponderada obtida por poliedros de Voronoi.

Podemos observar claramente, que em todas as variáveis, a média dos dados ponderados está muito próxima à média dos dados desagrupados, observando que os dados não se encontram amostrados preferencialmente. O ganho nesse método é que podemos usá-lo como referência posteriormente para checar a qualidade das estimativas (média desagrupada vs média estimada e análise de deriva) estudados no capítulo 5.

Tabela 3.6. Resumo da média desagrupada ponderada e a média dos dados ponderados.

VARIABLE	Dados Desagrupados		Dados Ponderados	
	Média Ponderada	Variância	Média Ponderada	Variância
NN_Alap14	49.3	13.64	49.17	14.71
NN_Fe14	9.23	23.47	9.49	24.8
NN_Sire14	4.49	3.44	4.44	3.89
NN_Ti14	1.51	0.21	1.53	0.2
NN_Resto14	35.47	7.43	35.38	8.31
NN_Alap400	46.51	29.65	47.66	29.18
NN_Fe400	9.04	27.93	8.11	27.32
NN_Sire400	4.81	3.45	4.84	3.39
NN_Ti400	1.53	0.33	1.37	0.34
NN_Resto400	38.1	21.76	38.01	21.81
NN_Rec14	67.28	248.11	67.63	250.76
NN_Rec400	9.53	28.08	9.36	27.73
NN_Rec-400	23.19	163.25	23	166.11

3.6. Transformação de coordenadas

Esse trabalho utilizou a transformação da coordenada Z (vertical) de cada amostra, passando-as de coordenadas cartesianas às coordenadas estratigráficas. Desta forma, “alinha-se” amostras que estejam em um mesmo horizonte geoquímico e melhorar a análise da continuidade espacial dos valores dos atributos e as estimativas desses valores. Duas metodologias de transformação da coordenada Z foram avaliadas: distancia das amostras a Capa ou a Lapa.

Considerou-se que a camada de bauxita apresenta leves mudanças de espessura ao longo de sua extensão, portanto a transformação utilizada por Deutsch (2002) foi um pouco modificada em relação a essa hipótese. A Capa ou Lapa foram aplainadas e as demais coordenadas foram alteradas medindo suas distancias a esses planos horizontalizados. A nova coordenada e dada pela equação 3.2:

$$Z(i)_{estr} = Z_{original} - Z_c \quad (3.2)$$

$$i = 1, \dots, n \quad n = \text{número de amostras}$$

Onde:

- $Z(i)_{estr}$ = Nova coordenada Z da amostra;
- $Z_{original}$ = Coordenada Z atual da amostra;
- Z_c = Coordenada Z (vertical) medida pela distância até a superfície horizontalizada (Capa ou Lapa).

A figura 3.5 ilustra o processo de correção estratigráfica feita pela distância à superfície da Lapa. As linhas coloridas se referem aos contatos geológicos das camadas que ficam acima da bauxita. Para ser mais específico, a linha rosa representa o topo da camada de bauxita nodular, a linha verde o topo da camada de laterita, a linha vermelha o topo da camada de bauxita e a linha cinza o topo da camada de argila de fundo.

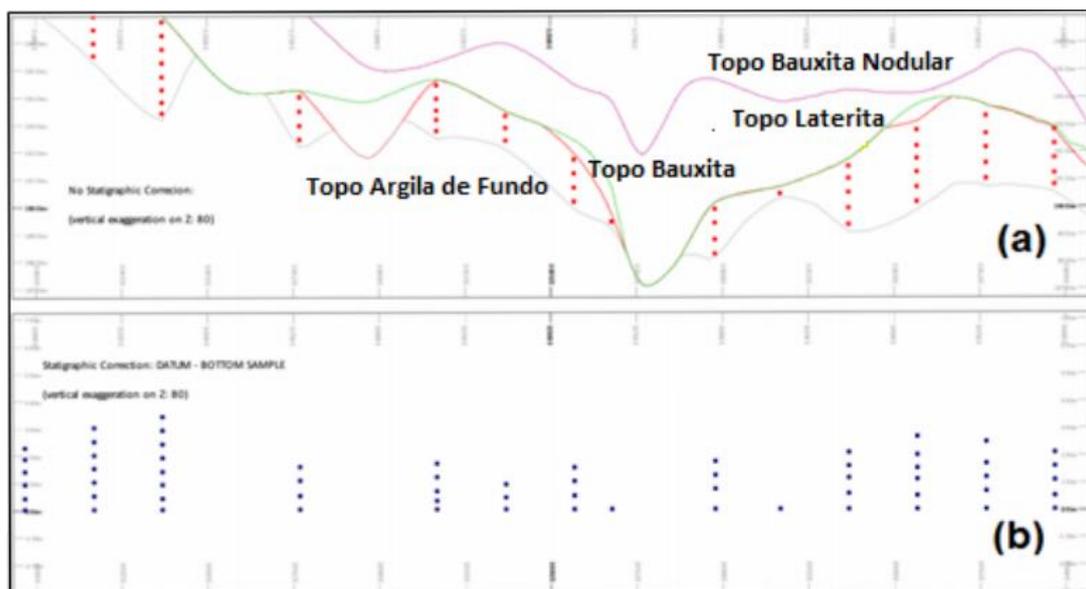


Figura 3.5. (a) Camada sem correção, (b) Camada corrigida pela Lapa. Imagem tomada da apresentação final do projeto MRN 2013.

A fim de avaliar as duas correções e escolher a que será utilizada no processo de análise da continuidade espacial e estimativa, criaram-se dois bancos de dados com novas coordenada em Z (distancia a Capa ou Lapa). Utilizou-se o método de validação cruzada (Isaaks e Srivastava, 1989) para checar a qualidade das estimativas para as duas bases de dados. A figura 3.6b, mostra que a variável Rec14 tem uma pequena melhoria da distribuição do erro com a correção estratigráfica pela distancia a Lapa (-0.02) fica mais próxima de zero que a média da Capa (0.30).

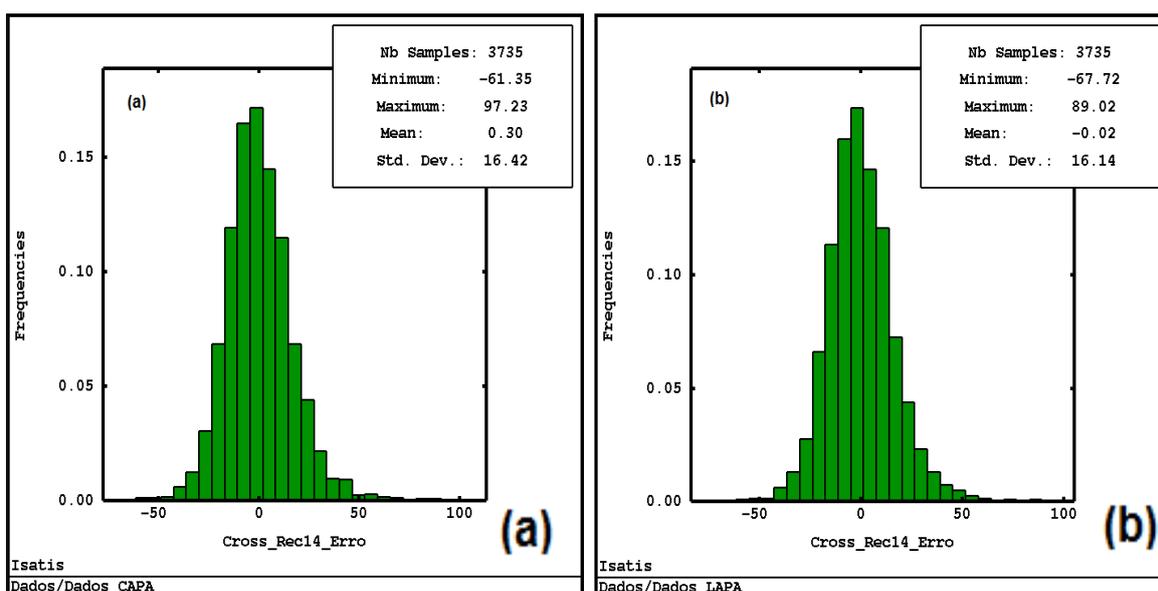


Figura 3.6. (a) Distribuição do Erro Rec14 na CAPA, (b) Distribuição do Erro Rec14 na LAPA.

3.7. Continuidade espacial

O conhecimento das direções de maior, intermediária e menor continuidade das variáveis analisadas é primordial na próxima etapa de interpolação (Krigagem e Cokrigagem). A fim de obter uma representação mais realista do depósito, foi utilizada a modelagem da continuidade espacial para as variáveis acumuladas, pois elas se encontram todas no mesmo suporte amostral. A presença de alguns valores extremos levou à utilização do correlograma não ergódico, pois ajuda a minimizar o impacto gerado por alguns pares amostrados que pode provocar muito ruído na variografia, dificultando sua modelagem.

Duas metodologias foram consideradas:

1. Estimar as variáveis Alap14A, Fe14A, Sire14A, Ti14A e Resto14A utilizando o modelo da recuperação na faixa +14# (Rec14) e, estimar as variáveis Alap400A, Fe400A, Sire400A, Ti400A e Resto400A, utilizando o modelo da recuperação na faixa +400# (Rec400). Como estamos trabalhando com acumulações, quer-se evitar gerar valores finais (desacumulados) em desacordo com as amostras originais (teores além dos valores mínimo ou máximo dos dados iniciais de cada atributo de interesse); portanto, esse procedimento é recomendável. O problema nesse caso é que provavelmente o modelo ajustado seja aderente para algumas variáveis e em algumas direções, mas não muito bom para outras.
2. Estimar cada uma das variáveis químicas e mássicas utilizando o seu próprio modelo de continuidade espacial. O problema aqui é que quando trabalhamos com acumulações, utilizar modelos variográficos diferentes, para cada variável, pode gerar distorções nas estimativas, devido aos pesos diferentes para a variável acumulada. Porém, nesse caso, o ganho é que cada modelo é representativo da realidade física do depósito.

Na análise variográfica, as variáveis acumuladas não apresentaram uma anisotropia clara no plano XY, ou seja, o banco de dados mostrou uma isotropia (mesma continuidade em todas as direções principais). Portanto, foi feito um correlograma omnidirecional no plano XY, obtendo assim o alcance das direções principais nesse mesmo plano, a direção de menor continuidade foi para o plano vertical, ou seja, ao longo da direção dos furos.

O efeito pepita foi obtido através da modelagem do correlograma vertical, pois ele auxilia a determinação da variabilidade de amostras a curtas distâncias. A figura 3.7 e 3.8, mostra o correlograma experimental nas 8 direções principais do plano XY e o modelo ajustado das variáveis (Rec14 e Rec-400).

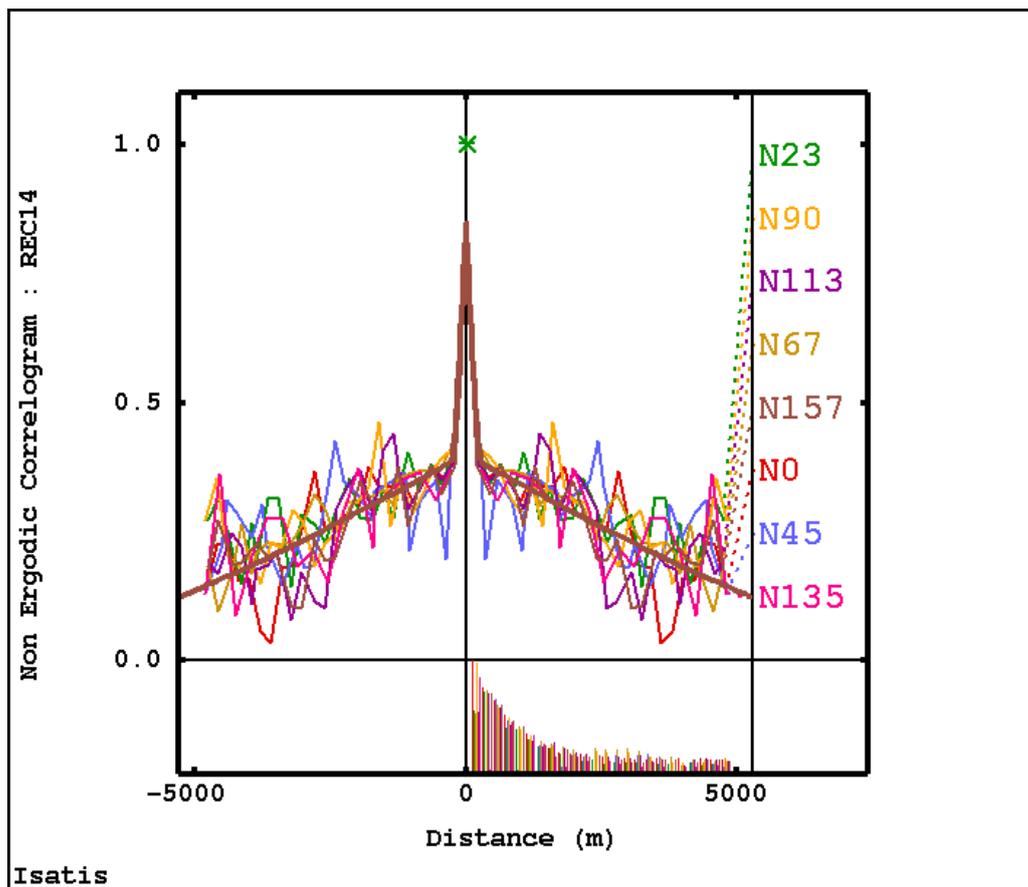


Figura 3.7. Correlograma experimental Rec14 nas 8 direções principais e o modelo ajustado.

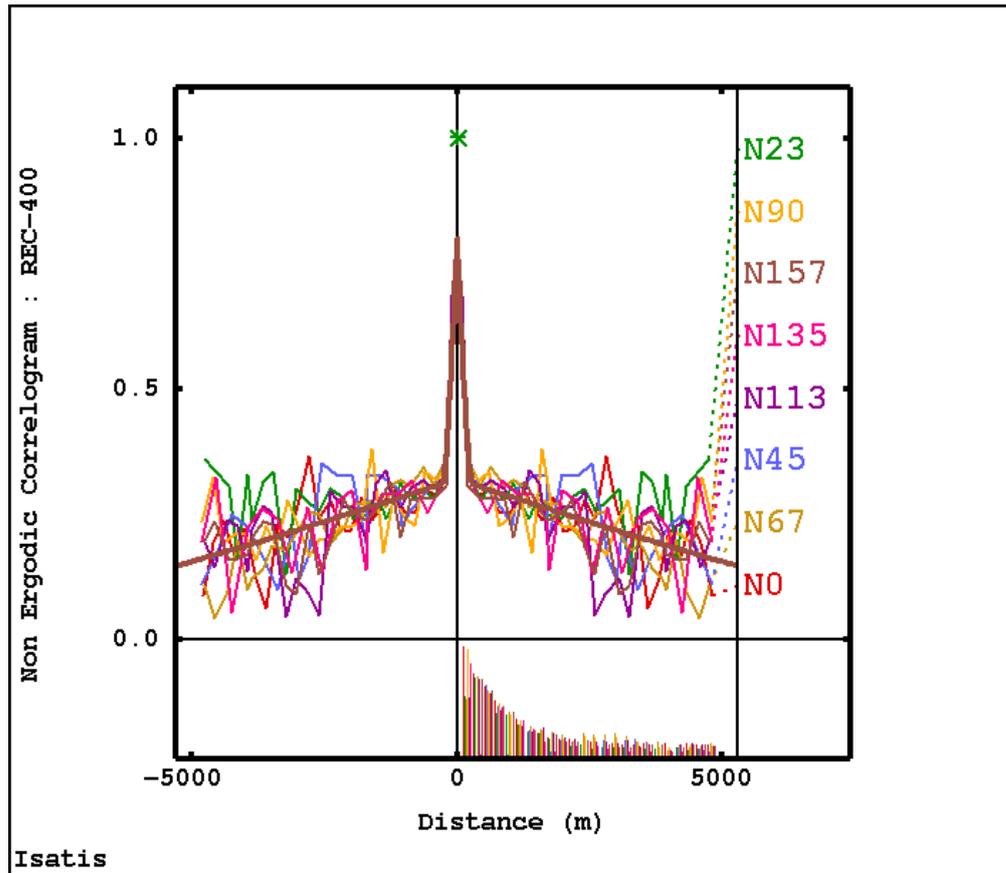


Figura 3.8. Correlograma experimental Rec-400 nas 8 direções principais e o modelo ajustado.

Na tabela 3.7, podemos observar um resumo dos parâmetros utilizados na variografia, ressaltando que para o caso dos correlogramas direcionais foi utilizada uma largura da banda vertical não maior que 0.5 m para evitar a mistura de amostras de diferentes níveis estratigráficos, pois lembramos que os dados em Z foram transformados em coordenadas estratigráficas.

Tabela 3.7. Parâmetros utilizados na Variografia Omnidirecional e Direcional.

Variografia								
Tipo	Direção	Mergulho	No. Lag (m)	Lag (m)	Tolerância Lag (m)	Tolerância Angular	Largura da Banda (m)	Tipo de função de continuidade espacial
Omni	Maior	N 0° Dip 0°	25	200	100	90°	1x10e8	Correlograma Não Ergódico
	Intermediária	N 90° Dip 0°	25	200	100	90°	1x10e8	
	Menor	N 0° Dip 90°	10	0.5	0.25	90°	0.5	
Dir	8	0° até 157.5°	25	200	100	22.5°	100	Correlograma Não Ergódico

Para a modelagem das variáveis acumuladas foram necessárias duas estruturas com modelos esféricos (além do efeito pepita). Na tabela 3.8, é feito um resumo detalhado onde podemos observar os alcances do correlograma medido em metros, assim como o efeito pepita e as contribuições de cada estrutura, chegando a um patamar cuja soma é 1.

Tabela 3.8. Resumo dos parâmetros encontrados na modelagem da variografia.

Variável	Modelo de Continuidade										
	c0	Primeira Estrutura					Segunda Estrutura				
		Modelo	c1	a (D1)	a (D2)	a (D3)	Modelo	c2	a (D1)	a (D2)	a (D3)
ALAP14A	0.15	0.47	300	300	3	0.38	12000	12000	4.8		
FE14A	0.13	0.54	350	350	2.7	0.33	8000	8000	8		
SIRE14A	0.1	0.67	350	350	2.4	0.23	5000	5000	9		
TI14A	0.05	0.47	300	300	3.3	0.48	8000	8000	5.7		
RESTO14A	0.17	0.51	300	300	2.2	0.32	7500	7500	7.5		
ALAP400A	0.12	0.43	300	300	3.3	0.45	4400	4400	17		
FE400A	0.17	0.58	350	350	2.6	0.25	3000	3000	80		
SIRE400A	0.15	0.45	300	300	4	0.4	6200	6200	9		
TI400A	0.27	0.58	350	350	3.3	0.15	9000	9000	100		
RESTO400A	0.15	0.4	300	300	3.5	0.45	5800	5800	10		
REC14	0.15	0.45	300	300	3	0.4	10500	10500	7		
REC400	0.15	0.4	300	300	3.5	0.45	5000	5000	14		
REC-400	0.2	0.48	300	300	3	0.32	14000	14000	5		

- D1 – Direção de Maior Continuidade – N 0° Dip 0°.
- D2 – Direção de Continuidade Intermediária - N 90° Dip 0°.
- D3 – Direção de Menor Continuidade - N 0° Dip 90°.
- c0 - Efeito pepita.
- a – Alcance.
- c1 – Patamar ou Contribuição da Primeira Estrutura.
- c2 – Patamar ou Contribuição da Segunda Estrutura.
- OBS 1: Foi utilizado como medida de continuidade espacial o Correlograma não Ergódico.
- OBS 2: Os parâmetros dos modelos de continuidade consideram patamar total igual a 1.

Note-se que existem algumas variáveis onde o alcance na segunda estrutura é relativamente grande em relação às dimensões do depósito, por exemplo, Alap14A no plano XY de 12000m e Ti400A na vertical de 100m. Isto devido à existência de uma anisotropia zonal (Wackernagel, 1995), onde as amostras, em determinada direção, apresentam baixa variância, nunca

chegando ao patamar do variograma ou correlograma. Para conseguir obter um modelo de continuidade espacial, com as mesmas estruturas das demais direções, utiliza-se um alcance muito grande, dando a impressão de ter um patamar baixo.

Nas figuras 3.9 (a, b, c), podemos observar a modelagem do correlograma experimental omnidirecional da variável Rec14 nas 3 direções principais (maior, intermediária e menor) respectivamente. O mapa variográfico (d) foi feito no plano XY, observando que não foi possível captar uma anisotropia geométrica nesse plano, ou seja, os dados apresentaram uma isotropia (mesma continuidade em todas as direções).

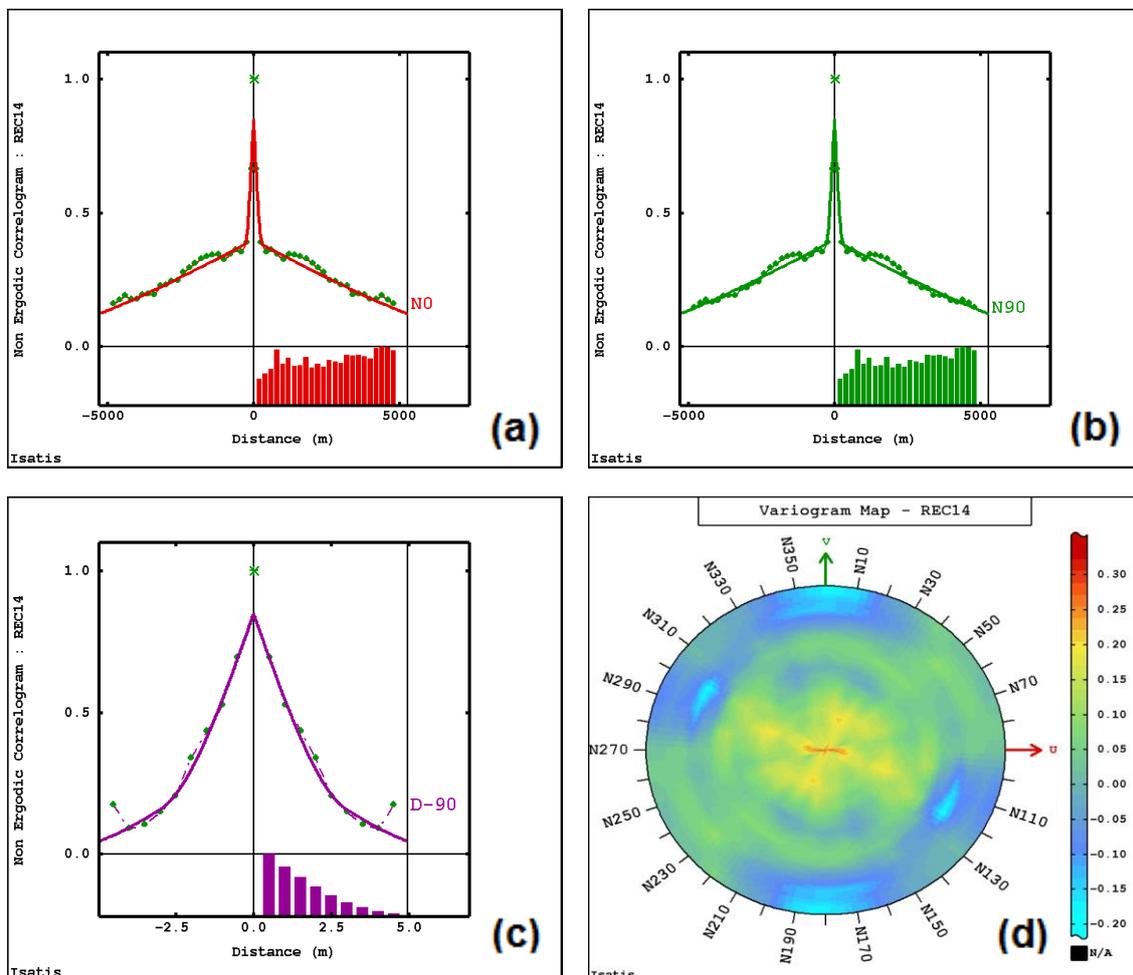


Figura 3.9. Correlograma não ergódico variável Rec14 (a) Maior, (b) Intermediária, (c) Menor, (d) Mapa variográfico no plano XY.

3.8. Estimativa

O foco principal nesse capítulo é fazer uso das técnicas de estimativas geoestatísticas geralmente mais utilizadas no dia a dia quando se apresenta a necessidade de trabalhar com dados multivariados, cuja finalidade é o fechamento (a soma das estimativas de todas as variáveis analisadas é 100%). As técnicas utilizadas nessa dissertação são a Krigagem Ordinária e a Cokrigagem Ordinária utilizando o Modelo de Correção Intrínseca (MCI) definidos por Wackernagel (1995).

No processo de estimativas, foram testadas 5 opções diferentes, fazendo variações no tipo de banco de dados e o modelo variográfico.

1. Banco de dados:

- **Isotópico:** Todas as amostras estão disponíveis, para todas as variáveis, nos mesmos locais. Como sucede em nosso banco de dados em particular.
- **Heterotópico:** Algumas amostras (não todas), das diferentes variáveis, compartilham locações amostrais outras não. Foi gerada essa situação eliminando 626 amostras aleatoriamente do total disponível para cada espécie química, com a condição de ser uma amostra por furo. Na fração mássica, foi mantido o banco de dados isotópico.

2. Modelo variográfico:

- *Modelos Individuais:* Cada variável é interpolada com seu próprio modelo de continuidade bem definido anteriormente na variografia. Por exemplo, a variável Alap14A utilizou o seu próprio modelo, a variável Fe400a utilizou o seu modelo e assim para as demais variáveis.
- *Modelos Únicos:* Cada variável é interpolada com o modelo da recuperação em cada fração granulométrica, ou seja, as variáveis Alap14A, Fe14A, Sire14A, Ti14A, Resto14A utilizaram o modelo

variográfico da Rec14 e as variáveis Alap400A, Fe400A, Sire400A, Ti400A, Resto400A utilizaram o modelo variográfico da Rec400.

Na tabela 3.9, é feito um resumo de cada uma das metodologias utilizadas no processo de estimativa e suas características principais.

Tabela 3.9. Resumo das características metodológicas utilizadas na estimativa.

Característica	Metodologia				
	1	2	3	4	5
Dados	Isotópico	Isotópico	Heterotópico	Heterotópico	Isotópico-Heterotópico
Modelo	Individuais	Único	Individuais	Único	Único
Estimativa	Krigagem Ordinária	Krigagem Ordinária	Krigagem Ordinária	Krigagem Ordinária	Cokrigagem (MCI)

Em cada uma das etapas da interpolação foram feitos 4 testes para krigagem ordinária, cuja finalidade é obter várias alternativas de estimativas e escolher aquela que gere mais aderência nas validações discutidas no capítulo 5, sendo a estratégia ou teste 4 que obteve melhores resultados. Na tabela 3.10, foi feito um resumo dos parâmetros utilizados para cada um desses testes, considerando uma discretização de 25 pontos por bloco 5x5x1, em leste, norte e vertical respectivamente, já que as compostas têm em média o comprimento vertical do bloco.

Tabela 3.10. Resumo dos parâmetros para os 4 testes da krigagem.

	Teste 1	Teste 2	Teste 3	Teste 4
No. mín. de amostras	3	3	3	3
No. máx. de amostras	16	24	12	24
Setores Angulares	4	8	4	8
No. máx. amostras por setor	4	3	3	3
Setores não vazios	-	3	-	-
Discretização do bloco	5x5x1	5x5x1	5x5x1	5x5x1

Outro aspecto relevante foi a definição da estratégia de busca do elipsoide, ou seja, até quantos metros irá a busca por amostras para obter a estimativa num determinado bloco. Geralmente, é utilizada a mesma estratégia do modelo variográfico, pois ela representa a continuidade espacial dos dados. Nesse caso em particular algumas variáveis apresentam uma anisotropia zonal, aonde as amostras nunca chegam ao patamar do modelo variográfico.

Portanto, para poder modelar adequadamente um variograma que possua uma anisotropia zonal, com as mesmas estruturas das demais direções, utiliza-se um alcance muito grande (artificial) para poder atingir a variância global no infinito.

Na tabela 3.11, se mostra o alcance máximo dos modelos de cada variável na segunda estrutura. Para nosso caso em particular, utilizaremos o chamado modelo prático, que consiste em utilizar como alcance máximo onde começa a anisotropia zonal.

Tabela 3.11. Resumo dos parâmetros utilizados na busca da krigagem.

Alvo	Litologia	Variável	Modelo Original			Modelo Prático			Variância		
			Segunda Estrutura			Busca até			%		
			a (D1)	a (D2)	a (D3)	a (D1)	a (D2)	a (D3)	a (D1)	a (D2)	a (D3)
JAMARI	Bauxita	ALAP14A	12000	12000	4.8	4500	4500	4	83	83	98
		FE14A	8000	8000	8	4000	4000	3	90	90	85
		SIRE14A	5000	5000	9	5000	5000	3	100	100	88
		TI14A	8000	8000	5.7	5000	5000	3.5	92	92	91
		RESTO14A	7500	7500	7.5	4000	4000	4.5	91	91	93
		ALAP400A	4400	4400	17	4400	4400	4	100	100	71
		FE400A	3000	3000	80	3000	3000	2.5	100	100	76
		SIRE400A	6200	6200	9	4500	4500	4	96	96	85
		TI400A	9000	9000	100	4500	4500	2.5	95	95	81
		RESTO400A	5800	5800	10	3500	3500	3.5	91	91	78
		REC14	10500	10500	7	4800	4800	4	86	86	91
		REC400	5000	5000	14	5000	5000	4	100	100	74
		REC-400	14000	14000	5	4800	4800	3	84	84	93

D1 – Direção de Maior Continuidade – N 0° Dip 0°

D2 – Direção de Continuidade Intermediária - N 90° Dip 0°

D3 – Direção de Menor Continuidade - N 0° Dip 90°

a – Alcance

Um exemplo claro se faz evidente na variável Alap14A que tem um alcance no modelo original de 12000m no plano XY, mas o alcance prático é de apenas 4500m onde a anisotropia zonal se faz evidente. Além disso, foi apresentada a percentagem da variância estandardizada (patamar ao utilizar um correlograma) que atinge esse modelo prático, nesse caso seria de 83%.

Outro exemplo mais evidente ainda de uma anisotropia zonal se encontra para a variável Ti400A onde o alcance máximo na vertical (ao longo do furo) é de 100m. Claramente se analisarmos que as amostras estratificadas por furo chegam aproximadamente a 5m na vertical (pois estão de 0.5 em 0.5 m), não é possível encontrar uma continuidade até 100 metros. Portanto, o alcance prático é somente de 2.5m que representa 81% do patamar.

A figura 3.10 e 3.11, mostram um exemplo claro da variável Alap14A (plano XY) e Ti400A (vertical) com uma forte anisotropia zonal e o alcance onde atinge o modelo prático. Os círculos no modelo variográfico mostram como se estabiliza a variância depois de certo alcance, nesse preciso momento onde começa a anisotropia zonal é obtido o modelo prático, tomando como referencia esse alcance para propósitos de estimativa.

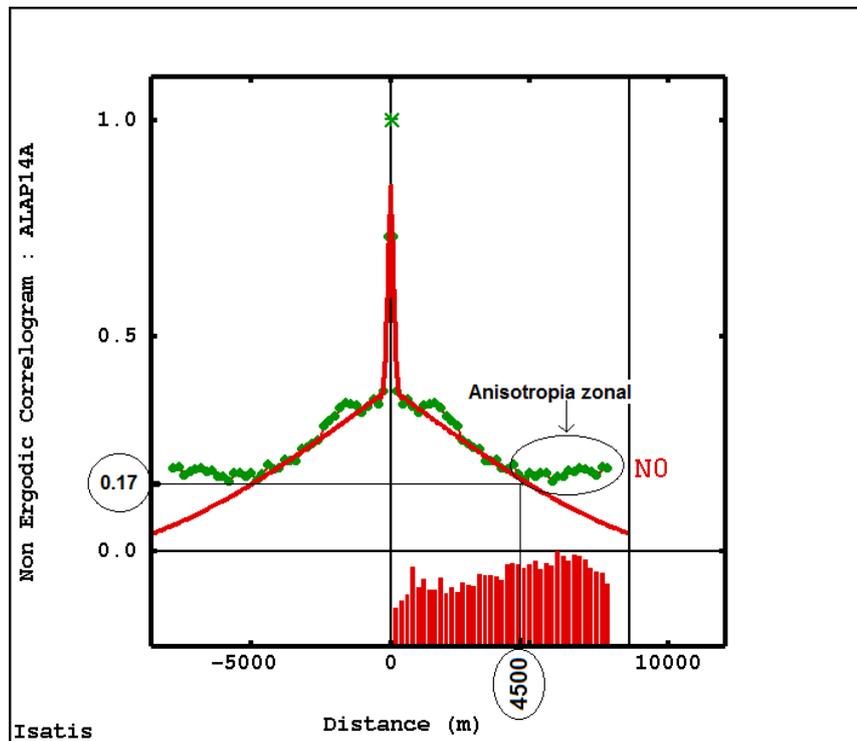


Figura 3.10. Anisotropia zonal da variável ALAP14A no plano XY.

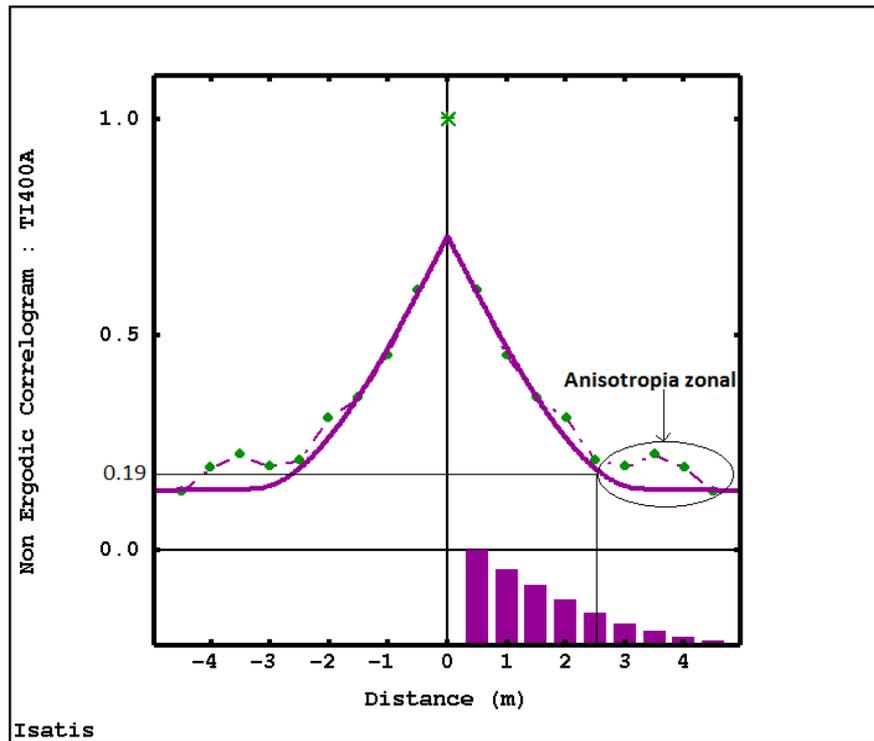


Figura 3.11. Anisotropia zonal da variável Ti400A no plano vertical.

3.9. Análise parcial dos resultados

Nessa parte, somente serão apresentados os resultados em relação ao fechamento dos balanços químico e mássico e os valores mínimos e máximos em cada uma das metodologias estudadas, além de um breve comentário. Mais adiante, no capítulo 5 será feita uma análise detalhada dos métodos de validações utilizadas nas estimativas.

Lembrando que, como estamos trabalhando com variáveis acumuladas é necessário fazer um processo de desacumulação para obter o teor final de cada uma das variáveis analisadas. Esse processo consiste em dividir cada bloco da variável estimada pela sua respectiva estimativa da fração granulométrica que lhe corresponde. Portanto, para obter o teor do Fe na fração +14#, dividimos as estimativas do Fe14A pelas estimativas da Rec14.

3.9.1. Metodologia 1 (isotópico – individuais - krigagem ordinária)

Na tabela 3.12, podemos observar os valores mínimos e máximos das variáveis em relação às estimativas obtidas por krigagem ordinária no banco de dados isotópico, utilizando cada uma o seu próprio modelo de continuidade espacial.

Note que, as estimativas das variáveis Alap14, Resto14, Alap400 e Resto400 se encontram fora do intervalo dos dados originais (mínimo e máximo), isso devido a que cada variável vai receber pesos diferentes em relação ao seu variograma. Como consequência, no momento de fazer-se o processo de desacumulação pode aparecer esse tipo de problema. Embora o número de blocos fora do intervalo é insignificante em relação ao depósito como um todo, não podemos apresentar estimativas finais maiores ou menores que os dados originais o que leva, por exemplo, a um pós-processamento.

Tabela 3.12. Comparação entre os valores mínimos e máximos dos dados estimados e os dados originais (Metodologia 1).

Metodologia 1							
Variável	Blocos estimados	Dados Estimados		Dados Originais		Blocos fora do Intervalo	
		Mínimo	Máximo	Mínimo	Máximo	Nº	%
1_Krig_Alap14	74448	25.61	65.62	13.07	58.78	10	0.01
1_Krig_Fe14	74448	0.79	39.96	0.69	54.65	0	0.00
1_Krig_Sire14	74448	0.79	10.68	0.27	14.9	0	0.00
1_Krig_Ti14	74448	0.32	3.4	0.27	3.57	0	0.000
1_Krig_Resto14	74448	25.69	55.33	21.47	52.32	2	0.003
1_Krig_Alap400	74448	13.54	87.32	0.2	58.32	293	0.39
1_Krig_Fe400	74448	0.98	34.14	0.56	46.46	0	0.00
1_Krig_Sire400	74448	0.73	20.5	0.46	28.9	0	0.00
1_Krig_Ti400	74448	0.38	4.7	0.32	7.13	0	0.00
1_Krig_Resto400	74448	19.29	87.78	24.25	79.24	17	0.02
1_Krig_Rec14	74448	13.83	94.42	3.21	97.52	0	0.00
1_Krig_Rec400	74448	1.95	37.6	0.55	44.86	0	0.00
1_Krig_Rec400	74448	3.55	76.47	1.04	94.14	0	0.00

A figura 3.12 mostra o histograma de fechamento tanto para as variáveis químicas nas frações +14# (a) e +400# (b) como para as variáveis mássicas (c), cada uma delas obtidas pela somatória das estimativas na sua respectiva fração granulométrica. Podemos notar que nos três casos apresentam problemas no fechamento químico e mássico, ou seja, a soma de todas as estimativas em cada faixa é diferente de 100%.

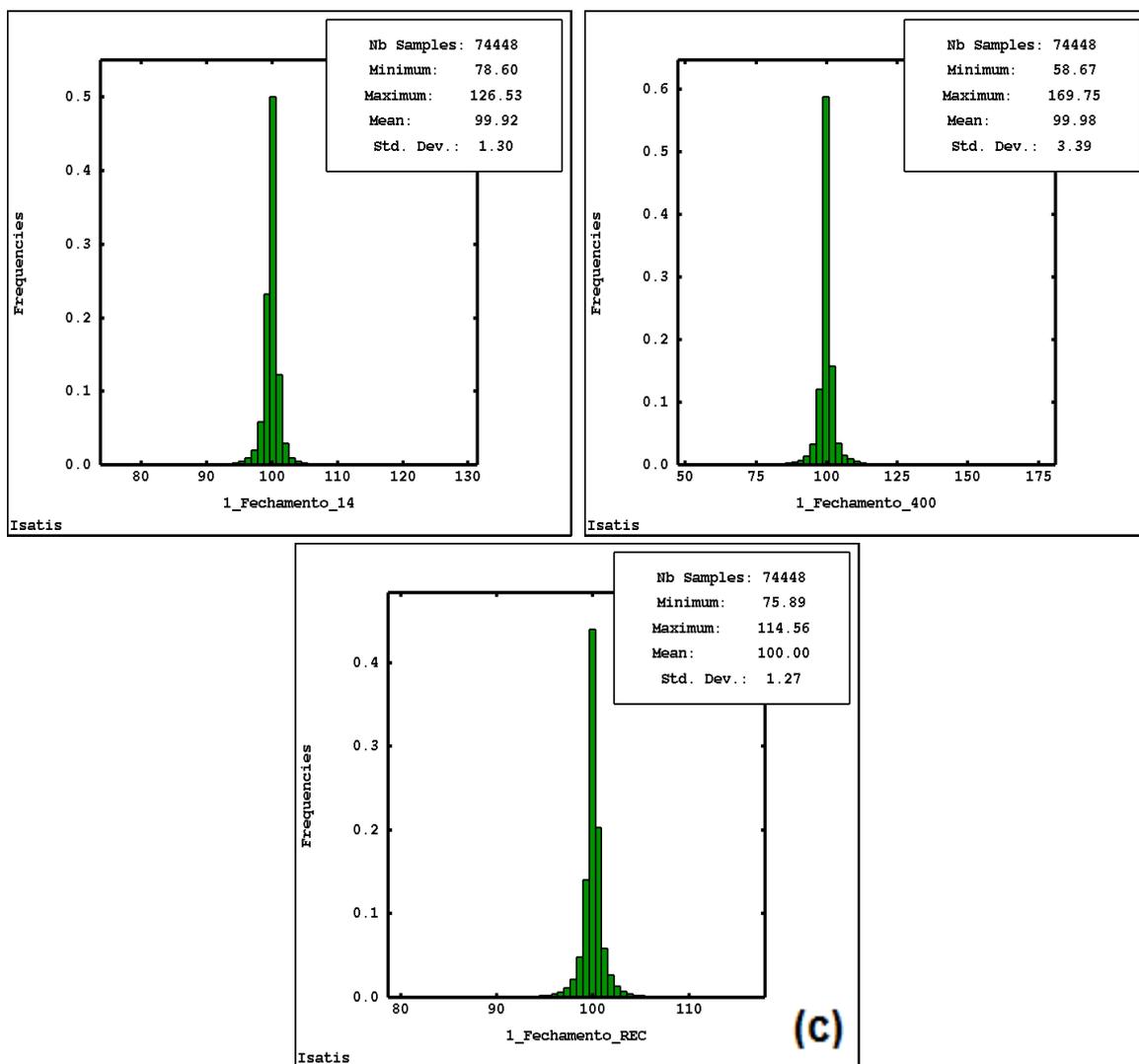


Figura 3.12. Histograma de fechamento metodologia 1, (a) faixa +14#, (b) faixa +400#, (c) frações mássicas.

3.9.2. Metodologia 2 (isotópico – único - krigagem ordinária)

Nesse caso, utilizaremos um modelo variográfico único (Rec14 e Rec400) para estimar as demais variáveis químicas nas suas respectivas frações granulométricas. Para o caso das recuperações, foi utilizado o modelo da Rec14. A tabela 3.13 apresenta os valores mínimos e máximos das variáveis em relação às estimativas obtidas por krigagem ordinária no banco de dados isotópico.

Note que nenhuma estimativa ficou fora do intervalo dos dados originais, devido à utilização de um mesmo modelo variográfico e mesma estratégia de krigagem para todas variáveis. A desvantagem está nos resultados dos modelos das demais variáveis, que não exibem muita aderência aos teores reais do depósito

Tabela 3.13. Comparação entre os valores mínimos e máximos dos dados estimados e os dados originais (Metodologia 2).

Metodologia 2							
Variável	Blocos estimados	Dados Estimados		Dados Originais		Blocos fora do Intervalo	
		Mínimo	Máximo	Mínimo	Máximo	Nº	%
2_Krig_Alap14	74448	25.69	56.8	13.07	58.78	0	0.00
2_Krig_Fe14	74448	1.52	37.98	0.69	54.65	0	0.00
2_Krig_Sire14	74448	0.88	10.11	0.27	14.9	0	0.00
2_Krig_Ti14	74448	0.33	3.21	0.27	3.57	0	0.00
2_Krig_Resto14	74448	25.97	48.51	21.47	52.32	0	0.00
2_Krig_Alap400	74448	15.56	56.28	0.2	58.32	0	0.00
2_Krig_Fe400	74448	1.11	32.38	0.56	46.46	0	0.00
2_Krig_Sire400	74448	0.83	20.23	0.46	28.9	0	0.00
2_Krig_Ti400	74448	0.38	4.57	0.32	7.13	0	0.00
2_Krig_Resto400	74448	27.9	68.11	24.25	79.24	0	0.00
2_Krig_Rec14	74448	13.83	94.42	3.21	97.52	0	0.00
2_Krig_Rec400	74448	1.95	37.67	0.55	44.86	0	0.00
2_Krig_Rec400	74448	3.28	78.16	1.04	94.14	0	0.00

A figura 3.13 mostra o histograma de fechamento para os três balanços e, a diferença da metodologia anterior a somatória das estimativas em cada balanço é 100%.

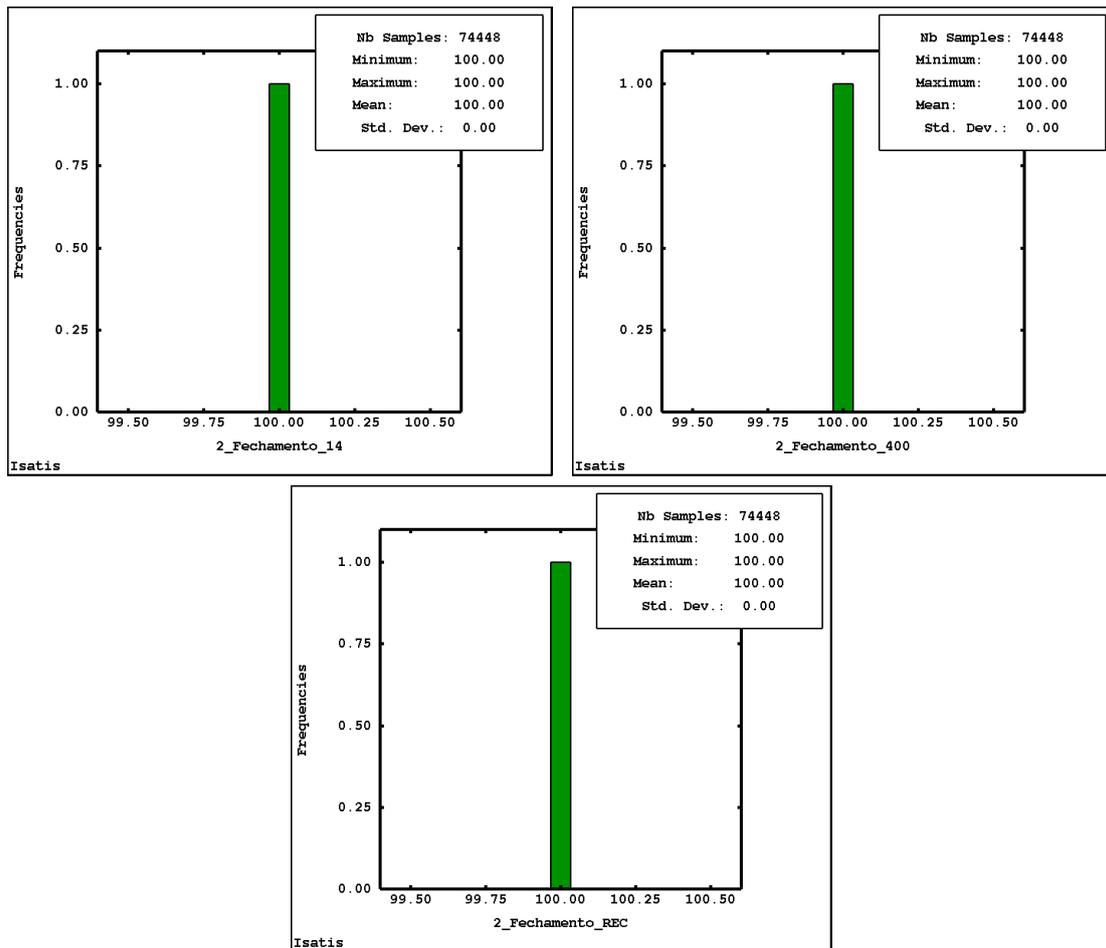


Figura 3.13. Histograma de fechamento metodologia 2, (a) faixa +14#, (b) faixa +400#, (c) frações mássicas.

3.9.3. Metodologia 3 e 4 (heterotópico – único – individual - krigagem ordinária)

As tabelas 3.14 e 3.15 apresentam os mesmos parâmetros da metodologia 1 e 2. A única diferença nesse caso é que o banco de dados é heterotópico (definido anteriormente) excetuando os dados das recuperações que foram mantidos isotópicos. Para ambas as metodologias, as estimativas das variáveis (Alap14, Resto14, Alap400, Fe400 e Resto400) ficaram fora do intervalo dos valores mínimo e máximo dos dados originais, com uma percentagem significativa de blocos fora do intervalo na variável Alap400 de aproximadamente 2.9% e 2.65% respectivamente.

Tabela 3.14. Comparação entre os valores mínimos e máximos dos dados estimados e os dados originais (Metodologia 3).

Metodologia 3							
Variável	Blocos estimados	Dados Estimados		Dados Originais		Blocos fora do Intervalo	
		Mínimo	Máximo	Mínimo	Máximo	Nº	%
3_Krig_Alap14	74448	25.83	134.81	13.07	58.78	236	0.32
3_Krig_Fe14	74448	0.79	40.01	0.69	54.65	0	0.000
3_Krig_Sire14	74448	0.79	13.56	0.27	14.9	0	0.00
3_Krig_Ti14	74448	0.32	3.19	0.27	3.57	0	0.00
3_Krig_Resto14	74448	17.25	59.52	21.47	52.32	41	0.06
3_Krig_Alap400	74448	13.64	127.76	0.2	58.32	2157	2.90
3_Krig_Fe400	74448	0.98	40.24	0.56	36.84	1	0.001
3_Krig_Sire400	74448	0.73	14.72	0.46	24.39	0	0.00
3_Krig_Ti400	74448	0.38	4.84	0.32	7.13	0	0.00
3_Krig_Resto400	74448	17.92	92.97	24.25	68.34	129	0.17
3_Krig_Rec14	74448	13.83	94.42	3.21	97.52	0	0.00
3_Krig_Rec400	74448	1.95	37.6	0.55	44.86	0	0.00
3_Krig_Rec400	74448	3.55	76.47	1.04	94.14	0	0.00

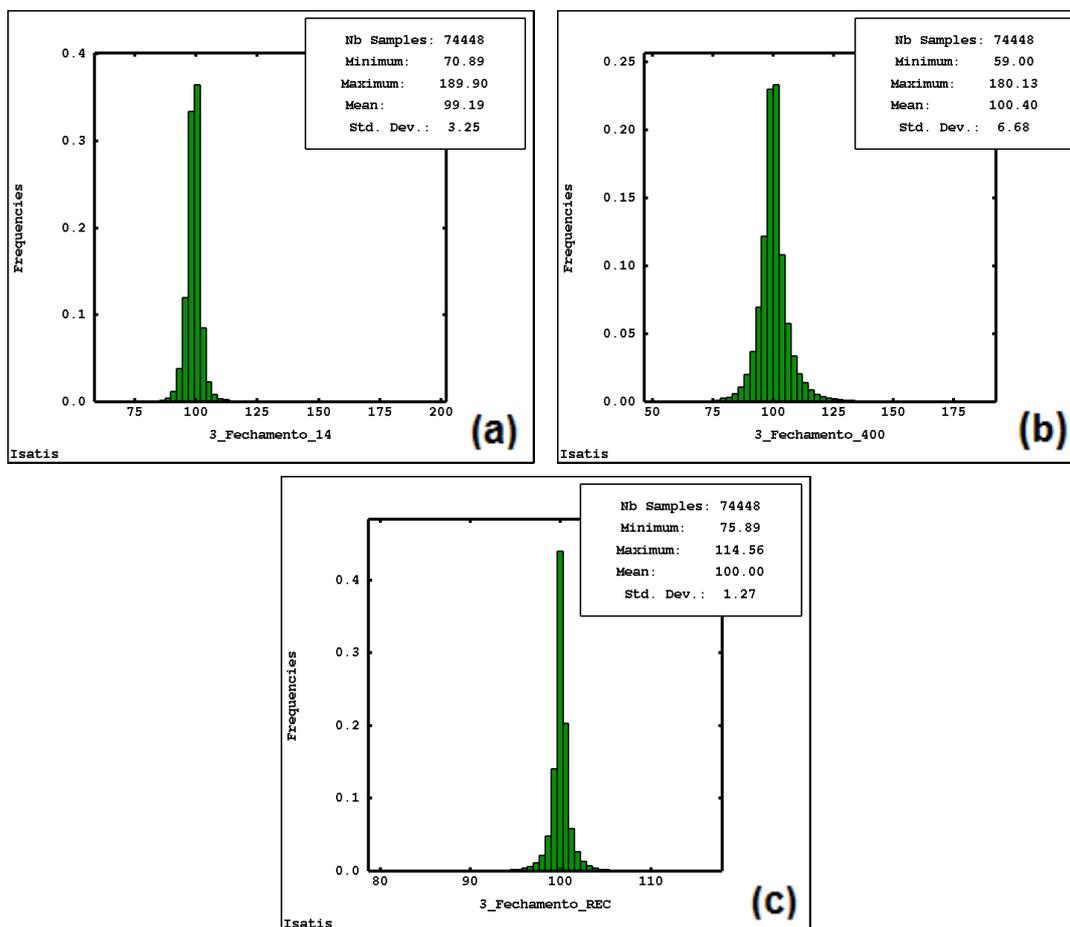


Figura 3.14. Histograma de fechamento metodologia 3, (a) faixa +14#, (b) faixa +400#, (c) frações mássicas.

Tabela 3.15. Comparação entre os valores mínimos e máximos dos dados estimados e os dados originais (Metodologia 4).

Metodologia 4							
Variável	Blocos estimados	Dados Estimados		Dados Originais		Blocos fora do Intervalo	
		Mínimo	Máximo	Mínimo	Máximo	Nº	%
4_Krig_Alap14	74448	25.92	134.48	13.07	58.78	232	0.31
4_Krig_Fe14	74448	1.52	38.07	0.69	54.65	0	0.00
4_Krig_Sire14	74448	0.88	13.5	0.27	14.9	0	0.00
4_Krig_Ti14	74448	0.35	3.15	0.27	3.57	0	0.00
4_Krig_Resto14	74448	15.61	58.98	21.47	52.32	29	0.04
4_Krig_Alap400	74448	15.67	124.9	0.2	58.32	1972	2.65
4_Krig_Fe400	74448	1.11	39.31	0.56	36.84	1	0.00
4_Krig_Sire400	74448	0.8	14.58	0.46	24.39	0	0.00
4_Krig_Ti400	74448	0.38	5.04	0.32	7.13	0	0.00
4_Krig_Resto400	74448	19.07	92.8	24.25	68.34	138	0.19
4_Krig_Rec14	74448	13.83	94.42	3.21	97.52	0	0.00
4_Krig_Rec400	74448	1.95	37.67	0.55	44.86	0	0.00
4_Krig_Rec400	74448	3.28	78.16	1.04	94.14	0	0.00

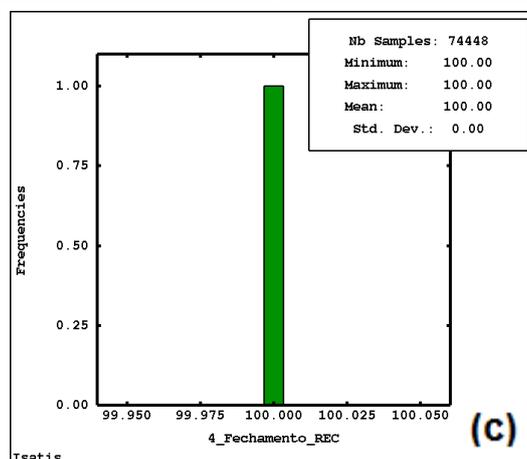
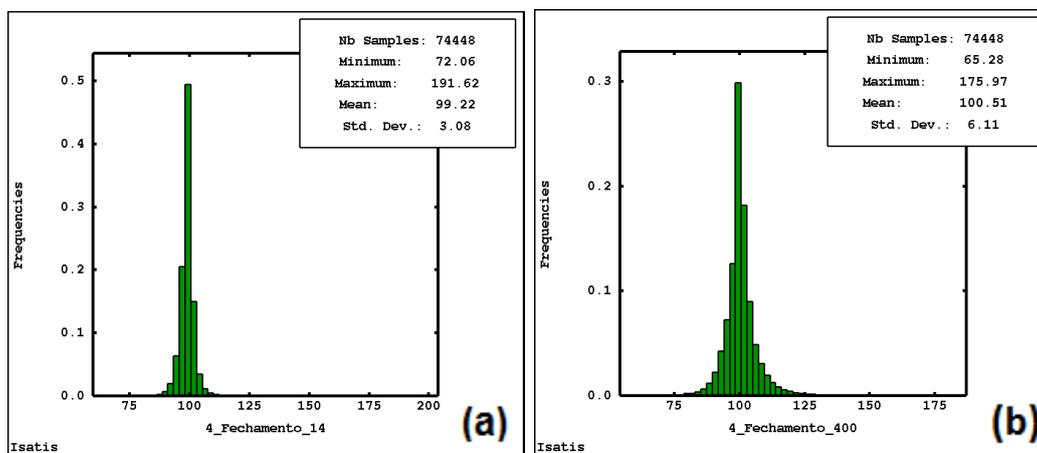


Figura 3.15. Histograma de fechamento metodologia 4, (a) faixa +14#, (b) faixa +400#, (c) frações mássicas.

De acordo com as figuras 3.14 e 3.15 observamos que em ambas as metodologias os histogramas do balanço químico e mássico não fecham 100%, exceto a recuperação da metodologia 4, pois o banco de dados é isotópico e utilizou o mesmo modelo da Rec14 para as demais recuperações.

3.9.4. Metodologia 5 (heterotópico – único – individual - cokrigagem)

Neste caso, utilizaremos a cokrigagem ordinária como método de estimativa tanto para as variáveis químicas em dados heterotópicos como para variáveis mássicas em dados isotópicos. Usaremos um modelo de correionalização intrínseca (MCI) que funciona como uma simplificação do modelo linear de correionalização (MLC), devido à grande quantidade de modelos variográficos diretos e cruzados. Uma leitura mais completa desses métodos está disponível em Wackernagel (1995), Chilès & Delfiner (1999), Goovaerts (1997), entre outros.

Como estamos trabalhando com variáveis acumuladas e, posteriormente teremos que fazer o processo de desacumulação, optou-se por usar o modelo da variável Rec14 na construção de todos os variogramas diretos e cruzados, fazendo somente um reescalamento no sill.

3.9.4.1. Cokrigagem de dados isotópicos (recuperações)

Wackernagel (1995) menciona que a cokrigagem de dados isotópicos é equivalente à krigagem, especialmente quando o determinante da matriz de correionalização é zero. Um caso particular sucede quando os variogramas diretos e cruzados são proporcionais a um mesmo variograma, ou seja, na utilização de um modelo MCI. Isto pode ser verificado na estimativa das recuperações, pois o banco de dados é isotópico.

Nas figuras 3.16 (a, b, c), podemos observar os gráficos de dispersão entre as estimativas das recuperações obtidas por krigagem ordinária e cokrigagem utilizando um modelo único (Rec14) rescalonando somente o sill. Note que o coeficiente de correlação é 1, garantindo que as estimativas são iguais nos dois métodos utilizados. Além disso, ao manter esse mesmo modelo, também se está garantindo o fechamento mássico das recuperações em 100% .

A figura 3.16d, mostra o histograma de fechamento obtido por cokrigagem (metodologia 5). A desvantagem novamente aqui é que os modelos variográficos não se ajustam adequadamente aos variogramas diretos e cruzados.

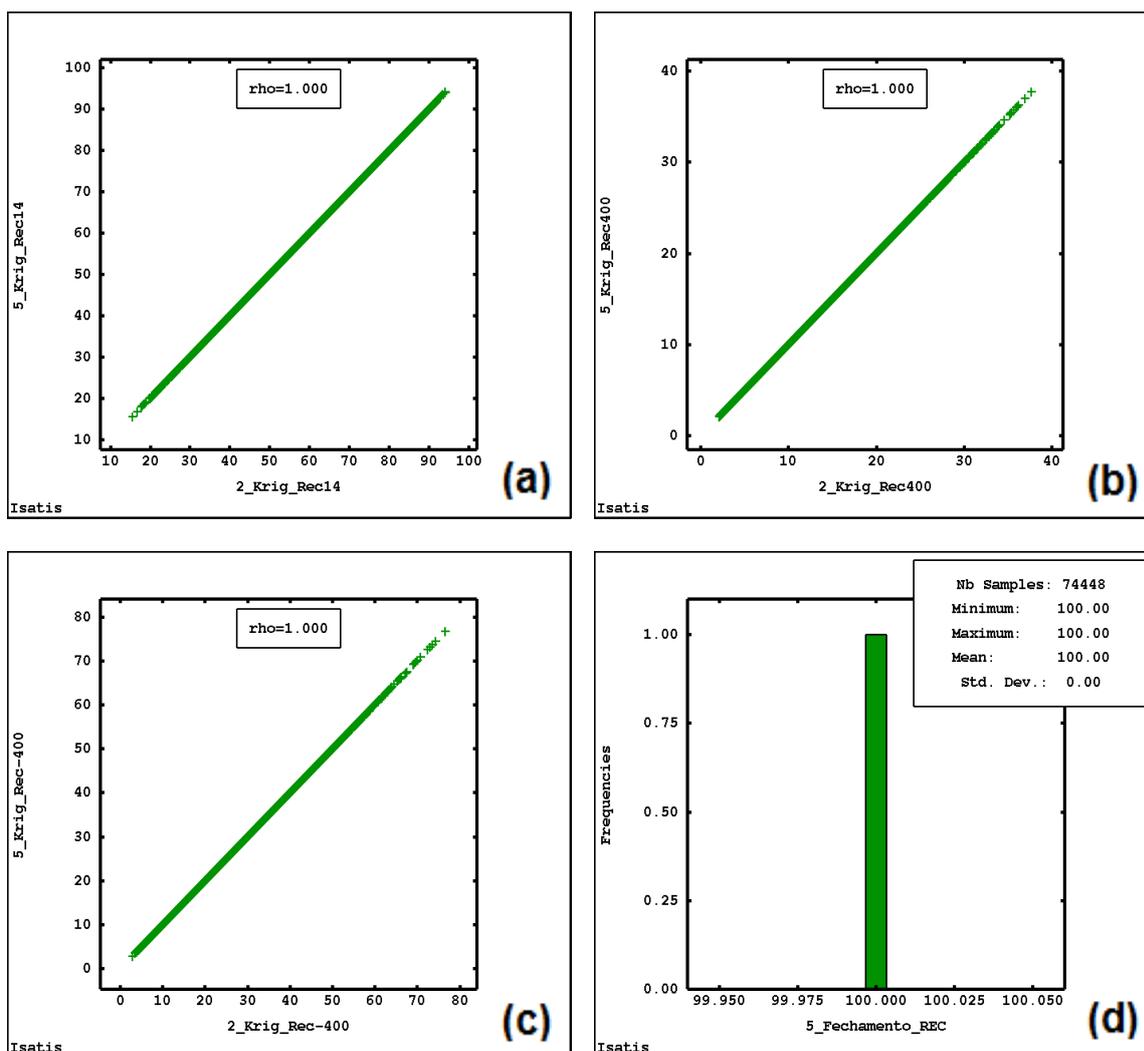


Figura 3.16. Diagramas de dispersão da krigagem ordinária comparada com cokrigagem utilizando o modelo da REC14, (a) Rec14, (b) Rec400, (c) Rec-400, (d) Histograma de fechamento Metodologia 5.

A tabela 3.16, mostra os parâmetros utilizados na criação do modelo MCI para cada uma das recuperações (+14#, +400# e -400#). Note que como foi utilizado o modelo de correionalização intrínseca da Rec14, todos os variogramas cruzados são proporcionais aos variogramas diretos.

Tabela 3.16. Parâmetros das contribuições e o efeito pepita do modelo MCI nas recuperações (Metodologia 5).

	Rec14 vs Rec400	Rec14 vs Rec-400	Rec400 vs Rec-400	Sill MCI Rec14, Rec400 e Rec-400
ρ	-0.67	-0.95	0.42	-
Efeito pepita	-0.1005	-0.1425	0.063	0.15
Sill 1	-0.3015	-0.4275	0.189	0.45
Sill 2	-0.268	-0.38	0.168	0.4

3.9.4.2. Cokrigagem de dados heterotópicos (químicas)

A diferença da krigagem, que utiliza apenas os valores das amostras para estimar uma variável de interesse, a cokrigagem incorpora informação secundária correlacionada com outras variáveis para ajudar a estimar nos lugares onde a variável principal não foi amostrada.

As tabelas 3.17 e 3.18 mostram os coeficientes de correlações (ρ) das variáveis acumuladas nas duas frações granulométricas (+14# e +400#), considerando as variáveis que têm uma boa ou uma aceitável correlação aquelas que possuam um $\rho \geq 0.55$.

Tabela 3.17. Coeficiente de Correlação (Metodologia 5) da Fração Granulométrica +14 Acumulada.

	Alap14A	Fe14A	Sire14A	Ti14A	Resto14A
Alap14A		0.4	-0.08	0.75	0.92
Fe14A	0.4		-0.19	0.43	0.38
Sire14A	-0.08	-0.19		0.004	0.2
Ti14A	0.75	0.43	0.004		0.74
Resto14A	0.92	0.38	0.2	0.74	

Tabela 3.18. Coeficiente de Correlação (Metodologia 5) da Fração Granulométrica +400 Acumulada.

	Alap400A	Fe400A	Sire400A	Ti400A	Resto400A
Alap400A		0.44	0.76	0.59	0.95
Fe400A	0.44		0.2	0.34	0.39
Sire400A	0.76	0.2		0.55	0.82
Ti400A	0.59	0.34	0.55		0.61
Resto400A	0.95	0.39	0.82	0.61	

Na tabela 3.19, podemos observar as estimativas obtidas por cokrigagem ordinária no banco de dados heterotópico, usando a mesma estratégia de busca da metodologia 4. Note, que para este caso, todas as estimativas ficaram fora dos intervalos mínimo e máximo excetuando as recuperações mássicas por serem isotópicas. Além disso, obteve-se uma maior percentagem de blocos fora do intervalo em algumas variáveis, caso a destacar das variáveis Ti14 e Ti400 que apresentam a maior percentagem de 8.58% e 11.21% respectivamente.

Tabela 3.19. Comparação entre os valores mínimos e máximos dos dados estimados e os dados originais (Metodologia 5).

Metodologia 5								
Variável	Blocos estimados	Dados Estimados		Dados Originais		Blocos fora do Intervalo		
		Mínimo	Máximo	Mínimo	Máximo	Nº	%	% -
5_Krig_Alap14	74448	25.79	73.13	13.07	58.78	43	0.06	0
5_Krig_Fe14	74448	0.62	38.16	0.69	54.65	1	0.00	0
5_Krig_Sire14	74448	-12.34	48.93	0.27	14.9	2330	3.13	2.45
5_Krig_Ti14	74448	-19.61	18.28	0.27	3.57	6389	8.58	3.25
5_Krig_Resto14	74448	22.97	57.4	21.47	52.32	4	0.005	0
5_Krig_Alap400	74448	16.06	80.63	0.2	58.32	277	0.37	0
5_Krig_Fe400	74448	-33.54	30.11	0.56	36.84	1767	2.37	2.01
5_Krig_Sire400	74448	-47.67	30.79	0.46	24.39	3348	4.50	3.87
5_Krig_Ti400	74448	-32.14	14.53	0.32	7.13	8346	11.21	8.20
5_Krig_Resto400	74448	19.25	87.15	24.25	68.34	37	0.05	0
5_Krig_Rec14	74448	13.83	94.42	3.21	97.52	0	0.00	0
5_Krig_Rec400	74448	1.95	37.67	0.55	44.86	0	0.00	0
5_Krig_Rec400	74448	3.28	78.16	1.04	94.14	0	0.00	0

Um caso atípico que podemos notar nesta metodologia é que foram geradas estimativas negativas com uma percentagem (%-) alta nas variáveis Sire14, Ti14, Fe400, Sire400 e Ti400. Uma das desvantagens apresentadas por Goovaerts (1997) na cokrigagem ordinária tradicional é que a soma dos pesos da variável secundária é zero, portanto, provavelmente alguns dados vão receber pesos negativos incrementando o risco de obter estimativas negativas. Para tratar de atenuar esse problema e obter o menor número possível de blocos negativos, foram criados 3 cenários diferentes para 6 testes de cokrigagem.

Os três cenários foram:

1. Usando a mesma vizinhança do modelo da Rec14, ou seja, uma busca até 4800 metros no plano XY e 4 metros na vertical.
2. Fazendo duas cokrigagens, uma primeira busca até 600 metros no plano XY e 1.7 metros na vertical que corresponde aproximadamente 64% da variância e a outra até o alcance da Rec14 para estimar os blocos que não foram estimados na primeira passada.
3. Fazendo duas cokrigagens, uma busca até 400 metros no plano XY e 1.6 metros na vertical que corresponde aproximadamente 62% da variância e a outra até o alcance da Rec14.

Os testes da cokrigagem foram baseados na quantidade de amostras utilizadas por octante e quadrante: máximo 3 amostras por octante e máximo 3 amostras por quadrante.

A tabela 3.19, mostra a percentagem de blocos negativos estimados pelas 6 metodologias descritas anteriormente. Note que não existe um cenário padrão na redução de blocos negativos em todas as variáveis, por exemplo, o cenário que produz o menor número de blocos negativos na variável Ti400 é o cenário 2 utilizando busca por octante (5.61%), mas gera 0.02% de blocos negativos na estimativa da variável Fe14 que em outros cenários não tinham, por exemplo, cenário 1 utilizando a mesma busca por octante.

Tabela 3.20. Percentagem de blocos negativos obtidos pelos 6 testes da Cokrigagem.

Variável	% Blocos - por Quadrante			% Blocos - por Octante		
	Cenário 1	Cenário 2	Cenário 3	Cenário 1	Cenário 2	Cenário 3
Krig_Alap14	0	0	0	0	0	0
Krig_Fe14	0	0.01	0.02	0	0.02	0.02
Krig_Sire14	2.45	1.21	1.00	2.02	1.04	0.99
Krig_Ti14	3.25	2.39	2.27	2.61	2.20	2.16
Krig_Resto14	0	0	0	0	0	0
Krig_Alap400	0	0	0	0	0	0
Krig_Fe400	2.01	1.30	1.35	1.47	1.09	1.19
Krig_Sire400	3.87	2.36	2.06	3.25	2.13	2
Krig_Ti400	8.20	6.27	6.12	6.84	5.61	5.78
Krig_Resto400	0	0	0	0	0	0
Krig_Rec14	0	0	0	0	0	0
Krig_Rec400	0	0	0	0	0	0
Krig_Rec400	0	0	0	0	0	0

A presença de pesos negativos e, portanto, estimativas negativas na cokrigagem ordinária tradicional é proporcional ao número de variáveis utilizadas na estimativa. Goovaerts (1997) menciona que uma maneira de atenuar esse efeito dos pesos negativos é fazendo uma cokrigagem ordinária não tradicional, na qual a soma dos pesos da variável primária e secundária é um (1). Até o momento, não há software que possa trabalhar com mais de 3 variáveis simultaneamente.

A operação de fechamento é dada pela somatória de todas as variáveis estimadas em cada faixa granulométrica ou nas espécies químicas, no entanto, a presença de alguns blocos com estimativas negativas inviabilizou a utilização dessa operação.

Capítulo 4

Transformações logarítmicas isométricas (ilr)

Esse capítulo apresenta a análise de dados composicionais (CODA) utilizando as transformações logarítmicas isométricas (*ilr*) como uma alternativa para lidar com estimativas cuja finalidade seja o balanço (fechamento) granulométrico e/ou das espécies químicas em dados multivariados. As estimativas dessas transformações são obtidas a partir da krigagem ordinária e, por último, é feito um processo de retro-transformação (ilr^{-1}) que traz novamente os dados estimados ao espaço original. Um procedimento similar é aplicado para a simulação seqüencial gaussiana e a simulação por bandas rotativas.

4.1. Transformações logarítmicas isométricas (ilr)

Sabe-se que, quando trabalhamos com dados multivariados (múltiplas variáveis) e, a soma desses atributos deva fechar uma constante, geralmente 1 ou 100%, surge a necessidade de manter essa condição (fechamento) após estimativas. Ou seja, todos os blocos estimados devem ter com os valores estimados o fechamento do balanço químico ou mássico das variáveis analisadas em cada faixa.

Nessa dissertação, utilizaremos as transformações logarítmicas isométricas (*ilr*) comentadas anteriormente no capítulo 2. A idéia principal é utilizar essas transformações para fazer todo o processo de análise da continuidade espacial e estimativa. No final, esses valores serão retro-transformados para voltar ao espaço dos dados originais pela sua função inversa (ilr^{-1}).

As etapas utilizadas para fazer as transformações logarítmicas isométricas (*ilr*) foram:

1. Fechamento do banco de dados: O banco de dados tinha problema no fechamento inicial dos dados, tanto para as variáveis químicas na fração granulométrica +14# como na +400# como na massa retida em cada peneira, ou seja:

$$Alap_{14} + Fe_{14} + Sire_{14} + Ti_{14} \neq 100\% \text{ (Fração + 14)}$$

$$Alap_{400} + Fe_{400} + Sire_{400} + Ti_{400} \neq 100\% \text{ (Fração + 400)}$$

$$Rec14 + Rec400 + Rec-400 \neq 100\% \text{ (massa retida em cada peneira)}$$

Para o caso das espécies químicas, foi criada uma variável nova que carregasse todo o erro de fechamento chamado Resto14 (na fração +14#) e Resto400 (na fração +400#). A razão principal de sua utilização é porque nesse banco de dados em particular não foram analisadas todas as espécies químicas que levassem ao fechamento, portanto, o novo balanço é dado da seguinte maneira:

$$Alap_{14} + Fe_{14} + Sire_{14} + Ti_{14} + Resto_{14} = 100\%$$

$$Alap_{400} + Fe_{400} + Sire_{400} + Ti_{400} + Resto_{400} = 100\%$$

Para o caso das recuperações (Rec14, Rec400 e Rec-400), foi feita uma distribuição proporcional do erro, que consiste em distribuir o erro faltante ou sobranse para fechar 100%. Esse erro é distribuído em todas as variáveis proporcionalmente de acordo à equação (2.26) definida no capítulo 2 como:

$$C(Z) = \left[\frac{k * Z_1}{\sum_{i=1}^D Z_i}, \frac{k * Z_2}{\sum_{i=1}^D Z_i}, \dots, \frac{k * Z_D}{\sum_{i=1}^D Z_i} \right] \quad (2.26)$$

Onde:

- $C(Z)$ = fechamento;
- k = constante de fechamento (geralmente 100%);
- Z_i = Valor da i -ésima amostra.

2. Criação da base ortonormal: A base ortonormal é obtida através de uma partição sequencial binária, definida por Pawlowsky (2005) como uma hierarquia das partes de uma composição. Na primeira ordem da hierarquia, todas as partes são divididas em dois grupos binários (+1 e -1). Nas etapas seguintes, cada grupo é dividido em dois novos grupos e o processo continua até que todos os grupos têm uma única parte. O número de partições binárias ao final do processo é D-1 (onde D é o número de dimensões que correspondem ao número de variáveis analisadas por cada faixa).

Na tabela 4.1, observa-se a partição sequencial binária utilizada nesse estudo, onde r(+) é dada pela somatória das partições binárias positivas e s(-) é dada pela somatória das partições binárias negativas. Note que, essa partição não faz referência aos teores e sim ao número de variáveis utilizadas, portanto, para cada fração química e mássica, o número de partições binárias foi reduzido numa dimensão, ou seja, a frações +14# e +400# ficaram com 4 partições das 5 espécies químicas analisadas. Para o caso das recuperações, há somente duas partições. O número de partições é relevante no processo de estimativa porque reduz numa dimensão o número de variáveis a ser estimadas por cada faixa.

Tabela 4.1. Partição sequencial binária para cada fração (+14#, +400#, -400#).

No. de Partições	Fração +14#					r(+)	s(-)
	Alap14	Fe14	Sire14	Ti14	Resto14		
1	+1	+1	+1	-1	-1	3	2
2	+1	+1	-1	0	0	2	1
3	+1	-1	0	0	0	1	1
4	0	0	0	+1	-1	1	1
No. de Partições	Fração +400#					r(+)	s(-)
	Alap400	Fe400	Sire400	Ti400	Resto400		
1	+1	+1	+1	-1	-1	3	2
2	+1	+1	-1	0	0	2	1
3	+1	-1	0	0	0	1	1
4	0	0	0	+1	-1	1	1
No. de Partições	Recuperações			r(+)	s(-)		
	Rec14	Rec400	Rec-400				
1	+1	+1	-1	2	1		
2	+1	-1	0	1	1		

3. Transformação das coordenadas ilr : Para obter a nova distribuição dos dados transformados, utiliza-se a seguinte expressão definida por Egozcue & Pawlowsky (2003) como:

$$ilr_i = \sqrt{\frac{r * s}{r + s}} \ln \left[\frac{(x_{i1}x_{i2} \dots x_{ir})^{1/r}}{(x_{j1}x_{j2} \dots x_{js})^{1/s}} \right] \quad (4.1)$$

Onde:

- ilr_i = transformação da nova coordenada;
- $r(+)$ = soma dos 1s (+1) positivos (calculados anteriormente);
- $s(-)$ = soma dos 1s (-1) negativos (calculados anteriormente);
- $(x_{i1}x_{i2} \dots x_{ir})^{1/r}$ = média geométrica das variáveis que foram selecionadas com (+1) na partição sequencial binária;
- $(x_{j1}x_{j2} \dots x_{js})^{1/s}$ = média geométrica das variáveis que foram selecionadas com (-1) na partição sequencial binária.

Essa nova transformação é aplicada diretamente aos teores em cada faixa cuja soma é 100%, tendo em conta a partição seqüencial binária obtida na Tabela 4.1.

Na figura 4.1, há um exemplo do procedimento para calcular as transformações ilr para uma amostra (amostra 1). Dividiremos a equação (4.1) em duas partes para explicar com detalhe o cálculo dessas transformações.

Para calcular a primeira parte da equação (4.1), definida como $\sqrt{(2 * 1)/(2 + 1)}$, devemos somar o número de vezes que aparecem +1 e -1 definidos na primeira partição binária (ver figura 4.1). Na segunda parte da equação (4.1), definida como $[\ln [(20 * 30)^{1/2}/(50)^{1/1}]]$ devemos aplicar a média geométrica utilizando como referência o valor dos teores definidos na primeira partição, ou seja, a primeira média geométrica $((20 * 30)^{1/2})$ é referente aos teores que foram carimbados como +1 na partição binária e, a

segunda média geométrica $((50)^{1/1})$ é referente aos teores que foram carimbados como -1 na mesma partição. A multiplicação dessas duas partes leva a obter a primeira transformação logarítmica isométrica ilr_{Rec_1} . O procedimento é similar para o cálculo da segunda transformação ilr_{Rec_2} , mudando unicamente a referência da partição binária.

Amostra	Rec14	Rec400	Rec-400
1	20	30	50
2	40	30	30
3	50	10	40
.	.	.	.
.	.	.	.
.	.	.	.

No. de Partições	Recuperações			r(+)	s(-)
	Rec14	Rec400	Rec-400		
1	+1	+1	-1	2	1
2	+1	-1	0	1	1

$$ilr_{REC_1} = \sqrt{\frac{2 * 1}{2 + 1}} \ln \left[\frac{(20 * 30)^{1/2}}{(50)^{1/1}} \right]$$

$$ilr_{REC_1} = -0.57$$

$$ilr_{REC_2} = \sqrt{\frac{1 * 1}{1 + 1}} \ln \left[\frac{(20)^{1/1}}{(30)^{1/1}} \right]$$

$$ilr_{REC_2} = -0.28$$

Figura 4.1. Processo de cálculo das transformações ilr .

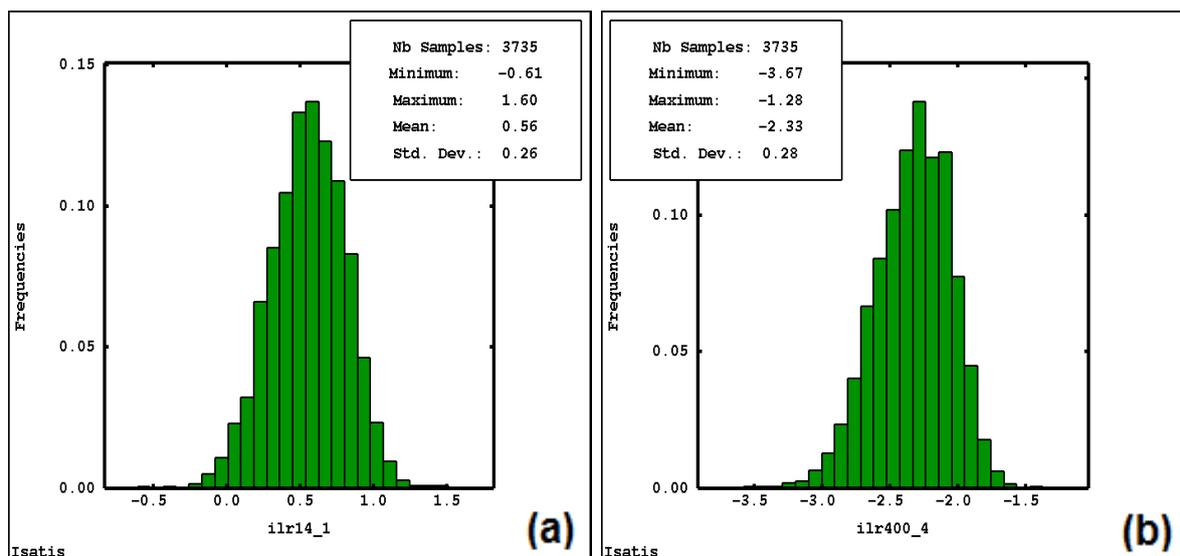
Como já mencionamos anteriormente, o novo espaço é reduzido uma dimensão (D-1), portanto, das 13 variáveis (5 espécies químicas em 2 faixas granulométricas mais a massa retida em 3 faixas granulométricas) que tínhamos inicialmente no espaço original, agora passamos a ter somente 10 variáveis. A tabela 4.2 mostra a estatística básica de cada variável transformada e nova nomenclatura utilizada.

Note que o novo espaço amostral possui uma estatística básica totalmente diferente ao espaço original dos dados, incluindo até números negativos. A razão principal é porque estamos trabalhando com transformações tipo logarítmicas que faz com que gere esse tipo de resultados.

Tabela 4.2. Estatística básica das transformações logarítmicas isométricas (ilr).

	Variável	Dados	Mínimo	Máximo	Média	Desvio Padrão	Variância	Q25	Q50	Q75
Fração +14	ilr14_1	3735	-0.61	1.6	0.56	0.26	0.07	0.38	0.57	0.73
	ilr14_2	3735	-0.28	3.88	1.24	0.54	0.29	0.86	1.19	1.56
	ilr14_3	3735	-1.01	2.98	1.29	0.44	0.2	0.99	1.25	1.57
	ilr14_4	3735	-3.65	-1.58	-2.28	0.25	0.06	-2.44	-2.26	-2.11
Fração +14	ilr400_1	3735	-1.39	1.46	0.5	0.29	0.09	0.32	0.52	0.7
	ilr400_2	3735	-2.48	3.46	1.15	0.52	0.27	0.8	1.08	1.43
	ilr400_3	3735	-2.73	3.22	1.3	0.49	0.24	0.97	1.27	1.6
	ilr400_4	3735	-3.67	-1.28	-2.33	0.28	0.08	-2.52	-2.31	-2.13
Rec	ilrRec_1	3735	-3.27	2.32	0.14	0.54	0.29	-0.22	0.09	0.49
	ilrRec_2	3735	-1.28	3.12	1.47	0.54	0.3	1.16	1.51	1.82

Ao trabalhar nesse novo espaço amostral não teremos problemas em criar estimativas negativas, já que estamos trabalhando com transformações logarítmicas. Mais adiante, quando for explicado o processo de retro-transformação (ilr^{-1}), as transformações voltarão ao espaço dos dados originais mantendo as estimativas coerentes. A figura 4.2 mostra os histogramas de freqüências da nova distribuição dos teores das variáveis *ilr14_1* e *ilr400_4* nesse novo espaço transformado.

Figura 4.2. Histogramas das transformações ilr, (a) *ilr14_1*, (b) *ilr400_4*.

Para o caso das transformações (ilr), foi aplicado o mesmo procedimento de acumulação definido anteriormente no capítulo 3, lembrando que para cada fração granulométrica foi reduzida uma dimensão, assim, as frações +14# e +400# possuem 4 variáveis cada uma. Além disso, foi adicionada a letra A ao final para representar que a variável está acumulada. Na tabela 4.3, foi feito um resumo estatístico das variáveis transformadas (ilr) acumuladas.

Tabela 4.3. Resumo estatístico das variáveis transformadas (ilr) acumuladas.

Variável	Dados	Mínimo	Máximo	Média	Desvio Padrão	Variância	Q25	Q50	Q75
ilr14_1A	3735	-38.76	109.52	36.39	17.53	307.3	24.66	36.39	48
ilr14_2A	3735	-6.16	371.67	88.92	50.28	2527.61	52.01	83.08	120.17
ilr14_3A	3735	-66.56	242.6	85.43	31.87	1015.76	64.09	82.07	104.39
ilr14_4A	3735	-253.13	-7.95	-152.3	31.87	1015.72	-173.96	-156.09	-136.38
ilr400_1A	3735	-12.02	36.43	4.92	4.47	20.01	1.9	3.9	6.83
ilr400_2A	3735	-29.99	67.24	10.61	7.94	63.12	5.37	8.53	13.43
ilr400_3A	3735	-33.02	104.31	12.96	11	121.02	6.07	10.02	16.21
ilr400_4A	3735	-130.78	-1.36	-22.48	14.64	214.27	-28.38	-18.58	-12.59

O processo de acumulação leva a novas variáveis com novas distribuições de probabilidade. A figura 4.3 mostra os histogramas de freqüências das variáveis ilr14_1A e ilr400_4A nesse novo espaço que foi transformado e acumulado pela recuperação. Note que, nesse caso a variável ilr400_4A foi a que teve maior impacto na distribuição dos dados transformados, apresentado uma assimetria negativa mais forte em comparação com os dados sem acumular. No caso da variável, ilr14_1A foi mantida a mesma simetria da distribuição original.

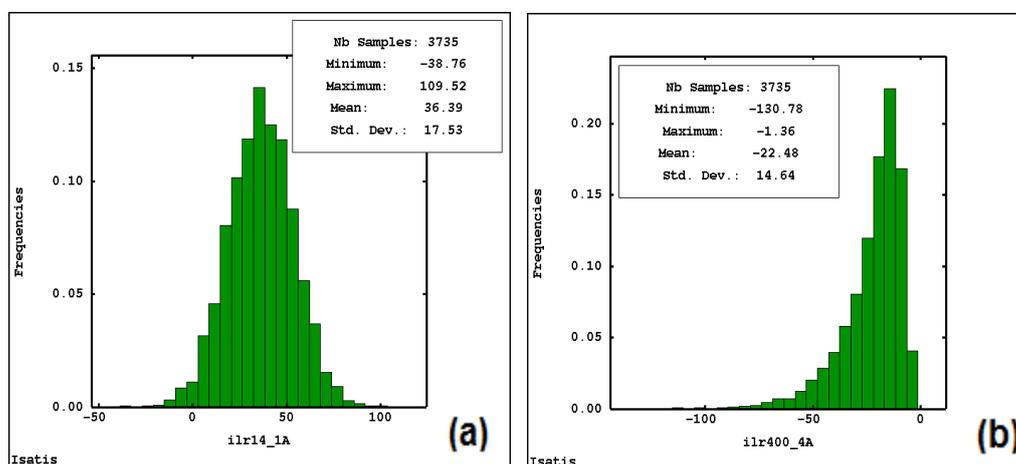


Figura 4.3. Histogramas das transformações ilr acumuladas, (a) ilr14_1A, (b) ilr400_4A.

4.2. Continuidade espacial

A modelagem da continuidade espacial das transformações logarítmicas isométricas (ilr) é baseada no uso de correlogramas não ergódicos. No caso das espécies químicas que foram transformadas e, posteriormente acumuladas pela fração correspondente em cada faixa, foi necessário o modelamento de cada variável acumulada para levar em consideração o fato de as análises químicas não terem o mesmo suporte amostral. No caso das frações granulométricas, cada modelo é baseado somente nas transformações, já que não é usado o processo de acumulação.

Para obter as direções principais de continuidade no plano XY, foram realizados 8 correlogramas nesse plano, variando de 0° até 157° com um intervalo de 22.5° em cada direção. Os demais parâmetros utilizados na variografia são os mesmos utilizados no capítulo anterior (ver tabela 3.7 do capítulo 3).

A figura 4.4 mostra o correlograma experimental nas 8 direções principais do plano XY e o modelo ajustado da variável ilr14_2A.

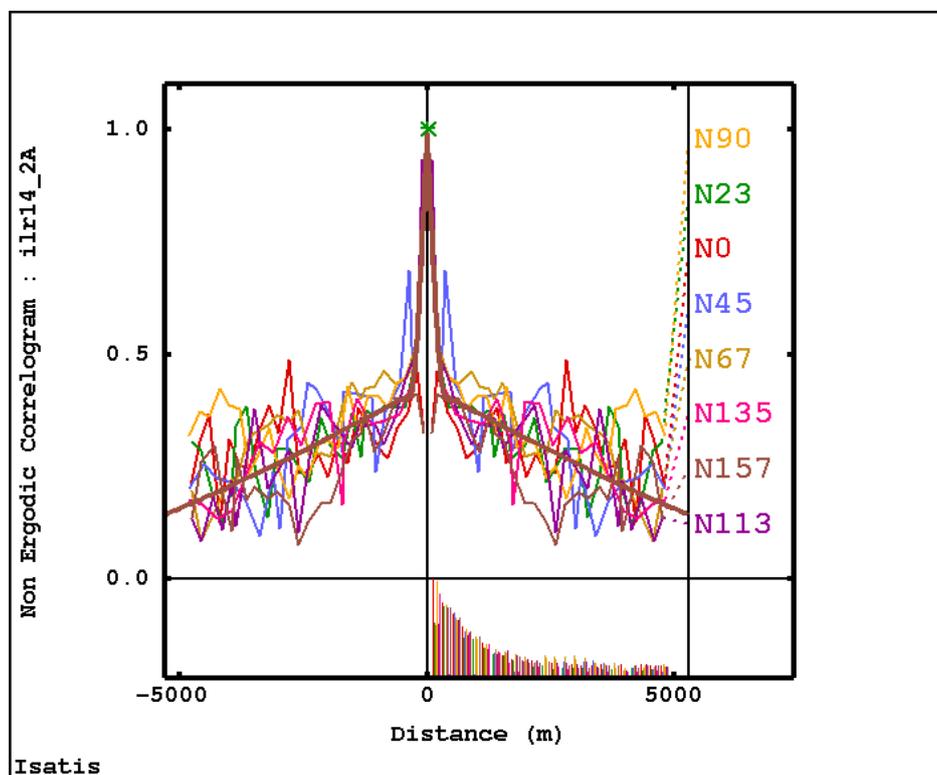


Figura 4.4. Correlograma experimental ilr14_2A nas 8 direções principais e o modelo ajustado.

Note, que o comportamento do correlograma em todas as direções é muito similar, ou seja, as variáveis acumuladas não apresentaram uma anisotropia claramente definida. Portanto, para modelar as direções principais foi considerado um correlograma omnidirecional no plano horizontal e a direção de menor continuidade é vertical.

Todos os modelos foram definidos com duas estruturas, utilizando em cada estrutura um modelo esférico. Na tabela 4.4, é apresentado um resumo detalhado dos alcances de cada variável, o efeito pepita e as contribuições de cada estrutura. Note que, existem algumas variáveis que apresentam uma forte anisotropia zonal, por exemplo, a variável ilr14_4A possui 11.000m de alcance no plano XY e a variável ilr1400_2A possui 300m de alcance na vertical. No entanto, para as distâncias de busca utilizadas nas estimativas, são utilizados alcances limitados (práticos).

Tabela 4.4. Resumo dos parâmetros encontrados na modelagem da variografia.

Variável	Modelo de Continuidade										
	c0	Primeira Estrutura					Segunda Estrutura				
		Modelo	c1	a (D1)	a (D2)	a (D3)	Modelo	c2	a (D1)	a (D2)	a (D3)
ilr14_1A	0.15		0.57	350	350	2.2		0.28	6500	6500	40
ilr14_2A	0.01		0.56	350	350	4.3		0.43	11000	11000	4.4
ilr14_3A	0.15	ESFÉRICO	0.67	350	350	2.2	ESFÉRICO	0.18	6500	6500	5
ilr14_4A	0.2		0.54	350	350	2		0.26	11000	11000	9.5
ilr400_1A	0.05		0.46	250	250	2.4		0.49	4200	4200	30
ilr400_2A	0.08		0.59	250	250	2.7		0.33	3200	3200	300
ilr400_3A	0.09		0.5	400	400	4		0.41	6000	6000	11.5
ilr400_4A	0.1		0.41	300	300	3.5		0.49	4900	4900	12
ilrRec_1	0.15		0.6	300	300	2.8		0.25	9500	9500	4.8
ilrRec_2	0.15		0.43	300	300	3.1		0.42	6500	6500	26

D1 – Direção de Maior Continuidade – N 0° Dip 0°.

D2 – Direção de Continuidade Intermediária - N 90° Dip 0°.

D3 – Direção de Menor Continuidade - N 0° Dip 90°.

c0 - Efeito pepita.

a – Alcance.

c1 – Patamar ou Contribuição da Primeira Estrutura.

c2 – Patamar ou Contribuição da Segunda Estrutura.

OBS 1: Foi utilizado como medida de continuidade espacial o Correlograma não Ergódico.

OBS 2: Os parâmetros dos modelos de continuidade consideram patamar total igual a 1.

A figura 4.5 mostra o modelamento dos correlogramas omnidirecionais da variável *ilr14_2A* nas principais direções do plano horizontal (a, b) e a direção de menor continuidade na vertical (c). Além disso, podemos observar que na mesma figura, o mapa variográfico no plano XY (d), mostra uma clara isotropia em todas as direções.

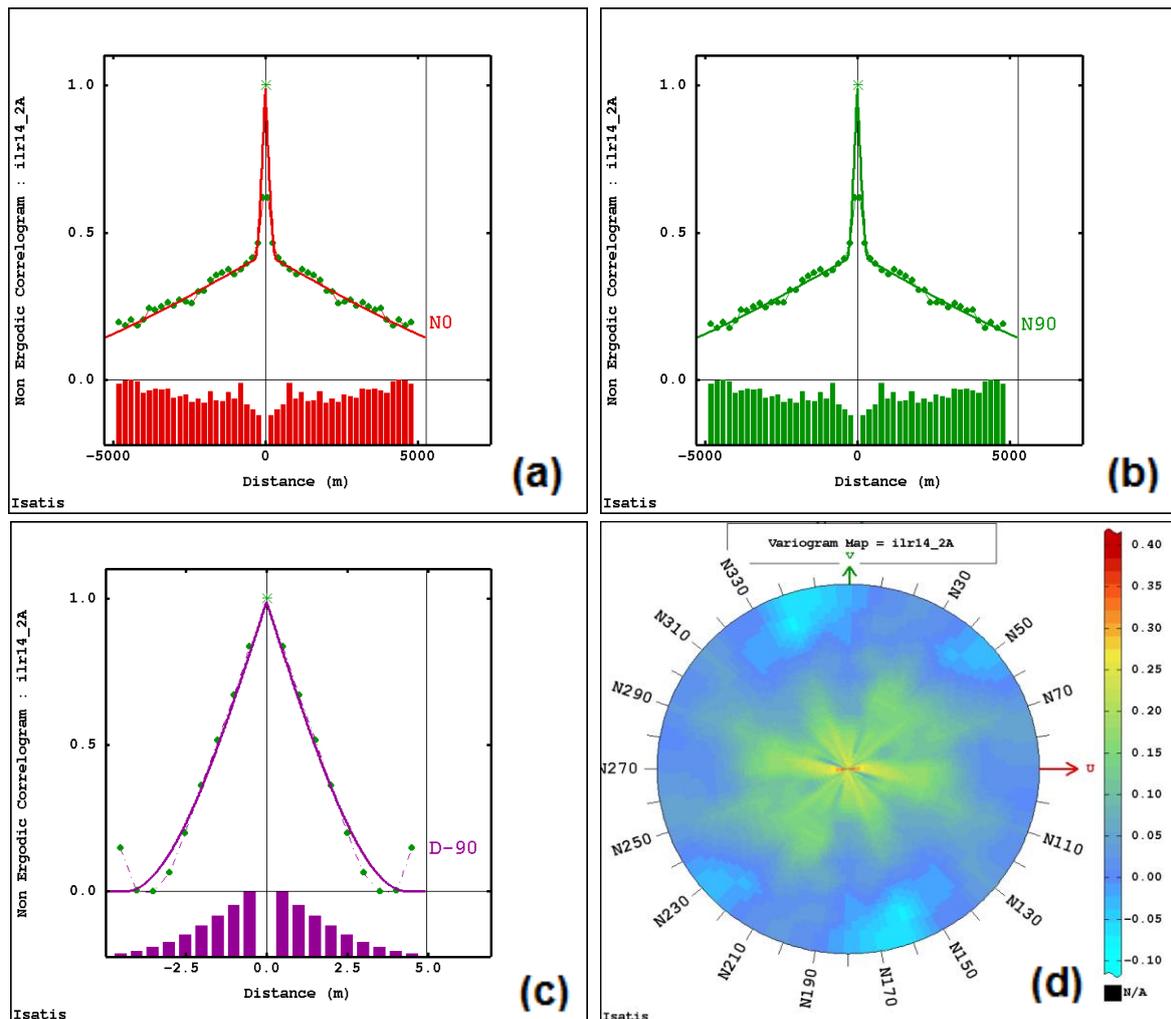


Figura 4.5. Correlograma não ergódico da variável *ilr14_2A* nas direções de (a) Maior, (b) Intermediária, (c) Menor continuidade, (d) Mapa variográfico no plano XY.

4.3. Estimativa

Quando utilizamos um banco de dados heterotópico cuja finalidade seja estimativa dos atributos aliado ao fechamento das espécies químicas e/ou mássico das frações retidas nos ensaios granulométricos, os resultados não se mostram satisfatórios independentemente da metodologia utilizada. Ou seja, utilizar um banco de dados heterotópico não garante que a soma das estimativas das variáveis em cada bloco sejam 100% devido aos pesos distintos que cada variável receberá associado a distintos modelos variográficos. Visto isso no capítulo anterior, decidiu-se trabalhar somente com o banco de dados isotópico (original) que se garante essa possibilidade.

A técnica geoestatística utilizada nessa etapa de interpolação foi a krigagem ordinária aplicada às transformações logarítmicas isométricas (ilr) acumuladas. A ideia principal é poder fazer futuras comparações dos resultados obtidos nessa fase, com os resultados obtidos no capítulo anterior em termos de qualidade de estimativas, validações e fechamento.

Como já sabemos, o banco de dados tinha variações de suporte amostral, portanto, foram estimadas as variáveis ilr acumuladas em cada faixa e, posteriormente, foi feito o processo de desacumulação. No caso das frações granulométricas, a estimativa foi feita diretamente nas transformações, pois não existe processo de acumulação.

Nessa etapa, somente foram consideradas duas metodologias para o banco de dados isotópico: primeiro, utilizamos modelos individuais onde cada variável é interpolada com seu próprio modelo de continuidade bem definido anteriormente na variografia. Segundo, utilizamos modelos únicos onde cada variável é interpolada com o modelo da recuperação em cada fração granulométrica, ou seja, as variáveis ilr14_1A, ilr14_2A, ilr14_3A, ilr14_4A utilizaram o modelo variográfico da variável REC14 e as variáveis ilr400_1A, ilr400_2A, ilr400_3A, ilr400_4A utilizaram o modelo variográfico da variável REC400.

Em cada etapa de interpolação, foram feitos os mesmos testes da krigagem utilizados no capítulo anterior, com a idéia principal de obter o melhor resultado possível fazendo pequenas variações em cada parâmetro. Por exemplo, o número mínimo e máximo de amostras para estimar um bloco, o número de sectores angulares, o número máximo de amostras por cada sector angular, entre outros (ver tabela 3.10 no capítulo 3).

A definição da estratégia de busca é primordial nesse tipo de depósito em particular, pois apresenta uma peculiaridade importante onde algumas variáveis mostram uma anisotropia zonal. Para o caso das transformações ilr, também modelou-se uma anisotropia zonal; portanto, decidiu-se utilizar o modelo prático que consiste em utilizar como alcance máximo onde começa o sill dessa anisotropia zonal.

A tabela 4.5, mostra o alcance máximo da segunda estrutura dos modelos que foram utilizados na krigagem (modelo prático) para cada uma das variáveis transformadas e, a variância que esse alcance representa em relação ao alcance dos modelos originais. Note que, a variável ilr14_4A possui um alcance de 11.000m no modelo original, mas para esse caso em particular utilizaremos um alcance de 4.500m que representa aproximadamente 90% da variância total.

Tabela 4.5. Resumo dos parâmetros utilizados na busca da krigagem.

Variável	Modelo Original			Modelo Prático			Variância		
	Segunda Estrutura (m)			Busca até (m)			% ²		
	a (D1)	a (D2)	a (D3)	a (D1)	a (D2)	a (D3)	a (D1)	a (D2)	a (D3)
ilr14_1A	6500	6500	40	3000	3000	2.5	90	90	75
ilr14_2A	11000	11000	4.4	4500	4500	4.4	82	82	100
ilr14_3A	6500	6500	5	4800	4800	2.5	98	98	95
ilr14_4A	11000	11000	9.5	4500	4500	4	90	90	90
ilr400_1A	4200	4200	30	4200	4200	2.5	100	100	57
ilr400_2A	3200	3200	300	3200	3200	2.5	100	100	68
ilr400_3A	6000	6000	11.5	4000	4000	4	95	95	80
ilr400_4A	4900	4900	12	4900	4900	3.5	100	100	72
ilrRec_1	9500	9500	4.8	4800	4800	3	93	93	96
ilrRec_2	6500	6500	26	4500	4500	3	93	93	65

D1 – Direção de Maior Continuidade – N 0° Dip 0°

D2 – Direção de Continuidade Intermediária - N 90° Dip 0°

D3 – Direção de Menor Continuidade - N 0° Dip 90°

a – Alcance

A figura 4.6 mostra um exemplo claro de anisotropia zonal da variável *ilr400_2A* na vertical. Note como a variância se estabiliza depois de determinado alcance, portanto, tomamos como referência o começo dessa anisotropia zonal para obter o modelo prático que vai ser utilizado no processo de estimativa.

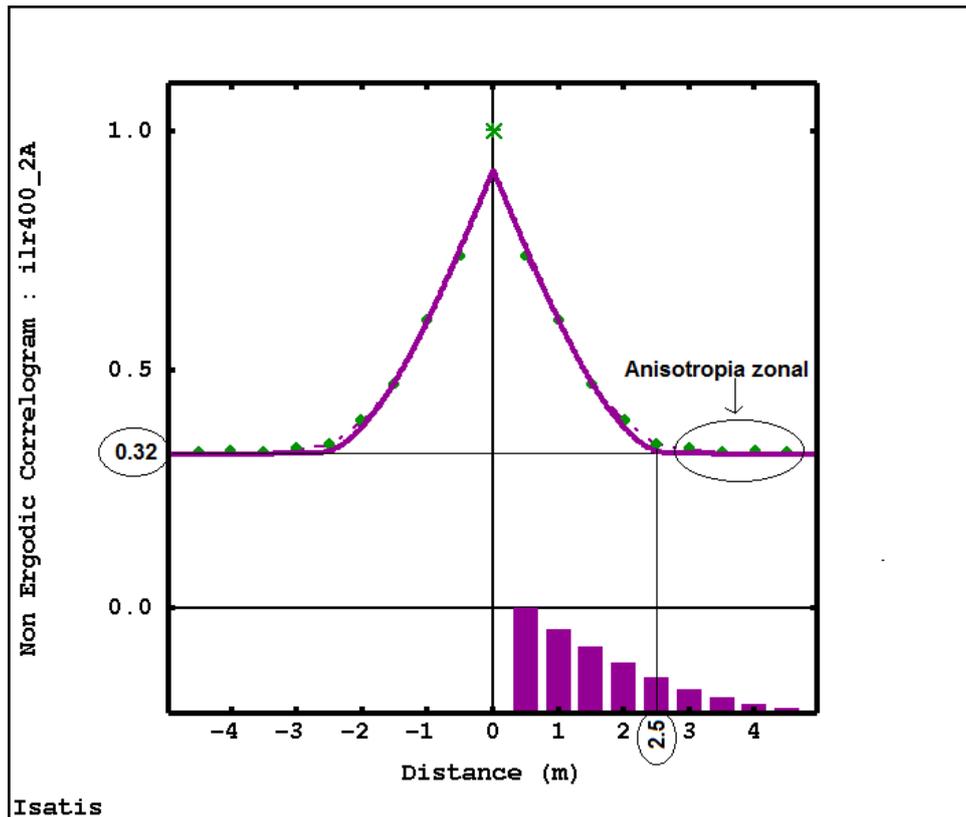


Figura 4.6. Anisotropia zonal da variável *ilr400_2A* no plano vertical.

4.4. Viés na média aritmética

Quando utilizamos transformações não lineares (logarítmicas, anamorfose gaussiana, etc.), existe a possibilidade de haver um viés na média estimada do teor dos blocos (*ilr*) ao retro-transformarmos as estimativas para o espaço original. Isto sucede quando os dados são transformados por uma função não linear e procede-se cálculos com essas variáveis transformadas (por exemplo krigagem). Ao voltar-se ao espaço dos dados originais, essas médias

são retro-transformadas pela sua função inversa gerando viés. Em caso de utilizar-se transformações lineares não haveria esse problema.

A tabela 4.6 mostra um exemplo da média aritmética obtida a partir de 4 amostras comparada com a média aritmética retro-transformada utilizando duas funções de transformação: uma linear e outra não linear.

Para o caso da transformação linear, foi multiplicado cada amostra por 2 e, o resultado da média aritmética (5.0), foi retro-transformado pela sua função inversa (dividido por 2). Note que, para esse caso, a média é reproduzida satisfatoriamente ($2.50 = 2.50$). No caso da transformação não linear, foi aplicada a função logaritmo natural (Ln) em cada amostra e, o resultado da média aritmética (0.79), foi retro-transformado pela sua função inversa (exponencial). Note que existe um viés na média ($2.50 \approx 2.21$), ou seja, o processo da transformação não linear não garante a reprodução da média aritmética na retro-transformação.

Tabela 4.6. Comparação da media aritmética usando transformações lineares e não lineares.

	Amostras	Transformação Linear (x2)	Transformação Logarítmica (Ln)
	1	2	0.0
	2	4	0.7
	3	6	1.1
	4	8	1.4
Média Aritmética	2.50	5.0	0.79
Média Retro-transformada	-	2.50	2.21

Isto ganha muita importância na hora de fazer estimativas, pois cada bloco estimado é representado pela média aritmética das estimativas pontuais utilizadas na discretização de cada bloco.

Neste estudo, foi utilizado um tamanho de bloco de 50mx50mx1m com uma discretização de 25 pontos por bloco (5x5x1). Cada bloco estimado é obtido a partir da média aritmética das estimativas pontuais com krigagem (que

também é uma média ponderada) que foram utilizadas nessa discretização, ou seja, o teor final de cada bloco é representado pela média aritmética dos 25 pontos estimados.

Para atenuar esse problema nas transformações logarítmicas inversas (ilr), foi feita uma krigagem pontual no mesmo suporte da discretização utilizada anteriormente (10mx10mx1m). Depois, esse modelo de pontos foi retro-transformado ao espaço dos dados originais aplicando a transformação logarítmica inversa (ilr^{-1}). Por último, para obter o teor final de cada bloco de 50mx50mx1m foi tomada a média aritmética desses 25 pontos retro-transformados. A figura 4.7 mostra todo o processo descrito anteriormente.

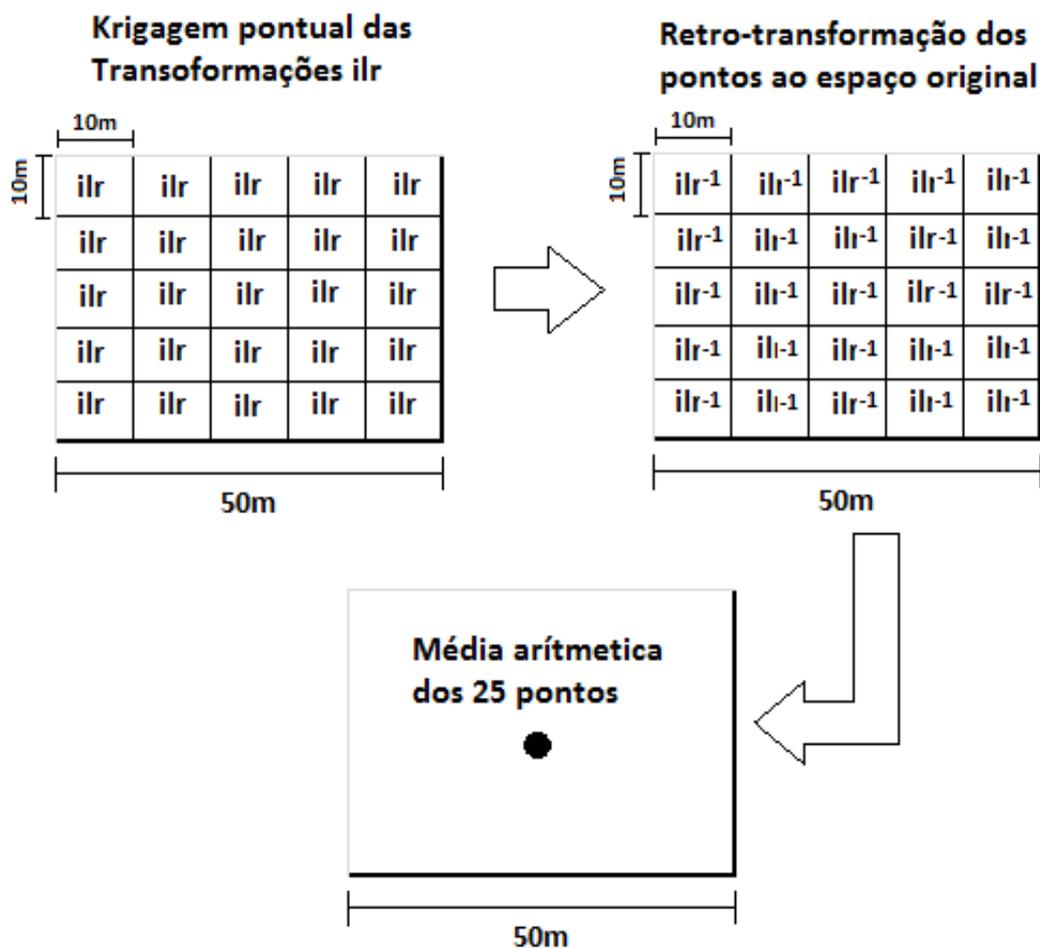


Figura 4.7. Processo para atenuar o viés na média aritmética utilizando transformações não lineares.

4.5. Reprodução da média geométrica

Pawlowsky & Egozcue (2001) mencionam que as transformações logarítmicas isométricas (*ilr*) são transformações não tendenciosas pelo fato de referirem-se a estimativa de média de coordenadas. Portanto, a média geométrica no simplex (geometria de Aitchison) é uma medida de tendência central não tendenciosa. Diferente da média aritmética, que produz um viés no valor esperado em cada variável que faz parte da composição, o valor da média geométrica não apresenta esse problema, ou seja, a média geométrica de um conjunto de dados se reproduz satisfatoriamente na retro-transformação de uma função não linear.

A tabela 4.7 mostra um exemplo da média geométrica obtida a partir de 4 amostras comparada com a média geométrica retro-transformada utilizando as mesmas transformações lineares e não lineares descritas anteriormente. Note que, para esse caso, a média geométrica é reproduzida satisfatoriamente ($3.31 = 3.31$) na função inversa da transformação não linear. Essa retro-transformação (função exponencial) é aplicada à média aritmética das transformações não lineares.

Tabela 4.7. Comparação da média geométrica usando transformações lineares e não lineares.

	Amostras	Transformação Linear (x2)	Transformação Logarítmica (Ln)
	2	4	0.69
	3	6	1.10
	4	8	1.39
	5	10	1.61
Média Aritmética	3.50	7.0	1.20
Média Geométrica	3.31	6.62	1.14
Média Retro-transformada	-	3.50	3.31

4.6. Retro-transformação (ilr^{-1})

Uma vez que os dados transformados são estimados, o procedimento pra trazer de volta essas estimativas ao espaço dos dados originais é mediante a transformação logarítmica isométrica inversa (ilr^{-1}) definida por Egozcue & Pawlowsky (2003), como:

$$ilr^{-1} = C(\text{Exp}(x \cdot \psi)) \quad (4.2)$$

Onde:

ilr^{-1} = transformação logarítmica isométrica inversa.

x = coordenadas (variáveis transformadas estimadas).

ψ = matriz que representa as coordenadas.

C = operação de fechamento

As etapas para o cálculo da transformação logarítmica isométrica inversa são:

1. Construção da matriz ψ : A matriz ψ é construída a partir da partição seqüencial binária utilizada anteriormente, portanto, cada balanço terá sua própria matriz de acordo ao número de partições em cada faixa. O cálculo é baseado nas seguintes expressões:

$$\psi_{i+} = + \sqrt{\frac{s_i}{r_i(r_i + s_i)}} \quad (4.3)$$

$$\psi_{i-} = - \sqrt{\frac{r_i}{s_i(r_i + s_i)}} \quad (4.4)$$

$$\psi_{i0} = 0 \quad (4.5)$$

onde ψ_{i+} e ψ_{i-} representam os valores da matriz ψ que foram selecionados como +1 e -1 na partição seqüencial binária respectivamente, r_i e s_i representam a somatória dos +1 e -1 obtidas na partição seqüencial binária. A tabela 4.8 mostra a matriz ψ que

representa as coordenadas obtidas a partir da partição seqüencial binária.

Tabela 4.8. Matriz ψ que representa as coordenadas para cada balanço.

No. de Partições	Fração +14#				
	Alap14	Fe14	Sire14	Ti14	Resto14
1	$+\sqrt{\frac{2}{3(3+2)}}$	$+\sqrt{\frac{2}{3(3+2)}}$	$+\sqrt{\frac{2}{3(3+2)}}$	$-\sqrt{\frac{3}{2(3+2)}}$	$-\sqrt{\frac{3}{2(3+2)}}$
2	$+\sqrt{\frac{1}{2(2+1)}}$	$+\sqrt{\frac{1}{2(2+1)}}$	$-\sqrt{\frac{2}{1(2+1)}}$	0	0
3	$+\sqrt{\frac{1}{1(1+1)}}$	$-\sqrt{\frac{1}{1(1+1)}}$	0	0	0
4	0	0	0	$+\sqrt{\frac{1}{1(1+1)}}$	$-\sqrt{\frac{1}{1(1+1)}}$

No. de Partições	Fração +400#				
	Alap400	Fe400	Sire400	Ti400	Resto400
1	$+\sqrt{\frac{2}{3(3+2)}}$	$+\sqrt{\frac{2}{3(3+2)}}$	$+\sqrt{\frac{2}{3(3+2)}}$	$-\sqrt{\frac{3}{2(3+2)}}$	$-\sqrt{\frac{3}{2(3+2)}}$
2	$+\sqrt{\frac{1}{2(2+1)}}$	$+\sqrt{\frac{1}{2(2+1)}}$	$-\sqrt{\frac{2}{1(2+1)}}$	0	0
3	$+\sqrt{\frac{1}{1(1+1)}}$	$-\sqrt{\frac{1}{1(1+1)}}$	0	0	0
4	0	0	0	$+\sqrt{\frac{1}{1(1+1)}}$	$-\sqrt{\frac{1}{1(1+1)}}$

No. de Partições	Recuperações		
	Rec14	Rec400	Rec-400
1	$+\sqrt{\frac{1}{2(2+1)}}$	$+\sqrt{\frac{1}{2(2+1)}}$	$-\sqrt{\frac{2}{1(2+1)}}$
2	$+\sqrt{\frac{1}{1(1+1)}}$	$-\sqrt{\frac{1}{1(1+1)}}$	0

Note que, os valores de ψ_{i+} e ψ_{i-} numa mesma partição são iguais, isto se deve ao cálculo desses valores serem obtidos a partir da somatória dos parâmetros de r_i e s_i . A matriz ψ , portanto, possui o mesmo número de filas e colunas que a partição seqüencial binária.

Os valores nessa tabela são expressos em termos de raízes para facilitar o entendimento de como foram aplicadas cada uma das equações (4.3, 4.4, 4.5) utilizadas nessa matriz. Alguns valores receberam o valor de 0, pois na partição seqüencial binária também possuíam esse valor.

2. Produto matricial $x \cdot \psi$: Para obter a transformação logarítmica isométrica inversa (ilr^{-1}), é preciso fazer uma operação de multiplicação de matrizes. As duas matrizes utilizadas correspondem às matrizes formadas pelas estimativas obtidas nas transformações (ilr) para cada fração (+14#, +400#, -400#) e a matriz ψ calculada anteriormente.
3. Operação de fechamento: Para fazer o processo de fechamento é preciso primeiro calcular o inverso da função logarítmica, ou seja, aplicar a função exponencial (\exp) aos valores calculados anteriormente no produto matricial. Por último, esse valor passa pelo processo de fechamento descrito na equação 2.26 anteriormente definida. A figura 4.8 mostra um exemplo de todo o processo descrito anteriormente para calcular a retro-transformação dos dados estimados no espaço dos dados originais.

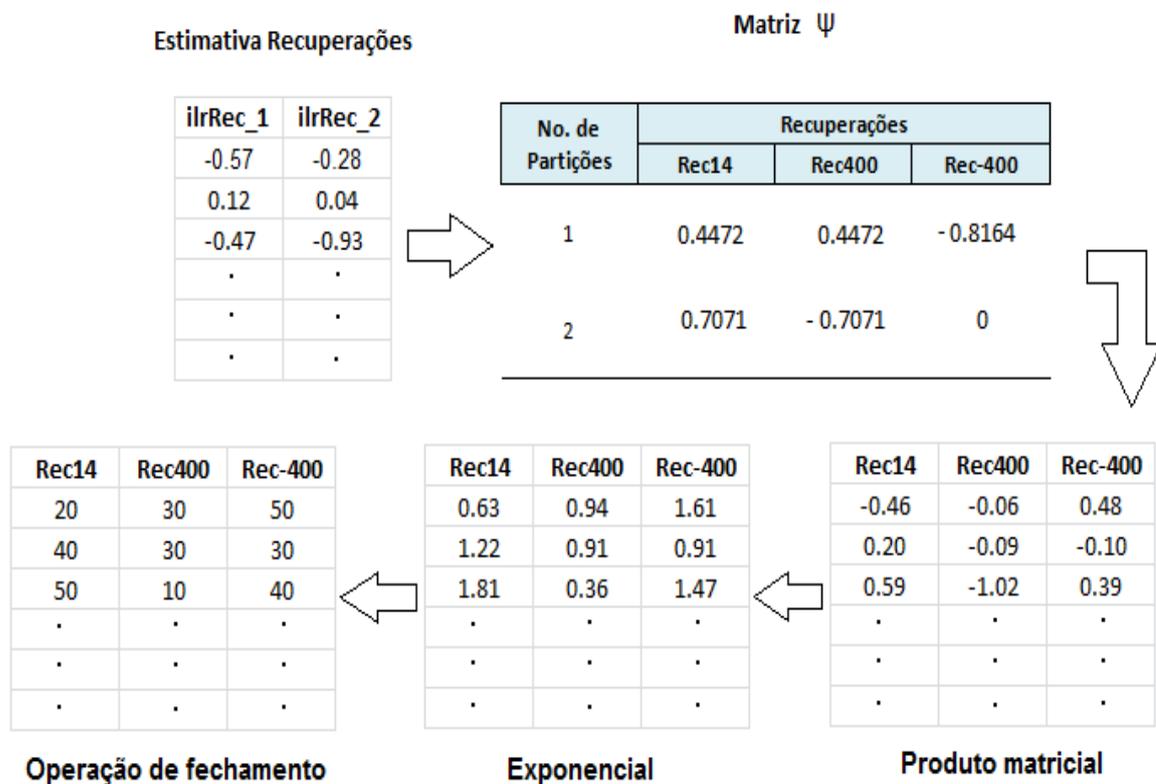


Figura 4.8. Processo de transformação logarítmica isométrica inversa (ilr^{-1}).

4.7. Análise preliminar dos resultados

Os resultados apresentados aqui fazem referência ao fechamento do balanço químico e mássico e, os valores mínimos e máximos em cada uma das metodologias estudadas, além de um breve comentário. Lembrando que, como as variáveis estimadas foram transformadas e posteriormente acumuladas, foi necessário fazer um processo de desacumulação. Nessa vez, esse processo foi feito antes que as estimativas fossem retro-transformadas pela sua função inversa (ilr^{-1}).

As duas metodologias de estimativas analisadas nesse capítulo foram baseadas nas metodologias 1 e 2 descritas anteriormente no capítulo 3. Essas metodologias utilizam o banco de dados isotópico, variando somente o tipo de modelo de continuidade espacial utilizado no processo de estimativa.

4.7.1. Resultados com metodologia 1 (ilr - isotópico – variogramas individuais - krigagem ordinária)

A tabela 4.9 mostra os valores mínimos e máximos das variáveis em relação às estimativas obtidas por krigagem ordinária fazendo uso das transformações (ilr). Note que, as estimativas das variáveis Alap14, Fe14, Resto14, Alap400, Fe400, Ti400 e Resto400 ficaram fora do intervalo dos dados originais. Isto é atribuído ao processo de desacumulação, pois cada variável utilizou o seu próprio modelo de continuidade e, por conseguinte, recebeu pesos diferentes.

O número de blocos fora dos valores mínimo e máximo não foram considerados significativos por representarem uma pequena percentagem (menor que 1%). Embora o número de blocos fora do intervalo seja insignificante em relação ao depósito como um todo, essas estimativas devem passar por um pós-processamento.

Tabela 4.9. Valores mínimos e máximos dos dados estimados usando uma krigagem pontual e os dados originais (Metodologia 1).

Metodologia 1							
Variável	Blocos estimados	Dados Estimados		Dados Originais		Blocos fora do Intervalo	
		Mínimo	Máximo	Mínimo	Máximo	Nº	%
1_Krig_ilr_Alap14	74448	22.98	73.58	13.07	58.78	83	0.11
1_Krig_ilr_Fe14	74448	0.37	34.88	0.69	54.65	1	0.001
1_Krig_ilr_Sire14	74448	0.43	12.12	0.27	14.9	0	0.00
1_Krig_ilr_Ti14	74448	0.31	3.08	0.27	3.57	0	0.00
1_Krig_ilr_Resto14	74448	22.58	55.29	21.47	52.32	1	0.001
1_Krig_ilr_Alap400	74448	1.08	72.6	0.2	58.32	415	0.56
1_Krig_ilr_Fe400	74448	0.48	33.33	0.56	46.46	4	0.01
1_Krig_ilr_Sire400	74448	0.52	21.73	0.46	28.9	0	0.00
1_Krig_ilr_Ti400	74448	0.17	5.05	0.32	7.13	6	0.01
1_Krig_ilr_Resto400	74448	22.19	74.18	24.25	79.24	2	0.00
1_Krig_ilr_Rec14	74448	8.74	95.92	3.21	97.52	0	0.00
1_Krig_ilr_Rec400	74448	1.46	38.98	0.55	44.86	0	0.00
1_Krig_ilr_Rec-400	74448	2.22	83.9	1.04	94.14	0	0.00

As metodologias 1 dos capítulos 3 e 4 produziram resultados similares em relação a estimar valores fora do intervalo mínimo e máximo dos dados originais. Ressaltamos uma vez mais, que isso é gerado por utilizar-se modelos variográficos distintos para cada variável acumulada e, portanto, quando é aplicado o processo de desacumulação, pode ser gerado esse tipo de valor. Para atenuar esse problema é que recorreremos a utilização de um modelo único para todas as variáveis e outorgar o mesmo peso a todas as espécies químicas que fazem parte da composição. Assim, é garantido que os blocos estimados não estimem valores fora dos limites dos dados. Isto é claramente visto nas metodologias 2 de ambos os capítulos.

A diferença do capítulo 3, o ganho nesta metodologia (ilr) está no fechamento químico e mássico das estimativas em de cada uma das faixas analisadas. As figuras 4.9 (a, b, c) mostram os histogramas de fechamento para os três balanços (faixa +14#, faixa +400# e recuperação). Note que, mesmo utilizando modelos diferentes para cada variável, é respeitada a soma das estimativas de cada faixa em 100%.

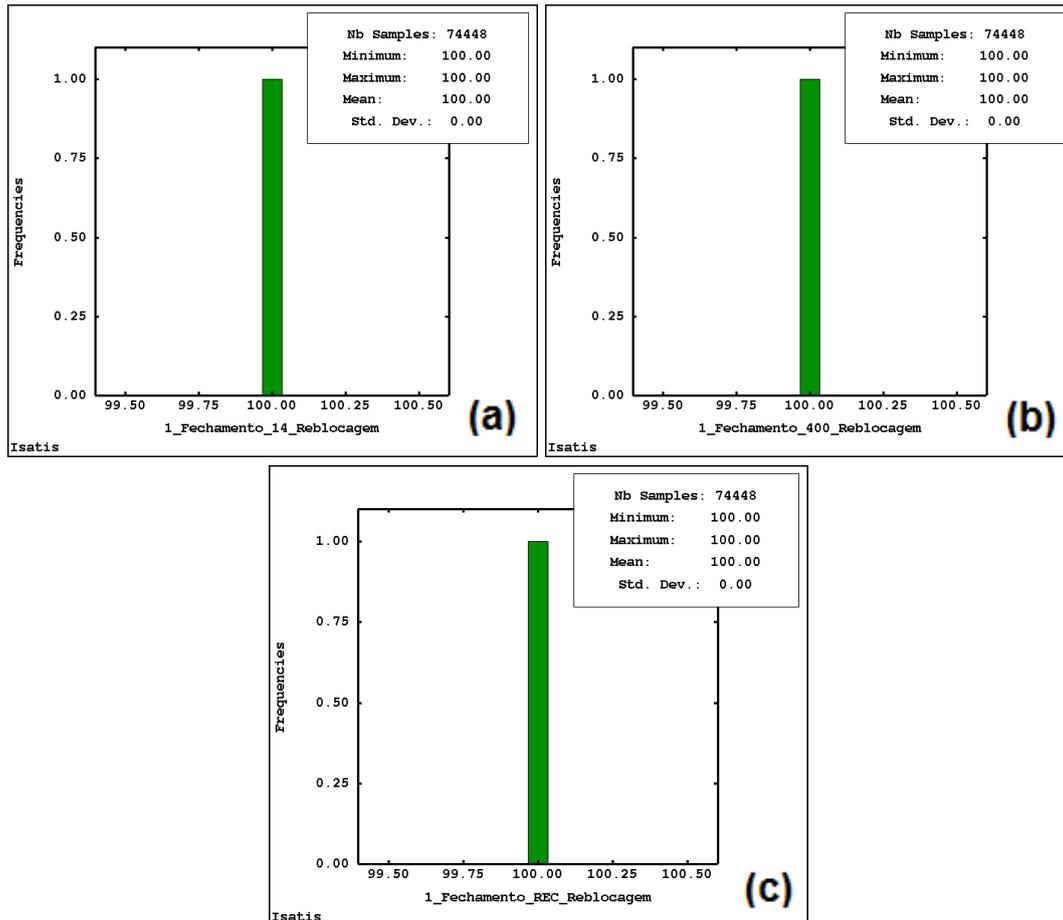


Figura 4.9. Histograma de fechamento metodologia 1 usando uma krigagem pontual, (a) faixa +14#, (b) faixa +400#, (c) frações mássicas.

4.7.2. Resultados com metodologia 2 (ilr- isotópico – variograma único - krigagem ordinária)

A tabela 4.10 não apresentou nenhum valor fora do intervalo esperado, isto devido ao modelo utilizado no processo de estimativa por cada faixa ser o mesmo, outorgando igual peso para todas as variáveis. Ou seja, as estimativas das variáveis que foram transformadas na fração +14# (ilr14_1A, ilr14_2A, ilr14_3A e ilr14_4A) utilizaram o modelo da Rec14 e, as estimativas das variáveis que foram transformadas na fração +400# (ilr400_1A, ilr400_2A, ilr400_3A e ilr400_4A) utilizou-se o modelo da Rec400.

Nessa metodologia, também foi respeitado o fechamento, ou seja, a soma das estimativas das variáveis em cada bloco é 100%. Na figura 4.10 (a, b, c), podemos observar os histogramas de fechamento para cada fração.

Tabela 4.10. Valores mínimos e máximos dos dados estimados usando uma krigagem pontual e os dados originais (Metodologia 2).

Metodologia 2							
Variável	Blocos estimados	Dados Estimados		Dados Originais		Blocos fora do Intervalo	
		Mínimo	Máximo	Mínimo	Máximo	Nº	%
2_Krig_ilm_Alap14	74448	24.87	57.13	13.07	58.78	0	0.00
2_Krig_ilm_Fe14	74448	1.39	33.18	0.69	54.65	0	0.00
2_Krig_ilm_Sire14	74448	0.72	9.96	0.27	14.9	0	0.00
2_Krig_ilm_Ti14	74448	0.32	3.14	0.27	3.57	0	0.00
2_Krig_ilm_Resto14	74448	27.67	48.53	21.47	52.32	0	0.00
2_Krig_ilm_Alap400	74448	2.04	56.35	0.2	58.32	0	0.00
2_Krig_ilm_Fe400	74448	0.87	29.84	0.56	46.46	0	0.00
2_Krig_ilm_Sire400	74448	0.69	18.97	0.46	28.9	0	0.00
2_Krig_ilm_Ti400	74448	0.37	4.69	0.32	7.13	0	0.00
2_Krig_ilm_Resto400	74448	29.22	70.52	24.25	79.24	0	0.00
2_Krig_ilm_Rec14	74448	9.39	95.75	3.21	97.52	0	0.00
2_Krig_ilm_Rec400	74448	1.66	39.11	0.55	44.86	0	0.00
2_Krig_ilm_Rec-400	74448	2.33	82.72	1.04	94.14	0	0.00

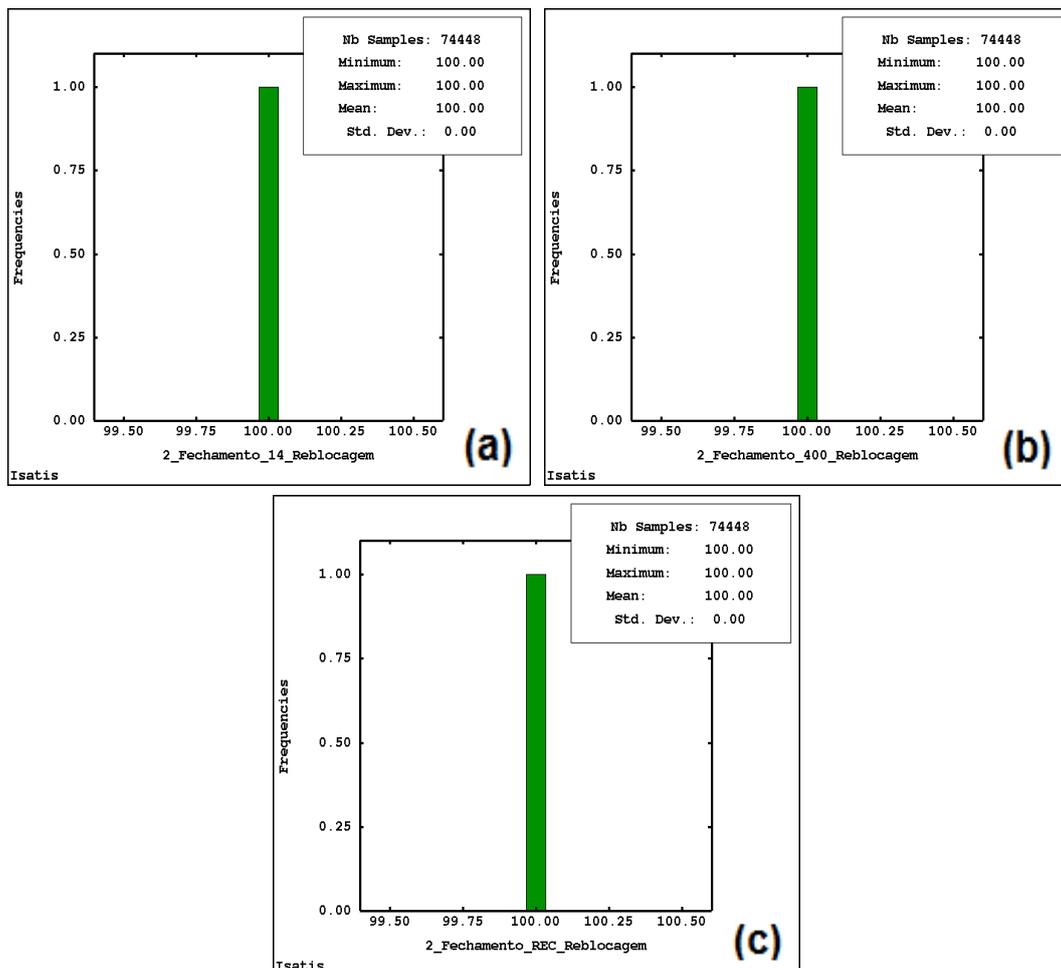


Figura 4.10. Histograma de fechamento metodologia 2 usando uma krigagem pontual, (a) faixa +14#, (b) faixa +400#, (c) frações mássicas.

4.7.3. Metodologia 1 e 2 utilizando uma krigagem em blocos

Para analisar o impacto do viés na média aritmética local das estimativas para obtenção dos teores em blocos, foi feita uma krigagem em blocos com uma discretização de 25 pontos (5x5x1). Essa discretização corresponde ao mesmo tamanho espaçamento utilizado anteriormente na krigagem pontual.

Os resultados em termos de valores mínimos e máximos são muito similares às estimativas obtidas por a krigagem pontual na metodologia 1, ou seja, foram geradas estimativas fora do intervalo dos dados originais devido ao processo de desacumulação. As variáveis que apresentaram esse problema foram: Alap14, Fe14, Resto14, Alap400, Fe400, Sire400, Ti400 e Resto400 (ver tabela 4.11). No caso da metodologia 2, os resultados não mostraram valores fora desse intervalo devido ao uso do mesmo modelo de continuidade espacial como foi explicado anteriormente (ver tabela 4.12).

Ambas as metodologias respeitaram os balanços das espécies químicas e granulométricas para cada faixa. As figuras 4.11 e 4.12 mostram o fechamento das estimativas em 100% nas duas metodologias utilizadas.

Tabela 4.11. Valores mínimos e máximos dos dados estimados usando uma krigagem em blocos e os dados originais (Metodologia 1).

Metodologia 1							
Variável	Blocos estimados	Dados Estimados		Dados Originais		Blocos fora do Intervalo	
		Mínimo	Máximo	Mínimo	Máximo	Nº	%
1_Krig_ilr_Alap14	74448	23.28	73.87	13.07	58.78	135	0.18
1_Krig_ilr_Fe14	74448	0.34	34.59	0.69	54.65	2	0.003
1_Krig_ilr_Sire14	74448	0.39	12.3	0.27	14.9	0	0.00
1_Krig_ilr_Ti14	74448	0.31	3.42	0.27	3.57	0	0.00
1_Krig_ilr_Resto14	74448	22.41	55.86	21.47	52.32	1	0.001
1_Krig_ilr_Alap400	74448	0.97	79.88	0.2	58.32	522	0.70
1_Krig_ilr_Fe400	74448	0.34	31.95	0.56	46.46	7	0.01
1_Krig_ilr_Sire400	74448	0.29	22.03	0.46	28.9	4	0.01
1_Krig_ilr_Ti400	74448	0.08	6.25	0.32	7.13	16	0.02
1_Krig_ilr_Resto400	74448	16.06	73.52	24.25	79.24	20	0.03
1_Krig_ilr_Rec14	74448	8.47	96.03	3.21	97.52	0	0.00
1_Krig_ilr_Rec400	74448	1.46	39.01	0.55	44.86	0	0.00
1_Krig_ilr_Rec-400	74448	2.14	84.29	1.04	94.14	0	0.00

Tabela 4.12. Valores mínimos e máximos dos dados estimados usando uma krigagem em blocos e os dados originais (Metodologia 2).

Metodologia 2							
Variável	Blocos estimados	Dados Estimados		Dados Originais		Blocos fora do Intervalo	
		Mínimo	Máximo	Mínimo	Máximo	Nº	%
2_Krig_ilr_Alap14	74448	24.81	57.33	13.07	58.78	0	0.00
2_Krig_ilr_Fe14	74448	1.38	33.39	0.69	54.65	0	0.00
2_Krig_ilr_Sire14	74448	0.71	9.96	0.27	14.9	0	0.00
2_Krig_ilr_Ti14	74448	0.31	3.13	0.27	3.57	0	0.00
2_Krig_ilr_Resto14	74448	27.73	48.55	21.47	52.32	0	0.00
2_Krig_ilr_Alap400	74448	1.87	56.41	0.2	58.32	0	0.00
2_Krig_ilr_Fe400	74448	0.86	28.41	0.56	46.46	0	0.00
2_Krig_ilr_Sire400	74448	0.65	19.72	0.46	28.9	0	0.00
2_Krig_ilr_Ti400	74448	0.36	4.69	0.32	7.13	0	0.00
2_Krig_ilr_Resto400	74448	29.61	70.5	24.25	79.24	0	0.00
2_Krig_ilr_Rec14	74448	9.1	95.97	3.21	97.52	0	0.00
2_Krig_ilr_Rec400	74448	1.65	39.07	0.55	44.86	0	0.00
2_Krig_ilr_Rec-400	74448	2.18	83.12	1.04	94.14	0	0.00

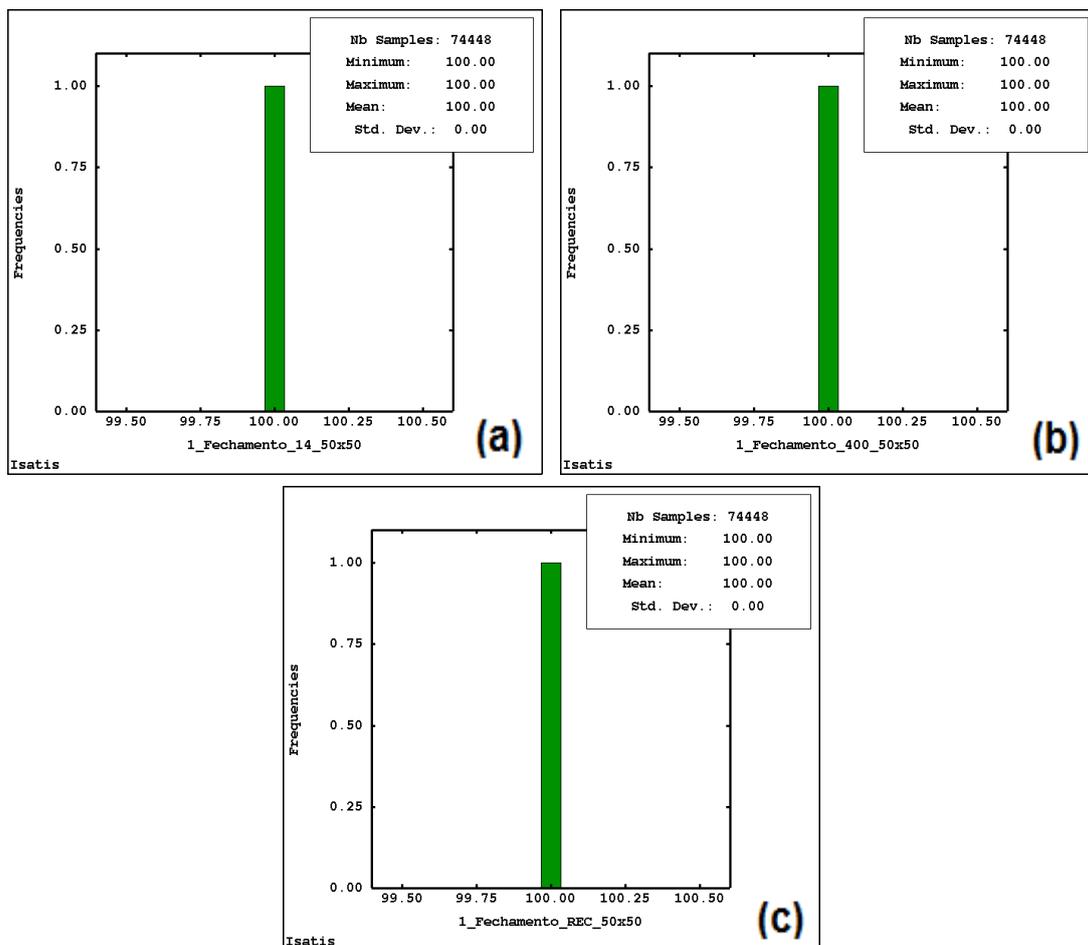


Figura 4.11. Histograma de fechamento metodologia 1 usando uma krigagem em blocos, (a) faixa +14#, (b) faixa +400#, (c) frações mássicas.

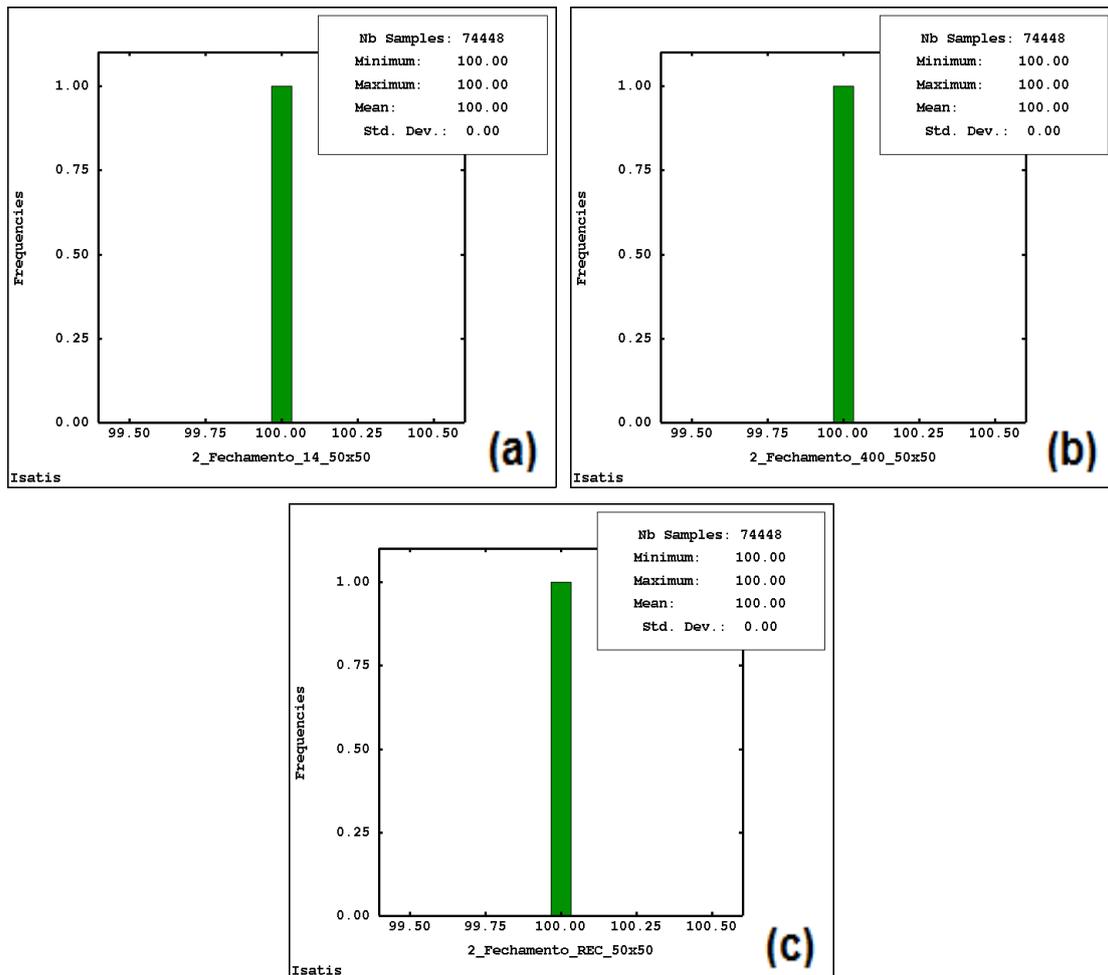


Figura 4.12. Histograma de fechamento metodologia 2 usando uma krigagem em blocos, (a) faixa +14#, (b) faixa +400#, (c) frações mássicas.

4.8. Simulação

Através da krigagem, foi possível atenuar o problema do viés da média aritmética nas transformações não lineares, mas surge a seguinte dúvida: e se queremos eliminar completamente esse viés?

A simulação surge como uma solução a esse problema, pois enquanto a krigagem produz uma estimativa ótima (um valor médio dos possíveis valores) em cada local, a simulação gera um possível valor de erro (resíduo) para essa estimativa ótima (em cada cenário), criando o aspecto aleatório (localmente) da variável regionalizada (Koppe, 2009). Em outras palavras, a simulação é baseada na geração de números randômicos dentro de uma distribuição acumulada para cada local, enquanto a krigagem, para poder estimar esse

mesmo local, ela utiliza de todas as amostras que se encontram na sua vizinhança multiplicada pelos pesos correspondentes.

4.8.1. Aspectos gerais

Esta dissertação utilizou dois algoritmos de simulação: Bandas Rotativas e Simulação seqüencial Gaussiana definidos no capítulo 2. O primeiro algoritmo foi aplicado nas frações granulométricas (Rec14, Rec400 e Rec-400) e o segundo algoritmo nas frações das espécies químicas (Alap, Fe, Sire, Ti e Resto) nas faixas +14# e +400#.

A figura 4.13 mostra o procedimento realizado no processo de simulação das transformações logarítmicas isométricas (*ilr*) em cada uma das variáveis analisadas. Este processo é válido para os dois algoritmos. Um aprofundamento destas teorias podem ser visitadas no Journal & Huijbregts (1978).

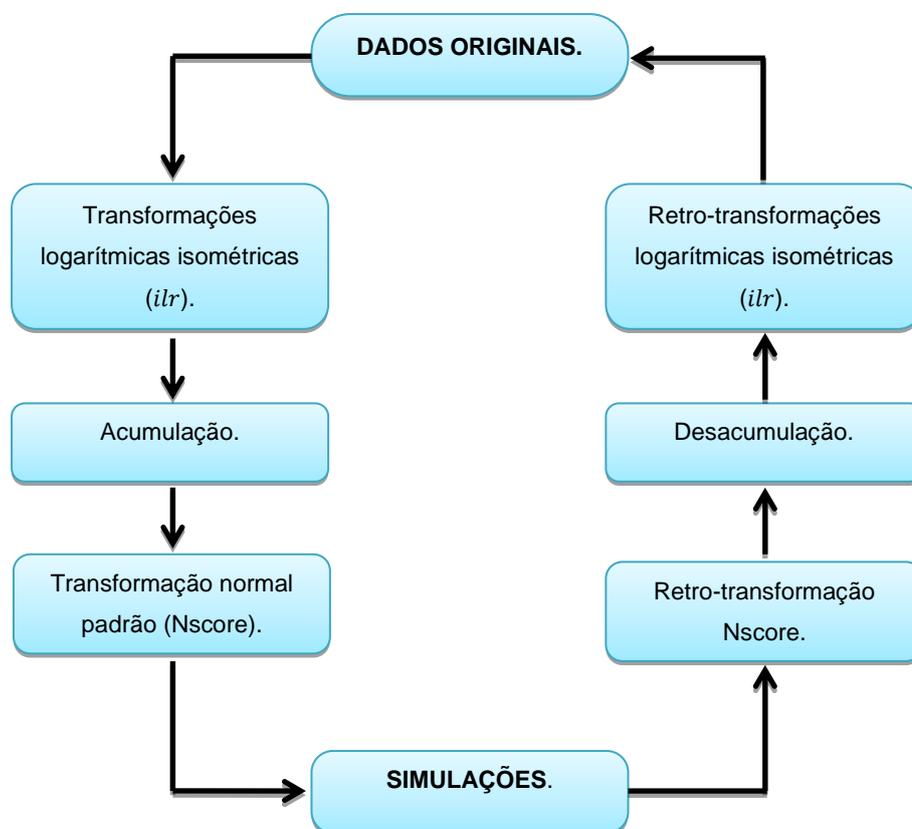


Figura 4.13. Processo de simulação para as transformações logarítmicas isométricas (*ilr*).

Uma das condições relevantes para uso dos processos de simulação selecionados é que os dados devem ter uma distribuição normal padrão. Nenhuma das variáveis analisadas apresentou uma distribuição normal padrão nas transformações logarítmicas inversas acumuladas, portanto, para poder simular foi necessário transformá-las ao espaço normal. Isto implica que, toda a análise espacial (correlogramas) foi baseada nessa nova distribuição dos dados (ver figura 4.14 a,b).

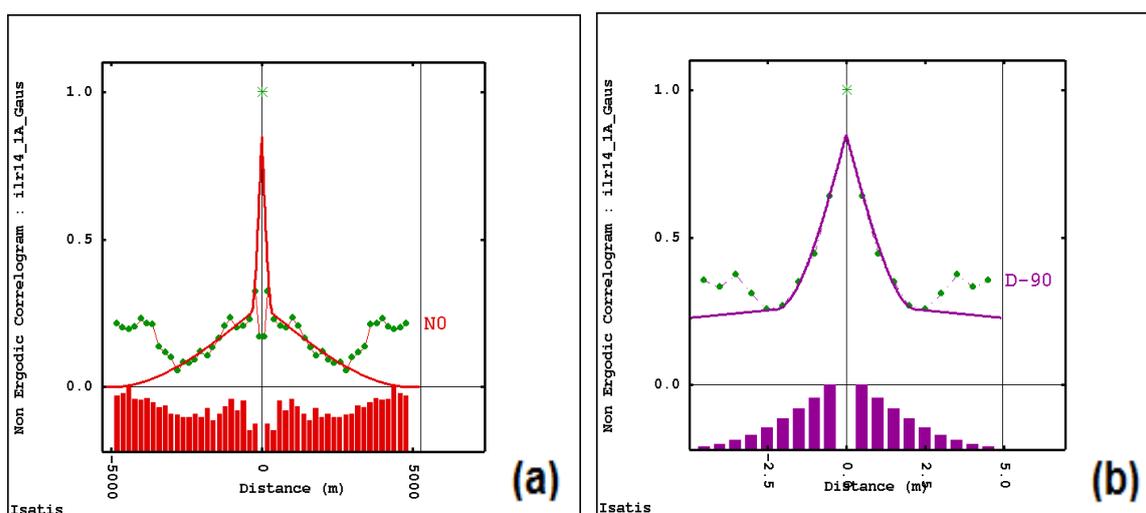


Figura 4.14. Correlograma da variável ilr14A gaussiana, (a) Plano horizontal N0, (b) Plano vertical N90.

Os parâmetros utilizados nas simulações foram:

- 30 realizações por cada variável.
- Blocos simulados de 10x10x0.5 m.
- Número mínimo e máximo de amostras: 3 e 24 respectivamente.
- Número de sectores angulares: 8
- Número de amostras por sector: 3
- A vizinhança de busca foi até o alcance prático, já que algumas variáveis apresentaram anisotropia zonal.

Considerando que a simulação é feita sobre pontos e não sobre blocos diretamente, cada variável foi simulada em um grid com espaçamento de 10x10x0.5 metros. Depois, cada uma desses cenários foram reblocados para um tamanho de bloco de 50x50x0.5 m similar ao que foi feito na krigagem pontual das (*ilrs*).

4.8.2. Análise preliminar dos resultados

Os resultados de cada uma das simulações estão baseados na utilização dos modelos individuais (metodologia 1), ou seja, cada variável é modelada e simulada em relação a sua própria continuidade. Além disso, o processo de simulação foi aplicado nas frações granulométricas (Rec14, Rec400 e Rec-400) e nas variáveis químicas (Alap14, Fe14, Sire14 Ti14 e Resto14) da faixa +14#.

A tabela 4.13 mostra os valores mínimos e máximos das variáveis em relação às simulações obtidas na simulação por bandas rotativas, que correspondem às frações granulométricas (Rec14, Rec400 e Rec-400) e a simulação seqüencial gaussiana correspondente às frações das espécies químicas (Alap, Fe, Sire, Ti e Resto) nas faixas +14#. Note que, nas frações das espécies químicas, quando utilizamos modelos diferentes, podemos obter valores fora dos limites mínimos e máximos dos dados originais. Isto é atribuído ao processo de desacumulação explicado anteriormente e não às técnicas de simulação em si.

Tabela 4.13. Valores mínimos e máximos dos dados simulados e os dados originais (Metodologia 1).

Metodologia 1										
Algoritmo	Variável	Blocos Simulados	Dados Originais		Simulação # 1		Simulação # 12		Simulação # 24	
			Mín.	Mín.	Mín.	Mín.	Mín.	Máx.	Mín.	Máx.
Simulação gaussiana	1_Sim_ilr_Alap14	74448	13.07	58.78	0.08	97.91	1.6	99.68	0.31	99.37
	1_Sim_ilr_Fe14	74448	0.69	54.65	0.004	26.04	0.003	27.36	0.002	27.76
	1_Sim_ilr_Sire14	74448	0.27	14.9	0.007	15.67	0.002	15.76	0.001	15.04
	1_Sim_ilr_Ti14	74448	0.27	3.57	3e-6	6.69	8e-6	6.81	3e-6	6.28
	1_Sim_ilr_Resto14	74448	21.47	52.32	1.55	99.1	0.22	97.14	0.47	99.32
Bandas Rotativas	1_Sim_ilr_Rec14	74448	3.21	97.52	5.69	96.67	5.03	95.9	4.23	97.23
	1_Sim_ilr_Rec400	74448	0.55	44.86	1.21	52.27	1.32	58.98	0.79	62.14
	1_Sim_ilr_Rec-400	74448	1.04	94.14	1.53	91.94	1.6	90.74	1.05	92.82

Cada cenário simulado deve reproduzir satisfatoriamente (considerando as flutuações ergódicas) tanto o histograma como o variograma dos dados originais. A figura 4.15 (a) mostra o histograma dos dados originais da Rec14 e, a figura 4.15 (b) mostra o histograma simulado para o cenário # 4 da mesma

variável. Note que, as distribuições de ambos os histogramas são muito parecidas, isto mostra que a simulação reproduz a distribuição dos dados originais. Nas figuras 4.15 (c,d,e,f), também podemos observar esse mesmo comportamento.

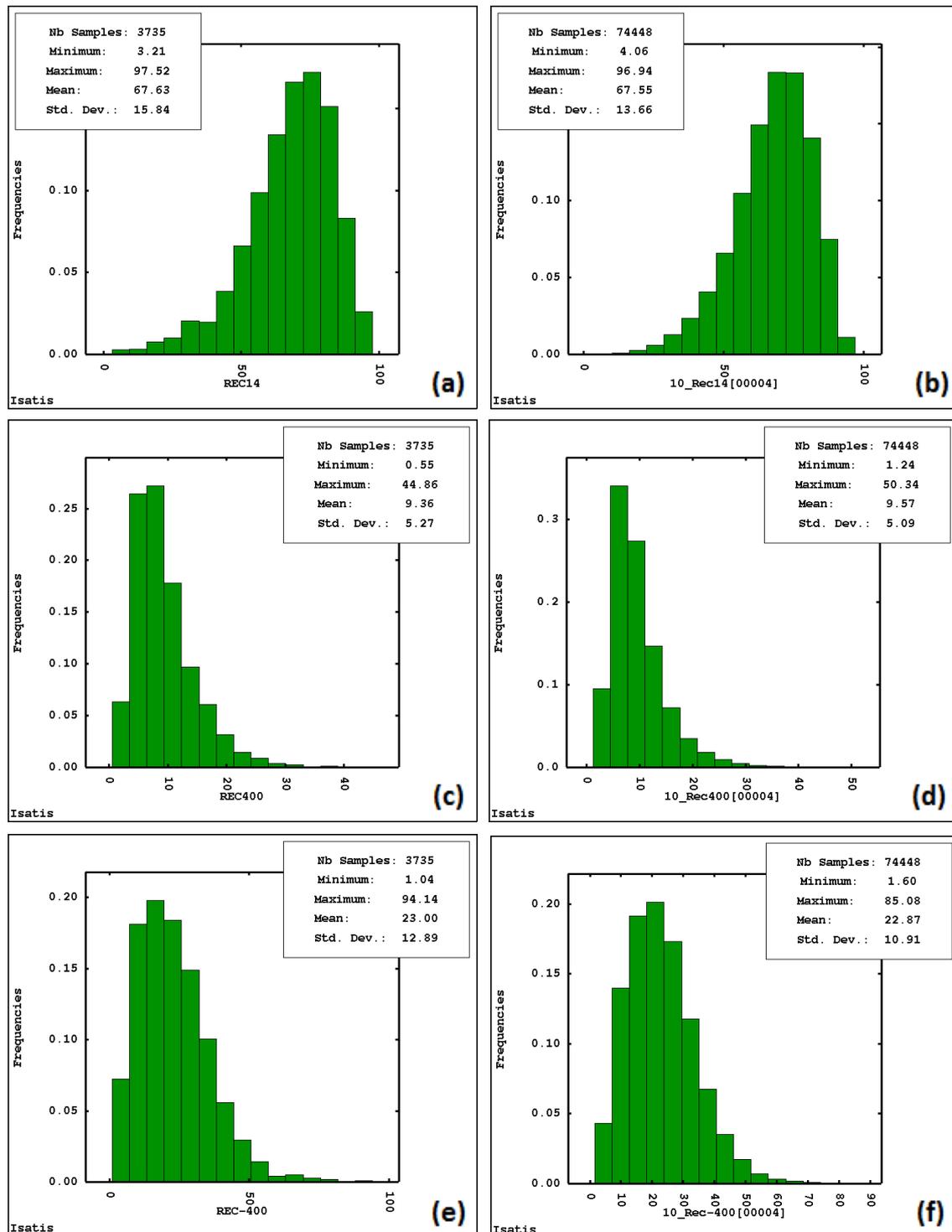


Figura 4.15. Histogramas dos dados originais e simulações nas frações granulométricas (a) Rec14, (b) Rec14_Sim, (c) Rec400, (c) Rec400_Sim, (e) Rec-400, (e) Rec-400_Sim.

O segundo método para validar as simulações é a reprodução do variograma dos dados originais. A figura 4.16 (a,b,c) mostra as 30 simulações das variáveis Rec14, Rec400 e Rec-400 (cor verde) e o modelo variográfico dos dados originais (cor vermelho). Note que, nas três variáveis analisadas o modelo é reproduzido ergodicamente em cada uma das simulações, apresentando uma boa aderência ao modelo.

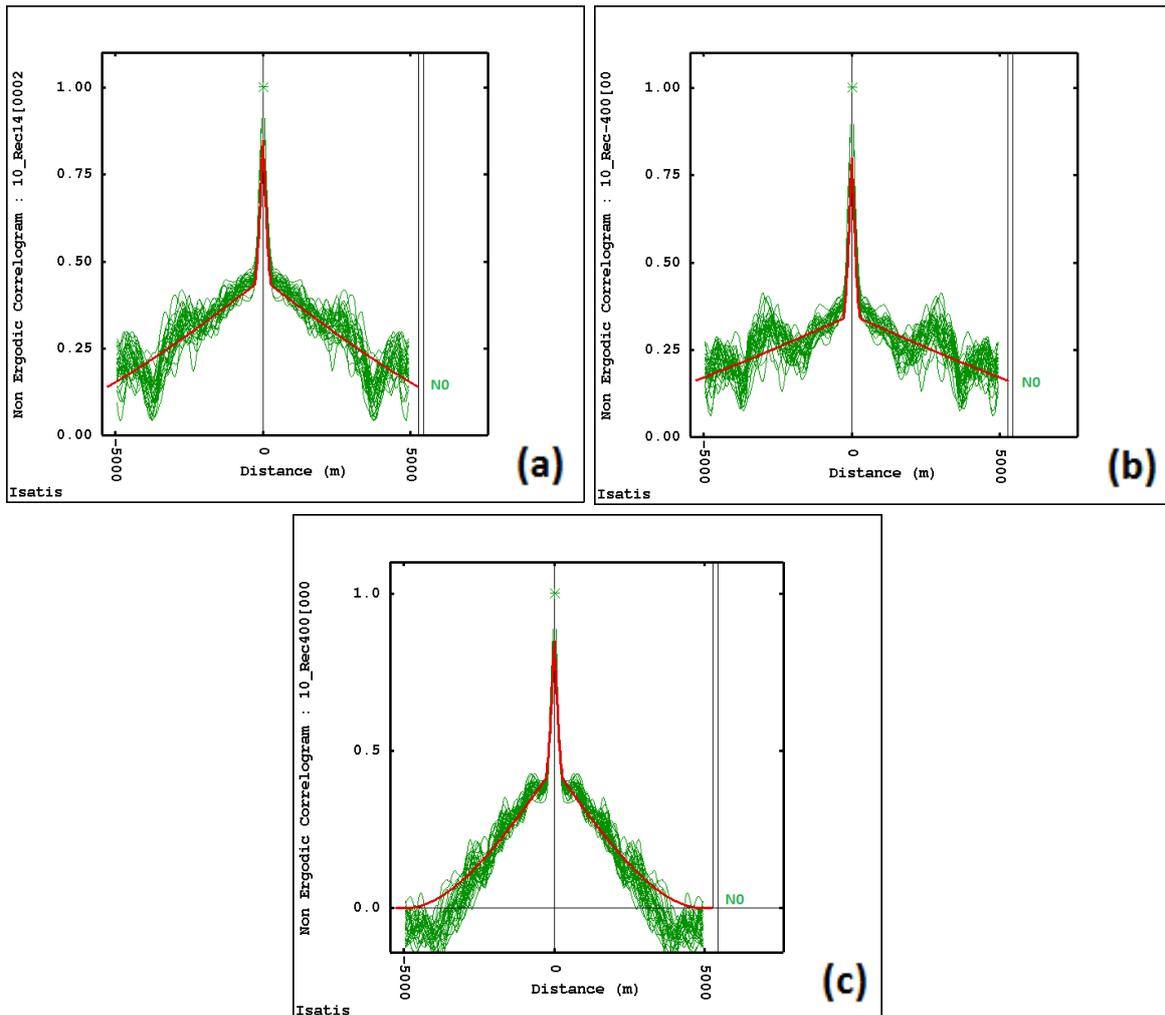


Figura 4.16. Correlograma dos dados originais (cor vermelha) vs Correlograma das simulações (a) Rec14, (b) Rec400, (c) Rec-400.

Por último, foi analisado o fechamento granulométrico para cada uma das simulações. Na figura 4.17 (a,b,c,d), podemos ver o fechamento dos cenários 8, 15, 21 e 27. O relevante nesta parte, é que mesmo utilizando modelos diferentes para cada variável, é respeitada a soma de cada uma das simulações em 100%.

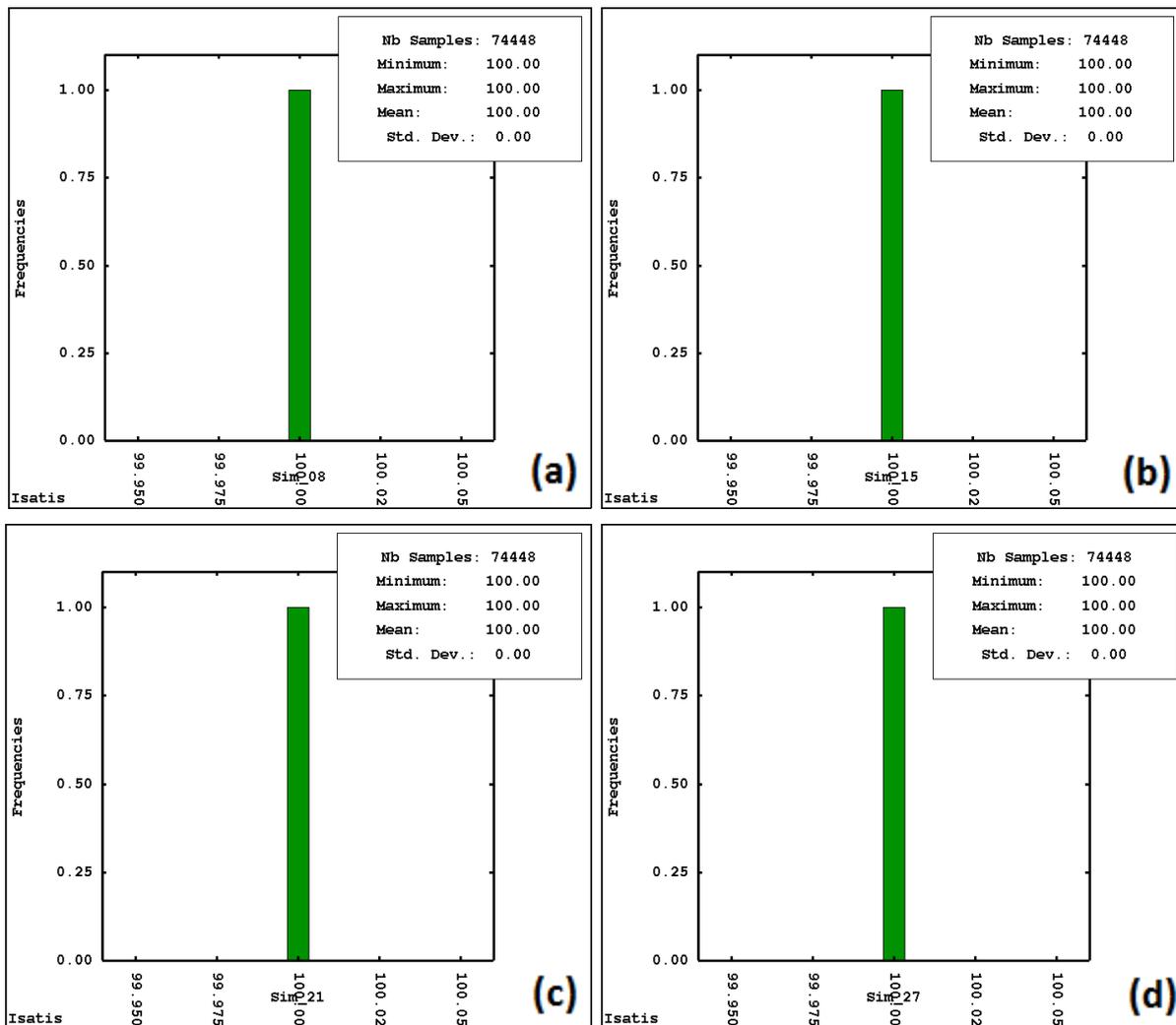


Figura 4.17. Histograma de fechamento nas simulações (a) Cenário 08, (b) Cenário 15, (c) Cenário 21, (d) Cenário 27

Capítulo 5 – Validações

Neste capítulo, é apresentado uma série de técnicas que são utilizadas para validar as estimativas e simulações feitas no capítulo anterior. As validações apresentadas correspondem às metodologias 1 e 2 referentes à utilização do banco de dados isotópico.

5.1. Reprodução da média global

O primeiro método de validação utilizado nesta dissertação é a reprodução da média global do depósito. Como já mencionamos anteriormente, muitas vezes os depósitos podem ter uma amostragem preferencial em zonas ricas. Portanto, para obter uma média mais realista do depósito, foi utilizado o método de desagrupamento da poligonal ou polígono de Voronoi, onde cada amostra recebe um peso proporcional a área do polígono (2D) ou poliedro (3D) formado na sua vizinhança (ver Isaaks e Srivastava, 1989, p. 238-241).

A tabela 5.1 mostra a média global desagrupada ponderada dos dados originais e cada uma das médias globais obtidas nas diferentes metodologias da krigagem, i.e. a tradicional (OK) e a krigagem das transformações logarítmicas isométricas (*ilr*), utilizando o banco de dados isotópico. Note que, tanto a krigagem tradicional como a krigagem (*ilr*) geraram resultados muito próximos em comparação com a média desagrupada. A diferença do erro relativo entre ambas às médias é menor que 5% na maioria das variáveis analisadas, excetuando o Fe400 na krigagem tradicional e o Fe14 e Fe400 na krigagem das transformações logarítmicas isométricas. Isto representa uma boa aderência das metodologias de estimativas abordadas neste trabalho em termos de valores globais.

Outro aspecto importante é que por meio da metodologia 2 (usando modelos únicos) pode-se quantificar o viés global, já que ambas as metodologias utilizaram a mesma estratégia de busca com o mesmo modelo de

estimativa (Rec 14 e Rec400). A tabela 5.1 mostra a percentagem do erro relativo entre as duas médias, mostrando que, globalmente esse erro não é maior que 7%.

Tabela 5.1. Média desagrupada vs Média das Estimativas.

Variável	# Blocos	Média Desagrupada	Metodologia 1 (Modelos Individuais)				Metodologia 2 (Modelos Únicos)				
			Média		% Erro Relativo		Média		% Erro Relativo		% Erro Viés
			OK	ilr(10x10)	OK	ilr(10x10)	OK	ilr(10x10)	OK	ilr(10x10)	
Alap14	74448	49.3	49.11	49.42	0.39	0.24	49.13	49.51	0.35	0.42	0.77
Fe14	74448	9.23	9	8.51	2.56	8.46	9.02	8.49	2.33	8.72	5.88
Sire14	74448	4.49	4.67	4.62	3.85	2.81	4.67	4.53	3.85	0.88	3.00
Ti14	74448	1.51	1.47	1.46	2.72	3.42	1.47	1.46	2.72	3.42	0.68
Resto14	74448	35.47	35.68	35.99	0.59	1.44	35.71	36.01	0.67	1.50	0.84
Alap400	74448	46.51	46.99	47.26	1.02	1.59	47.05	47.36	1.15	1.79	0.66
Fe400	74448	9.04	8.41	8.07	7.49	12.02	8.44	7.92	7.11	14.14	6.16
Sire400	74448	4.81	4.87	4.76	1.23	1.05	4.87	4.75	1.23	1.26	2.46
Ti400	74448	1.53	1.46	1.46	4.79	4.79	1.45	1.44	5.52	6.25	0.69
Resto400	74448	38.1	38.25	38.45	0.39	0.91	38.19	38.54	0.24	1.14	0.92
Rec14	74448	67.28	67.19	68.29	0.13	1.48	67.19	68.28	0.13	1.46	1.62
Rec400	74448	9.53	9.51	9.29	0.21	2.58	9.51	9.3	0.21	2.47	2.21
Rec400	74448	23.19	23.3	22.42	0.47	3.43	23.3	22.42	0.47	3.43	3.78

5.2. Reprodução da média local (análise de deriva)

O segundo método de validação refere-se à reprodução da média local. Basicamente, este método consiste em alinhar o modelo de blocos desagrupado e o modelo de blocos krigado em cada uma das metodologias estudadas. Depois, obtêm-se as médias locais ao longo do plano X, Y e Z em diferentes faixas e, por último, o resultado de cada média é plotado em um gráfico de média local vs a faixa analisada. A variabilidade da média local vai estar representada pela aderência entre ambas estimativas ao longo de faixa.

As figuras 5.1 e 5.2 mostram a análise de deriva para as variáveis Alap14 e Rec14 ao longo do plano X, Y e Z da metodologia 1 (*modelos individuais*). Note que, a média local mantém um comportamento muito similar nas duas variáveis ao longo de cada faixa. Isto mostra que existe uma boa aderência das médias locais entre o modelo de blocos estimado e o modelo de blocos desagrupado. O tamanho de cada faixa é de 200 metros no plano X, Y e de 1 metro no plano Z.

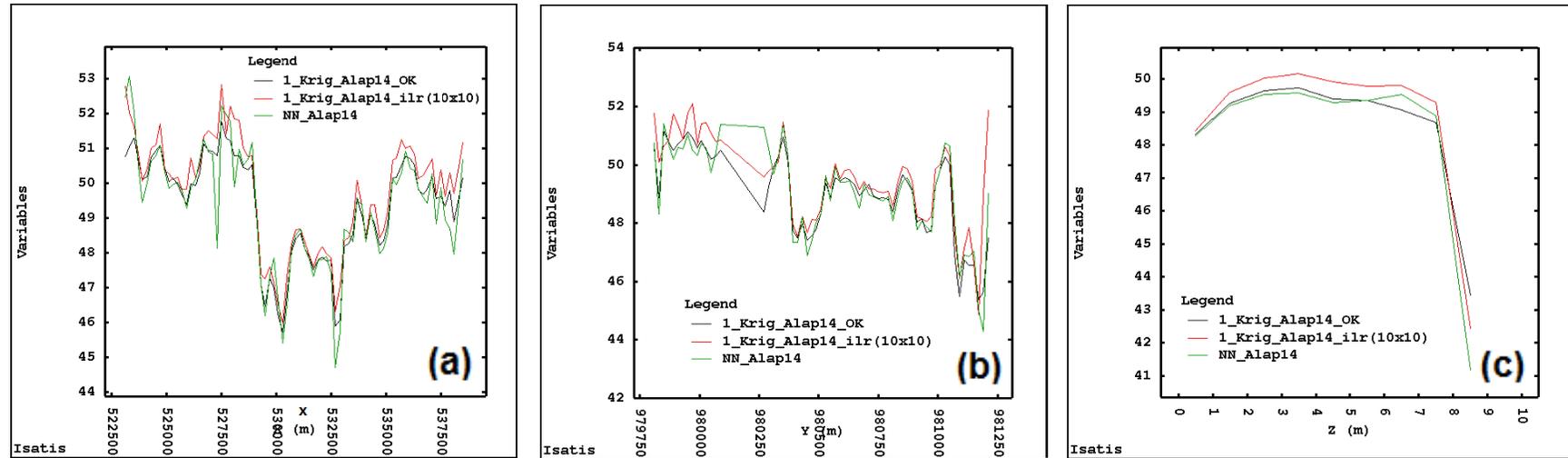


Figura 5.1. Análise de deriva da variável Alap14 entre a krigagem ordinária (OK), krigagem das (*itr*) e o desagrupamento por NN (a) Plano X, (b) Plano Y, (c) Plano Z.

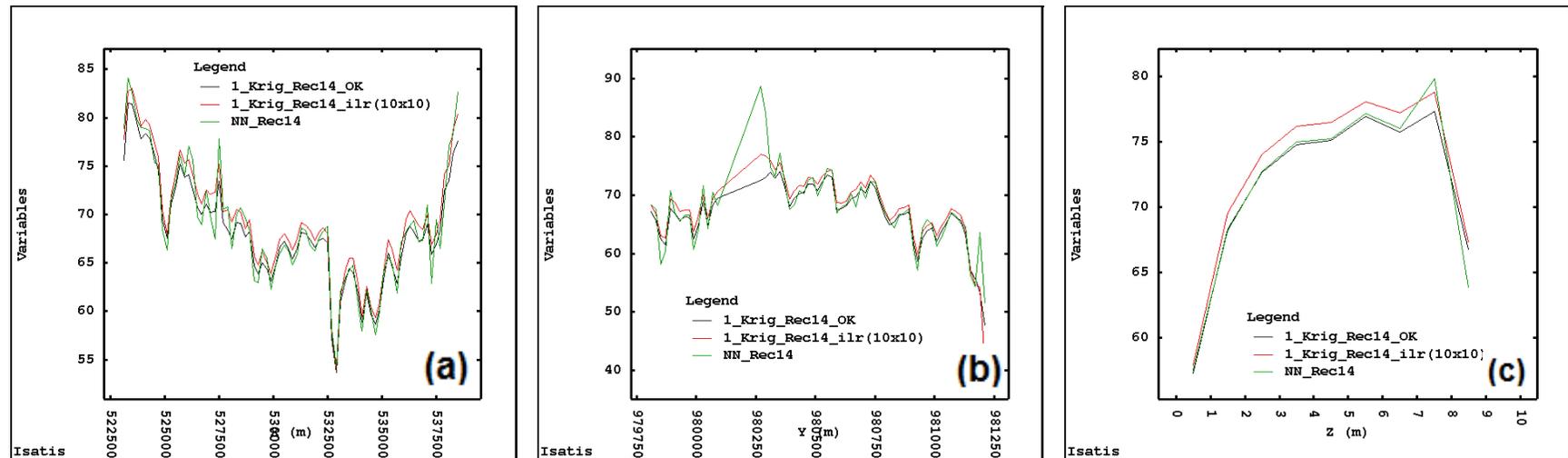


Figura 5.2. Análise de deriva da variável Rec14 entre a krigagem ordinária (OK), krigagem das (*itr*) e o desagrupamento por NN (a) Plano X, (b) Plano Y, (c) Plano Z.

5.3. Soma dos pesos positivos

O terceiro método de validação refere-se à percentagem de pesos positivos utilizados nas estimativas em relação à vizinhança da krigagem. A ideia principal desta validação é checar a existência de pesos negativos (indesejáveis).

As figuras 5.3 (a,b) mostram a soma dos pesos positivos da krigagem pontual das variáveis Rec14 e Rec400. Lembremos que essas variáveis são estimadas para serem usadas depois no processo de desacumulação das variáveis químicas tanto na faixa +14# como na faixa +400#. Note que, nos dois casos existe uma alta percentagem dos pesos positivos cuja soma é um. Isto garante uma boa distribuição dos pesos da krigagem na vizinhança adotada para este estudo com poucos pesos negativos. Um caso similar sucede nas figuras 5.4 (a,b) das variáveis *ilrRec_1* e *ilrRec_2* que correspondem à soma dos pesos positivos da krigagem pontual das transformações logarítmicas isométricas (*ilr*) respectivamente.

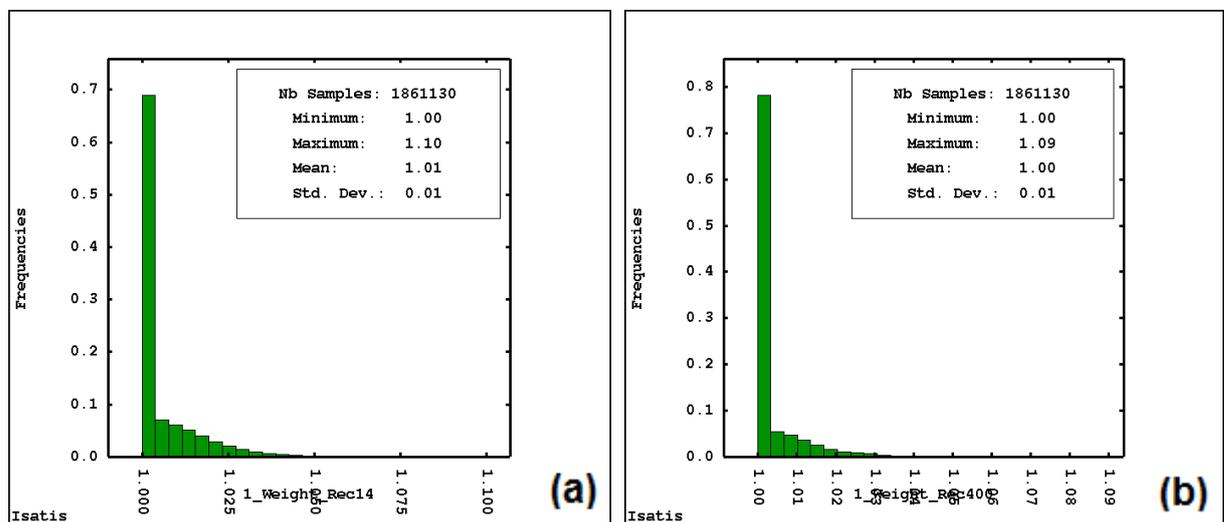


Figura 5.3. Soma dos pesos positivos da krigagem pontual (a) Rec14, (b) Rec400.

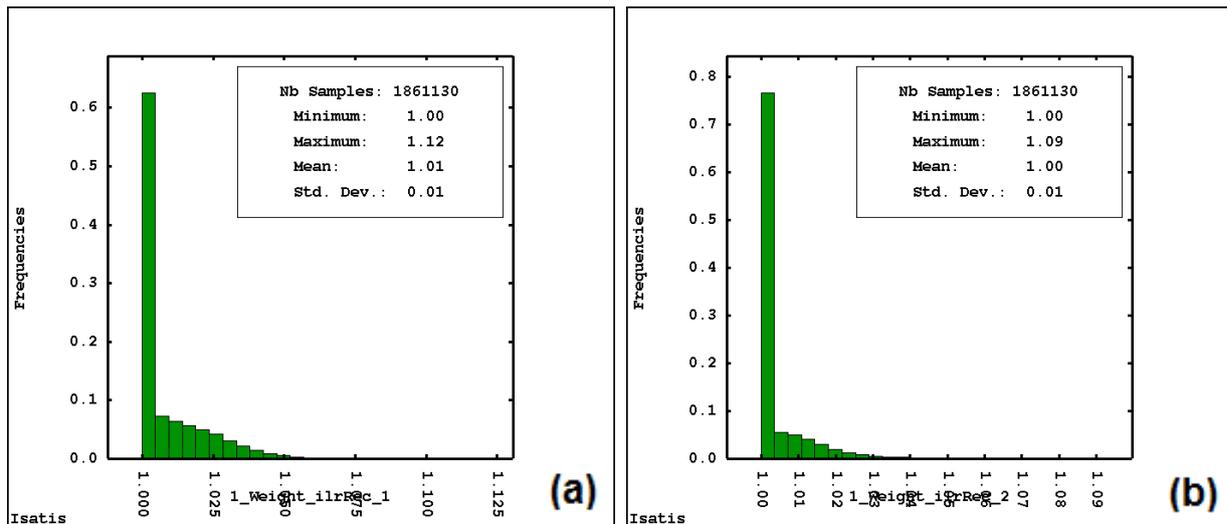


Figura 5.4. Soma dos pesos positivos da krigagem pontual (a) *ilrRec_1*, (b) *ilrRec_2*.

5.4. Análise visual

Outro método de validação utilizado nesta dissertação é a análise visual das estimativas em comparação com a posição espacial das amostras e seus teores. Sabe-se que, se existem regiões agrupadas com alto teor, é muito provável encontrar nessa mesma região, estimativas com essa mesma característica.

As figuras 5.5 (a-d) mostram as amostras das variáveis Alap14, Fe14, Sire14 e Rec14, comparadas com as estimativas obtidas por krigagem ordinária nas transformações (*ilr*). Note que, em todas as variáveis respeitou-se a distribuição das amostras nos locais onde havia teores altos e teores baixos, comparadas o modelo de blocos estimado em cada uma delas. O plano na qual foram plotadas as imagens é o plano XY.

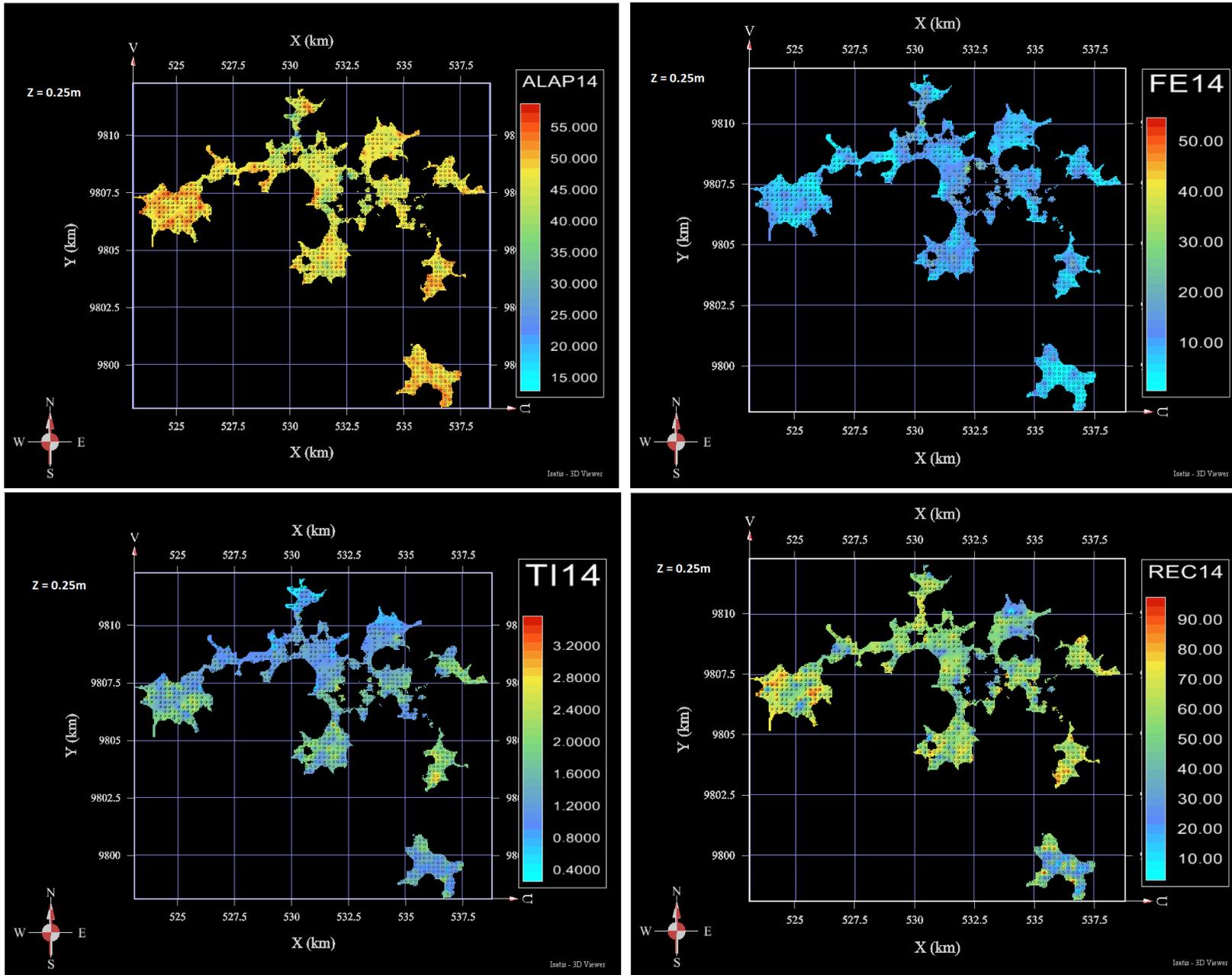


Figura 5.5. Análise visual das estimativas vs as amostras “Z=0,25m” (a) Alap14, (b) Fe14, (c) Ti14, (d) Rec14.

5.5. E-type das simulações vs estimativas e desagrupamento

Para observar se existe um viés significativo nas estimativas obtidas por krigagem das transformações (*ilr*), foi utilizada a média (E-type) das 30 simulações para cada variável utilizando a técnica de análise de deriva apresentado anteriormente. Lembremos que, como as transformações (*ilr*) são transformações não lineares, pode-se gerar um viés na retro-transformação após utilizar-se os dados transformados com interpoladores lineares como é o caso da krigagem. A questão está em observar se esse viés é realmente significativo e se causa um grande impacto nas estimativas.

A figuras 5.6 e 5.7 mostram a análise de deriva para as variáveis Alap14 e Rec14 ao longo do plano X, Y e Z da metodologia 1 (*modelos individuais*). Note que, o E-type das simulações (cor azul) segue um comportamento muito similar à krigagem das transformações (*ilr*) (cor vermelha). A média desagrupada (cor verde) e a média obtida por krigagem ordinária (cor preta) também seguem esse mesmo comportamento. Como conclusão, não há uma evidência significativa do viés gerado pelas estimativas por krigagem das transformações (*ilr*) que possa impactar nos teores estimados. O tamanho de cada faixa é de 200 metros no plano X, Y e de 1 metro no plano Z.

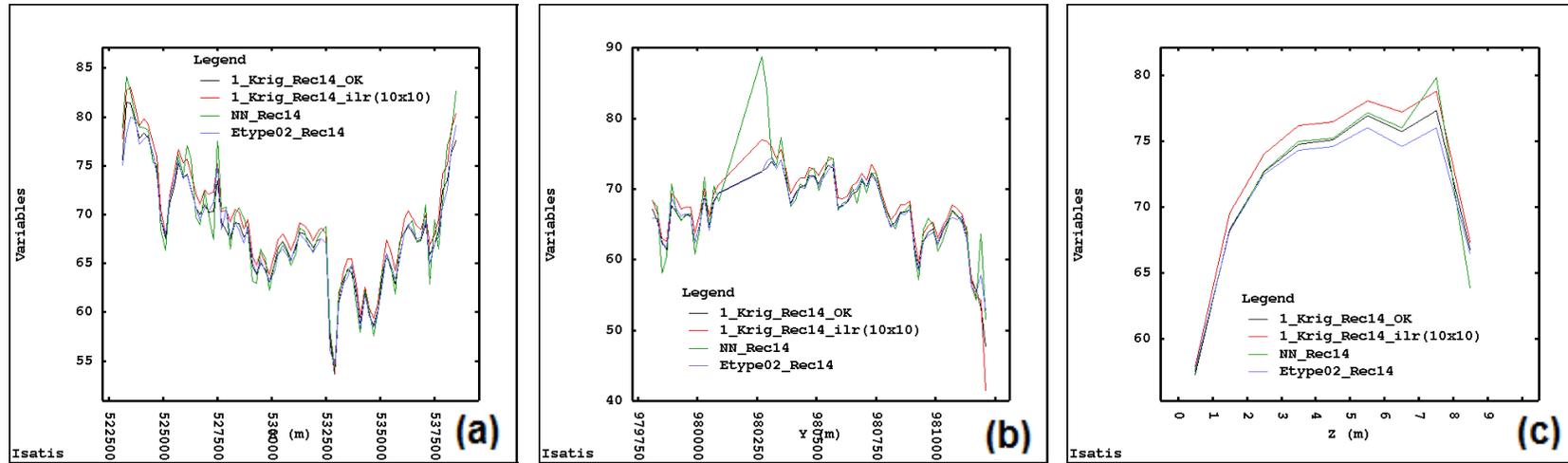


Figura 5.6. Análise de deriva da variável Rec14 entre o e-type das simulações, krigagem ordinária (OK), krigagem das (*ilr*) e o desagrupamento por NN (a) Plano X. (b) Plano Y. (c) Plano Z.

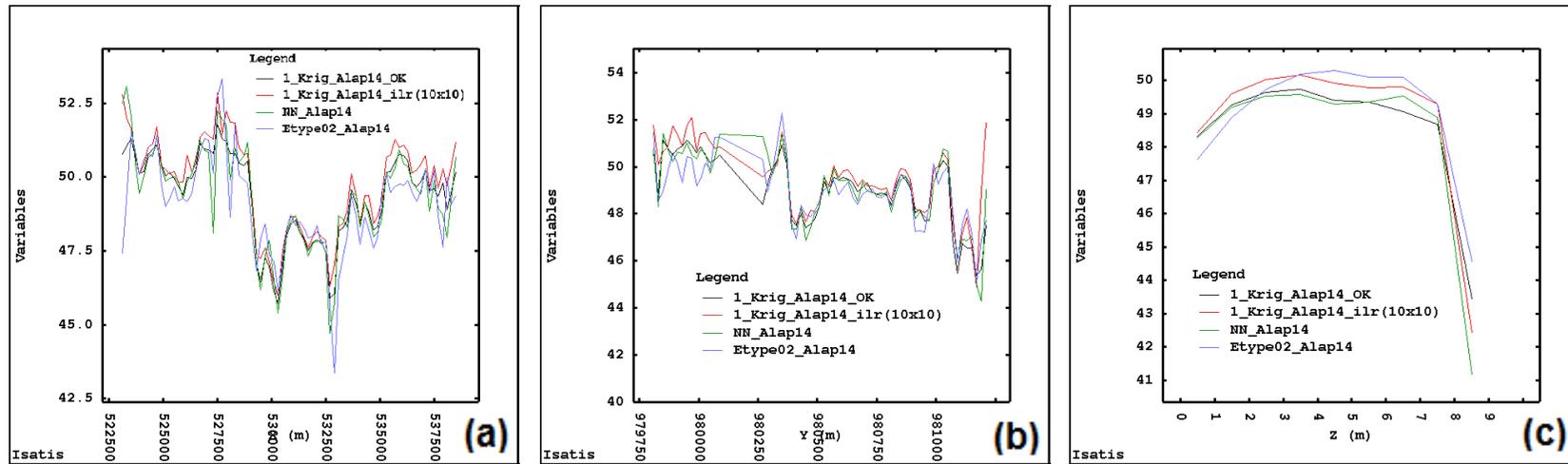


Figura 5.7. Análise de deriva da variável Alap14 entre o e-type das simulações, krigagem ordinária (OK), krigagem das (*ilr*) e o desagrupamento por NN (a) Plano X. (b) Plano Y. (c) Plano Z.

Capítulo 6

Conclusões e Recomendações

Neste capítulo, são apresentadas as conclusões obtidas durante o desenvolvimento deste trabalho. Além disso, será apresentada uma série de recomendações para trabalhos futuros.

5.1. Conclusões

Devido à grande variedade de metodologias abordadas neste trabalho, as conclusões são apresentadas em duas fases. Uma primeira refere-se às conclusões obtidas usando métodos geoestatísticos tradicionais (krigagem e cokrigagem) e a segunda refere-se à metodologia proposta (krigagem e simulação das transformações logarítmicas isométricas). As conclusões apresentadas neste trabalho estão baseadas no tipo de depósito que foi estudado.

5.1.1. Primeira fase

- A utilização da krigagem ordinária em depósitos multivariados isotópicos utilizando *modelos de continuidade espacial individuais* (cada variável com seu próprio modelo) gerou valores em alguns blocos fora dos limites mínimo e máximo dos dados originais. Isto é atribuído ao processo de desacumulação e não ao método interpolador em si, já que os dados se encontravam em suportes diferentes necessitando do uso da variável auxiliar acumulação. Além disso, não foram mantidos os balanços químicos e mássicos em cada uma das frações analisadas. A vantagem nesta metodologia é que cada variável é estimada respeitando o seu próprio modelo de continuidade espacial.

- Contrário ao caso anterior, a krigagem ordinária em depósitos multivariados isotópicos utilizando *modelos de continuidade espacial únicos* (cada variável utilizou o modelo do acumulador correspondente a sua própria faixa granulométrica) não gerou nenhum bloco fora do intervalo dos dados originais e manteve o fechamento dos balanços após estimativas. A desvantagem nesta metodologia é que as variáveis utilizam um único modelo de referência e muitas vezes esses modelos não se aderem muito bem ao variograma experimental calculado com os dados originais.
- A utilização da krigagem ordinária em depósitos multivariados heterotópicos utilizando *modelos únicos e modelos individuais*, não garante em nenhum dos casos, tanto o fechamento nos balanços como respeitar o intervalo delimitado pelo limite mínimo e máximo dos dados originais. A causa é o desequilíbrio na distribuição dos pesos da krigagem, já que o banco de dados heterotópico carece de informação de algumas variáveis em alguns lugares não analisados.
- A cokrigagem ordinária utilizando um modelo MCI nos dados isotópicos apresentou os mesmos resultados da krigagem ordinária utilizando um modelo único. A razão principal é que o modelo MCI usa o mesmo modelo para os variogramas diretos e cruzados, rescalonando unicamente o sill. Portanto, as condições de fechamento e limites dos dados foram mantidas.
- Uma última abordagem nesta primeira fase foi a utilização da cokrigagem ordinária em dados heterotópicos utilizando o mesmo modelo MCI. Contrário dos casos anteriores, esta metodologia gerou além de estimativas fora do intervalo dos dados, valores negativos. Isto devido a que a cokrigagem ordinária tradicional está condicionada a que a soma dos pesos da variável secundária seja zero, portanto, provavelmente alguns dados secundários recebem pesos negativos incrementando o risco de obter estimativas negativas.

5.2.2. Segunda fase

- A krigagem pontual das transformações logarítmicas isométricas (*ilr*) utilizando *modelos individuais* mostrou-se ser uma ferramenta eficiente para lidar com depósitos multivariados isotópicos, já que mesmo utilizando modelos diferentes no espaço simplex, as estimativas garantiram o fechamento dos balanços químicos e mássicos, além de produzir boas estimativas. Os blocos que ficaram fora do intervalo dos dados originais são atribuídos ao processo de desacumulação e não são representativos em relação ao modelo de blocos estimados (menos de 1%).
- Os resultados da krigagem pontual (*ilr*) utilizando *modelos únicos* mostraram-se muito similar aos resultados da krigagem tradicional, garantindo além dos valores mínimos e máximos dos dados, os balanços em cada uma das faixas analisadas. A utilização de um único modelo de referência garantiu que todas as variáveis recebessem os mesmos pesos da krigagem e, por conseguinte, o processo de desacumulação não foi afetado.
- A krigagem direta em blocos (*ilr*) utilizando os dois tipos de modelos (*únicos e individuais*) produziu resultados semelhantes à krigagem pontual (*ilr*), portanto, não foi possível observar uma diferença significativa entre ambos os métodos.
- Todas as técnicas de krigagem anteriormente mencionadas deveriam produzir um viés teórico na média aritmética por causa da utilização de transformações não lineares. No entanto, o e-type das simulações (que não gera viés) mostrou muita aderência com os resultados da krigagem, ou seja, não foi possível encontrar um viés significativo que impacte as estimativas. Assim, a simulação surge como uma alternativa para lidar com dados que exijam este tipo de transformações não lineares, retendo o E-type como estimativa do bloco.

5.2. Recomendações para trabalhos futuros

- Aplicar as transformações logarítmicas isométricas (*ilr*) em outros depósitos multivariados para contribuir com o desenvolvimento da metodologia proposta.
- As transformações logarítmicas isométricas (*ilr*) só podem ser aplicadas a depósitos multivariados isotópicos, ou seja, todas as variáveis possuem análise química ou mássica nos locais amostrados. Uma idéia interessante seria explorar algumas alternativas aonde o banco de dados seja heterotópico, já que a maioria dos depósitos minerais multivariados tem essa característica.
- A krigagem de transformações não lineares gera um viés na média aritmética quando os dados são retro-transformados. Uma alternativa interessante seria investigar a possibilidade de fazer outro tipo de transformações que, além de não apresentar esse problema, garanta o fechamento dos balanços após estimativas.

Referências Bibliográficas

Aitchison, J. (1981). *A new approach to null correlations of proportions*. *Mathematical Geology* 13 (2), pp. 175-189.

Aitchison, J. (1982). *The statistical analysis of compositional data (with discussion)*. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* 44 (2), pp. 139-177.

Aitchison, J. (1983). *Principal component analysis of compositional data*, *biometrika*, 70 (1), pp. 57-65.

Aitchison, J. (1984). *The statistical analysis of geochemical compositions*. *Mathematical Geology* 16 (6), pp. 531-564.

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data, Monographs on Statistics and Applied Probability*, Chapman & Hall Ltd, 416p.

Barceló-Vidal, C; Martín-Fernández, J. A. & Pawlowsky-Glahn, V. (2001). *Mathematical Foundations of Compositional Data Analysis*, Universitat de Girona, Spain, 20p.

Boezio, M. N. (2010). *Estudo das Metodologias Alternativas da Geoestatística Multivariada Aplicadas a Estimativas de Teores de Depósito de Ferro*, Universidade Federal Rio Grande do Sul, Porto Alegre – Brasil, 465p.

Boezio, M. N; Abichequer, L. A. & Costa, J. F. C. (2012). *Map Decomposition of Compositional Data to Estimate Grades in Iron Ore Deposits*. Universidade Federal Rio Grande do Sul, Brasil, 13p.

Boezio, M. N; Costa, J. F. C. & Koppe, J. C. (2012). *Cokriging of Additive Log-Ratios (alr) For Grade Estimation in Iron Ore Deposits*. Universidade Federal Rio Grande do Sul, Brasil, 11p.

Bragulat, E. J & Sala, C. H. (2003). *Comparison of Kriging Results of Regionalised Compositional Data Using Three Different Data Transformations*.

Case Study: Bauxites in Hungary. Universitat Politècnica de Catalunya (UPC), Spain. 6p.

Bragulat, E. J; Sala, C. H & Diblasi, A. M. (2002). *An Experimental Comparison of Cokriging of Regionalized Compositional Data Using Four Different Methods. Case Study: Bauxites in Hungary.* Universitat Politècnica de Catalunya (UPC), Spain. 6p.

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

Dagbert, M. (2001). *Comments on "The estimation of mineralized veins: a comparative study of direct and indirect approaches,"* by Marcotte and A. Boucher, *Exploration and Mining Geology*, 10(3):243-244.

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

Egozcue J. J. & Pawlowsky-Glahn, V. (2005). *Groups of Parts and Their Balances in Compositional Data Analysis.* Technical University of Catalonia, Spain, 34p.

Egozcue J. J. & Pawlowsky-Glahn, V. (2006). *Simplicial geometry for compositional data,* Technical University of Catalonia, Spain, 16p.

Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G. & Barceló-Vidal, C., (2003). *Isometric logratio transformations for compositional data analysis.* *Mathematical Geology* 35 (3), 279-300p.

Eynatten, H.; Pawlowsky-Glahn, V. & Egozcue, J. J. (2002). *Understanding Perturbation on the Simplex: A Simple Method to Better Visualize and Interpret Compositional Data in Ternary Diagrams.* Friedrich-Schiller-Universität Jena, Alemanha, 9p.

Filzmoser, P. & Hron, K. (2008). *Correlation Analysis for Compositional Data,* Vienna University of Technology, Austria, 15p.

Filzmoser, P; Hron, K. & Reimann, C. (2009). *Principal component analysis for compositional data with outliers*. Vienna University of Technology, Áustria, 12p.

Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*, Oxford University Press, Nova Iorque, 483p.

Gy, P. (1982). *Sampling of particulate materials, theory and practice, 2nd edition*, Developments in Geomathematics 4, Elsevier, ISBN 0-444-42079-7, Amsterdam, 431p.

Isaaks, E. H. e Srivastava, R. M. (1989). *An Introduction to Applied Geoestatistics*, Oxford University Press, New York, 561p.

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

Koppe, V C. (2005). *Análise de Incerteza Associada à Determinação da Velocidade de Onda Sônica em Depósitos de Carvão Obtida por Perfilagem Geofísica*. Universidade Federal Rio Grande do Sul, Porto Alegre – Brasil, 282p.

Koppe, V. C. (2009). *Metodologia para Comparar a Eficiência de Alternativas para disposição de Amostras*. Universidade Federal Rio Grande do Sul, Porto Alegre, 236p.

Kotschoubey, B.; Calaf, J. M.; Costa, A. C.; Leite, A. S. & Duarte, C. H. (2005). *Caracterização e Gênese dos Depósitos de Bauxita da Província Bauxitífera de Paragominas, Noroeste da Bacia do Grajaú, Nordeste do Pará/Oeste do Maranhão*. Universidade Federal do Pará, Pará – Brasil, 96p.

Krige, D. G. (1951). *A Statistical Approach to Some Mine Valuations and Allied Problems at the Witwatersrand*. MSc. Thesis, University of Witwatersrand, Johannesburg, South Africa.

Marat Abzalov. (2011). *Sampling Errors and Control of Assay Data Quality in Exploration and Mining Geology*, Applications and Experiences of

Quality Control, Prof. Ognyan Ivanov (Ed.), ISBN: 978-953-307-236-4, InTech, Available from: <http://www.intechopen.com/books/applications-and-experiences-of-qualitycontrol/sampling-errors-and-control-of-assay-data-quality-in-exploration-and-mining-geology>, 35p.

Marechal, A. (1970). *Cokrigage et Regression em Correlation Intrinsique*, Centre de Geostatistique de Fontainebleau, Fontainebleau, 40p.

Marques, D. M. (2010). *Desenvolvimento de um Algoritmo para Simular a Variabilidade do Minério em Pilhas de Homogeneização*. Universidade Federal Rio Grande do Sul, Porto Alegre, 130p.

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

Matheron, G. (1965). *Regionalized Variables and Their Estimation*, Paris, 306 p (Editions Masson: Paris).

Pawlowsky, G. V; Egozcue, J. J & Tolosana, D. R. (2010). *Lecture Notes on Compositional Data Analysis*. Technical University of Catalonia, Spain. 108p.

Pawlowsky-Glahn, V. & Buccianti, A. (2011). *Compositional Data Analysis - Theory and Applications*, University of Girona, Spain, 241p.

Pawlowsky-Glahn, V. & Egozcue J. J. (2006). *Compositional Data and their Analysis: An Introduction*. Geological Society, London. 11p.

Pawlowsky-Glahn, V. & Egozcue, J. (2001). *BLU Estimators and Compositional Data*. Universitat de Girona, Girona-Spain, 16p.

Pawlowsky-Glahn, V. & Olea, R. A. (2004). *Geostatistical Analysis of Compositional Data*, Studies in Mathematical Geology, Oxford University Press, 181p.

Pearson, K. (1897). *Mathematical Contributions to the Theory of Evolution on a Form of Spurious Correlation which May Arise when Indices are*

Used in the Measure of Organs, Anais da Royal Society of London, vol LX, p. 489-502.

Pitard, F.F. (1993). *Pierre Gy's sampling theory and sampling practice*, 2nd edition, CRC Press, ISBN 0-8493-8917-8, New York, 488p.

Rivoirard, J. (2003). *Course on Multivariate Geostatistics*. Ecole des Mines de Paris, Centre de Géostatistique, Fontainebleau – France, 76p.

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

Srivastava, R. M. (1987). *A Non-ergodic Framework for Variogram and Covariance Functions*. Master's thesis, Stanford University, Stanford, CA, 122p.

Thió-Henestrosa, S. & Martín-Fernández, J. A. (2005). *Dealing with Compositional Data: The Freeware CoDaPack*, Universitat de Girona, Spain, 21p.

Vieira, J. D. (2011). *Estudo de Alternativas de Estimativas para Jazidas de Ferro Utilizando Dados Heterotópicos*. Universidade Federal Rio Grande do Sul, Porto Alegre, 180p.

Wackernagel, H. (1994). *Cokriging versus Kriging in Regionalized Multivariate data Analysis*, Geoderma, 62, 83-92p.

Wackernagel, H. (1995). *Multivariate Geostatistics: An Introduction with Applications*, Springer-Verlag, Berlin, 274p.

Walvoort, D. J. & Gruijter, J. J. (2001). *Compositional Kriging: A Spatial Interpolation Method for Compositional Data*. Wageningen University and Research Centre, The Netherlands, 16p.