UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Escola de Engenharia

Programa de Pós-graduação em Engenharia de Minas, Metalúrgica e Materiais (PPGE3M)

USO DE DADOS DE DIFERENTE SUPORTE EM GEOESTATÍSTICA E DESENVOLVIMENTOS EM SIMULAÇÃO GEOESTATÍSTICA MULTIVARIADA

Marcel Antonio Arcari Bassani

Tese para obtenção do título de Doutor em Engenharia

Porto Alegre, RS
2018

Marcel Antonio Arcari Bassani

USO DE DADOS DE DIFERENTE SUPORTE EM GEOESTATÍSTICA E DESENVOLVIMENTOS EM SIMULAÇÃO GEOESTATÍSTICA MULTIVARIADA

Tese submetida ao Programa de Pós-graduação em Engenharia de Minas, Metalúrgica e Materiais (PPGE3M) da Universidade Federal do Rio Grande do Sul como requisito parcial à obtenção do título de Doutor em Engenharia

Orientador: Prof. PhD. João Felipe Coimbra Leite Costa

Porto Alegre, RS

2018

Marcel Antonio Arcari Bassani

USO DE DADOS DE DIFERENTE SUPORTE EM GEOESTATÍSTICA E DESENVOLVIMENTOS EM SIMULAÇÃO GEOESTATÍSTICA MULTIVARIADA

Esta tese foi julgada adequada para a obtenção do título de Doutor em Engenharia, na área de Tecnologia Mineral e Metalurgia Extrativa e aprovada em sua forma final pelo Orientador e pela Banca Examinadora designada pelo Programa de Pós-Graduação em Engenharia de Minas, Metalúrgica e Materiais (PPGE3M) da Universidade Federal do Rio Grande do Sul

_____

Prof. Dr. João Felipe Coimbra Leite Costa
PhD. pela University of Queensland
Orientador

_____

Prof. Dr. Carlos Pérez Bergmann
Dr. pela Rheinisch Westfälische Technische Hochschule Aachen
Coordenador do PPGE3M

BANCA EXAMINADORA

Dr. Diego Machado Marques – UFRGS _____

Dra. Maria Noel Morales Boezio – DINAMIGE _____

Dra. Vanessa Cerqueira Koppe – UFRGS _____

AGRADECIMENTOS

Ao professor PhD. João Felipe Coimbra Leite Costa pela orientação, ajuda, confiança e amizade durante meus estudos. João é realmente um excelente orientador, estando sempre disposto a discutir questões geoestatísticas.

Aos colegas do LPM, em especial Ricardo H., Cristina, Péricles, Áttila e Augusto pelos bons momentos de convivência.

Ao professor PhD. Clayton Deutsch pela orientação recebida durante meu período de estudos no exterior no *Centre for Computational Geostatistics* (CCG) na *University of Alberta*. As contribuições do professor Clayton enriqueceram bastante o trabalho.

Aos colegas do CCG, em especial Felipe, Diogo, Ana, Daniel, George e Connor. Felipe me ajudou bastante na minha chegada à cidade de Edmonton.

Aos integrantes do PPGE3M que trabalham para manter o nível de excelência do programa.

À minha família, meus pais Ernidio e Vania e meus irmãos Rafael, Ismael e Dinho.

Aos meus sogros José Renato e Roselaine.

À ALCOA e MRN pelos dados.

A CAPES pelo auxílio financeiro.

À minha noiva Waleska, que sempre esteve do meu lado me dando força e apoio. Serei sempre grato a ti!

RESUMO

Essa tese investiga três problemas: (1) o uso de dados de diferente suporte em geoestatística, (2) simulação multivariada com restrições e (3) verificação da distribuição multivariada.

Quando as amostras tem suporte diferente, essa diferença de suporte precisa ser considerada para construir um modelo de teores. A tese propõe a krigagem utilizando covariâncias médias entre as amostras para considerar dados de diferente suporte. A metodologia é comparada com dois métodos: (1) krigagem utilizando covariâncias pontuais entre os dados e (2) o método indireto. A krigagem utilizando covariâncias pontuais entre os dados ignora a diferença de suporte entre os dados. O método indireto trabalha com a variável acumulação, em vez do teor original. A krigagem com covariâncias médias resultou em estimativas mais precisas do que os outros dois métodos.

Depósitos minerais multivariados frequentemente têm variáveis que contém restrições de fração e soma. As restrições de fração ocorrem quando uma variável é parte da outra, como a Alumina Aproveitável e Alumina Total em um depósito de bauxita. A Alumina Aproveitável não pode ser maior do que a Alumina Total. Restrições de soma ocorrem quando a soma das variáveis não pode exceder um valor crítico. Por exemplo, a soma de teores não pode ser maior do que cem. A tese desenvolve uma metodologia para cosimular teores com restrições de soma e fração. As simulações reproduzem os histograms, variogramas e relações multivariadas e honram as restrições de soma e fração.

As simulações geoestatísticas multivariadas devem reproduzir as relações entre as variáveis. Dentro desse contexto, essa tese investiga a verificação da distribuição multivariada de simulações geoestatísticas. A tese desenvolve uma métrica de distância entre a distribuição multivariada dos dados e das simulações. A métrica desenvolvida foi efetiva para detectar erro e viés. Além disso, a métrica foi usada para comparar métodos de simulação geoestatística multivariada.

ABSTRACT

This thesis investigates three problems: (1) use of data of different support in geostatistics, (2) multivariate simulation with constraints and (3) verification of the multivariate distribution.

When the samples have different support, this difference in support must be considered to build a grade model. The thesis proposes kriging with average covariances between the data to consider data of different support. The methodology is compared with two methods: (1) kriging using point support covariances between the data and (2) the indirect approach. Kriging using point support covariances between the data ignores the difference in support between the data. The indirect approach works with the variable accumulation, instead of the original grade. Kriging with average covariances resulted in more precise estimates than the other two methods.

Multivariate mineral deposits often have variables that contain fraction and sum constraints. Fraction constraints occur when a variable is a fraction of the other, such as Recoverable and Total Alumina in a bauxite deposit. The Recoverable Alumina must not exceed Total Alumina. Sum constraints occur when the sum of the variables must not exceed a critical threshold. For instance, the sum of grades must not be above one hundred in a mineral deposit. The thesis develops a methodology to cosimulate grades with sum and fraction constraints. The simulations reproduce the histograms, variograms and multivariate relationships and honor the sum and fraction constraints.

Multivariate geostatistical simulations should reproduce the relationships between the variables. In this context, the thesis investigates the verification of the multivariate distribution of geostatistical simulations. The thesis develops a metric to measure the distance between the multivariate distributions of the data and the simulations. The metric developed was effective to detect error and bias. Moreover, the metric was used to compare multivariate simulation methods.

Keywords: support, kriging, multivariate simulation, constraints, checking

LISTA DE FIGURAS

Figura 1: Esquema da análise granulométrica. 3

Figura 2: Gráfico de dispersão esquemático entre duas variáveis com restrição de soma (a) e restrição de fração (b). 4

Figura 3: Esquema do cálculo de covariância bloco-a-bloco. Cada linha representa uma covariância ponto-a-ponto. 22

Figura 4: Mapa de localização das amostras. 27

Figura 5: Histograma de REC14 (a) diagrama quantil-quantil entre os dados originais e desagrupados (b) e histograma do comprimento das amostras (c). 28

Figura 6: Modelo de blocos e amostras. As linhas representam o modelo de blocos enquanto que os pontos representam as amostras. 31

Figura 7: Análise de deriva nas direções X (a), Y (b) e Z (c). 32

Figura 8: Cenário de referência. Estimativa do teor médio de um segmento de 3 m centrado no local **u** utilizando seis amostras quase pontuais. 35

Figura 9: Cenário das estimativas: estimativa do teor médio de um segmento de 3 m centrado no local **u** com uma amostra de linha e uma amostra quase pontual. Os pontos brancos na amostra na linha são pontos discretizantes usados para calcular as covariâncias entre os dados para a krigagem com amostras de diferente suporte. 36

Figura 10: Influência do efeito pepita no peso (a), na estimativa (b) e na variância do erro (c). O peso em (a) se refere ao peso da amostra de linha para o método indireto e para a krigagem com amostras de diferente suporte. Para a krigagem com amostras quase pontuais, o peso em (a) se refere à soma dos pesos das amostras 1-5 na figura 8. 39

Figura 11: Influência do alcance do variograma no peso (a), na estimativa (b) e na variância do erro (c). O peso em (a) se refere ao peso da amostra de linha para o método indireto e para a krigagem com amostras de diferente suporte. Para a krigagem com amostras quase pontuais, o peso em (a) se refere à soma dos pesos das amostras 1-5 na figura 8. 41

Figura 12: Influência do tipo de variograma no peso (a), na estimativa (b) e na variância do erro (c). O peso em (a) se refere ao peso da amostra de linha para o método indireto e para a krigagem com amostras de diferente suporte. No caso da krigagem com amostras quase pontuais, o peso em (a) se refere à soma dos pesos das amostras 1-5 na figura 8. 43

Figura 13: Exemplo de regularização pela espessura da camada. Furo de sondagem antes da regularização (a) e depois da regularização (b). 45

Figura 14: Variograma de Espessura (a) e variograma experimental da variável Acumulação junto com o modelo de variograma da variável Espessura (b) e diagrama de dispersão entre as variáveis Acumulação e Espessura (c). 48

Figura 15: Mapa de localização de um ponto a ser estimado e as três amostras mais próximas usadas na estimativa. A área dos círculos é proporcional ao comprimento das amostras.....50

Figura 16: Ilustração da função F(z).....54

Figura 17: Ilustração da função G-1(p).....55

Figura 18: Esquema gráfico da transformação *normal score*. 55

Figura 19: Representação de uma matriz de dados (a) e do vetor da primeira variável (b).....57

Figura 20: Esquema da restrição de soma para duas variáveis e vetor ortogonal à restrição de soma.....60

Figura 21: PPMT convencional e PPMT usando vetores ortogonais à restrição de soma. 61

Figura 22: Matriz de correlação das variáveis. 66

Figura 23: Workflows testados. 67

Figura 24: Gráfico de dispersão entre a variável original usada no numerador e a variável transformada: ST e A1 (a), ST e U2 (b), FE e A2 (c), FE e U3 (d), TI e A3 (e) e TI e U1 (f).....69

Figura 25: Reprodução do histograma da Alumina Total (AT) para os workflows testados.....71

Figura 26: Gráfico de dispersão das variáveis AT e AA dos dados originais (a) e das simulações para os 4 workflows testados: workflow I (b), workflow II (c), workflow III (d) e workflow IV (e). 73

Figura 27: Gráfico de dispersão das variáveis AT e FE dos dados (a) e das simulações para os 4 workflows testados: workflow I (b), workflow II (c), workflow III (d) e workflow IV (e). 74

Figura 28: Reprodução dos coeficientes de correlação para os 4 workflows: workflow I (a), workflow II (b), workflow III (c), e workflow IV (d).....75

Figura 29: Verificação da soma das variáveis AT, ST, FE e TI para os workflows testados.....76

Figura 30: Mínimo e máximo dos dados e das simulações. 78

Figura 31: Porcentagem de valores simulados fora do intervalo dos dados. 79

Figura 32: Mapa de localização das amostras com diferentes espaçamentos amostrais. 80

Figura 33: Relação entre a média desagrupada de AT e o tamanho de célula.....82

Figura 34: Fluxograma da metodologia. 83

Figura 35: Histogramas e gráficos de dispersão das duas primeiras variáveis transformadas PPMT. 84

Figura 36: Reprodução dos histogramas. 87

Figura 37: Reprodução dos variogramas na direção horizontal. 89

Figura 38: Reprodução dos variogramas na direção vertical. 90

Figura 39: Gráfico de dispersão entre os coeficientes de correlação dos dados e da primeira realização. 91

Figura 40: Gráfico de dispersão entre AT e FE dos dados originais (a) e da primeira realização (b). Gráfico de dispersão entre RC e ST dos dados originais (c) e da primeira realização (d). 92

Figura 41: Histograma das soma das variáveis AT, ST, FE e TI da primeira realização (a) e histograma das proporções de valores corrigidos para as 20 realizações (b). 93

Figura 42: Gráfico de dispersão entre AT e AA (a) e entre ST e SR (b) da primeira realização. A área azul corresponde à região que não viola a restrição de fração. 93

Figura 43: Mínimo e máximo dos dados e da primeira realização para todas as variáveis (a) e proporção de valores corrigidos para a variável AA (b). 94

Figura 44: Esquema do cálculo da cdf multivariada usando quantis (a) e os valores dos dados (b). A linha vermelha representa a envoltória convexa dos dados. 97

Figura 45: Esquema de duas variáveis com correlação negative. Os pontos cinza representam “zeros” da cdf multivariada. 100

Figura 46: Matriz de correlação das variáveis. 101

Figura 47: Porcentagem de zeros da cdf multivariada em função do número de variáveis. 102

Figura 48: *Boxplots* da cdf multivariada em função do número de variáveis. 102

Figura 49: Fluxograma da metodologia. 103

Figura 50: Gráfico de dispersão entre as variáveis X e Y dos dados originais (a) e dos dados com diferentes níveis de erro relativo: 2% (b), 4% (c), 6% (d), 8% (e) e 10% (f). 105

Figura 51: Gráfico de dispersão entre o erro relativo e o D90 (a), erro quadrático médio (b) e diferença entre coeficientes de correlação (c). 106

Figura 52: Gráfico de dispersão entre as variáveis X e Y dos dados originais (a) e dos dados com diferentes níveis de viés: 2% (b), 4% (c), 6% (d), 8% (e) e 10% (f). 107

Figura 53: Gráfico de dispersão entre o viés e o D90 (a), erro quadrático médio (b) e diferença entre coeficientes de correlação. 108

Figura 54: Gráfico de dispersão entre as variáveis X e Y dos dados originais (a) e dos dados com diferentes níveis de erro relativo: 2% (b), 4% (c), 6% (d), 8% (e) e 10% (f). 110

Figura 55: Gráfico de dispersão entre o erro relativo e o D90 (a), erro quadrático médio (b) e diferença entre coeficientes de correlação (c). 111

Figura 56: Gráfico de dispersão entre as variáveis X e Y dos dados originais (a) e dos dados com diferentes níveis de viés: 2% (b), 4% (c), 6% (d), 8% (e) e 10% (f). 112

Figura 57: Gráfico de dispersão entre o viés e o D90 (a), erro quadrático médio (b) e diferença entre coeficientes de correlação (c). 113

Figura 58: Matriz de correlação entre as variáveis. 116

Figura 59: Fluxograma das metodologias das simulações para o método direto e indireto. 118

Figura 60: Boxplots dos dados originais e primeira realização obtida com os métodos direto e indireto para a variável AT (a) e ST (b). 120

Figura 61: Histograma da proporção de valores corrigidos na correção de extrapolação para as variáveis AT (a) e ST (b). 121

Figura 62: Reprodução dos histogramas. 122

Figura 63: Gráfico de dispersão entre RC e ATA dos dados originais (a) e da primeira realização feita com o método indireto (b). 123

Figura 64: Reprodução do histograma de uma realização não condicional feita com o método direto e indireto. 124

Figura 65: Reprodução do variograma da variável AT. 125

Figura 66: Reprodução do variograma para a variável ST. 126

Figura 67: Gráfico de dispersão entre as variáveis AT e ST dos dados (a) e da primeira realização feita com o método direto (b) e indireto (c). 128

Figura 68: Histograma das soma das variáveis AT e ST da primeira realização feita com o método direto (a) e indireto (b). 129

Figura 69: Densidade de probabilidade da estatística D90 calculada sobre as 20 realizações para os dois métodos. 130

Figura 70: Arquivo de parâmetros do software *Block_Variogram*. 148

Figura 71: Arquivo com amostras de bloco. 149

Figura 72: Arquivo de saída do software *Block_Variogram*. 150

Figura 73: Arquivo de parâmetros do software Block_Vmodel. 151

Figura 74: Arquivo de saída do software *Block_Vmodel*. 152

Figura 75: Interface da aba *General and Data* do *plug-in Block_kriging_DH*. 153

Figura 76: Exemplo de arquivo de linha. 155

Figura 77: Amostras de linha carregadas no SGeMS. 155

Figura 78: Arquivo de parâmetros do software mv_d90. 157

Figura 79: Arquivo de parâmetros do software mvs_sum_check. 159

Figura 80: Arquivo de parâmetros do software mvs_frac_ratio. 160

LISTA DE TABELAS

Tabela 1: Comparativo entre a krigagem com amostras de diferente suporte e a krigagem utilizando covariâncias ponto-a-ponto 33

Tabela 2: Resumo dos parâmetros testados na análise de sensibilidade 37

Tabela 3: Sumário estatístico para REC14 e comprimento para os bancos de dados 3D e 2D. 46

Tabela 4: Comparativo entre a krigagem com amostras de diferente suporte e o método indireto. 49

Tabela 5: Estimativa de REC14 no ponto a ser estimado na figura 15. 51

Tabela 6: Variáveis e notações. 64

Tabela 7: Sumário estatístico dos dados. 65

Tabela 8: Etapas de pré-processamento dos dados. 70

Tabela 9: Variáveis utilizadas na etapa de PPMT. 70

Tabela 10: Modelos de variograma. 86

Tabela 11: Sumário estatístico das variáveis. 116

Tabela 12: Modelos de variograma das variáveis normal score. 117

Tabela 13: Variáveis de entrada e saída utilizadas na transformação PPMT. 119

LISTA DE SIGLAS

alr: *additive log-ratio* – razão logarítmica aditiva

ccdf: *conditional cumulative distribution function* – função de distribuição cumulativa condicional

cdf: *cumulative distribution function* – função de distribuição cumulativa

clr: *centred log-ratio* - razão logarítmica centrada

DSU: *direct semivariogram upscaling*

GSLIB: *geostatistical software library* – biblioteca de softwares geoestatísticos

ilr: *isometric log-ratio* – razão logarítmica isométrica

MAF: *minimum/maximum autocorrelation factors* – fatores de autocorrelação mínimo e máximo

MLC: modelo linear de coregionalização

PCA: *principal component analysis* – análise de componentes principais

PPMT: *projection pursuit multivariate transformation*

SGeMS: *Stanford Geostatistical Modeling Sofware*

SGS: *sequential Gaussian simulation* – simulação sequencial Gaussiana

SL: *scaling laws*

SMU: *selective mining unity* – unidade seletiva de lavra

SCT: *stepwise conditional transform*

LISTA DE SÍMBOLOS

$C(\mathbf{u},\mathbf{u}')$: covariância entre os pontos $\mathbf{u}$ e $\mathbf{u}'$

$\overline{C}(v_\alpha, v_\beta)$: covariância média entre os volumes v_α e v_β

$F(z)$: função de distribuição cumulativa dos dados originais

$G(y)$: função de distribuição cumulativa Gaussiana padrão

$G^{-1}(p)$: função quantil da distribuição Gaussiana padrão

$\gamma(\mathbf{h})$: modelo de variograma em suporte de ponto para o vetor de separação $\mathbf{h}$

$\hat{\gamma}(\mathbf{h})$: variograma experimental em suporte de ponto para o vetor de separação $\mathbf{h}$

$\gamma_v(\mathbf{h})$: modelo de variograma em suporte v para o vetor de separação $\mathbf{h}$

$\hat{\gamma}_v(\mathbf{h})$: variograma experimental em suporte v para o vetor de separação $\mathbf{h}$

$\bar{\gamma}(v,v)$: *gammabar* ou variograma médio para o volume v

$\bar{\gamma}(v,v_\mathbf{h})$: variograma médio entre o bloco v e o bloco v separado de uma distância $\mathbf{h}$

$\mathbf{h}$: vetor de separação

λ_α: peso associado ao dado na localização $\mathbf{u}_\alpha$

φ: transformação *normal score*

$\sigma_E^2(\mathbf{u})$: variância do erro de estimativa no local $\mathbf{u}$

$\mathbf{u}$: vetor de coordenadas

z: valor do atributo de interesse

y: valor *normal score*

$\mathbf{Z}$: matriz de dados multivariados

$\mathbf{z}$: vetor de dados correspondentes a uma variável

$\mathbf{S}^{-1/2}$: matriz utilizada na operação de *sphering*

$\mathbf{D}$: matriz de autovalores

$\mathbf{V}$: matriz de autovetores

$\sum 0$: matriz de covariância

$\mathbf{p}$: projeção dos dados multivariados ao longo de um vetor unitário

$\boldsymbol{\theta}$: vetor unitário

$\mathbf{X}$: matriz de dados antes da Gaussianização

I: índice utilizado para medir a não Gaussianeidade de uma projeção

$\mathbf{z}_{total}$: teor total

$\mathbf{z}_{frac}$: teor fracionário

$\mathbf{f}$: razão de fração

$D90$: medida de diferença entre distribuições multivariadas

SUMÁRIO

1 INTRODUÇÃO 1

1.1 PROBLEMA 1

1.1.1 Amostras com diferente suporte 1

1.1.2 Simulação geoestatística multivariada com restrições 3

1.1.3 Verificação da distribuição multivariada 5

1.2 OBJETIVOS 6

1.3 REVISÃO BIBLIOGRÁFICA 7

1.3.1 Estimativa com dados de diferente suporte 7

1.3.2 Mudança de suporte no variograma 9

1.3.3 Verificação da distribuição multivariada 11

1.3.4 Simulação geoestatística multivariada com restrições 12

1.4 ORGANIZAÇÃO DA TESE 15

2 TEORIA SOBRE ESTIMATIVA COM DADOS DE DIFERENTE SUPORTE E DECONVOLUÇÃO DO VARIOGRAMA 17

2.1 ESTIMATIVA COM DADOS DE DIFERENTE SUPORTE 17

2.1.1 Estimativa e variância do erro de estimativa 17

2.1.2 Método indireto 18

2.1.3 Krigagem com amostras de diferente suporte 20

2.2 REGULARIZAÇÃO E DECONVOLUÇÃO DO VARIOGRAMA 22

2.2.1 Variograma experimental para amostras de diferente suporte 22

2.2.2 Regularização do variograma para blocos de forma regular 23

2.2.3 Regularização do variograma para blocos de forma irregular 24

2.2.4 Deconvolução do variograma 25

3 ESTUDO DE CASO: KRIGAGEM COM AMOSTRAS DE DIFERENTE COMPRIMENTO 26

3.1 APRESENTAÇÃO DO BANCO DE DADOS 26

3.2 ANÁLISE E MODELAGEM DO VARIOGRAMA 28

3.3 ESTIMATIVA....29
3.4 VALIDAÇÃO DO MODELO....29
3.5 COMPARAÇÃO COM KRIGAGEM USANDO COVARIÂNCIAS PONTO-A-PONTO ENTRE OS DADOS....30
3.6 RESULTADOS....30
3.6.1 Validação do modelo....30
3.6.2 Comparação com krigagem utilizando covariâncias ponto-a-ponto entre as amostras....32
4 COMPARATIVO ENTRE MÉTODO INDIRETO E KRIGAGEM COM AMOSTRAS DE DIFERENTE SUPORTE....34
4.1 EXEMPLO SIMPLES....34
4.1.1 Referência: krigagem utilizando amostras quase pontuais....34
4.1.2 Cenário das estimativas: amostra de linha e amostra pontual....35
4.1.3 Análise de sensibilidade....36
4.1.4 Resultados....37
4.2 ESTUDO COMPARATIVO COM VALIDAÇÃO CRUZADA....44
4.2.1 Banco de dados....44
4.2.2 Variograma de REC14....47
4.2.3 Variogramas de Acumulação e Espessura....47
4.2.4 Estimativa e validação cruzada....48
4.2.5 Resultados da validação cruzada....49
4.3 OBSERVAÇÕES....51
5 SIMULAÇÃO GEOESTATÍSTICA E TRANSFORMAÇÕES MULTIVARIADAS....53
5.1 SIMULAÇÃO SEQUENCIAL GAUSSIANA....53
5.2 TRANSFORMAÇÃO *NORMAL SCORE*....54
5.3 *PROJECTION PURSUIT MULTIVARIATE TRANSFORM*....56
5.3.1 Notação dos dados....56
5.3.2 Sphering....57
5.3.3 Projeção....58
5.3.4 PPMT....58

5.4 PPMT COM VETORES CONTROLADOS 60
5.5 RAZÕES 61
5.5.1 Razões A *61*
5.5.2 Razões U *62*
5.5.3 Razão de fração *62*
6 COMPARAÇÃO DE TRANSFORMAÇÕES MULTIVARIADAS E SIMULAÇÃO GEOESTATÍSTICA PARA DADOS COM RESTRIÇÕES DE FRAÇÃO E SOMA 64
6.1 BANCO DE DADOS 64
6.2 COMPARAÇÃO ENTRE TRANSFORMAÇÕES MULTIVARIADAS PARA DADOS MULTIVARIADOS COM RESTRIÇÕES DE SOMA E FRAÇÃO 66
6.2.1 Workflows *66*
6.2.2 Resultados *71*
6.3 SIMULAÇÃO GEOESTATÍSTICA 80
6.3.1 Descrição espacial *80*
6.3.2 Desagrupamento *81*
6.3.3 Metodologia *82*
6.3.4 Análise da continuidade espacial *85*
6.3.5 Resultados *87*
6.4 OBSERVAÇÕES 94
7 VERIFICAÇÃO DE DISTRIBUIÇÕES MULTIVARIADAS 96
7.1 FUNÇÃO DE DISTRIBUIÇÃO CUMULATIVA MULTIVARIADA 96
7.2 MÉTRICAS DE DISTÂNCIA ENTRE CDFS MULTIVARIADAS 97
7.2.1 Estatística D90 ..... *97*
7.2.2 Erro quadrático médio entre cdfs multivariadas *98*
7.2.3 Diferença entre coeficiente de correlação *98*
7.3 ZEROS DA CDF MULTIVARIADA 99
7.4 COMPARAÇÃO ENTRE MÉTRICAS 103
7.4.1 Metodologia *103*
7.4.2 Caso I: dados com relação linear *104*
7.4.3 Caso II: dados com relação não linear *108*

7.5 OBSERVAÇÕES 114

8 COMPARATIVO ENTRE MÉTODO DIRETO E INDIRETO EM SIMULAÇÃO GEOESTATÍSTICA MULTIVARIADA 115

8.1 BANCO DE DADOS 115
8.2 ANÁLISE DA CONTINUIDADE ESPACIAL 117
8.3 SIMULAÇÃO GEOESTATÍSTICA 118
8.4 RESULTADOS 119
8.4.1 Correção de extrapolação no método indireto 119
8.4.2 Reprodução dos histogramas 121
8.4.3 Reprodução dos variogramas 124
8.4.4 Reprodução das relações bivariadas 126
8.5 OBSERVAÇÕES 130

9 CONCLUSÕES 132

9.1 CONTRIBUIÇÕES DA TESE 132
9.1.1 Amostras de diferente suporte 132
9.1.2 Simulação geoestatística multivariada com restrições 134
9.1.3 Verificação da distribuição multivariada 135
9.2 LIMITAÇÕES DAS TÉCNICAS 135
9.2.1 Krigagem com amostras de diferente suporte 135
9.2.2 Simulação geoestatística multivariada com restrições 136
9.2.3 Verificação da distribuição multivariada 137
9.3 TRABALHOS FUTUROS 137
9.3.1 Amostras de diferente suporte 137
9.3.2 Simulação geoestatística multivariada com restrições 138
9.3.3 Verificação da distribuição multivariada 139

REFERÊNCIAS 140

APÊNDICE A: *SOFTWARES* 147

A.1 DECONVOLUÇÃO DO VARIOGRAMA 147
A.1.1 Block_Variogram 147

A.1.2 Block_Vmodel *150*
A.2 ESTIMATIVA COM AMOSTRAS DE DIFERENTE SUPORTE: BLOCK_KRIGING_DH 152
A.2.1 Interface *152*
A.2.2 Banco de dados em formato de linha *154*
A.3 PÓS-PROCESSAMENTO DE SIMULAÇÃO GEOESTATÍSTICA MULTIVARIADA 156
A.3.1 mv_d90 ..... *156*
A.3.2 mvs_sum_check *157*
A.3.3 mvs_frac_ratio *159*

1 Introdução

A seção 1.1 mostra os problemas abordados na tese. Os desafios sobre uso de amostras de diferente suporte em geoestatística, simulação geoestatística multivariada com restrições e verificação da distribuição multivariada são explicados nessa seção. A seção 1.2 lista os objetivos da tese. A seção 1.3 revisa a literatura relevante e a seção 1.4 explica a organização da tese.

1.1 Problema

1.1.1 Amostras com diferente suporte

Em relação ao uso de amostras de diferente suporte nas estimativas, a abordagem tradicional trabalha com a variável acumulação (Journel e Huijbregts, 1978; Krige, 1978; Bertoli *et al.*, 2003; Marques *et al.*, 2014). Acumulação é o produto do teor pela espessura quando as amostras tem diferente comprimento. A estimativa de teor é obtida através da divisão da estimativa da acumulação pela estimativa da espessura. Essa abordagem elimina a componente vertical (Z), resultando em um modelo 2D. O problema desse método é que um modelo de blocos 2D não é adequado para testar diferentes seletividades de lavra verticais. Além disso, um modelo de teores em 3D é necessário para algoritmos de otimização de cava. Outro problema de trabalhar com acumulação é a possível ocorrência de teores fora do intervalo definido pelo mínimo e máximo dos dados.

Quando um modelo 3D é necessário, a abordagem comum para estimar teores utilizando amostras de diferente suporte envolve os seguintes passos: (1) regularizar as amostras e (2) estimar utilizando as amostras regularizadas. Entretanto, a regularização das amostras não é simples quando há grande variação no comprimento das amostras. Regularizar para um comprimento maior pode não ser possível nas regiões que a camada de minério é delgada. Por exemplo, se a camada de minério possui 0.50 m de espessura, não é possível criar uma composta de 1.00 m sem misturar teores de diferentes domínios geológicos. Por outro lado, regularizar para um comprimento menor

resulta em pedaços de igual teor, fato que é incorreto e ao mesmo tempo reduz artificialmente a variabilidade a curta distância.

Outro método de lidar com amostras de diferente suporte é através do uso de covariâncias médias no sistema de krigagem (Deutsch *et al.*, 1996; Yao e Journel, 2000; Tran *et al.*, 2001; Kyriakidis, 2004; Pardo-Igúzquiza *et al.*, 2006; Hansen e Mosegaard, 2008; Liu e Journel, 2009; Poggio e Gimona, 2013). Entretanto, a krigagem com covariâncias médias é pouco utilizada na mineração. Bassani *et al.* (2014) utilizaram a krigagem com covariâncias médias para considerar dados provenientes de stopes já minerados. Nessa tese, a krigagem com covariâncias médias será utilizada para incorporar amostras de sondagem com diferente comprimento.

Covariâncias médias não podem ser utilizadas diretamente quando o suporte das amostras não está definido como um volume no espaço. Isso ocorre quando os teores são analisados por faixa granulométrica. Nessa situação, a amostra de testemunho é fragmentada e feita uma análise granulométrica (figura 1). Análise granulométrica consiste em passar o material em um conjunto de peneiras empilhadas. As peneiras com abertura de malha maior ficam no topo. O resultado é uma massa retida em cada peneira. A massa retida em uma determinada peneira dividida pela massa total é chamada de recuperação mássica da faixa granulométrica. Para cada faixa granulométrica, o teor é analisado. O teor analisado por faixa granulométrica está associado ao suporte mássico representado pela recuperação mássica.

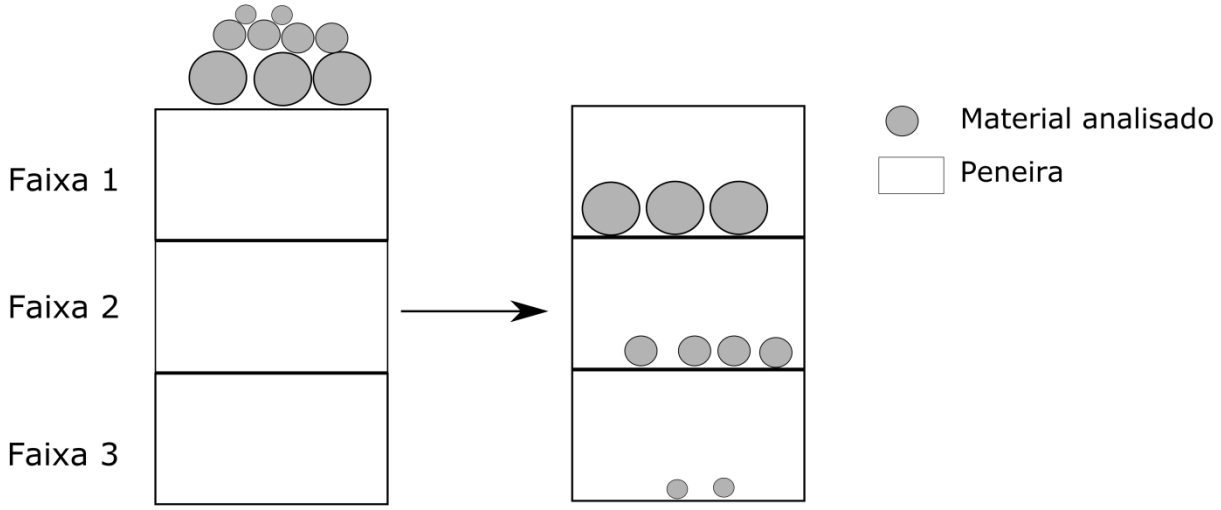

Figura 1: Esquema da análise granulométrica.

Depósitos de bauxita têm os teores analisados por faixa granulométrica. A principal espécie química desse tipo de depósito é o teor de Alumina. Essa espécie química em geral tem baixo coeficiente de variação. Marcotte e Boucher (2001) mostraram que os métodos direto e indireto resultam em estimativas similares quando a variável estimada tem pequeno coeficiente de variação e correlação baixa com a variável ponderadora. Entretanto, a diferença entre o método direto e indireto para simulação geoestatística não foi abordada. Nessa tese, é feito um estudo comparativo entre o método direto e indireto para simulação geoestatística multivariada em um depósito de bauxita.

1.1.2 Simulação geoestatística multivariada com restrições

A simulação geoestatística permite obter uma série de realizações que honram os dados amostrais, o histograma dos dados e a continuidade espacial. Essas múltiplas realizações permitem quantificar a incerteza dos teores, volumes e massa do minério. No caso de simulação geoestatística multivariada, os modelos de teores obtidos por simulação devem reproduzir as relações entre as variáveis. No caso das variáveis terem restrições de soma e fração, as simulações geoestatísticas devem também respeitar essas restrições.

Restrições de soma ocorrem quando a soma de algumas variáveis deve ser menor ou igual a uma constante. Por exemplo, a soma de todos os teores não pode ser superior a 100%. A figura 2a mostra o gráfico de dispersão de duas variáveis com restrição de soma, onde a soma das variáveis X e Y é sempre menor do que 100. Restrições de fração ocorrem quando uma variável é uma fração da outra. Por exemplo, em um depósito de bauxita, a Alumina Aproveitável é uma fração da Alumina Total. Como resultado, a Alumina Aproveitável não pode ser superior à Alumina Total em qualquer local de grid (simulado ou estimado). A figura 2b mostra o gráfico de dispersão de duas variáveis com restrição de fração, onde a variável X é sempre maior do que a variável Y.

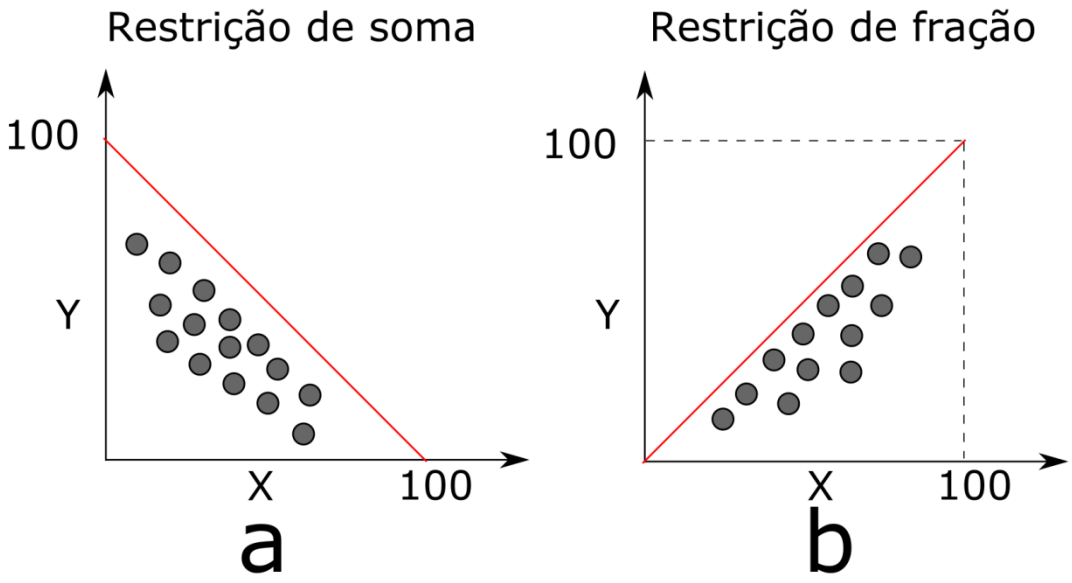

Figura 2: Gráfico de dispersão esquemático entre duas variáveis com restrição de soma (a) e restrição de fração (b).

A maioria dos algoritmos de simulação de múltiplas variáveis assume que os dados seguem uma distribuição multi-Gaussiana. Entretanto, os dados em geral não seguem essa distribuição. Em vista disso, há uma série de técnicas para transformar os dados de forma que eles sigam uma forma mais multi-Gaussiana. Uma prática comum é transformar os dados em dados Gaussianos univariados utilizando a transformação *normal score* (Deutsch e Journel, 1998) e assumir que eles são multi-Gaussianos. Considerando que os dados são multi-Gaussianos, as relações entre as variáveis são

então modeladas através do Modelo Linear de Corregionalização (MLC) (Journel e Huijbregts, 1978).

A modelagem do MLC é difícil à medida que o número de variáveis aumenta, porque é necessário modelar os variogramas diretos e cruzados. Como resultado, foram desenvolvidas diversas técnicas de descorrelação lineares, como *Principal Component Analysis* (PCA) (Goovaerts, 1993) e *Minimum/Maximum Autocorrelation Factors* (MAF) (Switzer e Green, 1984; Desbarats e Dimitrakopoulos, 2000). As variáveis descorrelacionadas podem então ser modeladas de maneira independente. Entretanto, as técnicas de descorrelação lineares não tornam as variáveis independentes no caso de relações complexas. Em vista disso, foram desenvolvidas transformações não lineares para tornar as variáveis multi-Gaussianas e independentes, como a *Stepwise Conditional Tranformation* (SCT) (Leuangthong e Deutsch, 2003) e a *Projection Pursuit Multivariate Transform* (PPMT) (Barnett *et al.*, 2014). A SCT exige uma grande quantidade de dados e não é prática em casos multivariados com várias variáveis. Por outro lado, a transformação PPMT é robusta em relação ao número de variáveis.

A PPMT provou reproduzir bem as relações entre variáveis (Barnett *et al.*, 2016). Entretanto, restrições de soma e fração não são consideradas diretamente na PPMT. Quando os dados tem restrição de soma, razões logarítmicas são frequentemente utilizadas (Pawlowsky-Glahn e Olea, 2004; Boisvert *et al*., 2013; Barnett e Deutsch, 2012). No caso de restrições de fração, Emery (2012) e Hosseini e Asghari (2015) usaram a SCT. Esses estudos não mostram simultaneamente restrições de soma e fração. A tese busca desenvolver uma metodologia para a simulação geoestatística de múltiplas variáveis com restrições de soma e fração. Além de reproduzir os histogramas, variogramas e coeficientes de correlação, as simulações devem reproduzir as relações entre as variáveis e respeitar as restrições de soma e fração.

1.1.3 Verificação da distribuição multivariada

As simulações geoestatísticas multivariadas devem reproduzir as relações entre as variáveis. Geomodeladores muitas vezes verificam se as simulações reproduzem os

coeficientes de correlação e variogramas cruzados dos dados (Leuangthong e Deutsch, 2003; Horta e Soares, 2010; Barnett *et al.*, 2014; Manchuk *et al.*, 2017; Mery *et al.*, 2017). Uma limitação dessas medidas é que elas medem apenas relações lineares entre as variáveis. Outra limitação é que elas medem apenas a relação entre duas variáveis. Dentro desse contexto, a tese busca desenvolver uma medida de diferença entre distribuições multivariadas que caracterize o seguinte: (1) relações lineares e não lineares e (2) relações entre duas ou mais variáveis.

1.2 Objetivos

A tese aborda três problemas principais:

- Integração de dados de diferente suporte em geoestatística;
- Simulação geoestatística multivariada com restrições;
- Verificação da distribuição multivariada.

Em relação ao problema da integração de dados de diferente suporte, os objetivos são os seguintes:

- Investigar o efeito de utilizar covariâncias médias no sistema de krigagem nas estimativas;
- Analisar o efeito de usar o método indireto nas estimativas;
- Adaptar *softwares* de krigagem para considerar a diferença de suporte no cálculo das covariâncias;
- Comparar o método direto e indireto no caso de simulação geoestatística multivariada de teores analisados por faixa granulométrica.

Em relação aos problemas da simulação geoestatística multivariada com restrições e verificação da distribuição multivariada, os objetivos são os seguintes:

- Desenvolver uma metodologia de simulação geoestatística multivariada que reproduza os dados condicionais, histogramas, variogramas, relações entre as variáveis e que respeitam as restrições de soma e fração;

- Aplicar a metodologia desenvolvida de simulação geoestatística multivariada com restrições em um estudo de caso;

- Desenvolver uma medida de diferença entre distribuições multivariadas que caracterize relações lineares e não lineares e relações entre duas ou mais variáveis;

- Aplicar a medida de diferença entre distribuições multivariadas para comparar métodos de simulação geoestatística multivariada.

1.3 Revisão bibliográfica

1.3.1 Estimativa com dados de diferente suporte

1.3.1.1 Krigagem

A krigagem de blocos (Journel e Huijbregts, 1978; Isaaks e Srivastava, 1989 e Goovaerts, 1997) utiliza covariâncias para relacionar amostras de suporte pontual com locais onde se deseja estimar definidos sobre um volume, que é chamado de bloco. A relação entre os dados de diferentes suportes é definida pelas covariâncias ponto-a-bloco e bloco-a-bloco. As covariâncias ponto-a-bloco e bloco-a-bloco são obtidas através de covariâncias ponto-a-ponto médias. A estimativa utilizando covariância ponto-a-bloco é igual à média das estimativas pontuais dentro do bloco (Journel e Huijbregts, 1978; Isaaks e Srivastava, 1989 e Goovaerts, 1997). Assim, o uso de covariâncias médias preserva a relação linear de uma variável aditiva definida em dois suportes distintos.

Quando as amostras estão definidas em suportes distintos, o sistema de krigagem pode ser generalizado para considerar a diferença de suporte entre as amostras (Journel e Huijbregts, 1978; Goovaerts, 1997; Isaaks e Srivastava, 1989). Nesse caso, a covariância entre as amostras são calculadas como covariâncias bloco-a-bloco.

O uso de covariâncias médias no sistema de krigagem para integrar dados de diferente suporte é bem estabelecido na literatura. Na área de sensoriamento remoto, Gotway e Young (2002) e Kyriakidis (2004) utilizaram krigagem com covariâncias

médias para estimar valores pontuais a partir de dados que representam uma área no espaço. Pardo-Igúzquiza *et al.* (2006) utilizaram cokrigagem com covariâncias médias para aumentar a resolução de uma imagem de satélite (*downscaling*). A imagem de satélite era composta por bandas multiespectrais de menor resolução (maior pixel) e de uma banda pancromática de maior resolução (menor pixel).

Goovaerts (2010) apresentou dois estudos de caso nos quais a krigagem de blocos foi utilizada para incorporar amostras de suporte pontual e areal. No primeiro estudo de caso, relacionado às ciências do solo, Goovaerts (2010) estimou a concentração de cromo. As amostras em suporte areal representavam o teor médio de cromo para cada unidade geológica. No segundo estudo de caso, relacionado às ciências médicas, a informação areal representava a incidência de câncer de mama em regiões censitárias. Nos dois estudos apresentados, a krigagem com amostras de diferente suporte foi comparada com a krigagem ordinária (feita apenas com as amostras pontuais) e a krigagem do resíduo. A krigagem com amostras de diferente suporte obteve estimativas mais acuradas do que as outras duas alternativas.

Liu e Journel (2009) publicaram um conjunto de *softwares* (*Bgeostats*) para integração de dados de diferentes suportes. Esse pacote foi implementado como uma série de *plug-ins* do *software SGeMS* (Remy *et al.*, 2008). O *plug-in* BKRIG faz estimativa com krigagem simples e ordinária e o *plug-in* BSSIM faz simulação sequencial direta.

1.3.1.2 Método indireto

Journel e Huijbregts (1978, pag. 199) afirmam que os dados da variável estudada tem que estar no mesmo suporte. Quando as amostras não estão no mesmo suporte, deve-se trabalhar com a variável acumulação. Duas situações são abordadas: (1) as amostras têm diferente comprimento e (2) os teores são analisados em diferentes faixas granulométricas. Quando a amostra tem diferente comprimento, acumulação é o produto do teor pelo comprimento. Quando as amostras tem diferente fração mássica, acumulação é o produto do teor pela recuperação mássica. A estimativa do teor é obtida pela divisão da estimativa da variável acumulação pela estimativa da variável

ponderadora (recuperação ou espessura). Essa abordagem é chamada de método indireto de estimativa.

As variáveis acumulação e espessura (ou recuperação mássica) precisam ser cokrigadas se elas são correlacionadas (Rossi e Deutsch, 2013). Dagbert (2001) recomenda utilizar o mesmo modelo variográfico e vizinhança de busca para estimar as variáveis acumulação e espessura para evitar o aparecimento de estimativas fora do intervalo dos dados. Quando o mesmo variograma é utilizado, os mesmos pesos de krigagem são utilizados para estimar as duas variáveis (acumulação e espessura). O mesmo efeito é obtido se as variáveis acumulação e espessura são cokrigadas utilizando um modelo de coregionalização intrínsica (Goovaerts, 1997), que modela os variogramas para acumulação, espessura e o variograma cruzado como proporcional ao mesmo modelo de variograma. Bertoli *et al.* (2003) utilizaram o modelo de coregionalização intrínsica para estimar as variáveis acumulação e espessura.

Marcotte e Boucher (2001), Roy *et al.* (2004) e Zuñiga e Emery (2010) compararam o método indireto de estimativa contra o método direto. No método direto, é feita a krigagem diretamente dos teores. Nesses estudos de caso, o suporte dos dados não foi considerado no sistema de krigagem no método direto.

1.3.2 Mudança de suporte no variograma

O uso de covariâncias médias no sistema de krigagem para incorporar dados de diferente suporte exige um modelo de variograma em suporte pontual. A inferência do variograma em suporte pontual não pode ser feita diretamente dos dados se esses estão em um suporte maior. Dentro desse contexto, é importante conhecer as relações entre variogramas de diferente suporte.

Journel e Huijbregts (1978, pag. 78) apresentaram a relação entre variogramas de diferente suporte de maneira genérica. Essa relação genérica foi chamada de *direct semivariogam upscaling* (DSU) por Babak *et al.* (2013). Journel e Huijbregts (1978) também mostraram uma abordagem mais simplificada, que foi chamada de *scaling laws* (SL) por Babak *et al.* (2013). SL assume que os dados volumétricos não se sobrepõem

e que a forma do variograma não muda. Dessa forma, SL é simplificada em encontrar os alcances e contribuições para o variograma de maior ou menor suporte.

Kupfersberger *et al.* (1998) utilizaram SL para obter um modelo linear de coregionalização (MLC) em suporte de ponto a partir de dados volumétricos provenientes de sísmica. Oz *et al.* (2002) utilizaram SL para obter um variograma em suporte de *log* (amostra obtida por perfilagem geofísica em um poço de petróleo) a partir de amostras de *core* (amostra de rocha obtida do poço de petróleo). Babak *et al.* (2013) comparou SL e DSU para obter variogramas em suporte de bloco a partir de dados em suporte de ponto. O método DSU produziu os melhores resultados. Nesses estudos, os dados de maior suporte tinham o mesmo tamanho. Babak *et al.* (2013) publicaram um *software* para fazer a mudança de suporte no variograma utilizando as duas metodologias (SL e DSU). No entanto, o *software* publicado lida apenas com blocos regulares e, portanto, não pode ser utilizado nessa tese. Na tese, as amostras de maior suporte possuem tamanhos variados.

Goovaerts (2008) detalhou a relação entre variogramas de diferente suporte de forma genérica, onde os dados de maior suporte possuem formas irregulares. Essa metodologia pode ser vista como uma forma genérica de DSU e será utilizada nessa tese.

A obtenção de um variograma em um dado suporte a partir de um variograma pontual é chamada de regularização (Journel e Huijbregts, 1978, pag. 77). A operação inversa (obter um variograma pontual a partir de um variograma em um dado suporte) é chamada de deconvolução (Jounel e Huijbregts, 1978, pag. 90).

Journel e Huijbregts (1978, pag. 91) apresentaram uma maneira genérica de fazer a deconvolução do variograma. Nesse método, o variograma em suporte pontual é ajustado iterativamente até que o variograma regularizado se ajuste ao variograma experimental definido em um determinado suporte. Goovaerts (2008) apresenta uma maneira automatizada de deconvoluir o variograma. Essa maneira está implementada no *software* comercial *SpaceStat* (*Biomedware*, EUA) e não será utilizada nessa tese. O algoritmo está detalhado em Goovaerts (2008) e envolve a modelagem automática de variogramas. Nessa tese, o método proposto por Journel e Huijbregts (1978, pag. 91)

será utilizado. Umas das contribuições esperadas da tese é um *software* auxiliar para fazer a deconvolução do variograma.

Um aspecto interessante é a obtenção de variogramas experimentais com amostras de diferente suporte. Journel (1986) afirma que o sistema de krigagem acomoda dados de diferente suporte, mas é necessário amostras com o mesmo suporte para calcular o variograma experimental. Por outro lado, Goovaerts (2008) calcula o variograma experimental para dados de diferente suporte.

1.3.3 Verificação da distribuição multivariada

Leuangthong *et al.* (2004) revisaram as verificações para simulações geoestatísicas. As simulações devem reproduzir o seguinte: (1) os dados nas suas localizações, (2) o histograma de referência, (3) as estatísticas de referência e (4) o modelo de covariância. No caso de simulação geoestatística multivariada, Leuangthong *et al.* (2004) afirmam que as simulações devem reproduzir a distribuição multivariada. Entretanto, Leuangthong *et al.* (2004) não mostram uma metodologia para verificar a reprodução da distribuição multivariada. Quando a distribuição multivariada deve ser reproduzida, a cosimulação geoestatística é feita.

Diversos estudos de caso de cosimulação geoestatística são encontrados na literatura (Leuangthong and Deutsch, 2003; Barnett *et al.*, 2014; Manchuk *et al.*, 2017; Horta and Soares, 2010; Mery *et al.*, 2017). Para verificar a relação entre as variáveis, os autores verificaram o seguinte: (1) reprodução dos gráficos de dispersão, (2) reprodução dos coeficientes de correlação e (3) reprodução dos variogramas cruzados. O problema dessas verificações é que elas consideram apenas a relação entre duas variáveis. Além disso, os coeficientes de correlação medem apenas relações lineares entre as variáveis.

A estatística de Kolmogorov-Smirnov (Massey, 1951) tem sido usada para quantificar a diferença entre duas funções de distribuição cumulativas univariadas (*cumulative distribution function* –cdf) e é chamada de estatística D. A estatística D corresponde ao máximo da diferença absoluta entre duas cdfs. Nessa tese, é proposto o 90º quantil da diferença absoluta entre cdfs multivariadas. O 90º quantil é mais

robusto do que o máximo na presença de valores extremos no conjunto de diferenças. A cdf multivariada das simulações geoestatísticas podem ser comparadas com a cdf multivariada dos dados.

1.3.4 Simulação geoestatística multivariada com restrições

As técnicas de simulação geoestatística permitem construir modelos numéricos de variáveis contínuas como propriedades petrofísicas e concentrações de metais. As técnicas de simulação geoestatística incluem a simulação sequencial direta (Soares, 2001), simulação sequencial dos indicadores (Gomez-Hernandez e Srivastava, 1990), simulação *p-field* (Froidevaux, 1993), simulação *annealing* (Deutsch, 1992) e as técnicas Gaussianas como simulação por bandas rotativas (Journel e Huijbregts, 1978) e simulação sequencial Gaussiana (Isaaks, 1990).

No caso de simulação de múltiplas variáveis correlacionadas, em geral as técnicas Gaussianas são utilizadas. As técnicas Gaussianas assumem que os dados seguem uma distribuição multi-Gaussiana. Dessa forma, as relações entre as variáveis são plenamente definidas pela matriz de covariância entre as variáveis. Como os dados em geral não seguem uma distribuição Gaussiana, várias transformações foram desenvolvidas para tornar os dados mais Gaussianos.

A transformação *normal score* (Deutsch e Journel, 1998) transforma os dados de forma que os dados transformados tenham um histograma Gaussiano. A transformação *normal score* torna os dados univariados Gaussianos, mas não os torna multi-Gaussianos. Uma prática comum é transformar os dados utilizando a transformação *normal score* e assumir que os dados transformados tenham uma distribuição multi-Gaussiana. Dessa forma, as relações entre as variáveis são obtidas através dos coeficientes de correlação provenientes do modelo linear de coregionalização (MLC) (Journel e Huijbregts, 1978) ou do modelo de Markov (Almeida e Journel, 1994).

O modelo de Markov causa problemas na reprodução do variograma (Babak e Deutsch, 2009). O MLC não é prático à medida que o número de variáveis aumenta devido à necessidade de modelar os variogramas diretos e cruzados. Em vista disso, técnicas de descorrelação como *Principal Component Analysis* (PCA) (Goovaerts,

1993) e *Minimum/Maximum Autocorrelation Factors* (MAF) (Switzer e Green, 1984; Desbarats e Dimitrakopoulos, 2000) foram desenvolvidas para descorrelacionar as variáveis. As variáveis descorrelacionadas podem ser então simuladas de maneira independente, sem a necessidade de modelar os variogramas cruzados.

A técnica de PCA (Hotelling, 1933) transforma um conjunto de N variáveis em N componentes que são descorrelacionados para h = 0. Os componentes são combinações lineares obtidos a partir da decomposição espectral da matriz de covariância. MAF é uma extensão de PCA e faz uma decomposição espectral duas vezes: (1) para a matriz de covariância de h = 0 e (2) para a matriz de covariância para h > 0. Se as relações entre as variáveis originais são plenamente caracterizadas por um MLC de duas estruturas, a transformação MAF descorrelaciona as variáveis para todas as distâncias (Desbarats e Dimitrakopoulos, 2000). Se as relações entre as variáveis originais são mais complexas, a transformação MAF remove boa parte da correlação espacial a pequenas distâncias (Desbarats e Dimitrakopoulos, 2000).

O problema das transformações PCA e MAF é que elas não tornam as variáveis independentes na presença de relações complexas. Em vista disso, foram desenvolvidas transformações multivariadas não lineares como S*tepwise Conditional Transformation* (SCT) (Leuangthong e Deutsch, 2003) e *Projection Pursuit Multivariate Transform* (PPMT) (Barnett e Deutsch, 2014). As transformações SCT e PPMT transformam os dados em multi-Gaussianos e independentes.

A transformação SCT (Leuangthong e Deutsch, 2003) começa com uma transformação *normal score* da primeira variável. A segunda variável é então particionada em uma série de classes de acordo com as classes de probabilidade da primeira variável. É então aplicada a transformação *normal score* para cada classe da segunda variável. No caso de n variáveis, a enésima variável é particionada de acordo com as n-1 primeiras variáveis. As variáveis transformadas por SCT são multi-Gaussianas e independentes, com uma matriz de correlação igual à matriz identidade. A limitação da SCT é que ela exige uma grande quantidade de dados. Leuangthong e Deutsch (2003) afirmam que é necessário de 10^n a 20^n dados, onde n corresponde ao número de variáveis, para obter distribuições condicionais bem definidas na SCT. Dentro desse contexto, a SCT muitas vezes não é viável para problemas com mais de

duas ou três variáveis. O fato que a SCT necessita de muitos dados motivou o desenvolvimento da PPMT (Barnett *et al.*, 2014; Barnett *et al.*, 2016).

De mesma forma que a SCT, a PPMT transforma os dados em multi-Gaussianos e independentes. Em uma distribuição multi-Gaussiana, qualquer projeção dos dados ao longo de um vetor tem uma distribuição Gaussiana. A transformação PPMT (Ryan *et al.*, 2016) começa com duas etapas de pré-processamento: (1) *normal score* e (2) *sphering*. A transformação *normal score* torna os dados Gaussianos univariados enquanto que a operação de *sphering* torna os dados descorrelacionados. Após isso, a PPMT busca de maneira iterativa as projeções dos dados que são mais não Gaussianas e aplica a transformação *normal score* nessas projeções. O usuário define o número de projeções e o nível de multi-Gaussianeidade como critério de parada para a busca das projeções.

A PPMT provou reproduzir as relações entre as variáveis (Barnett *et al.*, 2016). Entretanto, a transformação PPMT não considera explicitamente restrições de soma na transformação. Nessa tese, foi avaliada uma versão modificada da PPMT para considerar restrições de soma. Nesse caso, a primeira transformação *normal score* é aplicada à projeção ortogonal à restrição de soma. Essa projeção é uma combinação linear dos dados e corresponde à soma das variáveis que tem restrição de soma. Como a transformação *normal score* inversa evita extrapolação, espera-se que a soma de valores simulados esteja entre o mínimo e máximo da soma das variáveis.

Outro método para lidar com variáveis com restrição de soma é baseado no uso de razões e razões logarítmicas dos dados originais. Essa abordagem tem sido usada com sucesso com dados composicionais. Dados composicionais representam um conjunto de variáveis cuja soma é igual a uma constante. Quando a soma das variáveis não é igual a uma constante, uma variável *filler* pode ser adicionada ao banco de dados. A variável *filler* corresponde à constante menos a soma das variáveis remanescentes. As transformações mais comuns para dados composicionais são as razões logarítmicas aditivas (*additive log-ratio* – alr, Aitchison, 1986), razões logarítmicas centradas (*centred log-ratio* – clr, Aitchison 1986) e razões logarítmicas isométricas *(isometric log-ratio* - ilr, Egozcue *et al.*, 2003). Manchuck *et al.* (2016) usaram razões logarítmicas isométricas combinadas com PPMT para considerar a

restrição de soma do banco de dados. Mery *et al.* (2017) usaram razões que não utilizam logaritmos para considerar a restrição de soma dos dados. Mery *et al.* (2017) buscaram maximizar a correlação entre a variável original e a variável transformada. Nessa tese, duas razões foram avaliadas. A primeira razão é similar a alr. A diferença é que o logarítmico não foi calculado. A segunda razão analisada é a razão usada por Mery *et al.* (2017).

Para lidar com a restrição de fração, Emery (2012) e Hosseini e Asghari (2015) usaram a transformação S*tepwise Conditional Transformation* (SCT) (Leuangthong e Deutsch, 2003) para transformar as variáveis em variáveis Gaussianas antes da simulação geoestatística. As simulações honraram a restrição de fração. O problema é que a transformação SCT necessita de uma grande quantidade de dados quando várias variáveis são consideradas. Os estudos de caso mostrados por Emery (2012) e Hosseini e Asghari (2015) consideraram apenas duas variáveis.

1.4 Organização da tese

O capítulo 2 revisa a teoria sobre estimativa com dados de diferente suporte. O uso de covariâncias médias no sistema de krigagem e o método indireto são descritos. A krigagem com amostras de diferente de suporte exige um variograma em suporte de ponto, que é obtido através da deconvolução do variograma. O capítulo 2 detalha o processo de deconvolução do variograma.

O capítulo 3 apresenta um estudo de caso de krigagem com amostras de diferente comprimento em um depósito de bauxita. A krigagem com amostras de diferente suporte é comparada com a krigagem utilizando covariâncias ponto-a-ponto entre as amostras, que despreza a diferença de suporte entre as amostras. A precisão e acurácia das estimativas são avaliadas.

O capítulo 4 compara o método direto e indireto para estimativa utilizando amostras de diferente suporte. O impacto do variograma nas estimativas pelos dois métodos é investigado. Um estudo de caso em um depósito de bauxita é apresentado. A precisão e acurácia das estimativas obtidas com os dois métodos são avaliadas.

O capítulo 5 descreve a simulação sequencial Gaussiana. Além disso, o capítulo 5 detalha a transformação PPMT. Uma versão modificada da transformação PPMT para lidar com restrição de soma é apresentada. O capítulo 5 detalha também as razões utilizadas para lidar com as restrições de soma e fração.

O capítulo 6 mostra um comparativo de *workflows* para simulação multivariada com restrições de soma e fração. Os *workflows* utilizaram as razões e transformações apresentadas no capítulo 5. O comparativo foi feito através de simulação de Monte Carlo. As vantagens e desvantagens das técnicas são analisadas. Os resultados do estudo comparativo foram usados para escolher o *workflow* para a simulação geoestatística multivariada em um depósito de bauxita. Por último, a simulação geoestatística multivariada em um depósito de bauxita é apresentada com as devidas validações.

O capítulo 7 detalha a estatística $D90$, que é usada para medir a diferença entre distribuições multivariadas. O efeito do aumento do número de variáveis na função de distribuição acumulada multivariada é mostrado. A estatística $D90$ é analisada com dois bancos de dados sintéticos. O primeiro banco de dados consiste em duas variáveis com uma forte relação linear. O segundo banco de dados consiste em duas variáveis com uma relação não linear. Os dados foram corrompidos com adição de erro de precisão e viés. A estatística $D90$ é usada para identificar a distância entre os dados originais e corrompidos.

O capítulo 8 compara o método direto e indireto para a simulação geoestatística multivariada. A estatística $D90$ apresentada no capítulo 7 é usada para comparar os dois métodos.

O capítulo 9 resume as principais contribuições da tese. O capítulo 9 também lista sugestões para trabalhos futuros.

2 Teoria sobre estimativa com dados de diferente suporte e deconvolução do variograma

A seção 2.1 aborda o tema de estimativa com dados de diferente suporte. Nessa seção, primeiro são revisados os conceitos de estimativa e variância do erro de estimativa (seção 2.1.1). Posteriormente, são explicados dois métodos de estimativa que utilizam dados de diferente suporte: (i) método indireto (seção 2.1.2) e (ii) krigagem com amostras de diferente suporte (seção 2.1.3).

Como a krigagem utilizando amostras de diferente suporte exige um modelo de variograma em suporte de ponto, é necessária fazer a deconvolução do variograma quando não há uma quantidade suficiente de dados em suporte de ponto para inferir o modelo de variograma em suporte pontual. A deconvolução do variograma está presente na seção 2.2. Primeiro, a seção 2.2 mostra o cálculo dos variogramas experimentais para amostras de diferente suporte (seção 2.2.1). Segundo, a regularização do variograma para blocos de forma regular é apresentada na seção 2.2.2. A seção 2.2.3 generaliza a regularização do variograma para blocos de forma irregular. Posteriormente, a seção 2.2.4 apresenta a deconvolução do variograma.

2.1 Estimativa com dados de diferente suporte

2.1.1 Estimativa e variância do erro de estimativa

Considere o problema de estimar o valor de um atributo contínuo z na localização $\mathbf{u}$, ou seja, $z(\mathbf{u})$. Os dados consistem em um conjunto de n valores discretos definidos nas localizações $\mathbf{u}_\alpha \{z(\mathbf{u}_\alpha), \alpha = 1,...,n\}$. A estimativa de $z(\mathbf{u})$ é usualmente definida como uma combinação linear dos pesos λ_α e dos valores $z(\mathbf{u}_\alpha)$:

$$\begin{cases} z^*(\mathbf{u}) = \sum_{\alpha=1}^{n} \lambda_\alpha \cdot z(\mathbf{u}_\alpha) \\ \sum_{\alpha=1}^{n} \lambda_\alpha = 1 \end{cases} \quad (1)$$

Quando interpolação pelo inverso da distância é utilizada, os pesos usados na equação 1 são inversamente proporcionais à distância até o ponto $\mathbf{u}$. No caso de krigagem ordinária, os pesos são a solução do sistema de krigagem ordinária. A equação 2 define a variância do erro de estimativa (Pyrcz e Deutsch, 2014; Isaaks e Srivastava, 1989):

$$\sigma_E^2(\mathbf{u}) = C(\mathbf{0}) + \sum_{\alpha=1}^{n}\sum_{\beta=1}^{n} \lambda_\alpha \lambda_\beta C(\mathbf{u}_\alpha - \mathbf{u}_\beta) - 2 \cdot \sum_{\alpha=1}^{n} \lambda_\alpha C(\mathbf{u} - \mathbf{u}_\alpha) \quad (2)$$

onde $C(\mathbf{u}_\alpha - \mathbf{u}_\beta)$ é a covariância entre as localizações $\mathbf{u}_\alpha$ e $\mathbf{u}_\beta \cdot C(\mathbf{u} - \mathbf{u}_\alpha)$ é a covariância entre a localização $\mathbf{u}$ e $\mathbf{u}_\alpha$. Os parâmetros de entrada para a equação 2 são os pesos utilizados na estimativa e um modelo de covariância válido.

2.1.2 Método indireto

Acumulação é uma variável auxiliar obtida a partir do teor da amostra multiplicado pelo seu respectivo suporte. Em depósitos estratiformes, em que as amostras possuem comprimento diferente, a acumulação é obtida pelo produto do teor pela espessura. A estimativa de teor utilizando acumulação é feita através dos seguintes passos:

1. Estimativa da variável acumulação (produto do teor pela espessura);
2. Estimativa da variável espessura;
3. Divisão da estimativa de acumulação pela estimativa da espessura.

As estimativas das variáveis espessura e acumulação são feitas geralmente por krigagem ordinária.

Considere o problema de estimar o teor z na localização $\mathbf{u}$, ou seja, $z(\mathbf{u})$. Os dados consistem em duas amostras localizadas em $\mathbf{u}_\alpha$ e $\mathbf{u}_\beta$. Cada amostra possui uma espessura t. A equação 3 mostra a estimativa de $z(\mathbf{u})$ pelo método indireto para esse exemplo:

$$z^*(\mathbf{u}) = \frac{\lambda_\alpha \cdot t(\mathbf{u}_\alpha) \cdot z(\mathbf{u}_\alpha) + \lambda_\beta \cdot t(\mathbf{u}_\beta) \cdot z(\mathbf{u}_\beta)}{\lambda_\alpha \cdot t(\mathbf{u}_\alpha) + \lambda_\beta \cdot t(\mathbf{u}_\beta)} \tag{3}$$

onde λ_α e λ_β são os pesos de krigagem associados aos dados nas localizações $\mathbf{u}_\alpha$ e $\mathbf{u}_\beta$, respectivamente.

Quando os mesmos pesos são utilizados para estimar acumulação e espessura, o mesmo resultado é obtido com o seguinte procedimento: (1) multiplicar os pesos de krigagem pela espessura, (2) estandardizar os pesos para que eles somem 1 e (3) usar esses pesos estandardizados junto com a variável z no estimador definido pela equação 1. Quando os pesos são multiplicados pela espessura e estandardizados, o resultado são os seguintes pesos estandardizados (equação 4):

$$\lambda_\alpha^{STD} = \frac{\lambda_\alpha \cdot t(\mathbf{u}_\alpha)}{\lambda_\alpha \cdot t(\mathbf{u}_\alpha) + \lambda_\beta \cdot t(\mathbf{u}_\beta)}; \tag{4}$$

$$\lambda_\beta^{STD} = \frac{\lambda_\beta \cdot t(\mathbf{u}_\beta)}{\lambda_\alpha \cdot t(\mathbf{u}_\alpha) + \lambda_\beta \cdot t(\mathbf{u}_\beta)}$$

onde esses pesos estandardizados podem ser aplicados diretamente para estimar $z(\mathbf{u})$:

$$z^*(\mathbf{u}) = \lambda_\alpha^{STD} \cdot z(\mathbf{u}_\alpha) + \lambda_\beta^{STD} \cdot z(\mathbf{u}_\beta) \tag{5}$$

A equação 5 é equivalente à equação 3 contanto que os mesmos pesos sejam utilizados para estimar as variáveis acumulação e espessura. A comparação feita nessa tese considera esses pesos estandardizados, porque eles são aplicados diretamente na variável de interesse. Esses pesos mostram melhor a influência do método indireto nas estimativas dos teores. Esses pesos estandardizados podem ser utilizados na equação 2 para calcular a variância do erro para o método indireto. Journel e Huijbregts (1978, pag. 404-424) mostram o cálculo para a variância do erro no caso de uma divisão de forma geral. Entretanto, a abordagem considerada nessa tese é mais intuitiva para mostrar a influência do método indireto sobre o teor quando os mesmos pesos são utilizados para estimar as variáveis acumulação e espessura.

2.1.3 Krigagem com amostras de diferente suporte

Considere o problema de estimar o valor médio de um atributo contínuo z sobre um suporte V centrado em $\mathbf{u}$, ou seja, $z_V(\mathbf{u})$. Os dados consistem de um conjunto de valores definidos nos suportes v_α centrado nos locais $\mathbf{u}_\alpha$ $\{z_{v_\alpha}(\mathbf{u}_\alpha); \alpha = 1,...,n\}$. Suporte se refere ao tamanho no qual o atributo z foi medido. Por exemplo, o suporte V pode se referir a uma unidade seletiva de mina (*selective mining unity* – SMU) e o suporte v_α pode se referir a uma linha (por exemplo, testemunhos de sondagem medidos em diferentes comprimentos), uma área, ou um volume. Os dados no suporte v_α representam a média de teores pontuais sobre o volume v_α . Na literatura, valores médios sobre um suporte v são chamados de valores de bloco (Journel e Huijbregts 1978; Deutsch e Journel 1998; Goovaerts 1997). A equação 6 define o estimador de krigagem ordinária para o valor $z_V(\mathbf{u})$:

$$z_V^*(\mathbf{u}) = \sum_{\alpha=1}^{n} \lambda_\alpha z_{v_\alpha}(\mathbf{u}_\alpha) \tag{6}$$

onde λ_α é o peso de krigagem odinária associada ao dado $z_{v_\alpha}(\mathbf{u}_\alpha)$. Os pesos de krigagem são a solução do sistema de krigagem ordinária. A equação 7 define o sistema de krigagem ordinária considerando o suporte dos dados:

$$\begin{cases} \sum_{\beta=1}^{n} \lambda_\beta \overline{C}(v_\alpha, v_\beta) + \mu = \overline{C}(v_\alpha, V), \\ \qquad\qquad\qquad\qquad\qquad\qquad \alpha = 1,\ldots,n \\ \sum_{\beta=1}^{n} \lambda_\beta = 1, \end{cases} \tag{7}$$

onde μ é o multiplicador de Lagrange, $\overline{C}(v_\alpha, v_\beta)$ é a covariância entre o dado de bloco v_α e o dado de bloco v_β. $\overline{C}(v_\alpha, v_\beta)$ é calculada como a média de covariâncias ponto-a-ponto entre todos os pontos que discretizam o bloco v_α com todos os pontos que discretizam o bloco v_β:

$$\overline{C}(v_\alpha, v_\beta) = \frac{1}{N_i N_j} \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} C(\mathbf{u}_i^{'}, \mathbf{u}_j^{'}) \tag{8}$$

onde N_i é o número de dados discretizantes $\mathbf{u}_i^{'}$ do bloco v_α e N_j é o número de pontos discretizantes $\mathbf{u}_j^{'}$ do bloco v_β. A figura 3 mostra um esquema para o cálculo da covariância bloco-a-bloco. Cada linha na figura 3 é uma covariância ponto-a-ponto.

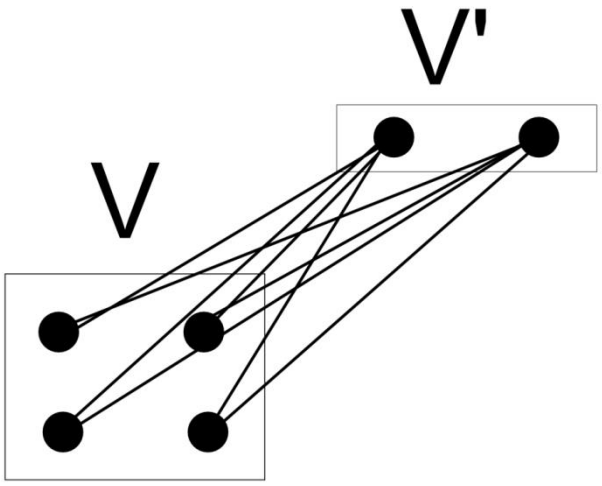

Figura 3: Esquema do cálculo de covariância bloco-a-bloco. Cada linha representa uma covariância ponto-a-ponto.

2.2 Regularização e deconvolução do variograma

2.2.1 Variograma experimental para amostras de diferente suporte

A equação 10 define o variograma experimental para amostras em suporte de ponto:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} \left[z(\mathbf{u}_\alpha) - z(\mathbf{u}_\alpha + \mathbf{h}) \right]^2 \tag{10}$$

onde $N(\mathbf{h})$ é o número de pares de amostras encontradas separadas pelo vetor $\mathbf{h}$. A equação 11 define o variograma experimental para amostras definidas em diferente suporte (Goovaerts, 2008):

$$\hat{\gamma}_v(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} \left[z_{v_\alpha}(\mathbf{u}_\alpha) - z_{v_{\alpha+\mathbf{h}}}(\mathbf{u}_\alpha + \mathbf{h}) \right]^2 \tag{11}$$

O uso de amostras de diferente suporte para o cálculo do variograma experimental pode ser questionado (Journel, 1986). No entanto, as amostras de diferente suporte podem ajudar a inferir a continuidade espacial quando não há um subconjunto de amostras de mesmo suporte para o cálculo do variograma experimental.

2.2.2 Regularização do variograma para blocos de forma regular

Sob a hipótese de estacionariedade, o variograma em suporte de ponto e o variograma regularizado são relacionados através da equação 12 (Journel e Huijbregts, pag. 77):

$$\gamma_v(\mathbf{h}) = \bar{\gamma}(v, v_{\mathbf{h}}) - \bar{\gamma}(v, v) \tag{12}$$

onde $\gamma_v(\mathbf{h})$ é o valor do variograma no suporte de bloco v para uma distância $\mathbf{h}$, $\bar{\gamma}(v, v_{\mathbf{h}})$ é o variograma médio entre o bloco v e o bloco v separado de uma distância $\mathbf{h}$ ($v_{\mathbf{h}}$). O termo $\bar{\gamma}(v, v)$ é o variograma médio dentro do bloco v e é conhecido na literatura como *gammabar*. A equação 13 mostra o cálculo do *gammabar* $\bar{\gamma}(v, v)$:

$$\bar{\gamma}(v, v) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \gamma(\mathbf{u}'_i, \mathbf{u}'_j) \tag{13}$$

onde N é o número de pontos discretizantes do bloco v. A equação 14 mostra o cálculo de $\bar{\gamma}(v, v_{\mathbf{h}})$:

$$\bar{\gamma}(v, v_{\mathbf{h}}) = \frac{1}{N_i N_j} \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} \gamma(\mathbf{u}'_i, \mathbf{u}'_j) \tag{14}$$

onde N_i é o número de pontos discretizantes do bloco v e N_j é o número de pontos discretizantes do bloco $v_{\mathbf{h}}$. As relações apresentadas assumem que todos os blocos possuem a mesma forma geométrica. Goovaerts (2008) apresentou uma forma genérica da regularização do variograma para blocos de diferentes formas, reproduzida na seção 2.2.3.

2.2.3 Regularização do variograma para blocos de forma irregular

Quando os blocos são regulares, o termo $\bar{\gamma}(v,v)$ é constante. Por outro lado, quando os blocos são irregulares, o *gammabar* depende da distância **h**. A equação 15 mostra a regularização do variograma de uma forma mais genérica (Goovaerts, 2008):

$$\gamma_v(\mathbf{h}) = \bar{\gamma}_{\mathbf{h}}(v, v_{\mathbf{h}}) - \bar{\gamma}_{\mathbf{h}}(v,v) \tag{15}$$

O valor de $\bar{\gamma}_{\mathbf{h}}(v,v)$ é calculado como a média aritmética de *gammabars* de blocos separados de uma distância **h** (equação 16):

$$\bar{\gamma}_{\mathbf{h}}(v,v) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} \left[\bar{\gamma}(v_\alpha, v_\alpha) + \bar{\gamma}(v_{\alpha+\mathbf{h}}, v_{\alpha+\mathbf{h}}) \right] \tag{16}$$

onde $N(\mathbf{h})$ é o número de pares de blocos para uma distância **h**. $\bar{\gamma}(v_\alpha, v_\alpha)$ é calculado através da Equação 13. Blocos menores formam pares para distâncias menores, enquanto que blocos maiores formam pares para distâncias maiores.

Nessa tese, a distância **h** é calculada como a distância média entre os centroides dos blocos. Goovaerts (2008) propõe utilizar uma distância média ponderada entre os pontos discretizantes. Entretanto, essa distância ponderada aumenta o cálculo computacional e é muito semelhante com a distância média entre os centroides dos

blocos. Goovaerts (2008) comparou a distância proposta com a distância média entre os centroides dos blocos e o coeficiente de correlação foi de 0.99.

A equação 17 define o cálculo de $\bar{\gamma}_{\mathbf{h}}(v, v_{\mathbf{h}})$:

$$\bar{\gamma}_{\mathbf{h}}(v, v_{\mathbf{h}}) = \frac{1}{N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} \bar{\gamma}(v_{\alpha}, v_{\alpha+\mathbf{h}}) \tag{17}$$

$\bar{\gamma}_{\mathbf{h}}(v, v_{\mathbf{h}})$ é interpretado como a média aritmética dos variogramas médios entre os blocos v separados de uma distância $\mathbf{h}$. O termo $\bar{\gamma}(v_{\alpha}, v_{\alpha+\mathbf{h}})$ é obtido através da equação 14.

2.2.4 Deconvolução do variograma

Journel e Huijbregts (1978, pag. 90) propõem o seguinte método para a obtenção do variograma em suporte de ponto a partir de amostras definidas em um suporte v:

1. Definir um modelo de variograma em suporte de ponto $\gamma(\mathbf{h})$ a partir dos variogramas experimentais em suporte de bloco $\hat{\gamma}_v(\mathbf{h})$.

2. Computar o modelo de variograma teórico $\gamma_v(\mathbf{h})$ usando a equação 15 e comparar com os valores experimentais $\hat{\gamma}_v(\mathbf{h})$.

3. Ajustar iterativamente os parâmetros do modelo de variograma em suporte de ponto $\gamma(\mathbf{h})$ para que o modelo teórico $\gamma_v(\mathbf{h})$ se ajuste aos pontos experimentais.

3 Estudo de caso: krigagem com amostras de diferente comprimento

O capítulo 3 mostra um estudo de caso da krigagem com amostras de diferente suporte com um banco de dados de um depósito mineral. As amostras de diferente suporte consistem de furos de sondagem amostrados em diferente comprimentos. O capítulo 3 também apresenta uma comparação da krigagem com amostras de diferente suporte com a krigagem utilizando covariâncias ponto-a-ponto entre as amostras, que ignora a diferença de suporte entre as amostras.

3.1 Apresentação do banco de dados

O banco de dados é proveniente de um depósito de bauxita e contém um total de 686 furos de sondagem localizados em um espaçamento quase regular de 200 x 200 m nas direções Leste e Norte. As coordenadas Z foram transformadas para coordenadas estratigráficas. A variável de interesse é a porcentagem da fração granulométrica retida na peneira de tamanho 14# (REC14). REC14 é uma variável aditiva, similar à teor.

Inicialmente, o banco de dados foi amostrado no intervalo de 0.50 metros. Para demonstrar a técnica, que considera amostras de diferente comprimento, 343 dos 686 furos foram selecionados aleatoriamente. Nesses furos de sondagem selecionados, REC14 foi regularizada pela espessura da camada. O comprimento da composta corresponde à espessura da camada de minério. O teor da composta representa o teor médio ao longo da espessura de minério. O banco de dados resultante contém 343 furos de sondagem com uma amostra cujo comprimento corresponde à espessura total de minério (pontos pretos na figura 4) e 343 furos de sondagem com amostras cujos comprimentos são aproximadamente 0.50 m (pontos brancos na figura 4). O banco de dados imita um banco de dados que contém amostras obtidas de diferentes campanhas de sondagem com comprimentos de amostragem distintos.

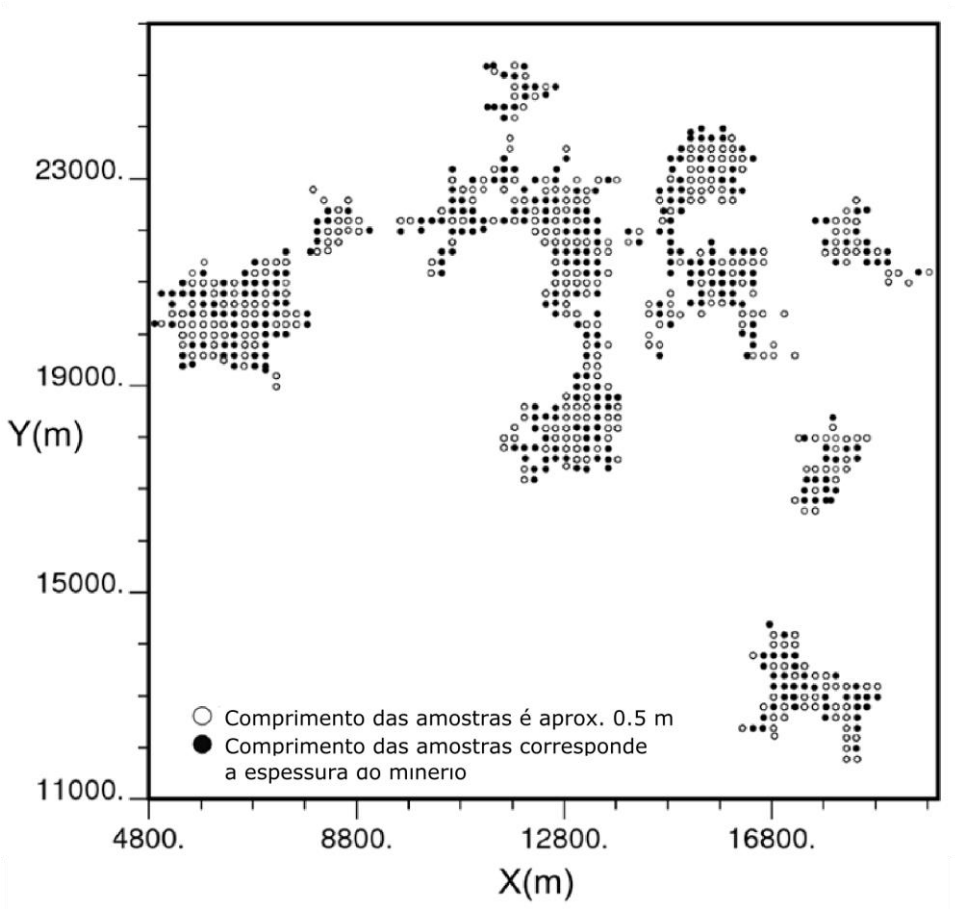

Figura 4: Mapa de localização das amostras.

A figura 5a mostra o histograma de REC14 ponderada pelo comprimento das amostras e o sumário estatístico. A distribuição é relativamente simétrica entorno da média com um baixo coeficiente de variação. Como os furos de sondagem estão regularmente espaçados, a estatística é representativa da área de estudo. A figura 5b mostra o diagrama quantil-quantil (traduzido pelo autor do termo original *QQ plot*) entre os dados e os dados desagrupados (obtidos com uma estimativa pelo método do polígono de influência). Os pontos na figura 5b estão próximos à reta y = x, mostrando que as duas distribuições são similares, como esperado. A figura 5c mostra o histograma do comprimento das amostras. O comprimento das amostras varia de 0.25 m até 7.88 metros. Como aproximadamente 80% das amostras tem comprimento menor ou igual a 0.50 m, o geomodelador pode se sentir tentado a usar apenas essas amostras para a estimativa. Entretanto, usar apenas essas amostras resulta em uma perda de informação excessiva.

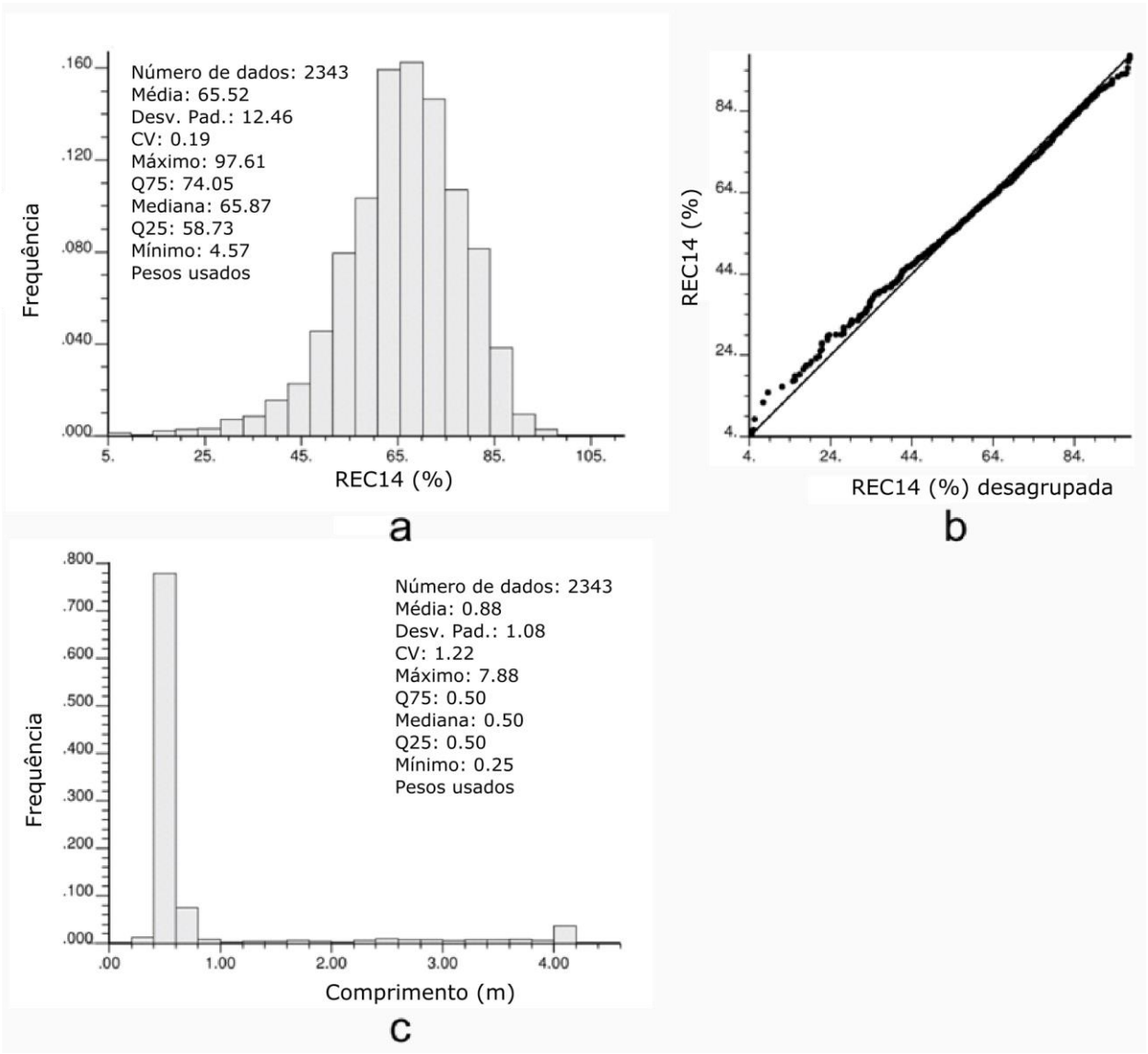

Figura 5: Histograma de REC14 (a) diagrama quantil-quantil entre os dados originais e desagrupados (b) e histograma do comprimento das amostras (c).

3.2 Análise e modelagem do variograma

A equação 18 descreve o modelo de variograma de REC14:

$$\gamma(\mathbf{h}) = 0.15 + 0.50 \cdot Sph(\frac{NS}{250m}, \frac{EW}{250m}, \frac{vert}{4.10m}) + 0.35 \cdot Sph(\frac{NS}{4500m}, \frac{EW}{4500m}, \frac{vert}{4.20m}) \quad (18)$$

Apenas as amostras com comprimento entre 0.25 e 0.50 m foram usadas para calcular o variograma experimental. Como a krigagem com amostras de diferente comprimento necessita de um modelo de variograma em suporte pontual, amostras longas (com comprimento maior do que 0.5 m) não foram usadas para calcular o variograma experimental.

3.3 Estimativa

A estimativa foi feita através de krigagem ordinária considerando a diferença de suporte entre os dados. As amostras de sondagem foram discretizadas ao longo da direção do furo. O espaçamento de discretização corresponde ao comprimento dos dados de menor comprimento utilizados para calcular o variograma experimental. Os pontos discretizantes das amostras foram utilizados para calcular as covariâncias bloco-a-bloco entre as amostras. A estimativa foi feita em um modelo de blocos com tamanho de bloco de 50 x 50 x 0.50 m em X, Y e Z. A discretização do bloco foi de 5 x 5 x 1. As estimativas foram limitadas para os blocos dentro do modelo geológico interpretado.

3.4 Validação do modelo

O modelo de teores foi verificado com as seguintes técnicas: (1) inspeção visual e (2) análise de deriva.

Inspeção visual consiste em comparar visualmente o modelo de teores com as amostras na mesma escala de cores. O modelo de teores precisa estar consistente com os dados.

Análise de deriva consiste em primeiro definir uma série de faixas ou fatias ao longo de X, Y e Z. Então, o teor médio do modelo de blocos e o teor médio dos dados desagrupados dentro de cada fatia são comparados. Os dados foram desagrupados com o método do polígono de influência.

3.5 Comparação com krigagem usando covariâncias ponto-a-ponto entre os dados

A krigagem com amostras de diferente suporte foi comparada contra krigagem usando covariâncias ponto-a-ponto, que ignora a diferença de suporte entre os dados. Nesse caso, todas as amostras são consideradas no mesmo suporte (pontual). As covariâncias entre os dados no sistema de krigagem são calculadas como covariâncias ponto-a-ponto. A comparação foi feita utilizando validação cruzada.

A validação cruzada consiste em primeiro remover uma amostra em uma localização particular. Segundo, o valor é estimado naquela localização utilizando as amostras restantes. Os mesmos parâmetros de estimativa utilizados na estimativa do modelo de blocos foram usados na validação cruzada. O erro de estimativa (diferença entre o valor estimado e real), o erro absoluto médio e o erro quadrático médio foram calculados. A estatística básica do erro de estimativa foi obtida. O erro médio mede a acurácia das estimativas. O erro médio absoluto, o erro médio quadrático e o desvio padrão do erro medem a precisão das estimativas.

3.6 Resultados

3.6.1 Validação do modelo

A figura 6 mostra uma vista em planta do modelo de blocos junto com as amostras. As áreas de alto teor do modelo de blocos estão próximas às amostras de alto teor, como esperado.

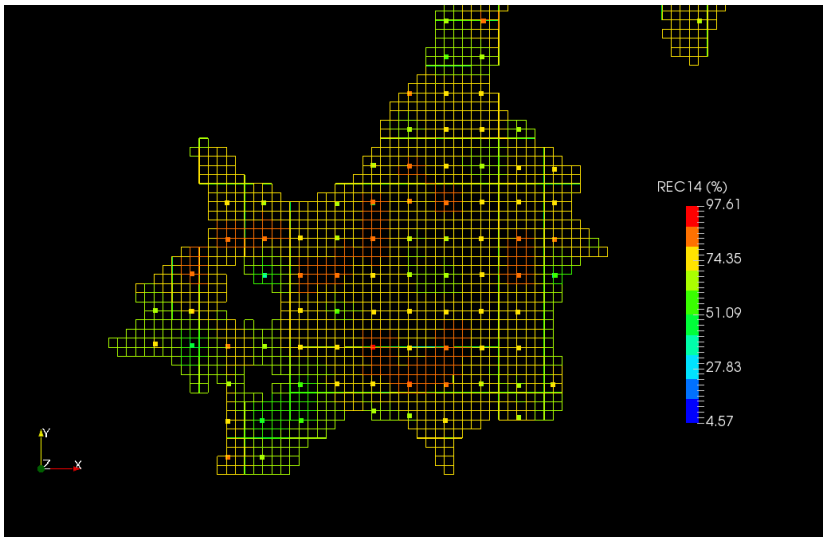

Figura 6: Modelo de blocos e amostras. As linhas representam o modelo de blocos enquanto que os pontos representam as amostras.

A análise de deriva mostra que o modelo de blocos reproduziu a tendência dos dados ao longo das direções X, Y e Z (veja as figuras 7a-c). As médias locais do modelo de blocos são similares às médias locais desagrupadas para as três direções. Além disso, as estimativas não subestimam ou superestimam sistematicamente a média local (veja as figuras 7a-c, a média local do modelo de blocos não está sempre acima ou abaixo da média local desagrupada).

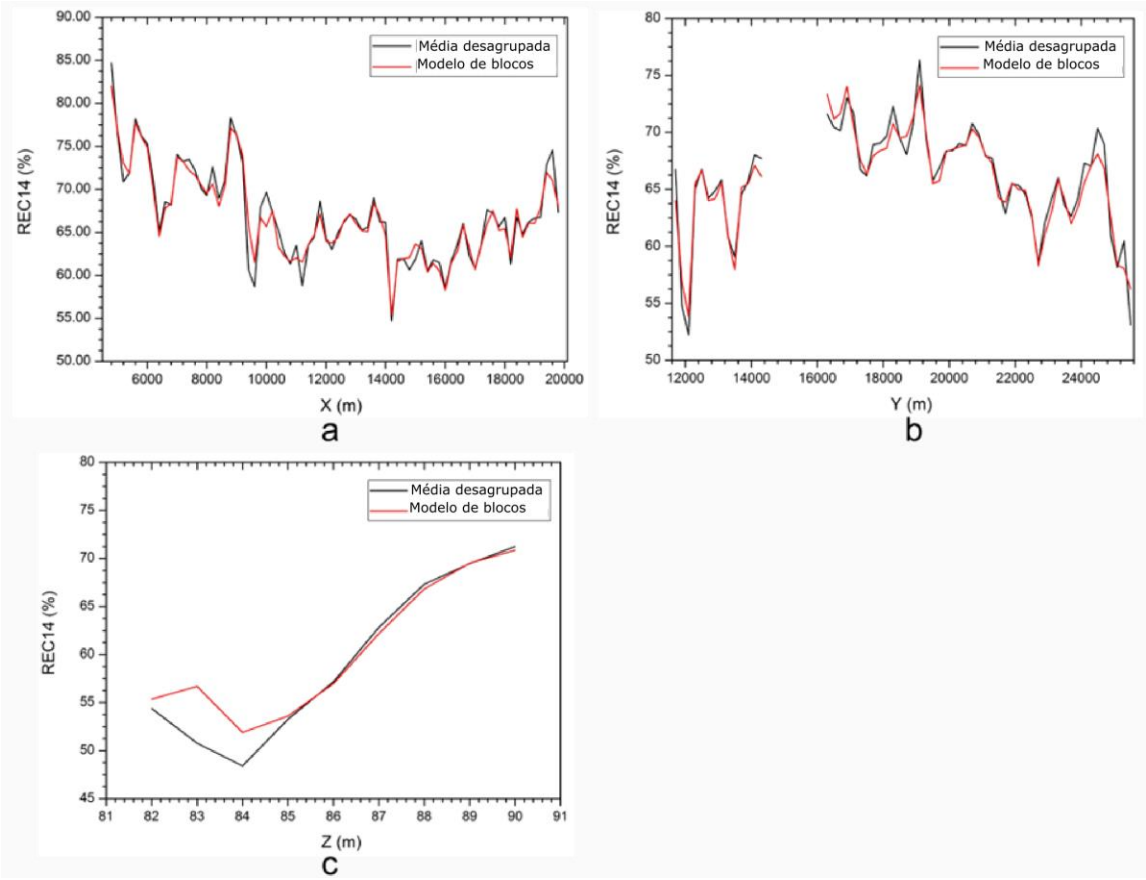

Figura 7: Análise de deriva nas direções X (a), Y (b) e Z (c).

3.6.2 Comparação com krigagem utilizando covariâncias ponto-a-ponto entre as amostras

A tabela 1 mostra a estatística básica do erro de estimativa para krigagem com amostras de diferente suporte e krigagem utilizando covariâncias ponto-a-ponto. A krigagem com amostras de diferente suporte resultou em estimativas mais acuradas, pois o erro médio é mais próximo de zero. Além disso, a krigagem com amostras de diferente suporte produziu estimativas mais precisas. Para essa metodologia, o desvio padrão do erro, o erro absoluto médio e o erro quadrático médio são menores. Especificamente, o erro quadrático médio obtido com a krigagem com amostras de diferente suporte é aproximadamente 5 % menor do que o erro quadrático médio obtido com a krigagem utilizando covariâncias ponto-a-ponto.

Tabela 1: Comparativo entre a krigagem com amostras de diferente suporte e a krigagem utilizando covariâncias ponto-a-ponto

	Krigagem com amostras de diferente suporte (covariâncias médias)	Krigagem utilizando covariâncias ponto-a-ponto
Número	2342	2342
Erro médio (%)	0.14	0.46
Desv. pad. do erro (%)	9.51	9.75
Erro absoluto médio (%)	7.09	7.38
Erro quadrático médio ($\%^2$)	90.37	95.26

Os resultados destacam a capacidade da krigagem de lidar com amostras de diferente suporte. Geomodeladores muitas vezes não percebem que o sistema de krigagem pode incorporar amostras de diferente suporte (Journel, 1986). No exemplo, a krigagem foi utilizada para estimar teores em um depósito mineral utilizando amostras de diferente comprimento.

Quando comparada com a krigagem utilizando covariâncias ponto-a-ponto, a krigagem com amostras de diferente suporte resultou em estimativas mais precisas e acuradas. No caso de estimativa com amostras de diferente suporte, a diferença de suporte entre as amostras precisa ser considerada no sistema de krigagem através do uso de covariâncias médias.

4 Comparativo entre método indireto e krigagem com amostras de diferente suporte

Esse capítulo compara o método indireto de estimativa e a krigagem com amostras de diferente suporte. A comparação foi feita primeiro em um ambiente controlado (seção 4.1) e depois foi feito um estudo de validação cruzada com um banco de dados proveniente de um depósito de bauxita (seção 4.2). Para visualizar melhor a influência do método indireto sobre a variável teor, os pesos estandardizados (mostrados na equação 4, seção 2.1.2) foram utilizados no método indireto para o cálculo da variância do erro de estimativa (equação 2, seção 2.1.1). Essa abordagem é válida quando os mesmos pesos são utilizados para estimar a variável acumulação e espessura.

4.1 Exemplo simples

4.1.1 Referência: krigagem utilizando amostras quase pontuais

O cenário de referência consiste na estimativa do teor médio de um segmento de 3 m centrado em um ponto $\mathbf{u}$ usando seis amostras de 1 m. As amostras 1-5 estão alinhadas na vertical e representam um furo de sondagem de 5 m amostrado a cada metro e a amostra 6 representa um furo de sondagem com comprimento de 1 m (veja a figura 8). As amostras de 1 m são consideradas no suporte quase pontual. O segmento de 3 m a ser estimado foi discretizado com 3 pontos para o cálculo das covariâncias bloco-a-bloco.

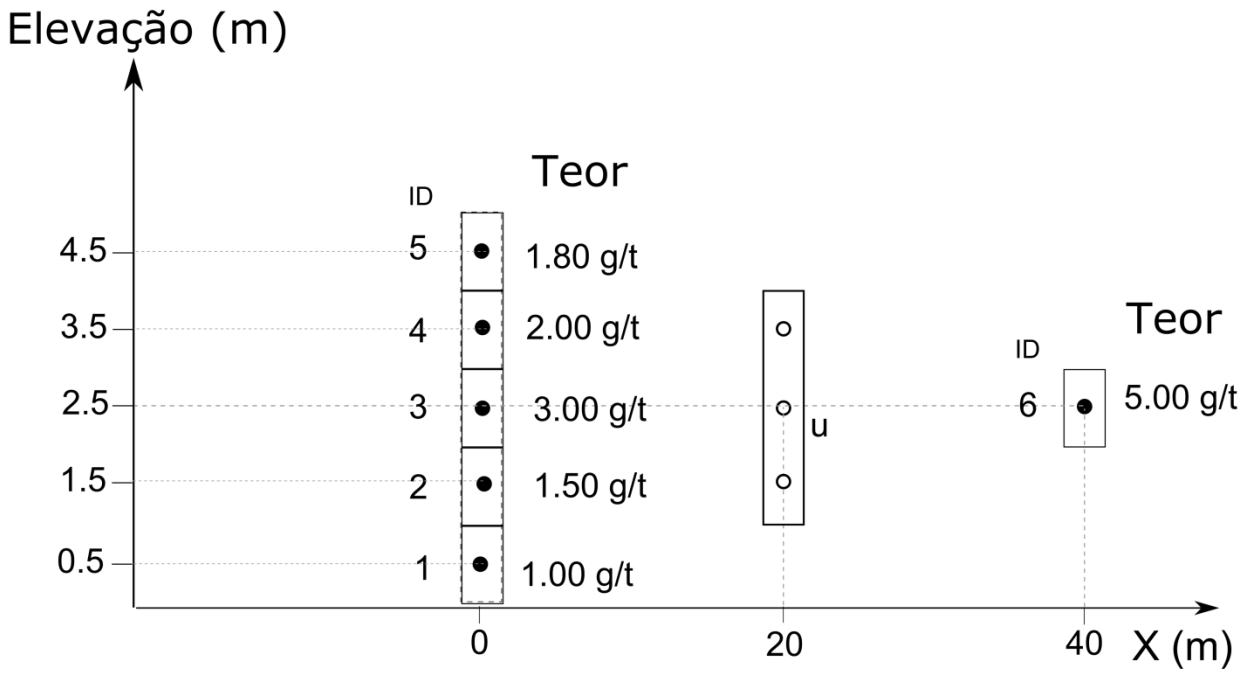

Figura 8: Cenário de referência. Estimativa do teor médio de um segmento de 3 m centrado no local **u** utilizando seis amostras quase pontuais.

4.1.2 Cenário das estimativas: amostra de linha e amostra pontual

As amostras 1-5 no cenário de referência foram regularizadas. O resultado é uma composta de 5 m de comprimento (veja a figura 9). O teor da composta (1.86 g/t) é o teor médio (ponderado pelo comprimento) das amostras 1-5 no cenário de referência e é chamada de amostra de linha. Duas metodologias foram consideradas para estimar o teor médio do segmento de 3 m centrado no ponto **u** usando essas amostras que tem diferente comprimento: (1) krigagem com amostras de diferente suporte e (2) método indireto (utilizando acumulação). A krigagem ordinária foi utilizada.

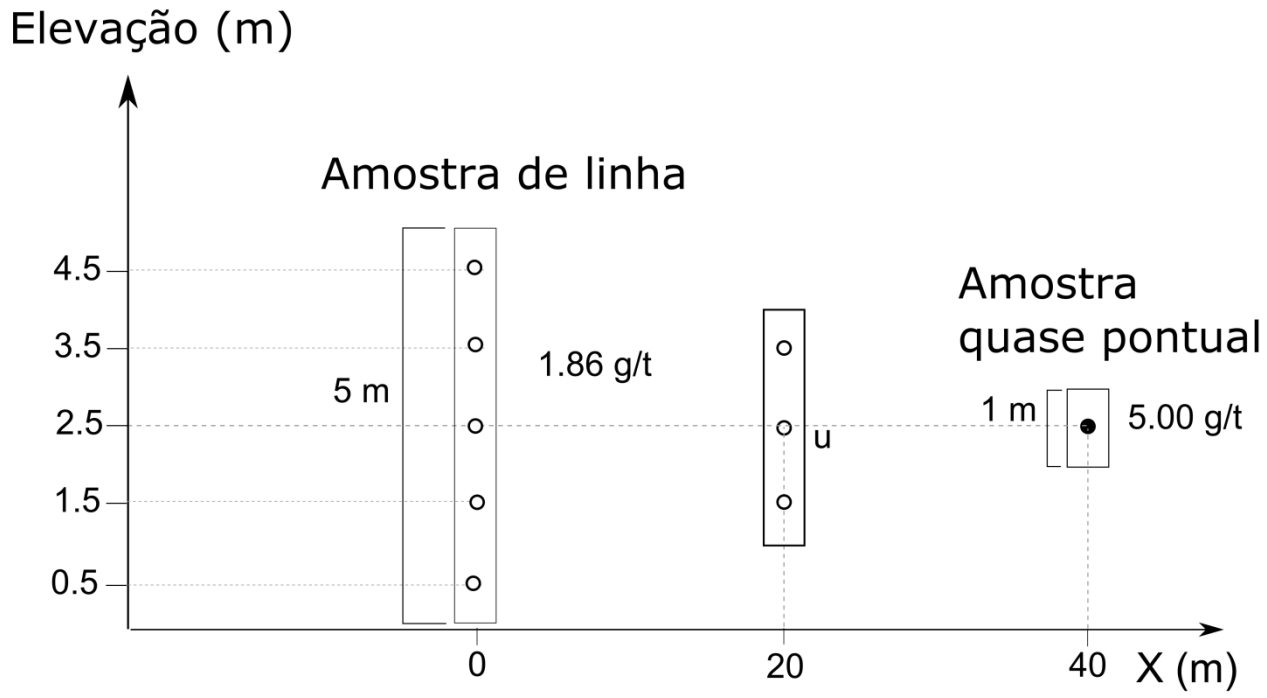

Figura 9: Cenário das estimativas: estimativa do teor médio de um segmento de 3 m centrado no local **u** com uma amostra de linha e uma amostra quase pontual. Os pontos brancos na amostra na linha são pontos discretizantes usados para calcular as covariâncias entre os dados para a krigagem com amostras de diferente suporte.

4.1.3 Análise de sensibilidade

A estimativa do teor médio do segmento de 3 m centrado no local $\mathbf{u}$ foi feita com diferentes modelos de variograma. Para analisar a influência do método de estimativa nas estimativas, o mesmo modelo de variograma foi usado para estimar as variáveis Teor, Acumulação e Espessura em cada comparação. O modelo de variograma é definido no suporte de 1 m, que é considerado um suporte quase pontual. Os efeitos no peso da amostra de linha, na estimativa e na variância do erro de estimativa foram medidos. O peso da amostra de linha foi comparado com a soma dos pesos das amostras 1-5 no cenário de referência (krigagem com amostras quase pontuais). A variância do erro $\sigma_V^2(\mathbf{u})$ foi dividida pela variância de bloco $\overline{C}(V,V)$. Comparar variâncias do erro calculadas com diferentes modelos de variograma pode gerar resultados enganosos. Uma variância do erro aparentemente baixa é alta se ela é

similar à variância de bloco. O interesse é na variância do erro em relação à variância total do bloco.

A tabela 2 resume a análise de sensibilidade. Quando um parâmetro era testado, os outros permaneciam constantes para cada rodada comparativa. No total, três rodadas comparativas foram feitas para testar os três parâmetros correspondentes (efeito pepita, alcance do variograma e tipo do variograma). Por exemplo, a primeira rodada comparativa testa o efeito pepita ($C0$). Nessa rodada, o modelo de variograma consistiu de um efeito pepita mais uma estrutura esférica com um alcance fixo de 60 m e uma contribuição ($C1$) de um menos o efeito pepita testado.

Tabela 2: Resumo dos parâmetros testados na análise de sensibilidade

Rodada comparativa	Parâmetro testado (variável)	Modelo de variograma
1	Efeito pepita (*C0*)	$\gamma(h) = C0 + C1 \cdot Sph(\frac{h}{60m})$ $C1 = 1 - C0$
2	Alcance (*a*)	$\gamma(h) = 0.1 + 0.9 \cdot Sph(\frac{h}{a})$
3	Tipo de variograma (*vtype*): Gaussiano, esférico, exponencial	$\gamma(h) = 0.1 + 0.9 \cdot vtype(\frac{h}{60m})$

4.1.4 Resultados

Efeito do efeito pepita

Quando o efeito pepita é zero, a soma dos pesos para as amostras 1-5 no cenário de referência é aproximadamente 0.50 (veja a figura 10a). Como o fenômeno é espacialmente contínuo (baixo efeito pepita), as amostras 1-5 atuam como um *cluster*,

ou seja, elas são redundantes. Como resultado, o sistema de krigagem reduz o peso dessas amostras. O mesmo efeito ocorre no caso da krigagem com amostras de diferente suporte em relação à composta de 5 m. A krigagem com amostras de diferente comprimento considerou a redundância dessa amostra de 5 m. Por outro lado, o método indireto dá mais peso para a amostra de linha em comparação com krigagem com amostras de diferente suporte. O método indireto atua em duas dimensões e não desagrupa ao longo da dimensão vertical.

À medida que o efeito pepita aumenta, a krigagem perde a sua propriedade de desagrupamento (Journel e Huijbregts, 1978). Os pesos das amostras 1-5 no cenário de referência se tornam menos redundantes. Como resultado, a soma dos pesos das amostras 1-5 aumenta. De maneira similar, o peso da amostra de linha aumenta no caso de krigagem com amostras de diferente suporte. Para o caso extremo de efeito pepita puro, o método indireto e krigagem com amostras de diferente suporte resultaram no mesmo peso para a amostra de linha.

De um modo geral, o peso da amostra de linha na krigagem com amostras de diferente suporte é similar à soma dos pesos das amostras 1-5 no cenário de referência. Como resultado, as estimativas no caso da krigagem com amostras de diferente suporte e no cenário de referência são bastante similares (figura 10b). A estimativa para o método indireto é menor porque ela resultou em um maior peso para amostra de linha. Como a amostra de linha possui teor menor do que a amostra pontual, a estimativa resultante é menor. No caso de efeito pepita puro, a estimativa é a mesma para os três casos.

No geral, a variância do erro obtida por krigagem com amostras de diferente suporte foi menor do que a variância do erro obtida pelo método indireto (figura 10c). Esse resultado é esperado, porque a krigagem produz os pesos que minimizam a variância do erro.

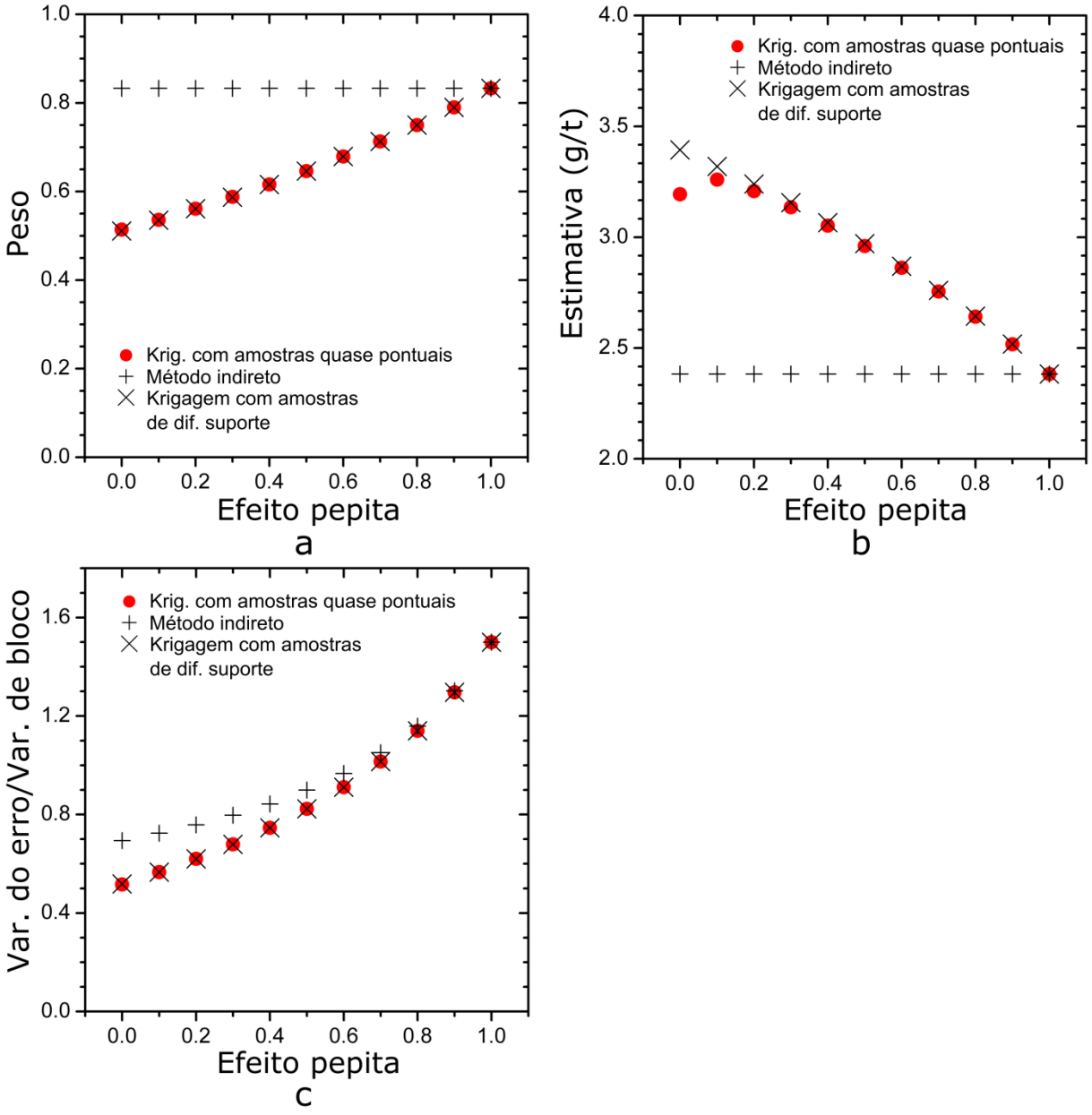

Figura 10: Influência do efeito pepita no peso (a), na estimativa (b) e na variância do erro (c). O peso em (a) se refere ao peso da amostra de linha para o método indireto e para a krigagem com amostras de diferente suporte. Para a krigagem com amostras quase pontuais, o peso em (a) se refere à soma dos pesos das amostras 1-5 na figura 8.

Efeito do alcance do variograma

De maneira geral, o peso da amostra de linha na krigagem com amostras de diferente suporte é parecido com a soma dos pesos 1-5 no cenário de referência (figura 11a). Em comparação com esses dois métodos, o método indireto resultou em um maior peso para a amostra de linha. O resultado é que a estimativa pelo método indireto discorda das outras duas estimativas (figura 11b). Para todos os alcances de variograma considerados, a variância do erro foi maior para o método indireto (figura 11c).

O peso da amostra de linha, estimativa e variância do erro foram pouco afetados pelos alcances de variograma considerados na krigagem com amostras de diferente suporte (figura 11). Isso ocorreu porque os alcances considerados pouco afetaram as covariâncias bloco-a-bloco. As covariâncias bloco-a-bloco são impactadas fortemente se o alcance é curto em relação ao suporte da amostra. Os alcances considerados são maiores do que o comprimento da amostra de linha (a amostra de linha possui 5 m de comprimento enquanto que o menor alcance considerado é de 20 m).

A variância do erro diminuiu à medida que o alcance do variograma aumentou (figura 11c). Conforme o alcance do variograma aumenta, o fenômeno é espacialmente mais contínuo e o decréscimo da variância do erro é esperado.

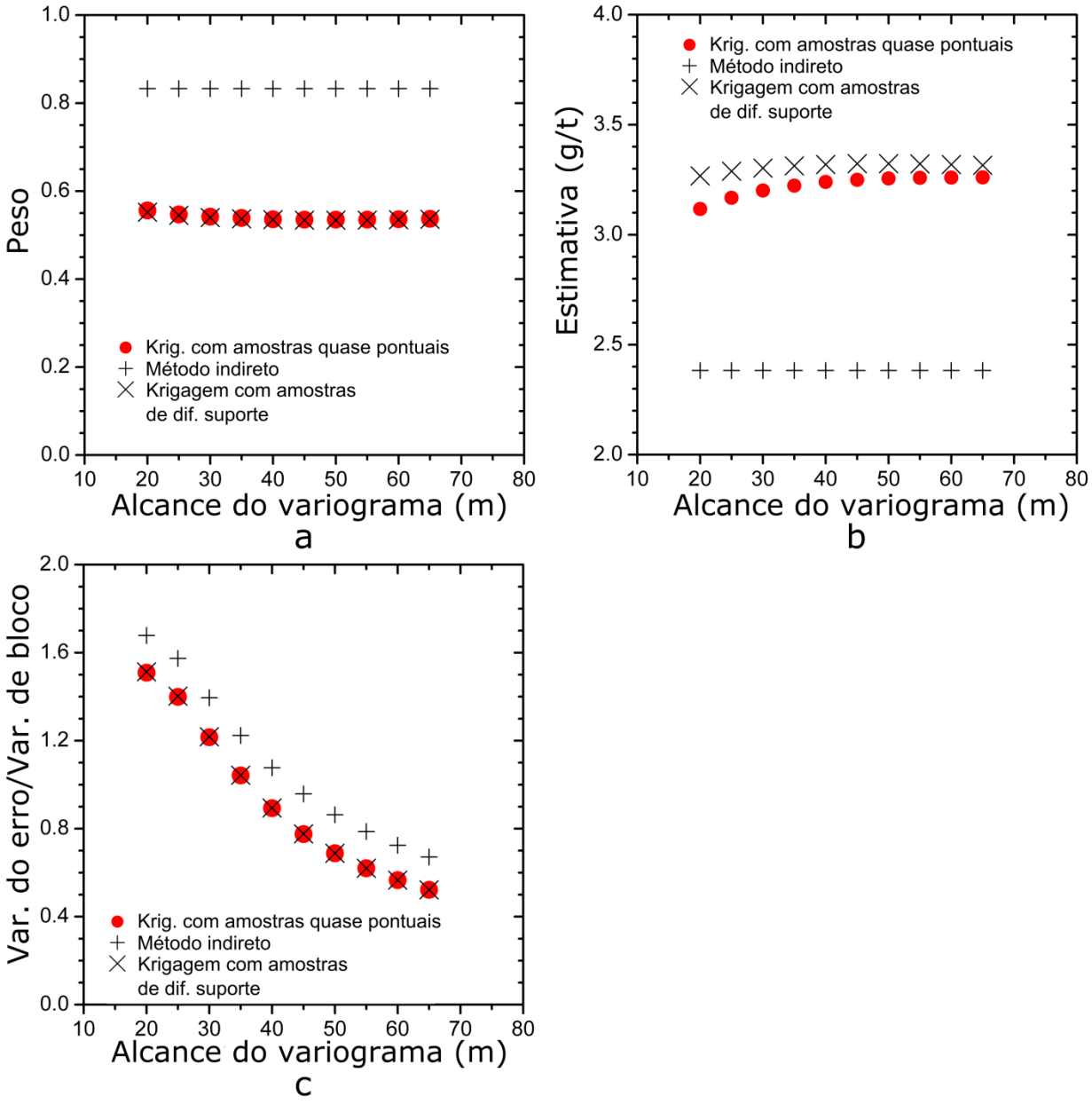

Figura 11: Influência do alcance do variograma no peso (a), na estimativa (b) e na variância do erro (c). O peso em (a) se refere ao peso da amostra de linha para o método indireto e para a krigagem com amostras de diferente suporte. Para a krigagem com amostras quase pontuais, o peso em (a) se refere à soma dos pesos das amostras 1-5 na figura 8.

Efeito do tipo de variograma

A figura 12 mostra o efeito do tipo de variograma no peso, na estimativa e na variância do erro. Para os tipos de variograma considerados, o método indireto resultou em maior peso para a amostra de linha em comparação com os outros dois casos (cenário de referência e krigagem com amostras de diferente suporte). Como a amostra de linha tem teor menor, a estimativa com o método indireto é menor do que a estimativa nos outros dois casos. Além disso, o tipo de variograma teve pouca influência na estimativa e no peso da amostra de linha.

O modelo de variograma Gaussiano resultou na menor variância do erro, o variograma esférico resultou em uma variância do erro intermediária e o variograma exponencial resultou na maior variância do erro (figura 12c). O variograma Gaussiano é espacialmente mais contínuo do que o variograma esférico, que é espacialmente mais contínuo do que o variograma exponencial. A variância do erro diminuiu à medida que a continuidade espacial aumentou (figura 12c). Para os tipos de variograma considerados, o método indireto gerou variâncias do erro maiores do que os outros dois casos (cenário de referência e krigagem com amostras de diferente suporte).

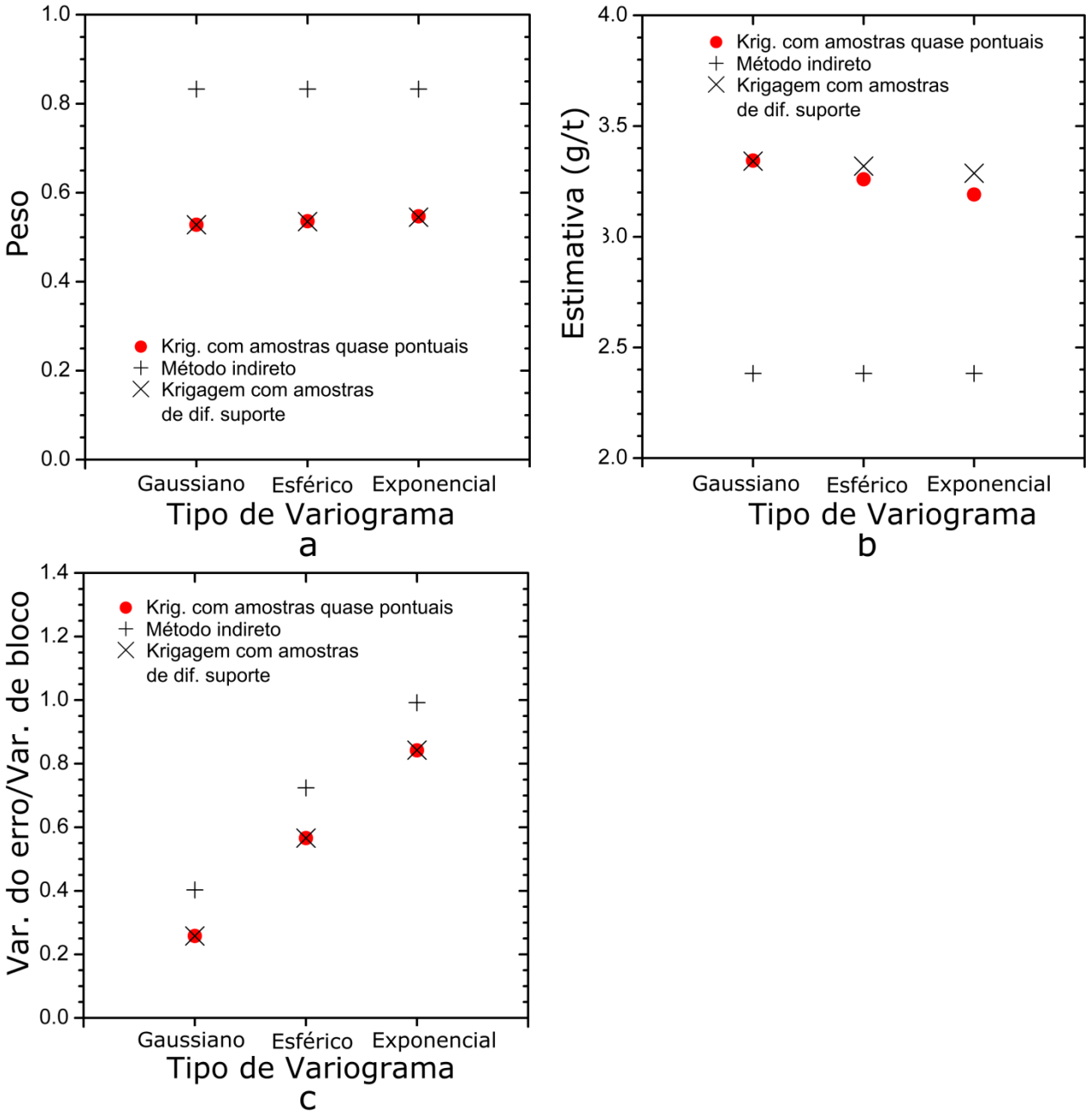

Figura 12: Influência do tipo de variograma no peso (a), na estimativa (b) e na variância do erro (c). O peso em (a) se refere ao peso da amostra de linha para o método indireto e para a krigagem com amostras de diferente suporte. No caso da krigagem com amostras quase pontuais, o peso em (a) se refere à soma dos pesos das amostras 1-5 na figura 8.

4.2 Estudo comparativo com validação cruzada

4.2.1 Banco de dados

O banco de dados é proveniente de um depósito de bauxita localizado no norte do Brasil. O banco de dados contém 686 furos de sondagem localizados em uma malha quase regular de 200 x 200 m ao longo de X e Y. As coordenadas Z originais foram transformadas em coordenadas estratigráficas. A variável de interesse é a fração mássica retida na peneira de 14# (REC14) expressa em porcentagem. REC14 ajuda a equipe de planejamento de mina a prever a fração grosseira da bauxita, que é recuperada para o processamento metalúrgico. Depois que a bauxita é minerada e peneirada, os finos são descartados. Os finos contêm principalmente minerais ricos em sílica. REC14 foi inicialmente amostrada no comprimento de 0.50 metro.

Para demonstrar as técnicas, que lidam com dados de diferente suporte, as amostras foram regularizadas ao longo da direção do furo com comprimentos diferentes. O banco de dados resultante contém 686 furos de sondagem e 1529 amostras cujos comprimentos variam de 0.30 m até 6.42 m (veja a tabela 3).

Como o método indireto necessita de um banco de dados em duas dimensões, REC14 foi regularizada ao longo da espessura do minério. A figura 13 mostra um exemplo ilustrativo de regularização pela espessura da camada. Cada furo de sondagem regularizado tem uma amostra que representa o valor médio ao longo da espessura do minério. O comprimento da amostra corresponde à espessura do minério. O banco de dados regularizado possui 686 furos de sondagem e 686 amostras cujos comprimentos variam de 0.30 m até 7.88 m (veja a tabela 3).

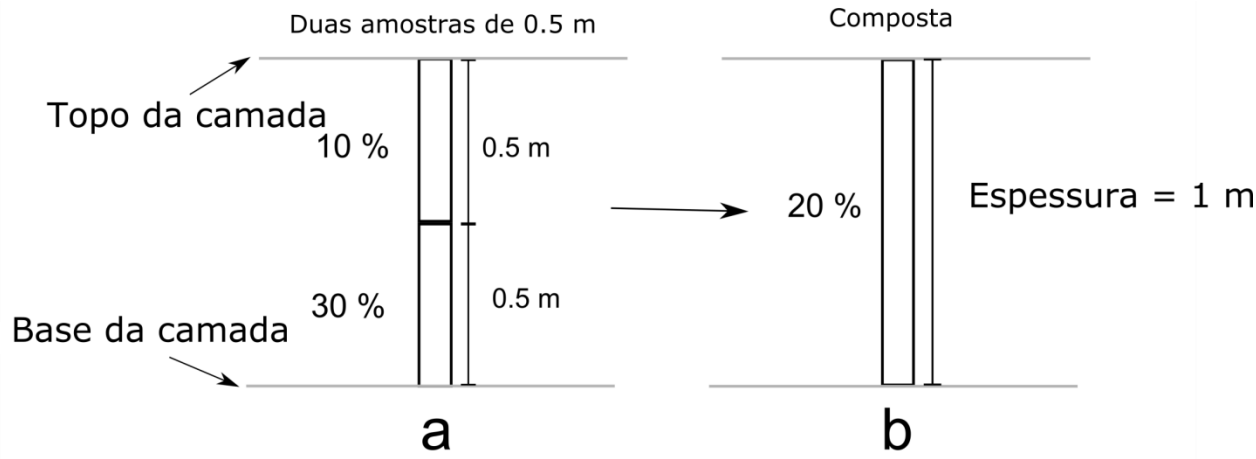

Figura 13: Exemplo de regularização pela espessura da camada. Furo de sondagem antes da regularização (a) e depois da regularização (b).

A tabela 3 mostra a estatística básica dos dois bancos de dados (antes e depois da regularização). A média de REC14 é a mesma para os dois bancos de dados, mas o desvio padrão é menor no banco de dados regularizado, como esperado. O comprimento máximo para o banco de dados regularizado é maior, pois a regularização combina amostras adjacentes em amostras maiores.

Tabela 3: Sumário estatístico para REC14 e comprimento para os bancos de dados 3D e 2D.

	Banco de dados 3D antes da regularização pela espessura da camada		Banco de dados 2D depois da regularização pelo espessura da camada	
	REC14 (%)	Comprimento (m)	REC14 (%)	Comprimento (m)
Número	1529	1529	686	686
Média	65.52	1.35	65.52	3.01
Desv. Pad.	12.43	1.05	8.93	1.59
CV	0.19	0.78	0.14	0.53
Mínimo	5.03	0.30	15.16	0.30
Quartil inferior	58.78	0.50	59.96	1.76
Mediana	66.95	1.00	65.56	3.00
Quartil superior	73.66	1.98	71.85	4.05
Máximo	97.09	6.42	89.28	7.88

A validação cruzada foi feita com o banco de dados regularizado para os dois métodos (método indireto e amostras com diferente suporte). Na prática, o banco de dados 3D com 1529 amostras seria usado para construir um modelo de teores usando krigagem com amostras de diferente suporte. Entretanto, a intenção é testar os dois métodos sob as mesmas condições, com a mesma quantidade de informação. Como o método indireto necessita de um banco de dados 2D, o banco de dados regularizado foi utilizado para testar os dois métodos.

O banco de dados 3D (com 1529 amostras) foi usado para obter o variograma de REC14 no suporte quase pontual para a krigagem com amostras de diferente suporte. Nesse estudo de caso, a krigagem com amostras de diferente suporte requer um modelo de variograma em três dimensões. As amostras foram discretizadas ao longo da direção vertical para calcular as covariâncias bloco-a-bloco. Portanto, um modelo de variograma ao longo da direção vertical é necessário. Não é possível modelar um variograma na vertical utilizando um banco de dados 2D.

4.2.2 Variograma de REC14

O variograma de REC14 foi obtido por deconvolução do variograma. O modelo de variograma é definido em um suporte quase pontual, que se refere a dados representando um segmento de 0.5 m. Journel e Huijbregts (1978, p. 231) afirmam que obter o verdadeiro modelo de variograma em suporte pontual é ilusório. O geoestatístico não consegue precisão maior do que o menor suporte dos dados sem assumir hipóteses que não podem ser verificadas (Journel e Huijbregts, 1978). A equação 15 descreve o modelo de variograma de REC14:

$$\gamma(h) = 0.22 + 0.66 \cdot Sph(\frac{NS}{250m}, \frac{EW}{250m}, \frac{vert}{3.00m}) + 0.12 \cdot Sph(\frac{NS}{8000m}, \frac{EW}{8000m}, \frac{vert}{5.50m}) \quad (15)$$

O modelo de variograma tem um alcance curto ao longo da vertical. Quanto mais espacialmente descontínuo é o atributo (alto efeito pepita e alcance curto), mais peso é atribuído a amostras longas na krigagem com amostras de diferente suporte. O variograma é estandardizado, com patamar igual a um.

4.2.3 Variogramas de Acumulação e Espessura

Os variogramas experimentais das variáveis Acumulação e Espessura foram calculados. A figura 14a mostra o variograma da Espessura no plano horizontal. Não há componente vertical para o variograma de Acumulação e Espessura. O modelo variográfico de Espessura mostrou um bom ajuste com os variogramas experimentais de Acumulação (veja a figura 14b). Como as variáveis Acumulação e Espessura possuem forte correlação (coeficiente de correlação de 0.96, veja a figura 14c), elas têm continuidade espacial semelhante.

A equação 16 define o modelo de variograma da Espessura:

$$\gamma(h) = 0.35 + 0.30 \cdot Sph(\frac{NS}{800m}, \frac{EW}{800m}) + 0.35 \cdot Sph(\frac{NS}{2300m}, \frac{EW}{2300m}) \quad (16)$$

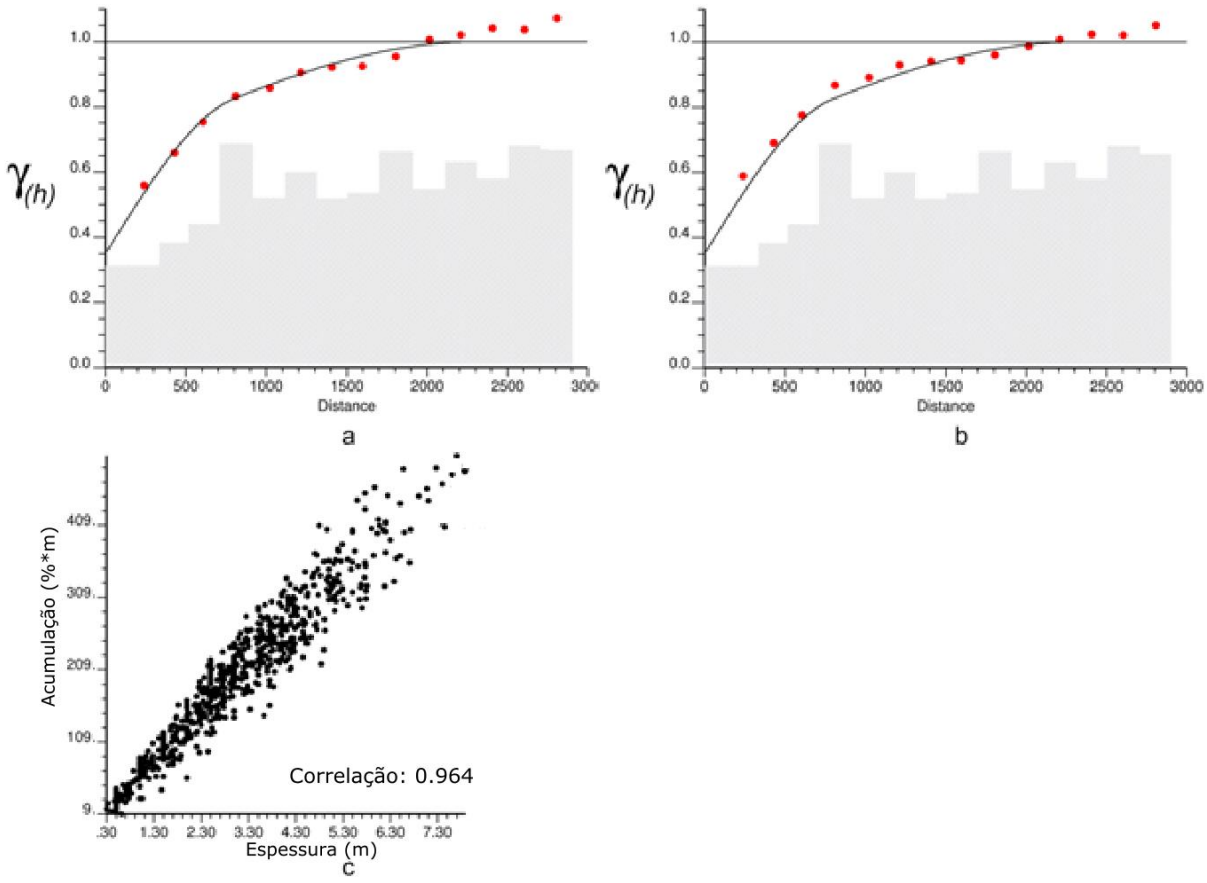

Figura 14: Variograma de Espessura (a) e variograma experimental da variável Acumulação junto com o modelo de variograma da variável Espessura (b) e diagrama de dispersão entre as variáveis Acumulação e Espessura (c).

4.2.4 Estimativa e validação cruzada

Validação cruzada consiste em primeiro remover uma amostra em um local particular. Segundo, o valor é estimado nesse local usando as amostras restantes. O erro médio, desvio padrão do erro, erro absoluto médio e erro quadrático médio foram calculados. O erro médio mede a acurácia das estimativas. O desvio padrão do erro, o erro absoluto médio e o erro quadrático médio medem a precisão das estimativas.

A estimativa foi feita com krigagem ordinária. Para a krigagem com amostras de diferente suporte, o modelo de variograma em suporte de ponto de REC14 foi utilizado. Para o método indireto, o modelo de variograma da Espessura foi utilizado para estimar

as variáveis Espessura e Acumulação. A mesma estratégia de busca foi utilizada para os dois métodos de estimativa.

4.2.5 Resultados da validação cruzada

A tabela 4 mostra a estatística básica do erro de estimativa. Os dois métodos são quase não enviesados, pois o erro médio é próximo de zero para os dois casos. A krigagem com amostras de diferente suporte produziu as estimativas mais precisas. O desvio padrão do erro, o erro absoluto médio, e o erro quadrático médio são menores para o caso da krigagem com amostras de diferente suporte. Especificamente, o erro quadrático médio é aproximadamente 5% menor do que o erro quadrático médio obtido com o método indireto.

Esses resultados são consistentes com aqueles encontrados no exemplo simples. No exemplo simples, a krigagem com amostras de diferente suporte mostrou maior precisão que o método indireto, porque a variância do erro foi menor para a krigagem com amostras de diferente suporte.

Tabela 4: Comparativo entre a krigagem com amostras de diferente suporte e o método indireto.

	Krigagem com amostras de diferente suporte	Método indireto
Número	686	686
Erro médio (%)	-0.08	0.08
Desv. Pad. do erro (%)	8.17	8.39
Erro absoluto médio (%)	6.39	6.55
Erro quadrático médio ($\%^2$)	66.68	70.34

Para analisar de perto a influência do método de estimativa nas estimativas, um ponto estimado na validação cruzada foi selecionado. A figura 15 mostra o mapa de localização do ponto estimado junto com as três amostras mais próximas utilizadas na estimativa. Cada amostra possui teor (representado pela cor na figura 15) e

comprimento (representado pela área do círculo na figura 15). Há duas amostras longas com alto teor e uma amostra curta com baixo teor (veja a figura 15).

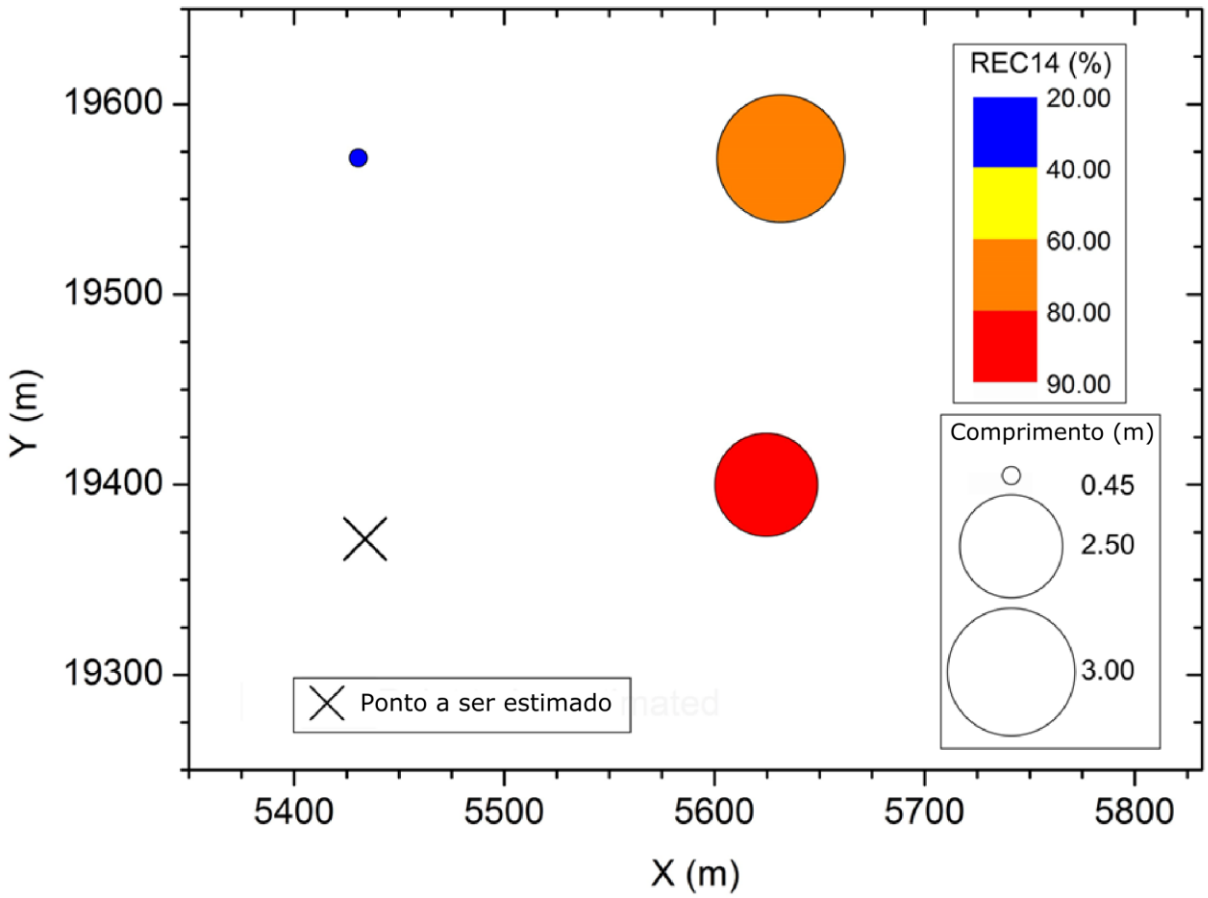

Figura 15: Mapa de localização de um ponto a ser estimado e as três amostras mais próximas usadas na estimativa. A área dos círculos é proporcional ao comprimento das amostras.

A tabela 5 mostra a estimativa no ponto a ser estimado na figura 15 pelos dois métodos (krigagem com amostras de diferente suporte e método indireto). A estimativa é maior para o método indireto. Isso ocorreu porque o método indireto resultou em maior peso para as amostras de maior comprimento. Como as amostras de maior comprimento têm maior teor, a estimativa é maior para o método indireto. O resultado é consistente com aqueles obtidos para o exemplo simples. No exemplo simples, o método indireto resultou em maior peso para as amostras de maior comprimento.

Tabela 5: Estimativa de REC14 no ponto a ser estimado na figura 15.

Método de estimativa	Estimativa de REC14 (%)
Método indireto	75.74
Krigagem com amostras de diferente suporte	69.63

4.3 Observações

Esse capítulo comparou o método indireto com a krigagem considerando amostras de diferente suporte para estimar teores usando amostras de diferente comprimento. Para o método indireto, as variáveis acumulação e espessura foram estimadas usando o mesmo modelo de variograma. Essa abordagem mitiga o problema de estimativas fora do intervalo definido pelo mínimo e máximo dos dados e é frequentemente utilizada por geomodeladores.

Em geral, o método indireto resulta em maior peso para amostras longas do que a krigagem com amostras de diferente suporte. Essa diferença aumenta quando o atributo é espacialmente contínuo (baixo efeito pepita, longo alcance). Quando o atributo é espacialmente contínuo, a krigagem com amostras de diferente suporte aparentou lidar melhor com a redundância de uma amostra longa. O resultado é que a krigagem com amostras de diferente suporte atribui menos peso para amostras longas. Por outro lado, o método indireto trabalha em duas dimensões e não desagrupa ao longo da dimensão vertical perdida. O resultado é que o método indireto tende a atribuir maior peso para amostras longas.

A krigagem com amostras de diferente suporte necessita de um modelo de variograma em suporte pontual. Os pesos obtidos na krigagem com amostras de diferente suporte são sensíveis ao efeito pepita do modelo de variograma. O peso atribuído às amostras longas aumenta à medida que o efeito pepita aumenta.

Sob um ponto de vista prático, o método indireto reforça a influência das amostras longas nas estimativas vizinhas. A diferença entre as estimativas geradas pelos dois métodos (método indireto e krigagem com amostras de diferente suporte) é maior quando as amostras na vizinhança possuem teores e suportes bastante distintos e o variograma em suporte quase pontual dos teores possui um longo alcance e baixo efeito pepita (espacialmente contínuo).

Um estudo com validação cruzada foi realizado utilizando os dois métodos de estimativa em um banco de dados de um depósito de bauxita. Os resultados mostraram que a krigagem com amostras de diferente suporte resultou em estimativas mais precisas do que o método indireto. O estudo de caso foi aplicado para uma variável com uma distribuição praticamente simétrica com um coeficiente de variação baixo.

5 Simulação geoestatística e transformações multivariadas

A seção 5.1 explica de forma sucinta a simulação sequencial Gaussiana (*sequential Gaussian simulation* – SGS). A SGS de uma variável exige a transformação *normal score*, que é apresentada na seção 5.2. A transformação *normal score* transforma os dados em dados Gaussianos univariados e é adequada para a SGS de uma variável. No caso de múltiplas variáveis, a SGS necessita de uma transformação que as torne variáveis multi-Gaussianas e independentes. A transformação *Projection Pursuit Multivariate Transform* (PPMT) é utilizada nessa tese para tornar as variáveis multi-Gaussianas e independentes. A PPMT é explicada na seção 5.3. A seção 5.4 apresenta uma versão modificada da PPMT feita para lidar com variáveis que tem restrições de soma. Quando as variáveis possuem restrições de soma e fração, é comum transformar os dados utilizando razões. A seção 5.5 apresenta razões utilizadas para lidar com restrições de soma e fração no banco de dados.

5.1 Simulação sequencial gaussiana

A simulação sequencial Gaussiana consiste nos seguintes passos:

1. Transformar os dados originais em dados gaussianos utilizando a transformação *normal score*;

2. Escolher aleatoriamente uma localização $\mathbf{u}$ e procurar por dados e nós previamente simulados no entorno de $\mathbf{u}$;

3. Calcular a média e variância da função de distribuição cumulativa condicional (*conditional cumulative distribution function* – ccdf) local Gaussiana por krigagem simples utilizando os dados e nós previamente simulados. A média corresponde à estimativa por krigagem simples e variância corresponde à variância de krigagem simples;

4. Sortear um valor da ccdf local Gaussiana e adicionar ao banco de dados;

5. Repetir os passos 2-4 até que todos os nós do grid sejam simulados.

6. Retro-transformar os valores simulados Gaussianos para o espaço dos dados originais.

O processo é repetido para uma série de realizações.

5.2 Transformação *normal score*

A transformação *normal score* transforma os dados originais z em valores gaussianos y. Os valores gaussianos y seguem uma função de distribuição cumulativa (*cumulative distribution function* – cdf) Gaussiana padrão e são chamados de valores *normal score*. A transformação *normal score* pode ser obtida através de uma correspondência gráfica entre a cdf dos dados originais $F(z)$ e a cdf de uma distribuição Gaussiana padrão $G(y)$.

Graficamente, $F(z)$ pode ser interpretada como uma função cujo parâmetro de entrada é um valor z dos dados originais e o resultado é a probabilidade acumulada correspondente, que está entre 0 e 1 (veja a figura 16). De maneira similar, a função $G(y)$ calcula a cdf (situada entre 0 e 1) associada a um valor *normal score* y.

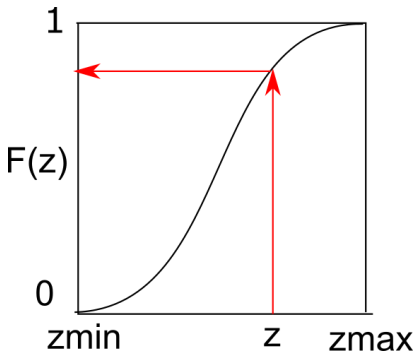

Figura 16: Ilustração da função F(z).

A equação 17 define a transformação *normal score*:

$$y = G^{-1}[F(z)] \tag{17}$$

onde G^{-1} é a função quantil da distribuição Gaussiana padrão. A função G^{-1} tem como parâmetro de entrada um valor de probabilidade acumulada (cdf) situado entre [0,

1] e retorna o valor *normal score* correspondente. A figura 167 mostra a função quantil da distribuição Gaussiana padrão para um valor *p*:

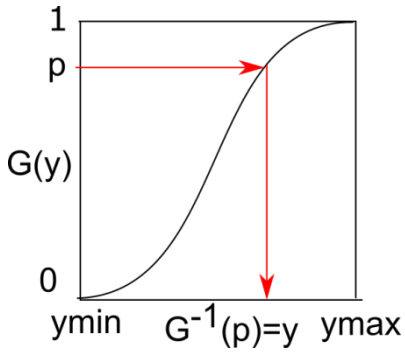

Figura 17: Ilustração da função $G^{-1}(p)$.

A transformação *normal score* pode ser vista graficamente como uma tabela de correspondência quantil-quantil. Cada quantil no espaço original z possui um quantil correspondente no espaço gaussiano y (veja a figura 18).

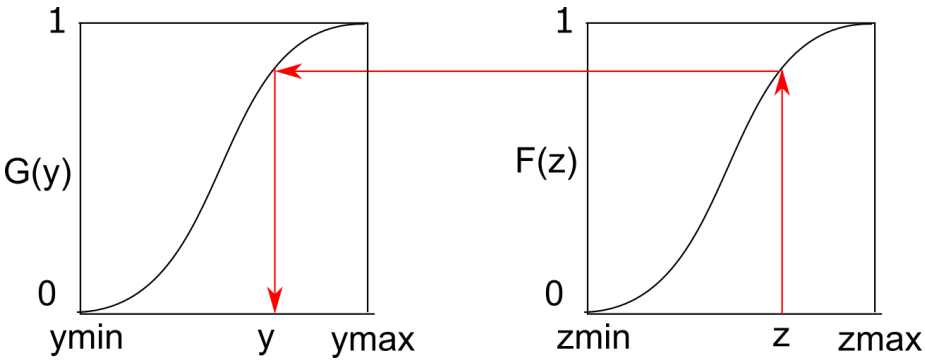

Figura 18: Esquema gráfico da transformação *normal score*.

A transformação *normal score* inversa é definida pela equação 18:

$$z = F^{-1}[G(y)] \tag{18}$$

A transformação *normal score* transforma os dados de forma que eles sejam univariados Gaussianos. Essa transformação é apropriada no caso de SGS de uma

variável. No caso de múltiplas variáveis, a SGS requer que os dados sejam multi-Gaussianos e independentes.

5.3 *Projection Pursuit Multivariate Transform*

A transformação *projection pursuit multivariate transform* (PPMT) lida com dados multivariados. A seção 5.3.1 apresenta a notação utilizada para representar um conjunto de dados multivariados. A PPMT envolve duas etapas de pré-processamento dos dados: (1) transformação *normal score* e (2) *sphering*. A seção 5.3.2 explica a operação de *sphering*. A PPMT envolve o conceito de projeção dos dados. A seção 5.3.3 explica a obtenção da projeção dos dados. A seção 5.3.4 explica o algoritmo de PPMT.

5.3.1 Notação dos dados

Um conjunto de dados com n amostras e m variáveis é definido pela matriz $\mathbf{Z}$. Cada coluna em $\mathbf{Z}$ representa uma variável e cada linha representa uma observação com as m variáveis. A figura 19a mostra a representação de uma matriz de dados com n amostras e m variáveis. A primeira variável é representada pelo vetor $\mathbf{z}_1$, que tem n amostras (figura 19b).

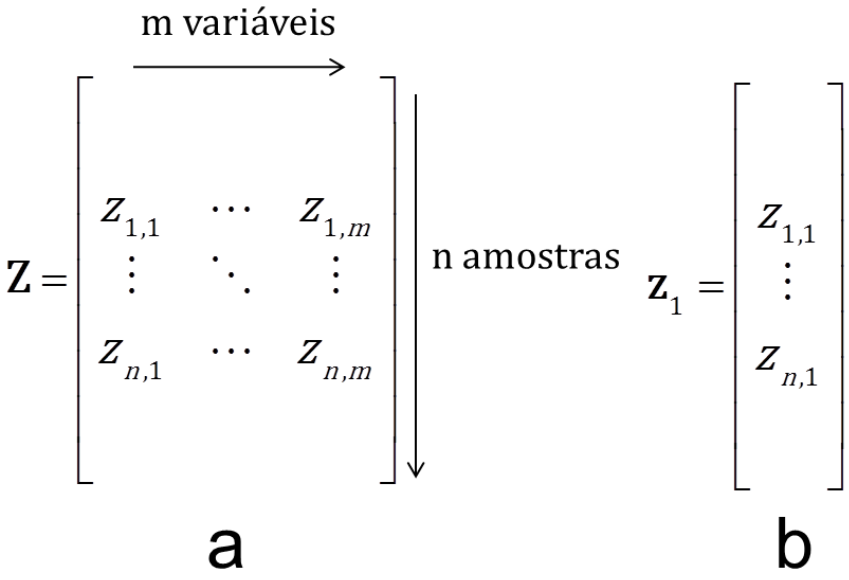

Figura 19: Representação de uma matriz de dados (a) e do vetor da primeira variável (b).

5.3.2 *Sphering*

A operação de *sphering* descorrelaciona as variáveis. Considere uma matriz de dados $\mathbf{W}$. Primeiro o vetor de médias $\boldsymbol{\mu}$ e dos dados $\mathbf{W}$ são calculados. *Sphering* dos dados $\mathbf{W}$ consiste na seguinte operação (equação 19):

$$\mathbf{X} = \mathbf{S}^{-1/2}(\mathbf{W} - \boldsymbol{\mu}) \tag{19}$$

onde $\mathbf{S}^{-1/2}$ é obtida pela através da equação 20:

$$\mathbf{S}^{-1/2} = \mathbf{V}\mathbf{D}^{-1/2}\mathbf{V}^{\mathrm{T}} \tag{20}$$

$\mathbf{D}$ e $\mathbf{V}$ são as matrizes de autovalores e autovetores obtidas a partir da decomposição espectral da matriz de covariância $\sum 0$:

$$\sum 0 = \mathbf{VDV}^{\mathrm{T}} \tag{21}$$

A operação de *sphering* utilizada nessa tese foi apresentada por Barnett *et al.* (2016). Essa implementação diminui a mistura entre as variáveis originais nas variáveis transformadas em comparação com a operação de *sphering* apresentada por Barnett *et al.* (2014). Como consequência, há uma maior correlação entre as variáveis originais e as variáveis transformadas depois da operação de *sphering*. A maior correlação entre as variáveis transformadas com as variáveis originais aumenta a chance que as simulações reproduzam as propriedades das variáveis originais, como o variograma (Barnett *et al.*, 2016).

5.3.3 Projeção

A equação 22 define a projeção $\mathbf{p}$ dos dados $\mathbf{Z}$ sobre um vetor unitário $\boldsymbol{\theta}$:

$$\mathbf{p} = \mathbf{Z}\boldsymbol{\theta} \tag{22}$$

5.3.4 PPMT

A transformação PPMT para um conjunto de variáveis isotópicas consiste nos seguintes passos:

1. Aplicar a transformação *normal score* nas variáveis originais;
2. Aplicar a operação de *sphering* nos dados *normal score*;
3. Procurar por uma projeção interessante dos dados e aplicar a transformação *normal score* ao longo dessa projeção. O passo 3 é repetido até que a distribuição seja multi-Gaussiana com uma matriz de covariância igual a matriz identidade.

Se os dados são multi-Gaussianos, qualquer projeção dos dados tem uma distribuição Gaussiana. Projeções interessantes são as projeções menos Gaussianas. O índice I proposto por Friedman (1987) é utilizado para medir a não-Gaussianeidade de uma projeção. O índice I calcula a diferença entre uma distribuição Gaussiana e a

distribuição dos dados ao longo de uma projeção. É usada uma busca otimizada para encontrar as projeções que maximizam o índice I. Detalhes sobre a busca otimizada podem ser encontrados em Friedman (1987).

Depois da operação de *sphering*, os dados $\mathbf{X}$ são transformados em $\tilde{\mathbf{X}}$ de modo que a projeção de $\tilde{\mathbf{X}}$ ao longo do vetor unitário $\boldsymbol{\theta}$ tenha uma distribuição Gaussiana. A transformação de $\mathbf{X}$ em $\tilde{\mathbf{X}}$ envolve os seguintes passos:

1. Construção de uma matriz ortonormal $\mathbf{U}$ em que a primeira coluna corresponde ao vetor $\boldsymbol{\theta}$ (equação 23):

$$\mathbf{U} = [\boldsymbol{\theta}, \boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_{m-1}] \tag{23}$$

onde os $m-1$ vetores unitários $\boldsymbol{\beta}_i$ são obtidos através do algoritmo de Gram-Schmidt.

2. Multiplicação das matrizes $\mathbf{X}$ e $\mathbf{U}$:

$$\mathbf{XU} = [\mathbf{p}, \mathbf{X}\boldsymbol{\beta}_1, ..., \mathbf{X}\boldsymbol{\beta}_{m-1}] \tag{24}$$

onde a primeira coluna da matriz $\mathbf{XU}$ corresponde à projeção $\mathbf{p}$.

3. *Normal score* da primeira coluna da matriz $\mathbf{XU}$. Essa transformação é denotada pelo símbolo Θ e está expressa na equação 25. Θ transforma a projeção $\mathbf{p}$ em uma variável Gaussiana $\tilde{\mathbf{p}}$:

$$\Theta(\mathbf{XU}) = [\tilde{\mathbf{p}}, \mathbf{X}\boldsymbol{\beta}_1, ..., \mathbf{X}\boldsymbol{\beta}_{m-1}] \tag{25}$$

3. Multiplicação da matriz $\Theta(\mathbf{XU})$ pela matriz $\mathbf{U}^{\mathrm{T}}$ (equação 26):

$$\tilde{\mathbf{X}} = [\Theta(\mathbf{XU})]\mathbf{U}^{\mathrm{T}} \tag{26}$$

A projeção dos dados transformados $\tilde{\mathbf{X}}$ ao longo do vetor $\boldsymbol{\theta}$ segue uma distribuição Gaussiana padrão.

5.4 PPMT com vetores controlados

O uso de projeções “especiais” na PPMT para lidar com restrições de soma é investigado nessa tese. Os vetores das projeções “especiais” são os vetores ortogonais aos vetores que definem as restrições de soma. A figura 20 ilustra a restrição em que soma das variáveis Z1 e Z2 deve ser menor ou igual a 100%. A restrição é representada pelo vetor [1, -1] (linha vermelha na figura 20). O vetor ortogonal à restrição é o vetor [1, 1] (linha azul na figura 20).

A projeção dos dados ao longo do vetor [1, 1] consiste na soma das variáveis Z1 e Z2. Então a transformação *normal score* é aplicada sobre essa soma. A transformação *normal score* inversa não permite extrapolação. Como a transformação *normal score* é aplicada para a soma das variáveis, espera-se que a soma dos valores simulados retro-transformados respeite a restrição de soma.

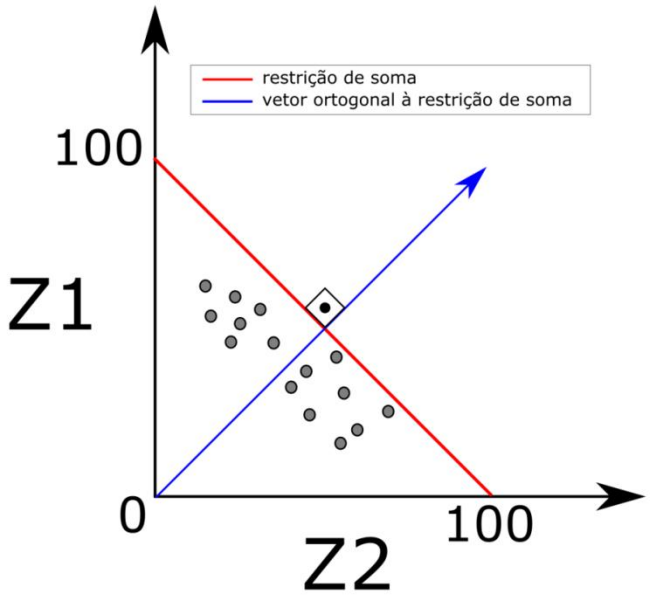

Figura 20: Esquema da restrição de soma para duas variáveis e vetor ortogonal à restrição de soma.

Depois das projeções “especiais” iniciais, a operação de *sphering* é aplicada nos dados e a PPMT prossegue usando a busca otimizada para encontrar as projeções menos Gaussianas (Barnett *et al.*, 2016). A figura 21 mostra a diferença entre a transformação PPMT convencional e a transformação PPMT proposta, que usa vetores ortogonais às restrições de soma.

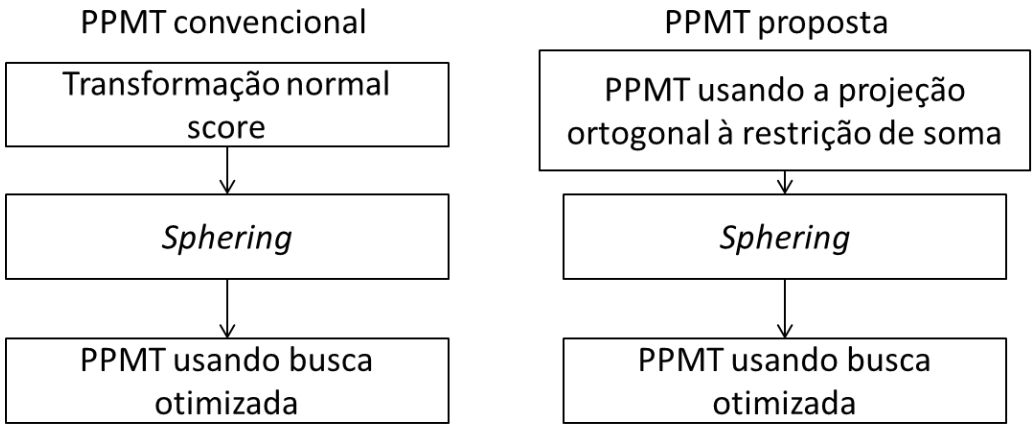

Figura 21: PPMT convencional e PPMT usando vetores ortogonais à restrição de soma.

5.5 Razões

5.5.1 Razões A

As razões A são similares às razões aditivas logarítmicas (*additive log-ratio* - alr), mas o logaritmo não é calculado. As razões A são aplicadas para dados composicionais, cuja soma seja igual a uma constante. Se a soma das variáveis não é igual a uma constante, uma variável *filler* pode ser adicionada. A variável *filler* é obtida através da constante menos a soma das variáveis restantes. A equação 27 mostra o cálculo das razões A:

$$\mathbf{a}_k = \frac{\mathbf{z}_k}{\mathbf{z}_m}, k = 1,...,m \tag{27}$$

A equação 28 mostra a operação inversa das razões A:

$$\mathbf{z}_k = \frac{\mathbf{a}_k}{\left(\sum_{k=1}^{m} \mathbf{a}_k\right) + 1}, k = 1,...,m \tag{28}$$

5.5.2 Razões U

As razões U foram aplicadas por Mery *et al.* (2017) para a simulação geoestatística de teores em um depósito de ferro. Primeiro, as variáveis devem ser reordenadas da variável com a menor média para a variável com a maior média. A equação 29 define o cálculo das razões U:

$$\begin{aligned} \mathbf{u}_1 &= \mathbf{z}_1 \\ \mathbf{u}_k &= \frac{\mathbf{z}_k}{100 - \sum_{i=1}^{k-1} \mathbf{z}_i}, k = 2,...,m \end{aligned} \tag{29}$$

A equação 30 define a operação inversa das razões U:

$$\begin{aligned} \mathbf{z}_1 &= \mathbf{u}_1 \\ \mathbf{z}_k &= \mathbf{u}_k \cdot \left(100 - \sum_{i=1}^{k-1} \mathbf{z}_i\right), k = 2,...,m \end{aligned} \tag{30}$$

5.5.3 Razão de fração

Em depósitos minerais, algumas variáveis são frações de outra variável. Por exemplo, depósitos de cobre muitas vezes possuem as variáveis cobre total e cobre solúvel. O cobre solúvel precisa ser sempre menor ou igual ao cobre total. Essa condição precisa ser respeitada em um modelo simulado dessas duas variáveis. O teor

total é denominado $\mathbf{z}_{total}$ e o teor fracionário é denominado $\mathbf{z}_{frac}$. Simular diretamente as duas variáveis $\mathbf{z}_{total}$ e $\mathbf{z}_{frac}$ não garante que o valor simulado de $\mathbf{z}_{total}$ seja maior do que o valor simulado de $\mathbf{z}_{frac}$. A metodologia proposta é simular o teor total $\mathbf{z}_{total}$ e a razão de fração $\mathbf{f}$ definida pela equação 31:

$$\mathbf{f} = \frac{\mathbf{z}_{frac}}{\mathbf{z}_{total}} \tag{31}$$

As simulações são feitas com o teor total $\mathbf{z}_{total}$ e a razão de fração $\mathbf{f}$. O teor fracionário original $\mathbf{z}_{frac}$ é obtido através da multiplicação do teor total simulado e da razão de fração simulada.

6 Comparação de transformações multivariadas e simulação geoestatística para dados com restrições de fração e soma

A seção 6.1 apresenta o banco de dados. A seção 6.2 faz um estudo comparativo de quatro *workflows* para lidar com dados multivariados com restrições de fração e soma. O comparativo entre os resultados dos *workflows* foi feito utilizando modelos derivados de simulação de Monte Carlo. Os resultados obtidos na seção 6.2 foram usados para escolher o *workflow* para realizar a simulação geoestatística das variáveis, que é descrita na seção 6.3.

6.1 Banco de dados

Os dados foram obtidos de um depósito de bauxita no Brasil. A tabela 6 mostra as variáveis de interesse com as notações e definições.

Tabela 6: Variáveis e notações.

Variável	Unidade	Notação	Definição
Alumina total	%	AT	Teor de alumina
Alumina recuperável	%	AA	Teor de alumina recuperável
Sílica total	%	ST	Teor de sílica total
Sílica reativa	%	SR	Teor de sílica reativa
Óxido de ferro	%	FE	Teor de óxido de ferro
Óxido de titânio	%	TI	Teor de óxido de titânio
Recuperação	%	RC	Proporção mássica de fragmentos grosseiros

A tabela 7 mostra o sumário estatístico dos dados. O banco de dados é isotópico (veja a tabela 7, todas as variáveis possuem o mesmo número de amostras). Os maiores valores ocorrem para a Alumina Total e Alumina Recuperável. A Sílica Total e Sílica Reativa são contaminantes e possuem valores menores. Os teores totais

(Alumina Total e Sílica Total) possuem médias maiores do que os teores fracionários (Alumina Recuperável e Sílica Reativa), como esperado.

Tabela 7: Sumário estatístico dos dados.

	RC	AT	AA	ST	SR	FE	TI
Número	6267	6267	6267	6267	6267	6267	6267
Média (%)	16.82	50.56	46.40	4.35	3.51	17.17	1.03
Desv. pad. (%)	5.72	4.81	5.39	2.49	2.24	6.82	0.35
CV	0.34	0.10	0.12	0.57	0.64	0.40	0.34
Mínimo (%)	0.27	20.46	18.02	0.10	0.06	1.05	0.09
Q25 (%)	12.73	48.16	43.45	2.40	1.67	12.62	0.81
Q50 (%)	16.23	51.37	47.11	4.07	3.30	15.88	0.99
Q75 (%)	20.24	53.92	50.20	5.85	5.03	20.25	1.17
Máximo (%)	45.78	62.73	60.16	27.95	19.90	61.88	4.28

A figura 22 mostra a matriz de correlação entre as variáveis. Os teores totais AT e ST têm forte correlação positiva com os respectivos teores fracionários AA e SR, como esperado. As variáveis AT e FE tem forte correlação negativa.

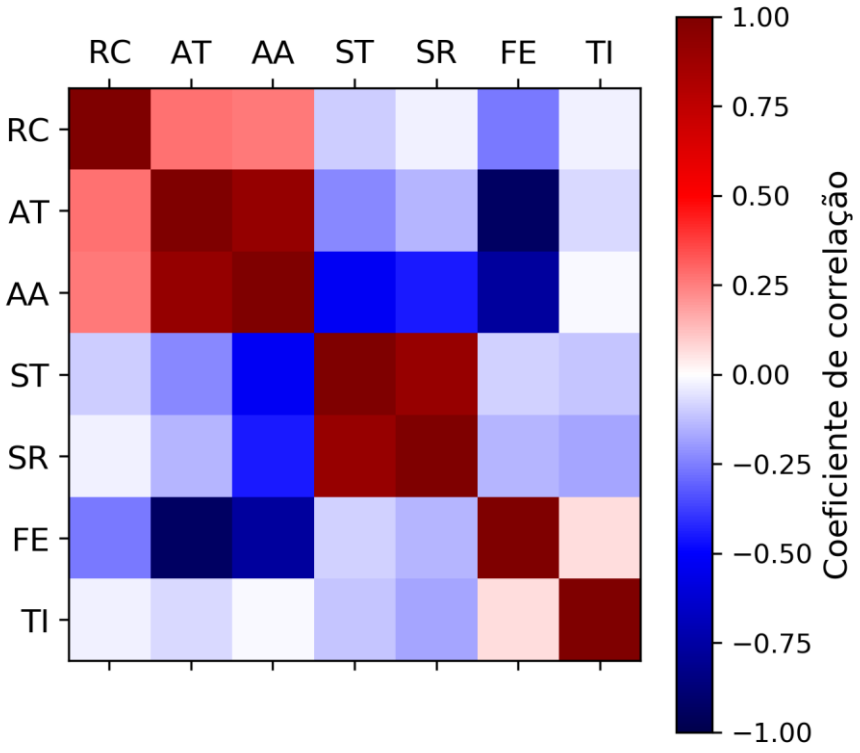

Figura 22: Matriz de correlação das variáveis.

O banco de dados contém três restrições: (1) a soma de AT, ST, FE e TI não pode ser maior do que 100%, (2) AA não pode ser maior do que AT e (3) SR não pode ser maior do que ST.

6.2 Comparação entre transformações multivariadas para dados multivariados com restrições de soma e fração

6.2.1 *Workflow*s

Foram testados quatro *workflows*. A figura 23 mostra o fluxograma desses *workflows*.

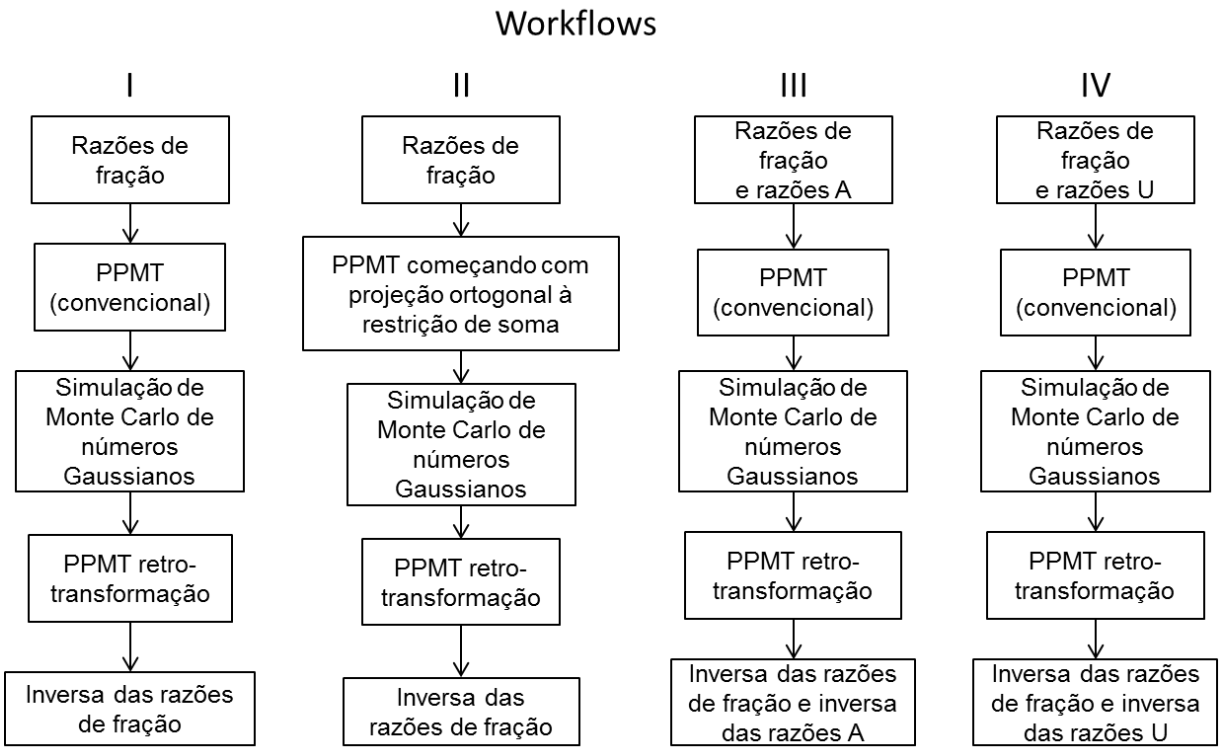

Figura 23: Workflows testados.

Todos os *workflows* usaram razões de fração. Nesse passo, os teores fracionários (AA e SR) foram substituídos pelas respectivas razões de fração (equação 32):

$$FR\_AA = \frac{AA}{AT};\ FR\_SR = \frac{SR}{ST} \tag{32}$$

No *workflow* III, as razões A foram aplicadas para as variáveis AT, ST, FE e TI, pois elas têm restrição de soma. Para obter as razões A, primeiro uma variável *filler* foi obtida (equação 33):

$$FILLER = 100 - (AT + ST + FE + TI) \tag{33}$$

Posteriormente, as razões A foram obtidas através da equação 34:

$$A1 = \frac{ST}{AT};\ A2 = \frac{FE}{AT};\ A3 = \frac{TI}{AT};\ A4 = \frac{FILLER}{AT} \tag{34}$$

A variável AT foi escolhida para estar no denominador porque ela tem a maior média. A intenção é evitar a ocorrência de valores próximos de zero no denominador. No *workflow* IV, as razões U foram aplicadas para as variáveis com restrição de soma (AT, ST, FE e TI). A equação 35 mostra o cálculo das razões U:

$$U1 = TI;\ U2 = \frac{ST}{100 - TI};\ U3 = \frac{FE}{100 - TI - STI};\ U4 = \frac{AT}{100 - TI - ST - FE} \tag{35}$$

A figura 24 mostra a correlação entre as variáveis originais usadas no numerador e as variáveis transformadas por razões A e razões U. As razões U mostraram maior correlação com as variáveis originais que estavam no numerador. Por exemplo, a correlação entre U3 e FE é aproximadamente 1 (figura 24d) enquanto que a correlação entre A2 e FE é 0.98 (figura 24c). Isso ocorreu porque as variáveis TI, ST e FE possuem valores muito menores do que 100. Como resultado, as divisões com as razões U preservaram melhor a correlação com as variáveis originais utilizadas no numerador.

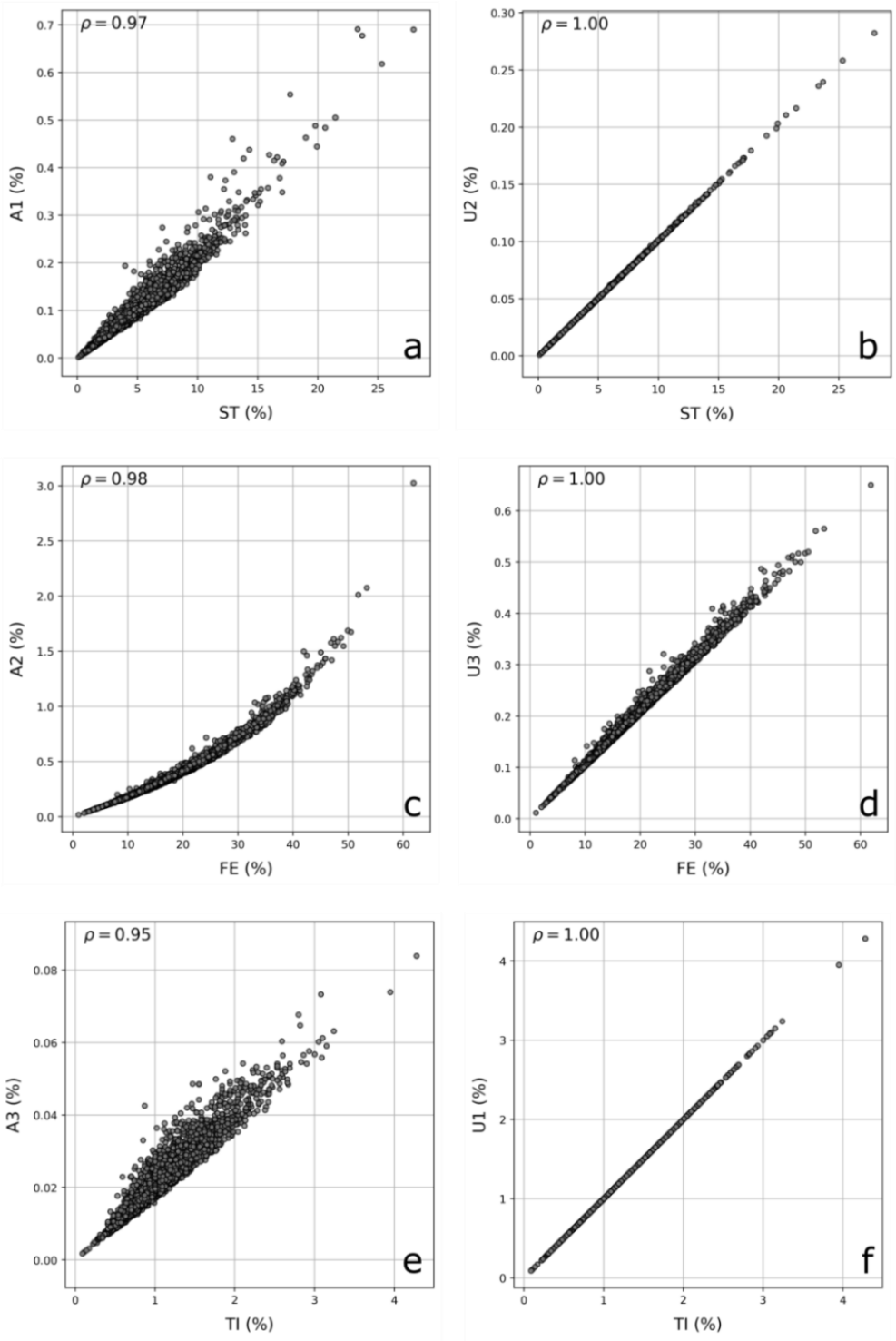

Figura 24: Gráfico de dispersão entre a variável original usada no numerador e a variável transformada: ST e A1 (a), ST e U2 (b), FE e A2 (c), FE e U3 (d), TI e A3 (e) e TI e U1 (f).

A tabela 8 resume as etapas de pré-processamento dos dados nos diversos *workflows*. A tabela 9 resume as variáveis utilizadas para a etapa de PPMT. Depois da

etapa de PPMT, as variáveis são independentes e seguem uma distribuição multi-Gaussiana. Como resultado, números Gaussianos aleatórios podem ser sorteados de maneira independente para cada variável. Para cada variável, um milhão de números Gaussianos foram simulados por Monte Carlo e retro-transformados.

Tabela 8: Etapas de pré-processamento dos dados.

Workflow	Transformação	Variáveis de entrada	Variáveis de saída
I, II, III, IV	Razões de fração	AA, SR	FR_AA, FR_SR
III	Razões A	AT, ST, FE, TI	A1, A2, A3, A4
IV	Razões U	AT, ST, FE, TI	U1, U2, U3, U4

Tabela 9: Variáveis utilizadas na etapa de PPMT.

Workflow	Transformação	Variáveis de entrada
I	PPMT (convencional)	RC, FR_AA, FR_SR, AT, ST, FE, TI
II	PPMT começando com projeção ortogonal à restrição de soma	RC, FR_AA, FR_SR, AT, ST, FE, TI
III	PPMT (convencional)	RC, FR_AA, FR_SR, A1, A2, A3, A4
IV	PPMT (convencional)	RC, FR_AA, FR_SR, U1, U2, U3, U4

6.2.2 Resultados

6.2.2.1 Reprodução dos histogramas

As sete variáveis combinadas com os quatro *workflows* resultam em 28 histogramas a serem verificados. Para tornar o capítulo sintético, apenas a reprodução do histograma da variável AT é mostrada na figura 25. A linha vermelha na figura 25 mostra o histograma dos dados enquanto que a linha preta representa o histograma da simulação. O histograma foi bem reproduzido para todos os *workflows* (veja as figuras 25a-d, a linha preta é coincidente com a linha vermelha).

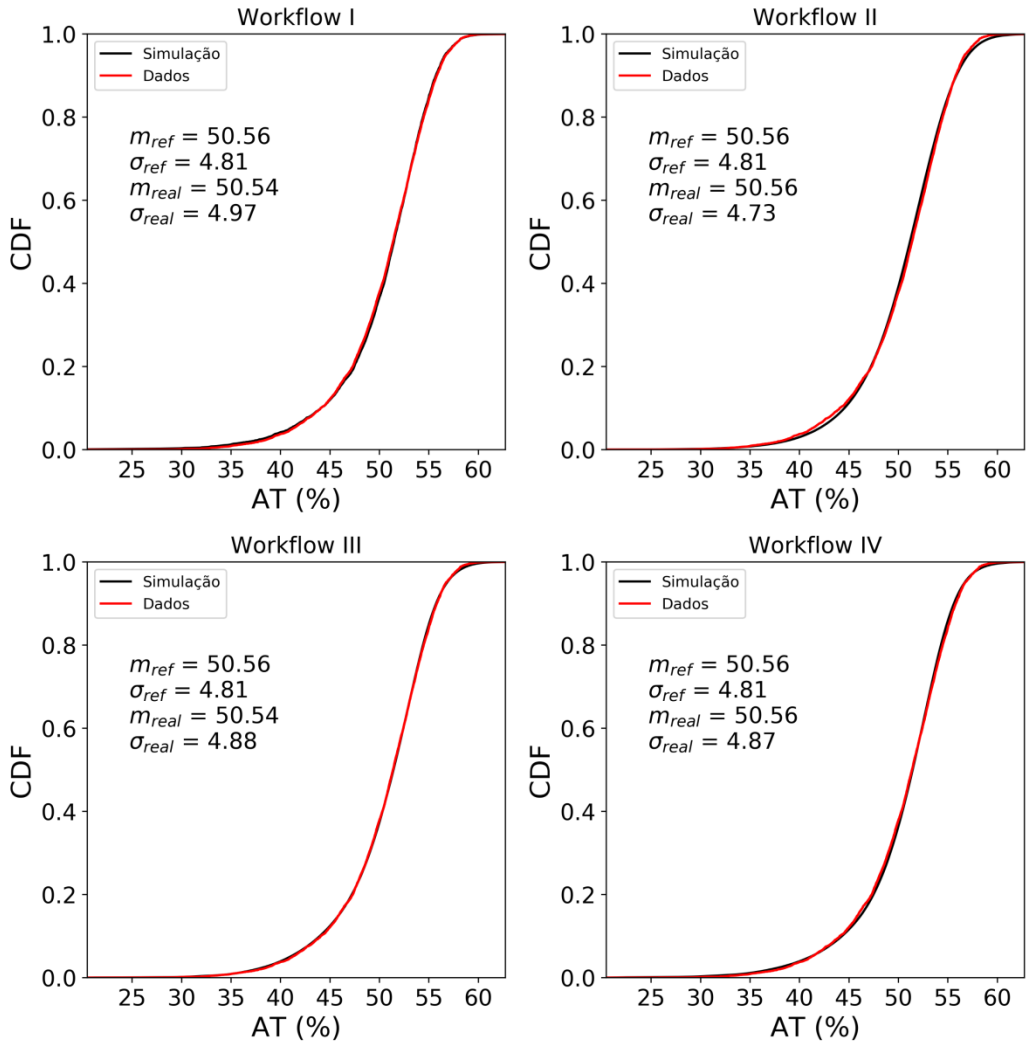

Figura 25: Reprodução do histograma da Alumina Total (AT) para os workflows testados.

6.2.2.2 Reprodução das relações bivariadas

A figura 26 mostra o gráfico de dispersão de AT e AA dos dados (figura 26a) e das simulações para os 4 *workflows* testados (figuras 26b-e). Todos os *workflows* reproduziram bem a relação bivariada entre essas duas variáveis. A restrição que AT deve ser maior do que AA foi respeitada para os *workflows* I, III e IV. Para o *workflow* II, tem um ponto que viola a restrição de fração (veja o círculo vermelho na figura 26c, o valor de AA está maior do que AT). Isso ocorreu, porque o *workflow* II não aplicou a transformação *normal score* diretamente na razão de fração de AA. Como resultado, a razão de fração simulada teve valores acima de 1 para o *workflow* II (houve extrapolação para a razão de fração). A consequência é que alguns valores violaram a restrição de fração. Para os *workflows* I, III e IV, a transformação *normal score* foi aplicada diretamente para a razão de fração de AA. A transformação *normal score* inversa evitou que os valores simulados da razão de fração de AA extrapolassem. Como consequência, os valores simulados para os *workflows* I, III e IV respeitaram a restrição de fração.

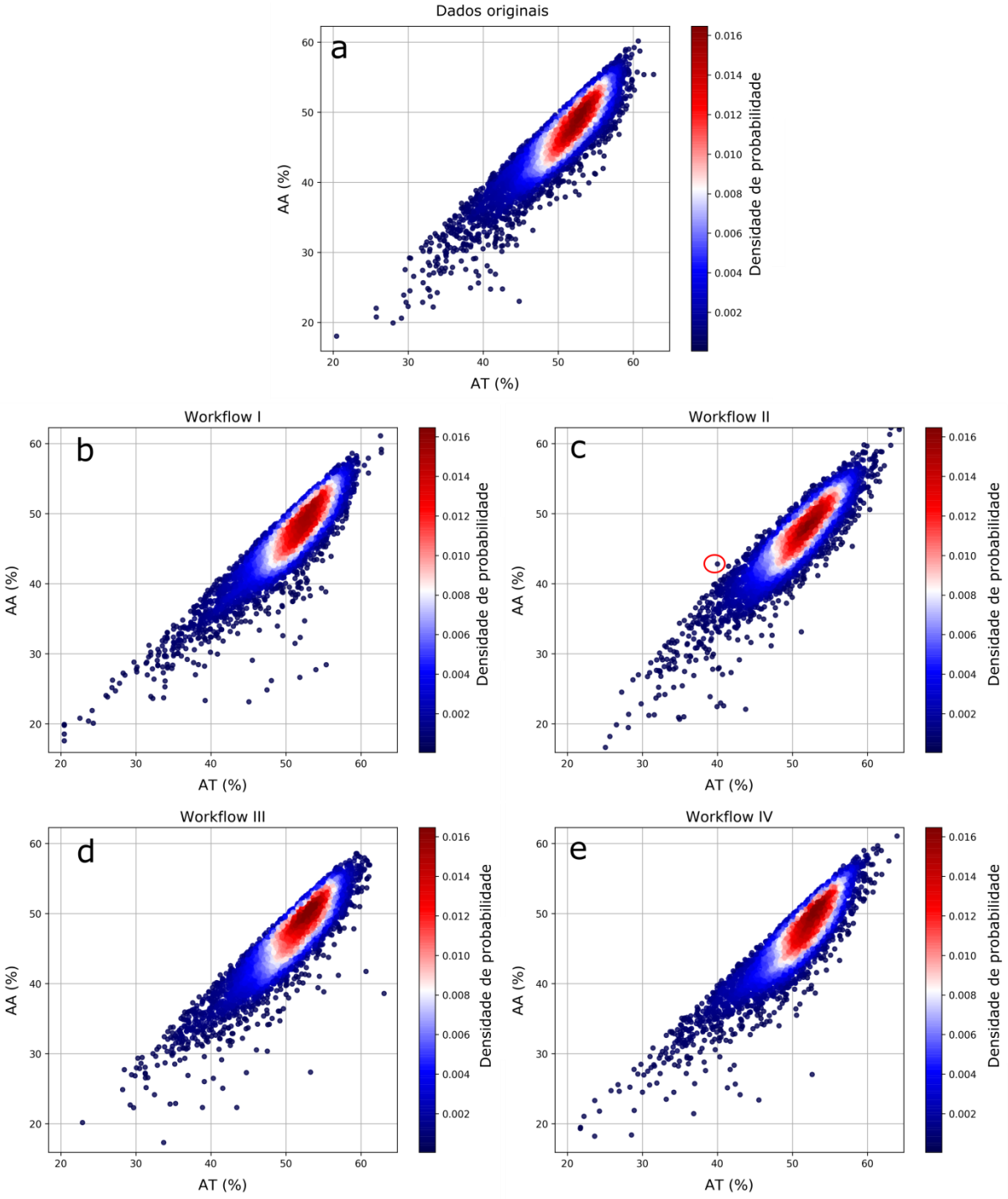

Figura 26: Gráfico de dispersão das variáveis AT e AA dos dados originais (a) e das simulações para os 4 workflows testados: workflow I (b), workflow II (c), workflow III (d) e workflow IV (e).

Considerando a soma de AT, ST, FE e TI, os dois atributos com as maiores médias são AT e FE. Nesse caso, é interessante checar a relação bivariada entre AT e FE. A figura 26 mostra o gráfico de dispersão entre AT e FE dos dados (figura 27a) e

das simulações para os 4 *workflows* testados (figuras 27b-e). A relação bivariada entre AT e FE foi bem reproduzida para os 4 *workflows*. O *workflow* I (figura 27b) teve maior espalhamento na região com AT menor do que 30% e FE maior do que 50%.

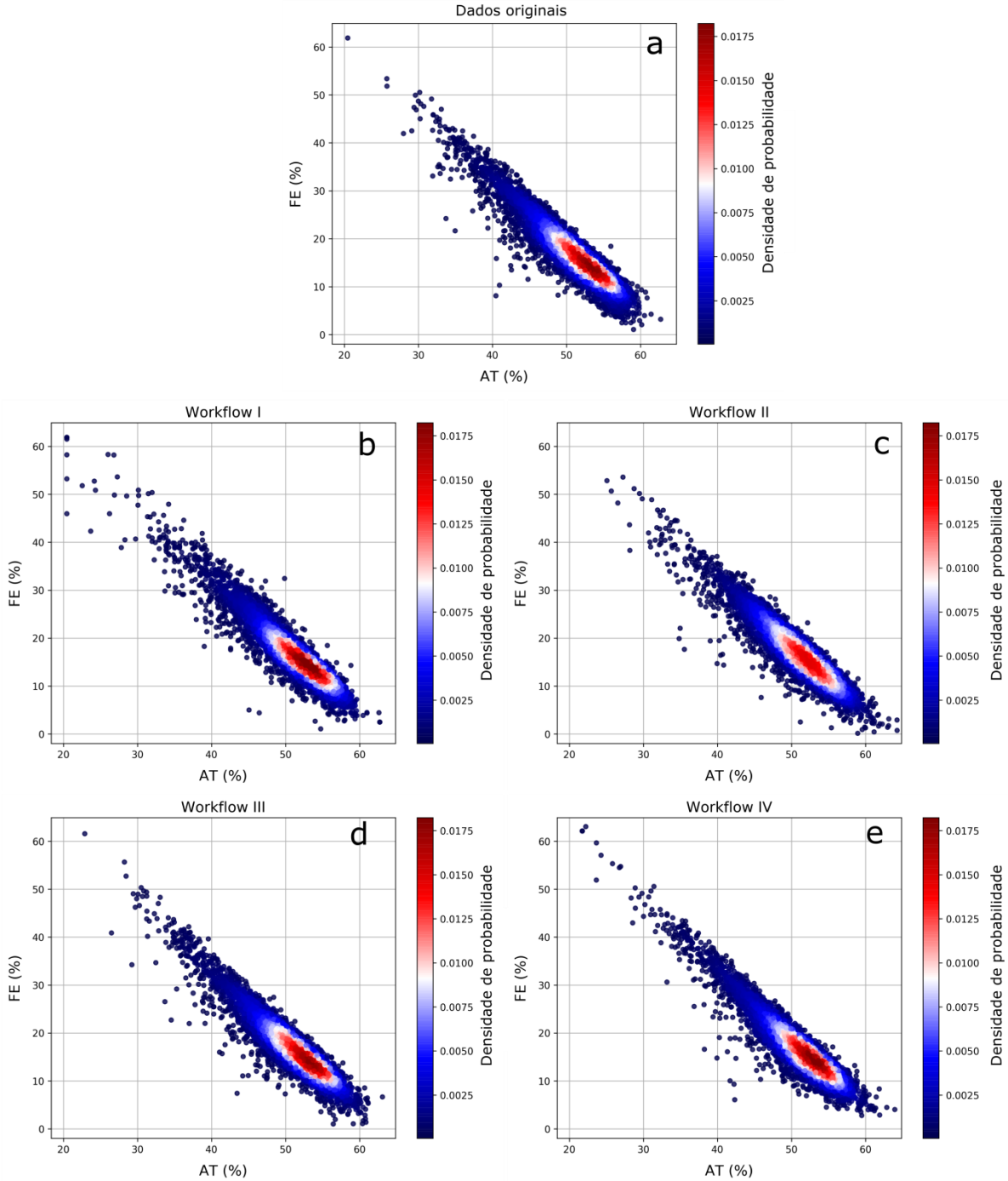

Figura 27: Gráfico de dispersão das variáveis AT e FE dos dados (a) e das simulações para os 4 workflows testados: workflow I (b), workflow II (c), workflow III (d) e workflow IV (e).

As figuras 28a-d mostram os gráficos de dispersão entre os coeficientes de correlação dos valores simulados e dos dados para os *workflows* I-IV. Todos os *workflows* reproduziram bem os coeficientes de correlação (veja as figuras 28a-d, os pontos estão próximos da linha y = x).

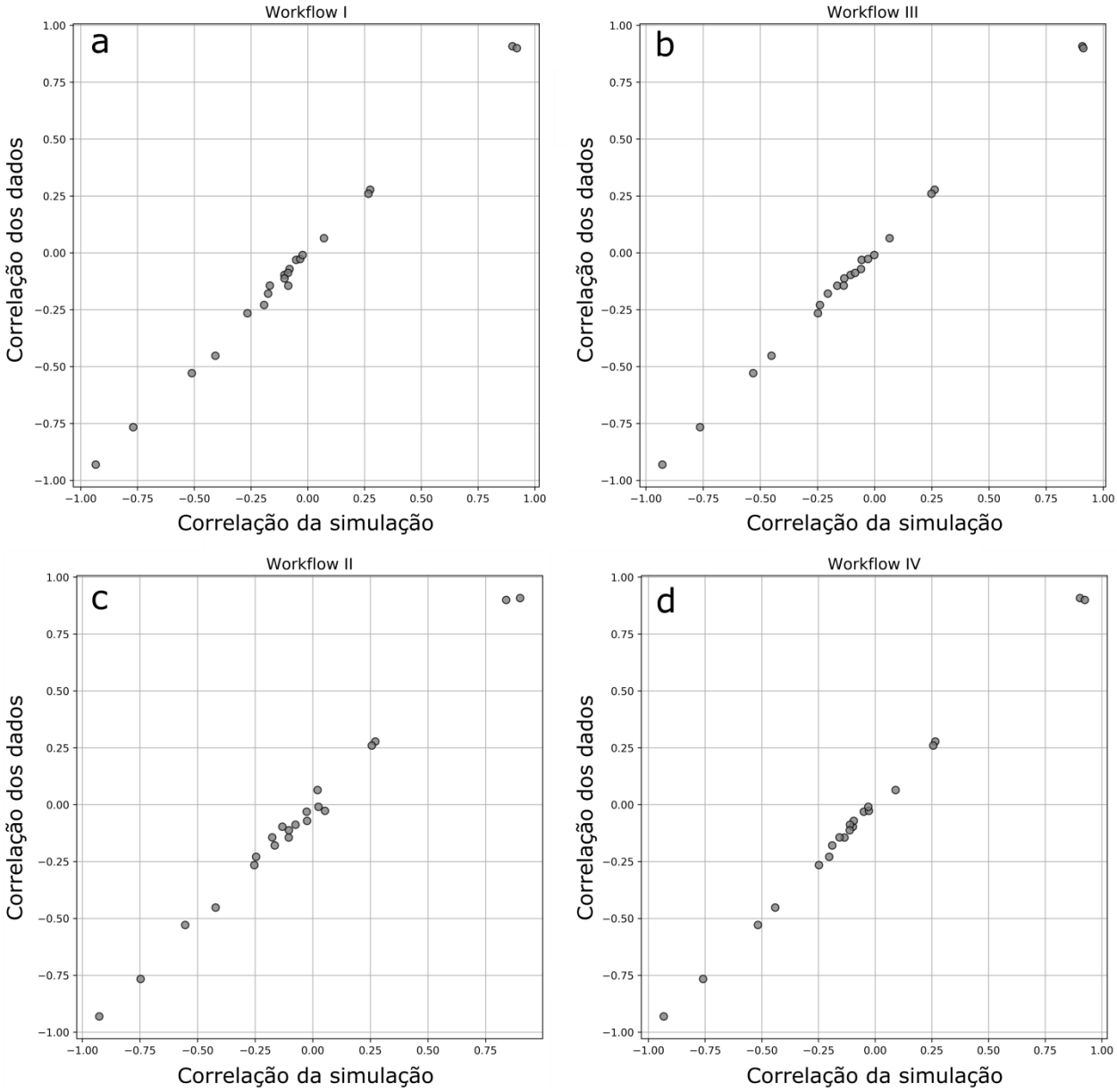

Figura 28: Reprodução dos coeficientes de correlação para os 4 workflows: workflow I (a), workflow II (b), workflow III (c), e workflow IV (d).

6.2.2.3 Verificação da restrição de soma

A figura 29 mostra o máximo e o percentil 99.9 da soma das variáveis AT, ST, FE e TI das simulações para os 4 *workflows*. O *workflow* I produziu uma pequena porcentagem de valores simulados cuja soma foi superior a 100%. O percentil 99.9 para o *workflow* I foi menor do que 90% (veja a figura 29). Isso indica que menos do que 0.1% dos valores simulados mostraram soma de AT, ST, FE e TI acima de 100%. A soma das variáveis no banco de dados tem uma média de aproximadamente 80%. Se a soma das variáveis AT, ST, FE e TI no banco dados fosse mais próxima a 100%, os valores simulados para o *workflow* I provavelmente mostrariam maior porcentagem de valores com soma acima de 100%. Por outro lado, os *workflows* II, III e IV não produziram valores simulados com soma de AT, ST, FE e TI acima de 100%.

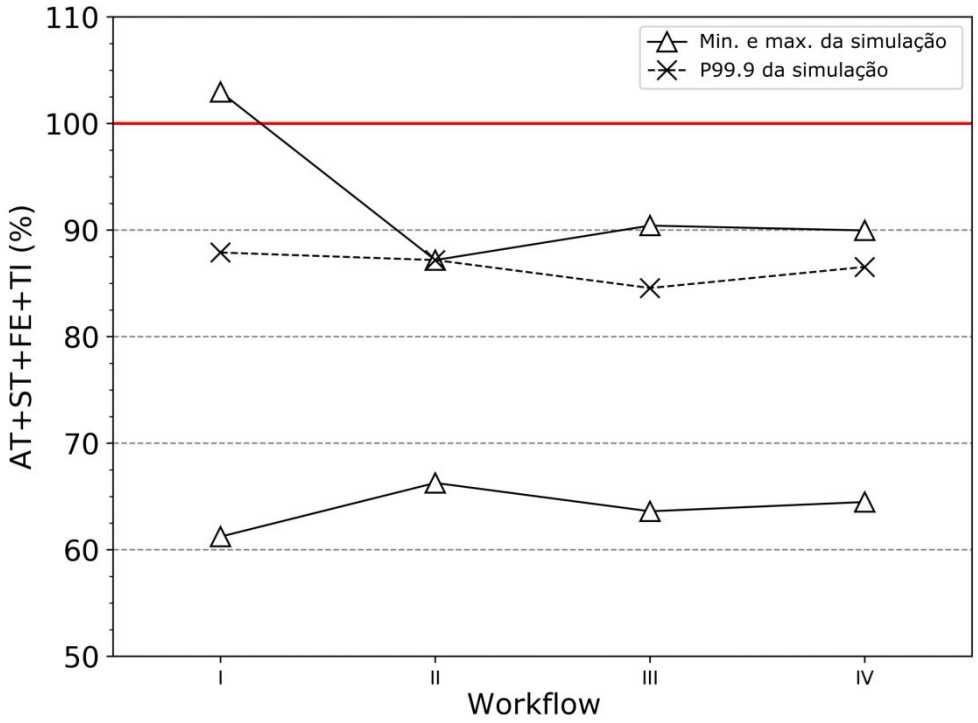

Figura 29: Verificação da soma das variáveis AT, ST, FE e TI para os workflows testados.

6.2.2.4 Verificação de extrapolação

A figura 30 mostra o mínimo e máximo dos dados e das simulações para todos os *workflows* considerados. Os teores fracionários AA e SR tiveram valores simulados fora do alcance dos dados para todos os *workflows*. Isso ocorreu porque as variáveis AA e SR não foram simuladas diretamente. Foram simuladas as razões de fração de AA e SR. Depois que a razão de fração foi simulada, a multiplicação da razão de fração simulada pelo teor total simulado produziu valores fora do intervalo definido pelo mínimo e máximo dos dados.

Em relação às variáveis AT, ST, FE e TI, o *workflow* I não gerou valores simulados além do intervalo dos dados. Isso ocorreu porque *workflow* I é o único *workflow* em que a transformação *normal score* foi aplicada diretamente para as variáveis originais. A transformação *normal score* inversa assegura que os valores retro-transformados estejam entre o mínimo e máximo dos dados. Para o *workflow* II, a transformação *normal score* foi aplicada para a soma das variáveis AT, ST, FE e TI. Para os *workflows* III e IV, a transformação *normal score* foi aplicada para as razões A e razões U, respectivamente.

O *workflow* II mostrou os piores resultados em relação à extrapolação. O *workflow* II foi o único *workflow* que resultou em valores negativos (figura 29). Em comparação com os *workflows* I, III e IV, o *workflow* II possui mínimo e máximo de valores simulados que mais diferem do mínimo e máximo dos dados (veja a figura 29). Além disso, o uso de um vetor ortogonal à restrição de soma afetou todas as variáveis. A variável RC não está considerada na restrição de soma, mas os valores simulados de RC pelo *workflow* II mostraram valores além do mínimo e máximo dos dados.

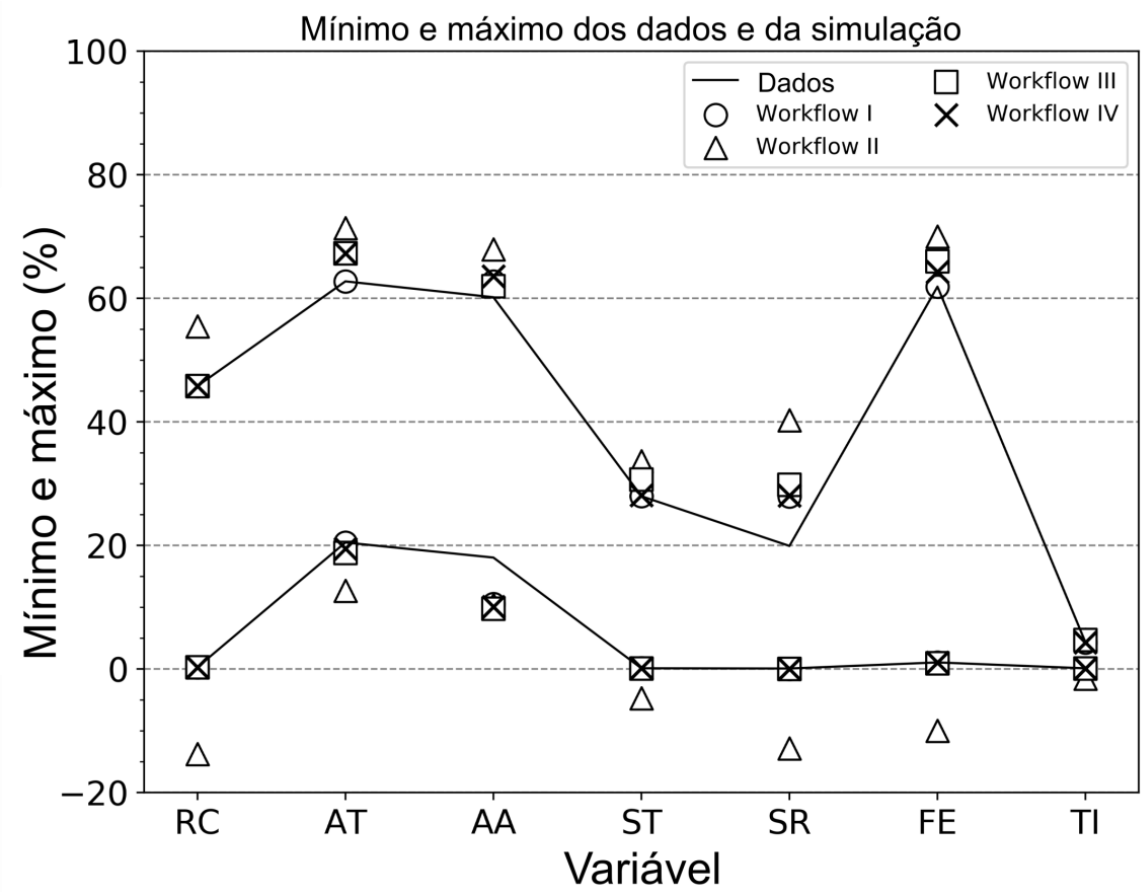

Figura 30: Mínimo e máximo dos dados e da simulação.

A figura 31 mostra a porcentagem de pontos que estão fora do intervalo dos dados para os 4 *workflows* testados. O *workflow* II teve a maior porcentagem de valores fora do intervalo dos dados para todas as variáveis.

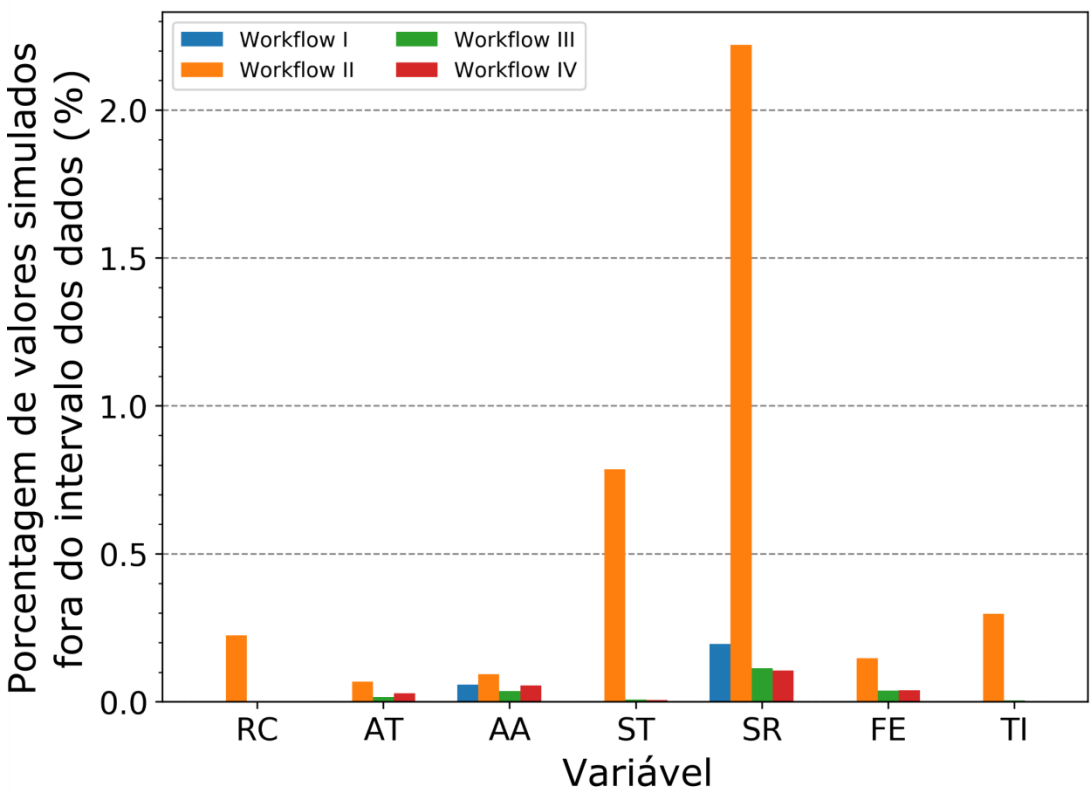

Figura 31: Porcentagem de valores simulados fora do intervalo dos dados.

A comparação dos métodos foi feita para testar as transformações. Números Gaussianos aleatórios foram sorteados e retro-transformados. Em vista dos resultados obtidos, o *workflow* I foi escolhido para fazer a simulação geoestatística das variáveis, que é apresentada na seção 6.3.

A proporção de valores simulados que violou a restrição de soma foi pequena no *workflow* I. Além disso, o *workflow* I utiliza uma transformação a menos nas variáveis originais do que os *workflow*s III e IV, que utilizam razões A e razões U. A ideia é diminuir o número de transformações para preservar melhor a estrutura espacial das variáveis. As razões A e U causam uma mistura das variáveis, que afeta a continuidade espacial.

6.3 Simulação geoestatística

6.3.1 Descrição espacial

A figura 32 mostra o mapa de localização dos dados. A amostragem foi feita de maneira irregular, com diferentes espaçamentos amostrais. O maior espaçamento amostral é de aproximadamente 200 x 200 m enquanto que o menor espaçamento amostral é de 25 x 25 m nas direções X e Y. Além disso, a área possui espaçamentos amostrais intermediários de 100 x 100 m e 50 x 50 m nas direções X e Y. As amostras são provenientes de furos de sondagem verticais e o comprimento das amostras é de 0.50 metros.

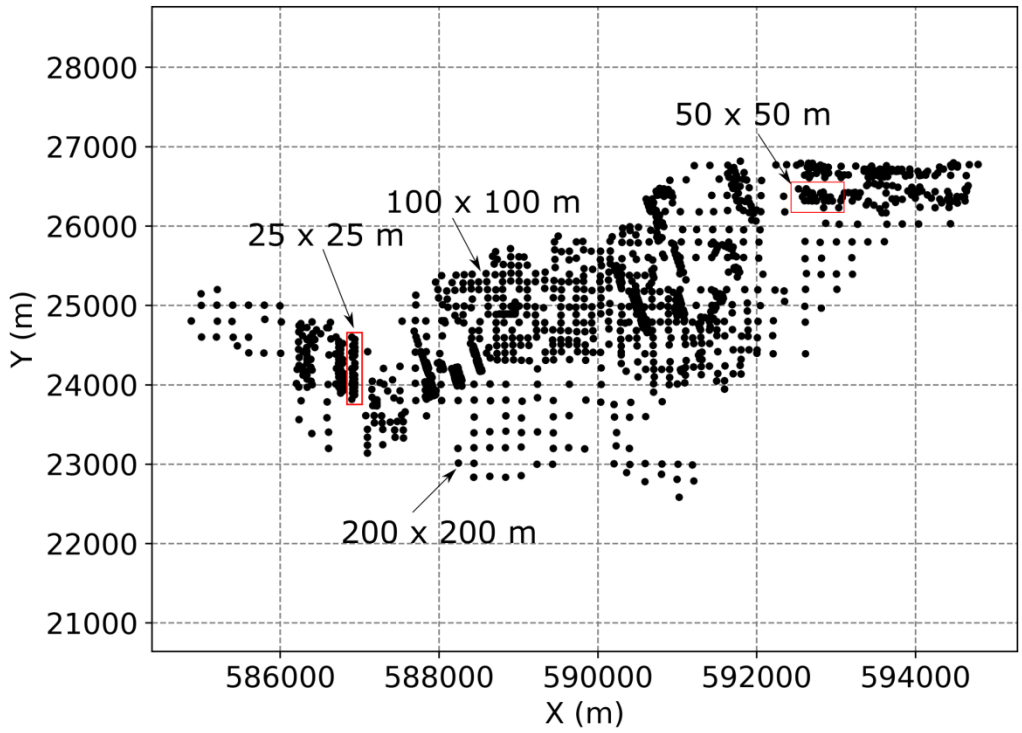

Figura 32: Mapa de localização das amostras com diferentes espaçamentos amostrais.

As coordenadas Z foram transformadas para coordenadas estratigráficas. A equação 36 define as coordenadas estratigráficas:

$$z_{estrat} = z_{original} - z_{topo} \quad (36)$$

onde z_{estrat} corresponde à coordenada Z estratigráfica, $z_{original}$ corresponde à coordenada Z original e z_{topo} corresponde à coordenada Z do topo da camada de minério. A modelagem dos variogramas e as simulações geoestatísticas foram feitas utilizando as coordenadas estratigráficas z_{estrat}.

6.3.2 Desagrupamento

O desagrupamento foi feito em duas dimensões utilizando o método das células móveis (Deutsch e Journel, 1998). Pyrcz e Deutsch (2014) recomendam fazer o desagrupamento por células móveis em duas dimensões quando os furos são verticais. Foram testados diversos tamanhos de célula e foi escolhido o tamanho que diminuísse a média de AT. A figura 33 mostra o gráfico de dispersão entre o tamanho de célula e a média desagrupada de AT. A média desagrupada de AT praticamente estabiliza a partir do tamanho de célula de 200 metros. O tamanho de célula escolhido foi de 200 m nas direções X e Y e coincide aproximadamente com o maior espaçamento amostral.

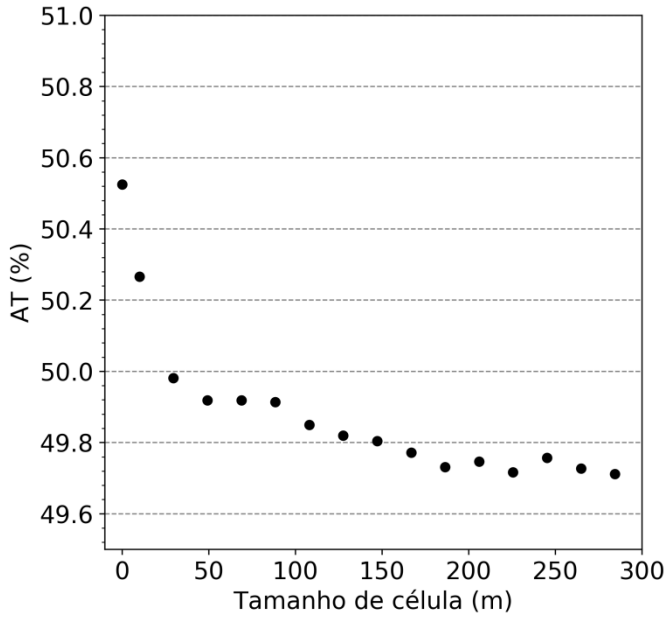

Figura 33: Relação entre a média desagrupada de AT e o tamanho de célula.

6.3.3 Metodologia

Os resultados na seção 6.4 mostraram que o *workflow* I teve uma pequena porcentagem de valores simulados que não respeitaram a restrição de soma. Além disso, os teores fracionários AA e SR tiveram problemas de extrapolação. Em vista disso, foram incluídas duas etapas de correção no *workflow* I para a simulação geoestatística de teores: (1) correção de soma e (2) correção de extrapolação. A figura 34 mostra o fluxograma da metodologia utilizada.

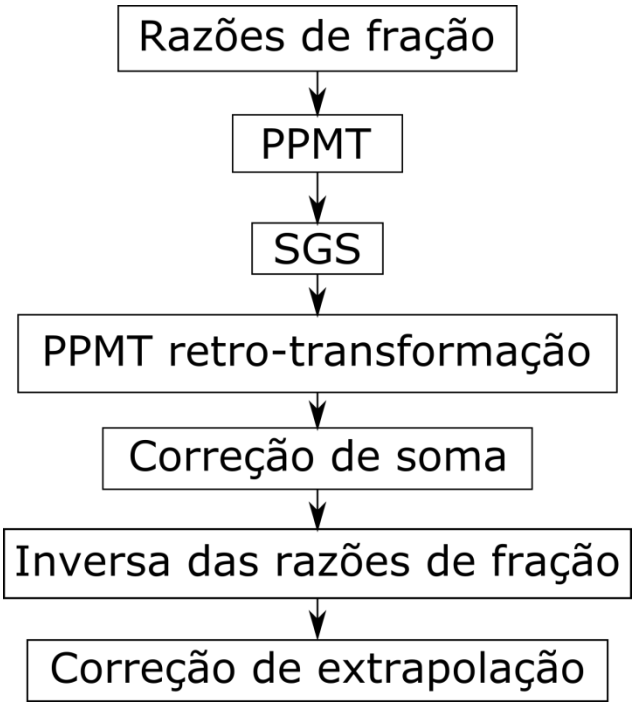

Figura 34: Fluxograma da metodologia.

Primeiros os teores fracionários AA e SR foram transformados para as respectivas razões de fração FR_AA e FR_SR. Segundo as variáveis foram transformadas usando PPMT. A transformação PPMT foi feita utilizando pesos de desagrupamento. As variáveis transformadas PPMT seguem uma distribuição multi-Gaussiana e são independentes. A figura 35 mostra os histogramas e gráficos de dispersão das primeiras duas variáveis PPMT. As duas variáveis seguem uma distribuição Gaussiana padrão, com média zero e desvio padrão um e são independentes. As variáveis transformadas PPMT foram simuladas de maneira independente utilizando simulação sequencial Gaussiana (*sequential Gaussian simulation* – SGS). Vinte realizações foram realizadas. Quarenta amostras foram utilizadas para a construção do sistema de krigagem na SGS.

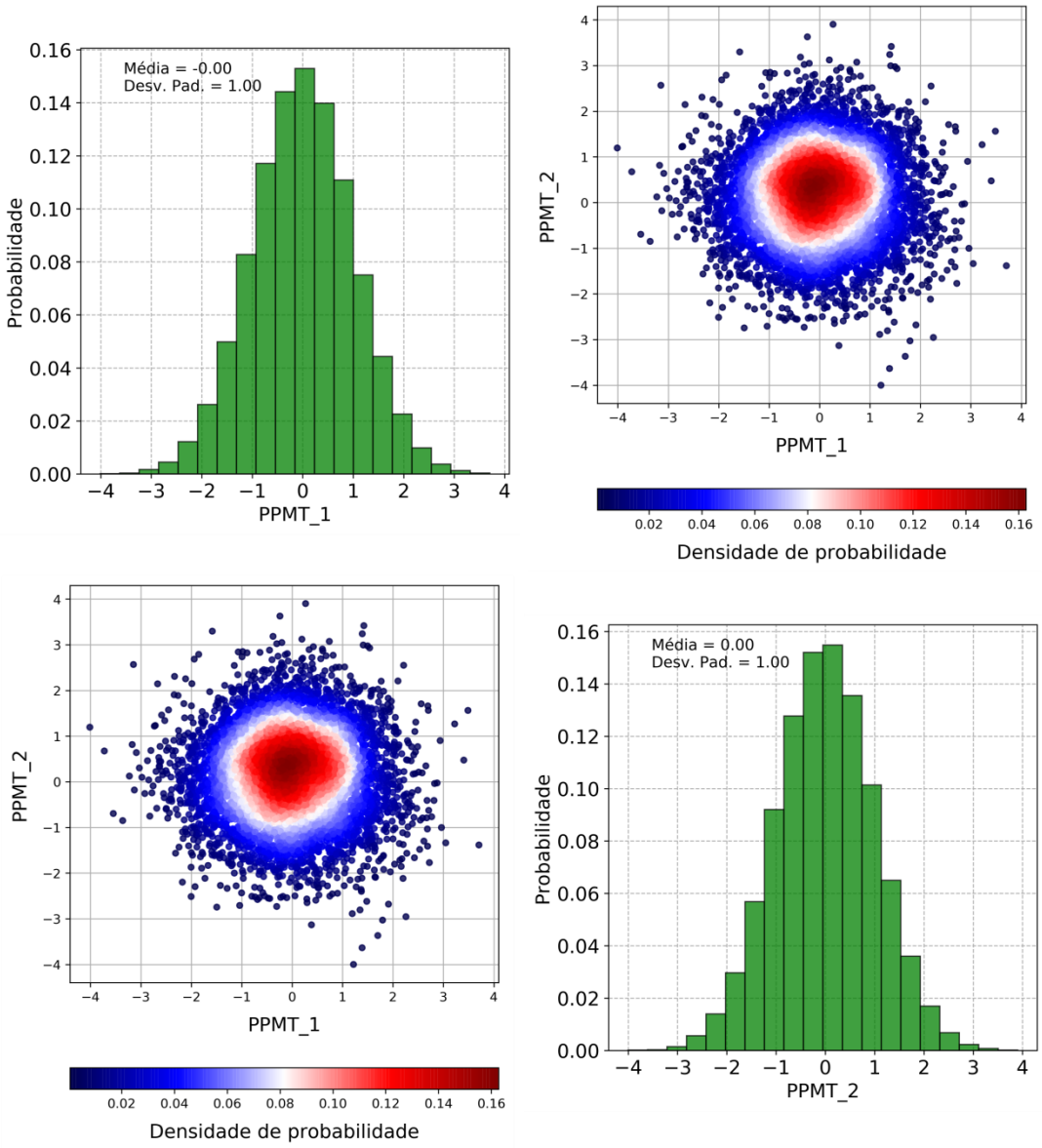

Figura 35: Histogramas e gráficos de dispersão das duas primeiras variáveis transformadas PPMT.

Os teores simulados são transformados para o espaço dos dados originais utilizando a PPMT retro-transformação. O resultado da PPMT retro-transformação são valores simulados dos teores totais (AT, ST, FE e TI), de RC e das razões de fração (FR_AA e FR_SR). A correção de soma corrige os teores totais simulados (AT, ST, FE e TI) e foi feita usando o software *mvs_sum_check*, cuja documentação está no apêndice A. Os teores totais corrigidos são chamados de AT', ST', FE' e TI'. A inversa das razões de fração utiliza as razões de fração FR_AA e FR_SR e os respectivos teores totais corrigidos AT' e ST'. A inversa das razões de fração resulta nos teores

simulados de AA e SR. Por último, é feita uma correção de extrapolação dos teores fracionários AA e SR. A inversa das razões de fração e a correção de extrapolação são feitas utilizando o software *mvs_frac_ratio*, cuja documentação está no apêndice A.

6.3.4 Análise da continuidade espacial

Os variogramas modelados são dos valores *normal score* das variáveis. A transformação *normal score* foi aplicada nas variáveis originais, exceto para os teores fracionários AA e SR. Nesse caso, a transformação *normal score* foi aplicada para as razões de fração FR_AA e FR_SR, respectivamente. A transformação *normal score* para o cálculo dos variogramas foi feita sem utilizar pesos de desagrupamento. Embora a simulação seja feita com as variáveis transformadas PPMT, Barnett *et al.* (2016) recomendam utilizar o variograma dos *normal scores* para melhorar a reprodução do variograma das realizações. Os variogramas experimentais foram calculados utilizando as coordenadas estratigráficas. A tabela 10 mostra o modelo dos variogramas das 7 variáveis *normal score*. A direção de maior continuidade é no plano horizontal e a direção de menor continuidade é na direção vertical.

Tabela 10: Modelos de variograma.

Variável	Estrutura	Variância	Alcance NS (m)	Alcance EW (m)	Alcance vert. (m)
NS_RC	Efeito pepita	0.00	x	x	x
	Esférico	0.60	65	65	2.60
	Esférico	0.30	430	430	3.50
	Esférico	0.10	10000	10000	3.70
NS_AT	Efeito pepita	0.20	x	x	x
	Esférico	0.52	65	65	2.40
	Esférico	0.28	2000	2000	2.50
NS_FR_AA	Efeito pepita	0.00	x	x	x
	Esférico	0.70	65	65	2.20
	Esférico	0.19	800	800	7.70
	Esférico	0.11	10000	10000	8.00
NS_ST	Efeito pepita	0.00	x	x	x
	Esférico	0.57	60	60	2.65
	Exponencial	0.43	1900	1900	2.70
NS_FR_SR	Efeito pepita	0.05	x	x	x
	Esférico	0.54	70	70	3.50
	Esférico	0.30	1300	1300	3.80
	Esférico	0.11	2000	2000	3.90
NS_FE	Efeito pepita	0.04	x	x	x
	Esférico	0.66	65	65	1.75
	Esférico	0.30	4000	4000	2.20
NS_TI	Efeito pepita	0.00	x	x	x
	Esférico	0.47	90	90	3.00
	Esférico	0.33	1000	1000	3.10
	Esférico	0.20	5000	5000	3.20

6.3.5 Resultados

6.3.5.1 Reprodução dos histogramas

A figura 36 mostra a reprodução dos histogramas das sete variáveis. Os histogramas das variáveis originais foram em geral bem reproduzidos. As maiores discrepâncias ocorreram para as variáveis SR e TI.

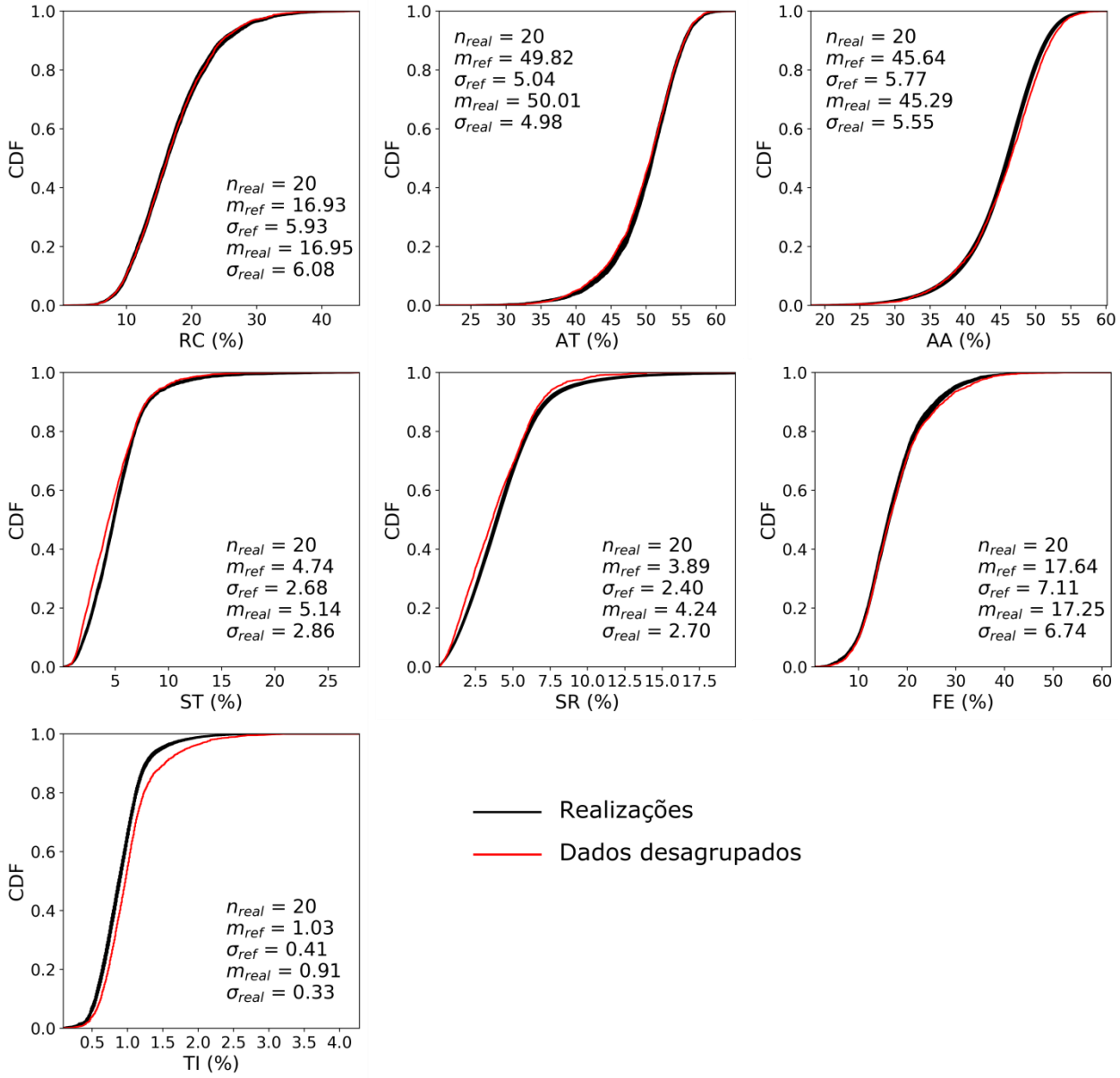

Figura 36: Reprodução dos histogramas.

6.3.5.2 Reprodução dos variogramas

A figura 37 mostra a reprodução do variograma na direção horizontal. Os variogramas estão estandardizados, com patamar igual a um. A reprodução do variograma na direção horizontal foi boa na estrutura de curto alcance. Na estrutura de longo alcance, o variograma das simulações ficou menos contínuo do que o variograma dos dados para as variáveis ST e SR. No caso da variável RC, a estrutura de longo alcance das realizações ficou mais contínua que a estrutura de longo alcance dos dados. Esses resultados estão de acordo com aqueles apresentados por Safikhani *et al.* (2017) e Paravarzar *et al.* (2015). Safikhani *et al.* (2017) compararam a reprodução do variograma da SGS em função dos parâmetros de busca. Safikhani *et al.* (2017) recomendam utilizar pelo menos 50 amostras. Em alguns casos, 200 amostras são necessárias para ter uma boa reprodução do variograma (Safikhani *et al.*, 2017). As realizações feitas nesse estudo utilizaram um máximo de 40 amostras. Não foi utilizado um maior número de amostras porque isso iria aumentar o tempo de processamento. Paravarzar *et al.* (2015) identificaram problemas na reprodução do variograma na SGS devido ao uso de uma vizinhança de busca muito restrita para a construção das distribuições condicionais locais.

A figura 38 mostra a reprodução do variograma na direção vertical. Os variogramas das variáveis AT, AA e FE foram bem reproduzidos na distância até 1.5 metros na direção vertical. Para as variáveis RC, ST, SR e TI, o variograma das simulações ficou mais descontínuo do que o variograma dos dados originais. Isso aconteceu devido à vizinhança de busca restrita e também porque a transformação PPMT descorrelaciona as variáveis para h = 0. Essa descorrelação causa uma desestruturação das variáveis a pequenas distâncias. O resultado é uma má reprodução do variograma na direção vertical. Esses resultados estão de acordo com os resultados obtidos por Barnett (2015). A má reprodução dos variogramas das simulações na direção vertical é observada também quando a transformação MAF é utilizada na simulação (Desbarats e Dimitrakipoulos, 2000).

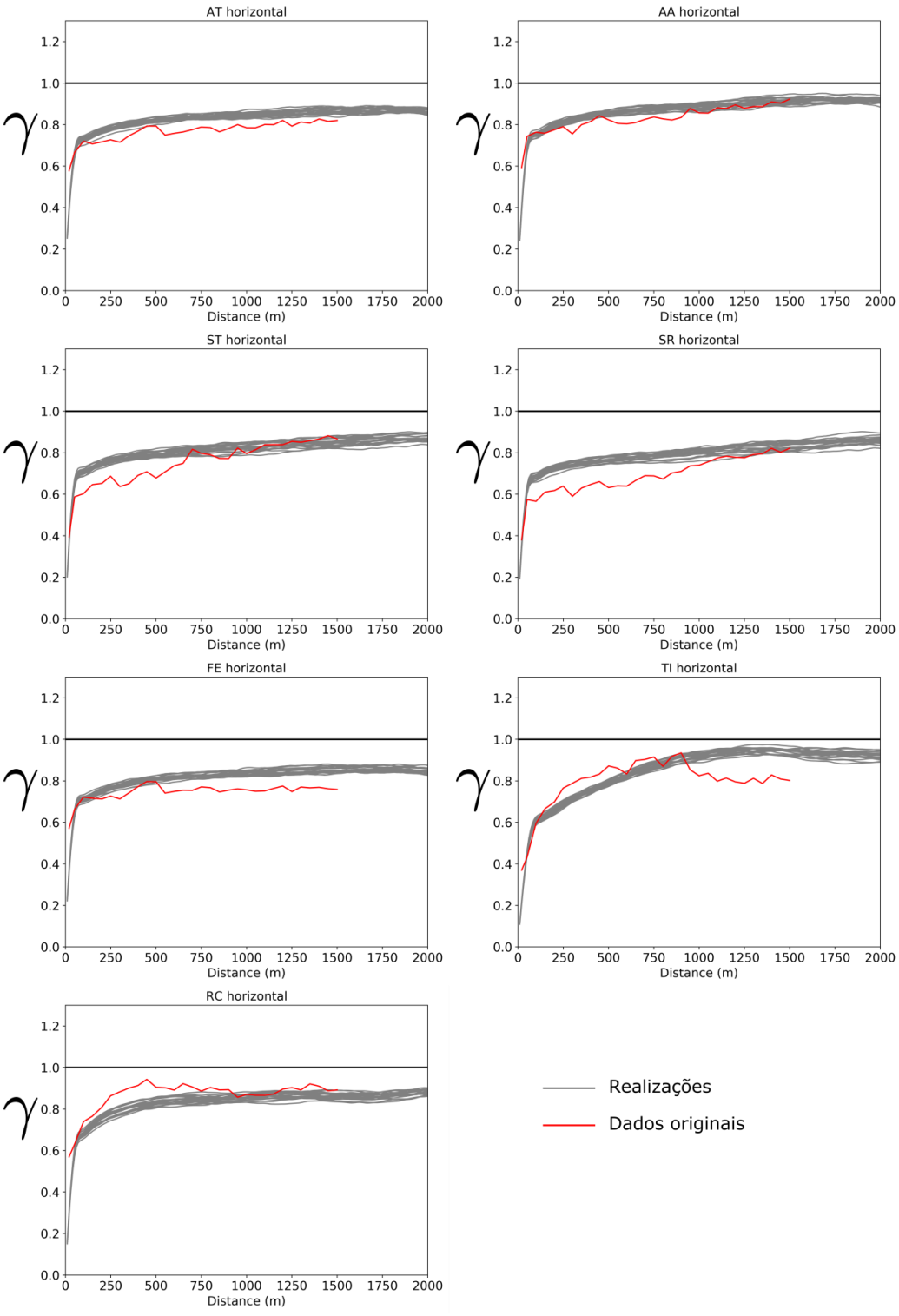

Figura 37: Reprodução dos variogramas na direção horizontal.

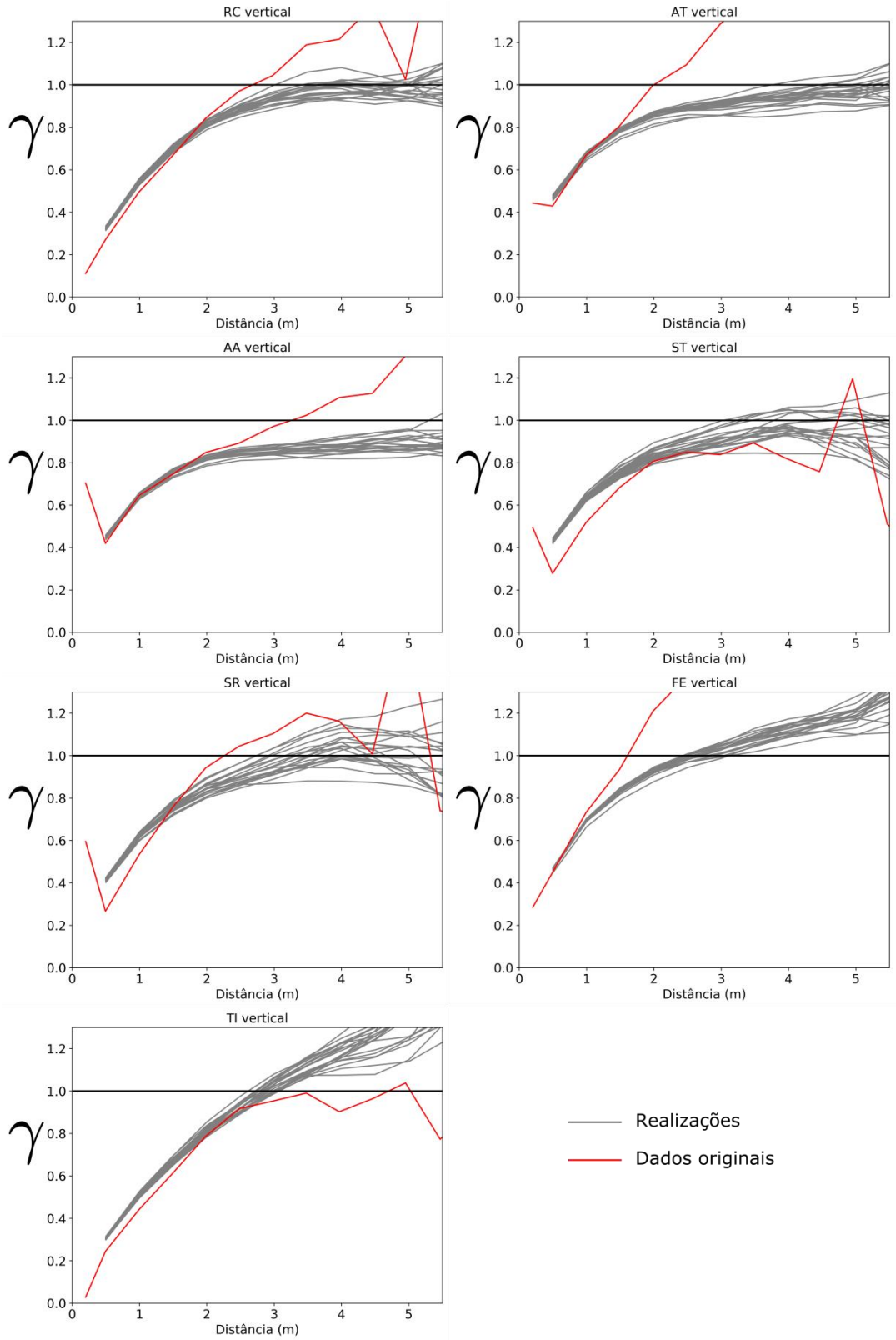

Figura 38: Reprodução dos variogramas na direção vertical.

6.3.5.3 Reprodução das relações bivariadas

A figura 39 mostra o gráfico de dispersão entre os coeficientes de correlação obtidos da primeira realização com os coeficientes de correlação dos dados. Os coeficientes de correlação foram reproduzidos, pois todos os pontos estão próximos da primeira bissetriz.

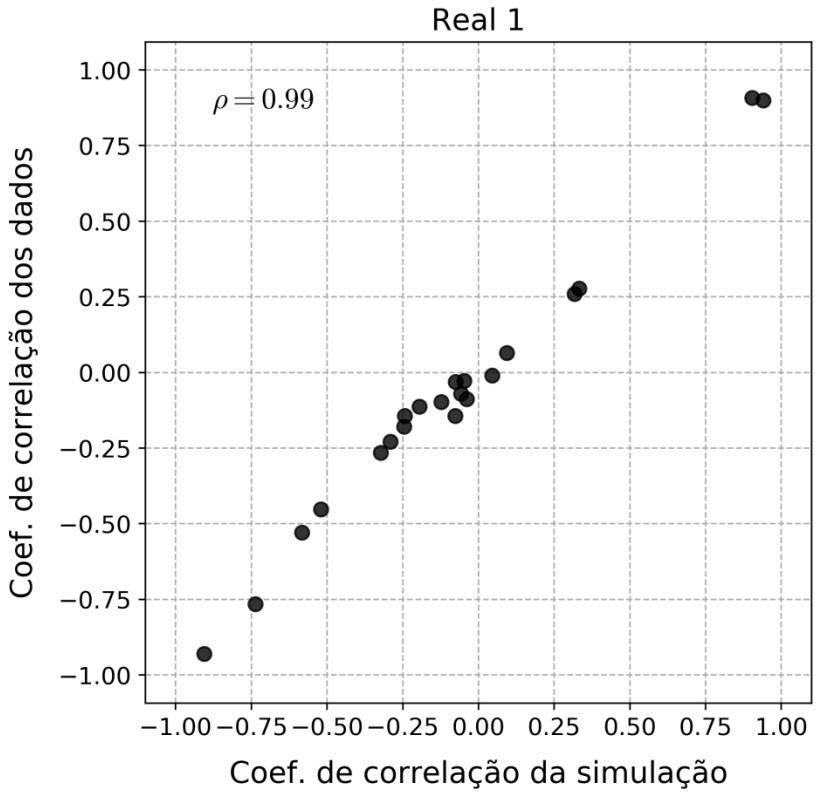

Figura 39: Gráfico de dispersão entre os coeficientes de correlação dos dados e da primeira realização.

Os gráficos de dispersão dos dados e os gráficos de dispersão da primeira realização foram comparados. As sete variáveis simuladas resultam um total de 21 gráficos de dispersão. Para tornar o capítulo sintético, apenas os gráficos de dispersão entre AT e FE (figuras 40a-b) e entre RC e ST (figuras 40c-d) são mostrados. AT e FE tem uma forte correlação negativa enquanto que RC e ST tem baixa correlação. Os gráficos de dispersão das simulações estão de acordo com os gráficos de dispersão

dos dados. Isso mostra que as relações bivariadas entre as variáveis foram reproduzidas.

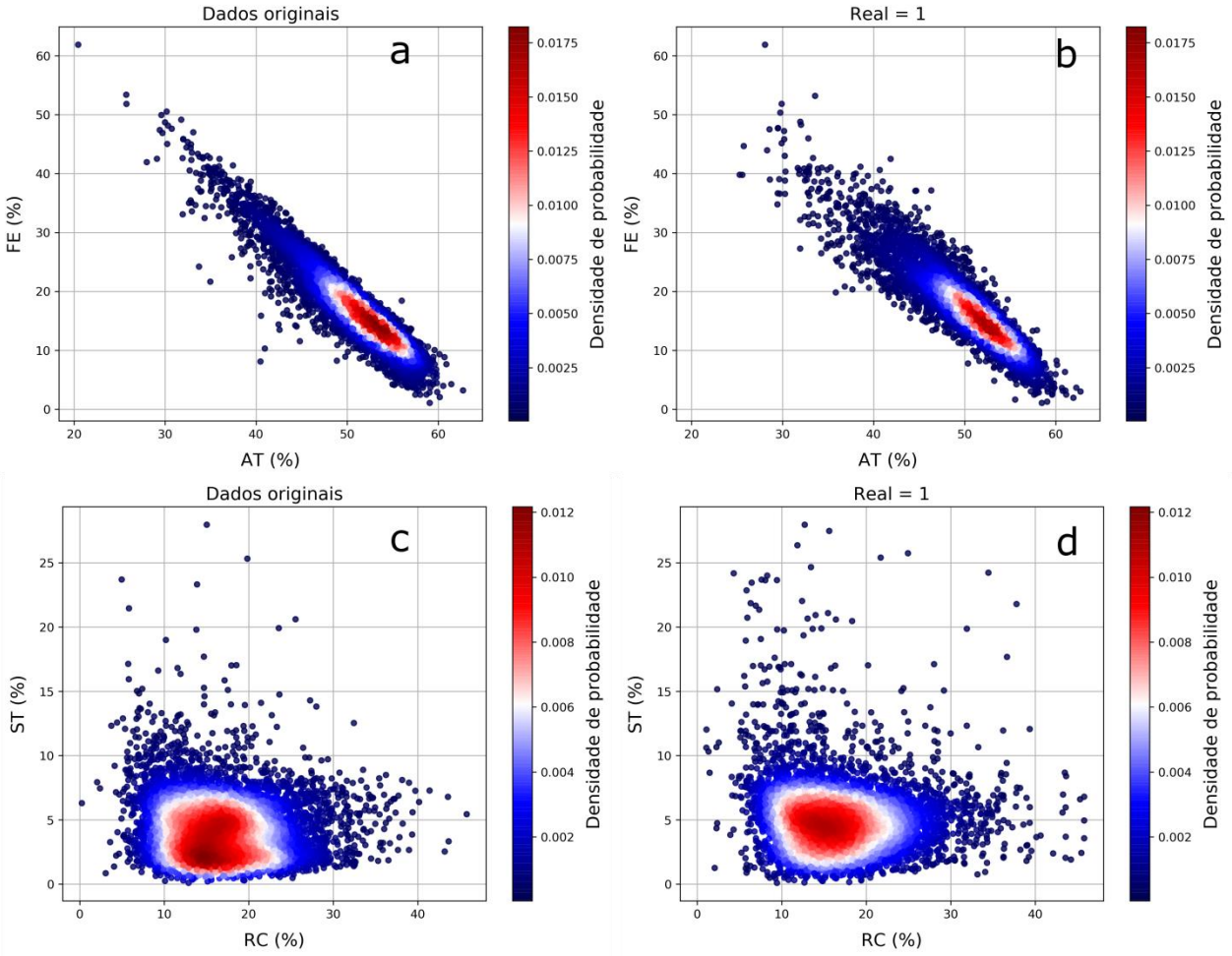

Figura 40: Gráfico de dispersão entre AT e FE dos dados originais (a) e da primeira realização (b). Gráfico de dispersão entre RC e ST dos dados originais (c) e da primeira realização (d).

6.3.5.4 Verificação da restrição de soma

A figura 41a mostra o histograma da soma das variáveis AT, ST, FE e TI para a primeira realização. O máximo da soma é igual a 100. Isso indica que nenhum nó simulado teve soma maior do que 100. A figura 41b mostra a proporção de nós de grid corrigidos na correção de soma para as 20 realizações. A proporção de valores corrigidos foi baixa, com média de 0.002%.

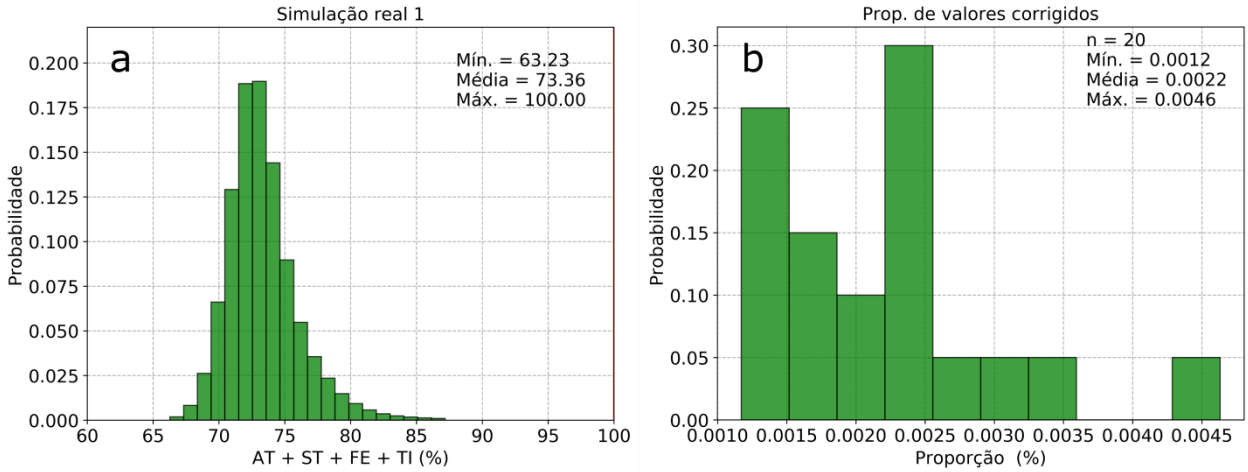

Figura 41: Histograma das soma das variáveis AT, ST, FE e TI da primeira realização (a) e histograma das proporções de valores corrigidos para as 20 realizações (b).

6.3.5.5 Verificação das restrições de fração

A figura 42a mostra o gráfico de dispersão entre AT e AA para a primeira realização. A variável AT está sempre acima de AA (região azul na figura 42a). A figura 41b mostra o gráfico de dispersão entre ST e SR. ST está sempre acima de SR (região azul na figura 42b). Isso mostra que as duas restrições de fração foram respeitadas.

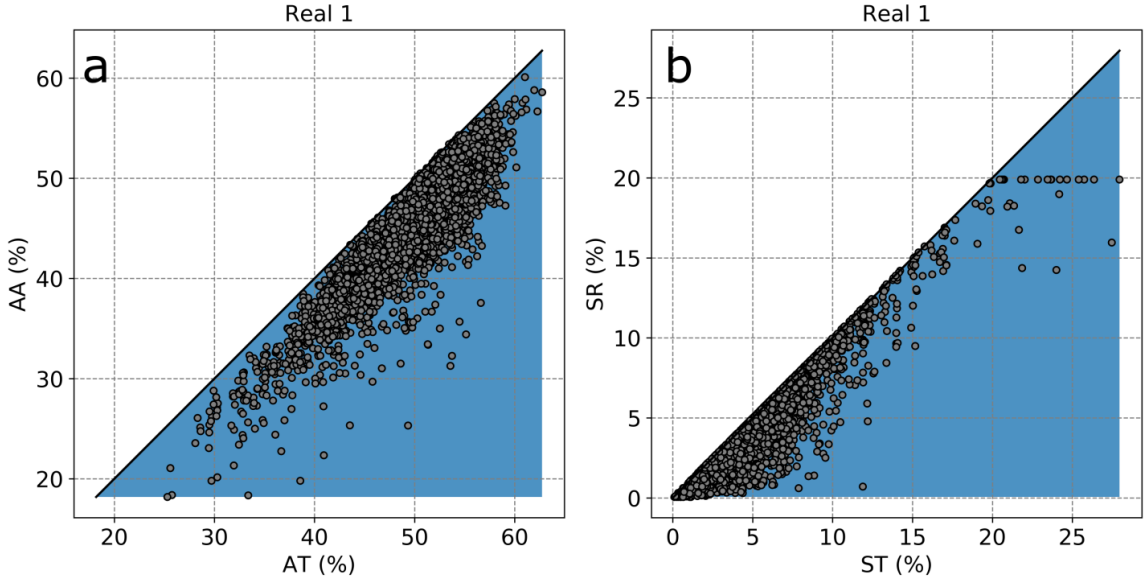

Figura 42: Gráfico de dispersão entre AT e AA (a) e entre ST e SR (b) da primeira realização. A área azul corresponde à região que não viola a restrição de fração.

6.3.5.6 Verificação de extrapolação

A figura 43a mostra o mínimo e máximo das variáveis da primeira realização e dos dados. O mínimo e máximo da primeira realização não é inferior ou superior ao mínimo e máximo dos dados para todas as variáveis. A figura 43b mostra a proporção de valores corrigidos na correção de extrapolação para a variável AA sobre as 20 realizações. A proporção de valores corrigidos para a variável AA foi pequena, com um máximo de aproximadamente 0.12%.

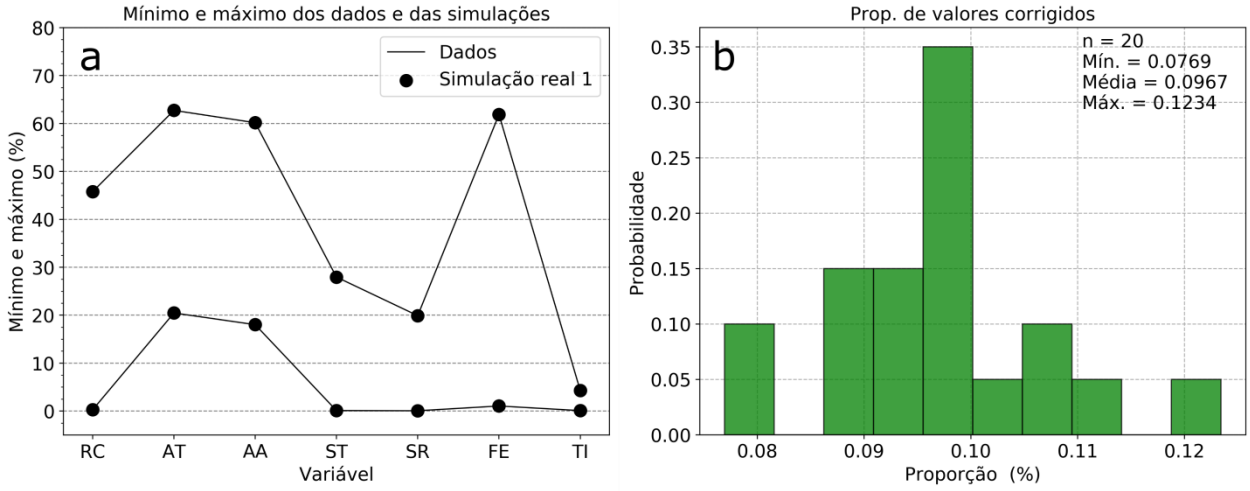

Figura 43: Mínimo e máximo dos dados e da primeira realização para todas as variáveis (a) e proporção de valores corrigidos para a variável AA (b).

6.4 Observações

Esse capítulo comparou quatro *workflows* para lidar com bancos de dados multivariados que tem restrições de fração e de soma. Para lidar com a restrição de fração, foram utilizados o teor total e a razão entre o teor fracionário e o teor total. Para lidar com a restrição de soma, quatro abordagens foram testadas para as variáveis com restrição de soma: (1) PPMT das variáveis originais, (2) PPMT com projeção ortogonal à restrição de soma, (3) razões A e PPMT e (4) razões U e PPMT.

As razões de fração foram efetivas para lidar com a restrição de fração. Os valores simulados honraram a restrição de fração. A desvantagem é a ocorrência de valores simulados além do mínimo e máximo dos dados para os teores fracionários.

PPMT reproduziu as relações bivariadas entre as variáveis. Todos os *workflows* mostraram boa reprodução dos gráficos de dispersão e coeficientes de correlação.

A abordagem convencional de PPMT aplicada às variáveis originais não gerou extrapolação. Menos de 0.1% dos valores simulados por PPMT (abordagem convencional) não respeitaram a restrição de soma. Como PPMT reproduz as relações multivariadas, PPMT implicitamente lida com a restrição de soma.

PPMT das variáveis originais usando projeção ortogonal à restrição de soma foi efetiva para lidar com a restrição de soma. A maior desvantagem da técnica é a ocorrência de valores além do mínimo e máximo dos dados, incluindo a ocorrência de valores negativos. A extrapolação dos valores simulados ocorreu para todas as variáveis, mesmo aquelas que não estavam inclusas na restrição de soma. Uma maior investigação é necessária para desenvolver uma técnica de PPMT que considere a restrição de soma e evite extrapolação.

As razões A e razões U foram eficazes para lidar com a restrição de soma. Para os *workflows* que usaram as razões A e razões U, a porcentagem de valores simulados além do mínimo e máximo dos dados foi menor do que 0.2%. As variáveis transformadas por razões U tiveram maior correlação com as variáveis originais usadas no numerador do que as variáveis transformadas por razões A.

As razões de fração e a abordagem convencional de PPMT foram utilizadas na simulação geoestatística das variáveis. Foram incluídas duas correções: (1) correção de soma e (2) correção de extrapolação. As proporções de nós de grid corrigidos foram baixas para as duas correções. As realizações reproduziram os histogramas e relações bivariadas. As restrições de fração e a restrição de soma foram honradas. O variograma foi bem reproduzido na estrutura de curto alcance na direção horizontal. Na direção vertical, o variograma foi bem reproduzido as variáveis AT, AA e FE para distâncias até 1.50 metros. Para as variáveis RC, ST, SR e TI, as realizações tiveram menor continuidade espacial do que os dados originais devido à descorrelação dos dados para $h = 0$.

7 Verificação de distribuições multivariadas

A seção 7.1 explica o cálculo da função de distribuição cumulativa (*cumulative distribution function* – cdf) para múltiplas variáveis. A cdf multivariada é utilizada para o cálculo das métricas de distância entre cdfs multivariadas, que são apresentadas na seção 7.2. A seção 7.3 explica como a proporção de valores na cauda inferior da distribuição multivariada muda à medida que o número de variáveis aumenta. A seção 7.4 faz uma comparação entre as métricas apresentadas na seção 7.2. A seção 7.5 resume os resultados obtidos.

7.1 Função de distribuição cumulativa multivariada

Considere duas variáveis aleatórias X e Y com os respectivos limiares x e y. A função de distribuição cumulativa (*cumulative distribution function* – cdf) dessas duas variáveis é definida como:

$$F(x,y) = \frac{\text{número de amostras que } X \le x \text{ e } Y \le y}{\text{número total de amostras}} \tag{36}$$

A cdf multivariada pode ser calculada para múltiplas variáveis. Os valores dos dados podem ser utilizados para os limiares x e y. Usar os valores dos dados como limiares para calcular a cdf multivariada torna o tempo de processamento viável. Por exemplo, considere uma distribuição multivariada com k variáveis. Se os percentis das variáveis são usados para calcular a cdf multivariada, 100^k pontos seriam usados para calcular a cdf multivariada. Essa abordagem é inviável para muitas variáveis. O uso dos valores dos dados limita o conjunto de limiares ao número de dados para qualquer número de variáveis.

As figuras 44a-b ilustram o cálculo da cdf multivariada para duas variáveis usando uma série de quantis e os valores dos dados, respectivamente. Cada “X” preto na figura 44 representa um ponto no espaço multidimensional usado para calcular a cdf multivariada. Quando os quantis são usados, muitos pontos estão fora da envoltória

convexa (traduzido pelo autor do termo original *convex hull*) dos dados (linha vermelha na figura 44). Os pontos fora da envoltória convexa são zeros ou uns da cdf multivariada. Por outro lado, quando os valores dos dados são utilizados, os pontos utilizados para o cálculo da cdf multivariada estão entre zero e um.

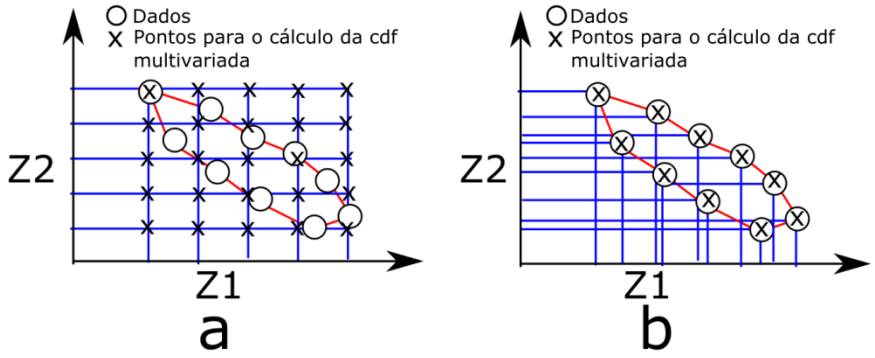

Figura 44: Esquema do cálculo da cdf multivariada usando quantis (a) e os valores dos dados (b). A linha vermelha representa a envoltória convexa dos dados.

7.2 Métricas de distância entre cdfs multivariadas

7.2.1 Estatística D90

A estatística Kolmogorov-Smirnov (chamada de estatística D) é uma medida de diferença entre cdfs. Considere duas cdfs bivariadas $F_{dados}(x,y)$ e $F_{sim}(x,y)$. $F_{dados}(x,y)$ corresponde à cdf dos dados originais para duas variáveis X e Y. $F_{sim}(x,y)$ corresponde a cdf bivariada para a simulação das variáveis X e Y. A equação 37 define a estatística D entre essas duas distribuições bivariadas:

$$D = max\left(\left|F_{dados}(x,y) - F_{sim}(x,y)\right|\right) \quad (37)$$

onde max é o máximo.

A estatística $D90$ é similar à estatística D. A diferença é que o 90º percentil é calculado, em vez do máximo (equação 38):

$$D90 = P90\left(\left|F_{dados}(x,y) - F_{sim}(x,y)\right|\right) \quad (38)$$

A estatística $D90$ é mais robusta que a estatística D na presença de valores extremos no conjunto de diferenças. A estatística $D90$ mede a relação entre n variáveis se as cdfs são calculadas para n variáveis.

7.2.2 Erro quadrático médio entre cdfs multivariadas

A equação 39 mostra o erro quadrático médio entre duas cdfs bivariadas:

$$MSE_{cdf} = E[\left(F_{dados}(x,y) - F_{sim}(x,y)\right)^2] \quad (39)$$

onde E é o valor esperado. Similar ao $D90$, o MSE_{cdf} mede a relação entre n variáveis se as cdfs são calculadas para n variáveis.

7.2.3 Diferença entre coeficiente de correlação

A equação 40 define o coeficiente de correlação entre duas variáveis aleatórias X e Y:

$$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \quad (40)$$

onde $Cov(X,Y)$ é a covariância entre X e Y, σ_X é o desvio padrão de X e σ_Y é o desvio padrão de Y.

Para verificar a simulação multivariada, o geomodelador pode calcular a diferença absoluta entre o coeficiente de correlação da simulação ρ_{sim} e o coeficiente de correlação dos dados ρ_{dados} (equação 41):

$$\left|\rho_{dados} - \rho_{sim}\right| \tag{41}$$

O coeficiente de correlação mede a relação linear entre duas variáveis. Se o banco de dados apresenta uma relação linear forte, os valores simulados devem reproduzir essa relação linear forte.

Uma desvantagem do coeficiente de correlação é que ele mede a relação apenas entre duas variáveis. O coeficiente de correlação não consegue caracterizar relações entre três ou mais variáveis, ao contrário do $D90$ e do erro quadrático médio das cdfs multivariadas. Outra desvantagem do coeficiente de correlação é que ele não descreve relações não lineares.

7.3 Zeros da cdf multivariada

Quando os valores dos dados são usados como limiares para calcular a cdf multivariada, os valores da cdf multivariada decrescem e tendem a zero à medida que o número de variáveis aumenta. Por exemplo, considere uma observação com dois valores $x1$ e $y1$. Quando uma variável é considerada, a cdf representa a proporção de pontos em que a primeira variável é menor ou igual à $x1$. Quando duas variáveis são consideradas, a cdf bivariada representa a proporção de pontos em que a primeira variável é menor ou igual à $x1$ e a segunda variável é menor ou igual à $y1$. A proporção para o caso de uma variável é maior ou igual do que a proporção para o caso de duas variáveis. À medida que o número de variáveis aumenta, a proporção de pontos abaixo de múltiplos limiares diminui. A região onde a cdf multivariada é próxima de zero é chamada de cauda inferior da distribuição multivariada.

A tendência de representar a cauda inferior da distribuição multivariada à medida que o número de variáveis aumenta é reforçada se as duas variáveis tem correlação negativa. A figura 45 mostra um esquema de duas variáveis com correlação negativa. A cdf multivariada para os pontos cinza na figura 45 é $1/n$, onde n é o número de dados. For exemplo, além do ponto $(x1, y1)$, não há outro ponto em que a variável X é menor

ou igual a $x1$ e a variável Y é menor ou igual a $y1$ (veja a figura 45, não há nenhum ponto à esquerda da linha pontilhada vertical e abaixo da linha pontilhada horizontal). Os pontos em que a cdf multivariada é menor ou igual a $1/n$ são chamados nessa tese de “zeros” da cdf multivariada (a cdf multivariada não é exatamente zero, mas se torna bem próxima à zero à medida que o número de dados aumenta). Na figura 45, 8 dos 16 pontos (50% dos dados) são zeros da cdf multivariada.

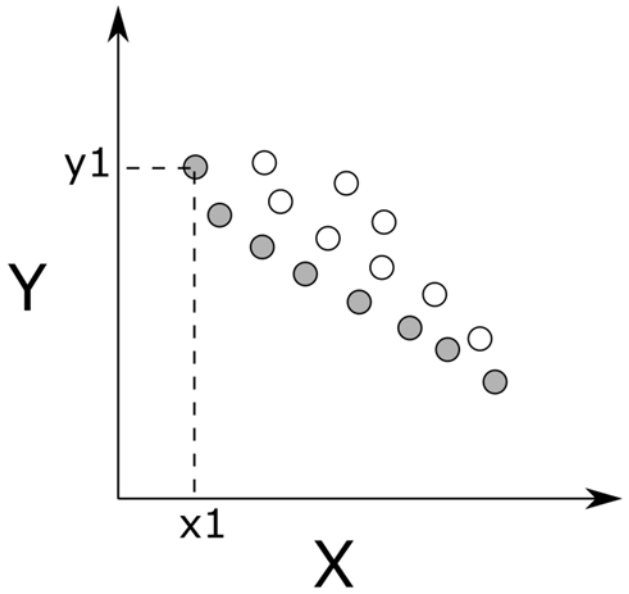

Figura 45: Esquema de duas variáveis com correlação negative. Os pontos cinza representam “zeros” da cdf multivariada.

Uma cdf multivariada que só contém valores próximos de zero não é informativa. Uma cdf multivariada informativa deve conter valores entre zero e um. À medida que o número de variáveis aumenta, uma maior quantidade de dados é necessária para ter uma cdf multivariada informativa. Nesse contexto, é interessante checar a quantidade de dados necessária para ter uma cdf multivariada representativa.

A porcentagem de zeros na cdf multivariada em função do número de variáveis foi analisada para diferentes quantidades de dados. Os dados foram simulados por Monte Carlo de uma distribuição multi-Gaussiana com nove variáveis. Todas as variáveis seguem uma distribuição Gaussiana padrão, com média igual a zero e variância igual a um. A figura 46 mostra a matriz de correlação das variáveis.

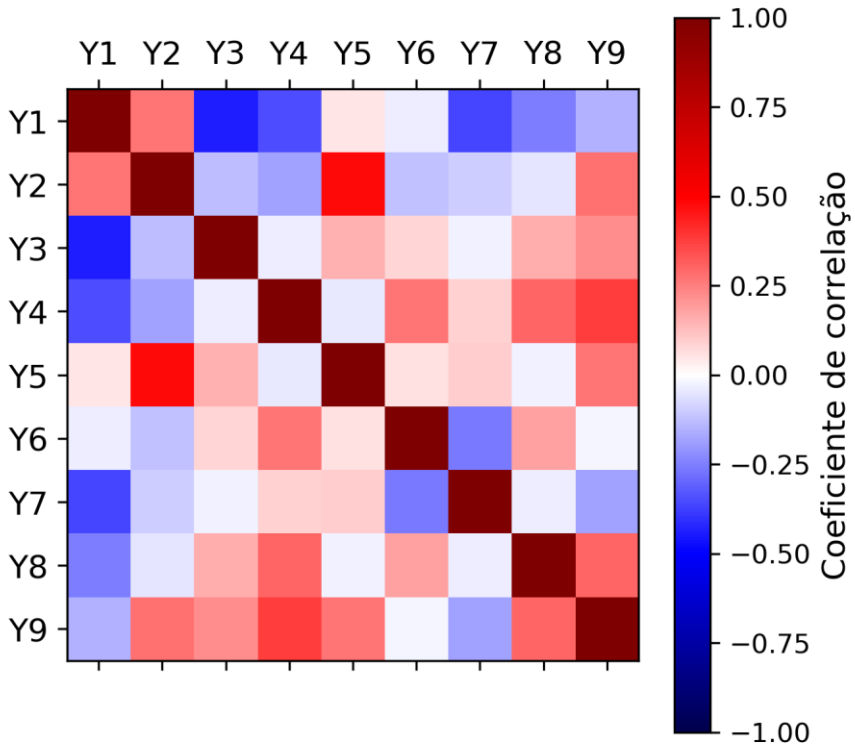

Figura 46: Matriz de correlação das variáveis.

Para cada k variáveis consideradas, as primeiras k variáveis foram usadas para calcular a cdf multivariada. Por exemplo, a cdf bivariada foi calculada com as primeiras duas variáveis. Escolher as primeiras k variáveis simplifica o experimento, pois não é necessário calcular a cdf multivariada para todas as possíveis combinações. Por exemplo, para o caso de duas variáveis de um total de nove, 36 combinações de pares são possíveis. A figura 47 mostra a porcentagem de zeros para diferentes números de variáveis. À medida que o número de variáveis aumenta, uma maior quantidade de dados é necessária para ter uma cdf multivariada informativa, com baixa porcentagem de zeros.

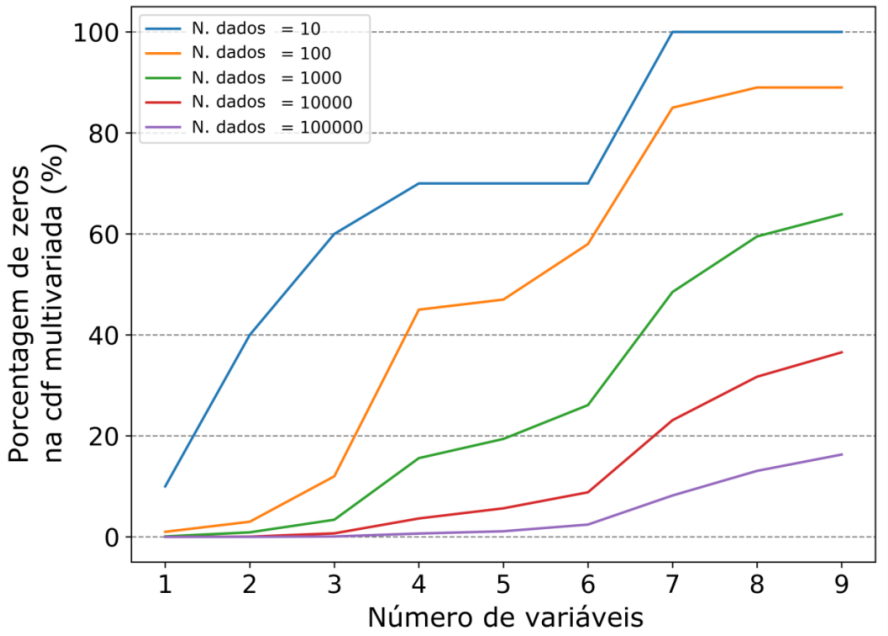

Figura 47: Porcentagem de zeros da cdf multivariada em função do número de variáveis.

A figura 48 mostra os *boxplots* da cdf multivariada em função do número de variáveis. À medida que o número de variáveis aumenta, a cdf multivariada tende a ter apenas valores próximos de zero.

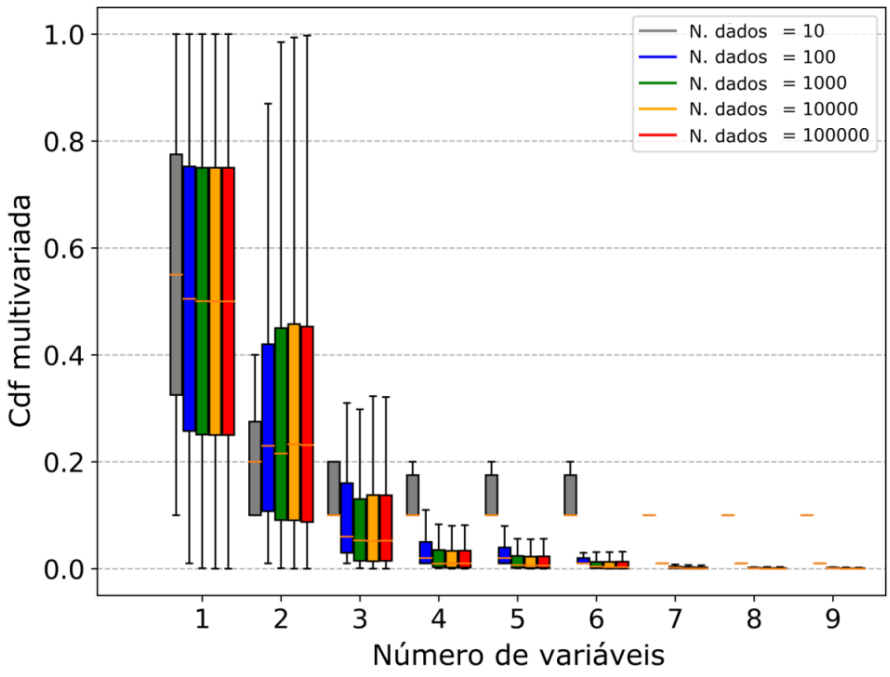

Figura 48: *Boxplots* da cdf multivariada em função do número de variáveis.

7.4 Comparação entre métricas

Essa seção compara a estatística $D90$ contra a diferença entre coeficientes de correlação e o erro quadrático médio da cdf multivariada.

7.4.1 Metodologia

A figura 49 mostra o fluxograma da metodologia. Primeiro, os dados foram simulados por Monte Carlo de uma distribuição bivariada conhecida. Segundo, os dados foram modificados de duas maneiras: (1) adicionando erro e (2) adicionando viés. Então, os dados modificados foram comparados com os dados originais usando o $D90$, a diferença entre coeficientes de correlação e o erro quadrático médio da cdf multivariada. Dois casos foram realizados. O caso I considera duas variáveis com uma relação linear. O caso II considera duas variáveis com uma relação não linear.

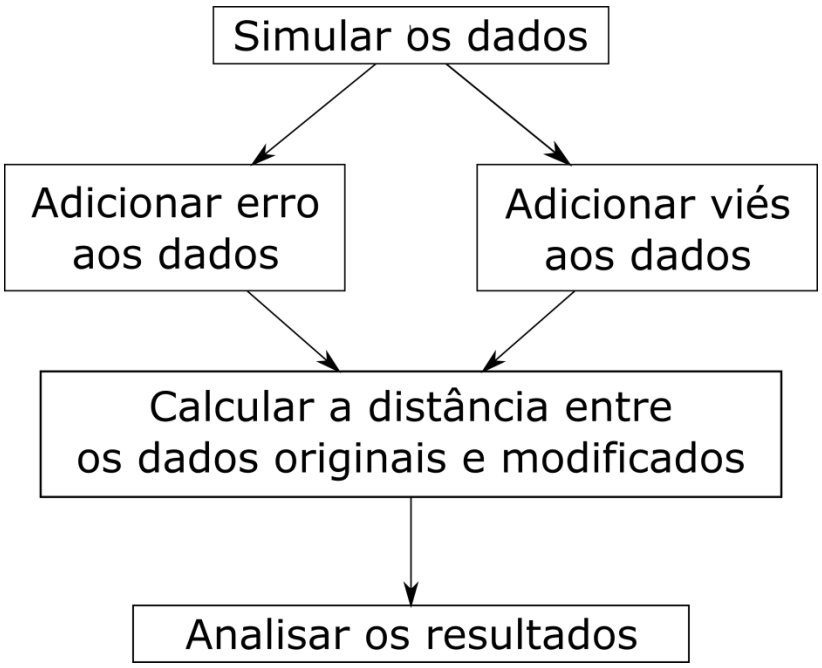

Figura 49: Fluxograma da metodologia.

Adicionando erro aos dados

Erros aleatórios Gaussianos com médias iguais a zero e diferentes coeficientes de variação foram adicionados aos dados. Os coeficientes de variação escolhidos foram os seguintes: 2, 4, 6, 8 e 10%. Para cada amostra, o desvio padrão do erro corresponde ao coeficiente de variação multiplicado pelo valor da amostra. O erro foi adicionado às duas variáveis.

Adicionando viés aos dados

Foram adicionados 2, 4, 6, 8 e 10% de viés às duas variáveis. Para cada amostra, os valores das duas variáveis foram multiplicados por um mais o viés.

7.4.2 Caso I: dados com relação linear

Os dados foram simulados por Monte Carlo de uma distribuição Gaussiana bivariada com coeficiente de correlação igual a -0.94. A primeira variável é chamada X e tem uma média de 52.64 e desvio padrão de 4.57. A segunda variável é chamada de Y e tem uma média de 13.78 e desvio padrão de 7.07. Essas duas variáveis tem uma forte relação linear, com coeficiente de correlação próximo de menos um.

7.4.2.1 Dados com erro

A figura 50a o gráfico de dispersão dos dados originais e as figuras 50b-f mostram os gráficos de dispersão dos dados com erro relativo adicionado.

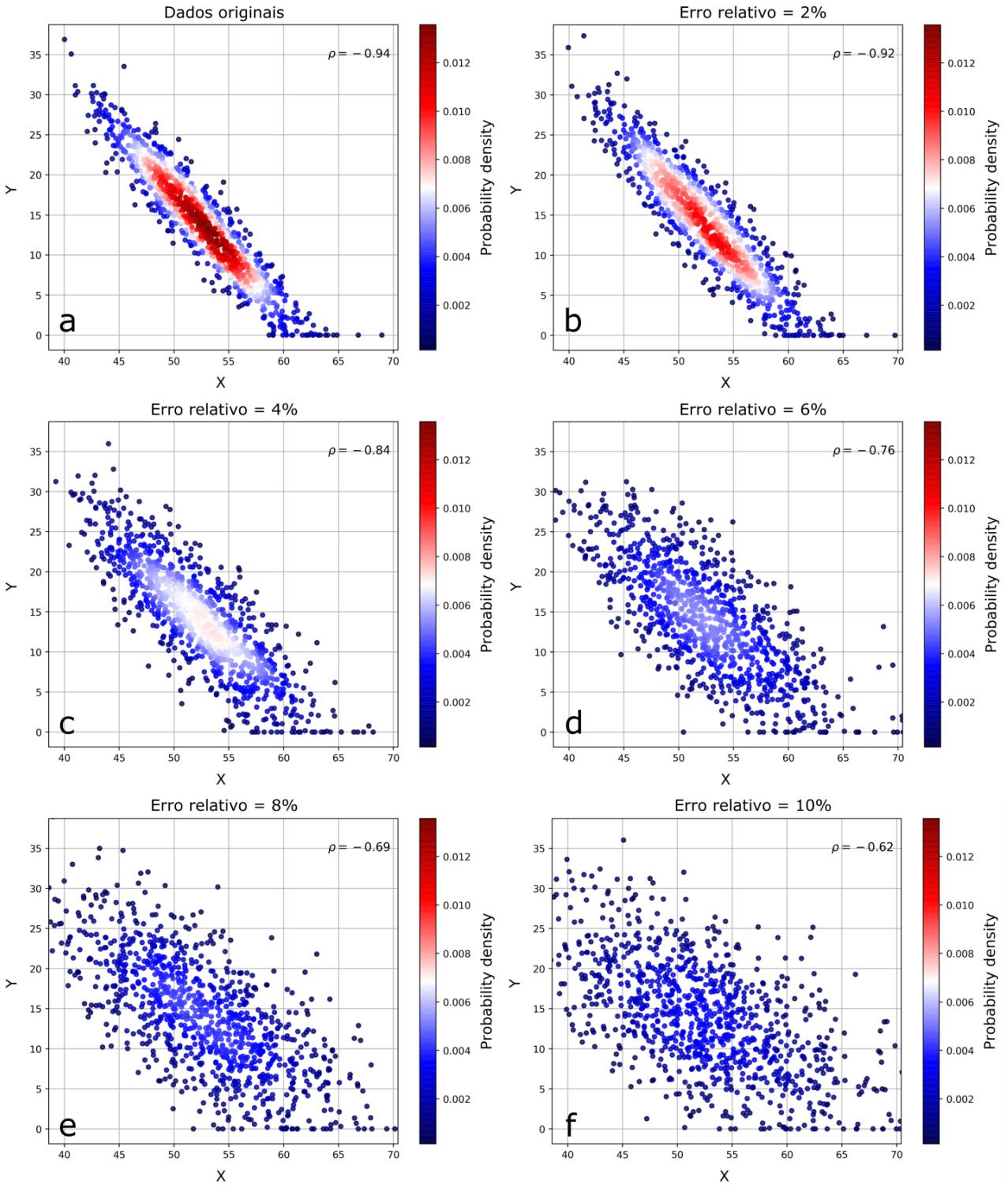

Figura 50: Gráfico de dispersão entre as variáveis X e Y dos dados originais (a) e dos dados com diferentes níveis de erro relativo: 2% (b), 4% (c), 6% (d), 8% (e) e 10% (f).

As figuras 51a-c mostram a relação entre o erro relativo e o $D90$, o erro quadrático médio e a diferença entre coeficientes de correlação, respectivamente. O $D90$ (figura 51a), o erro quadrático médio da cdf multivariada (figura 51b) e a diferença entre os coeficientes de correlação (figura 51c) aumentaram à medida que o erro relativo aumentou.

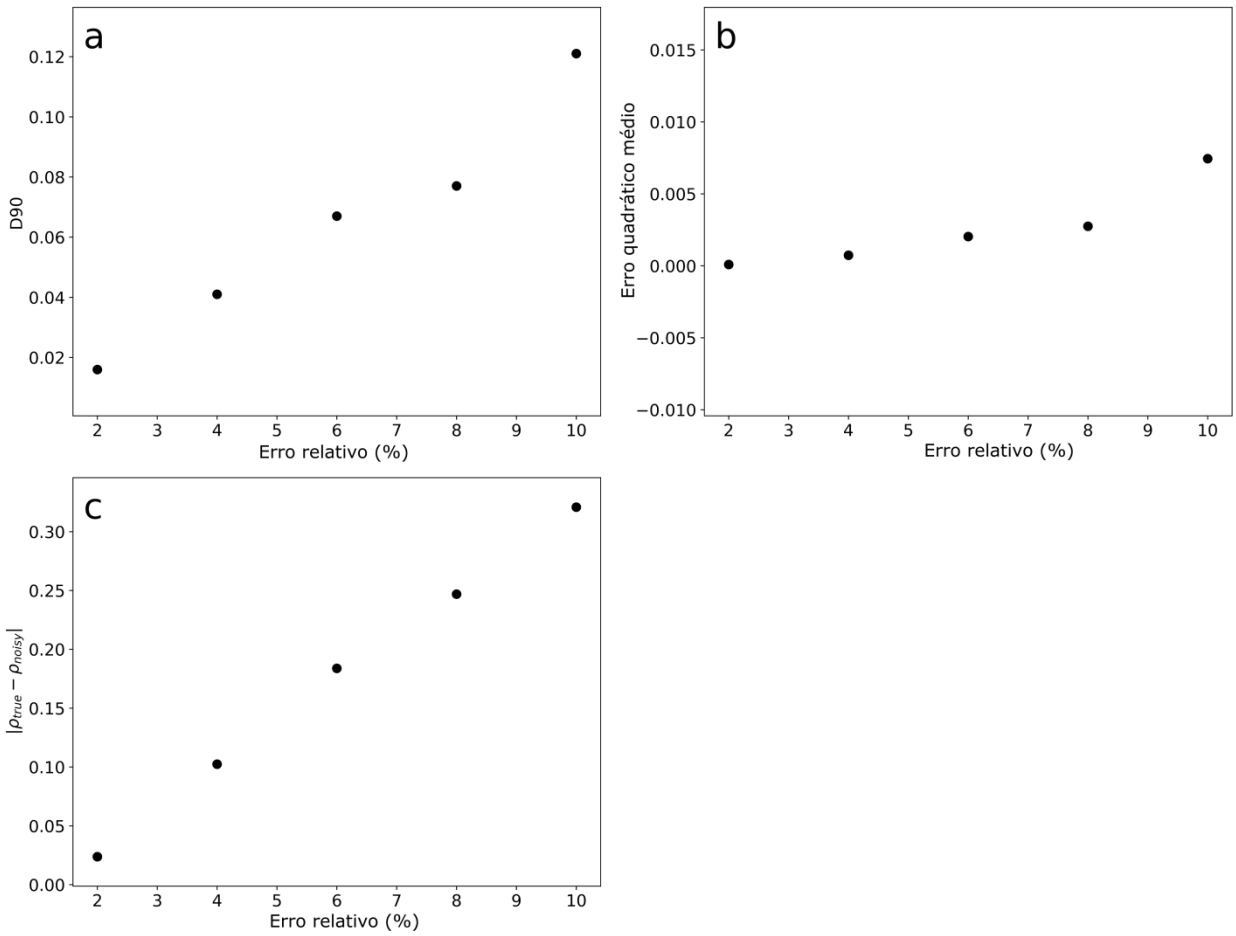

Figura 51: Gráfico de dispersão entre o erro relativo e o D90 (a), erro quadrático médio (b) e diferença entre coeficientes de correlação (c).

7.4.2.2 Dados com viés

A figura 52a mostra o gráfico de dispersão dos dados originais e as figuras 52b-f mostram os gráficos de dispersão dos dados com viés.

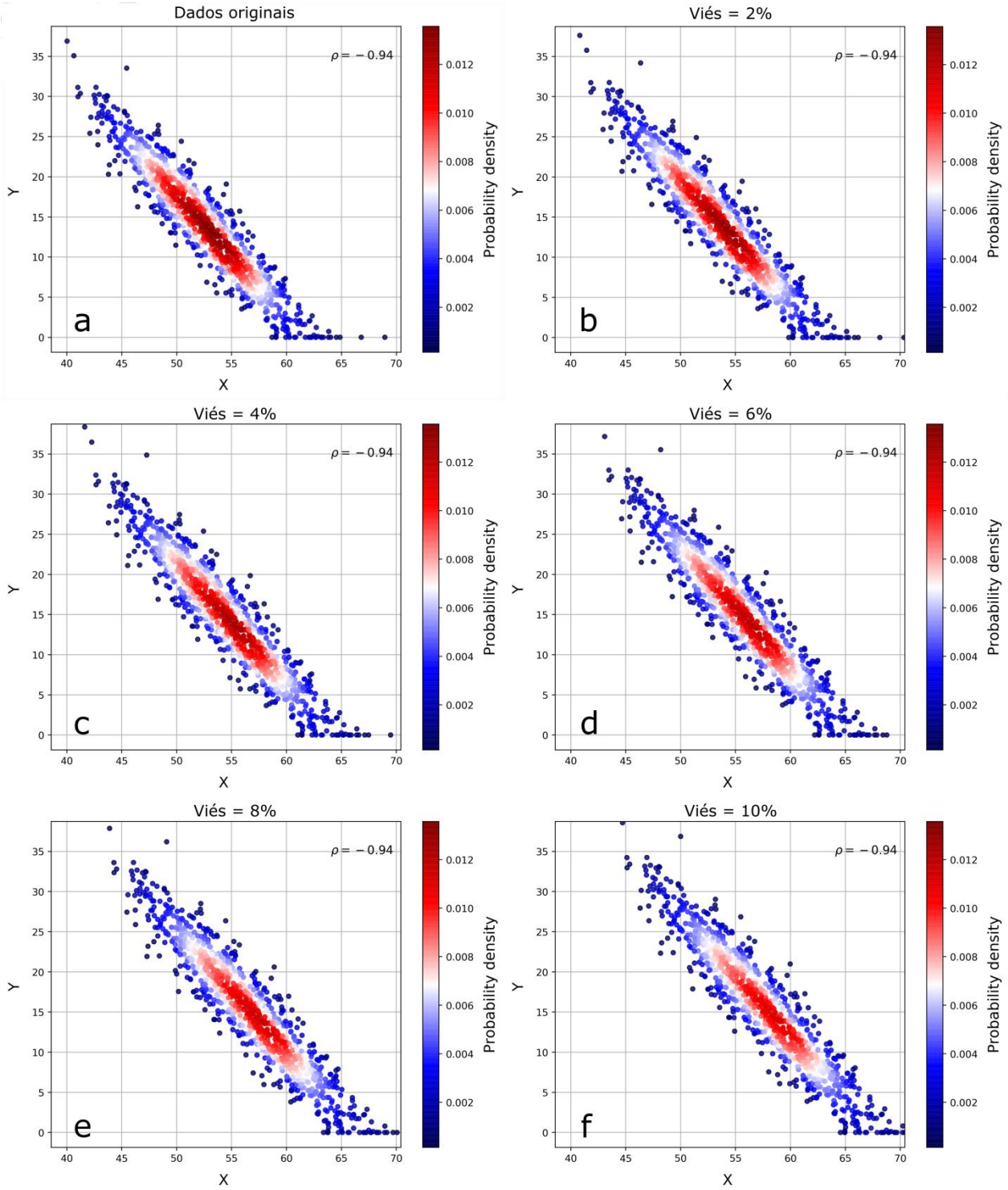

Figura 52: Gráfico de dispersão entre as variáveis X e Y dos dados originais (a) e dos dados com diferentes níveis de viés: 2% (b), 4% (c), 6% (d), 8% (e) e 10% (f).

A figura 53 mostra a relação entre o viés e o $D90$, erro quadrático médio e a diferença entre coeficientes de correlação. O $D90$ e o erro quadrático médio

aumentaram à medida que o viés aumentou (figuras 53a-b). Por outro lado, a diferença entre coeficientes de correlação permaneceu igual à zero para os diferentes níveis de viés (figura 53c).

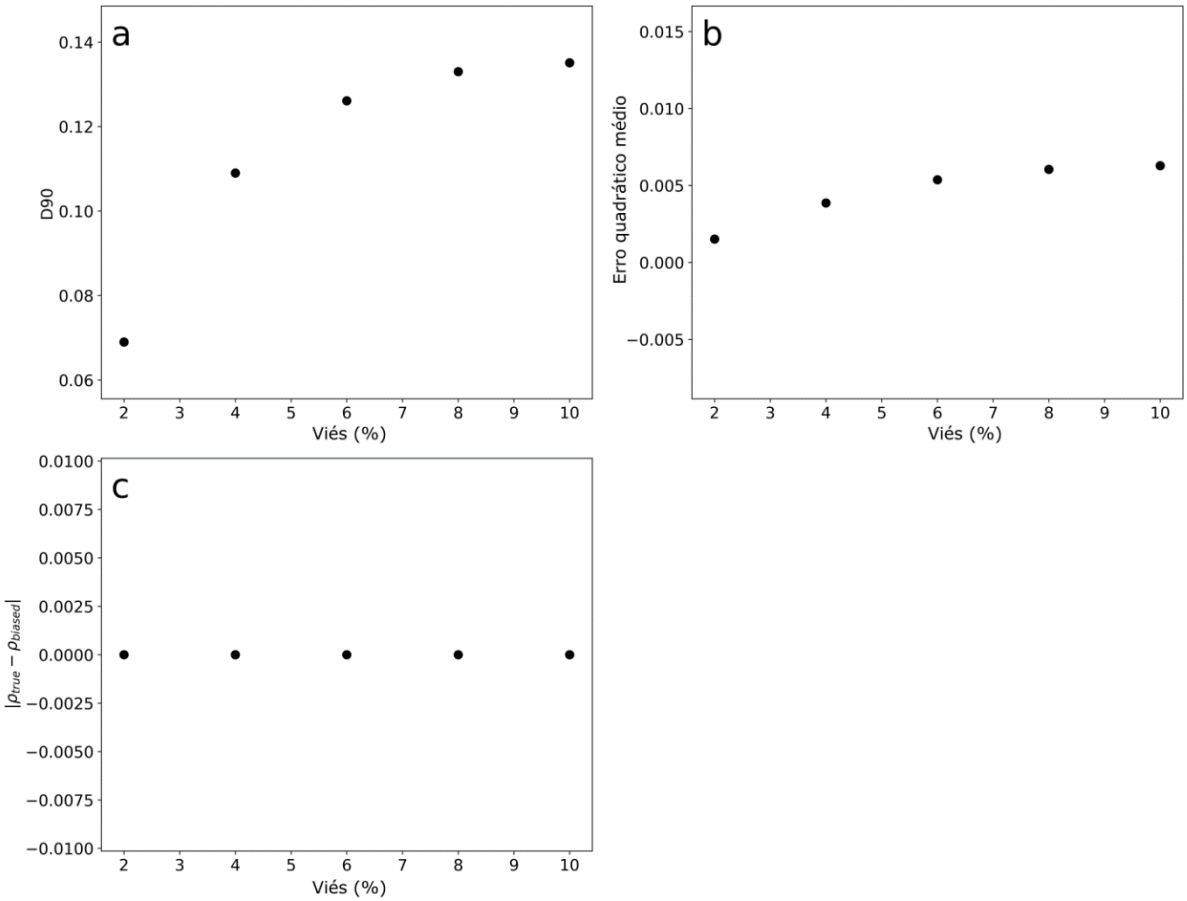

Figura 53: Gráfico de dispersão entre o viés e o D90 (a), erro quadrático médio (b) e diferença entre coeficientes de correlação.

7.4.3 Caso II: dados com relação não linear

Nesse teste, duas variáveis chamadas X e Y foram usadas. X foi simulada por Monte Carlo de uma distribuição Gaussiana com média igual a 30 e desvio padrão igual a 5. A variável Y foi calculada através da equação 42:

$$Y = X^2 - 60X + 900 \quad (42)$$

As variáveis X e Y tem uma relação quadrática.

7.4.3.1 Dados com erro

A figura 54a mostra o gráfico de dispersão dos dados originais e as figuras 54b-f mostram os gráficos de dispersão dos dados com erro.

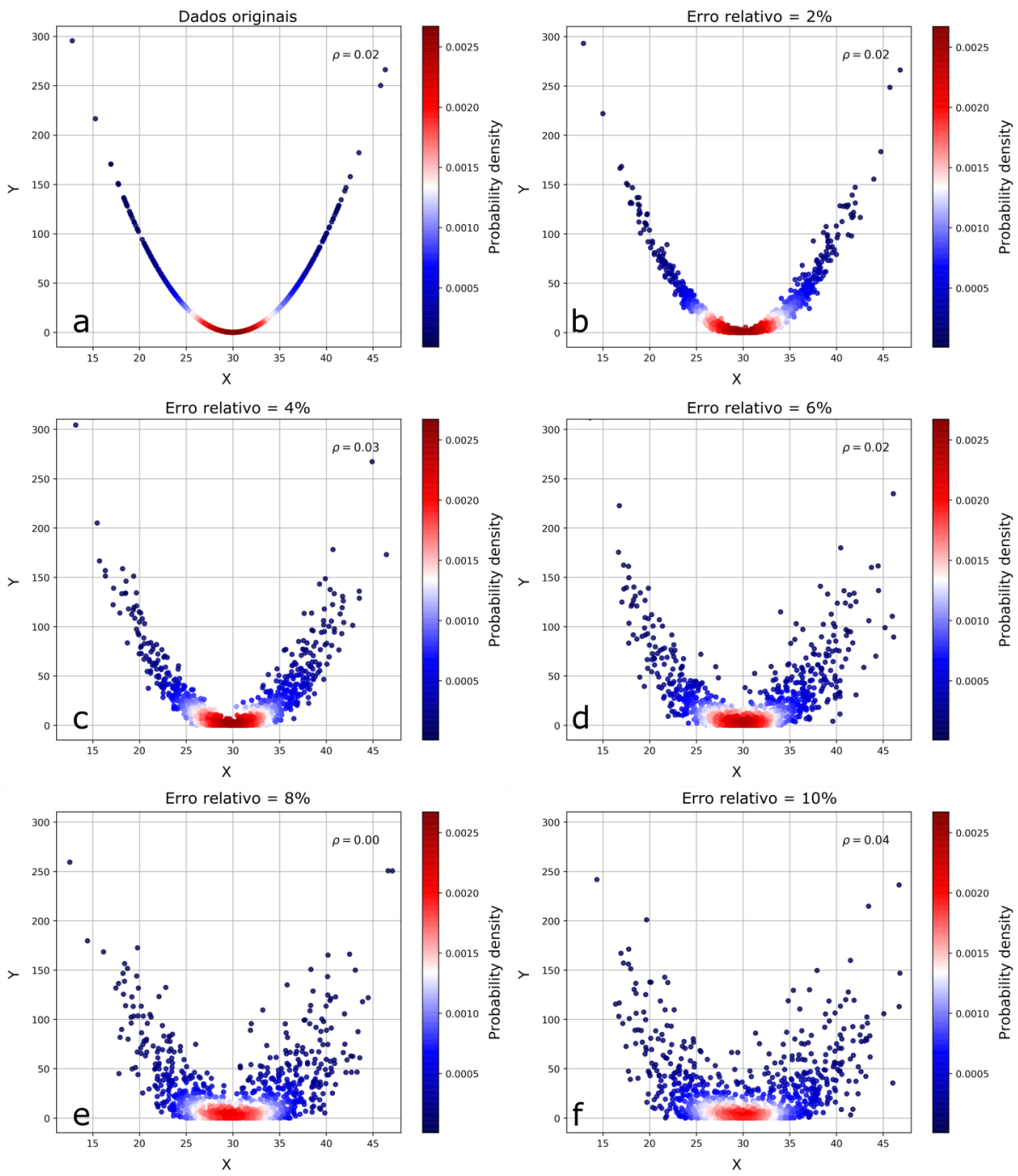

Figura 54: Gráfico de dispersão entre as variáveis X e Y dos dados originais (a) e dos dados com diferentes níveis de erro relativo: 2% (b), 4% (c), 6% (d), 8% (e) e 10% (f).

As figuras 55a-c mostram a relação entre o erro relativo e o $D90$, o erro quadrático médio e a diferença entre coeficientes de correlação, respectivamente. O $D90$ e o erro quadrático médio aumentaram à medida que o erro relativo aumento

(figuras 55a-b). Entretanto, a diferença entre coeficientes de correlação ficou próxima de zero para os diferentes erros relativos considerados (veja a escala do eixo vertical na figura 55c).

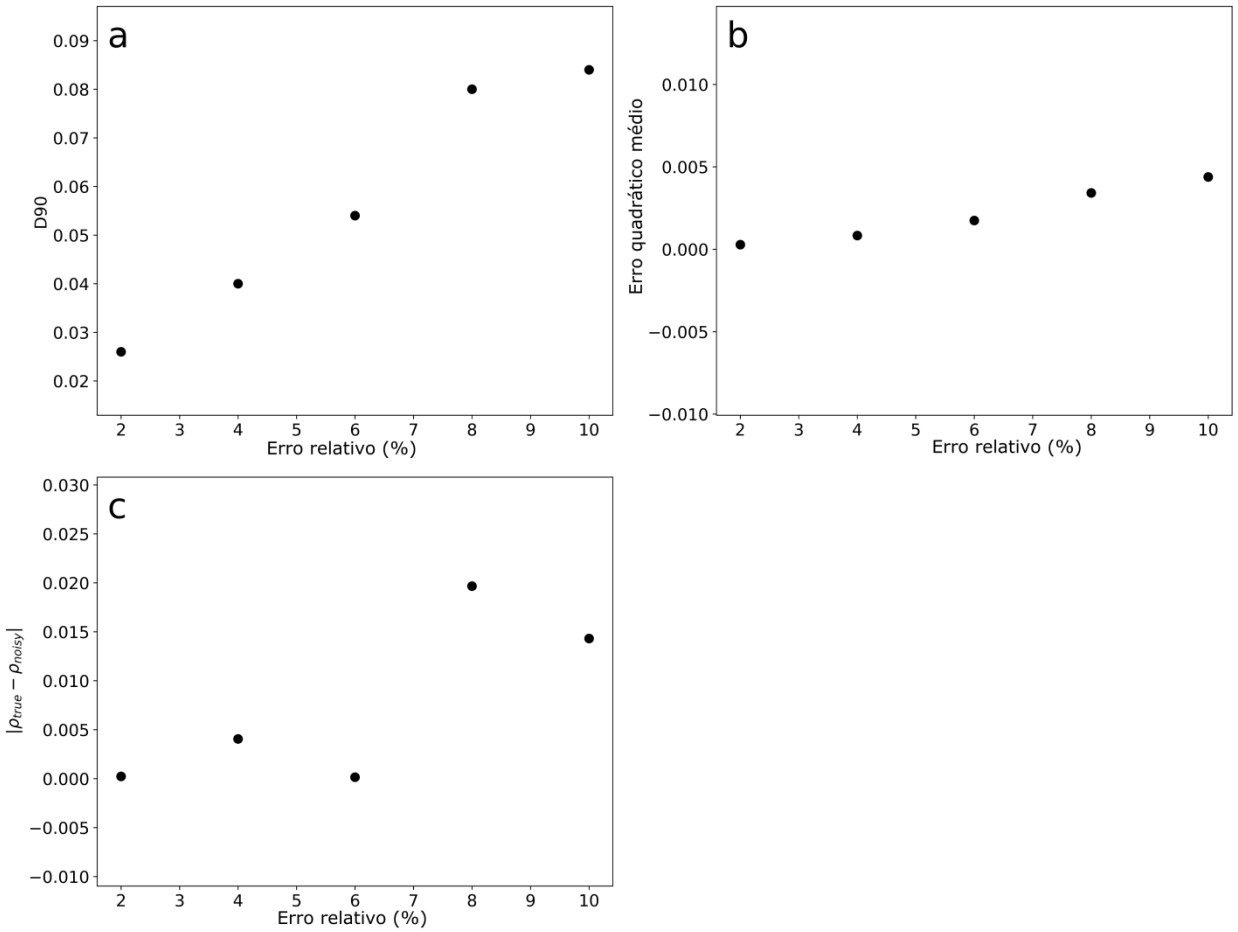

Figura 55: Gráfico de dispersão entre o erro relativo e o D90 (a), erro quadrático médio (b) e diferença entre coeficientes de correlação (c).

7.4.3.2 Dados com viés

A figura 56a mostra o gráfico de dispersão dos dados originais enquanto que as figuras 56b-f mostram os gráficos de dispersão dos dados com diferentes níveis de viés adicionados.

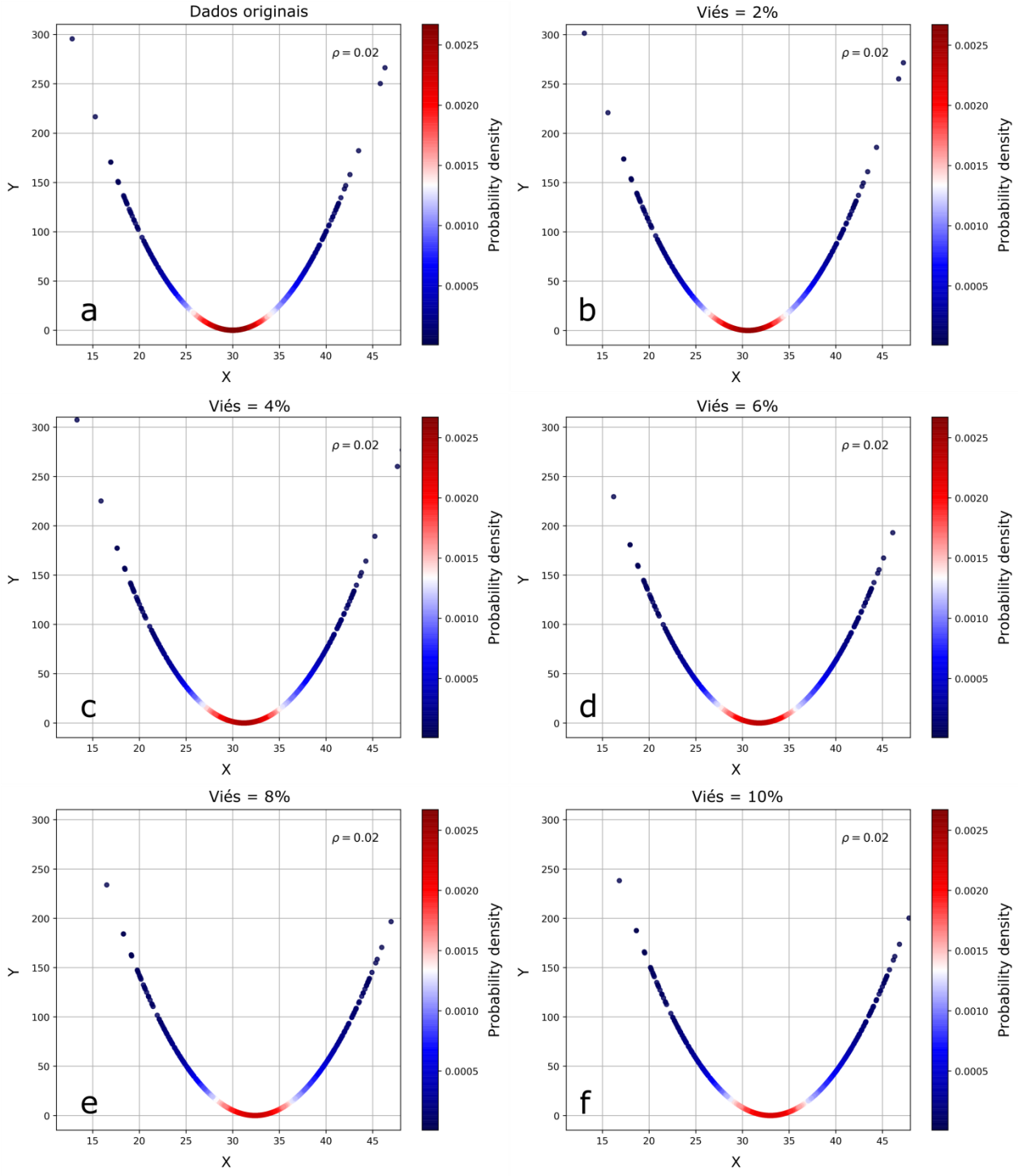

Figura 56: Gráfico de dispersão entre as variáveis X e Y dos dados originais (a) e dos dados com diferentes níveis de viés: 2% (b), 4% (c), 6% (d), 8% (e) e 10% (f).

A figura 57 mostra a relação entre o viés e as diferentes métricas observadas. O $D90$ e o erro quadrático médio aumentaram à medida que o viés aumentou (figuras

57a-b). Por outro lado, a diferença entre coeficientes de correlação foi praticamente zero para os diferentes níveis de viés (Figura 57c).

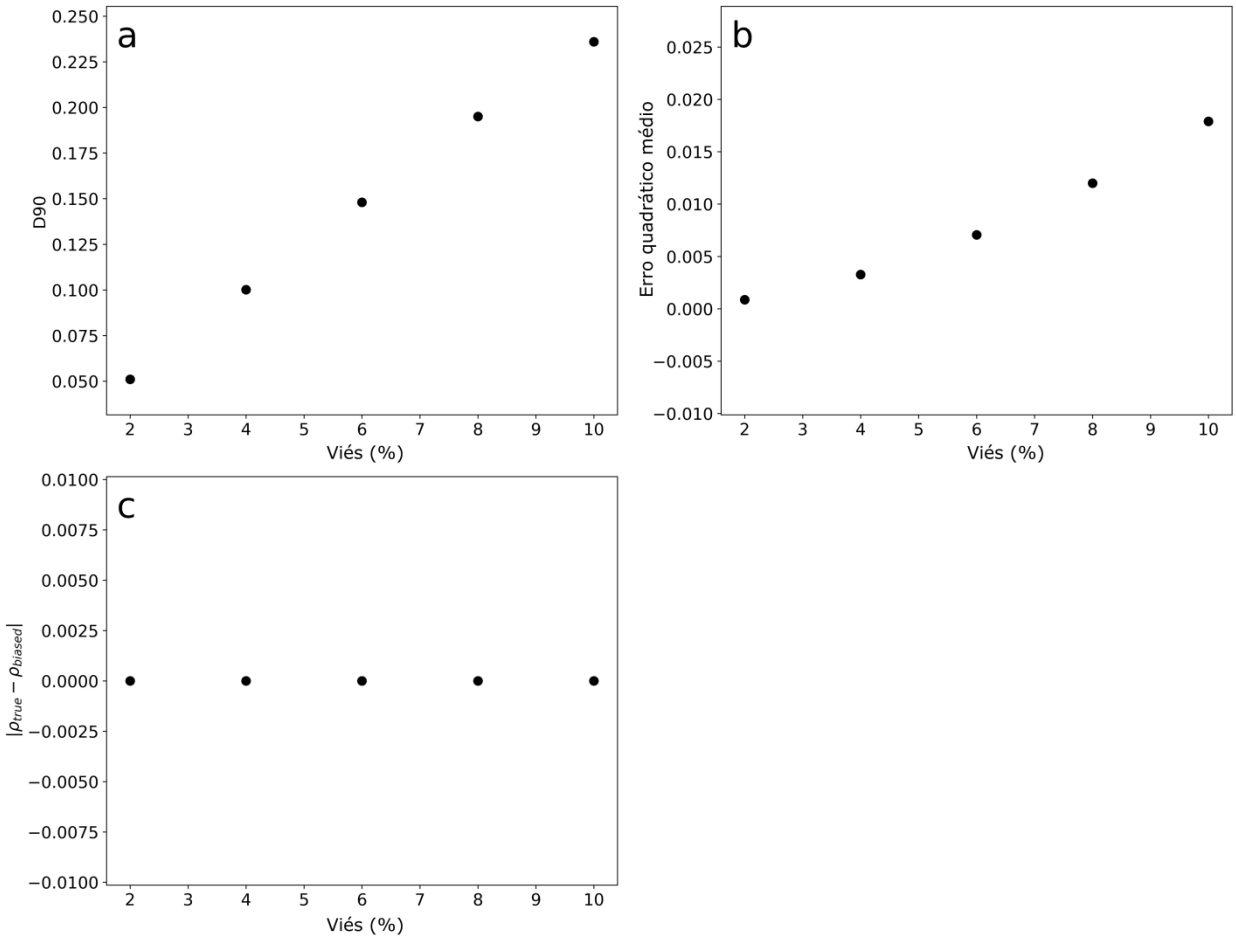

Figura 57: Gráfico de dispersão entre o viés e o D90 (a), erro quadrático médio (b) e diferença entre coeficientes de correlação (c).

7.5 Observações

Esse capítulo apresentou o $D90$, que é a métrica proposta para medir a distância entre duas distribuições multivariadas. O $D90$ é definido como o 90º percentil da diferença absoluta entre duas cdfs multivariadas. O $D90$ foi comparado com o erro quadrático médio da cdf multivariada e com a diferença entre coeficientes de correlação. As comparações foram feitas para dois casos: (1) duas variáveis com uma forte relação linear e (2) duas variáveis com uma relação quadrática. Para esses dois casos, os dados foram modificados de duas maneiras diferentes. A primeira modificação consistiu em adicionar diferentes níveis de erro aos dados originais. A segunda modificação consistiu em adicionar diferentes níveis de viés aos dados originais. As métricas foram usadas para calcular a diferença entre as distribuições multivariadas dos dados originais e modificados (dados com erro e dados com viés).

No caso de dados com relação linear, a diferença entre coeficientes de correlação foi efetiva para identificar o erro relativo, mas não detectou o viés. Para o caso de dados com relação não linear, a diferença entre coeficientes de correlação não mediu o erro relativo e também não mediu o viés.

O $D90$ e o erro quadrático médio da cdf multivariada foram melhores do que a diferença entre coeficientes de correlação. O $D90$ e o erro quadrático médio foram sensíveis tanto para o erro relativo quanto para o viés para os dois casos: (1) dados com relação linear e (2) dados com relação quadrática.

8 Comparativo entre método direto e indireto em simulação geoestatística multivariada

O capítulo 8 apresenta um comparativo entre o método direto e indireto para simulação geoestatística multivariada. O banco de dados é o mesmo utilizado no estudo de caso de simulação geoestatística multivariada com restrições apresentado no capítulo 6. As seções de descrição espacial dos dados e desagrupamento presentes no capítulo 6 não serão repetidas. As variáveis utilizadas nesse comparativo foram a Alumina Total (AT) e Sílica Total (ST).

8.1 Banco de dados

Os teores de Alumina Total (AT) e Sílica Total (ST) foram analisados por faixas granulométricas. A porcentagem mássica retida na fração granulométrica é chamada de Recuperação (RC). No caso do método indireto, é necessário obter as variáveis acumuladas. As variáveis acumuladas foram obtidas através da multiplicação dos teores pela Recuperação (equação 37):

$$\begin{aligned} ATA &= AT \cdot RC \\ STA &= ST \cdot RC \end{aligned} \tag{37}$$

A tabela 11 mostra o sumário estatístico das variáveis. As variáveis acumuladas ATA e STA possuem maior coeficiente de variação do que as variáveis originais AT e ST (tabela 11). Essa situação é diferente do que ocorre em geral para depósitos tabulares finos ou formados por veios. Nesses tipos de depósitos, a variável acumulada é obtida pela multiplicação do teor pela espessura. Os teores altos e de alta variabilidade geralmente estão associados a pequenas espessuras. Nessa situação, as variáveis acumuladas em geral tem menor coeficiente de variação menor do que o teor de ouro (Rossi e Deutsch, 2013) e são mais fáceis de estimar. Além disso, as variáveis acumuladas diminuem a área de influência de teores altos associados e pequenas espessuras (Rossi e Deutsch, 2013).

Tabela 11: Sumário estatístico das variáveis.

	AT (%)	ST (%)	RC (%)	ATA (%^{2})	STA (%^{2})
Número	6267	6267	6267	6267	6267
Média	50.56	4.35	16.82	857.88	71.78
Desv. pad.	4.81	2.49	5.72	320.3	47.72
CV	0.10	0.57	0.34	0.37	0.66
Mínimo	20.46	0.1	0.27	10.19	1.29
Q25	48.16	2.4	12.73	627.53	36.87
Q50	51.37	4.07	16.23	829.84	61.88
Q75	53.92	5.85	20.24	1047.21	94.5
Máximo	62.73	27.95	45.78	2652.04	526.12

A figura 58 mostra a matriz de correlação entre as variáveis. A variável RC tem correlação baixa com os teores originais AT e ST. Por outro lado, RC tem alta correlação com a variável acumulada ATA, com um coeficiente de correlação de 0.97.

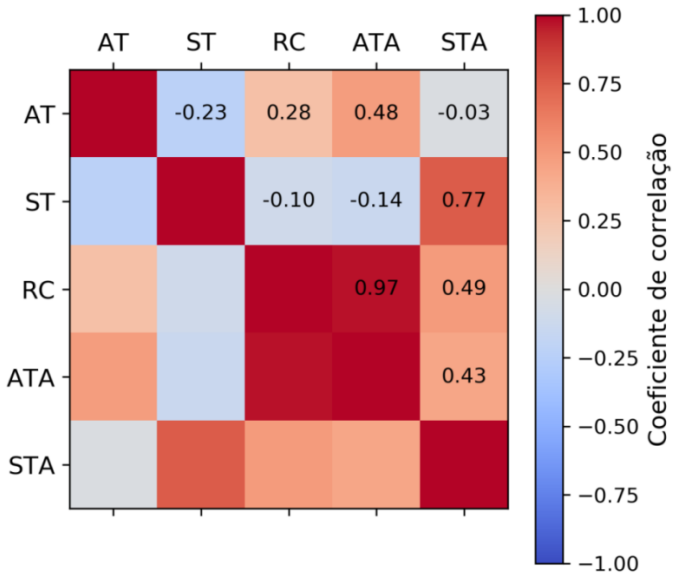

Figura 58: Matriz de correlação entre as variáveis.

8.2 Análise da continuidade espacial

Os variogramas modelados são dos valores *normal score* das variáveis. Barnett *et al.* (2016) recomendam usar os variogramas dos valores *normal score* para as simulações geoestatísticas feitas utilizando a transformação PPMT. No método direto, a transformação *normal score* foi aplicada sobre as variáveis AT e ST. No método indireto, a transformação *normal score* foi aplicada sobre as variáveis RC, ATA e STA. A transformação *normal score* foi feita sem utilizar pesos de desagrupamento. A tabela 12 mostra os modelos dos variogramas utilizados.

Tabela 12: Modelos de variograma das variáveis normal score.

Método	Variável	Estrutura	Variância	Alcance NS (m)	Alcance EW (m)	Alcance vert. (m)
Direto	NS_AT	Efeito pepita	0.20	x	x	x
		Esférico	0.52	65.00	65.00	2.40
		Esférico	0.28	2000.00	2000.00	2.50
	NS_ST	Efeito pepita	0.00	x	x	x
		Esférico	0.57	60.00	60.00	2.65
		Exponencial	0.43	1900.00	1900.00	2.70
Indireto	NS_RC	Efeito pepita	0.00	x	x	x
		Esférico	0.60	65.00	65.00	2.60
		Esférico	0.30	430.00	430.00	3.50
		Esférico	0.10	10000.00	10000.00	3.70
	NS_ATA	Efeito pepita	0.00	x	x	x
		Esférico	0.55	40.00	40.00	2.50
		Esférico	0.30	380.00	380.00	3.30
		Esférico	0.15	3000.00	3000.00	3.50
	NS_STA	Efeito pepita	0.00	x	x	x
		Esférico	0.60	65.00	65.00	2.55
		Esférico	0.40	1600.00	1600.00	2.70

8.3 Simulação geoestatística

A figura 59 mostra o fluxograma das metodologias para o método direto e indireto.

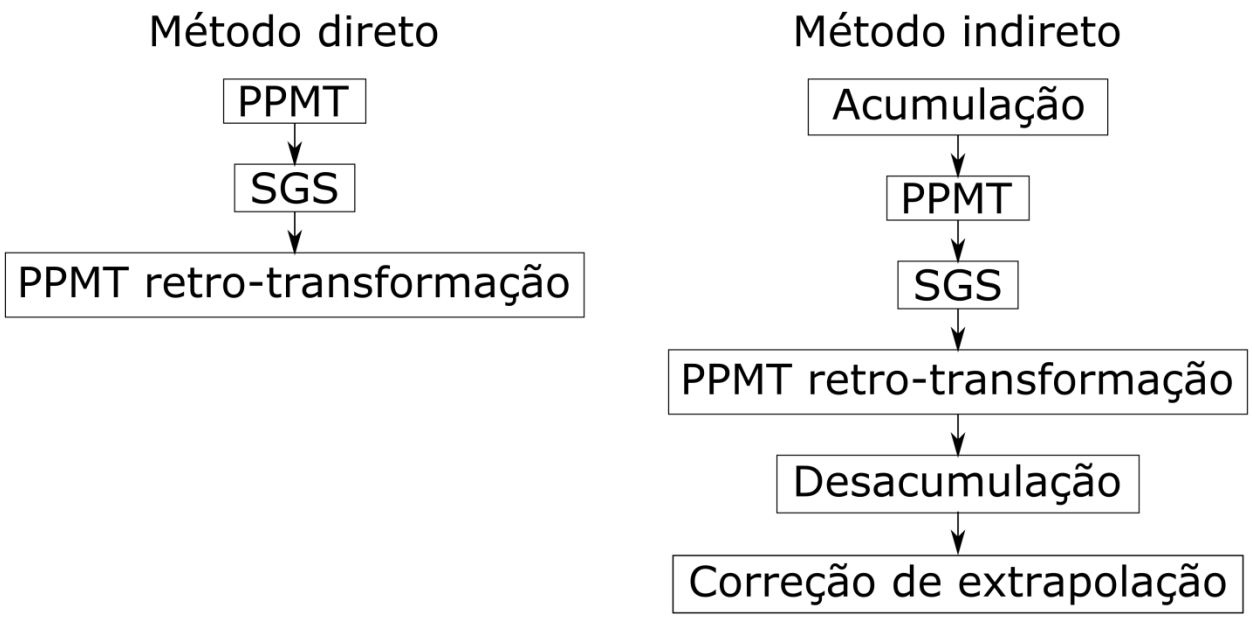

Figura 59: Fluxograma das metodologias das simulações para o método direto e indireto.

As variáveis transformadas PPMT foram utilizadas na simulação. A transformação PPMT foi feita utilizando pesos de desagrupamento. O método direto simula as variáveis transformadas correspondentes às variáveis originais AT e ST. O método indireto simula as variáveis transformadas correspondentes às variáveis acumuladas ATA e STA e à variável RC. A simulação da variável RC é necessária para a etapa de desacumulação no método indireto. A tabela 13 descreve a transformação PPMT para os dois métodos. As variáveis de saída na tabela 13 foram simuladas utilizando simulação sequencial Gaussiana (*sequential Gaussian simulation* – SGS).

Tabela 13: Variáveis de entrada e saída utilizadas na transformação PPMT.

Metodologia	Variáveis de entrada	Transformação	Variáveis de saída
Método direto	AT	PPMT	PPMT_AT
	ST		PPMT_ST
Método indireto	RC	PPMT	PPMT_RC
	ATA		PPMT_ATA
	STA		PPMT_STA

As simulações foram feitas utilizando o software *usgsim* (Manchuk e Deutsch, 2012). Foram feitas 20 realizações. Foram utilizadas 40 amostras para o sistema de krigagem. As simulações das variáveis PPMT foram transformadas para o espaço dos dados originais fazendo a transformação inversa da PPMT. No caso do método indireto, foi necessário fazer a etapa de desacumulação. Nessa etapa, as simulações das variáveis acumuladas ATA e STA foram divididas pelas simulações da variável RC. A etapa de desacumulação pode resultar em valores fora do intervalo definido pelo mínimo e máximo dos dados. Em vista disso, foi feita uma correção de extrapolação após a desacumulação. A correção de extrapolação muda os valores simulados maiores do que o máximo dos dados para o máximo dos dados. De maneira similar, a correção de extrapolação muda os valores simulados menores do que o mínimo dos dados para o mínimo dos dados.

8.4 Resultados

8.4.1 Correção de extrapolação no método indireto

As figura 60a-b mostram os *boxplots* dos dados e da primeira realização para as variáveis AT e ST. No caso do método indireto, os *boxplots* são das simulações sem a correção de extrapolação. No caso da AT, o método indireto gerou um máximo acima de 400% (figura 60a). No caso da ST, o máximo obtido no método indireto foi acima de 80%, que está bem acima do máximo dos dados (figura 60b). Não houve extrapolação das variáveis AT e ST na simulação feita pelo método direto. Isso ocorre porque a

transformação *normal score,* presente na etapa de PPMT, foi aplicada nas variáveis originais no método direto. A transformação *normal score* inversa evita a extrapolação das variáveis originais no método direto. Por outro lado, no método indireto a transformação *normal score* foi aplicada sobre as variáveis acumuladas.

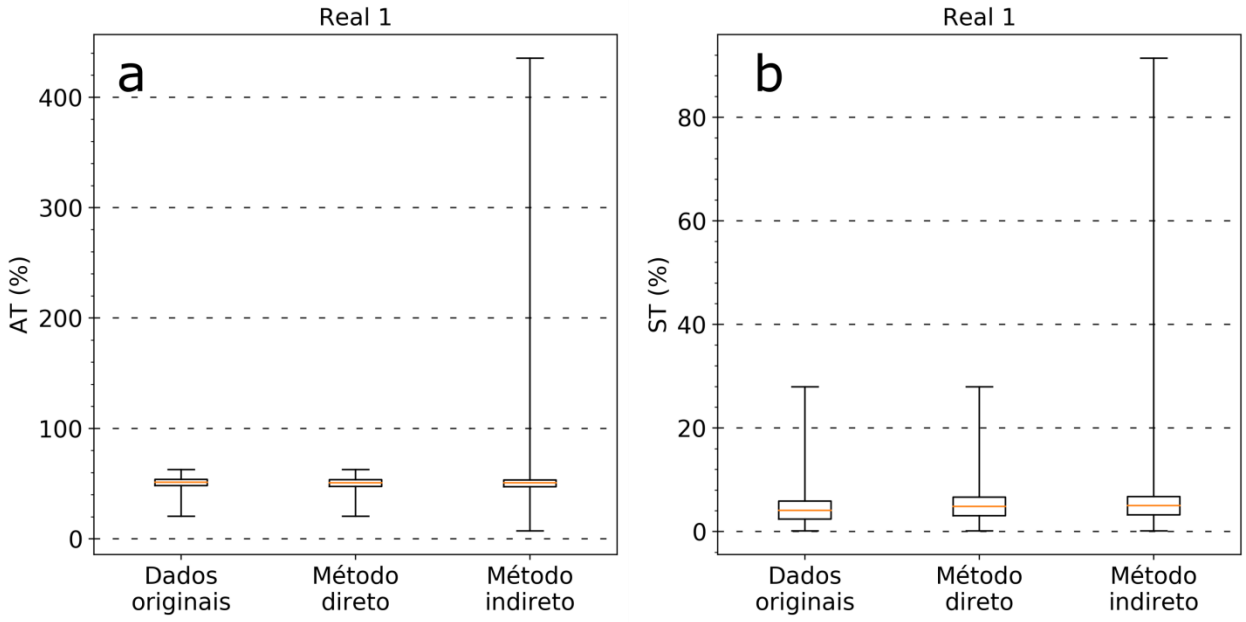

Figura 60: Boxplots dos dados originais e primeira realização obtida com os métodos direto e indireto para a variável AT (a) e ST (b).

As figuras 61a-b mostram os histogramas das proporções de nós de grid corrigidos no método indireto para as 20 realizações para as variáveis AT e ST. As proporções de valores corrigidos foram baixas (abaixo de 0.1%) para as duas variáveis.

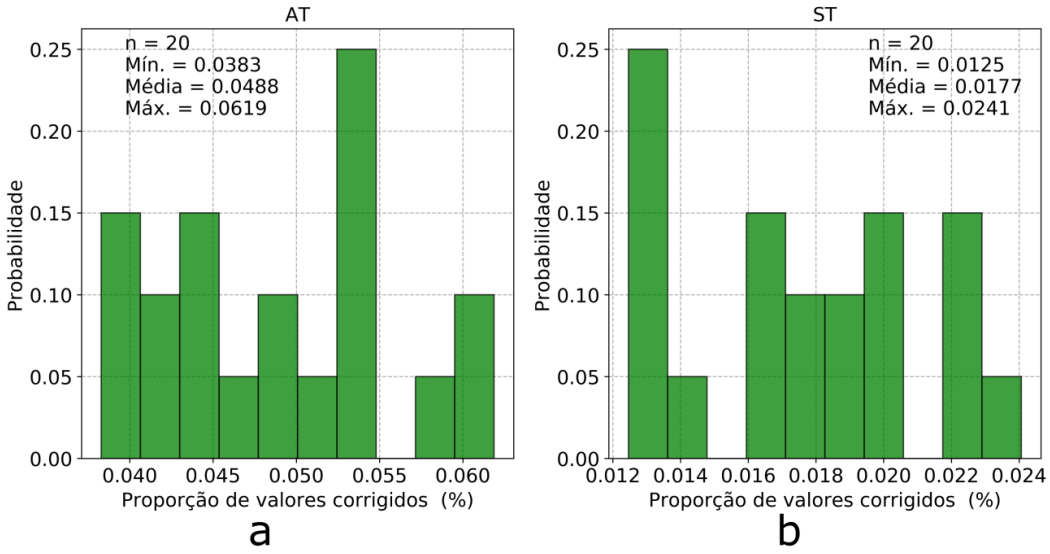

Figura 61: Histograma da proporção de valores corrigidos na correção de extrapolação para as variáveis AT (a) e ST (b).

8.4.2 Reprodução dos histogramas

A figura 62 mostra a reprodução dos histogramas para os métodos direto e indireto. Os dois métodos reproduziram bem os histogramas das variáveis originais. No caso da ST, as simulações utilizando o método direto tiveram média e desvio padrão mais próximos da média e desvio padrão dos dados desagrupados. Os histogramas cumulativos das simulações são similares entre si para os dois métodos (figura 62). Essa similitude é resultado da grande quantidade de dados e indica uma baixa incerteza global.

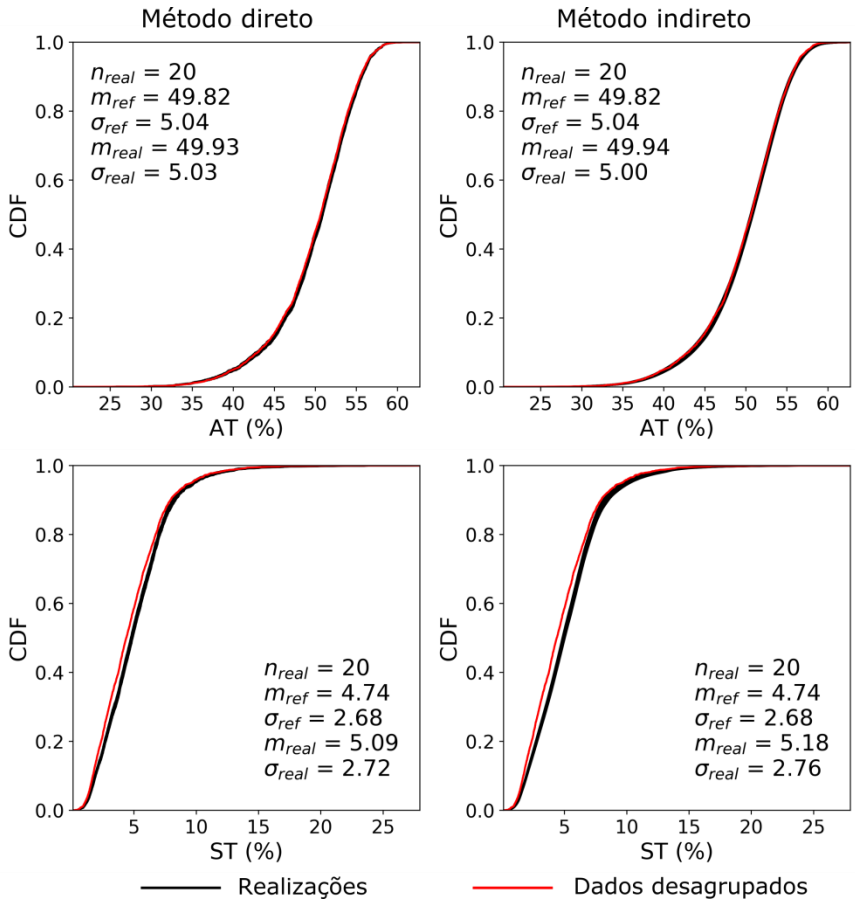

Figura 62: Reprodução dos histogramas.

A reprodução do histograma das variáveis originais pelo método indireto foi melhor do que o esperado, pois o método indireto não utiliza o histograma das variáveis originais na transformação PPMT. O método indireto utiliza o histograma das variáveis acumuladas e da variável ponderadora na transformação PPMT.

A reprodução do histograma das variáveis originais no método indireto foi favorecida pela grande quantidade de dados. Outra razão para a boa reprodução do histograma no método indireto é o fato da transformação PPMT ter preservado as relações entre as variáveis acumuladas (ATA e STA) e a variável ponderadora (RC). As figuras 63a-b mostram o gráfico de dispersão entre as variáveis RC e ATA dos dados originais (figura 63a) e da primeira realização feita com o método indireto (figura 63b). As simulações das variáveis acumuladas ficaram consistentes com as simulações da variável ponderadora. Dessa forma, a etapa de desacumulação resultou em simulações de teores que reproduziram os histogramas dos teores.

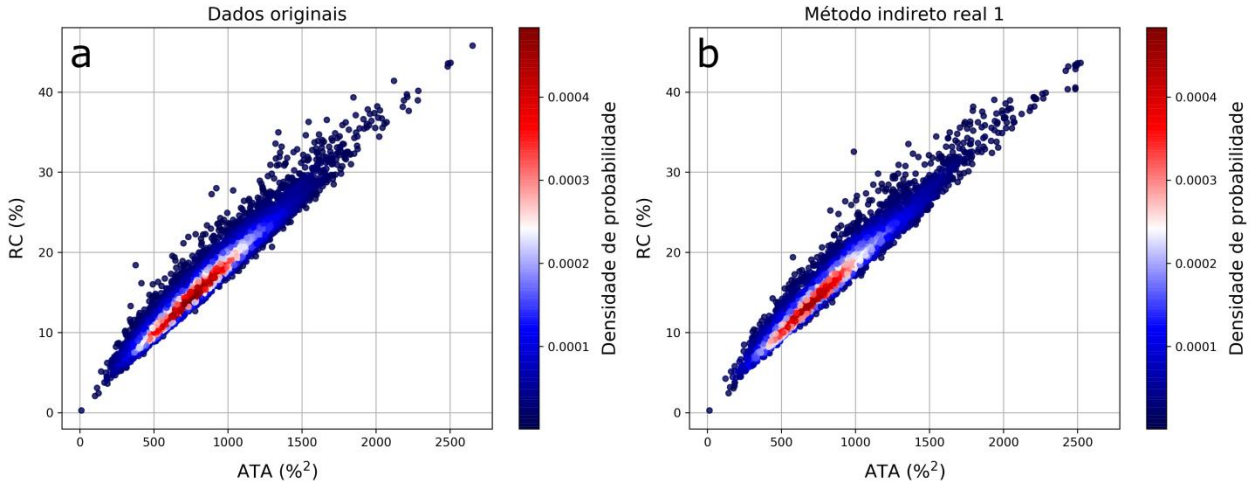

Figura 63: Gráfico de dispersão entre RC e ATA dos dados originais (a) e da primeira realização feita com o método indireto (b).

Para testar as metodologias em um cenário com pouca informação, foi feita uma realização não condicional utilizando SGS com os métodos direto e indireto. A simulação não condicional não utiliza as amostras no sistema de krigagem. Apenas nós previamente simulados são utilizados no sistema de krigagem. As amostras foram usadas apenas na transformação PPMT. A figura 64 mostra a reprodução do histograma da realização não condicional feita com os métodos direto e indireto. As realizações não condicionais feitas pelos dois métodos reproduziram o histograma dos dados desagrupados. O resultado mostra a robustez das metodologias em um cenário com menos informação.

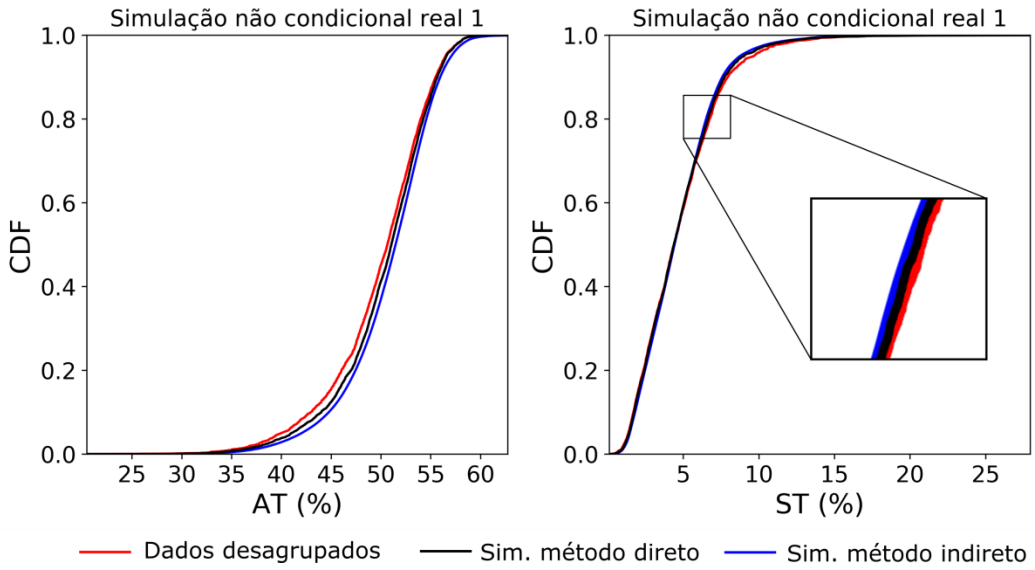

Figura 64: Reprodução do histograma de uma realização não condicional feita com o método direto e indireto.

8.4.3 Reprodução dos variogramas

A figura 65 mostra a reprodução do variograma da variável AT nas direções horizontal e vertical para os dois métodos de simulação (método direto e indireto). Para os dois métodos, o patamar das realizações ficou acima do patamar dos dados na direção horizontal. Na direção vertical, o método direto reproduziu melhor o variograma da variável AT.

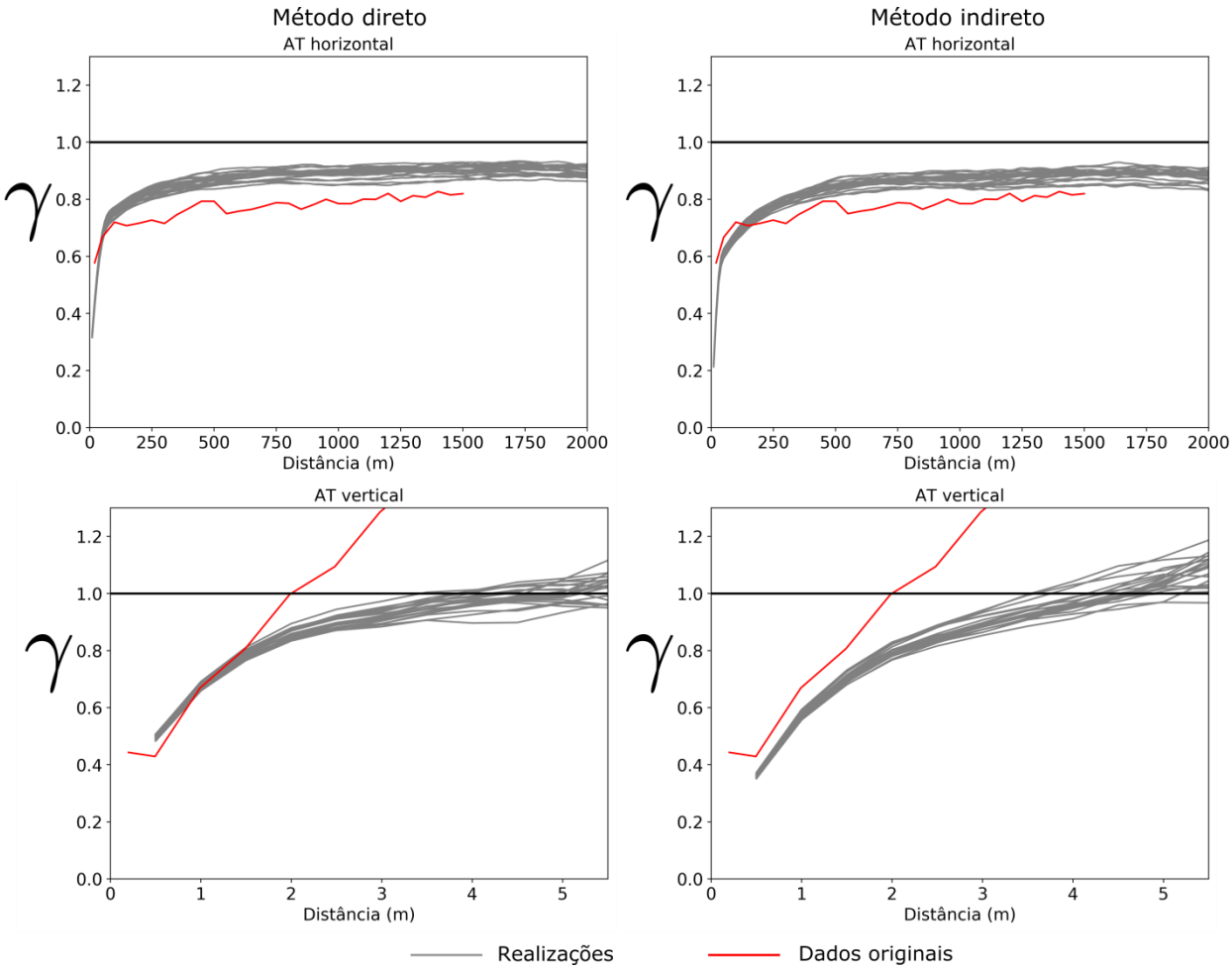

Figura 65: Reprodução do variograma da variável AT.

A figura 66 mostra a reprodução do variograma da variável ST nas direções horizontal e vertical para os dois métodos de simulação (método direto e indireto). Na direção horizontal, os dois métodos reproduziram o variograma na estrutura de curto alcance. Na estrutura de longo alcance, as realizações ficaram espacialmente mais descontínuas do que os dados. Na direção vertical, as realizações ficaram espacialmente mais descontínuas do que os dados. Além disso, o método indireto reproduziu melhor o variograma da variável ST na direção vertical.

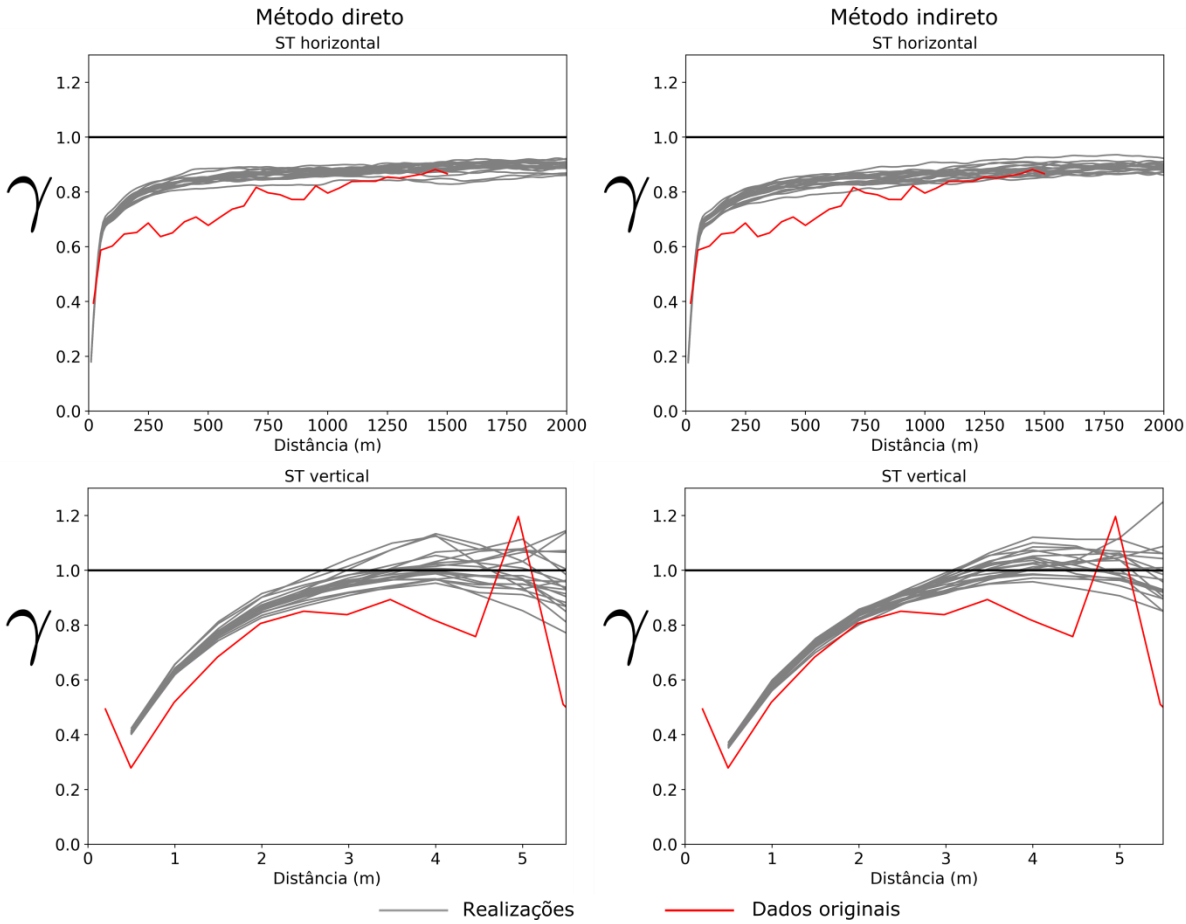

Figura 66: Reprodução do variograma para a variável ST.

8.4.4 Reprodução das relações bivariadas

As figuras 67a-c mostram o gráfico de dispersão entre AT e ST dos dados originais e das simulações para os dois métodos (método direto e indireto). A simulação pelo método indireto gerou pontos fora da distribuição bivariada (veja os círculos na figura 67c). Ao contrário da simulação pelo método indireto (círculo vermelho na figura 67c), os dados originais não têm pontos na região em que AT é maior do que 50% e ST é maior do que 15%). De maneira similar, a simulação pelo método indireto gerou valores na região em que AT é menor do que 30% e ST é aproximadamente 10% (círculo azul na figura 67c). Essa região não possui pontos no gráfico de dispersão dos dados originais (figura 67a).

Um dos motivos dos artefatos é que a transformação PPMT foi feita sobre as variáveis acumuladas no método indireto. No método direto, a transformação PPMT foi

feita sobre as variáveis originais. Como resultado, o método direto reproduziu melhor a relação bivariada das variáveis originais. Além disso, o método indireto fez a transformação PPMT de três variáveis (AT, ST e RC) enquanto que o método direto fez a transformação PPMT de duas variáveis (AT e ST). A transformação PPMT é mais eficaz quando a razão entre a quantidade de dados e o número de variáveis é maior (Barnett *et al.*, 2016).

Outro motivo para a presença de artefatos no método indireto é a extrapolação causada na etapa de desacumulação. Quando o valor simulado da variável ponderadora é próximo de zero, os valores simulados tendem a ser maior do que o máximo dos dados para mais de uma variável. A correção de extrapolação faz com que os valores simulados que extrapolaram recebam o máximo das variáveis. Essa correção de extrapolação para várias variáveis simultaneamente resulta em artefatos no gráfico de dispersão (círculo vermelho na figura 67c).

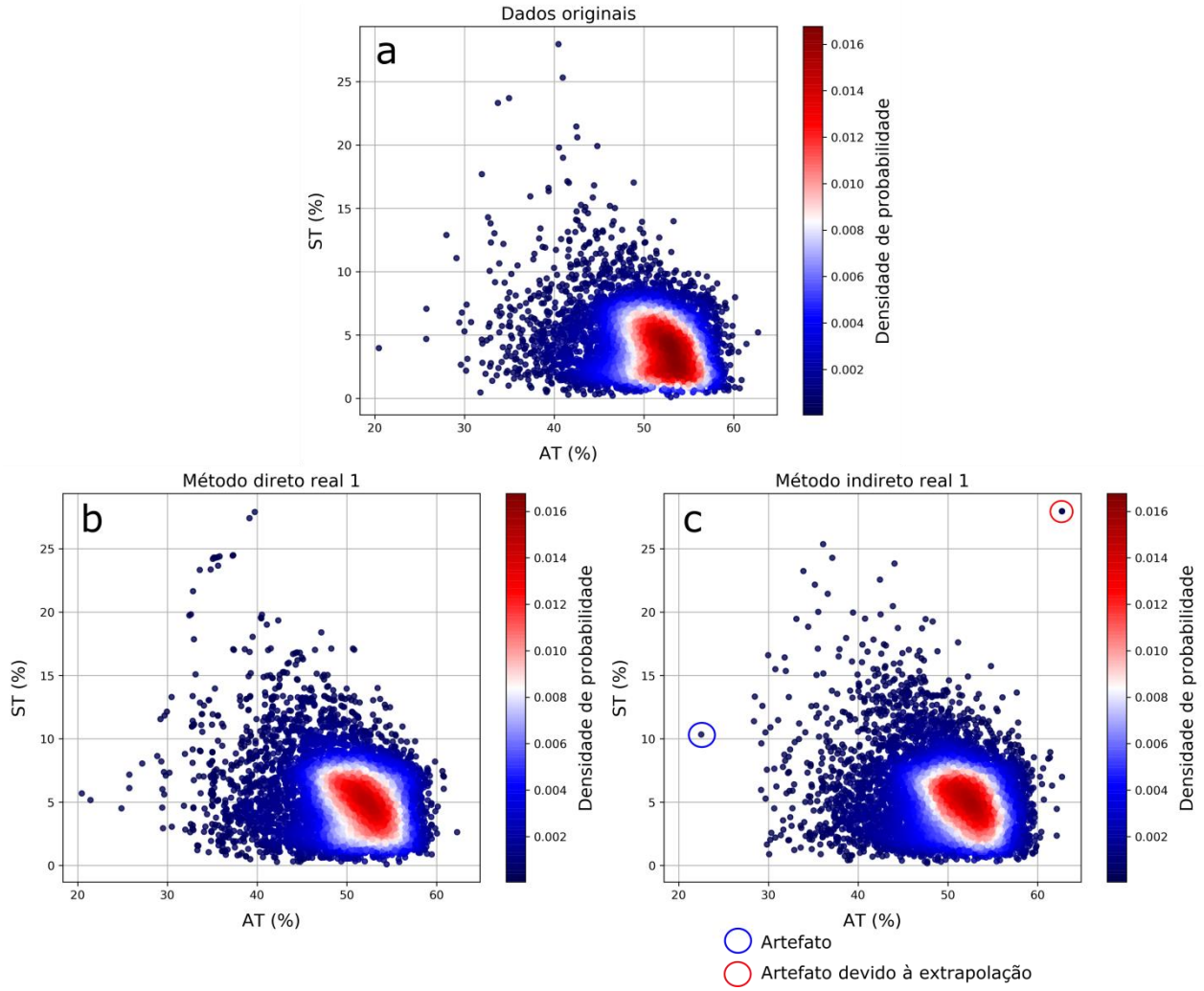

Figura 67: Gráfico de dispersão entre as variáveis AT e ST dos dados (a) e da primeira realização feita com o método direto (b) e indireto (c).

A piora na reprodução da relação entre AT e ST causada pelo método indireto aumenta a chance das simulações violarem restrições de soma. As figuras 68a-b mostram os histogramas da soma das variáveis AT e ST da primeira realização obtida com os métodos direto e indireto. O máximo da soma foi maior para o método indireto (figura 68b).

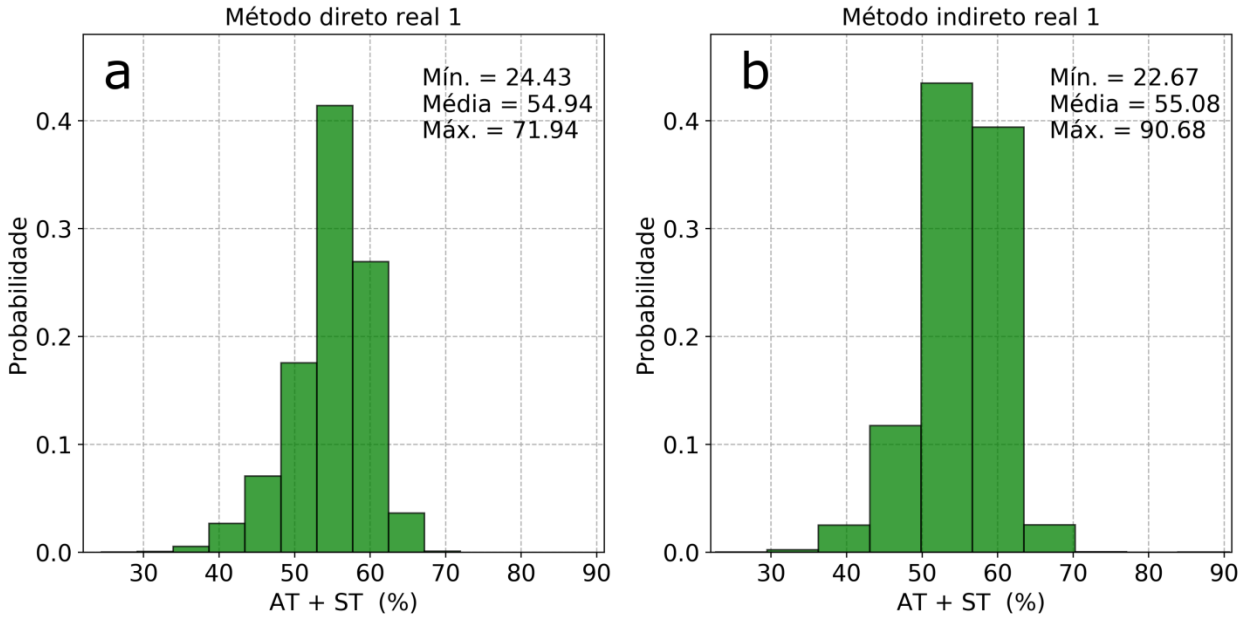

Figura 68: Histograma das soma das variáveis AT e ST da primeira realização feita com o método direto (a) e indireto (b).

A figura 69 mostra a densidade de probabilidade da estatística $D90$ calculada sobre as 20 realizações para os dois métodos. A estatística $D90$ mede a diferença entre a distribuição bivariada de AT e ST das simulações e dos dados originais. A estatística $D90$ para o método indireto foi em média menor do que aquela obtida pelo método indireto. Esse resultado mostra que o método direto reproduziu melhor a distribuição bivariada entre AT e ST.

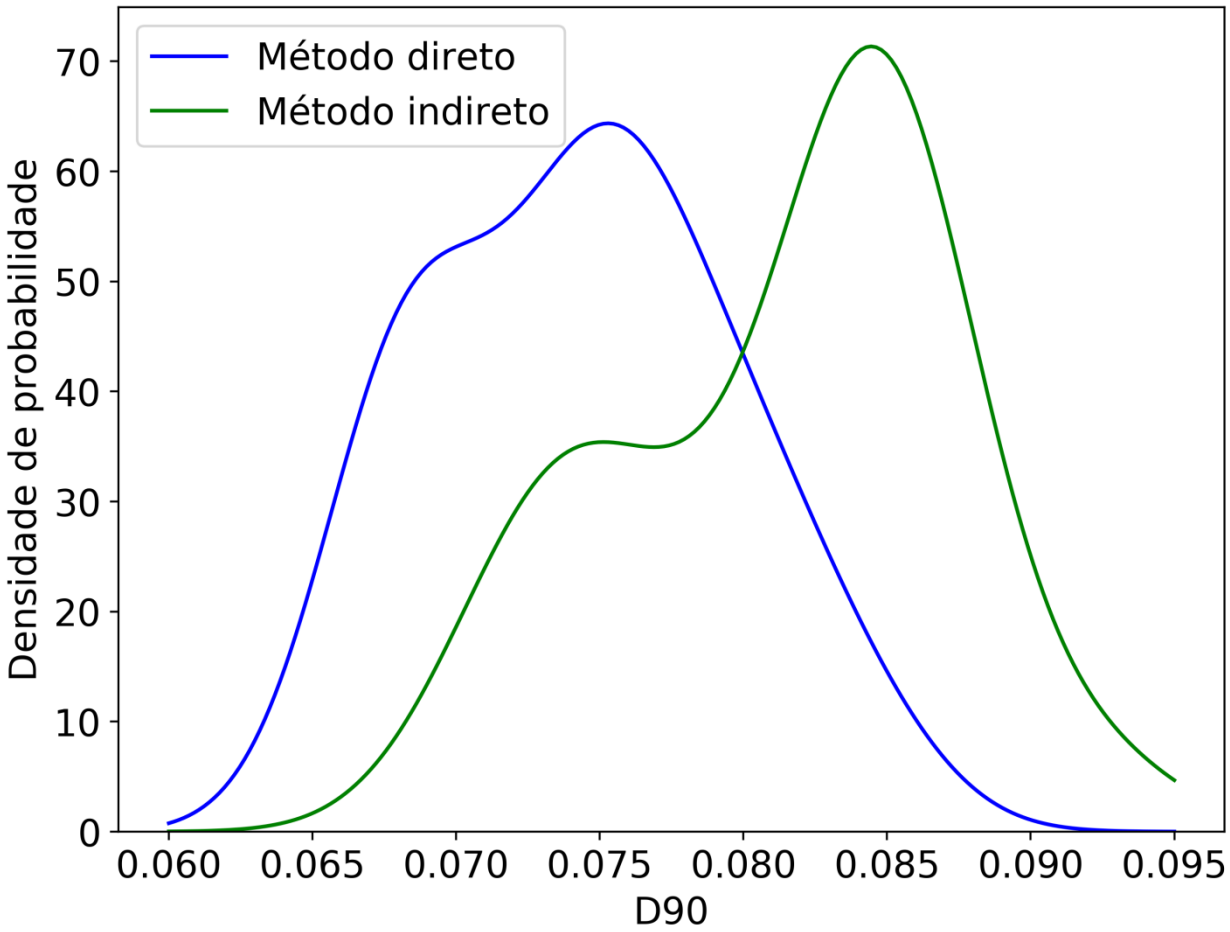

Figura 69: Densidade de probabilidade da estatística D90 calculada sobre as 20 realizações para os dois métodos.

8.5 Observações

Esse capítulo comparou o método direto e indireto para a simulação geoestatística de múltiplas variáveis. O banco de dados é proveniente de um depósito de bauxita. As variáveis Alumina Total (AT) e Sílica Total (ST) foram consideradas. Foram avaliadas a reprodução dos histogramas, variogramas e relações bivariadas das simulações.

Os métodos direto e indireto de simulação reproduziram bem os histogramas das variáveis originais. Os variogramas na direção horizontal foram semelhantes para os dois métodos de simulação. Na direção vertical, o método direto reproduziu melhor o

variograma da variável AT enquanto que o método indireto reproduziu melhor o variograma da variável ST.

O método direto reproduziu melhor a relação bivariada entre AT e ST. A estatística $D90$ foi calculada para medir a diferença entre a distribuição bivariada dos dados e das simulações. As simulações pelo método direto tiveram, em média, uma estatística $D90$ menor do que as simulações obtidas com o método indireto. Diferente do método indireto, o método direto não apresentou artefatos no gráfico de dispersão entre AT e ST. No caso de simulação geoestatística multivariada, o método direto deve ser escolhido.

9 Conclusões

A tese abordou três assuntos: (1) uso de amostras de diferente suporte em geoestatística, (2) simulação geoestatística multivariada com restrições e (3) verificação da distribuição multivariada. A seção 9.1 mostra as contribuições da tese em cada assunto. A seção 9.2 alerta sobre as limitações dos estudos desenvolvidos. Por último, a seção 9.3 lista sugestões para trabalhos futuros.

9.1 Contribuições da tese

9.1.1 Amostras de diferente suporte

Krigagem com amostras de diferente suporte

A tese mostrou o uso de covariâncias médias no sistema de krigagem para incorporar amostras de diferente comprimento na estimativa. A técnica foi aplicada em um estudo de caso de bauxita. Além disso, a técnica foi comparada com a krigagem utilizando covariâncias ponto-a-ponto entre as amostras. A krigagem com covariâncias médias resultou em estimativas mais precisas e acuradas. Isso mostra que a técnica é eficaz para integrar dados de diferente suporte. Com essa técnica, o usuário não precisa criar informação ao quebrar uma amostra grande em amostras pequenas de igual teor ou trabalhar com um subconjunto dos dados que tenha suporte parecido.

A tese mostrou também o efeito do variograma nos pesos de estimativa na krigagem com covariâncias médias entre as amostras. Quanto mais espacialmente descontínuo a curtas distâncias é o fenômeno, maior é o peso para amostras longas. Particularmente, o peso das estimativas para amostras longas é sensível ao efeito pepita. Para fenômenos espacialmente contínuos (alcance longo em relação ao tamanho das amostras e baixo efeito pepita), o peso para amostras de diferentes comprimentos é bastante similar. Nesse caso, a krigagem com covariâncias médias é similar à krigagem utilizando covariâncias ponto-a-ponto entre as amostras. Esse resultado mostra que o geomodelador pode, através do variograma, analisar o quanto o

uso de covariâncias médias no sistema de krigagem irá afetar as estimativas. Se o fenômeno tiver baixo efeito pepita e longo alcance em relação ao suporte das amostras, as estimativas serão pouco afetadas pelo uso de covariâncias médias no sistema de krigagem.

Variograma com amostras de diferente suporte

A krigagem com amostras de diferente suporte necessita de um modelo de variograma em suporte pontual. A obtenção de um modelo de variograma em suporte pontual, a partir de amostras de diversos tamanhos, é chamada de deconvolução do variograma. Nessa tese, foi desenvolvido um conjunto de *softwares* para a deconvolução do variograma junto com a documentação pertinente. Os *softwares* permitem utilizar amostras de suporte diferente com tamanhos variados, ao contrário do *software* publicado por Babak *et al.* (2013), que trabalha apenas com blocos regulares. Especificamente, o *software Block_Variogram* foi desenvolvido para calcular o variograma experimental para amostras de diferente suporte e o *software Block_Vmodel* faz a regularização do variograma. A tese mostra a deconvolução do variograma para amostras de diferente comprimento em um depósito de bauxita.

Método indireto

O capitulo 2 mostra o cálculo da variância do erro de estimativa para o método indireto quando o mesmo modelo de variograma e estratégia de busca é utilizado para estimar a variável acumulação e a variável ponderadora. Essa variância de estimativa pode ser utilizada para auxiliar a classificação de recursos minerais.

O capítulo 4 ilustra o que acontece com os pesos de estimativa no método indireto quando o mesmo variograma é utilizado para estimar a variável acumulação e a variável ponderadora. Na maioria dos casos, o método indireto resulta em maior peso para as amostras de maior suporte do que a krigagem com covariâncias médias. A exceção ocorre quando o modelo de variograma é efeito pepita puro. Sob um ponto de

vista prático, o método indireto reforça a influência de amostras de maior suporte nas estimativas.

As estimativas pelo método indireto foram comparadas com as estimativas obtidas com krigagem com amostras de diferente suporte no caso de amostras de diferente comprimento. As estimativas feitas por krigagem com amostras de diferente suporte foram mais precisas e acuradas do que as estimativas feitas pelo método indireto. O estudo de caso foi aplicado para uma variável com uma distribuição praticamente simétrica com um coeficiente de variação baixo.

O capítulo 8 mostra um comparativo entre o método direto e indireto no caso de simulação geoestatística multivariada de teores analisados por faixa granulométrica. A reprodução dos variogramas e histogramas foram semelhantes para os dois métodos. Por outro lado, a reprodução da distribuição multivariada foi melhor para o método direto. Os gráficos de dispersão das simulações feitas pelo método direto não apresentaram artefatos, ao contrário das simulações feitas pelo método indireto.

9.1.2 Simulação geoestatística multivariada com restrições

O capítulo 6 mostrou um comparativo de *workflows* para simulação geoestatística multivariada com restrições de soma e fração. A diferença dos *workflows* consiste na etapa de transformação dos dados antes da simulação geoestatística. Para cada tipo de restrição há uma transformação. O comparativo foi feito por simulação de Monte Carlo. A tese mostra a aplicação da simulação de Monte Carlo para testar transformações multivariadas de maneira rápida. É mais rápido simular por Monte Carlo do que simular um grid inteiro com simulação sequencial Gaussiana, por exemplo.

A principal contribuição da tese é uma metodologia para a simulação geoestatística multivariada com restrições de soma e fração. As razões de fração foram utilizadas para considerar as restrições de fração. A transformação PPMT foi utilizada para considerar as relações entre as variáveis. As variáveis transformadas foram simuladas por simulação sequencial Gaussiana e retro-transformadas. As devidas correções foram feitas para evitar que houvesse problemas de extrapolação ou de violação da restrição de soma.

9.1.3 Verificação da distribuição multivariada

A tese propõe o uso da estatística $D90$, que é o 90º percentil da diferença absoluta entre cdfs multivariadas, para medir a diferença entre distribuições multivariadas. Os valores dos dados são utilizados como limiares para o cálculo das cdfs multivariadas. A estatística $D90$ foi eficaz para detectar erros de viés e precisão em bancos de dados com relações lineares e não lineares.

Além disso, a tese mostra a aplicação da estatística $D90$ para comparar o método direto e indireto na simulação geoestatística multivariada de teores analisados por faixa granulométrica. A estatística $D90$ permitiu verificar que as simulações feitas pelo método direto reproduzem, em média, melhor a distribuição multivariada dos dados.

9.2 Limitações das técnicas

9.2.1 Krigagem com amostras de diferente suporte

A krigagem com amostras em diferente suporte requer um modelo de variograma em suporte pontual, que muitas vezes é difícil de ser inferido. A deconvolução do variograma e os pesos na krigagem com amostras de diferente suporte são sensíveis ao efeito pepita. Quando o efeito pepita não pode ser inferido de maneira adequada, as estimativas feitas por krigagem com amostras de diferente suporte podem ser questionadas.

A krigagem com amostras de diferente suporte assume que as amostras de maior suporte representam o valor médio ao longo de um volume. Dessa forma, não é correto fazer uma transformação não linear dessas amostras. Isso limita a aplicação da transformação PPMT combinada com técnicas Gaussianas de simulação geoestatística. Uma alternativa é obter informações em suporte pontual a partir das amostras de diferente suporte (*downscaling*) utilizando simulação sequencial direta (Soares, 2010). Depois que as amostras estão em suporte pontual, é possível utilizar-se a transformação PPMT. Entretanto, o processo de *downscaling* no caso de múltiplas

variáveis exige a inferência da distribuição multivariada em suporte de ponto, que não é trivial. Além disso, o processo de *downscaling* não garante que as amostras em suporte de ponto vão respeitar as restrições de soma e fração.

9.2.2 Simulação geoestatística multivariada com restrições

As transformações utilizadas no capítulo 6 têm vantagens e desvantagens. Cada transformação funciona bem para o propósito que ela foi desenvolvida. As razões de fração respeitam a restrição de fração, mas podem causar extrapolação dos valores das variáveis originais. De maneira similar, as razões A e razões U respeitam a restrição de soma, mas podem causar extrapolação dos valores das variáveis originais. Por outro lado, a transformação PPMT convencional reproduz as relações entre as variáveis e não causa extrapolação. Entretanto, a transformação PPMT pode resultar em nós de grid que não respeitam as restrições de soma e fração. A transformação PPMT começando com a projeção ortogonal à restrição de soma respeitou a restrição de soma, mas causou bastante extrapolação, incluindo a presença de valores negativos.

A metodologia proposta utilizou a transformação PPMT para lidar com a restrição de soma. No estudo de caso apresentado, a soma das variáveis com restrição de soma tem uma média de 80% no banco de dados original. Se a soma das variáveis com restrição de soma for mais próxima de 100%, as razões A, razões U ou razões logarítmicas podem ser mais adequadas para lidar com a restrição de soma.

O estudo de caso de simulação geoestatística multivariada com restrições apresentado no capítulo 6 mostrou problemas na reprodução do variograma. Na direção horizontal, houve problema na reprodução do variograma na estrutura de longo alcance. Na direção vertical, as realizações ficaram mais espacialmente descontínuas do que os dados originais. A transformação PPMT causa uma desestruturação das variáveis a pequenas distâncias e isso acaba afetando os variogramas (Barnett, 2015). Além disso, a simulação sequencial Gaussiana causa problemas na reprodução do variograma (Safikhani *et al.*, 2017).

9.2.3 Verificação da distribuição multivariada

Uma das limitações da estatística $D90$ é que ela exige uma grande quantidade de dados no caso de muitas variáveis. À medida que o número de variáveis aumenta, a estatística $D90$ fica muito próxima de zero, pois todos os pontos dos dados se situam na cauda inferior da distribuição multivariada. Uma alternativa é checar a distribuição multivariada em menores dimensões. Por exemplo, considere um total de 7 variáveis. O espaço de duas dimensões corresponde ao total de combinações de 2 variáveis em um total de 7, que é 21. O geomodelador pode calcular a estatística $D90$ sobre todas as 21 combinações e obter a média. Além disso, a distribuição multivariada para dimensões maiores do que 2 fica difícil de ser visualizada. Em 2 dimensões, a distribuição multivariada pode ser visualizada através de gráficos de dispersão.

Outra limitação da estatística $D90$ é que ela demanda bastante recurso computacional no caso de grids grandes. A estatística $D90$ necessita conhecer a proporção de nós de grid que seja menor ou igual a cada observação (dado amostral). Esse processo se repete para cada realização. Considere a seguinte situação: um banco de dados com N amostras, um grid de simulação de M nós e um conjunto com K realizações. O cálculo da estatística $D90$ para todas as realizações exige um laço com N*M*K iterações.

9.3 Trabalhos futuros

9.3.1 Amostras de diferente suporte

Em relação ao uso de amostras de diferente suporte, as sugestões para trabalhos futuros são as seguintes:

- Realizar um estudo de caso de krigagem com amostras de diferente suporte para um banco de dados multivariado. O banco de dados multivariado aumenta a complexidade da deconvolução do variograma, pois é necessário fazer a deconvolução dos variogramas cruzados;

- Adaptar os *softwares Block_Variogram* e *Block_Vmodel* para poder incorporar múltiplas variáveis. Atualmente, esses *softwares* só consideram uma variável.

9.3.2 Simulação geoestatística multivariada com restrições

As sugestões de trabalhos futuros para o tema de simulação geoestatística multivariada com restrições são as seguintes:

- Repetir o estudo de caso de simulação geoestatística apresentado no capítulo 6 utilizando simulação por bandas rotativas (*Turning Bands Simulation* – TBS) em vez de simulação sequencial Gaussiana (*Sequential Gaussian Simulation* – SGS). O estudo de caso do capítulo 6 apresentou problemas na reprodução do variograma. A TBS reproduz melhor os variogramas do que a SGS (Paravarzar *et al.*, 2015). Além disso, é interessante incorporar a transformação MAF após a transformação PPMT para remover alguma correlação espacial remanescente das variáveis transformadas PPMT. Barnett *et al*. (2014) melhoraram a reprodução do variograma com a utilização da transformação MAF após a transformação PPMT.

- Avaliar a metodologia para um banco de dados em que a soma das variáveis com restrição de soma é mais próxima de 100%. No estudo de caso realizado, a soma das variáveis com restrição de soma tem uma média de 80% aproximadamente. Nesse caso, o uso das razões U para considerar a restrição de soma pode ser mais adequado do que utilizar diretamente a transformação PPMT.

- Desenvolver uma transformação que torne as variáveis multi-Gaussianas e independentes e que considera restrições de soma, restrições de fração e que não causa extrapolação. A tese avaliou uma versão modificada da transformação PPMT que começa com a projeção ortogonal à restrição de soma. Essa transformação respeitou a restrição de soma, mas causou problemas de extrapolação.

9.3.3 Verificação da distribuição multivariada

A pesquisa sobre a verificação da distribuição multivariada pode se beneficiar dos seguintes estudos:

- Aplicar a estatística $D90$ para comparar algoritmos de cosimulação geoestatística em um estudo de caso com maior número de variáveis. A estatística $D90$ foi utilizada no capítulo 8 para comparar a cosimulação de 2 variáveis apenas.

- Melhorar o desempenho computacional da estatística $D90$ para avaliar distribuições multivariadas com grande número de variáveis.

- Fazer um estudo de sensibilidade calculando a estatística $D90$ utilizando um subconjunto dos nós de grid e dos dados escolhidos aleatoriamente. O impacto de utilizar um subconjunto dos dados e nós de grid na estatística $D90$ e na velocidade de processamento pode ser analisado.

Referências

Aitchison, J., 1986. *The statistical analysis of compositional data*. Chapman & Hall, London.

Almeida, A. S. e Journel, A. G., 1994. Joint simulation of multiple variables with a Markov-type coregionalization model. *Mathematical Geology*, 26, (5), 565–588.

Babak, O., Cuba, M. A., e Leuangthong, O., 2013. Direct upscaling of semivariograms and cross semivariograms for scale-consistent geomodeling. *Transactions of the Society for Mining, Metallurgy, and Exploration*, *334,* (1), 544-552.

Barnett, R. M., 2015. *Managing complex multivariate relations in the presence of incomplete spatial data*. Tese de doutorado, University of Alberta.

Barnett, R. M. e Deutsch, C. V., 2012. Practical implementation of non-linear transforms for modeling geometallurgical variables. Em *Geostatistics Oslo 2012* (pag. 409-422). Springer Netherlands.

Barnett, R. M., Manchuk, J. G., e Deutsch, C. V., 2014. Projection pursuit multivariate transform. *Mathematical Geosciences*, 46, (3), 337-359.

Barnett, R. M., Manchuk, J. G., e Deutsch, C. V., 2016. The projection-pursuit multivariate transform for improved continuous variable modeling. *SPE Journal*.

Bassani, M.A.A., Machado, P.L., Costa, J.F.C.L., e Rubio, R.J.H., 2014. Using production data to improve grade estimation in underground mining. *Applied Earth Science: Transactions of the Institute for Mineralogy and Metallurgy Section B*, 122, (4), 243-248.

Bertoli, O., Vann, J., e Dunham, S., 2003. Two-dimensional geostatistical methods—theory, practice and a case study from the 1A shoot nickel deposit, Leinster, Western Australia. Em *Proceedings of the 5th international mining geology conference. The Australian Institute of Mining and Metallurgy, Melbourne* (pag. 189-195).

Boisvert, J. B., Rossi, M. E., Ehrig, K., e Deutsch, C. V., 2013. Geometallurgical modeling at Olympic Dam Mine, South Australia. *Mathematical Geosciences*, 45, (8), 901–925.

Dagbert, M., 2001. Comments on the paper “The estimation of mineralized veins: A comparative study of direct and indirect approaches” by D. Marcotte and A. Boucher. *Exploration and Mining Geology*, 10, pag. 243-244.

Desbarats, A. J., e Dimitrakopoulos, R., 2000. Geostatistical simulation of regionalized pore-size distributions using min/max autocorrelation factors. *Mathematical Geology*, 32*,* (8), 919-942.

Deutsch, C. V. e Journel, A. G., 1998. *Geostatistical software library and user’s guide*. Oxford University Press, New York.

Deutsch, C.V., Srinivasan, S., e Mo, Y., 1996. Geostatistical reservoir modeling accounting for precision and scale of seismic data. *SPE Annual Technical Conference*, Denver, Colorado, U.S.A, 6-9 October 1996.

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

Emery, X., 2012. Co-simulating total and soluble copper grades in an oxide ore deposit. *Mathematical Geosciences*, 44, (1), 27-46.

Friedman, J. H., 1987. Exploratory projection pursuit. *Journal of the American statistical association*, 82, (397), 249-266.

Froidevaux, R., 1993. Probability field simulation. Em *Geostatistics Tróia'92* (pag. 73-83). Springer, Dordrecht.

Gómez-Hernández, J. J., e Srivastava, R. M., 1990. ISIM3D: An ANSI-C three-dimensional multiple indicator conditional simulation program. *Computers & Geosciences*, 16, (4), 395-440.

Goovaerts, P., 1993. Spatial orthogonality of the principal components computed from coregionalized variables. *Mathematical Geology*, *25*(3), 281-302.

Goovaerts, P., 1997. Geostatistics for natural resource evaluation; Oxford, Oxford University Press.

Goovaerts, P., 2008. Kriging and semivariogram deconvolution in the presence of irregular geographical units. *Mathematical Geosciences*, 40, (1), 101-128.

Goovaerts, P., 2010. Combining areal and point data in geostatistical interpolation: applications to soil science and medical geography. *Mathematical and Geosciences*, 42, (5), 535-554.

Gotway, C. A. e Young, L. J., 2002. Combining incompatible spatial data. *Journal of the American Statistical Association*, 97, (458), 632–648.

Hansen, T.M. e Mosegaard, K., 2008. Visim: sequential simulation for linear inverse problems. *Computers & Geosciences*, 34, (1), 53–76.

Horta, A. e Soares, A., 2010. Direct sequential co-simulation with joint probability distributions. *Mathematical Geosciences*, 42, (3), 269–292.

Hosseini, S. A. e Asghari, O., 2015. Simulation of geometallurgical variables through stepwise conditional transformation in Sungun copper deposit, Iran. *Arabian Journal of Geosciences*, 8, (6), 3821-3831.

Hotelling, H., 1933. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24, (6), 417–441.

Isaaks, E. H., 1990. The application of Monte Carlo methods to the analysis of spatially correlated data. Tese de doutorado, Stanford University

Isaaks, H. E. e Srivastava, M. R., 1989. *An introduction to applied geostatistics*. Oxford, Oxford University Press.

Journel, A.G., 1986. Geostatistics: models and tools for the earth sciences. *Mathematical Geology*, 18, 119–140.

Journel, A. G. e Huijbregts, C. J., 1978. *Mining geostatistics*. Academic press.

Krige, D. G., 1978. Lognormal–de Wijsian geostatistics for ore evaluation; Johannesburg. South African Institute of Mining and Metallurgy Monograph Series.

Kupfersberger, H., Deutsch, C. V., e Journel, A. G., 1998. Deriving constraints on small-scale variograms due to variograms of large-scale data. *Mathematical geology*, *30,* (7), 837-852.

Kyriakidis, P. C., 2004. A geostatistical framework for area-to-point spatial interpolation. *Geographical Analysis*, 36*,* (3), 259-289.

Leuangthong, O. e Deutsch, C. V., 2003. Stepwise conditional transformation for simulation of multiple variables. *Mathematical Geology*, 35, (2), 155-173.

Leuangthong, O., McLennan, J. A., e Deutsch, C. V., 2004. Minimum acceptance criteria for geostatistical realizations. *Natural Resources Research*, 13, (3), 131-141.

Liu, Y. e Journel, A. G., 2009. A package for geostatistical integration of coarse and fine scale data. *Computers & Geosciences*, 35, (3), 527-547.

Manchuk, J. G., e Deutsch, C. V., 2012. A flexible sequential Gaussian simulation program: USGSIM. *Computers & geosciences*, *41*, 208-216.

Manchuk, J. G., Barnett, R. M., e Deutsch, C. V., 2017. Reproduction of secondary data in projection pursuit transformation. *Stochastic Environmental Research and Risk Assessment*, 31, (10), 2585-2605.

Marcotte, D. e Boucher, A., 2001. The estimation of mineralized veins: a comparative study of direct and indirect approaches. *Exploration and Mining Geology*, 10, (3), 235-242.

Marques, D. M., Rubio, R. H., Costa, J. F. C. L., e Silva, E. M. A. D., 2014. The effect of accumulation in 2D estimates in phosphatic ore. *Rem: Revista Escola de Minas*, 67, (4), 431-437.

Massey Jr, F. J., 1951. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American statistical Association*, 46, (253), 68-78.

Mery, N., Emery, X., Cáceres, A., Ribeiro, D., e Cunha, E., 2017. Geostatistical modeling of the geological uncertainty in an iron ore deposit. *Ore Geology Reviews*, *88*, 336-351.

Oz, B., V. Deutsch, C., e Frykman, P., 2002. A visualbasic program for histogram and variogram scaling. *Computers and Geosciences*, 28, (1), 21–31.

Paravarzar, S., Emery, X., e Madani, N., 2015. Comparing sequential Gaussian and turning bands algorithms for cosimulating grades in multi-element deposits. *Comptes Rendus Geoscience*, 347, (2), 84-93.

Pardo-Igúzquiza, E., Chica-Olmo, M., e Atkinson, P. M., 2006. Downscaling cokriging for image sharpening. *Remote Sensing of Environment*, 102, (1), 86-98.

Pawlowsky-Glahn, V. e Olea, R. A., 2004. *Geostatistical analysis of compositional data* (Vol. 7). Oxford University Press.

Poggio, L. e Gimona, A., 2013. Modelling high resolution RS data with the aid of coarse resolution data and ancillary data. *International Journal of Applied Earth Observation and Geoinformation*, 23, 360-371.

Pyrcz, M. J. e Deutsch, C. V., 2014. *Geostatistical reservoir modeling*. Oxford University Press.

Remy, N., Boucher, A., Wu, J., 2008. *Applied geostatistics with SGeMS: a users guide*. Cambridge University Press.

Rossi, M. E. e Deutsch, C. V., 2013. *Mineral resource estimation*. Springer Science & Business Media.

Roy, W., Butt, S. D., e Frempong, P. K., 2004. Geostatistical resource estimation for the Poura narrow-vein gold deposit. *CIM Bulletin,* 97, 1077, 47–51.

Safikhani, M., Asghari, O., e Emery, X., 2017. Assessing the accuracy of sequential gaussian simulation through statistical testing. *Stochastic Environmental Research and Risk Assessment*, 31, (2), 523-533.

Soares, A., 2001. Direct sequential simulation and cosimulation. *Mathematical Geology*, 33, (8), 911–926.

Switzer, P. e Green, A., 1984. Min/max autocorrelation factors for multivariate spatial imagery: Dept. *Stat., Stanford Univ., Stanford, CA, Tech. Rep*, *6*.

Tran, T. T., Deutsch, C. C. e Xie, Y. 2001. Direct geostatistical simulation with multiscale well, seismic, and production data. SPE Annual Technical Conference and Exhibition held in New Orleans, Louisiana, 30 September–3 October.

Yao, T. e Journel, A. G., 2000. Integrating seismic attribute maps and well logs for porosity modeling in a west Texas carbonate reservoir: addressing the scale and precision problem. *Journal of Petroleum Science and Engineering*, 28, (1), 65-79.

Zuñiga, R., Emery, X., 2010. Evaluating mineral resources in a narrow vein–type deposit. Em: MINIM 2010, *Proceedings of the 4th International Conference on Mining and Inovation*.

Apêndice A: *softwares*

A.1 Deconvolução do variograma

A.1.1 *Block_Variogram*

O *software Block_Variogram* segue um estilo GSLIB e é usado para calcular o variograma experimental de amostras de bloco. Basicamente, ele realiza o cálculo que o *gamv* (Deutsch e Journel, 1998) faz. A diferença é no formato do arquivo de entrada dos dados. Enquanto que o *gamv* calcula o variograma experimental para amostras de ponto, o *Block_Variogram* calcula o variograma experimental para amostras de bloco.

Parâmetros

A figura 70 mostra o arquivo de parâmetros do *software Block_Variogram*. A descrição linha a linha do arquivo de parâmetros é a seguinte:
Linha 5: arquivo de dados com as amostras de blocos;
Linha 6: colunas das coordenadas X, Y, Z, teor e *Block_ID* das amostras de bloco;
Linha 7: limites mínimo e máximo de corte dos dados;
Linha 8: arquivo de saída com os variogramas experimentais dos blocos;
Linha 9: número de direções;
Linha 10: azimute, tolerância do azimute, largura de banda horizontal, mergulho, tolerância do mergulho, largura de banda vertical e *rake*;
Linha 11: número de *lags*, tamanho de *lag* e tolerância do *lag*;
As linhas 12-13 definem a direção 2 e são análogas às linhas 10-11.
Linha 14: opção para estandardizar o patamar do variograma;
Linha 15: número de variogramas a serem calculados;
Linha 16: tipo de variograma.

```
1                Parameters for Block_Variogram
2                ******************************
3
4  START OF PARAMETERS:
5  ../data/blocks.dat                  - file with block data
6  1   2   3   4   5                   - columns for X, Y, Z, grade, block_id
7  -1.0e21     1.0e21                  - trimming limits
8  block_variogram.out                 - file for variogram output
9  2                                   - number of directions
10 0. 90. 9999.    0. 90. 9999.  0.00  - Dir 01: azm,atol,bandh,dip,dtol,bandv, rake
11 10  5.0  2.50                       - nlag,xlag,xtol
12 0. 20. 9999.    0. 90. 9999. 0.00   - Dir 02: azm,atol,bandh,dip,dtol,bandv, rake
13  8  7.0  4.0                        - nlag,xlag,xtol
14 0                                   - standardize sills? (0=no, 1=yes)
15 2                                   - number of variograms
16 1                                   - variogram type
17 2                                   - variogram type
```

Figura 70: Arquivo de parâmetros do software *Block_Variogram*.

Amostras de bloco

A figura 71 exemplifica o arquivo com as amostras de bloco. O arquivo segue o formato GeoEAS e possui pelo menos 5 colunas: X, Y, Z, Teor e *Block_ID*. Cada bloco é representado por um conjunto de pontos no espaço. As colunas X, Y e Z definem as coordenadas X, Y e Z dos pontos discretizantes. A coluna teor define a propriedade de interesse. Se os pontos discretizantes possuem teores diferentes, o teor médio é atribuído ao bloco. A coluna *Block_ID* define a qual bloco os pontos discretizantes pertencem. Todos os pontos com o mesmo *Block_ID* pertencem ao mesmo bloco. Na figura, o bloco com *Block_ID* igual a 5 possui cinco pontos discretizantes. O teor do bloco é a propriedade de interesse.

```
Block Samples File
5
X
Y
Z
Grade
Block_ID
1.00 0.00 0.00  10.00 5
2.00 10.00 0.00 10.00 5
3.00 20.00 0.00 10.00 5
4.00 30.00 0.00 10.00 5
5.00 40.00 0.00 10.00 5
6.00 50.00 0.00 10.00 2
7.00 60.00 0.00 10.00 2
8.00 70.00 0.00 10.00 2
9.00 80.00 0.00 10.00 2
```

Figura 71: Arquivo com amostras de bloco.

Arquivo de saída

A figura 72 mostra o arquivo de saída do software *Block_Variogram*. O arquivo de saída do *Block_Variogram* contém 8 colunas para cada direção. As colunas representam o seguinte:

Coluna 1: índice do *lag;*

Coluna 2: distância média do *lag;*

Coluna 3: valor do variograma;

Coluna 4: número de pares;

Coluna 5: média dos valores presentes no fim do vetor;

Coluna 6: média dos valores presentes no início do vetor;

Coluna 7: variância dos valores presentes no fim do vetor;

Coluna 8: variância dos valores presentes no início do vetor.

O arquivo de saída do *Block_Variogram* pode ser utilizado pelo *software vargplt* do GSLIB (Deutsch e Journel, 1998) para a visualização do variograma experimental em suporte de bloco.

index_lag	dist	gam_value	npairs	t_mean	h_mean	t_variance	h_variance
1	44.541	0.333	7	75.26	60.00	105.35	114.73
2	199.031	0.858	1194	64.38	64.48	222.74	222.24
3	399.448	0.880	953	64.71	64.97	210.58	203.41
4	599.627	0.877	836	63.95	64.86	193.63	202.80
5	800.429	0.865	708	64.45	64.35	183.92	224.19
6	998.724	0.872	552	65.56	64.98	210.13	213.05
7	1199.519	0.900	468	65.37	63.60	170.35	205.67
8	1398.031	0.777	379	64.31	63.53	173.26	203.72
9	1600.408	0.845	311	64.61	62.97	198.43	175.69
10	1802.727	1.022	264	64.76	62.49	159.32	182.70
11	2001.106	0.886	279	64.18	60.61	178.30	142.71
12	2199.428	0.914	273	65.52	61.43	144.97	155.89
13	2398.942	0.986	242	66.13	61.42	168.23	147.28
14	2598.935	0.987	216	66.17	61.65	134.21	164.83

Figura 72: Arquivo de saída do software *Block_Variogram*.

A.1.2 *Block_Vmodel*

O *software Block_Vmodel* calcula o variograma médio entre blocos, o variograma médio dentro do bloco e o variograma em suporte de bloco com base em um modelo de variograma em suporte de ponto. Ele é usado para obter o modelo de variograma em suporte de ponto a partir de amostras de diferente suporte.

Parâmetros

A figura 73 mostra o arquivo de parâmetros do *software Block_Vmodel*. A descrição linha a linha do arquivo de parâmetros é a seguinte:

Linha 5: arquivo de dados com as amostras de blocos;

Linha 6: colunas das coordenadas X, Y, Z, Teor e *Block_ID* das amostras de bloco;

Linha 7: limites mínimo e máximo de corte dos dados;

Linha 8: arquivo de saída com os variogramas experimentais dos blocos;

Linha 9: número de direções;

Linha 10: azimute, tolerância do azimute, largura de banda horizontal, mergulho, tolerância do mergulho, largura de banda vertical e *rake;*

Linha 11: número de *lags*, tamanho de *lag* e tolerância do *lag.*

As linhas 12-13 definem a direção 2 e são análogas às linhas 10-11.

Linha 14: número de estruturas e efeito pepita do modelo de variograma em suporte de ponto;

Linha 15: tipo de variograma, contribuição, azimute, mergulho e rake;

Linha 16: alcance na direção de maior continuidade, na direção de menor continuidade e alcance na direção vertical.

```
1                Parameters for Block Vmodel
2                **********************
3
4  START OF PARAMETERS:
5  ../data/blocks.dat                  - file with block data
6  1   2   3   4   5                   - columns for X, Y, Z, grade, block_id
7  -1.0e21     1.0e21                  - trimming limits
8  block_vmodel.out                    - file for variogram output
9  2                                   - number of directions
10 0. 90. 9999.    0. 90. 9999.  0.00 - Dir 01: azm,atol,bandh,dip,dtol,bandv, rake
11 10  5.0  2.50                       - nlag,xlag,xtol
12 0. 20. 9999.    0. 90. 9999. 0.00  - Dir 02: azm,atol,bandh,dip,dtol,bandv, rake
13 8  7.0  4.0                         - nlag,xlag,xtol
14 1    0.2                           - nst, nugget effect
15 1    0.8  0.0   0.0   0.0          - it,cc,ang1,ang2,ang3
16          10.0  10.0  10.0          - a_hmax, a_hmin, a_vert
17
```

Figura 73: Arquivo de parâmetros do software *Block_Vmodel*.

Arquivo de saída

A figura 74 mostra o arquivo de saída do *Block_Vmodel*. Para cada direção, o *software Block_Vmodel* calcula o variograma médio entre blocos, o variograma médio dentro do bloco e o variograma em suporte de bloco. O arquivo de saída do *software Block_Vmodel* pode ser utilizado no software vargplt do GSLIB (Deutsch e Journel, 1998) para a visualização do variograma.

```
Avg_Gamma_Between_Blocks  DIR_ 1:  index_lag   dist  gam_value  npairs
   1       23.985      0.72355       26
   2      240.603      1.26717     8446
   3      431.193      1.29253    10525
   4      607.365      1.29612    12173
   5      810.266      1.30075    19316
   6     1022.142      1.30587    14298
Avg_Gamma_Within_Block    DIR_ 1:  index_lag   dist  gam_value  npairs
   1       23.985      0.48294       26
   2      240.603      0.42515     8446
   3      431.193      0.42286    10525
   4      607.365      0.42268    12173
   5      810.266      0.41986    19316
   6     1022.142      0.41601    14298
Block_Variogram           DIR_ 1:  index_lag   dist  gam_value  npairs
   1       23.985      0.24061       26
   2      240.603      0.84202     8446
   3      431.193      0.86967    10525
   4      607.365      0.87344    12173
   5      810.266      0.88089    19316
   6     1022.142      0.88986    14298
```

Figura 74: Arquivo de saída do software *Block_Vmodel*.

A.2 Estimativa com amostras de diferente suporte: Block_kriging_DH

Block_kriging_DH é um *plug-in* do SGeMS que realiza krigagem com amostras de diferente comprimento. As covariâncias entre amostras são calculadas como covariâncias bloco-a-bloco.

A.2.1 Interface

A interface do *plug-in* é dividida em três abas: *General and Data*, *Search* e *Variogram*. A figura 75 mostra a aba General e Data. Os números na figura 75 representam o seguinte:

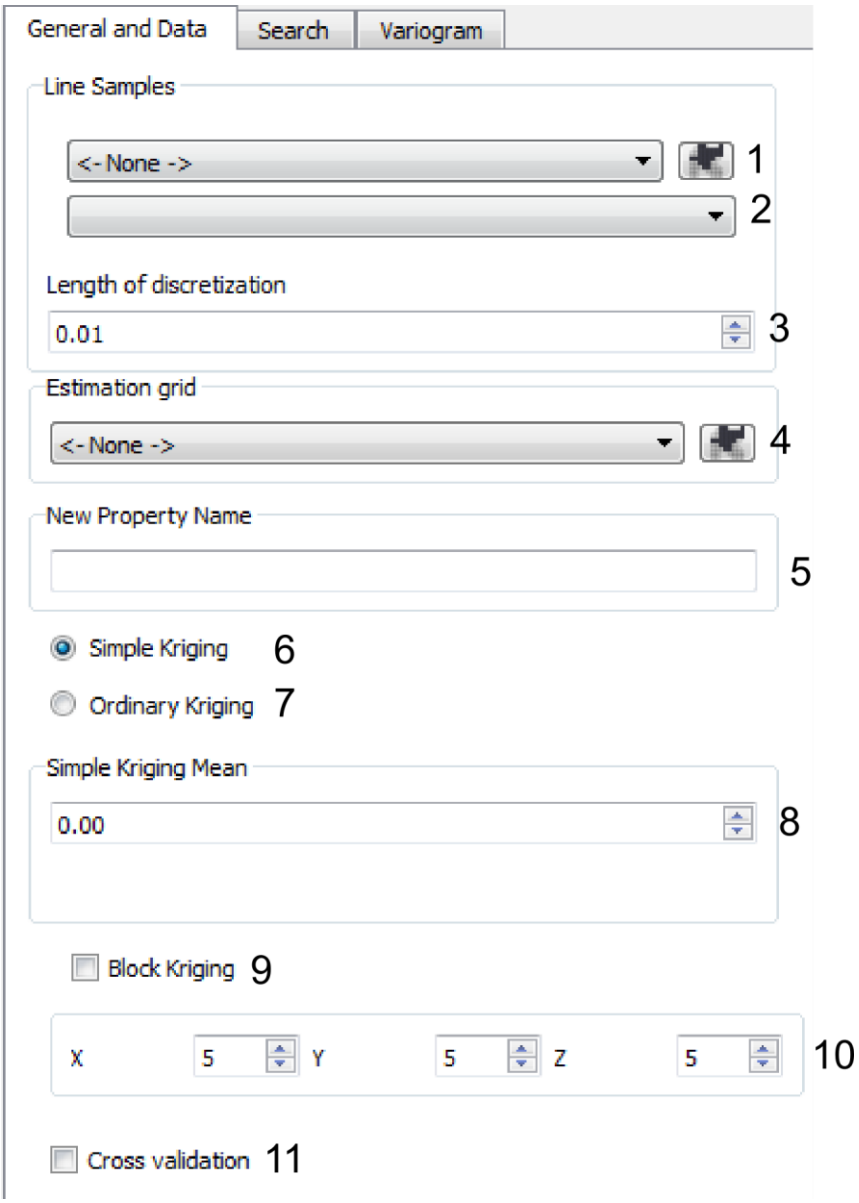

Figura 75: Interface da aba *General and Data* do *plug-in Block_kriging_DH*.

1. Seletor do banco de dados (precisa estar no formato de linhas);
2. Propriedade a ser estimada;
3. Comprimento de discretização. As amostras de linhas são discretizadas em uma série de pontos. O comprimento de discretização define o espaçamento entre os pontos de discretização. As amostras mais curtas do que metade do comprimento de discretização não são consideradas na estimativa;
4. Grid de estimativa;

6. Opção para krigagem simples;
7. Opção para krigagem ordinária;
8. Média para a krigagem simples;
9. Opção para krigagem de blocos. Se essa opção é selecionada, o valor estimado corresponde ao teor médio do bloco definido pelo grid de estimativa;
10. Discretização do bloco em X, Y e Z para a krigagem de blocos;
11. Opção para validação cruzada.

As abas *Search* e *Variogram* são iguais às abas utilizadas nos demais algoritmos do SGeMS. O modelo de variograma na aba *Variogram* é o modelo de variograma em suporte pontual. Esse modelo é utilizado no cálculo das covariâncias bloco-a-bloco.

A.2.2 Banco de dados em formato de linha

O *plug-in Block_kriging_DH* necessita que o banco de dados esteja em formato de linha. Cada amostra representa uma linha no espaço. O arquivo do banco de dados consiste em um arquivo csv (arquivo separado por vírgula – *comma separated values*) em que a primeira linha contém as seguintes palavras-chave:
dhid: inteiro que define o número do furo de sondagem;
topx: coordenada x do topo da amostra;
topy: coordenada y do topo da amostra;
topz: coordenada z do topo da amostra;
botx: coordenada x do fundo da amostra;
boty: coordenada y do fundo da amostra;
botz: coordenada z do fundo da amostra;
from: distância da boca do furo até o início da amostra;
to: distância da boca do furo até o fim da amostra.

A figura 76 mostra um exemplo de um arquivo de linha. Para carregar no SGeMS, basta o usuário arrastar o arquivo para a tela do SGeMS.

```
dhid,topx,topy,topz,midx,midy,midz,botx,boty,botz,from,to,length,Grade_Au
1,0,0,0,0,0,-0.5,0,0,-1,0,1,1,8.852578801
1,0,0,-1,0,0,-3,0,0,-5,1,5,4,15.08936538
1,0,0,-5,0,0,-6.5,0,0,-8,5,8,3,14.4167314
1,0,0,-8,0,0,-9,0,0,-10,8,10,2,7.076061898
1,0,0,-10,0,0,-15,0,0,-20,10,20,10,11.68406064
2,0,100,0,0,100,-0.5,0,100,-1,0,1,1,10.14345419
2,0,100,-1,0,100,-3,0,100,-5,1,5,4,11.53808395
2,0,100,-5,0,100,-6.5,0,100,-8,5,8,3,12.23045565
2,0,100,-8,0,100,-9,0,100,-10,8,10,2,13.97579034
```

Figura 76: Exemplo de arquivo de linha.

A figura 77 mostra os furos de sondagem carregados no SGeMS. Cada amostra representa uma linha no espaço.

Figura 77: Amostras de linha carregadas no SGeMS.

A.3 Pós-processamento de simulação geoestatística multivariada

A.3.1 *mv_d90*

O *software mv_d90* é um executável estilo GSLIB que calcula a estatística D90 para múltiplas realizações.

Parâmetros

A figura 78 mostra o arquivo de parâmetros do *software mv_d90*. A descrição linha a linha do arquivo de parâmetros do é a seguinte:
Linha 6: arquivo do banco de dados em formato GeoEAS;
Linha 7: número de variáveis;
Linha 8: colunas das variáveis no banco de dados;
Linha 9: limites mínimo e máximo de corte;
Linha 10: arquivo com as realizações;
Linha 11: colunas das variáveis no arquivo das simulações;
Linha 12: número de células em x, y e z e número de realizações;
Linha 13: número de nós de grid aleatórios para o cálculo do D90. Se o valor for menor do que zero, todos os nós de grid são usados para o cálculo do D90;
Linha 14: arquivo de saída. Contém o D90 calculado para cada realização.

```
1  ------------------------------------------------------------
2                        MULTIVARIATE D90
3   ***********************************************************
4
5  START OF PARAMETERS:
6  data_file.txt          - data file
7  4                      - number of variables
8  1 2 3 4                - cols of variables
9  -98.00 1.0e21          - trimming limits
10 sim_file.txt           - simulation file
11 1 2 3 4                - cols of variables
12 50 10 1 2              - nx, ny, nz, nreal
13 10000                  - n_random_samples from simulation
14 mv_d90.out             - output file with multivariate cdf
```

Figura 78: Arquivo de parâmetros do software mv_d90.

A.3.2 *mvs_sum_check*

O *software mvs_sum_check* é utilizado para verificar e corrigir simulações multivariadas que tem restrição de soma.

Algoritmo

Considere um nó de grid simulado com *n* variáveis cuja soma precisa ser menor ou igual a certo valor Max_{sum} mais uma tolerância tol_{sum}. O algoritmo do *software mvs_sum_check* é o seguinte:

1. A soma das *n* variáveis é calculada:

$$Sum = \sum_{i=1}^{n} z_i$$

2. Se Sum é maior do que Max_{sum} mais tol_{sum}, os valores simulados são corrigidos:

$$z_i' = z_i \cdot \left(\frac{Max_{sum}}{Sum} \right), i = 1, \ldots, n$$

3. Os valores corrigidos são verificados. Se os valores estão abaixo do mínimo dos dados ou acima do máximo dos dados, os valores simulados são corrigidos novamente:

$$z_i'' = z_i^{max} se(z_i' > z_i^{max}), i = 1,...,n$$
$$z_i'' = z_i^{min} se(z_i' < z_i^{min}), i = 1,...,n$$

4. Os valores originais z_i recebem os valores corrigidos z_i'' e os passos 1-3 são repetidos para um número de iterações.

A prática tem demonstrado que 20 iterações são suficientes para atingir a convergência dos resultados.

Parâmetros

A figura 79 mostra o arquivo de parâmetros do *software mvs_sum_check*. A descrição linha a linha do arquivo de parâmetros do é a seguinte:

Linha 5: arquivo das simulações no formato GeoEAS;

Linha 6: limites mínimo e máximo de corte;

Linha 7: número de nós de grid em x, y e z e número de realizações;

Linha 8: número de restrições de soma e número de iterações;

Linhas 9-13 definem a primeira restrição de soma;

Linha 9: Valor máximo para a soma e tolerância. Se a soma das variáveis é maior do que o valor máximo mais a tolerância, os teores são corrigidos;

Linha 10: número de variáveis para a primeira restrição de soma;

Linha 11: colunas das variáveis;

Linha 12: valores mínimos das variáveis;

Linha 13: valores máximos das variáveis;

As linhas 14-17 definem a segunda restrição de soma e são similares às linhas 10-13.

Linha 18: arquivo com relatório que informa a proporção de nós de grid corrigidos;

Linha 19: arquivo de saída com as realizações corrigidas.

```
1
2                  MULTIVARIATE SIMULATION CHECK FOR SUM CONSTRAINTS
3                   ************************************************
4  START OF PARAMETERS:
5  ./Sgsim/sgsim.out           - simulation file
6  -98.00 1.0e21               - trimming limits
7  100 100 1 1                 - nx, ny, nz, nreal
8  2 10                        - n of constraints and n of iterations
9  100.00 0.001                - max sum of variables and tolerance
10 4                           - number of variables for sum constraint 1
11 1 2 3 4                     - cols of variables  for sum constraint 1
12 0.00 0.00 0.00 0.00         - min of data for sum constraint 1
13 90.00 90.00 90.00 90.00     - max of data for sum constraint 1
14 4                           - number of variables for sum constraint 2
15 5 6 7 8                     - cols of variables  for sum constraint 2
16 0.00 0.00 0.00 0.00         - min of data for sum constraint 2
17 90.00 90.00 90.00 90.00     - max of data for sum constraint 2
18 Error_report_sum.out        - report of the errors for sum constraints
19 ./Sgsim/sgsim_corrected.out - file with corrected simulated values
```

Figura 79: Arquivo de parâmetros do software mvs_sum_check.

A.3.3 *mvs_frac_ratio*

O *software* *mvs_frac_ratio* é utilizado para fazer simulações de um teor fracionário a partir da simulação de um teor e de uma razão de fração. Além disso, o software *mvs_frac_ratio* corrige problemas de extrapolação dos teores fracionários.

Parâmetros

A figura 80 mostra o arquivo de parâmetros do software *mvs_frac_ratio*. A descrição linha a linha do arquivo de parâmetros é a seguinte:

Linha 5: arquivo da simulação em formato GeOEAS;

Linha 6: limites mínimo e máximo de corte;

Linha 8: número de razões de fração;

Linha 9: colunas do teor e da razão de fração, tipo de cálculo e fator. Se o tipo de cálculo é igual a 1, o teor resultante $teor_{res}$ é calculado da seguinte forma:

$$teor_{res} = \left(\frac{teor}{razão} \right) \cdot fator$$

Se o tipo de cálculo é igual a 2, o teor resultante $teor_{res}$ é calculado da seguinte forma:

$$teor_{res} = teor \cdot razão \cdot fator$$

Linha 10: mínimo e máximo dos dados. Se o teor resultante é acima do máximo, o valor é mudado para o máximo. Se o teor resultante é menor do que o mínimo, o valor é mudado para o mínimo;

As linhas 11-12 representam outra razão de fração e são similares às linhas 9-10.

Linha 13: arquivo com relatório que informa a proporção de nós de grid corrigidos;

Linha 14: arquivo de saída das simulações com o teor resultante. O teor resultante é armazenado na coluna da razão de fração.

```
1                          ****************************
2                          MULTIVARIATE FRACTION RATIO
3                          ****************************
4  START OF PARAMETERS:
5  ./Sgsim/sgsim.out           - simulation file
6  -98.00 1.0e21               - trimming limits
7  100 100 1 1                 - nx, ny, nz, nreal
8  2                           - number of fraction ratios
9  1  2  1  0.01               - col_grade, col_ratio, calc_type
10 0.00     100.00             - min and max for variable
11 3  4  1  0.01               - col_grade, col_ratio, calc_type
12 0.00     100.00             - minimum and maximum for variable
13 Error_report_ratio.out      - report of the errors for min and max
14 ./Sgsim/sgsim_cor.out       - file with corrected simulated values
```

Figura 80: Arquivo de parâmetros do software mvs_frac_ratio.