

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Escola de Engenharia

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

USO DE INFORMAÇÃO SECUNDÁRIA IMPRECISA COMO DISTRIBUIÇÃO
DE PROBABILIDADE LOCAL EM CONJUNTOS DE DADOS
COMPLETAMENTE HETEROTÓPICOS

Cristina da Paixão Araújo

Tese para obtenção do título de Doutor em Engenharia

Porto Alegre, RS

2019

Cristina da Paixão Araújo

USO DE INFORMAÇÃO SECUNDÁRIA IMPRECISA COMO DISTRIBUIÇÃO
DE PROBABILIDADE LOCAL EM CONJUNTOS DE DADOS
COMPLETAMENTE HETEROTÓPICOS

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. João Felipe Coimbra Leite Costa, PhD

Co-orientador: Prof. Amilcar Soares, PhD

Porto Alegre, RS

2019

Cristina da Paixão Araújo

USO DE INFORMAÇÃO SECUNDÁRIA IMPRECISA COMO DISTRIBUIÇÃO
DE PROBABILIDADE LOCAL EM CONJUNTOS DE DADOS
COMPLETAMENTE HETEROTÓPICOS

Esta tese foi julgada adequada para 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. Amílcar Soares
PhD. pelo Instituto Superior Técnico
Co-Orientador

Prof. Dr. Afonso Reguly
Dr. pela Universidade Federal do Rio Grande do Sul
Coordenador do PPGE3M

BANCA EXAMINADORA:

Dr. Luiz Eduardo Varela (EXP/GEOP/TGEO/PETROBRAS): _____

Dr. Diego Machado Marques (UFRGS): _____

Dr. Marcel Antônio Arcari Bassani (UFRGS): _____

AGRADECIMENTOS

Ao Professor Dr. João Felipe Coimbra Leite Costa, por toda orientação, disponibilidade e acima de tudo pelos ensinamentos.

Ao professor PhD. Amílcar Soares pela orientação recebida durante meu período de estudos no Centro de Recursos Naturais e Ambiente (CERENA) no Instituto Superior Técnico em Lisboa, Portugal. As contribuições e convívio com o professor Amílcar enriqueceram bastante o trabalho desenvolvido.

A minha família, especialmente minha mãe e minha irmã Simone por ser a minha estrutura, por acreditar em mim e me fazer acreditar em todos os meus sonhos.

Aos colegas Vanessa Koppe e Marcel Bassani pelas discussões, motivação e apoio incondicional.

A todos os colegas do LPM-Laboratório de Pesquisa Mineral da UFRGS em especial, Diego Marques, Roberto Rolo, Péricles, Augusto, Anuar, George, Carlos Schneider, Tais Câmara, Ricardo Hundelshausen, Carla De Carli, Rochana e Samantha, pelos bons momentos de convivência.

Aos colegas do CERENA, em especial João Narciso, João Neves, Catarina, Rubens, Eduardo, Alzira, Pedro, Helga, Francesco, Luan e Ângela. Aos professores Maria João Pereira e Leonardo Azevedo que me ajudaram na minha chegada à cidade de Lisboa. Obrigada pelo acolhimento e convívio.

Aos amigos Anneliese, Gabriela Barboza, Jaqueline Mohr, Gabriela Cybis, Mariana Neumann, Marcelo Mutti, Waleska, Noris e Heitor, por me acolherem em suas famílias em Porto Alegre.

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

A CAPES e ERASMUS pelo apoio financeiro.

RESUMO

Decisões na indústria mineral são geralmente baseadas em modelos estimados/simulados obtidos a partir de amostras. Durante a fase de exploração, as amostras são extraídas por furos de sondagem diamantados (DDH), que tendem a ser precisas e acuradas, mas são de alto custo e, conseqüentemente, escassas. Essas amostras são consideradas "*hard data* ou variável primária". Na fase de produção, as amostras podem ser coletadas por pó de perfuratriz ou amostras de fragmentos de rocha. Essas amostras são de baixo custo, abundantes e com alta incerteza. Portanto, tendem a existir mais dados com má qualidade, se comparado aos dados da fase de exploração. Essas amostras são consideradas "*soft data* ou variável secundária". Abordagens clássicas em geoestatística, para integrar as variáveis primárias e secundárias, utilizam os algoritmos de cokrigagem e co-simulação, que requerem o modelo linear de coregionalização. A obtenção desse modelo pode ser considerado uma tarefa difícil quando existem duas ou mais variáveis secundárias disponíveis. Este trabalho propôs uma metodologia que, inicialmente, transforma os dados da variável secundária em variável primária a partir da construção distribuição probabilidade local *a priori*, incorporando a incerteza das medidas nas amostras, e em seguida executa o algoritmo de simulação estocástica.

Para dados que podem ter diferentes suporte, foi investigado o algoritmo de simulação sequencial direta com distribuições pontuais. Foram comparadas três metodologias: Metodologia 01: Simulação sequencial direta utilizando somente a variável primária, Metodologia 02: Simulação sequencial direta com distribuição pontual utilizando krigagem simples e na Metodologia 03: Simulação sequencial direta com distribuições pontuais utilizando cokrigagem simples com modelo intrínseco para inferir os *pseudohards*. Os resultados mostraram que quando os *pseudohards* são inferidos com o algoritmo de cokrigagem, e a correlação entre os dados é considerada, melhora a reprodução da distribuição local e os modelos são mais precisos e acurados .

Para dados com mesmo suporte, tantos primários como secundários, foram avaliadas quatro metodologias definidas como: Metodologia 01: Simulação estocástica utilizando apenas a variável primária, Metodologia 02:

Simulação estocástica utilizando a variável primária e secundária como distribuição bivariada normal, Metodologia 03: Simulação estocástica utilizando a variável primária e secundária como distribuição inferida por cokrigagem e Metodologia 04: Simulação estocástica utilizando inserção múltipla com atualização bayesiana no local da secundária. Verificou-se que a amostragem da distribuição de probabilidade pelo algoritmo de Monte Carlo (*MCS*), tem-se a reprodução da continuidade espacial dos *pseudohards* inferidos com flutuações ergódicas menores que o algoritmo de simulação por campos de probabilidade (*p-field*). Sobre a construção da distribuição, verificou-se que a distribuição local quando foram utilizados os dados da variável secundária no local e na vizinhança e a variável primária, os modelos finais foram mais precisos e acurados. Além disto, esta abordagem pode ser utilizada com qualquer algoritmo de simulação estocástica.

As abordagens propostas apresentam duas vantagens: i) os erros da variável secundária são mitigados e não são transferidos para o modelo final ii) para incorporar a variável secundária é necessário somente o modelo de continuidade espacial da variável primária, evitando a modelagem do modelo linear de coregionalização (MLC). As metodologias propostas foram ilustradas em um banco de dados real, em 3D, em uma mina subterrânea de cobre e zinco, em que a variável primária e secundária possuem heterotopia total.

Verificou-se, nas metodologias investigadas, que a informação secundária foi incorporada sem que o viés e a imprecisão fossem transferidos para os modelos. As metodologias produzem modelos mais precisos e acurados quando comparados aos gerados usando somente dados primários.

ABSTRACT

Decisions in the mining industry are generally based on estimated/simulated models obtained from samples. During the exploration stage, samples obtained by diamond drilling holes (DDH) tend to be precise and accurate, but are expensive and, consequently, sparse. These samples are considered "hard data" or "primary variable". In the production stage, the samples are collected by reverse circulation or chip samples. These samples are cheaper, abundant and higher uncertainty than DDH. These samples are considered "soft data". Thus, there is a tendency to collect data in the production stage, instead of in the exploration stage, as the quantity of samples overcome its low. Classical approaches in geostatistics, to integrate primary and secondary variables, suggests that the use of cokriging and co-simulation algorithms requires a linear model of coregionalization (LMC). It can be considered a difficult task to obtain this model when two or more secondary variables are available. This study proposes a methodology that initially transforms the secondary variable into primary variable from local probability distribution by incorporating the uncertainty of the measurements in the samples. This approach is used to perform the stochastic simulation algorithm.

To integrate data with different support, Direct sequential simulation with local distributions algorithm was investigated. Three methodologies were compared: Methodology 01: Direct sequential simulation using only the primary variable; Methodology 02: Direct sequential simulation with local distribution using simple kriging; Methodology 03: Direct sequential simulation with local distributions using simple cokriging with intrinsic model to infer pseudohards. The results showed that the pseudohards were inferred with the cokriging algorithm and the correlation between variables was considered. It improves the reproduction of the local distribution and the final models are more precise and accurate.

To incorporate primary and secondary variables with the same support, four methodologies were evaluated: Methodology 01: Stochastic simulation using only the primary variable; Methodology 02: Stochastic simulation using the primary and secondary variable as normal bivariate distribution; Methodology

03: Stochastic simulation using the primary and secondary variables as a distribution inferred by cokriging with intrinsic model; Methodology 04: Stochastic simulation using multiple imputation with Bayesian update at secondary variable location. The results showed that the spatial continuity reproduction of the pseudohards has less ergodic fluctuations in the probability distribution sampling by the Monte Carlo algorithm (MCS) than the sampling by p-field simulation algorithm. To build the distribution, the local distribution was verified when both secondary and primary variables were used. The secondary variable considered locally and in the neighborhood. The final models were more precise and accurate. In addition, this approach can be used with any stochastic simulation algorithm.

The proposed approaches have two main advantages: i) the errors of the secondary variable are mitigated and are not transferred to the final model; ii) to incorporate the secondary variable it is necessary only the spatial continuity model of the primary variable, avoiding the modeling of the LMC. The developed methodologies were applied in a real case study, copper and zinc underground mine, where the primary and secondary variables are completely heterotopic.

It was verified that bias and imprecision of the secondary data were not transferred to the final models. The methodologies produce more precise and accurate models when compared to models obtained using only primary data.

LISTA DE FIGURAS

Figura 2. 1- Amostragem da distribuição global F_{zz} definida pelos intervalos da média local e variância de $z(x_i)$. FONTE: SOARES (2001).....	21
Figura 2. 2-Exemplo de amostras primárias (pontos azuis) e variáveis secundárias (pontos vermelhos) em configuração de heterotopia total	25
Figura 2. 3-Cálculo dos valores simulados no local da distribuição probabilidade local (<i>pseudohard</i>).....	26
Figura 2. 4- a) Banco de dados que será utilizado na simulação sequencial direta–b) cenário final com todos os nós dos grids simulados	27
Figura 2. 5- Histograma e mapa de localização da variável V	27
Figura 2. 6- Histograma, gráfico de dispersão e resultados do teste Kolmogorov-Smirnov para as distribuições a) d_1 , b) d_2 e c) d_3	30
Figura 2. 7- Exemplo da configuração amostral entre os dados da variável primária (pontos azuis) precisos e acurados e a distribuição de probabilidade dos dados (vermelho) representando os dados imprecisos e acurados	31
Figura 2. 8- Histograma, gráfico de dispersão para as distribuições a) d_1 , b) d_2 e c) d_3	36
Figura 2. 9 - Fluxograma da metodologia simulação sequencial direta com distribuições pontuais.....	37
Figura 2. 10- Mapa de localização dos dados utilizados no estudo de caso, no plano zy . (pontos azuis: variável primária, pontos vermelhos: variável secundária e área verde corresponde ao domínio geológico dos dados)	38
Figura 2. 11 - Correlograma cruzado entre os dados primários e secundários (linha vermelha) e a extrapolação do correlograma cruzado (linha azul) para $h = 0$	39
Figura 2. 12- Histograma dos dados <i>pseudohard</i> (linhas pretas) gerados comparados com o histograma da variável primária (Cu_DDH) (linha vermelha) a) Método 02 b) Método 03.	41
Figura 2. 13- Comparativo entre a reprodução dos variogramas dos modelos finais (linhas preta), o modelo do variograma (linha vermelha) e o variograma experimental dos dados (linha azul) nas direções de continuidade espacial a) direção principal (h_1), b) direção intermediária (h_2), e c) direção vertical (h_3). 42	
Figura 2. 14- Mapa de localização dos dados utilizados no estudo de caso, no plano zy . (pontos azuis: variável primária, pontos vermelhos: variável secundária e pontos verdes: dados da variável primária excluídos do banco de dados e área cinza corresponde ao domínio geológico dos dados).....	43
Figura 2. 15- Gráfico de dispersão dos valores reais da variável primária (Cu_DDH) comparado ao valor médio obtido pelo E-type pelo os métodos a) Método 01, b) Método 02 e c) Método 03	44
Figura 2. 16- Gráfico de dispersão do erro absoluto do Método 03 comparado com as outras metodologias a) Método 01 e b) Método 02.	45

Figura 3. 1- Fluxograma das metodologias a serem propostas para inserir a variável secundária com distribuição de probabilidade local	50
Figura 3. 2- Exemplo da localização espacial completamente heterotópica da variável primária (distribuições azuis) e da distribuição da variável secundária (distribuições vermelhas).	54
Figura 3. 3- Exemplo da amostragem realizada na distribuição local da variável secundária para gerar o <i>pseudohard</i>	55
Figura 3. 4- Valores de <i>pseudohard</i> gerados e a variável primária que serão utilizados no algoritmo de simulação estocástica.....	56
Figura 3. 5 – Distribuição com intervalo de probabilidade simétrico para $p = 0,80$	61
Figura 3. 6- Codificação do valor verdadeiro em distribuições com intervalo de probabilidade simétrico.	61
Figura 3. 7 – Gráfico de precisão entre os valores dentro do intervalo de probabilidade e os intervalos de probabilidade simétrico.	62
Figura 3. 8- Histograma dos dados após a transformação gaussiana a) variável primária Cu_DDH desagrupada e b) variável secundária Cu_Chip desagrupada.	64
Figura 3. 9 – Histograma dos dados <i>pseudohard</i> (linhas pretas) gerados comparados com o histograma da variável primária (Cu_DDH) (linha vermelha) a) Método 02 b) Método 03 e c) Método 04.	65
Figura 3. 10 – Reprodução dos variogramas dos <i>pseudohards</i> gerados e a variável primária (linhas preta), modelo do variograma dos dados da variável primária (linha vermelha) e variograma experimental da variável primária (linha azul) nas direções de continuidade espacial a) direção principal (h_1), b) direção intermediária (h_2), e c) direção vertical (h_3).....	66
Figura 3. 11 - Histograma dos modelos simulados (linhas pretas) comparados com o histograma da variável primária a) Método 01, b) Método 02, c) Método 03 e d) Método 04.	67
Figura 3. 12 – Reprodução dos variogramas dos modelos finais (linhas preta) e o modelo do variograma do variograma (linha vermelha) nas direções de continuidade espacial a) direção principal (h_1), b) direção intermediária (h_2), e c) direção vertical (h_3).....	68
Figura 3. 13- Histograma da variância condicional do modelos simulados para os métodos a) Método 01, b) Método 02, c) Método 03 e d) Método 04.....	69
Figura 3. 14 - Gráfico de dispersão dos valores reais da variável primária (Cu_DDH) comparado ao valor médio obtido pelo E-type pelo os métodos a) Método 01, b) Método 02, c) Método 03 e d) Método 04	72
Figura 3. 15 - Gráfico de dispersão do erro absoluto do Método 03 comparado com as outras metodologias a) Método 01, b) Método 02 e c) Método 04.....	73
Figura 3. 16 - Gráfico de acuracidade das metodologias a) Método 01 com 100 realizações, b) Método 01 com 50 realizações, c) Método 02, d) Método 03 e e) Método 04.	76

LISTA DE TABELAS

Tabela 2. 1- Estatística descritiva das distribuições d1, d2 e d3.....	28
Tabela 2. 2 - Pesos recebidos pela amostra retirada da distribuição de probabilidade com alta correlação (0,90) com a variável primária.	34
Tabela 2. 3 Pesos recebidos pela amostra retirada da distribuição de probabilidade com fraca correlação (0,10) com a variável primária.	34
Tabela 2. 4- Resultados do teste Komolgorov-Smirnov considerando que a distribuição de probabilidade possui forte correlação (0,90) com a variável primária	36
Tabela 2. 5- Estatística descritiva da variável cobre amostrada por furos de sondagem (DDH) e fragmento de rocha (<i>chip sample</i>).	39
Tabela 3. 1-Tabela comparativa do erro relativo das metodologias investigadas.	74
Tabela 3. 2-Tabela comparativa do coeficiente de “goodness”	75

SUMÁRIO

INTRODUÇÃO	1
1.1 Contextualização do problema a ser investigado.....	1
1.1.1 Amostragem e a diferença entre a qualidade das amostras.....	1
1.1.2 Construção dos modelos de teores	2
1.2 Revisão bibliográfica	4
1.2.1. Sobre a integração da informação secundária	4
1.2.1.1 Modelos obtidos por estimativa.....	4
1.2.1.2 Modelos obtidos por simulação.....	7
1.2.2. Sobre a integração de informação secundária imprecisa ou enviesada	10
1.2.3 Abordagens recentes em geoestatística sobre a integração de informação secundária.....	13
1.3 Meta	15
1.4 Contribuição da tese	17
1.5 Objetivo.....	15
1.6 Organização da tese	16
2 SIMULAÇÃO SEQUENCIAL DIRETA COM DISTRIBUIÇÕES PONTUAIS (DISTRIBUIÇÃO DE PROBABILIDADE NO LOCAL DA SECUNDÁRIA).....	19
2.1 Fundamentação teórica sobre as metodologias abordadas.....	20
2.1.1. Simulação sequencial direta (<i>Direct sequential simulation</i>)	20
2.1.2 Simulação sequencial direta com distribuição pontual.....	24
2.1.3 Investigação da reprodução da distribuição (<i>pseudohard</i>) inferidos por krigagem simples.....	27
2.1.4 Considerações sobre a reprodução da distribuição local (<i>pseudohard</i>)	30
2.2 Metodologia proposta.....	32
2.2.1 Investigação da reprodução da distribuição (<i>pseudohard</i>) inferidos por cokrigagem intrínseca.....	33
2.3 Estudo de caso	37
2.3.1 Descrição geológica do depósito estudado	37
2.3.2 Apresentação do banco de dados	38
2.3.3 Construção da distribuição de probabilidade	40
2.4 Validação	41
2.4.1 Validação dos <i>Pseudohard</i>	41

2.4.2. Avaliando a acuracidade e precisão dos modelos	43
2.5 Considerações	45
3 INCORPORAÇÃO DA VARIÁVEL SECUNDÁRIA COMO DISTRIBUIÇÃO PROBABILIDADE LOCAL UTILIZANDO SIMULAÇÃO ESTOCÁSTICA	48
3.1 Fundamentação teórica sobre as metodologias abordadas.....	50
3.1.1 Simulação por campos de probabilidade (<i>p-field</i>).....	50
3.2 Metodologias propostas para construção da distribuição de probabilidade (<i>pdf</i>) no local da variável secundária.....	52
3.2.1 Distribuição 01: distribuição da variável primária obtida a partir de uma distribuição normal bivariada	53
3.2.2 Distribuição 02: distribuição da variável primária obtida por cokrigagem simples utilizando modelo linear de correionalização intrínseco (MCI)	53
3.3 Metodologias propostas para amostrar a distribuição probabilidade da variável secundária e incorporação da incerteza nos modelos.....	53
3.4 Inserção múltipla no local da variável secundária obtida por atualização bayesiana (<i>Bayesian Updated</i>)	56
3.5 Verificação da incerteza das distribuições condicionais nos modelos estocásticos	60
3.6 Estudo de caso	63
3.6.1 Validação	63
3.6.1.1 Validação <i>Pseudohards</i>	63
3.6.1.2 Validação dos modelos finais	67
3.6.2. Avaliando a acuracidade e precisão dos modelos obtidos.....	71
3.7 Considerações	77
4 CONSIDERAÇÕES FINAIS.....	79
4.1 Comparativo entre as metodologias investigadas.....	79
4.2 Contribuições da tese	85
4.2.1 Incorporação da incerteza na variável secundária considerada como distribuição de probabilidade (<i>pdf</i>).....	85
4.2.2 Metodologias que não utilizam o modelo linear de correionalização (MLC).....	85
4.3 Limitações das metodologias propostas	87
4.4 Sugestões para trabalhos Futuros	88
REFERÊNCIAS BIBLIOGRÁFICAS	90

LISTA DE SIGLAS

- BU: *Bayesian Updated* - atualização bayesiana
- ccdf: *Conditional cumulative distribution function* - função de distribuição cumulativa condicional
- cdf: *Cumulative distribution function* - função de distribuição cumulativa
- COK: *Ordinary cokriging* - cokrigagem ordinária
- E-type: média das simulações estocásticas
- DBSIM: *Direct block simulation* - simulação direta de blocos
- DDH: *Diamond drill hole* - furos de sondagem diamantada
- Chip samples: amostragem de fragmentos de rocha a serem lavradas
- DSS: *Direct Sequential Simulation* - simulação sequencial direta
- FFT: *Fast Fourier transform* - transformada rápida de Fourier
- GSLIB: *Geostatistical software library* - biblioteca de softwares geoestatísticos
- GTSIM: *Gaussian truncated simulation* - simulação truncada Gaussiana
- ICCK: *intrinsic collocated cokriging* - cokrigagem colocada intrínseca
- ICK: *intrinsic cokriging* - cokrigagem intrínseca
- IP: *Interval probability* - intervalo de probabilidade
- KED: *Kriging with external drift* - krigagem com tendência externa
- KWS: *Kriging within strata* - krigagem dentro de domínio estratiformes
- MAF: *Minimum/maximum autocorrelation factors* – fatores de autocorrelação mínimo e máximo
- MCCOK: *Multicollocated cokriging* – cokrigagem multicolocada
- MCI: *Intrinsic model of coregionalization* - modelo de corregeionalização intrínseco
- MCS: *Monte Carlo simulation* – simulação Monte Carlo
- MI: *Multiple Imputation* - Inserção múltipla
- MM1: *Markov model 1* - modelo de Markov 1
- MM2: *Markov model 2* - modelo de Markov 2
- OK: *Ordinary kriging* - krigagem ordinária
- OKVME: *Ordinary kriging with variance of measurement error* - krigagem ordinária com a medida da variância do erro

PCA: Principal component analysis - análise de componentes principais

pdf: *probability distribution function* - função densidade de probabilidade

P-field: Campos de probabilidade

p-value: *probability value* – probabilidade de significância

SCOK: *Standardized ordinary cokriging* - cokrigagem ordinária estandardizada

SGS: Sequential Gaussian simulation - simulação sequencial Gaussiana

SGSBK: sequencial gaussiana com krigagem de blocos

SGSCS: Sequential Gaussian co-simulation - co-simulação sequencial Gaussiana

SIS: Sequential indicator simulation - simulação sequencial indicadores

SKLVM: *Simple kriging with locally varying mean* - krigagem simples com medias locais

SOK: *Standardized ordinary kriging* - krigagem ordinária estandardizada

Sph: *Spherical model* - modelo esférico

SK: Teste Smirnov-Kolmogorov

TLC: Teorema do limite central

Trn: Tabela de transformação quantil-quantil

XRF: Fluorescência de raios X portátil

ÍNDICE DE SÍMBOLOS

$F_z'(z)$: Função de distribuição cumulativa local dos dados originais

$\bar{y}_L(x_0)$: Estimativa por krigagem simples da distribuição *likelihood*

$\bar{y}_P(x_0)$: Estimativa por krigagem simples da distribuição *a priori*

$F_n(x)$: Distribuição dos pseudohards gerados

$F_z(z)$: Função de distribuição cumulativa dos dados originais

H_0 : condição da Hipótese nula

d_1 , d_2 e d_3 : Distribuição de referência

$p^l(x_0)$: Realizações não condicionais com dados espacialmente correlacionadas

p_{max} : Intervalo máximo de probabilidade

p_{min} : Intervalo mínimo de probabilidade

x_0 : Vetor de coordenadas

$y(x_0)^*$: Equivalente Gaussiano da estimativa por krigagem simples

$y^s(x_0)$: Dado simulado no espaço Gaussiano no local não amostrado

$z^l(x_0)$: Dado simulado no espaço original no local não amostrado

$\gamma_i(h)$: Variograma da variável primária

$\gamma_{ij}(h)$: Variograma cruzado ($i \neq j$) entre a variável primária e secundária

$\lambda_\alpha(x_0)$: λ_α peso associado ao local não amostrado x_0

$\bar{\xi}(p)$: Função indicadora média para todos os n valores da distribuição em cada local x_0

$\sigma_L^2(x_0)$: Variância da estimativa por krigagem simples da distribuição *likelihood*

$\sigma_P^2(x_0)$: Variância da estimativa por krigagem simples da distribuição *a priori*

$\sigma_{SK}^2(x_0)$: Variância da estimativa por krigagem simples

$\sigma_U^2(x_0)$: Variância da estimativa por krigagem simples da distribuição *updated*

φ_{ij} : Coeficiente de proporcionalidade para o variograma $\gamma_{ij}(h)$

h : Vetor de separação

y : Valor *normal score*

$C(x_0 - x_i)$: Covariância entre o local a ser estimado x_0 e as amostras utilizadas na estimativa x_i .

D : Medida de diferença entre a distribuição de referência e os *pseudohards* gerados

$F(z(x_\beta))$: Função de distribuição de probabilidade local, definidos nos locais dos dados experimentais x_β .

$F(x)$: Distribuição dos dados da distribuição de referência

$G(y(x_0)^*, \sigma_{SK}^2(x_0))$: Distribuição Gaussiana local centrada na média do equivalente gaussiano da estimativa e variância de krigagem simples

p : Probabilidade

$z(x_0)^*$: Estimador por krigagem simples

$z(x_i)$: Valores experimentais

$z(x_\beta)$: Dados experimentais da função da distribuição de probabilidade $F(z(x_\beta))$

$\xi(x_0; p)$: Função cada valor da *ccdf* em um local (x_0) é identificado por uma função de indicadores

CAPÍTULO 1

INTRODUÇÃO

Na geoestatística, a quantidade e qualidade dos dados influenciam na precisão e acuracidade de estimativas e simulações. Embora medições diretas com dados precisos e acurados são em quantidade limitada, existe frequentemente uma quantidade considerável de informação menos precisa. O uso combinado dessas duas informações representa um desafio para incorporar todas as informações disponíveis e obter um modelo final mais preciso e acurado.

1.1 Contextualização do problema a ser investigado

1.1.1 Amostragem e a diferença entre a qualidade das amostras

A amostragem é um processo contínuo ao longo da vida da mina. Seu objetivo é ajudar a identificar áreas de interesse com maior potencial de lavra e reduzir a incerteza na previsão de teores ou de outra propriedade físico-química relevante na cadeia produtiva mineral. Para isto, o planejamento de lavra requer amostragem com espaçamento adensado para garantir previsões acuradas e precisas.

Durante a fase de exploração, as amostras são extraídas por furos de sondagem diamantados (DDH), geralmente possuem maior custo e são acompanhados por programas de controle de qualidade de amostras. Nesta fase, existem poucos dados, em geral, demasiadamente espaçados, mas com alta qualidade. Estes são referidos como dados primários ou *hard data*.

Com o desenvolvimento e operação da mina, mais amostras estarão disponíveis. Nesta fase, existem amostras coletadas com diferentes técnicas e possuem protocolos diferentes, assim como, análises químicas realizadas por diferentes laboratórios ou métodos. Em geral, estas informações não têm a mesma qualidade dos ditos dados primários. Tende a haver muitos dados com má qualidade, se comparado aos dados da fase de exploração. Dados obtidos por sensores, como fluorescência de raio X portátil ou geofísica, têm ganho popularidade recentemente para controles de teor nas fase de lavra. Esses

dados contêm ruído, mas são numerosos e rápidos de serem obtidos. São denominados dados secundários ou *soft data*.

Em geoestatística, os dados primários são honrados nos modelos teores, ignorando os erros de medida. Já, os dados *soft data*, que são considerados dados de baixa precisão, para serem incorporados no processo de estimativa ou simulação dos dados primários como informação auxiliar, é necessário que exista correlação entre eles. O benefício da informação secundária é verificado em situações onde falte informação primária. A influência da informação secundária nos processos de estimativa ou simulação é determinada pelo modelo de covariância cruzada que considera a correlação, a distância e anisotropia das variáveis que afetam os pesos recebidos pelas amostras. A existência de erros nos dados secundários diminui a correlação entre os dados primários e secundários. Assim, qualquer decisão em ignorar os erros dos dados secundários pode comprometer os modelos propostos.

Em resumo, cada campanha de amostragem está associada a diferentes técnicas amostrais e níveis de precisão e acuracidade, além de possuírem diferentes tipos de suporte. A escolha da técnica de amostragem, na maioria das vezes, está relacionado ao baixo custo da mesma, que permite mais amostras, com consequência perda de acuracidade e precisão dos dados. A confiabilidade do modelo de teores está relacionada à precisão dos dados de entrada. Assim, a integração de dados, pode tornar-se uma tarefa difícil, já que existem dados de entrada com diferentes acuracidade e precisão podem gerar modelos resultantes incompatíveis com o fenômeno geológico.

1.1.2 Construção dos modelos de teores

Na indústria mineral, as decisões são feitas baseadas no modelo teores. O valor de um depósito mineral está intimamente associado as medidas do recurso e a reserva. O recurso corresponde a quantidade de toneladas e teor do mineral de interesse que tem uma possibilidade realista de extração. Já a reserva é definida como a quantidade de recurso calculada que pode ser extraída economicamente dentro de um planejamento da mina (ROSSI; DEUTSCH,

2014). Assim, a tomada de decisão são baseadas no modelo teores para que os recursos e reservas possam ser quantificados.

A construção dos modelos de teores pode ser realizado a partir dos métodos de estimativa tradicionais, que utilizam os algoritmos de krigagem. Nestes métodos, no local não amostrado, em suporte de blocos é atribuído a melhor estimativa local baseado nos dados disponíveis, ou seja, é criado um único modelo determinista, com um valor para cada local. Nestes métodos, ocorre a suavização das estimativas e o enviesamento condicional. Um modelo de bloco determinista não permite compreender as incertezas locais do fenômeno.

Os algoritmos de simulação também podem ser utilizados para construção destes modelos. Estes permitem quantificar a incerteza local em suporte *quasi-ponto*, afim de avaliar o risco e auxiliar na tomada de decisões, produzindo um modelo que honre as estatísticas dos dados originais. Souza (2007) propôs técnicas para comparar os resultados obtidos pelos métodos tradicionais com os erros obtidos via utilização de cenários simulados de referência. A análise da incerteza e seus impactos nos recursos foram avaliados para cada um dos métodos estudados. Os resultados mostraram que a simulação permite abordar a quantificação da incerteza na classificação de recursos e reservas, e observar as limitações de cada técnica e de seus parâmetros-chave. Assim, os métodos de simulação estocástica, podem ser considerados apropriados para acessar a incerteza espacial dos teores (Journal, 1974, Journal and Huijbregts, 1978).

Nestes modelos estimados ou simulados são utilizados apenas os dados primários disponíveis, que possuem maior precisão e acuracidade. Em geral, as demais amostras são excluídas devido a imprecisão ou viés incorporados durante o processo de amostragem. A exclusão de uma ou mais fontes de informação de dados é realizada para minizar os erros. Segundo Emery (2005), a quantidade de dados prevalece sobre a qualidade dos dados, portanto a informação secundária não deve ser descartada. Assim, a utilização dos dados secundários na construção dos modelos como informação adicional auxilia nas estimativas.

Considerando que os dados são medidos em diferentes suportes e que estão sujeitos a vários tipos de erro, construir um modelo único que seja capaz de lidar com todas as formas de dados de uma maneira adequada para modelar o fenômeno geológico torna-se um desafio.

Em muitos casos, a escolha de um modelo simples incorporando os dados mais importantes é suficiente para proporcionar modelos que satisfaçam as exigências. Isso pode ser porque os outros dados são respeitados implicitamente pela simulação/estimativa, ou simplesmente porque eles não são relevantes para as aplicações realizadas. Sendo assim, o que deve ser avaliado é qual a melhor metodologia a ser utilizada sua incorporação e potencial ganho do uso dessa informação.

1.2 Revisão bibliográfica

A informação a ser estimada é denominada primária e as demais secundárias. A integração da informação secundária torna-se viável quando duas ou mais variáveis são amostradas dentro de um mesmo domínio espacial e apresentam significativo grau de correlação. O objetivo é, portanto, melhorar a estimativa da variável subamostrada (primária) utilizando aquelas mais densamente amostradas (secundárias). Diversas formas de integração de informação secundária têm sido propostas ao longo do tempo com diferentes objetivos. Serão abordadas as metodologias baseadas na teoria das variáveis regionalizadas (MATHERON, 1965).

1.2.1. Sobre a integração da informação secundária

1.2.1.1 Modelos obtidos por estimativa

No que se refere à incorporação de informação secundária para estimar a variável primária, quando a informação secundária é exaustivamente amostrada três algoritmos de krigagem podem ser utilizados para a incorporação: krigagem simples com médias locais (SKLVM), krigagem com tendência externa (KED), krigagem dentro de domínio estratiformes (KWS). Informações sobre estes algoritmos estão disponíveis em Goovaerts (1997), Wackernagel (1994) e outros. Dentre os algoritmos de krigagem, a krigagem

simples com médias locais (SKLVM) e a krigagem com tendência externa (KED) utilizam a variável secundária para caracterizar a tendência espacial da variável primária. Os dois estimadores se diferem na definição da tendência da variável primária. Na krigagem simples com médias locais (SKLVM), a média da primária é determinada pela combinação linear dos resíduos. E, na krigagem com tendência externa (KED), os coeficientes da tendência são determinados implicitamente pelo sistema de krigagem da região próxima ao local a ser estimado. Para realizar a estimativa por krigagem estratificada (KWS), inicialmente a área de estudo é dividida em estratos baseada na informação secundária. Os dados primários devem então ser tratados como um estrato ou população separada. É necessário que cada estrato tenha dados primários suficientes para que possa ser inferido o modelo de covariância. Em seguida, cada estrato é estimado usando o modelo de covariância e os dados primários próximos de cada estrato.

Em situações em que a variável secundária não está exhaustivamente amostrada, os algoritmos de cokrigagem (Marechal, 1970) são utilizados. Estes são extensão natural dos algoritmos de krigagem para que uma variável secundária possa auxiliar nas estimativas. É recomendado em situações que a variável secundária possui uma densidade amostral superior a variável primária, desde que exista correlação entre elas para que possa ser incorporada. Para modelar a continuidade espacial, é necessário assegurar a positividade do modelo de correionalização. Um dos modelos mais usados é o modelo linear de correionalização (MLC) no qual todos os modelos de variogramas diretos e cruzados são expressos como combinações lineares das mesmas estruturas básicas. A modelagem, portanto, não é simples, dificultando com o aumento do número de variáveis. A dificuldade no ajuste do MLC é a principal desvantagem da cokrigagem. No caso isotópico, onde todas as variáveis são amostradas nos mesmos locais, a cokrigagem e krigagem possuem estimativas semelhantes. (WACKERNAGEL, 1994, p.148, GOOVAERTS, 1997, p. 215). Salienta-se, que a cokrigagem só irá melhorar as estimativas realizadas por krigagem, quando as variáveis apresentarem alta correlação e quando a variável primária estiver amostrada em menor proporção que as secundárias. Xu et al. (1992)

desenvolveram a cokrigagem colocada que propõe reter somente a variável secundária colocado no nó do grid em que será estimado a variável primária.

A respeito da continuidade espacial, o modelo intrínseco de correlação (MCI) e os modelos de Markov, são considerados simplificações do modelo linear de coregionalização (MLC). Estas propostas assumem que o variograma e covariância entre a variável primária e secundária cruzados são proporcionais, mas existe diferença significativa entre eles. Nos modelos de Markov por MM1 (Almeida e Journel, 1994) é utilizado a correlação espacial da variável aleatória primária e o coeficiente de correlação entre as variáveis para construir os modelos de continuidade espacial. Para o MM2, contrariamente ao caso do MM1, se faz necessária a modelagem de todas as covariâncias diretas e cruzadas. Já, o modelo de correlação intrínseco (MCI) assume que todas as variâncias e covariâncias são proporcionais a um mesmo modelo de coregionalização. Este modelo exige mais informações sobre a continuidade das variáveis em estudo e resulta em melhores modelos de heterogeneidade sem aumento da variância. Detalhes podem ser encontrados, entre outros, em Isaaks e Srivastava (1989), Wackernagel (1995) Goovaerts (1997), Wackernagel (1995) e Babak e Deustch (2009).

As metodologias de krigagem não linear consiste em utilizar os estimadores tradicionais (krigagem simples ou krigagem ordinária) aos dados, após esses sofrerem transformações matemáticas não lineares. Já, a krigagem e a simulação de indicadores são capazes de estimar distribuições das variáveis primárias e secundárias (ALABERT, 1987; JOURNEL, 1986; ZHU; JOURNEL, 1993). A combinação destes dados foi proposta por Zhu e Journel (1993) utilizando o formalismo da cokrigagem ordinária dos indicadores. Uma das limitações desta metodologia, segundo Goovaerts (1997), é que a prática cokrigagem dos indicadores possui poucos benefícios em relação a krigagem de indicadores, já que os indicadores acumulados já carregam informações de outras classes. Em situações em que os dados primários encontram-se nos mesmos locais que os dados secundários, os dados secundários possuem pouca influência.

Uma das alternativas para incorporação da informação secundária diferente dos métodos clássicos de krigagem e cokrigagem, consiste em

métodos de fatoração, em que as variáveis correlacionadas são transformadas em fatores descorrelacionados. Nesta linha, há o método das componentes principais (Principal Component Analysis – PCA) e mínimos-máximos fatores de auto correlação (Minimum-Maximum Autocorrelation Factors-MAF). Para uma descrição detalhada destas metodologias, podem ser consultadas: Davis e Greenes (1983), Goovaerts (1987), Desbarats e Dimitrakopoulos (2000) entre outros. Estas metodologias consistem em uma transformação linear dos atributos correlacionados, levando-os a um espaço em que estejam descorrelacionados. A diferença entre as metodologias consiste no processo de descorrelação: para o PCA ocorre em vetores de separação nulos. Já, para o MAF que é derivado do PCA, as variáveis são descorrelacionadas para vetores de separação não nulos. Assim, as estimativas podem ser realizadas individualmente por krigagem ou simulação, evitando o uso do modelo linear de correção regionalização. Silva (2013) analisou a aplicabilidade do método de Mínimos-Máximos Fatores de Auto correlação à estimativa de teores em depósitos multivariados, discutindo suas vantagens e limitações. Os resultados mostraram que, ao comparar a técnica com o método de cokrigagem, a solução MAF proveu resultados de maneira mais rápida e simples sem perda de precisão. Uma das limitações desta metodologia é que os dados primários e secundários devem ser isotópicos.

1.2.1.2 Modelos obtidos por simulação

Diferente dos modelos estimados que têm como objetivo incorporar o comportamento médio das variáveis. Os modelos obtidos por simulação estocástica é uma abordagem geoestatística para avaliar a incerteza local e espacial de uma variável regionalizada no local não amostrado e produzir cenários equiprováveis (Journel, 1974; Journel e Huijbregts, 1978). Estes modelos reproduzem as estatísticas das amostras (histograma e modelo de continuidade espacial), honrando os dados em seus locais originais.

Na indústria de petróleo, dentre as variáveis disponíveis para modelamento de reservatórios existem os dados de porosidade e permeabilidade obtidos de amostra extraídas de poços, dados “hard” que são considerados como primária ou secundária, em suporte de pontos, esparsos,

com menor volume e alta resolução vertical. Existem também os dados de sísmica, incertos, correlacionados com a porosidade e permeabilidade, que possuem baixa resolução vertical e são considerados como informações médias da região a ser modelada. Behrens *et al.* (1996) propuseram usar a metodologia chamada simulação sequencial gaussiana com krigagem de blocos (SGSBK). Nesta proposta, o mapa sísmico é considerado informação secundária, como mapa suavizado médio do reservatório. Xu *et al.* (1992) propuseram utilizar o algoritmo de simulação sequencial gaussiana (SGS) com cokrigagem colocada considerando que este algoritmo seria mais eficiente embora menos rigoroso que SGS com cokrigagem completa (*full cokriging*). Tran *et al.* (2001) afirmam que não é recomendado o uso de transformações não lineares, como a normalização dos dados, para integrar informações com diferentes escalas/suporte, pois a relação linear entre as variáveis não seria mantida. Eles sugeriram a incorporação das informações com maior suporte utilizando simulação sequencial direta através da cokrigagem de blocos.

Já os algoritmos de co-simulação, utilizam a informação secundária para condicionar a simulação da variável primária. A informação secundária a ser utilizada pode referir-se à mesma variável primária, mas apresentada em um formato diferente, ou qualquer outra variável à qual a variável primária está correlacionada. A diferença deste algoritmo, consiste no condicionamento dos parâmetros da função da distribuição condicional, a média é condicional ao valor da estimativa por cokrigagem e a variância corresponde a variância da estimativa por cokrigagem. No entanto, este algoritmo possui algumas variações que estão relacionados ao modelo de continuidade espacial dos dados a serem utilizados na estimativa. Nos algoritmos de co-simulação, a cokrigagem simples colocada é comumente utilizada devido a sua simplicidade de implementação, pois é necessário somente a modelo de continuidade espacial da variável primária e, com a aproximação Markov-Bayes, o coeficiente de correlação entre a variável primária e os dados secundários (XU *et al.*, 1992; ALMEIDA; JOURNAL, 1994; GOOVAERTS, 1997). Segundo Babak e Deustch (2008), o uso de cokrigagem colocada nos algoritmos de co-simulação, em algumas situações em que o banco de dados possui várias variáveis secundárias pode ocorrer o aumento da variância estimada. Para solucionar este problema, eles propuseram uma

correção da variancia numa abordagem conhecida como cokrigagem intrínseca (ICCK). Nesta técnica, são utilizados os dados primários e secundários colocados e o modelo de correionalização é construído com o coeficiente de correlação entre os dados primários e secundários e os dados secundários no local a ser estimado e nos outros locais.

Um método que pode ser estendido a diversas variáveis e utilizado para incorporar dados secundários como informação sísmica (VERLY, 1992) e permeabilidades dos poços (GOMEZ-HERNANDEZ; JOURNAL, 1992) é conhecido como SGSCS (Co-simulação sequencial gaussiana). Este método é considerado simples e pode ser utilizado para incorporar vários tipos de informação, por exemplo geológicas (valores de permeabilidade), sísmicas (impedância acústica) e poços (porosidade). Uma limitação do algoritmo deve-se ao modelo e a hipótese multiguassiana remover qualquer correlação espacial significativa e específica entre os valores extremos altos ou baixos. Sendo assim, nestes casos, não garante a reprodução dos padrões espaciais dos valores extremos (GOOVAERTS, 1997).

Yao e Journal (2000) propuseram utilizar quatro algoritmos para integrar informações de diferentes escalas (dados sísmicos 2-D para estimar/simular a porosidade em 3-D) no modelamento propriedades petrofísicas: i) uso de FFT para calcular as covariâncias diretas e cruzadas, ii) integração de dados com escalas diferentes por cokrigagem colocada de blocos da variável porosidade em 2D a partir dos poços e dados sísmicos. Esta estimativa foi utilizada para condicionar a simulação sequencial direta da variável porosidade, iii) uso da variância da estimativa para verificar a precisão dos dados e iv) simulação sequencial direta para integrar os dados.

Soares (2005), verificou que em alguns estudos de caso, que nos algoritmos de co-simulação sequencial, as realizações foram globalmente validadas nas estatísticas de primeira ordem, variogramas, histogramas e coeficiente de correlação, mas o método não respeitou a correlação entre as variáveis primária e secundária para determinadas classes. Isso pode levar a conclusões equivocadas, principalmente quando funções de custo não linear (usado em aplicações ambientais e de mineração) são aplicadas ao conjunto de valores de co-simulados. Neste contexto, na melhoria de algoritmos de co-

simulação sequencial, Horta e Soares (2010) apresentaram uma implementação a nova versão do algoritmo co-simulação sequencial direta com distribuição conjunta (Co-DSS) para melhorar a representatividade das realizações baseado na lei Bayes. Neste trabalho verificou-se que o algoritmo reproduz a correlação entre as variáveis e as distribuições condicionais, mesmo quando o coeficiente de correlação entre as variáveis é baixo.

Liu e Journel (2009), propuseram o uso do algoritmo de simulação sequencial direta para integração dos dados em diferentes suportes (dados geofísicos e dados de poços). Já que o algoritmo utiliza os dados no espaço original e preserva a linearidade dos dados no bloco. Além disto, esta abordagem permite utilizar blocos com diferentes geometrias e layouts.

1.2.2. Sobre a integração de informação secundária imprecisa ou enviesada

Emery (2003) apresentou uma metodologia de estimativa não linear em que foram utilizados dados imprecisos e estes dados foram estimados por krigagem disjuntiva. Na metodologia proposta, o conjunto de dados imprecisos foi considerado como um conjunto de dados *pseudohard* simulados a partir de sua distribuição *a posteriori*, construída com os dados da variável primária e os dados secundários. Neste estudo, como as variáveis são independentes, os pesos de krigagem são calculados uma única vez, o que permitiu rápido processamento computacional. Três metodologias foram comparadas: a esperança condicional sob a hipótese de multigaussianidade, krigagem disjuntiva e krigagem de indicadores utilizando os dados secundários (*soft data*). Os estudos mostraram que a incorporação dos dados secundários depende do nível de imprecisão dos dados utilizados. Dentre as metodologias investigadas, a krigagem disjuntiva produz melhores resultados, desde que o modelo da distribuição dos dados *pseudohard* selecionado seja adequado aos padrões estruturais da variável de interesse.

Emery (2005) estudou o impacto dos dados imprecisos na avaliação de recursos e reservas em dois estudos de caso: um banco de dados real de um depósito de cobre que possui dados de exploração e produção e outro banco de

dados sintético de ouro criado a partir de uma simulação não condicional. Foram comparados os resultados das metodologias: krigagem ordinária (OK), cokrigagem tradicional (COK), krigagem lognormal (LogNK) e krigagem dos indicadores (IK). Concluiu-se que os dados imprecisos não devem ser descartados devido à baixa qualidade, pois a quantidade de dados prevalece à qualidade dos dados (efeito da informação).

Deraisme e Stradom (2009), em seu estudo, avaliaram a incorporação de amostras de ferro com distintos métodos amostrais na estimativa de recurso. As amostras analisadas possuíam heterotopia total e foram obtidas por furos de sondagem diamantados (DDH) e perfuração percussiva. Para incorporação, foram avaliadas as metodologias cokrigagem tradicional (COK) e a krigagem ordinária com a medida da variância do erro (OKVME). Esta metodologia é uma simples modificação no algoritmo de krigagem ordinária que consiste em adicionar a variância do erro de medição específico de cada amostra na diagonal da matriz do sistema de krigagem. Assim, este método requer o input do valor do ruído dos dados chamado de variância do erro de medida. Neste artigo, foi utilizado a variância do erro da cokrigagem. A consequência é que os pesos atribuídos aos dados com o aumento da incerteza/variância, ficam mais próximos dos pesos da média aritmética. Os resultados mostraram situações, que as amostras secundárias possuem apenas erros de imprecisão, a metodologia krigagem ordinária com a medida da variância do erro (OKVME) é satisfatória e garante menor viés e melhor qualidade das estimativas.

Sobre a utilização de cokrigagem para integração de dados imprecisos, enviesados e heterotópicos no processo de estimativa, Minnitt e Deutsch (2014), em seu estudo, criaram um banco de dados artificial a partir de simulações, com espaçamento amostral regular. Utilizam a cokrigagem ordinária standardizada, que combina diferentes conjuntos de dados com configuração heterotópica a partir de dados com diferentes qualidades. Os resultados mostraram que a cokrigagem ordinária standardizada fornece melhores estimativas de recursos minerais em escalas local e global, em comparação com krigagem ordinária da variável primária. Vê-se o benefício altamente significativo da incorporação da variável secundária, pois o viés e a imprecisão existentes devido à má amostragem nos dados secundários não são transferidos para os resultados das

estimativas. É importante ressaltar, que os limites admissíveis do coeficiente de correlação entre as variáveis primária e secundária para aplicação desta técnica afetam os resultados das estimativas. Se o coeficiente de correlação for igual ou superior a 0,70, o método sugerido funciona muito bem, abaixo de 0,20, significa que não existe correlação entre as variáveis, portanto, os resultados obtidos por krigagem e cokrigagem simples serão semelhantes. Para dados com correlação entre 0,20 e 0,70, os resultados podem ser questionáveis. Outros autores também utilizaram a metodologia cokrigagem ordinária estandardizada para avaliar o efeito da informação secundária imprecisa e inacurada nas estimativas, considerando dados heterotópicos em outros bancos de dados (Donovan e Deutsch, 2014 ; Araújo e Costa, 2015).

Araújo *et al.* (2018) estudaram um depósito de ouro com dois tipos de dados: amostras provenientes furos de sondagem (DDH) considerados como dados primários e circulação reversa como dados secundários (RC) em um mesmo domínio geológico. Estas amostras possuíam moderada correlação e pouca continuidade espacial. Neste estudo, foram investigadas duas metodologias: cokrigagem ordinária não-tradicional estandardizada (SCOK), krigagem ordinária (OK) e krigagem ordinária estandardizada (SOK). Para que os dados primários e secundários possam ser utilizados juntos na estimativa, foi proposto a correção do viés e da imprecisão dos dados secundários a partir da média e desvio padrão da variável primária. Essa metodologia foi denominada krigagem ordinária estandardizada (SOK). Foram avaliados a curva teor-tonelagem, análise de deriva e a classificação errônea dos blocos para comparar os modelos gerados. Os resultados da estimativa produzida mostraram que a correção do viés da variável secundária combinada à krigagem ordinária, produziram melhores resultados com estimativas menos enviesadas e melhor classificação dos blocos como minério e estéril.

Cornah e Machaka (2015) avaliaram a integração de amostras secundárias imprecisas e enviesadas obtidas por diferentes campanhas de amostragem e diferentes tipos de suporte. O estudo foi realizado em um banco de dados sintético de minério de ferro, e os erros foram adicionados artificialmente. Foram avaliadas as metodologias: krigagem ordinária (OK), cokrigagem tradicional (COK), cokrigagem multicolocada (MCCOK) e a krigagem

ordinária com a medida da variância do erro (OKVME). Os resultados mostraram que se as amostras secundárias forem imprecisas, sem viés e abundantes a integração dos dados pode ser feita por krigagem ordinária (OK), que diminui o efeito da falta de informação (exclusão dos dados secundários). Para dados secundários imprecisos e sem viés, em que a imprecisão pode ser avaliada e quantificada a partir dos dados de controle de qualidade, recomenda-se a metodologia krigagem ordinária com a medida da variância do erro (OKVME), que melhora a precisão das estimativas quando comparado a krigagem ordinária (OK) e ao cenário de referência. Em casos que os dados secundários possuem viés e imprecisão, a estimativa por cokrigagem tradicional (COK) é recomendada.

1.2.3 Abordagens recentes em geoestatística sobre a integração de informação secundária

Após o levantamento do referencial teórico acima, verificou-se que a integração de informação secundária em geoestatística nos últimos anos está sendo abordada de forma diferente que devem ser investigadas nessa tese.

Em situações em que os dados secundários são mais amostrados do que os primários, técnicas de inserção múltipla (MI) estão disponíveis e consistem em simular primeiro as localizações do secundário e em seguida, no restante dos nós do grid de simulação (Barnett e Deutsch, 2015; Soares *et al.*, 2017; Silva e Deutsch, 2018). Barnett e Deutsch (2015) propuseram duas metodologias de múltiplas imputações: paramétricas e não-paramétricas para imputar dados geológicos faltantes com base na atualização bayesiana para conjuntos de dados isotópicos gerados. O método paramétrico assume que a distribuição conjunta é Gaussiana multivariada, para construir a distribuição conjunta é necessário o coeficiente de correlação entre as variáveis. Já, para a abordagem não-paramétrica calcula a distribuição conjunta a partir dos gráficos de dispersão. Da mesma forma, os métodos utilizados por Soares *et al.* (2017) e Silva e Deutsch (2018) utilizaram os gráficos de dispersão para obter a distribuição conjunta. A limitação é que a construção de um gráfico de dispersão requer um subconjunto em que os dados primários e secundários sejam isotópicos. Como resultado, esses métodos não são aplicáveis quando o

conjunto de dados é completamente heterotópico. É importante ressaltar que, amostras com heterotopia total ou completa, ocorrem quando dois tipos diferentes de amostragem são realizados na mina, devido ao custo operacional da amostragem ser superior ao custo de análise química. Portanto, as variáveis primárias e secundárias são medidas em locais diferentes não tendo nenhum ponto em comum de coleta.

Sobre os algoritmos de simulação estocástica, a metodologia por simulação sequencial direta (DSS), ao contrário das metodologias simulação sequencial gaussiana (SGS) e simulação sequencial de indicador (SIS), não necessita de qualquer transformação da variável, utiliza os dados e os variogramas originais (SOARES, 2001). Tran *et al.* (2001) afirmam, que como a simulação sequencial direta não utiliza transformações não lineares, a relação linear entre as amostras de diferente suporte é mantida. Para isto, é necessário utilizar as covariâncias médias. A partir deste algoritmo, foram desenvolvidas novas implementações. Horta e Soares (2010), propuseram o algoritmo de co-simulação sequencial direta com distribuição conjunta, para a reprodução das distribuições bivariadas entre as variáveis primárias e secundárias. Em extensão deste algoritmo, Soares *et al.* (2017) propuseram a metodologia simulação sequencial direta com distribuições locais. Nesta abordagem, os dados experimentais incertos são modelados por uma distribuição de probabilidade em cada local de amostra. A idéia em caracterizar os dados secundários como *pdfs* e espacialmente correlacionados foi baseada no algoritmo de simulação p-field (Srivastava, 1992 e Froidevaux, 1993). A inovação desta metodologia é que esta abordagem pode ser aplicada em qualquer algoritmo de simulação estocástica. Neves *et al.* (2018) utilizaram o algoritmo de simulação sequencial direta com distribuições pontuais em uma mina subterrânea de cobre dentro da abordagem do projeto *Real Time Mining* (Osterholt e Benndorf, 2015). Foram utilizadas amostras secundárias obtidas por fluorescência de raios X portátil (XRF). Os resultados mostram resultados promissores com modelos de recursos minerais mais precisos quando as amostras foram integradas.

1.3 Meta

Neste contexto, o presente trabalho propõe investigar e propor novas metodologias para incorporar informação secundária completamente heterotópica cuja incerteza é considerada como distribuição de probabilidade espacialmente correlacionadas.

Hipótese Investigativa: *Avaliar se há redução da incerteza e aumento da acuracidade e precisão dos modelos de teores pela integração dos dados secundários, considerando-os como uma distribuição de probabilidade local, com configuração espacial heterotópica, abordando a qualidade e quantidade dos dados disponíveis.*

Analisar-se-á o ganho na qualidade da estimativa de teores (ARAÚJO, 2015), assim como a acuracidade e precisão, com o uso adicional de informação, avaliando os benefícios na melhoria do modelo de teores. Nesta proposta, a incorporação da incerteza da informação secundária é feita através de leis de distribuição de probabilidades. Acredita-se que com a entrada da variável secundária com várias possibilidades de valores (distribuição local), poderá ser utilizada para caracterizar a incerteza do fenômeno estudado. Esta abordagem é inovadora, já que a maioria dos softwares e metodologias em geoestatística não aborda esta questão

1.4 Objetivo

O objetivo geral desta tese é avaliar diferentes formas de incorporar a incerteza da informação secundária. Na primeira etapa destas metodologias apresentadas, sugere-se o pré-processamento da variável secundária denominado *pseudohard* que possui a distribuição acumulada e a continuidade espacial da variável primária. Em seguida, os dados primários são adicionados aos *pseudohard* gerados e o algoritmo de simulação estocástica é executado.

Para atingir a meta proposta são considerados os seguintes objetivos específicos:

- Desenvolver métodos para incorporar os erros presentes nas amostras secundárias no processo de estimativa/simulação nos modelos de teores.
- Investigar melhoria no processo com a incorporação das amostras secundárias como distribuição de probabilidade, obtendo um modelo que represente de maneira mais confiável os teores do minério a ser lavrado.
- Desenvolver (adaptar) metodologias que considere a informação secundária como uma distribuição de probabilidade.

1.5 Organização da tese

A tese propõe investigar metodologias para incorporar dados secundários como distribuição de probabilidade e avaliar a precisão e acuracidade dos modelos obtidos. Neste trabalho, foram avaliadas duas possíveis metodologias que serão discutidas nos capítulos a seguir:

O capítulo 2 revisa e discute o algoritmo de simulação sequencial direta com distribuições pontuais proposta por Soares *et. al* (2017) em que o *pseudohard* é inferido por krigagem simples. Neste capítulo, propõe uma nova implementação em que o *pseudohard* é inferido pelo algoritmo de cokrigagem multicolocada proposto por Babak e Deutsch (2009). Nesta proposta será discutido, sobre a reprodução da distribuição pontual local, suas limitações e vantagens. Esta metodologia é recomendada para dados de diferentes suportes.

O capítulo 3 propõe uma metodologia para incorporar dados secundários como distribuição local, pode ser executada com qualquer algoritmo de simulação estocástica e para dados que estejam no mesmo suporte. Utilizamos a técnica de simulação por bandas rotativas (Journel e Huijbregts, 1978). Foram avaliadas 03 metodologias para construção da distribuição de probabilidade. Na primeira abordagem, é utilizada somente a variável secundária no local da distribuição. Na segunda, é utilizada a variável secundária no local e a variável primária e secundária na vizinhança da distribuição para construir a distribuição de probabilidade. Nestas propostas, a distribuição de probabilidade da variável

secundária (*cdf*'s) é amostrada aleatoriamente com o algoritmo de simulação *p-field* (Srivastava, 1992 e Froidevaux, 1993). Na terceira abordagem, a distribuição da variável é construída a partir da atualização bayesiana (BU). Detalhes sobre atualização bayesiana são encontrados em Barnett e Deutsch (2015). A partir da distribuição posterior obtida por atualização bayesiana, que possui informações da variável primária e secundária, valores da variável primária são amostrados sequencialmente e inferidos em cada local da variável secundária. Nesta proposta, os dados amostrados das distribuições foram considerados *pseudohard*. Assim, um novo conjunto de dados foi criado contendo os dados da variável primária e os *pseudohard*. Em seguida, o algoritmo de simulação estocástica é executado para gerar cenários equiprováveis. Para ilustrar a metodologia, foi utilizado o mesmo banco de dados do capítulo 02. Os resultados foram comparados com os modelos obtidos por simulação estocástica utilizando apenas a variável primária honrando o modelo de continuidade espacial da variável primária.

O capítulo 4 resume as principais contribuições da tese, avalia as vantagens e limitações de cada metodologia proposta e a sugestões para trabalhos futuros.

1.6 Contribuição da tese

Muitos trabalhos no tópico de combinar informação primária e secundária foram realizados, por exemplo em Isaaks e Srivastava, (1989); Goovaerts, (1997); Soares, (2001); Rossi e Deutsch (2014), e serão discutidos no item 1.5-Revisão bibliográfica. No entanto, há ainda muitos questionamentos quanto à maneira para incorporação dos dados secundários e sobre como incluir os dados secundários como uma função da distribuição de probabilidade. Considerando os trabalhos publicados sobre o tema semelhante ao dessa tese, algumas contribuições desta tese seriam:

- Avaliar a incorporação dos dados secundários como uma função de distribuição de probabilidade (incerteza dos dados secundário). Na abordagem tradicional dos algoritmos de interpolação, somente o algoritmo de krigagem de indicadores para integração de informação secundária (*soft indicator kriging*), considera os dados secundários como

uma distribuição local (*cdf*'s), quando a imprecisão dos dados não pode ser ignorada. Os demais algoritmos assumem a variável secundária local como um *datum*, e ignoram o erro da informação. Assim, acredita-se, que ao representar a amostra secundária como uma distribuição probabilidade, a incerteza no local a ser estimado/simulado será caracterizada e incorporada no modelo final.

- Verificar a possibilidade de incorporar dados secundários com configuração espacial completamente heterotópica, garantindo a reprodução da continuidade espacial da variável primária.

2 SIMULAÇÃO SEQUENCIAL DIRETA COM DISTRIBUIÇÕES PONTUAIS (DISTRIBUIÇÃO DE PROBABILIDADE NO LOCAL DA SECUNDÁRIA)

A simulação estocástica é uma abordagem geoestatística para avaliar a incerteza local e espacial das amostras em local não amostrados e produzir cenários igualmente prováveis (Journel, 1974; Journel e Huijbregts, 1978). Os modelos obtidos reproduzem as estatísticas das amostras (histograma e modelo de continuidade espacial) e honram os dados da amostra em seus locais originais. Já os algoritmos de co-simulação utilizam a informação secundária para condicionar a simulação da variável primária. A informação secundária a ser utilizada pode referir-se à mesma variável primária, mas apresentada em um formato diferente, ou qualquer outra variável à qual a variável primária está correlacionada.

Dentre os algoritmos de simulação sequencial, a primeira abordagem foi baseada na simulação sequencia Gaussiana (Isaaks, 1990) e a co-simulação (*Sequential Gaussian Co-simulation*) foi proposta por Verly (1992). O último requer as covariâncias diretas e cruzadas das variáveis explicadas em Journel e Huijbregts (1978), Goovaerts (1997) e Wackernagel (2003). Situações que os dados secundários são conhecidos em todos os locais a serem estimados, são utilizados os modelos de hipótese Markov-Bayes para descrever o modelo de continuidade espacial entre as variáveis coregionalizadas conhecidos como MM1 e MM2 que foram propostos por Almeida e Journel (1994) e Journel, (1999).

A simulação sequencial com formalismo de indicadores, conhecida como simulação sequencial de indicadores (SIS), foi proposta por Alabert (1987) e permite a abordagem da cokrigagem completa e dos modelos de Markov-Bayes (Zhu 1991; Zhu e Journel 1992) para integrar diferentes tipos de dados.

A abordagem que usa as variáveis originais, sem qualquer transformação prévia ou posterior e tem mais aplicações em estudos de caso de petróleo e situações onde o *downscale* dos dados é necessário foi proposta Soares (2001) conhecida como simulação e co-simulação sequencial direta (*DSS-Direct*

Sequential Simulation). Melhorias desses algoritmos foram propostas por Horta e Soares (2010) e Soares et al. (2017) que serão investigadas a seguir.

Este capítulo investiga o algoritmo de simulação sequencial direta com ênfase na sua última implementação, o algoritmo de simulação sequencial direta com distribuições pontuais. É proposto uma forma alternativa de integrar dados secundários com um algoritmo híbrido que combina duas abordagens diferentes para construir as distribuições condicionais (krigagem/cokrigagem) para melhorar a reprodução da distribuição local. Para ilustrar a metodologia proposta, será aplicado em um estudo de caso real em uma mina subterrânea de cobre, com dados totalmente heterotópicos, destacando como construir uma distribuição de ponto local. Para implementação é sugerido um fluxograma de trabalho.

Para avaliar a acurácia e precisão das metodologias investigadas, será utilizada a técnica de validação *Jackknife* (Efron, 1982). Serão comparados com os modelos obtidos pela simulação direta sequencial (DSS) condicionado somente com a variável primária, simulação sequencial direta com distribuição pontual com a abordagem proposta por Soares *et. al* (2017) utilizando krigagem simples e o algoritmo de simulação sequencial direta com a implementação proposta usando cokrigagem simples.

2.1 Fundamentação teórica sobre as metodologias abordadas

2.1.1. Simulação sequencial direta (*Direct sequential simulation*)

O desenvolvimento dos algoritmos de simulação geoestatística iniciou-se com Journel (1994). Ele verificou que a reprodução do modelo de covariância dos dados ocorre desde que os valores simulados sejam amostrados de distribuições locais centradas na média e variância das estimativas de krigagem simples. O algoritmo de simulação sequencial direta (DSS) foi desenvolvido por Soares (2001), não requer a hipótese de dados terem uma distribuição multiGaussiana. Depois da média e variâncias locais serem estimadas, o valor

simulado é gerado, não a partir de uma cdf global, mas a partir da distribuição de probabilidades local. Assim, as vantagens deste algoritmo consistem em:

- i) reproduzir o variograma dos modelos simulados no espaço original,
- ii) preservar as propriedades lineares dos dados com diferentes suportes e volumes, o que não é possível em algoritmos que utilizam técnicas gaussianas, como: simulação sequencial Gaussiana, método das bandas rotativas. Maiores detalhes podem ser consultados em Tran *et al.* (2001).

A figura 2.1 exemplifica a amostragem da distribuição global $F_z(z)$, dos dados (linha azul) no algoritmo simulação sequencial direta. Este algoritmo considera que os dados possam seguir qualquer distribuição de probabilidades, como por exemplo: uniforme, gaussiana, etc. Para amostrá-la, foi escolhido um intervalo da distribuição identificado com o número 1 (linha cinza) da variável z (x) da distribuição global $F_z(z)$ (linha azul), que é centrado na média e variância de krigagem da variável z (x_0). Em seguida, o correspondente gaussiano $y(x_0)^*$ da estimativa local intervalo $z(x_0)^*$ é identificado na distribuição gaussiana global dos dados (linha vermelha) construída a partir dos dados da variável primária. Este intervalo foi identificado como número 2 (linha cinza) da distribuição gaussiana global (linha vermelha). Assim, o valor simulado $z^s(x_0)$ é amostrado deste intervalo local $G(y(x_0)^*, \sigma_{SK}^2(x_0))$.

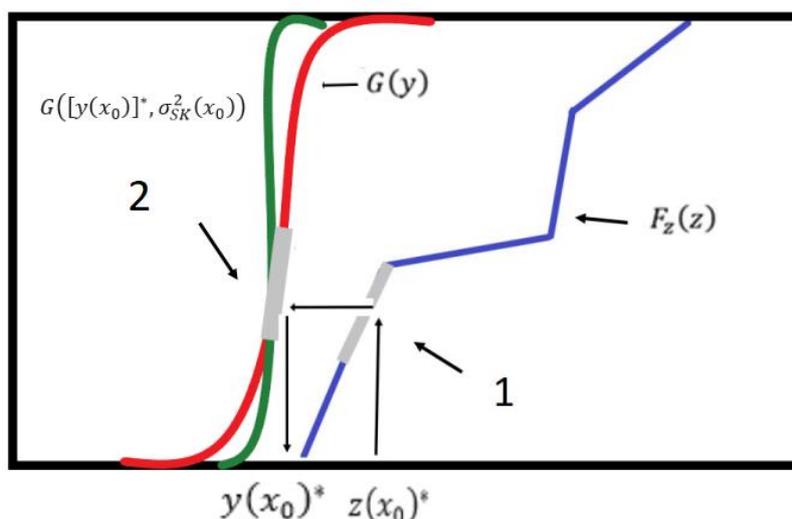


Figura 2. 1- Amostragem da distribuição global $F_z(z)$ definida pelos intervalos da média local e variância de $Z(x_i)$. FONTE: SOARES (2001)

Em resumo, após calcular a função $F_z'(z)$, que corresponde ao intervalo da distribuição no local a ser simulado (item 1 da figura 2.1), a simulação sequencial direta procede como segue:

- i) Definir o percurso aleatório sobre o conjunto de nós x_0 de uma malha regular a ser simulada.
- ii) Estimar a média e variância no local x_0 , identificadas, respectivamente, como o estimador por krigagem simples $[z(x_0)]^*$ e a variância da estimativa $\sigma_{SK}^2(x_0)$, condicionadas aos valores experimentais $z(x_i)$ e aos valores previamente simulados $z^l(x_i)$.
- iii) Determinar o intervalo da $F_z(z)$ a ser re-amostrado tendo por base a distribuição gaussiana:

$$G([y(x_0)]^*, \sigma_{SK}^2(x_0)) \text{ em que } [y(x_0)]^* = \varphi([z(x_0)]^*).$$

- iv) Gerar valor simulado $z^s(x_0)$ da *cdf* $F_z(z)$.
 - a) Gerar um valor simulado p de uma distribuição uniforme entre $U(0,1)$.
 - b) Gerar um valor y^l de $G([y(x_0)]^*, \sigma_{SK}^2(x_0))$:

$$y^l = G^{-1}([y(x_0)]^*, \sigma_{SK}^2(x_0), p)$$
 - c) Cálculo do valor simulado $z^l(x_0) = \varphi^{-1}(y^s)$
- v) Retorna a i) até todos os nós forem simulados.

Bourgault (1997) e Caers (2000a) fizeram demonstrações práticas sobre os vários tipos de distribuições locais. Verificaram que, com exceção de algumas distribuições paramétricas (por exemplo, a gaussiana) esta abordagem de simulação não reproduz os histogramas. Segundo Rossi e Deustch (2014), esta limitação se deve ao fato que a forma do histograma no local a ser simulado ser influenciado por três fatores: a formato do histograma dos dados originais, a escolha do formato da distribuição em que será amostrada de forma aleatória e a distribuição gaussiana local resultante. Para obter um histograma local adequado, vários autores propuseram metodologias para corrigir o histograma resultante como: Caers (2000b); Nowak e Srivastava (1997); Deutsch *et al.* (2001); Oz *et al.* (2003) e Soares (2001). Com as sugestões desses autores e a reprodução do histograma é garantida.

A respeito da reprodução média da variável original, a simulação sequencial direta não consegue reproduzir as médias em classes com baixa frequência. Soares (1998), mostra que este fato é comum nos algoritmos de simulação sequencial (SGS e SIS), este fato se deve ao caminho aleatório durante a visita dos nós próximos aos valores extremos. Para isto, foi proposto um algoritmo para a correção das probabilidades locais, o que implica a reprodução das proporções de uma variável categórica nos mapas simulados finais. A partir desta ideia foi proposto um parâmetro para correção da média local e para as classes de baixa frequência.

$$E\{y^s(x_0)\} \neq z(x_0)^* \quad (2.1)$$

A correção de viés pode ser corrigido pela equação 2.1:

$$z^s(x_0) = z^s(x_0) + [z^*(x_0) - z^s(x_0)] \quad (2.2)$$

onde $z^s(x_0)$ é o valor simulado para o nó (x_0) , é corrigido pelo desvio entre $z^*(x_0) - z^s(x_0)$. Esta equação permite a correção do viés e melhor reprodução do histograma $z(x)$, em que existem poucos dados.

Para variáveis categóricas, Almeida e Journel (1994) utilizam a mesma correção para as probabilidades locais para reproduzirem a tendência dos locais para as variáveis categóricas. A partir deste conceito, o padrão não estacionário foi caracterizado para compreender melhor os locais $z(x)$, corrigindo as estimativas locais utilizando a média local.

Supondo que m_z corresponde a média dos dados originais $z(x)$ e esta deve ser reproduzida no mapa final das simulações. O desvio e^s pode ser calculado a partir da diferença entre a média dos dados originais (m_z) e a média dos dados simulados $z(x_0)^s$ (equação 2.3).

$$e^s = m_z - m_z^s \quad (2.3)$$

onde m_z^s é media dos valores simulados e N_s o número de valores simulados pode ser calculada da seguinte forma (equação 2.4):

$$m_z^s = \frac{1}{N_s} \sum_{i=1}^{N_s} z^s(x_i) \quad (2.4)$$

Assim, a média local pode ser corrigida tanto na simulação quanto na co-simulação conforme a equação 2.5.

$$z(x_u)_s^* = z(x_u)^s + e^s \quad (2.5)$$

2.1.2 Simulação sequencial direta com distribuição pontual

Neste estudo, será utilizado simulação sequencial direta com distribuições pontuais que é uma extensão da simulação sequencial direta. Esta metodologia consiste, em durante o processo da simulação sequencial, visitar primeiro os locais onde há dados impreciso expresso por distribuições pontuais (*pseudohard*) e em seguida completar o restante dos nós do grid.

Nesta abordagem, Soares *et al.* (2017) propõe caracterizar a incerteza dos dados como uma distribuição de probabilidade (*pdf*). Nesta metodologia, a incerteza da medição de dados em um dado local x_β será interpretada como uma distribuição de probabilidade local identificada ($F(z(x_\beta))$), e representado por um conjunto de z possíveis valores. A partir da definição da incerteza na localização de dados x_β , expressa através de uma *pdf* local, são geradas realizações de valores possíveis de dados, tirados da função de distribuição de probabilidade local $F(z(x_\beta))$, definidos nos locais dos dados experimentais x_β . A simulação de valores estruturados dos dados incertos é anterior à simulação dos demais nós do grid. Assim, os dados experimentais são definidos como x_α e também como a distribuição $F(z(x_\beta))$ que está associado as medidas de dados incertos, no caso a variável secundária definidos como distribuições pontuais.

As vantagens desta metodologia proposta são que:

- i) pode ser utilizada para incorporar variáveis contínuas, categóricas e dados com suporte de blocos.
- ii) reproduz a continuidade espacial da variável primária.
- iii) Integra o dado secundário incerto sem necessidade de modelo linear de correção regionalização.

Para que este algoritmo seja executado, é necessário que o banco de dados possua a variável primária e sua incerteza caracterizada por uma distribuição probabilidade (*pdf*). Soares *et al.* (2017) e Neves *et al.* (2018) utilizaram os gráficos de dispersão para obter a distribuição probabilidade locais, mas outras metodologias podem ser investigadas. A figura 2.2a exemplifica um banco de dados com configuração espacial totalmente heterotópica que possuem dados primários (x_α), representados por pontos azuis e dados secundários (x_β) que correspondem aos pontos vermelhos. A figura 2.2b mostra o banco de dados que será utilizado na simulação, onde os pontos azuis representam a variável primária e os histogramas em vermelho representam a distribuição de probabilidade da variável secundária.

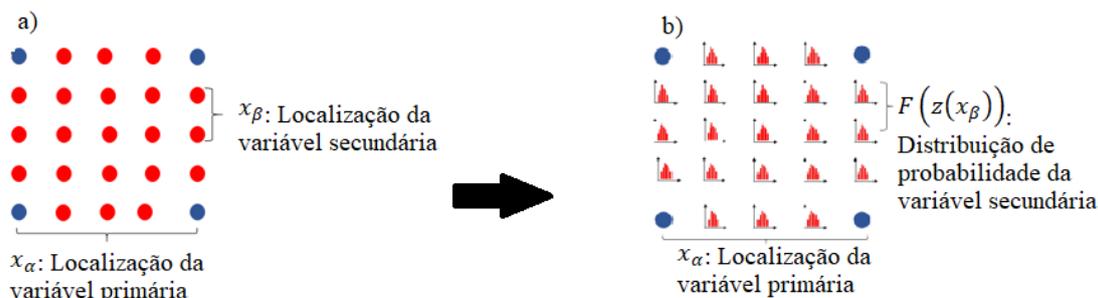


Figura 2. 2-Exemplo de amostras primárias (pontos azuis) e variáveis secundárias (pontos vermelhos) em configuração de heterotopia total

A grande vantagem desta metodologia é gerar valores simulados auto correlacionados espacialmente a partir das distribuições de probabilidade, chamados de *pseudohard*. Este processo é realizado em duas etapas. Na primeira etapa, a média local e variância no local dos dados experimentais $z(x_\beta)$ são calculados e condicionados aos dados primários $z(x_\alpha)$ na vizinhança e também com aos dados previamente simulados da distribuição de probabilidade $F(\mathbf{Z})$ em locais diferentes de x_β . Em seguida, são amostrados da distribuição que é centrada na média local e variância, conforme procedimento da simulação seqüencial direta (SOARES, 2001), descrito na figura 2.1. Assim, estes valores simulados (*pseudohard*) irão reproduzir a continuidade espacial dos dados.

Nesta etapa geram-se primeiro os valores N_d nos locais de dados de amostras experimentais, $x_\beta = 1, \dots, N_d$, usando a distribuição local de $F(z(x_\beta))$. No local da amostra x_β , a média e a variância de $z(x_\beta)$ são calculados por krigagem simples baseado nos dados primários conhecidos x_α e nos nós previamente simulados z^l dos dados secundários x_β . O valor $z^l(x_0)$ na localização experimental é retirado randomicamente a partir da distribuição local $F(z(x_0))$ centrado na estimativa krigagem simples da média local.

$$z(x_0)^* - m = \sum_{\alpha} \lambda_{\alpha}(x_0) [z(x_{\alpha}) - m] + \sum_{\beta} \lambda_{\beta}(x_0) [z^l(x_{\beta}) - m] \quad (2.6)$$

A figura 2.3a mostra o banco de dados que possui amostras primárias (pontos azuis) e distribuição de probabilidade da variável secundária (histograma vermelho). A figura 2.3b mostra o banco de dados após a primeira etapa do algoritmo, a simulação de valores nos locais das distribuições pontuais para gerar os *pseudohard*, que são autocorrelacionados e possuem informações da variável primária e secundárias obtidas por simulação sequencial direta.

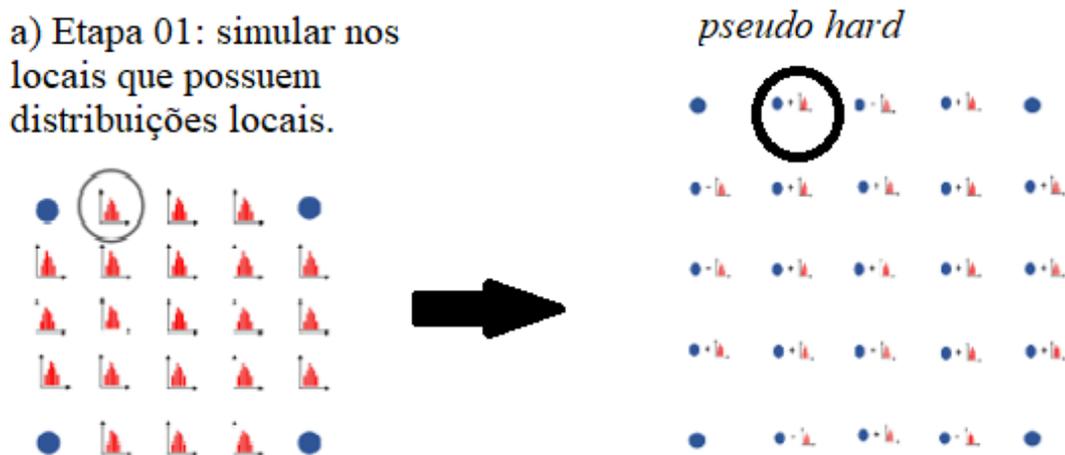


Figura 2. 3-Cálculo dos valores simulados no local da distribuição probabilidade local (*pseudohard*)

Na segunda etapa, são utilizados a variável primária e os valores simulados nos locais que possuem distribuições pontuais (*pseudohard*) para simular o restante de nós do grid com o algoritmo de simulação sequencial direta. Para

cada realização, um novo banco de dados é gerado antes do cenário (grid) completo ser simulado. A figura 2.4a mostra os dados da variável primária e os *pseudohard* dados simulados das distribuições. A figura 2.4b mostra a segunda etapa da metodologia em que os nós do grid restante foram simulados utilizando o algoritmo de simulação sequencial direta.

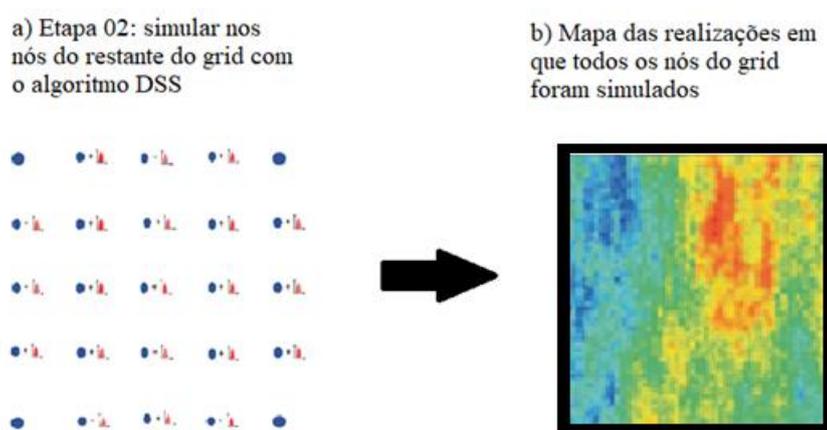


Figura 2. 4- a) Banco de dados que será utilizado na simulação sequencial direta—b) cenário final com todos os nós dos grids simulados

2.1.3 Investigação da reprodução da distribuição (*pseudohard*) inferidos por krigagem simples

Para investigar a reprodução da distribuição foi criado um banco com dados sintéticos, em 2D, com 30 amostras. Neste conjunto de dados, a mediana e a média são iguais, o que indica que não possui agrupamento preferencial, pois as amostras estão regularmente espaçadas (Figura 2.5).

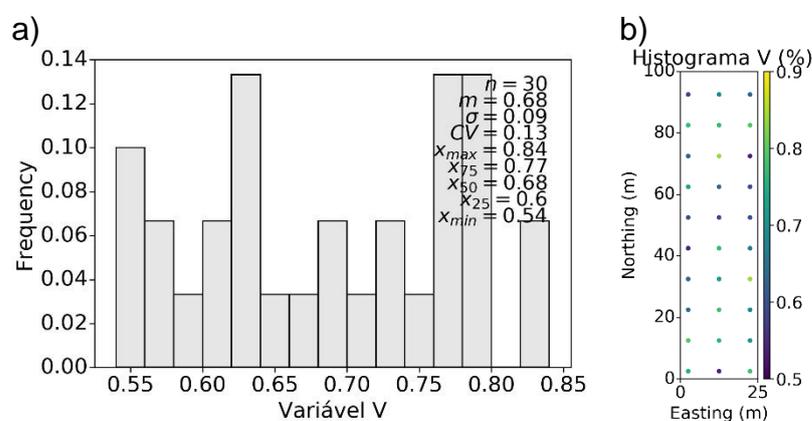


Figura 2. 5- Histograma e mapa de localização da variável V

Para avaliar a reprodução da distribuição, foram criadas três distribuições, identificadas como: d_1 , d_2 e d_3 . A tabela 2.1 mostra a estatística descritiva das distribuições a serem avaliadas. Na distribuição d_1 , a média da distribuição é próximo a média da variável V ($m_d \approx m_V$), a distribuição d_2 , a média da distribuição é maior que a média da variável V ($m_d > m_V$) e na distribuição d_3 , a média da distribuição é menor que a média da variável V ($m_d < m_V$).

Tabela 2. 1- Estatística descritiva das distribuições d_1, d_2 e d_3

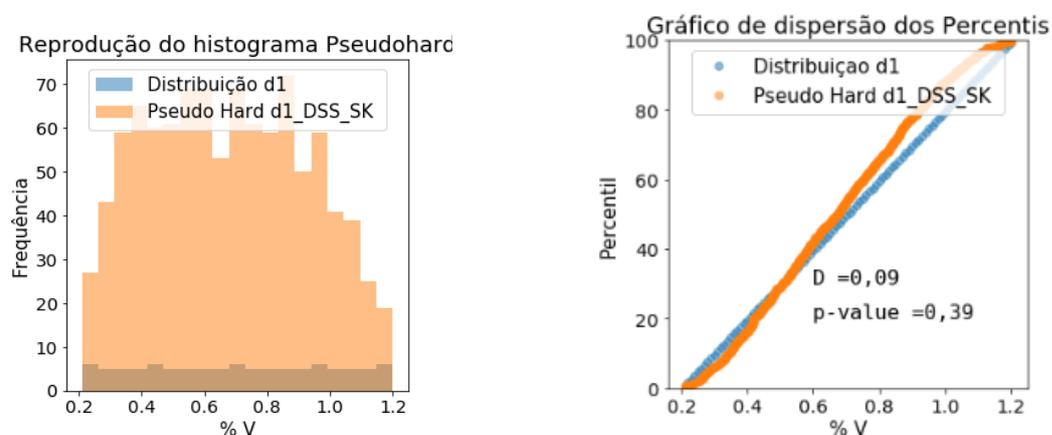
Identificação da distribuição	Nº de dados	Média	Mínimo	Máximo
d_1	100	0,70	0,21	1,20
d_2	100	1,01	0,52	1,51
d_3	100	0,45	0,25	0,61

Para verificar a reprodução das distribuições, foram criados 1000 dados de *pseudohard* e foi realizado o teste Kolmogorov-Smirnov (teste SK), chamada de estatística D . Este é um teste utilizado para comparar distribuições de probabilidade com uma distribuição de probabilidade de referência. O valor D (Equação 2.7) pode ser definido como a diferença entre a distribuição acumulada de referência, no caso, a distribuição dos dados $F(x)$ e a distribuição acumulada dos *pseudohard* gerados $F_n(x)$. Quando a hipótese nula (H_0) é verdadeira, significa que os dados das duas distribuições seguem a mesma distribuição de probabilidade. Para isto, o valor da probabilidade de significância (*p-value*) deve ser maior que o valor da estatística D . Isto significa que, em situações onde a probabilidade de significância (*p-value*) é pequena, a probabilidade de obter-se um valor da distribuição analisada similar à distribuição de referência é muito improvável, levando assim à rejeição da hipótese nula. Foi comparado também o formato das distribuições a partir do gráfico de dispersão entre os percentis das distribuições de referência (d_1, d_2 e d_3) e as distribuições de *pseudohard* gerados.

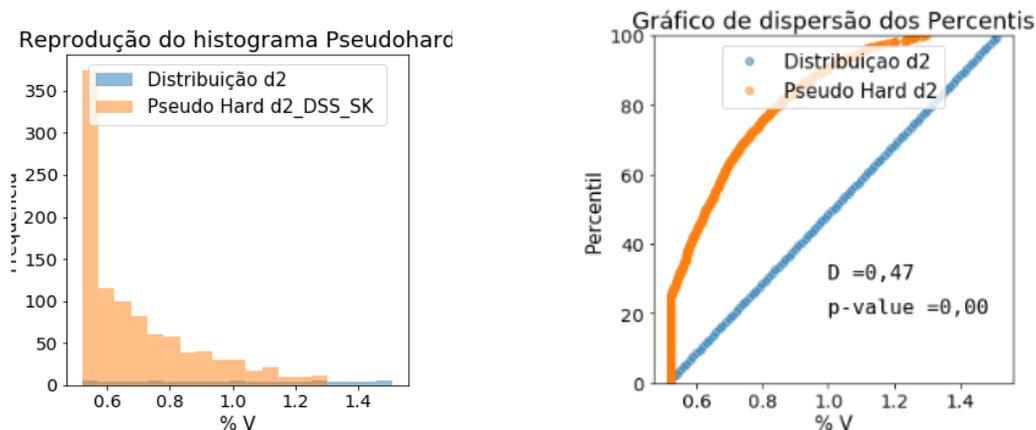
$$D = |F(x) - F_n(x)| \quad (2.7)$$

A figura 2.6 mostra o histograma, o gráfico de dispersão entre os percentis das distribuições de referência (representado em azul) e o *pseudohard* gerado (representado em laranja) e os resultados para o teste Kolmogorov-Smirnov, o valor D e a sua probabilidade de significância (p -value). Para gerar os pseudohards, foi utilizada na estratégia de busca o mesmo alcance do variograma. Sobre a distribuição d_1 (figura 2.6a) em que a média da distribuição é próxima à média dos dados, com o teste Kolmogorov-Smirnov, podemos afirmar que a distribuição dos *pseudohard* gerados é estatisticamente (p -value $> D$) igual e possui mesmo formato que a distribuição de referência d_1 . Para distribuições em que a média da distribuição é maior ou menor que a média dos dados da variável primária, isto não ocorre. Para a distribuição d_2 (figura 2.6b), a maior frequência de *pseudohard* gerados encontra-se nos valores próximos ao mínimo da distribuição (0,52) que é próximo a média da variável primária (0,68). Na distribuição d_3 (figura 2.6c), ocorre o contrário, a maior frequência de *pseudohard* gerados é próxima ao máximo da distribuição (0,61) que é próximo a média da variável primária (0,68). Portanto, a distribuição dos *pseudohard* gerados não são estatisticamente iguais (p -value $< D$) e não possuem o mesmo formato da distribuição dos dados originais. Sobre a reprodução da distribuição usando esta abordagem, podemos concluir que a distribuição gerada será estatisticamente semelhante e com o mesmo formato da distribuição de referência quando a média da distribuição de probabilidade local for próxima da média da estimativa naquele local.

a)



b)



c)

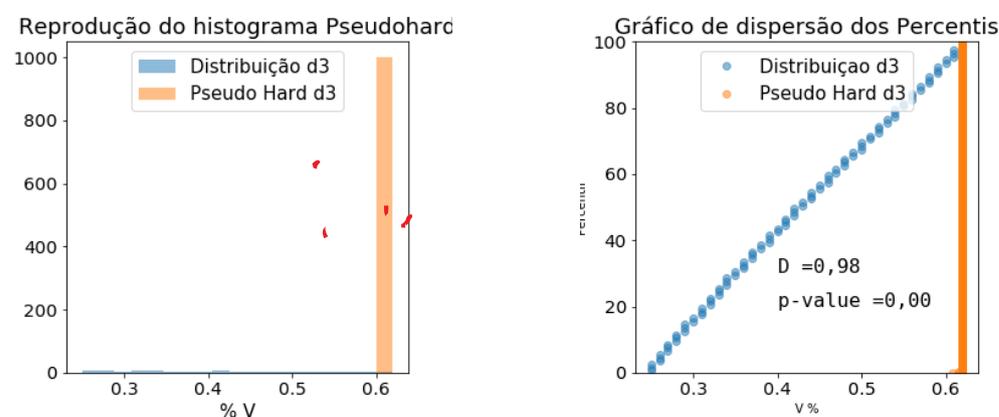


Figura 2. 6- Histograma, gráfico de dispersão e resultados do teste Kolmogorov-Smirnov para as distribuições a) d_1 , b) d_2 e c) d_3

2.1.4 Considerações sobre a reprodução da distribuição local (*pseudohard*)

Sobre a reprodução da continuidade espacial da distribuição local, conforme explicado no item 2.1.3, os *pseudohard* gerados são amostrados do intervalo de uma distribuição Gaussiana local centrada $G([y(x_0)]^*, \sigma_{SK}^2(x_0))$. Isto garante que valores de *pseudohard* inferidos a partir da distribuição de probabilidade reproduzam a continuidade espacial da variável de interesse (Journel, 1994). Os algoritmos de simulação sequencial consideram também que os dados originais e os dados previamente simulados como sendo dados primários e estes auxiliam na reprodução do variograma.

Sobre a reprodução da distribuição foi verificado que alguns fatores podem influenciar termos um bom ou não tão bons resultados. Dentre eles podemos destacar:

i) os dados da variável primária na vizinhança utilizados para condicionar a distribuição de probabilidade local. Conforme mostrado na equação 2.1, a estimativa local depende da média da variável primária e o peso recebidos pelos dados no entorno do local onde encontram-se os *pseudohards*. Os pesos recebidos (Equação 2.8) estão condicionados ao modelo de continuidade espacial e a configuração dos dados. Quando o local a ser estimado encontra-se distante dos dados, os pesos recebidos pelas amostras tendem a diminuir e a média da variável de interesse recebe peso maior. Portanto, a estimativa $Z^*(u)$ fica a média dos dados. Em situações, que a distribuição de probabilidade local tiver uma escala de valores diferente da média da variável primária condicionante (Figura 2.7a), a média global dos dados receberá peso maior e o *pseudohard* gerado ficará próximo a média dos dados primários condicionantes, no caso *hard data* (Figura 2.7b).

$$\lambda_m = 1 - \sum_{i=1}^n \lambda_{\alpha}(x_0) \quad (2.8)$$

onde λ_m é o peso recebido pela média λ_{α} são os pesos recebidos nas amostras do entorno do local a ser estimado

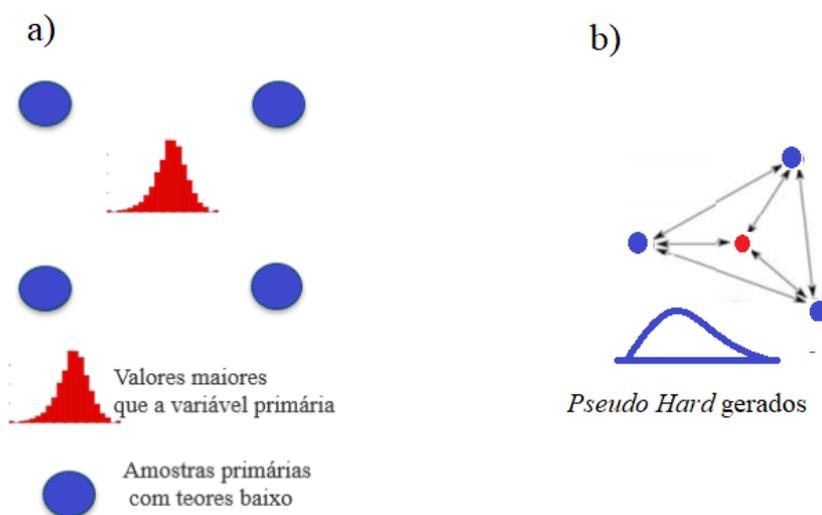


Figura 2. 7- Exemplo da configuração amostral entre os dados da variável primária (pontos azuis) precisos e acurados e a distribuição de probabilidade dos dados (vermelho) representando os dados imprecisos e acurados

ii) caminho aleatório: consiste em localizações pré-determinadas para se desenvolver a simulação estocástica. Este caminho é feito de forma aleatória garantindo que a simulação nunca seguirá a mesma sequência de visitação dos nós do grid. Assim, os dados da variável primária que serão utilizados para condicionar os *pseudohard* gerados serão diferentes em cada simulação, cada caminho aleatório poderá gerar uma vizinhança diferente, o que influenciará no valor do *pseudohard* gerado.

2.2 Metodologia proposta

A partir das considerações sobre a reprodução da distribuição locais, abordadas no item 2.1.3 e 2.1.4, para inferir os *pseudohard* foi investigado a utilização do algoritmo de cokrigagem. Espera-se melhorar a reprodução de distribuições que a média da distribuição (d_2 e d_3) são diferentes da média dos dados. A proposta será baseada na abordagem proposta por Babak e Deustch (2009), do algoritmo de cokrigagem intrínseca (ICCK), em que a variável secundária encontra-se exaustiva e leva em consideração a correlação entre a variável primária e a secundária e a variável secundária tanto nos locais da secundária quanto da primária. Neste estudo, as amostras primárias e secundárias possuem configuração amostral de heterotopia total, em que as amostras primárias e secundárias nunca estão no mesmo local. A variável secundária neste estudo serão os valores amostrados da distribuição local de $F(z(x_\beta))$. Assim, a variável primária e os valores amostrados das distribuições locais recebem peso no processo de estimativa. A cokrigagem colocada intrínseca (ICK) é uma cokrigagem colocada com correção da variância de estimativa. São utilizados os dados secundários no local a ser estimado e nos outros locais colocados, os dados primários e o modelo de correionalização é construído com o coeficiente de correlação entre os dados primários e secundários. No estudo realizado por Babak e Deustch (2009), os resultados mostraram que esta abordagem evita o aumento da variância, melhora a reprodução da correlação entre a variável primária e secundária e também a reprodução do variograma da variável de interesse no modelo final.

Nesta implementação, quanto maior a correlação entre os dados, maior será o peso recebido dado retirado dela distribuição ($F(z(x_\beta))$). Nesta proposta, será implementado um algoritmo híbrido. Na primeira etapa é realizada amostragem estruturada dos valores simulados da distribuição probabilidade da variável secundária ($F(z(x_\beta))$) e em seguida, é executado o algoritmo de cokrigagem colocada simples (Equação 2.3), para construir a distribuição que os valores *pseudohards* serão amostrados. Na segunda etapa, a variável primária e os n -cenários de *pseudohard* são utilizados para completar o restante do grid utilizando o algoritmo de simulação sequencial direta.

Nesta nova abordagem, pretende utilizar mais amostras primárias e secundárias correlacionadas para condicionar a distribuição de probabilidade local a ser construída. Para isto, a construção da distribuição probabilidade em cada local da variável secundária será realizado com a estimativa por cokrigagem simples local. Assim, a distribuição de probabilidade que será utilizada para amostrar os *pseudohards* será centrada no valor da estimativa e na variância de cokrigagem local.

2.2.1 Investigação da reprodução da distribuição (*pseudohard*) inferidos por cokrigagem intrínseca

Para verificar o efeito da correlação nos pesos atribuídos à variável primária e na variável secundária retirada das distribuições, foi realizado um teste com as distribuições de referência (d_1 , d_2 e d_3), que se encontram em três posições diferentes. A tabela 2.2 mostra os pesos recebidos pela variável primária e a retirada da distribuição de probabilidade da variável secundária que possuem alta correlação (0,90). Nestas situações, os valores retirados das distribuições recebem mais peso que a variável primária e a média da variável primária recebe peso nulo, conforme esperado.

Tabela 2. 2 - Pesos recebidos pela amostra retirada da distribuição de probabilidade com alta correlação (0,90) com a variável primária.

Pesos recebidos				
	Variável primária	Distribuição de probabilidade	Média	Soma dos pesos
d_1	0,15	0,85	0,00	1,00
d_2	0,16	0,84	0,00	1,00
d_3	0,13	0,87	0,00	1,00

A tabela 2.3 mostra os pesos recebidos pela variável primária e pela amostra retirada da distribuição de probabilidade da variável secundária que possuem fraca correlação (0,10). Nestas situações, os pesos recebidos são similares aos *pseudohard* gerados por krigagem simples, a variável e a média da variável primária recebem mais peso que a amostra retirada da distribuição local.

Tabela 2. 3 Pesos recebidos pela amostra retirada da distribuição de probabilidade com fraca correlação (0,10) com a variável primária.

Pesos recebidos				
	Variável primária	Distribuição de probabilidade	Média	Soma dos pesos
d_1	0,57	0,00	0,43	1,00
d_2	0,56	0,00	0,44	1,00
d_3	0,50	0,00	0,50	1,00

A figura 2.8 mostra o histograma e o gráfico de dispersão dos percentis da distribuição do *pseudohard* gerados. Em situações em que a distribuição da variável secundária possui forte correlação (0,90) com a primária (figura 2.8a), para a distribuição $d_1(m_d \approx m_v)$, o formato da distribuição dos *pseudohard* gerados (linha verde) é semelhante ao algoritmo proposto por Soares *et al.* (2017). Para as distribuições $d_2(m_d > m_v)$ e $d_3(m_d < m_v)$, ilustradas nas figuras 2.8b e 2.8c, a distribuição dos *pseudohard* gerados (linha verde) encontram-se mais próximos aos valores da distribuição de referência (linha azul), melhorando o formato da reprodução da distribuição *pseudohard* gerados.

Por definição, quando a variável primária possui fraca correlação com a variável secundária, a variável secundária receberá pouco peso, conforme demonstrado na tabela 2.3. Assim, os resultados das estimativas por cokrigagem simples são semelhantes a estimativa por krigagem simples. Portanto, a reprodução das distribuições de referência (d_1 , d_2 e d_3), serão semelhantes a implementação proposta por Soares *et al.* (2017), discutidas no item 2.1.3- Investigação da reprodução da distribuição (*pseudohard*) inferidos por krigagem simples.

a)

Reprodução do histograma Pseudohard

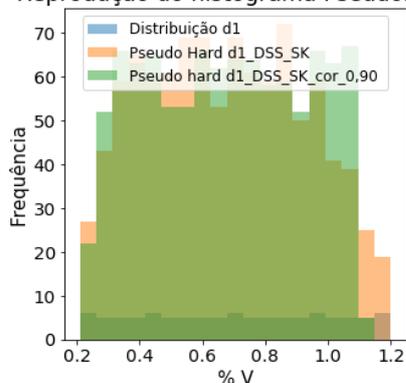
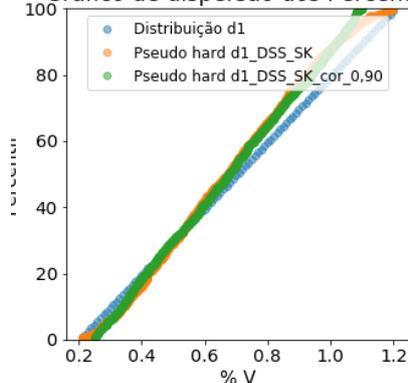


Gráfico de dispersão dos Percentis



b)

Reprodução do histograma Pseudohard

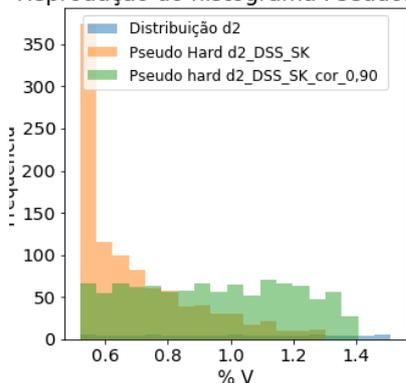
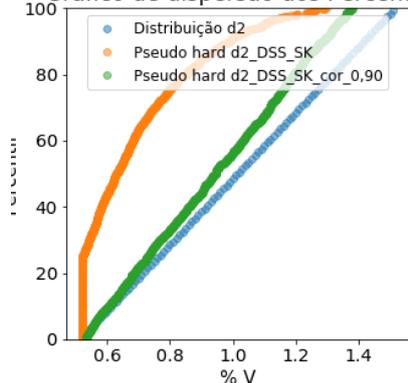


Gráfico de dispersão dos Percentis



c)

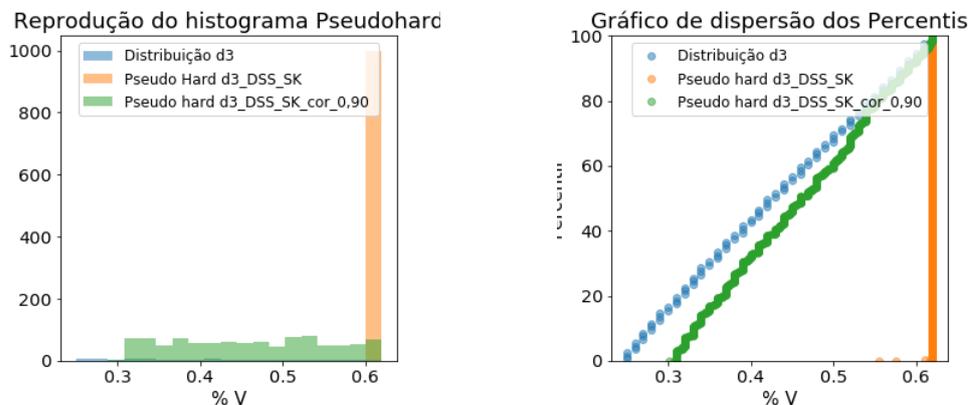


Figura 2. 8- Histograma, gráfico de dispersão para as distribuições a) d_1 , b) d_2 e c) d_3

A tabela 2.4 mostra os resultados para o teste Komolgorov-Sminorv, para as três distribuições investigadas, a hipótese nula foi verdadeira somente para a distribuição d_1 , em que a média da distribuição é próxima média dos dados.

Tabela 2. 4- Resultados do teste Komolgorov-Smirnov considerando que a distribuição de probabilidade possui forte correlação (0,90) com a variável primária

	D	p-value	Hipótese nula (H_0)
d_1	0,11	0,16	Verdadeira
d_2	0,14	0,04	Falsa
d_3	0,17	0,00	Falsa

Em resumo, a figura 2.9 mostra o fluxograma da metodologia proposta. Inicialmente, será necessário avaliar a correlação entre a variável primária e a secundária e construir as distribuições de probabilidade da variável secundária. Em seguida, os parâmetros de entrada: estratégia de busca, número de amostras e nós previamente simulados e o modelo de continuidade espacial da variável primária devem ser definidos. Em seguida, o algoritmo é executado e os modelos finais validados a partir da reprodução da continuidade espacial e distribuição de probabilidade de dados da variável primária. Detalhes sobre a

construção da distribuição serão discutidas no item 2.3.3 - Construção da distribuição de probabilidade deste capítulo e no capítulo 03 no item 3.2 - Metodologias propostas para construção da distribuição de probabilidade (*pdf*) no local da variável secundária desta tese.

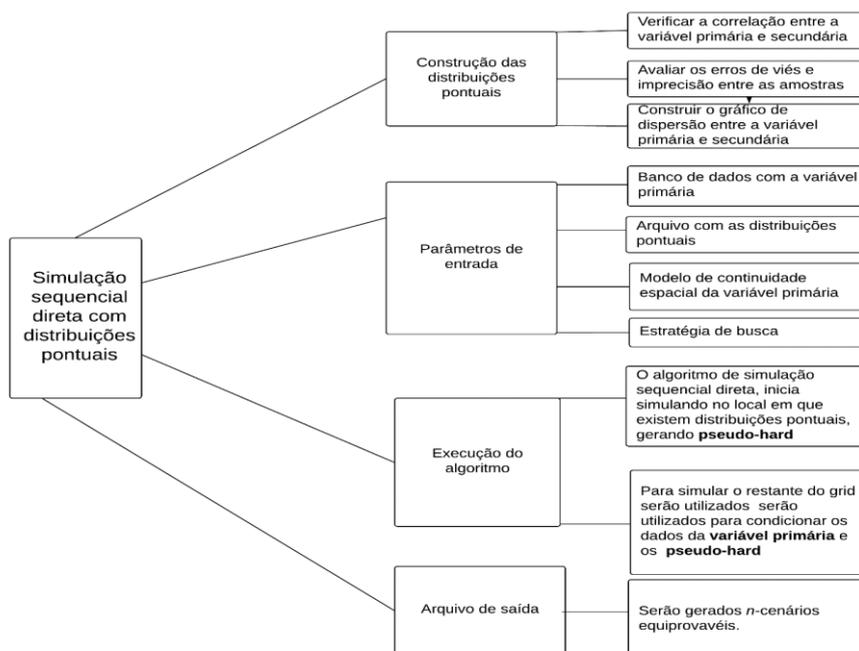


Figura 2. 9 - Fluxograma da metodologia simulação sequencial direta com distribuições pontuais

2.3 Estudo de caso

2.3.1 Descrição geológica do depósito estudado

O estudo de caso será realizado com dados de uma mina subterrânea onde se extrai minério de cobre e zinco. É um depósito tipo sulfetos maciços vulcanogênicos e ocorrem tipicamente como lentes polimetálicos (Cu, Zn, Sn, Pb), sulfetos maciços e *stockworks* que se formaram no fundo do mar ou perto dele ambientes vulcânicos submarinos. Os depósitos estão localizados perto do topo de um vulcão predominantemente da idade tardia do Devoniano-Carbonífero, 360-342Ma. Sobrepondo a mineralização, existe uma repetição de unidades vulcânico-sedimentares e *Fysch*, com aproximadamente 350m de espessura (Owen e Meyer, 2013).

2.3.2 Apresentação do banco de dados

A área de estudo as amostras foram coletadas por duas técnicas de amostragem diferentes: chip (fragmentos de rocha a serem lavradas) e furos de sondagem diamantada (DDH) no mesmo domínio geológico. As amostras possuem diferentes malhas amostrais. As amostras de furo de sondagem (Cu_DDH), considerada variável primária, possui 470 amostras com espaçamento médio 10x10 m (nas direções leste e norte e em média 1 metro longo do furo). As amostras de chip (Cu_Chip), denominada variável secundária, possui 1473 amostras com espaçamento de 1x1 m (nas direções leste e norte).

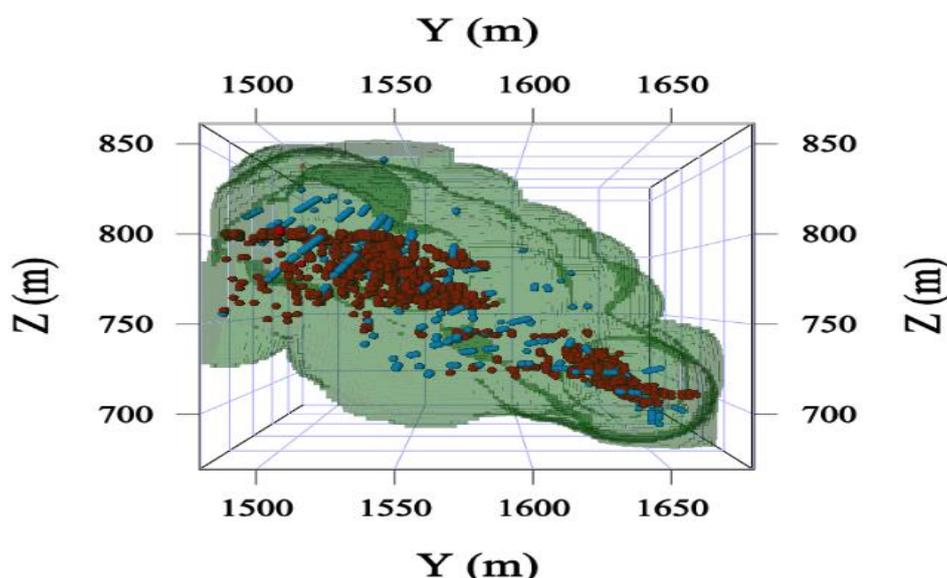


Figura 2. 10- Mapa de localização dos dados utilizados no estudo de caso, no plano zy. (pontos azuis: variável primária, pontos vermelhos: variável secundária e área verde corresponde ao domínio geológico dos dados)

As amostras primárias e secundárias possuem configuração espacial completamente heterotópica, e as amostras secundárias estão mais densamente amostradas que as primárias. Para avaliar a estatística descritiva das amostras em uma mesma região, foi feito um gráfico de dispersão entre as variáveis e foram retirados os pares das amostras separados até uma distância de 3 metros nas direções x, y e z.

A Tabela 2.5 mostra as principais parâmetros estatísticos dos conjunto de dados, a serem utilizados neste trabalho. Foi verificado, que a variável primária (Cu_DDH) possui média aritmética 15% e desvio padrão 5 % maiores que a variável secundária (Cu_Chip). Acredita-se que esta diferença é devido

aos erros incorporados no processo de amostragem (erro de delimitação, principalmente, Gy, 1992). Assim, assumimos que a variável primária (Cu_DDH) é precisa e acurada e a variável secundária (Cu_Chip) imprecisa e inacurada.

Tabela 2. 5- Estatística descritiva da variável cobre amostrada por furos de sondagem (DDH) e fragmento de rocha (*chip sample*).

Banco de dados	Nº de amostras	Média	Desvio Padrão	Coefficiente de variação	Mínimo	Máximo
Cu_DDH	470	5,71	5,86	1,02	0,70	22,92
Cu_Chip sample	1473	4,82	5,62	1,16	0,70	32,91

Considerando que o conjunto de dados são completamente heterotópicos, (a variável primária nunca está no mesmo local que a variável secundária), a correlação entre as amostras foi calculada a partir da extrapolação do correlograma cruzado para $h \rightarrow 0$ (Figura 2.11), proposto por Minnitt e Deustch (2014). O resultado mostra que, a partir da projeção do correlograma cruzado (linha azul), as variáveis possuem uma boa correlação entre elas de aproximadamente 0,90.

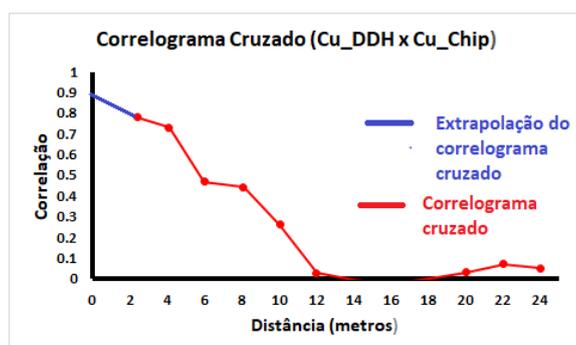


Figura 2. 11 - Correlograma cruzado entre os dados primários e secundários (linha vermelha) e a extrapolação do correlograma cruzado (linha azul) para $h = 0$.

Para definir o modelo de continuidade espacial, foi utilizado o variograma estandardizado (Equação 2.9). O efeito pepita mede a descontinuidade a curta distância e equivale a 0,25 (25%) da variância total, combinado com duas estruturas (0,44 e 0,31) esféricas (*Sph*). As direções de continuidade espacial principal, intermediária e vertical são identificadas como h_1 , h_2 e h_3 , e os

alcances em metros, respectivamente. Este modelo foi utilizado para gerar todos os modelos obtidos por simulação estocástica.

$$\gamma_{Cu}(h) = 0.25 + 0.44 \cdot \text{Sph}(1) \cdot \left(\frac{h1}{22 \text{ m}}, \frac{h2}{16 \text{ m}}, \frac{h3}{15 \text{ m}} \right) + 0.31 \cdot \text{Sph}(2) \cdot \left(\frac{h1}{52 \text{ m}}, \frac{h2}{31 \text{ m}}, \frac{h3}{25 \text{ m}} \right) \quad (2.9)$$

2.3.3 Construção da distribuição de probabilidade

Para ser consistente com o algoritmo, o arquivo de distribuição pontual deve possuir a localização das distribuições, ou seja, as coordenadas x, y, z e os n - possíveis valores da distribuição em cada local. Nessa abordagem, assumimos que as variáveis secundárias junto com a primária seguem uma distribuição bivariada normal. Para construir a distribuição de probabilidade, as variáveis primárias e secundárias foram transformados para o espaço normal gaussiano, a partir do executável NSCOREMV.exe (software GSLIB). Para que fosse evitado problemas de extrapolação e dados com valores negativos, as distribuições foram construídas no espaço gaussiano e foram retro-transformada para o espaço original, utilizando como referência a tabela de transformação quantil-quantil (*trn*) da variável primária. Assim, a média e variância das distribuições foram definidas conforme equações abaixo. Considerando que a esperança matemática de X_1 dado X_2 é:

$$E(X_1 | X_2 = x_2) = \rho x_2 \quad (2.10)$$

e a variância condicional é:

$$\text{var}(X_1 | X_2 = x_2) = 1 - \rho^2 \quad (2.11)$$

onde X_1 é a variável primária, X_2 é a variável secundária e ρ é o coeficiente de correlação entre as variáveis X_1 e X_2 .

2.4 Validação

2.4.1 Validação dos *Pseudohard*

Para avaliar a metodologia proposta, a continuidade espacial e o histograma acumulado dos *pseudohard* gerados devem ser semelhantes aos equivalentes da variável primária. A figura 2.12 mostra o histograma acumulado da variável primária (linha vermelha) e os 50 cenários simulados de valores *pseudohard* (linhas pretas) gerados por cokrigagem intrínseca (ICK). Podemos verificar que a metodologia 02_DSS_PD_SK (figura 2.12a) ocorreu uma pequena superestimativa dos valores do *pseudohard* gerados. Para a metodologia 03_DSS_PD_SCK (figura 2.12b), a distribuição dos *pseudohard* gerados são próximos a distribuição da variável primária, mas com uma pequena subestimativa. Este fato, ocorre porque foram utilizadas mais amostras para na metodologia 03, o que reduz a incerteza dos modelos.

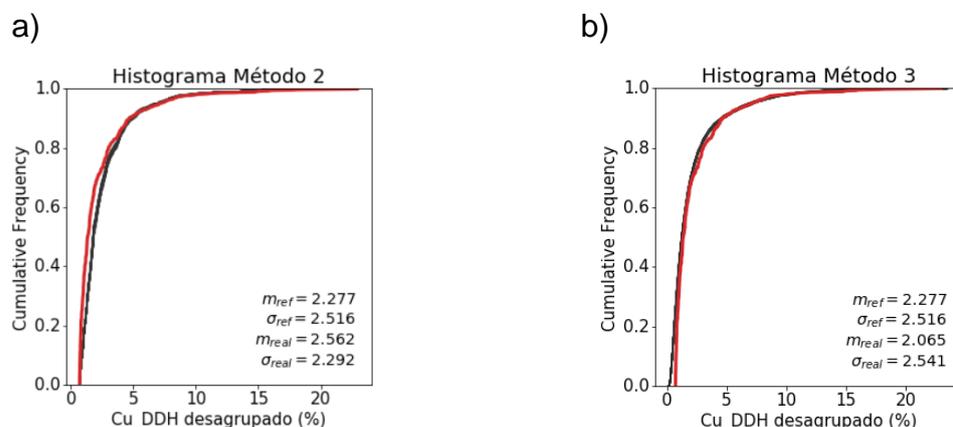


Figura 2. 12- Histograma dos dados *pseudohard* (linhas pretas) gerados comparados com o histograma da variável primária (Cu_DDH) (linha vermelha) a) Método e 02_DSS_PD_SK b) Método 03_DSS_PD_SCK.

A figura 2.13 mostra a reprodução da continuidade espacial dos *pseudohard* gerados pela metodologia 02_DSS_PD_SK (figura 2.13a) e 03_DSS_PD_SCK (figura 2.13b) que foram investigadas. A linha vermelha

corresponde ao modelo do variograma inferido, a linha azul ao variograma experimental dos dados e as linhas pretas ao variograma experimental dos *pseudohard* gerados. Podemos verificar que os *pseudohard* reproduzem o modelo de continuidade espacial da variável primária, nas direções principal, intermediária e vertical.

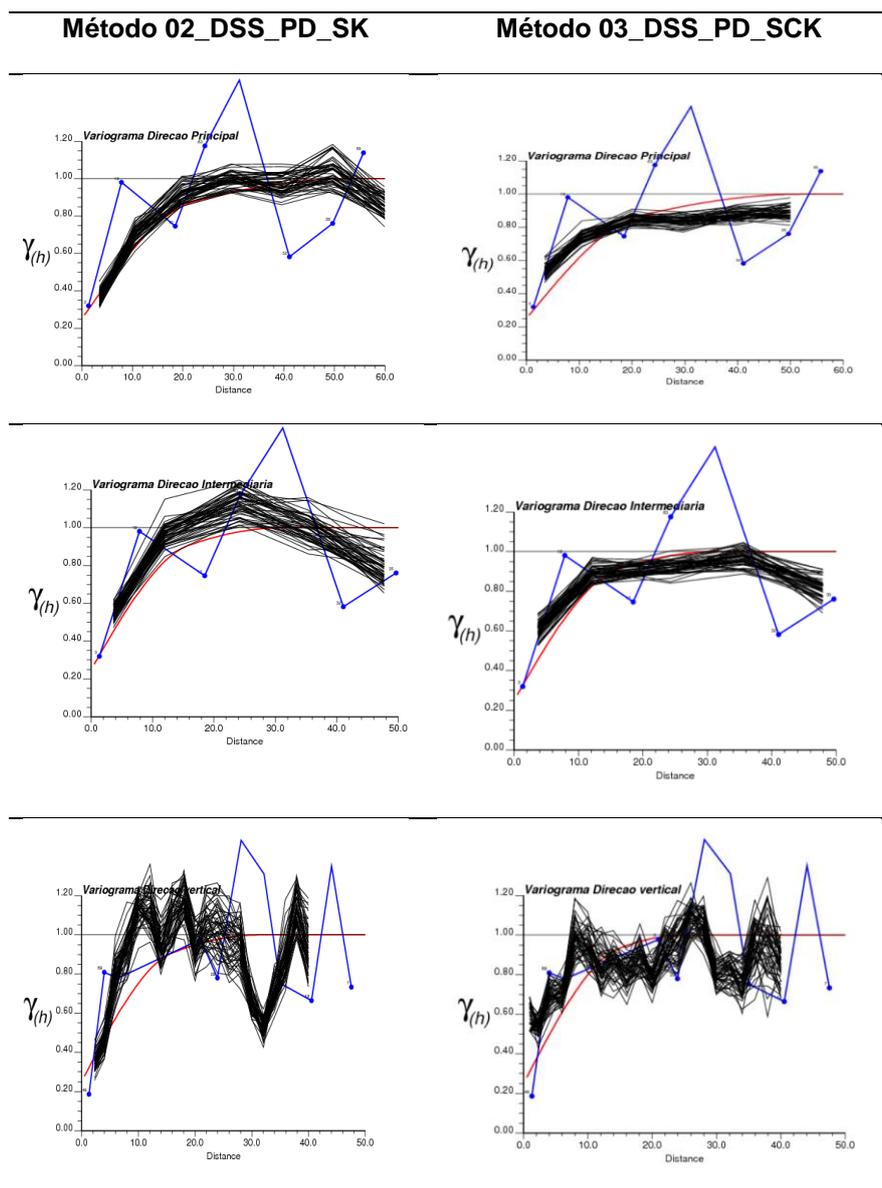


Figura 2. 13- Comparativo entre a reprodução dos variogramas dos modelos finais (linhas preta), o modelo do variograma (linha vermelha) e o variograma experimental dos dados (linha azul) nas direções de continuidade espacial a) direção principal (h_1), b) direção intermediária (h_2), e c) direção vertical (h_3)

2.4.2. Avaliando a acuracidade e precisão dos modelos

A fim de verificar se a metodologia proposta melhorou a exatidão e precisão dos modelos, os resultados obtidos pela simulação foram comparados com o valor considerados verdadeiros da variável primária cada local com base na conhecida ferramenta estatística chamada *Jackknife*.(Efron,1982) Foram excluídos da variável primária, 8 furos completos de perfuratriz próximos da variável secundária (que representam 55 dados da variável primária, ou seja, 12% dos dados da variável primária) e 50 simulações foram realizadas nestes locais (Figura 2.14). Foram comparadas três metodologias: Metodologia 01_DSS_Hard_only: Simulação sequencial direta utilizando somente a variável primária, Metodologia 02_DSS_PD_SK: Simulação sequencial direta com distribuição pontual utilizando krigagem simples e na Metodologia 03_DSS_PD_SCK: Simulação sequencial direta com distribuições pontuais utilizando cokrigagem simples com modelo intrínseco para inferir os *pseudohards*.

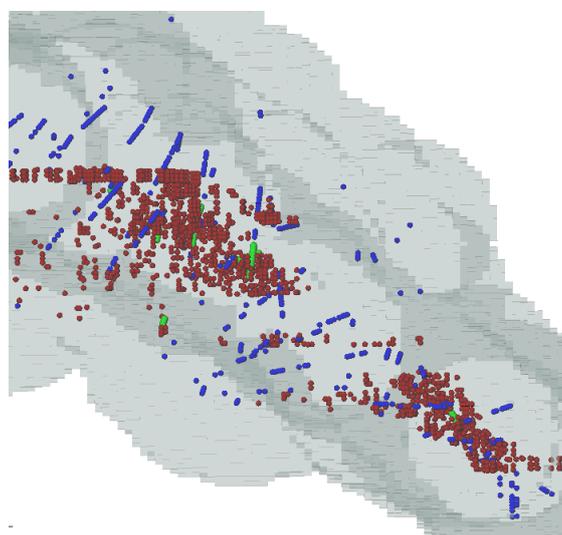


Figura 2. 14- Mapa de localização dos dados utilizados no estudo de caso, no plano zy. (pontos azuis: variável primária, pontos vermelhos: variável secundária e pontos verdes: dados da variável primária excluídos do banco de dados e área cinza corresponde ao domínio geológico dos dados)

A figura 2.15 mostra o gráfico de dispersão entre os valores verdadeiros da variável primária (Cu_DDH) comparados aos valores médios obtidos nos 50 cenários simulados (Cu_E-type), na mesma localização dos dados omitidos. Verificamos, que quando foram utilizados somente os dados da variável primária, na metodologia 01_DSS_Hard_only (Figura 2.15a), houve uma fraca correlação (0,25) entre os valores reais e estimados e as estimativas destes locais foram subestimadas. Para as metodologias 02_DSS_PD_SK e 03_DSS_PD_SCK, foram utilizadas mais amostradas para construir os modelos, portanto, a correlação entre os valores verdadeiros e os estimados aumentou. Podemos verificar que no método 03_DSS_PD_SCK, os valores estimados possuem maior correlação com os valores considerados verdadeiros. Nesta abordagem, tanto os dados primários e os secundários receberam pesos para condicionar a distribuição em que o pseudohard foi amostrado, o que auxiliou na construção do modelo.

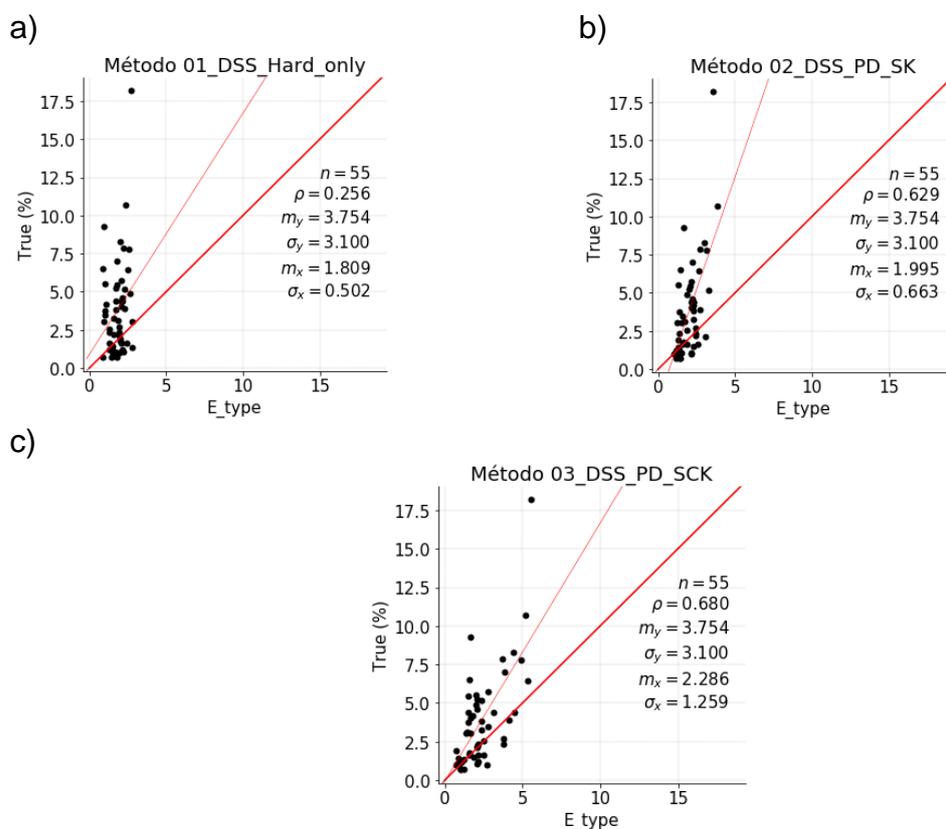


Figura 2. 15- Gráfico de dispersão dos valores reais da variável primária (Cu_DDH) comparado ao valor médio obtido pelo E-type pelo os métodos a) Método 01_DSS_Hard_only, b) Método 02_DSS_PD_SK e c) Método 03_DSS_PD_SCK

A magnitude do erro (Equação 2.12) que corresponde ao módulo da diferença entre o valor real da variável primária (Cu_{DDH}) e a média dos valores simulados para o mesmo local definido pelo tipo E -type (Cu_{E-type}) foi calculada para cada metodologia.

$$\text{Erro}_{\text{abs}} = |Cu_{DDH} - Cu_{E-type}| \quad (2.12)$$

A figura 2.16 mostra o gráfico de dispersão entre erros absolutos da metodologia 03_DSS_PD_SCK com as metodologias 01_DSS_Hard_only e 02_DSS_PD_SK investigadas. A figura 2.16a mostra que a média dos erros absolutos para a metodologia 01_DSS_Hard_only é 48% e a metodologia 02_DSS_PD_SK (figura 16b) é 32% maior que a metodologia 03_DSS_PD_SCK. Assim, podemos concluir que a metodologia 03_DSS_PD_SCK é mais acurada que as demais metodologias.

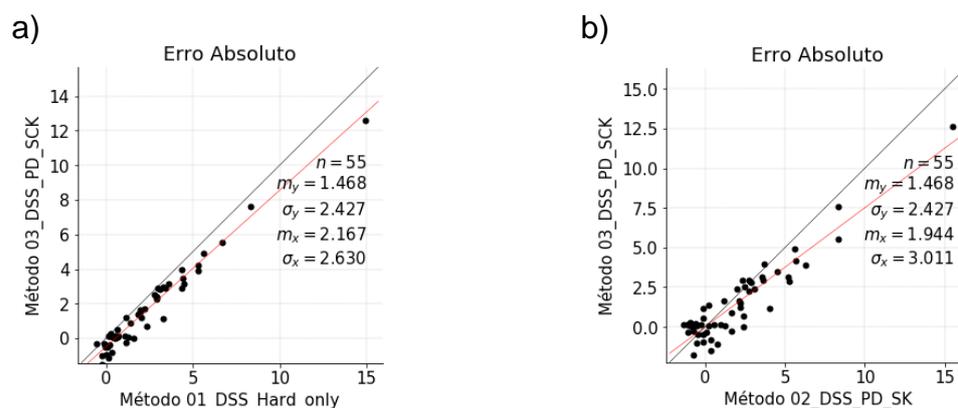


Figura 2. 16- Gráfico de dispersão do erro absoluto do Método 03_DSS_PD_SCK comparado com as outras metodologias a) Método 01_DSS_Hard_only e b) Método 02_DSS_PD_SK.

2.5 Considerações

Em geoestatística, para que os dados primários e secundários possam ser integrados nos algoritmos de krigagem e simulação, é necessário que estejam no mesmo suporte/volume. Para isto, é necessária operar mudanças na estatística dos dados utilizando métodos de mudança de suporte. Dentre os algoritmos de simulação estocástica disponíveis, a simulação sequencial direta (Soares, 2001), que utiliza os dados no seu espaço original, permite incorporar dados de diferente suporte e preserva a relação linear entre as amostras de

diferente suporte nos modelos estocásticos. Portanto, este algoritmo possui grande potencial aplicação na área de geoestatística do petróleo, mineração e sensoriamento remoto que podem utilizar amostras em diferentes suportes volumétricos.

A nova abordagem deste algoritmo, denominada simulação sequencial direta com distribuições pontuais (Soares *et.al.*, 2017), propõe integrar informação secundária com incerteza de medida associada na forma de distribuições de probabilidade. Naquela implementação, foi utilizado o algoritmo de krigagem simples, que atribui pesos a média dos dados e para variável primária no entorno (vizinhança local) para inferir os *pseudohards* (valor sorteado da distribuição de probabilidade construída no local do dado secundário incerto). Neste capítulo, foi implementado o algoritmo de cokrigagem simples, que considera a correlação espacial entre as variáveis primárias e secundárias, atribuindo a ambas pesos no processo de estimativa.

Para ilustrar esta implementação, foi utilizado um banco de dados de mina subterrânea de cobre e zinco. A acuracidade e precisão desta metodologia foi checada com a validação do tipo *Jackknife*. Nas duas abordagens investigadas, os resultados mostram que, independentemente da qualidade, os modelos resultantes não foram afetados pelo viés e imprecisão dos dados secundários adicionados e o fato de se utilizar mais informação levou a modelos mais acurados.

Os resultados mostraram que os erros absolutos foram menores e a correlação entre os dados verdadeiros (Cu_DDh) e estimados (Cu_E-type) foi maior. Isto indica, que os modelos resultantes foram mais acurados quando comparados aos modelos obtidos somente utilizando a variável primária (Método 01_DSS_Hard_only). Verificou-se que a incerteza dos dados secundários foi incorporada de forma adequada. A metodologia 03_DSS_PD_SCK, que utiliza o algoritmo de cokrigagem simples e atribui pesos para a amostra primária e para a variável secundária amostrada, produziu melhores resultados.

Por definição, em situações que a variável primária e secundária possui baixa correlação, o algoritmo de cokrigagem e krigagem simples possuem

resultados semelhantes. Portanto, os resultados obtidos pela metodologia 02_DSS_PD_SK, que utiliza o algoritmo de krigagem simples para inferir o *pseudohards*, podem emular resultados em que os dados possuem baixa correlação. É importante ressaltar, que valores baixos de correlação podem estar relacionados a protocolos de amostragem pobres, análises químicas inaccuradas, ou seja a erros incorporados na etapa de amostragem. Nestas situações, esta metodologia pode ser considerada para incorporar a variável secundária ao adicionar corretamente a esses dados seu viés e imprecisão.

É importante ressaltar, que o algoritmo de simulação sequencial direta com distribuições pontuais possui outros aprimoramentos e desenvolvimentos para a reprodução da distribuição (AZEVEDO, 2018) ¹.

¹ Comunicação pessoal em reunião via Skype®, com o professor Leonardo Azevedo em outubro de 2018.

3 INCORPORAÇÃO DA VARIÁVEL SECUNDÁRIA COMO DISTRIBUIÇÃO PROBABILIDADE LOCAL UTILIZANDO SIMULAÇÃO ESTOCÁSTICA

Para integrar as múltiplas variáveis com a correlação espacial, as metodologias clássicas em geoestatística sugerem modelagem dos variogramas diretos e cruzados usando o modelo linear de correionalização (MLC) para atribuir pesos aos valores correlacionados na cokrigagem. O desafio desta abordagem, é modelar os variogramas quando mais de duas variáveis são consideradas. Algumas simplificações do MLC foram propostas na literatura. Almeida e Journel (1994) sugeriram uma aproximação de Markov-Bayes de correionalização para modelar as covariâncias cruzadas. Na aproximação de Markov-Bayes, o modelo de variograma cruzado é proporcional ao variograma da variável primária (MM1) ou da variável secundária (MM2). Os modelos de Markov são geralmente usados para os algoritmos de cokrigagem colocada, que usa apenas o dado secundário colocado para estimativa. A limitação desta metodologia é que a variância é aumentada (inflada) quando utilizada nos algoritmos de co-simulação sequencial (Deutsch e Journel, 1998). Para evitar a aumento da variância, Babak e Deutsch (2009) propuseram usar os dados secundários colocados e também os na sua vizinhança junto com os dados da variável primária para estimar. Essa abordagem é denotada como cokrigagem colocada intrínseca (ICCK). A limitação de cokrigagem colocada intrínseca é que os dados secundários devem ser amostrados em todos os nós do grid (exaustivo).

Técnicas de inserção múltiplas, tem sido investigada para situações em que os dados secundários são mais amostrados do que os primários. Nesta abordagem, o algoritmo de simulação estocástica inicia-se nos locais da variável secundária em seguida, no restante dos nós do grid de simulação. Isto implica na melhor eficiência computacional dos algoritmos, já que não simula a variável secundária em todos os nós do grid. Alguns autores têm investigado estas metodologias como Barnett e Deutsch (2015); Soares et al. (2017) e Silva e Deutsch (2018).

Neste estudo, será abordado situações em que os dados secundários são mais amostrados do que os dados primários, mas não abrangem toda o grid. Almeida e Journel (1994) sugeriram simular as variáveis de forma hierárquica num contexto de co-simulação. Neste caso, a variável secundária é simulada primeiro e usada como informação secundária com o algoritmo de cokrigagem colocada intrínseca (ICCK) na simulação da variável primária. No entanto, essa abordagem exige mais memória do computador e tempo de processamento do que simular diretamente a variável primária.

A metodologia proposta é uma alternativa para incorporar informação secundária como distribuição probabilidade local que possa ser executada por qualquer algoritmo de simulação estocástica. A ideia de incorporar dados secundários como uma distribuição local de probabilidades para caracterizar falta de precisão e exatidão das medidas foi baseada em Soares *et al.* (2017) e Neves *et al.* (2018). Em seus trabalhos, os autores sugerem para integrar os dados secundários, criar uma função de distribuição conjunta entre a variável primária e secundária com os erros de medida incorporados. Essa abordagem foi implementada no algoritmo de simulação sequencial direta (DSS) com distribuições pontuais (Soares *et al.* 2017). Aqui, será mantida a vantagens deste algoritmo que são: utilizar apenas o modelo de continuidade espacial da variável primária e incorporar a incerteza dos dados secundários como uma distribuição probabilidade local.

Em resumo, a figura 3.1 mostra o fluxograma da metodologia proposta para incorporar a variável secundária como distribuição de probabilidade local. Inicialmente, será necessário construir a distribuição de probabilidade local. Para isto, foram investigadas três diferentes abordagens: i) distribuição da variável primária considerada como bivariada normal, ii) distribuição da variável primária construída a partir da estimativa por cokrigagem simples no local da variável secundária e iii) distribuição construída a partir da atualização bayesiana. Em seguida, estas distribuições são amostradas garantindo a correlação espacial entre a variável primária e secundária seja mantida, gerando n-possíveis cenários de pseudohards (cenários A). Sucessivamente, são gerados novos banco de dados, que consistem em: os dados da variável primária e os pseudohards (cenários B), que serão utilizados para acessar a incerteza dos

modelos no algoritmo de simulação estocástica. É importante ressaltar que cada cenário é utilizado um conjunto de pseudohards diferentes. Para o algoritmo de simulação estocástica são utilizados os mesmos parâmetros de entrada utilizado para a variável primária.

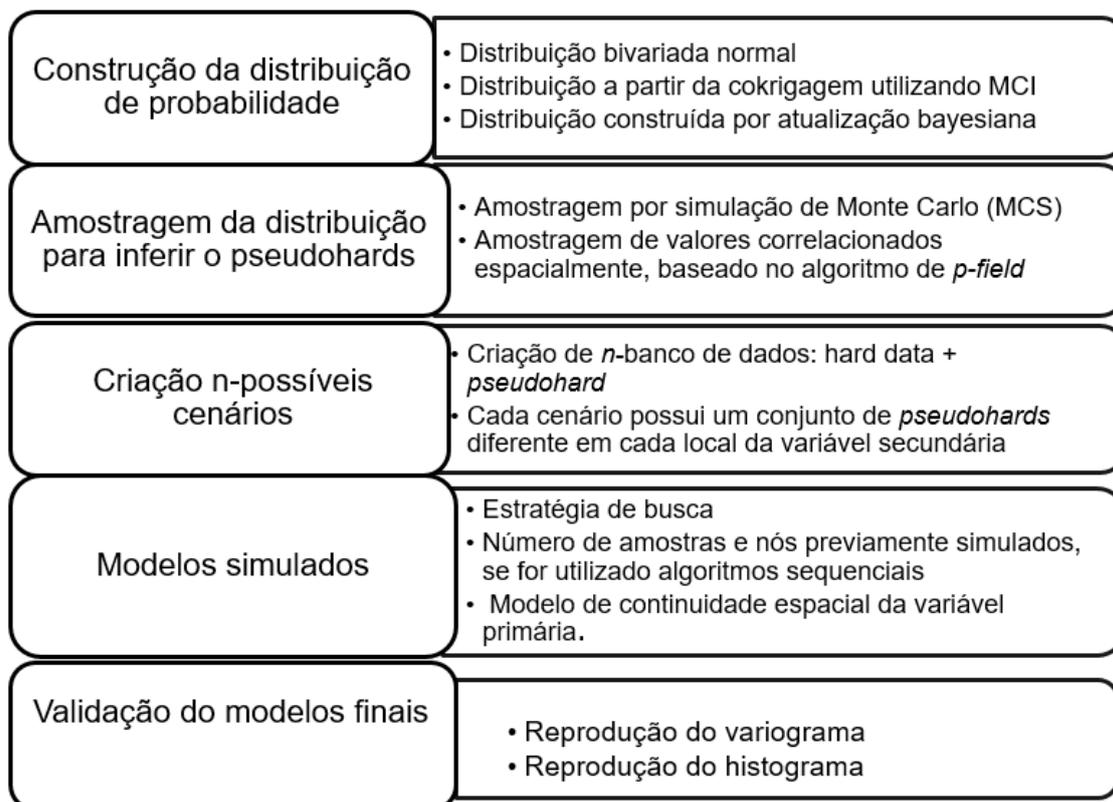


Figura 3. 1- Fluxograma das metodologias a serem propostas para inserir a variável secundária com distribuição de probabilidade local

3.1 Fundamentação teórica sobre as metodologias abordadas

3.1.1 Simulação por campos de probabilidade (*p-field*)

O algoritmo de simulação de campos de probabilidade (*p-field*) foi proposto por Srivastava (1992) e Froidevaux (1993). Esse método considera que as distribuições locais (*ccdf's*) são conhecidas no local a ser simulado. Em seguida, as simulações condicionais são obtidas a partir da amostragem aleatória de valores de probabilidade correlacionados espacialmente (*p-values*) destas distribuições locais, de maneira que o modelo de covariância dos dados seja reproduzido por esse campo de probabilidade.

O algoritmo está descrito nos passos abaixo:

- i. Em cada local a ser simulado, construir um modelo de distribuição local condicional de probabilidades *ccdf*'s, utilizando um algoritmo adequado, como por exemplo, krigagem multiGaussiana ou krigagem dos indicadores (Alabert,1987) para a variável (principal) a ser simulada.
- ii. Gerar realizações não condicionais de campos de probabilidade (Probability Field) $p^l(x_0)$ com dados espacialmente correlacionadas com modelo de covariância equivalente ao da variável que se deseja simular.
- iii. Em cada local a ser simulado, sorteia-se um valor simulado da ccdf, utilizando o valor *p-value* simulado na mesma posição.

Múltiplos cenários são gerados repetindo os passos (ii) e (iii).

Journel e Ying (2001) compararam os algoritmos de simulação sequencial Gaussiana (SGS), simulação Gaussiana truncada (GTSIM) (Journel e Isaaks, 1984; Matheron *et al.*, 1987; Xu e Journel, 1993) e os de simulação por campos de probabilidade (p-field). Os autores afirmaram que a simulação por campo de probabilidade (p-field) pode ser considerada transformação quantil-quantil de uma distribuição uniforme, já o algoritmo de simulação Gaussiana truncada (GTSIM), é uma transformação quantil-quantil de uma distribuição gaussiana. Sobre o algoritmo simulação sequencial gaussiana (SGS), os autores consideram que a retro-transformação dos dados normalizados pode ser considerada como uma série de truncamentos gaussianos. O algoritmo GTSIM é usado somente para variáveis categóricas (Xu e Journel, 1993; Johann *et al.*, 1997; Bahar e Kelkar, 1997; Fichtl *et al.*, 1997). Já o algoritmo p-field, pode ser utilizado para variáveis categóricas e contínuas (Srivastava, 1992; Cassiraga, 1999). A principal vantagem do algoritmo de simulação por campo de probabilidades é velocidade para se gerar múltiplas realizações e a facilidade em integrar variáveis secundárias como distribuição probabilidade local *a priori* da simulação.

Pyrzcz e Deustch (2001) alertaram sobre duas limitações sobre o método quando há dados condicionais: i) quando os dados de condicionamento local aparecem como mínimos locais ou máximos dos valores simulados, ii) os valores

simulados geralmente possuem continuidade espacial maior que a continuidade espacial da variável primária. Ortiz (2003) propôs uma solução global para corrigir a reprodução do variograma. Considerando que o modelo de covariância utiliza todas as amostras disponíveis no domínio, o autor sugere que a covariância do modelo simulado seja corrigida para reproduzir o variograma esperado. Esta correção foi implementada no algoritmo GAMVPFIELD.exe (software GSLIB).

3.2 Metodologias propostas para construção da distribuição de probabilidade (*pdf*) no local da variável secundária

Em projetos de mineração e extração de petróleo, as relações entre variáveis geológicas são muitas vezes complexas, o que representa um desafio para reproduzir essas relações usando as ferramentas convencionais de modelagem geoestatística.

Neste trabalho, serão utilizadas apenas duas variáveis (primária e secundária) que se encontram em um mesmo domínio geológico, possuem correlação entre si e se comportam como uma distribuição de probabilidade (*pdf*), ou seja, a probabilidade da variável aleatória é definida por um intervalo de possíveis valores. A ideia de probabilidade condicional também é utilizada nos métodos bayesianos e simulação sequencial gaussiana. Nas abordagens a seguir, serão discutidas duas possibilidades para a construção da distribuição de probabilidade da variável secundária. Estas serão construídas a partir dos dados normalizados das variáveis, para evitar extrapolações e presença de valores negativos nas distribuições. Para isto, foi realizada a transformação *nscore* das variáveis primárias e secundária. Assim, assumiu-se que os dados são Gaussianos multivariados.

Na primeira abordagem, denominada distribuição 01, a distribuição de probabilidade da variável secundária será considerada uma distribuição normal bivariada. Nesta proposta, será considerado apenas a variável secundária no local para construção da distribuição de probabilidade e a correlação entre as variáveis primária e secundária. Já a segunda proposta, distribuição 02, será

considerado a variável secundária no local e as variáveis primárias e secundárias próximas ao local que será construída a distribuição. Para isto, será utilizado o algoritmo de cokrigagem com modelo intrínseco para construção da distribuição.

3.2.1 Distribuição 01: distribuição da variável primária obtida a partir de uma distribuição normal bivariada

Esta proposta é a mesma utilizada no item 2.3.3 - Construção da distribuição de probabilidade. A diferença é que as distribuições foram construídas no espaço gaussiano.

3.2.2 Distribuição 02: distribuição da variável primária obtida por cokrigagem simples utilizando modelo linear de correionalização intrínseco (MCI)

A ideia é utilizar todas as informações primárias e secundárias disponíveis. Para isto, será utilizado o algoritmo de cokrigagem simples com o modelo de coreionalização intrínseca (MCI), que é considerado um caso particular do modelo de linear de correionalização (MLC). Assim, para definir o modelo de continuidade espacial, todos os variograma diretos e cruzados são proporcionais a um mesmo variograma. Neste caso, os variograma cruzados ou diretos são dados por:

$$\gamma_{ij}(h) = \omega_{ij}[\gamma_i(h)] \quad (3.3)$$

onde $\gamma_{ij}(h)$ é o variograma cruzado ($i \neq j$) entre a variável primária e secundária, $\gamma_i(h)$ é o variograma com uma mais estruturas da variável primária, no qual todos os variogramas diretos e cruzados são proporcionais e ω_{ij} é o coeficiente de proporcionalidade para o variograma $\gamma_{ij}(h)$, que será considerado o coeficiente de correlação entre as amostras.

3.3 Metodologias propostas para amostrar a distribuição probabilidade da variável secundária e incorporação da incerteza nos modelos

A ideia chave é incorporar a variável secundária, neste contexto considerado como uma distribuição de probabilidade, nos modelos obtidos por

simulação estocástica. A abordagem proposta é flexível e pode ser realizada em qualquer algoritmo de simulação estocástica, como por exemplo: simulações sequenciais ou métodos de bandas de rotativas. A metodologia é descrita nas etapas abaixo:

Etapa 01: Normalizar os dados da variável primária e secundária e criar uma distribuição probabilidade em cada local da variável secundária, conforme descrito nos itens acima 3.2.1 e 3.2.2. Neste estudo de caso, os dois conjuntos amostrados são completamente heterotópicos, o coeficiente de correlação entre as variáveis, para $h=0$ foi obtido a partir da extrapolação do correlograma cruzado experimental dos dados (Minnitt e Deustch, 2014).

A figura 3.2 exemplifica o mapa de localização das distribuições locais, onde o eixo vertical corresponde aos teores das amostras e eixo horizontal a coordenada espacial x . Considerando as amostras da variável primária (distribuição azul), que estão localizadas nas coordenadas 1, 3, 5 e 7, precisas e acuradas, ou seja, variância nula, as distribuições serão estreitas e finas. Já, a variável secundária (distribuição vermelha), localizadas nas coordenadas 2, 4 e 6, os erros foram incorporados no processo de amostragem. Assim, nestes locais foram construídas distribuições locais para incorporar estes erros.

Etapa 01 :

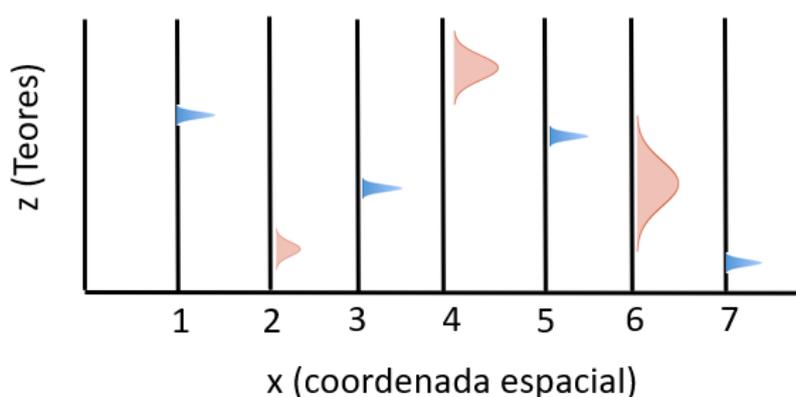


Figura 3. 2- Exemplo da localização espacial completamente heterotópica da variável primária (distribuições azuis) e da distribuição da variável secundária (distribuições vermelhas).

Etapa 02: Gerar valores das curvas de probabilidade da variável secundária (cenário A) com o mesmo modelo de continuidade espacial obtido da variável primária e correlacionados com a variável primária. Para obter esses valores, foram gerados cenários simulados não condicionados pela variável primária e secundária, mas com o mesmo modelo de continuidade espacial da variável primária nos locais da secundária. Para amostrar os valores de probabilidade correlacionados e aleatórios das distribuições locais (*cdf's*), foi utilizado o algoritmo de simulação *p-field* (Srivastava, 1992 e Froidevaux, 1993). A figura 3.3 ilustra a amostragem aleatória da distribuição da variável secundária. Na figura 3.3a, é sorteado um valor aleatório e correlacionado, identificado na figura como **a** na distribuição local de probabilidade da variável secundária com o algoritmo de simulação *P-field*. Em seguida, a área sob a curva do valor sorteado **a** será menor ou igual ao valor na *cdf* corresponde desta distribuição (Figura 3.3b). Assim, os valores de probabilidade aleatórios amostrados, são gerados entre 0 e 1 e são associados a um valor **a** chamado de *pseudohard*.

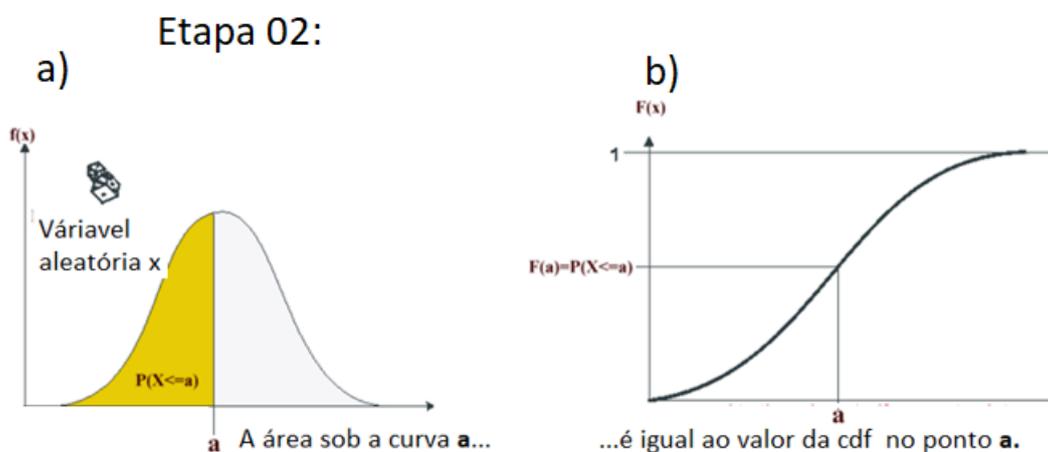


Figura 3. 3- Exemplo da amostragem realizada na distribuição local da variável secundária para gerar o *pseudohard*.

Etapa 3: Gerar n possíveis cenários do conjunto de dados *pseudohard* (cenário A) nos locais da variável secundária. Cada um destes n cenários, foi construído com a amostragem de distribuições geradas na etapa 1 com os valores de probabilidade amostrados no mesmo local (etapa 2).

Etapa 04: Para cada um dos n cenários gerados (cenário A, contendo os dados amostrados da distribuição probabilidade), foram adicionados os valores

da variável primária (cenário B que contém dados da variável primária e *pseudohard* amostrados da distribuição de probabilidade) para gerar cenários simulados. Este conjunto de cenários (cenário A + B) foi utilizado para acessar a incerteza da variável primária em cada local não amostrado.

A figura 3.4 exemplifica os cenários a serem utilizados para acessar a incerteza. Foram gerados três cenários A. Para gerar o primeiro cenário (Etapa 03), foram amostrados valores aleatórios e correlacionados das distribuições da variável secundária (triângulos laranja). Este procedimento foi repetido para o segundo cenário (quadrados verdes) e para o terceiro cenário (losangos azuis). Em seguida, para cada cenário, foi adicionado a variável primária (cenário B), que corresponde aos pontos azuis (Etapa 04). Em seguida, foi executado o algoritmo de simulação estocástica usando o banco de dados completo (*hard e pseudohard*).

Etapa 04:

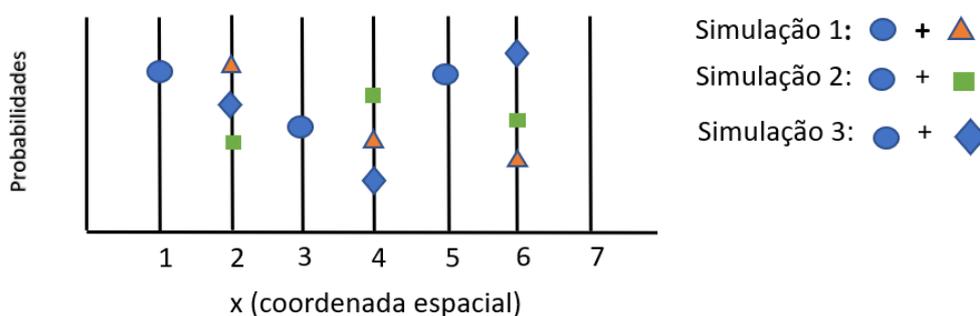


Figura 3. 4- Valores de *pseudohard* gerados e a variável primária que serão utilizados no algoritmo de simulação estocástica.

3.4 Inserção múltipla no local da variável secundária obtida por atualização bayesiana (*Bayesian Updated*)

As técnicas de inserção múltipla consistem em incorporar a incerteza dos dados faltantes por inferência n possíveis valores prováveis de uma distribuição (Rubin, 1987). Esta metodologia permite criar n conjuntos de dados, incorporando variáveis auxiliares correlacionadas no processo de interesse. Técnicas de inserção múltiplas e metodologias geoestatísticas tem sido combinada para incorporar a estrutura espacial dos dados inferidos e adicionar

variabilidade aos modelos. Essa abordagem permite que o pesquisador explore diferentes modelos que possam representar melhor os dados do estudo, que pode ser definida em três passos:

- i) inferir a distribuição dados faltantes que podem ser construídas por exemplo, modelos de regressão linear, atualização bayesiana,
- ii) realizar a amostragem estocástica para gerar múltiplos cenários
- iii) executar o algoritmo de simulação estocástica utilizando os dados disponível da variável de interesse e os dados inferidos.

Barnett e Deutsch (2015) propuseram duas metodologias de múltiplas imputações: paramétricas e não-paramétricas para imputar dados geológicos faltantes com base na atualização bayesiana para conjuntos de dados isotópicos gerados. No contexto de completamente heterotópicos, será utilizada a abordagem paramétrica. Esta abordagem requer o coeficiente de correlação entre as variáveis primária e secundária. Neste trabalho, combinamos a abordagem de Minnitt e Deutsch (2014) para obter o coeficiente de correlação com o método de imputação múltipla paramétrica de Barnett e Deutsch (2015). O método paramétrico apresentado por Barnett e Deutsch (2015) usa atualização bayesiana para inferir a incerteza dos dados secundários em seu local.

A atualização Bayesiana é considerada uma alternativa para a cokrigagem colocada (Doyen, 1996; Deutsch e Zanon, 2004; Ren, 2007) e não requer o modelo linear de correlogramização (MLC), apenas o coeficiente de correlação entre as variáveis primária e secundária. Assumimos, que obter o coeficiente de correlação é mais simples do que modelar o MLC. Como os dados são completamente heterotópicos, o coeficiente de correlação foi obtido através da extrapolação do correlograma experimental cruzado para $h = 0$. Nessa abordagem, a cokrigagem colocada é decomposta em dois modelos: (i) *a priori* que é construído os dados da variável primária e (ii) o *likelihood* é construído com os dados da variável secundária. A partir destas duas distribuições, uma nova distribuição é construída e atualizada para caracterizar a incerteza, denominada como *a posteriori* ou *updated*.

Para realizar a atualização bayesiana, assume-se a hipótese do modelo de multigaussianidade para a primária e secundária. Neste contexto, a distribuição *a priori* é Gaussiana e obtida por krigagem simples da variável primária utilizando os dados originais e previamente simulados/inseridos, sendo parametrizada na média e variância de krigagem simples. A estimativa de krigagem simples da variável primária ($\bar{y}_P(x_0)$) é dada por:

$$\bar{y}_P(x_0) = \sum_{i=1}^n \lambda_i y(x_i) \quad (3.4)$$

onde x_0 é o local a ser estimado e λ_i , ($i=1, \dots, n$) são os pesos recebidos pelas amostras primárias ($y(x_i)$) utilizadas nas estimativas. Os pesos de estimativa são obtidos através da resolução do sistema de krigagem simples. E a variância de krigagem simples ($\sigma_P^2(x_0)$) é definida como (Equação 3.5):

$$\sigma_P^2(x_0) = 1 - \sum_{i=1}^n \lambda_i C(x_0 - x_i) \quad (3.5)$$

onde $C(x_0 - x_i)$ é a covariância entre o local a ser estimado x_0 e as amostras utilizadas na estimativa x_i .

A distribuição condicional, denominada *likelihood*, será definida também por estimativa utilizando todas as variáveis secundárias disponíveis na área a ser modelada. Assim, a média ($\bar{y}_L(x_0)$) da distribuição *likelihood* será definida por (Equação 3.6):

$$\bar{y}_L(x_0) = \sum_{j=1}^m \lambda_j x_j(x_0) \quad (3.6)$$

onde λ_j , ($j = 1, \dots, m$) são os pesos recebidos pelas amostras secundárias no entorno do local a ser estimado ($x_j(x_0)$). A variância ($\sigma_L^2(x_0)$) da distribuição é dada por:

$$\sigma_L^2(x_0) = 1 - \sum_{j=1}^m \lambda_j \rho_{j,0} \quad (3.7)$$

onde $\rho_{j,0}$ é a correlação entre a variável primária e secundária no local a ser estimado.

As distribuições *a priori* e *likelihood* são combinadas a partir da estatística bayesiana para atualizar a distribuição *a posteriori* ou *updated*. Assim, em cada local não amostrado x_0 é construído uma distribuição condicionada pela variável primária e secundária. A equação 3.8 é utilizada para definir a média $\bar{y}_U^2(x_0)$ e a equação 3.9, para a variância $\sigma_U^2(x_0)$ da distribuição *a posteriori* ou *updated*.

$$\bar{y}_U^2(x_0) = \frac{\bar{y}_L(x_0)\sigma_P^2(x_0) + \bar{y}_P(x_0)\sigma_L^2(x_0)}{\sigma_P^2(x_0) - \sigma_P^2(x_0)\sigma_L^2(x_0) + \sigma_L^2(x_0)} \quad (3.9)$$

$$\sigma_U^2(x_0) = \frac{\sigma_L^2(x_0)\sigma_P^2(x_0)}{\sigma_P^2(x_0) - \sigma_P^2(x_0)\sigma_L^2(x_0) + \sigma_L^2(x_0)} \quad (3.10)$$

Para inferir n -possíveis valores, a distribuição *updated* é amostrada de forma estocástica e sequencial utilizando o algoritmo de simulação de Monte Carlo (MCS), para cada interação. O valor sorteado é utilizado para condicionar a distribuição *prior* do próximo local. Em seguida, análises estatísticas ou geoestatísticas podem ser executados. Na nossa abordagem, os dados primários inferidos serão denominados como *pseudohard*, adicionados as informações da variável primária e em seguida, o algoritmo de simulação estocástica será executado para que sejam obtidos n -cenários (Z^l) equiprováveis.

Sobre a quantidade de cenários gerados, trabalhos mais recentes, mostraram que 20-100 conjuntos de dados são justificados em termos de informações obtidas para o mínimo gasto computacional (Graham *et al.* 2007). É prática comum com a simulação geoestatística gerar 100-200 realizações para amostragem efetiva do espaço de incerteza.

3.5 Verificação da incerteza das distribuições condicionais nos modelos estocásticos

As técnicas alternativas de interpolação espacial, validação cruzada e *jackknife* podem ser usadas para avaliar os modelos de incerteza. Na validação cruzada, cada *datum* é removido individualmente (um por vez) e re-estimado com os remanescentes. Já o *Jackknife*, um conjunto de dados são removidos e os dados remanescentes são utilizados para reestimá-los. Após a estimativa com estes interpoladores, os valores reais podem ser comparados com estimados e a correlação entre eles avaliada e o erro calculado a partir da diferença entre o estimativa local e o valor real da amostra removida. Estas técnicas permitem também criar um conjunto de medições e *ccdf's* independentes nos locais em que foram removidos as amostras. Segundo Goovaerts (2005), para que estas *ccdf's* sejam consideradas precisas e acuradas, o valor verdadeiro da amostra $z(x)$ deve estar dentro do intervalo de probabilidade simétrico desta distribuição condicional.

Uma distribuição será considerada acurada, se o intervalo dos valores verdadeiros cair dentro do intervalo p simétrico de probabilidade (IP) da *ccdf* local para todos os p entre $[0,1]$. A precisão de uma distribuição de probabilidade acurada é medida considerando a proximidade dos intervalos do valores verdadeiros para todos os p entre $[0,1]$.

Portanto, para se avaliar a distribuição de probabilidade, o primeiro passo consiste em calcular o intervalo de probabilidade simétrico para cada distribuição de probabilidade (*ccdf*). Por exemplo, para um intervalo probabilidade (IP) $p = 0,80$, o intervalo mínimo (Equação 3.11) e máximo (Equação 3.12) de probabilidade simétrico da distribuição pode ser calculado da seguinte forma:

$$p_{min} = \frac{(1 - p)}{2} = \frac{(1 - 0,80)}{2} = \frac{0,20}{2} = 0,10 \quad (3.11)$$

$$p_{max} = \frac{(1 + p)}{2} = \frac{(1 + 0,80)}{2} = \frac{1,80}{2} = 0,90 \quad (3.12)$$

A figura 3.5 mostra exemplo de uma distribuição com intervalo de probabilidade simétrico para $p = 0,80$. Isto significa que para $p = 0,80$, os valores

verdadeiros da amostras devem cair entre 0,10 e 0,90, da *ccdf* local para que seja considerado acurado.

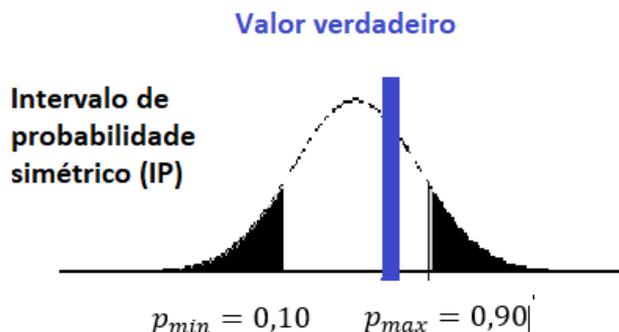


Figura 3. 5 – Distribuição com intervalo de probabilidade simétrico para $p = 0,80$.

Para isto, cada valor da *ccdf* em um local (x_0) é identificado por uma função de indicadores (Equação 3.13). Se o valor da *ccdf* estiver entre os intervalos de probabilidade mínimo e máximo será identificado como 1, caso contrário será 0.

$$\xi(x_0; p) = \begin{cases} 1 & \text{se } F^{-1}\left(x_0; \frac{(1-p)}{2}\right) < z(x_0) \leq F^{-1}\left(x_0; \frac{(1+p)}{2}\right) \\ 0 & \text{caso contrário} \end{cases} \quad (3.13)$$

A figura 3.6 mostra como é realizado a codificação do valor verdadeiro para n distribuições em que 1 representa que o valor verdadeiro está dentro e 0 quando está fora da distribuição.

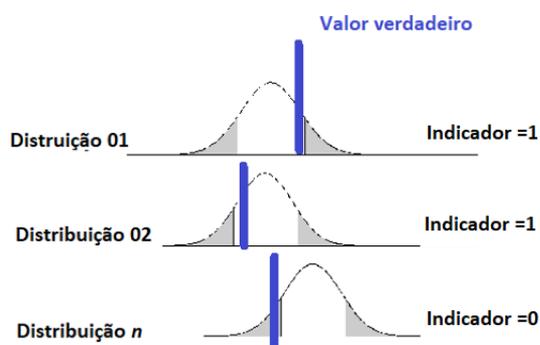


Figura 3. 6- Codificação do valor verdadeiro em distribuições com intervalo de probabilidade simétrico.

Em seguida, é calculada a média da função indicadora ($\overline{\xi(p)}$) para todos os n valores da distribuição em cada local x_0 :

$$\overline{\xi(p)} = \frac{1}{n} \xi(x_0; p) \quad (3.14)$$

Portanto, modelos incerteza são considerados acurados quando $\overline{\xi(p)} \geq p$. A precisão é definida a partir da proximidade $\overline{\xi(p)}$ com p , quando os dados são acurados. O gráfico de dispersão ou *accuracy plot* entre os valores estimados $\overline{\xi(p)}$ versus p verdadeiros é uma ferramenta utilizada para avaliar a precisão e acuracidade. A figura 3.7 exemplifica o gráfico de precisão. Os resultados são analisados a partir dos pontos plotado em relação a linha de 45°. Em situações, que todos os pontos estão acima da linha 45°, os modelos são considerados acurados e imprecisos. Se os pontos estão sobre a linha, os modelos são considerados precisos e acurados. Caso, os pontos estejam abaixo da linha, os modelos não são precisos e nem acurados.

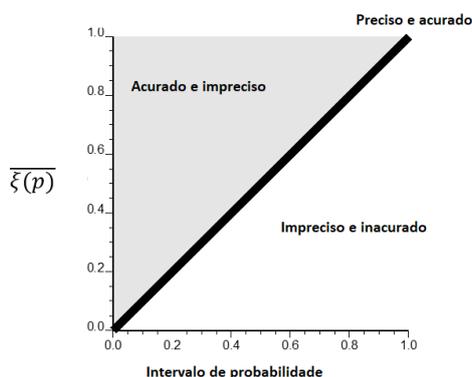


Figura 3. 7 – Gráfico de precisão entre os valores dentro do intervalo de probabilidade e os intervalos de probabilidade simétrico.

Deustch (1997) propôs o coeficiente de ajuste (*goodness* (G)) para avaliar a proximidade das frações estimadas $\overline{\xi(p)}$ e teóricas p , definido como:

$$G = 1 - \frac{1}{K} \sum_{k=1}^K w(p_k) |\xi(p_k) - p_k| \quad (3.15)$$

Em situações, que entre dois modelos probabilísticos com estatísticas *goodness* são semelhantes, a distribuição com menor espalhamento é considerada menos incerto.

3.6 Estudo de caso

Para ilustrar a metodologia, foi utilizado um banco de dados real de mina subterrânea descrito no item 2.3 - Estudo de caso. Neste estudo de caso, foram avaliadas quatro metodologias definidas como Método 01_TB_Hard_only - Simulação estocástica utilizando apenas a variável primária, Método 02_DBN_PF_TB - Simulação estocástica utilizando a variável primária e secundária como distribuição bivariada normal, Método 03_ICK_PF_TB - Simulação estocástica utilizando a variável primária e secundária como distribuição inferida por cokrigagem e Método 04_BU_TB - Simulação estocástica utilizando inserção múltipla com atualização bayesiana no local da secundária. Foram amostrados 50 *pseudohard* das distribuições da variável secundária com cada metodologia proposta e em seguida foram adicionados os dados da variável primária e executado o algoritmo de simulação estocástica.

Neste estudo, foram utilizados os mesmos parâmetros para realizar a simulação estocástica (estratégia de busca, modelo de continuidade espacial da variável primária, busca de amostras para realizar cada realização) para ambos os métodos. Foi utilizado o algoritmo de simulação bandas rotativas (*Turning Bands*) desenvolvido por Matheron (1973) e Journel (1974) e foram criados múltiplos cenários para cada método.

3.6.1 Validação

3.6.1.1 Validação *Pseudohards*

A figura 3.8 mostra o histograma Gaussiano univariado das variáveis primárias (figura 3.8a) e secundárias (figura 3.8b) obtidas pela transformação *n*score necessária para construção das distribuições de probabilidade para realizar a atualização Bayesiana e executar o algoritmo de simulação estocástica. Os resultados mostram que os dados após a transformação os dados são normalmente distribuídos (média zero, variância igual a um).

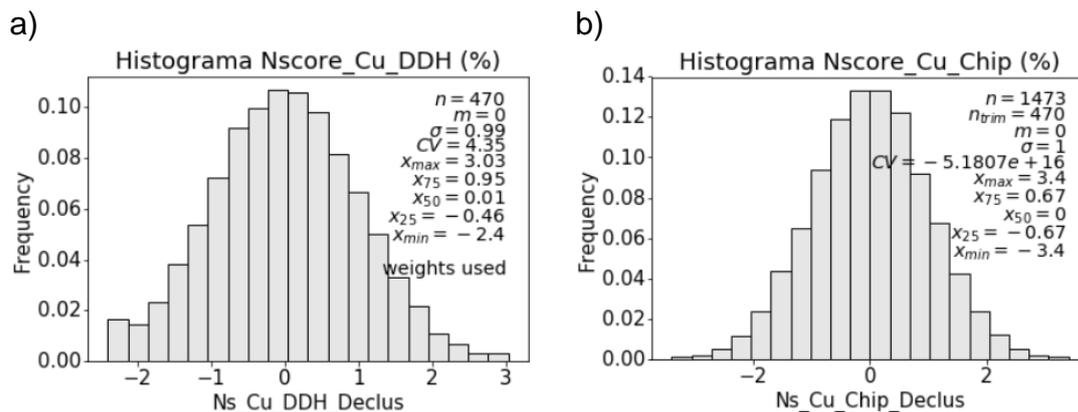
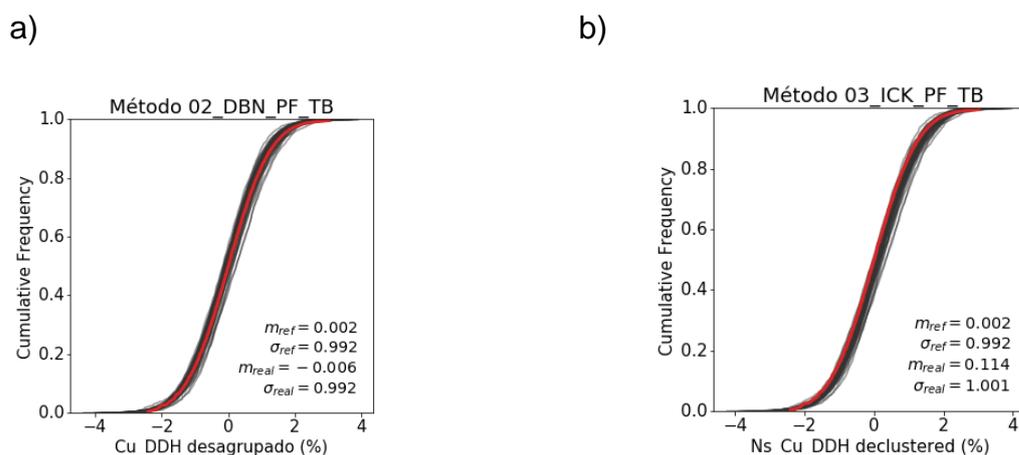


Figura 3. 8- Histograma dos dados após a transformação gaussiana a) variável primária Cu_DDH desagrupada e b) variável secundária Cu_Chip desagrupada.

A figura 3.9 mostra o histograma acumulado dos *pseudohards* (linhas pretas) gerados pelas metodologias propostas e histograma da variável primária (Cu_DDH) desagrupada (linha vermelha). Podemos verificar, que nas três metodologias propostas 02_DBN_PF_TB, 03_ICK_PF_TB e 04_BU_TB, a média (m_{real}) e desvio padrão (σ_{real}) dos *pseudohards* gerados estão próximos aos dados da variável primária. Portanto, podemos afirmar que a distribuição dos *pseudohards* gerados possuem a mesma distribuição acumulada dos dados da variável primária.



c)

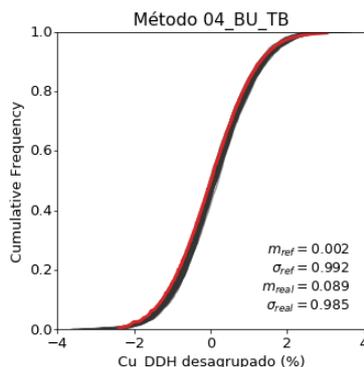


Figura 3. 9 – Histograma dos dados *pseudohard* (linhas pretas) gerados comparados com o histograma da variável primária (Cu_DDH) (linha vermelha) a) Método 02_DBN_PF_TB b) Método 03_ICK_PF_TB e c) Método 04_BU_TB.

Considerando, que as metodologias foram aplicadas em uma parte do depósito estudado, para avaliar a reprodução da continuidade espacial dos *pseudohard* gerados, foram comparados com o modelo de variograma do depósito (linha vermelha) e o variograma experimental da variável primária (linha azul), os variogramas *pseudohards* gerados (linhas pretas) nas direções principal (h_1), intermediária (h_2) e vertical (h_3) continuidade espacial.

A figura 3.10 mostra que nas metodologias investigadas, a continuidade espacial dos *pseudohards* gerados (linha preta) é semelhante ao variograma experimental da variável primária (linha azul), apesar da pouca quantidade de dados da variável primária (470 dados). Para as metodologias 02_DBN_PF_TB e 03_ICK_PF_TB, é possível verificar flutuações ergódicas maiores. Isto ocorre porque os *pseudohards* foi amostrado por campos de probabilidade (*p-field*) de forma independente. Ao compararmos a metodologia 04_BU_TB, as flutuações são menores, porque os *pseudohards* gerados são amostrados de forma estocástica e sequencial. Isto significa que, para cada nova amostragem, os *pseudohards* já amostrados são utilizados para condicionar o *pseudohards* seguinte.

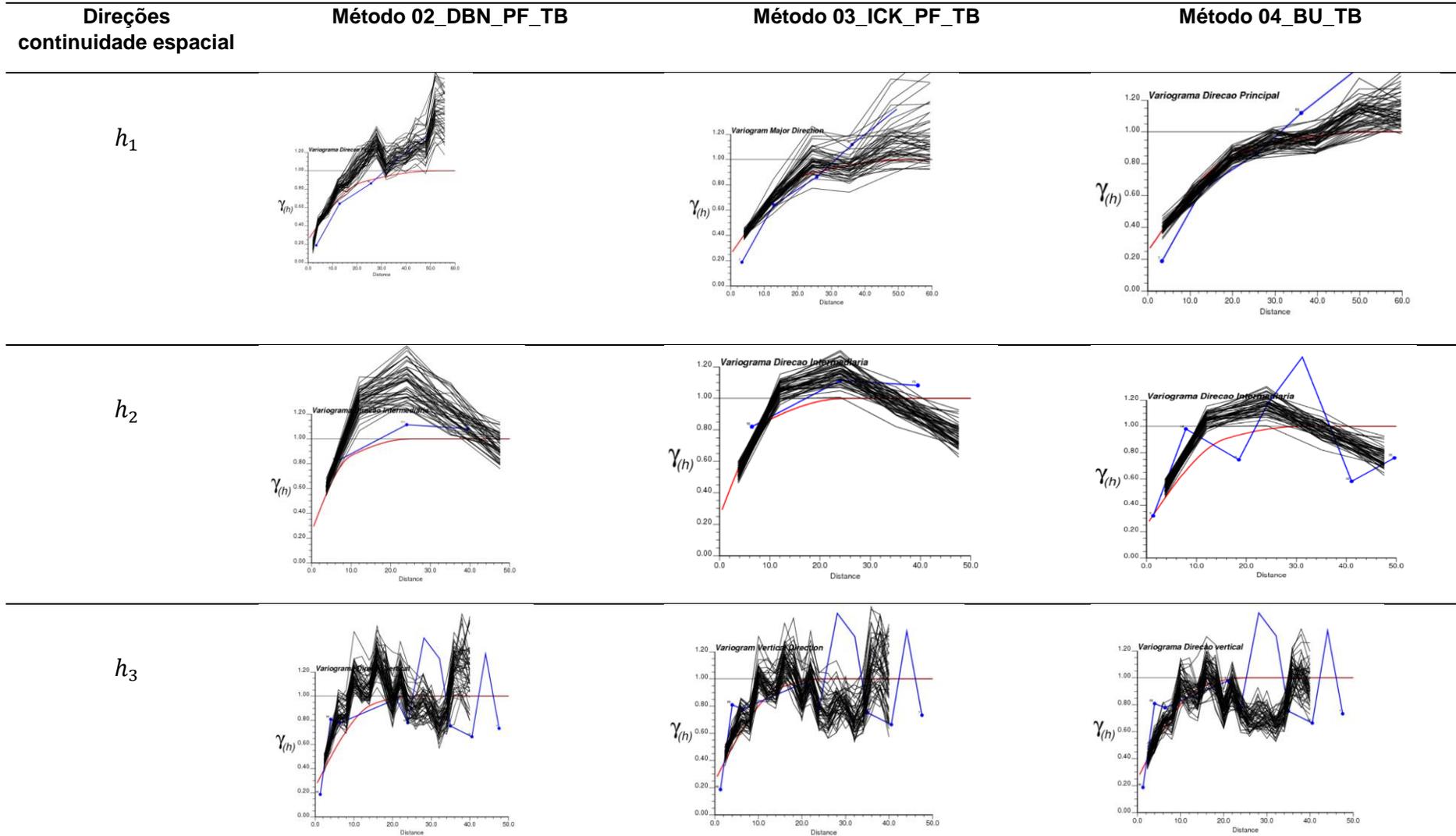


Figura 3. 10 – Reprodução dos variogramas dos *pseudohards* gerados e a variável primária (linhas preta), modelo do variograma dos dados da variável primária (linha vermelha) e variograma experimental da variável primária (linha azul) nas direções de continuidade espacial a) direção principal (h_1), b) direção intermediária (h_2), e c) direção vertical (h_3)

3.6.1.2 Validação dos modelos finais

Na figura 3.11, a linha vermelha representa o histograma cumulativo da variável primária (Cu_DDH) desagrupada e as linhas pretas, os histogramas de cada realização obtida pelos métodos investigados. Verificamos, que houve a reprodução estatística da variável primárias nas metodologias investigadas. Assim, podemos afirmar que, nas abordagens propostas, a variável secundária imprecisa e inacurada foi incorporada, seus os erros de viés e imprecisão foram atenuados (filtrados) e não foram transferidos para os modelos finais para todas as metodologias propostas.

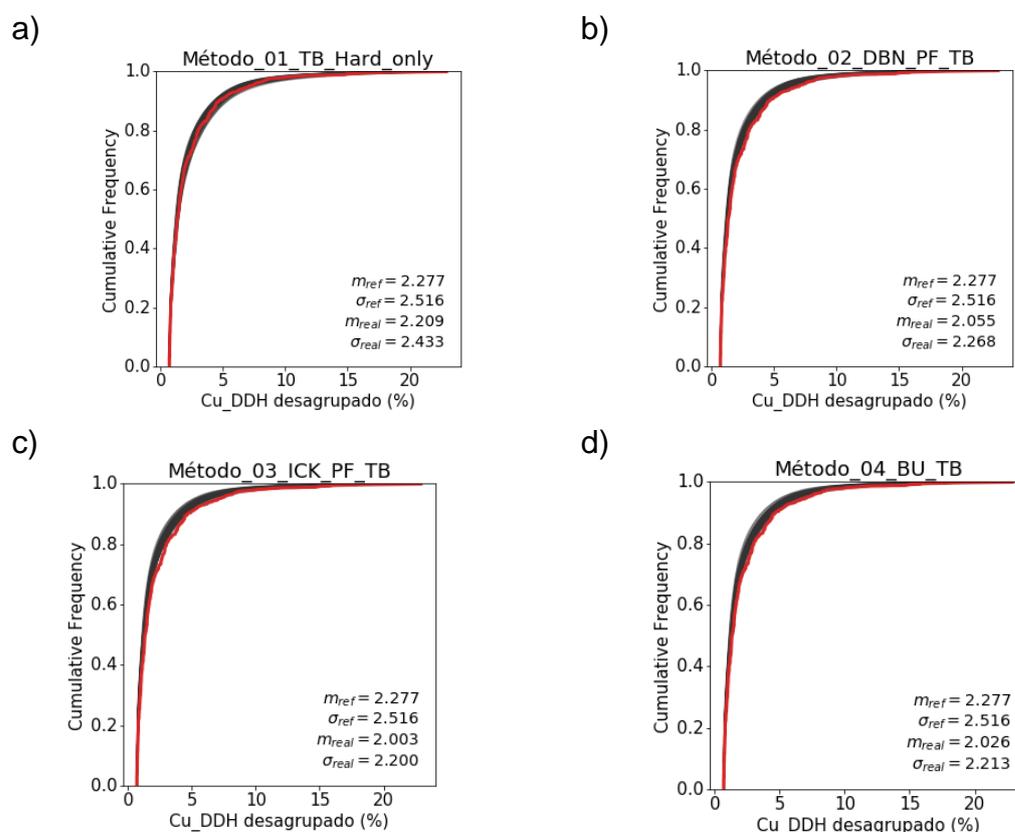


Figura 3. 11 - Histograma dos modelos simulados (linhas pretas) comparados com o histograma da variável primária a) Método 01_TB_Hard_only, b) Método 02_DBN_PF_TB, c) Método 03_ICK_PF_TB e d) Método 04_BU_TB.

A figura 3.12 mostra a reprodução dos variogramas pelos modelos obtidos, a linha vermelha representa o modelo de variograma da variável primária (Equação 2.6) e as linhas preta, as realizações obtidas pelas metodologias 02_DBN_PF_TB, 03_ICK_PF_TB e 04_BU_TB investigadas, nas direções principal, intermediária e vertical. Os resultados mostram que todas as metodologias reproduzem a continuidade espacial da variável primária.

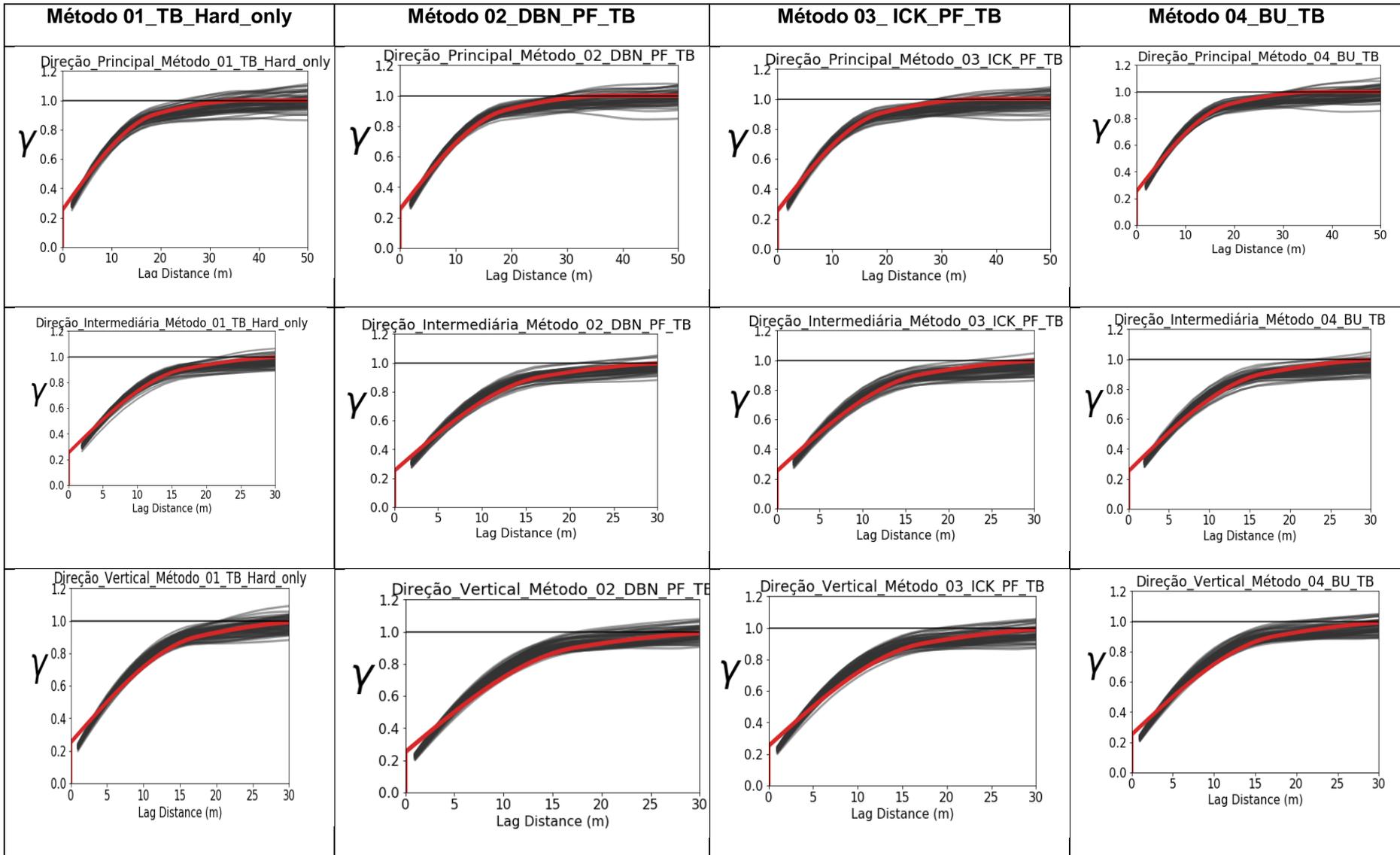


Figura 3. 12 – Reprodução dos variogramas dos modelos finais (linhas preta) e o modelo do variograma do variograma (linha vermelha) nas direções de continuidade espacial a) direção principal (h_1), b) direção intermediária (h_2), e c) direção vertical (h_3)

A figura 3.13 mostra o histograma da variância condicional dos modelos obtidos pelas quatro metodologias. A média da variância condicional (5,47) obtida usando simulação estocástica apenas com a variável primária (figura 3.13a) é maior quando comparados as outras metodologias. Na metodologia 02_DBN_PF_TB, a variância condicional dos modelos finais diminui em 17% (5,47 para 4,56) e na metodologia 04_BU_TB reduz em 18% (5,47 para 4,45). Para a construção da distribuição nestas abordagens, são utilizadas somente a variável secundária local, o que justifica as reduções na variância com valores próximos entre elas. Já na metodologia 03_ICK_PF_TB, a redução foi de 20% (5,47 para 4,38). Nesta abordagem, a distribuição construída utiliza todas as amostras primárias e secundárias dentro da vizinhança de busca da krigagem, assim mais amostras são utilizadas para condicionar a distribuição de onde será sorteado o pseudohard. Portanto, reduz a incerteza global e local, melhorando o modelo final.

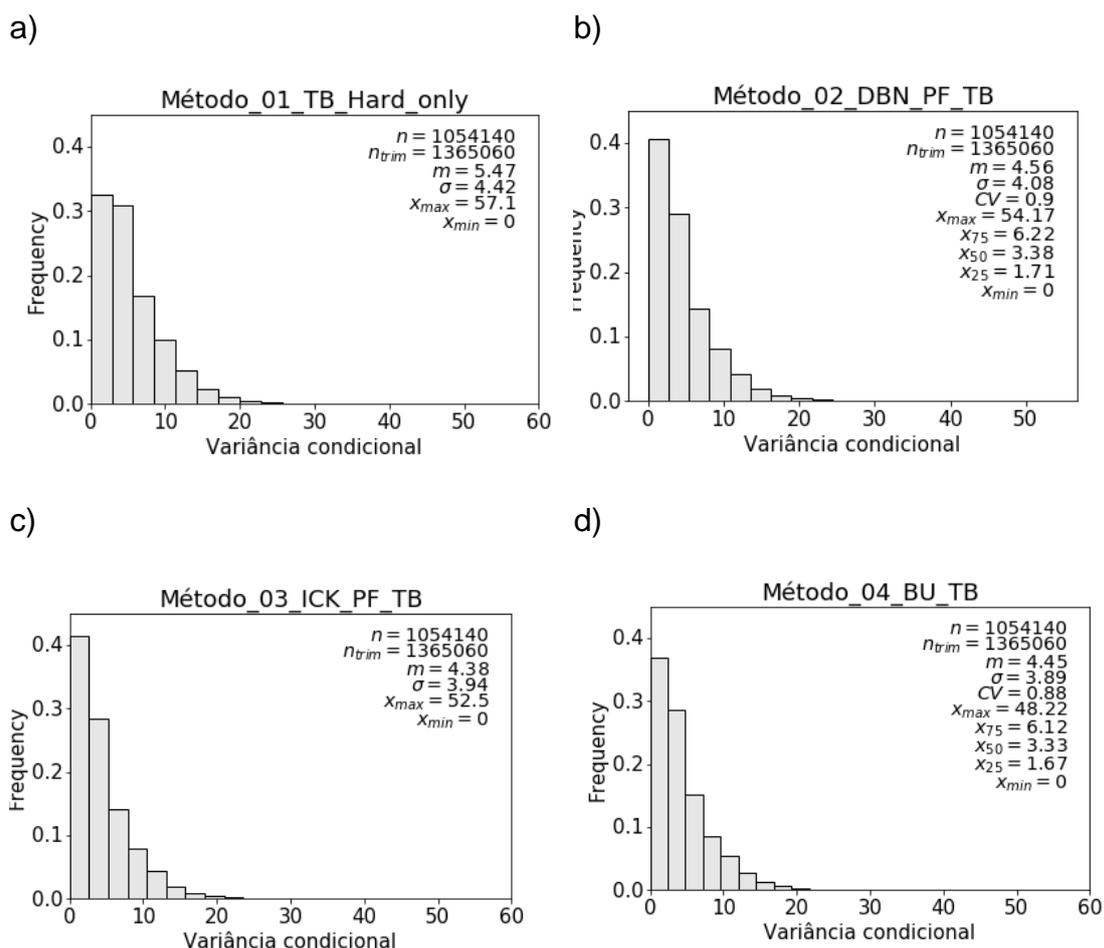
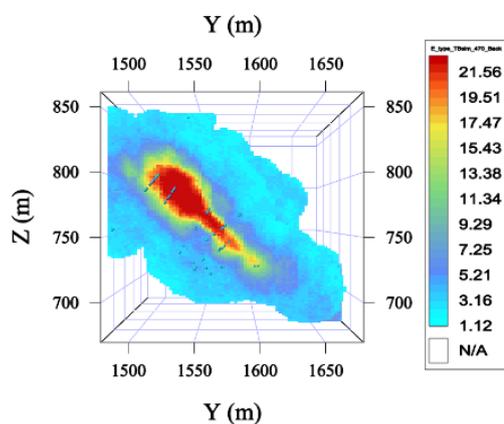


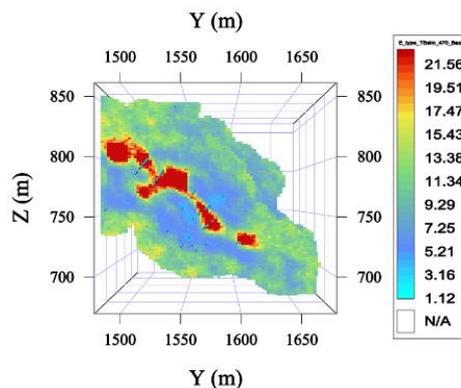
Figura 3. 13-Histograma da variância condicional do modelos simulados para os métodos a) Método 01_TB_Hard_only, b) Método 02_DBN_PF_TB, c) Método 03_ICK_PF_TB e d) Método 04_BU_TB.

A figura 3.14 mostra o modelo médio, conhecido como E-type do conjunto das 50 realizações obtidas para cada metodologia investigada utilizando a mesma estratégia de busca. Para comparar os modelos obtidos foi selecionada uma mesma região, com a mesma escala de cor que representa os teores para todas as metodologias investigadas. Podemos verificar que o Método 01_TB_Hard_only (figura 3.14a) que foi utilizado somente a variável primária, ou seja somente 470 amostras, o modelo médio foi suavizado e região com de alto teor foi superestimada quando comparada as demais metodologias. Nas metodologias que utilizaram as amostras primárias e os pseudohards obtidos a partir da amostragem da distribuição de probabilidade, a região de alto teor foi melhor delimitada. Nas metodologias 02_DBN_PF_TB (figura 3.14b) e 04_BU_TB (figura 3.14d), que utiliza somente a variável secundária para construir a distribuição probabilidade local, no Método 02_DBN_PF_TB, a área de alto teor foi superestimada quando comparada ao Método 04_BU_TB. Para o método 03_ICK_PF_TB (figura 3.14c), em que foram utilizadas mais amostras para construir a distribuição de probabilidade no local da variável secundária, foi verificado que a região de alto teor possui melhor resolução e definição quando comparada aos métodos. Portanto, podemos verificar que o Método 03_ICK_PF_TB produz modelos mais precisos e acurados.

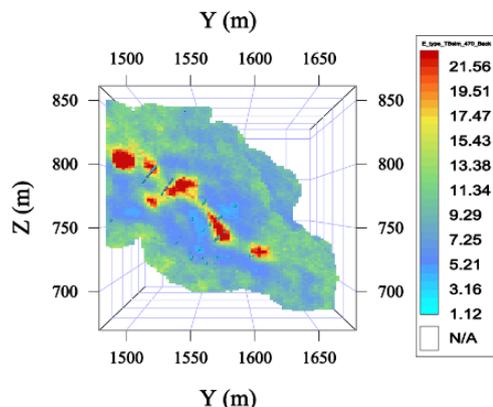
a)



b)



c)



d)

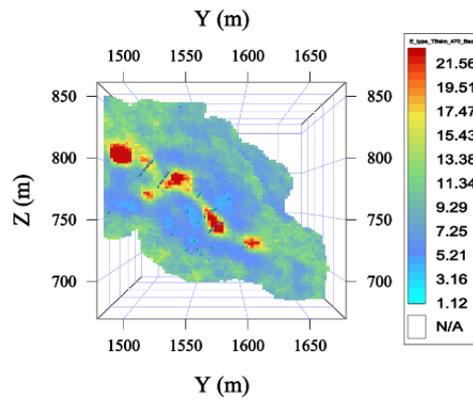


Figura 3. 14- Mapa dos modelos médios (E-type) de 50 realizações obtidos pelos métodos: a) Método 01_TB_Hard_only, b) Método 02_DBN_PF_TB, c) Método 03_ICK_PF_TB e d) Método 04_BU_TB.

3.6.2. Avaliando a acuracidade e precisão dos modelos obtidos

Para avaliar a acuracidade e precisão das metodologias investigadas, foi realizado o procedimento descrito no item 2.4.2 - Avaliando acuracidade e precisão dos modelos. A figura 3.15 mostra o gráfico de dispersão entre os valores verdadeiros da variável primária (Cu_{DDH}) comparados aos valores médios obtidos por simulação estocástica (Cu_{E-type}), na mesma localização dos dados omitidos, pelas metodologias propostas e os coeficientes de correlação de Pearson (ρ) e Spearman (ρ_s) entre elas. Para a metodologia 02_DBN_PF_TB, os coeficientes de correlação foram 0,75 e 0,77 e na metodologia 04_BU_TB foram 0,71 e 0,77. Já a metodologia 03_ICK_PF_TB, apresentou coeficientes de correlação (0,80 e 0,76) maiores que as outras metodologias. Isto indica que valores previstos (Cu_{E-type}) são mais próximos dos valores considerados verdadeiros (Cu_{DDH}) e os modelos obtidos por esta metodologia é mais acurado quando comparados com os modelos obtidos pelas outras metodologias propostas.

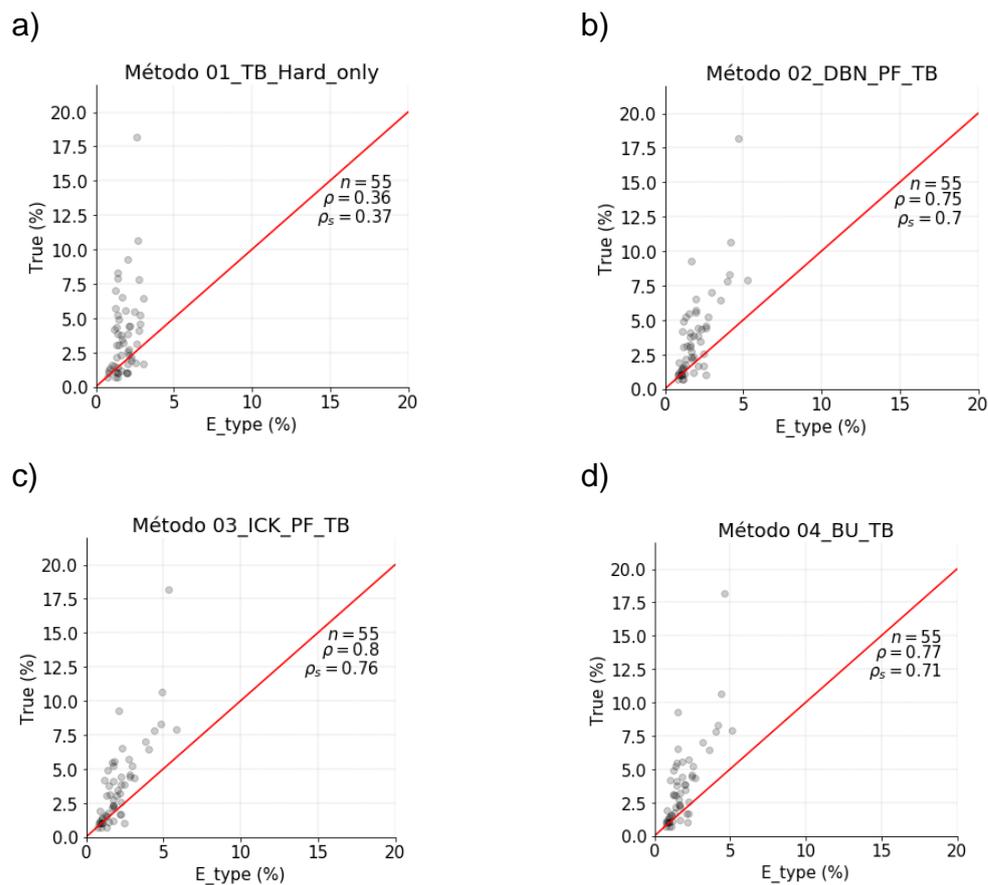
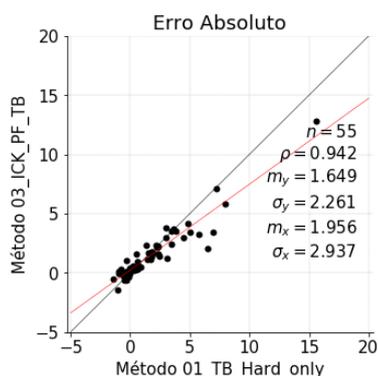


Figura 3. 15 - Gráfico de dispersão dos valores reais da variável primária (Cu_DDh) comparado ao valor médio obtido pelo E-type pelo os métodos: a) Método 01_TB_Hard_only, b) Método 02_DBN_PF_TB, c) Método 03_ICK_PF_TB e d) Método 04_BU_TB.

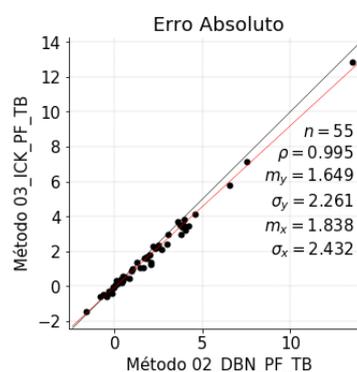
Para avaliar o erro absoluto (Equação 2.6) para cada metodologia investigada, foi construído gráfico de dispersão entre a metodologia 03_ICK_PF_TB, que foi a metodologia considerada mais acurada e as demais metodologias. Para a metodologia 01_TB_Hard_only (Figura 3.16a), a média (1,95) e o desvio padrão (2,94) do erro absoluto são maiores do que os obtidos usando a metodologia 03_ICK_PF_TB. Os dados estão mais dispersos e a maioria dos pontos estão abaixo da linha de 45°. Nesta metodologia, foi utilizado apenas a variável primária para condicionar as informações locais nos modelos. Isto indica que o uso de pouca informação (dados condicionais) produz modelos com erros absolutos maiores. Ao comparar com a metodologia 03_ICK_PF_TB, as metodologias 02_DBN_PF_TB (Figura 3.16b) e 04_BU_TB (Figura 3.16c) (em que foram utilizados a mesma quantidade de pseudohards gerados para condicionar as simulações e somente a variável secundária no local para construir a distribuição local de probabilidade) as médias (1,83 e 1,89) do erro absoluto e o desvio padrão (2,43 e 2,40) são maiores. Para a metodologia

03_ICK_PF_TB, a distribuição foi construída utilizando a variável secundária local, na vizinhança e a variável primária, a média foi 1,65 e desvio padrão 2,26 foram menores. Isto indica que os erros absolutos são menores. Esta redução do erro absoluto pode ser explicada porque mais informações são utilizadas no processo de simulação estocástica para condicionar as informações locais e, conseqüentemente, os resultados são mais precisos.

a)



b)



c)

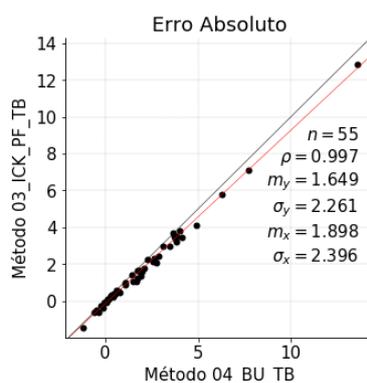


Figura 3. 16 - Gráfico de dispersão do erro absoluto do Método 03_ICK_PF_TB comparado com as outras metodologias a) Método 01_TB_Hard_only, b) Método 02_DBN_PF_TB e c) Método 04_BU_TB.

O erro relativo, definido como o erro absoluto (Equação 3.16) dividido pelo valor considerado verdadeiro (Cu_DDH), foi calculado em cada local para avaliar a porcentagem do erro considerando a magnitude do valor da variável primária.

$$\text{Erro}_{\text{rel}} = \left| \frac{Cu_{DDH} - Cu_{E\text{-type}}}{Cu_{DDH}} \right| * 100 \quad (3.16)$$

A tabela 3.1 o comparativo do erro relativo médio das metodologias investigadas. Podemos observar que na metodologia 01_TB_Hard_only em que foi utilizada apenas a variável primária o erro relativo foi 51,04%. Nas metodologias em que foram utilizadas mais informações, o erro relativo médio diminuiu. A redução maior ocorreu na metodologia 03_ICK_PF_TB (41,33%) em que foram utilizadas amostras primária e secundárias para condicionar a distribuição local, mostrando que é mais precisa que as outras metodologias investigadas.

Tabela 3. 1-Tabela comparativa do erro relativo das metodologias investigadas.

Metodologia	Média (%)	Desvio Padrão	Mínimo	Máximo
Método 01_TB_Hard_only	51,04	26,63	0,52	106,12
Método 02_DBN_PF_TB	45,27	26,41	0,82	168,46
Método 03_ICK_PF_TB	41,33	24,35	2,29	149,17
Método 04_BU_TB	45,09	22,24	5,03	124,47

A tabela 3.2 mostra o coeficiente de *Goodness* que mede a similaridade entre os valores reais e os estimados dos intervalos de probabilidade nos gráficos de precisão. Podemos verificar, que as metodologias 02_DBN_PF_TB, 03_ICK_PF_TB e 04_BU_TB, onde foram utilizados mais informação para condicionar as realizações, possuem coeficientes de *Goodness* maiores que a metodologia 01_TB_Hard_only. Quando comparadas a metodologia 04_BU_TB possui incerteza (variância) 5% menor. Entretanto, a acurácia, nas metodologias 02_DBN_PF_TB e 03_ICK_PF_TB são maiores (26 %). Portanto, estas metodologias são mais precisas e acuradas que as outras investigadas.

Tabela 3. 2-Tabela comparativa do coeficiente de "goodness"

Metodologia	Acurácia	Precisão	Coeficiente Goodness	Variância
Método 01 TB_Hard_only _50	1,00	0,78	0,89	3,62
Método 01 TB_Hard_only _100	1,00	0,78	0,89	3,57
Método 02_DBN_PF_TB	0,89	0,95	0,96	2,63
Método 03_ICK_PF_TB	0,89	0,95	0,96	2,63
Método 04_BU_TB	0,63	0,95	0,96	2,46

A exatidão e precisão local dos modelos obtidos foram avaliados pelo gráfico de acuracidade conhecido também como *Accuracy Plot*. Estes foram calculados usando o software *Accplot-sim.exe* da Geostatistical Software Library (GSLIB). Mais detalhes sobre modelos de precisão podem ser encontrados em Goovaerts (2001). A figura 3.17 mostra os gráficos de acuracidade para as quatro metodologias investigadas. Ao comparar o gráfico de acuracidade com 100 realizações (figura 3.17a) e 50 realizações (figura 3.17b) da metodologia 01 em que foram utilizados somente a variável primária no algoritmo de simulação estocástica, notamos resultados semelhantes. Portanto, podemos afirmar que 50 realizações foram suficientes para acessar a incerteza do depósito. Sobre a acuracidade dos modelos podemos verificar que o Método 01, que utiliza somente a variável primária, os modelos são considerados acurados (os pontos estão acima da linha de 45°), mas não são precisos (os pontos encontram afastados da linha de 45°). Para todos os intervalos de probabilidade para a metodologia 02 e 03 e os intervalos 0,30, 0,50 e 0,90 na metodologia 04 estão acima da linha de 45°. Nestes intervalos, os modelos obtidos são acurados, mas não são precisos. As metodologias 02 e 03, os intervalos de probabilidade são próximos a linha de 45°. Isto indica que estes modelos são mais precisos e acurados que os demais.

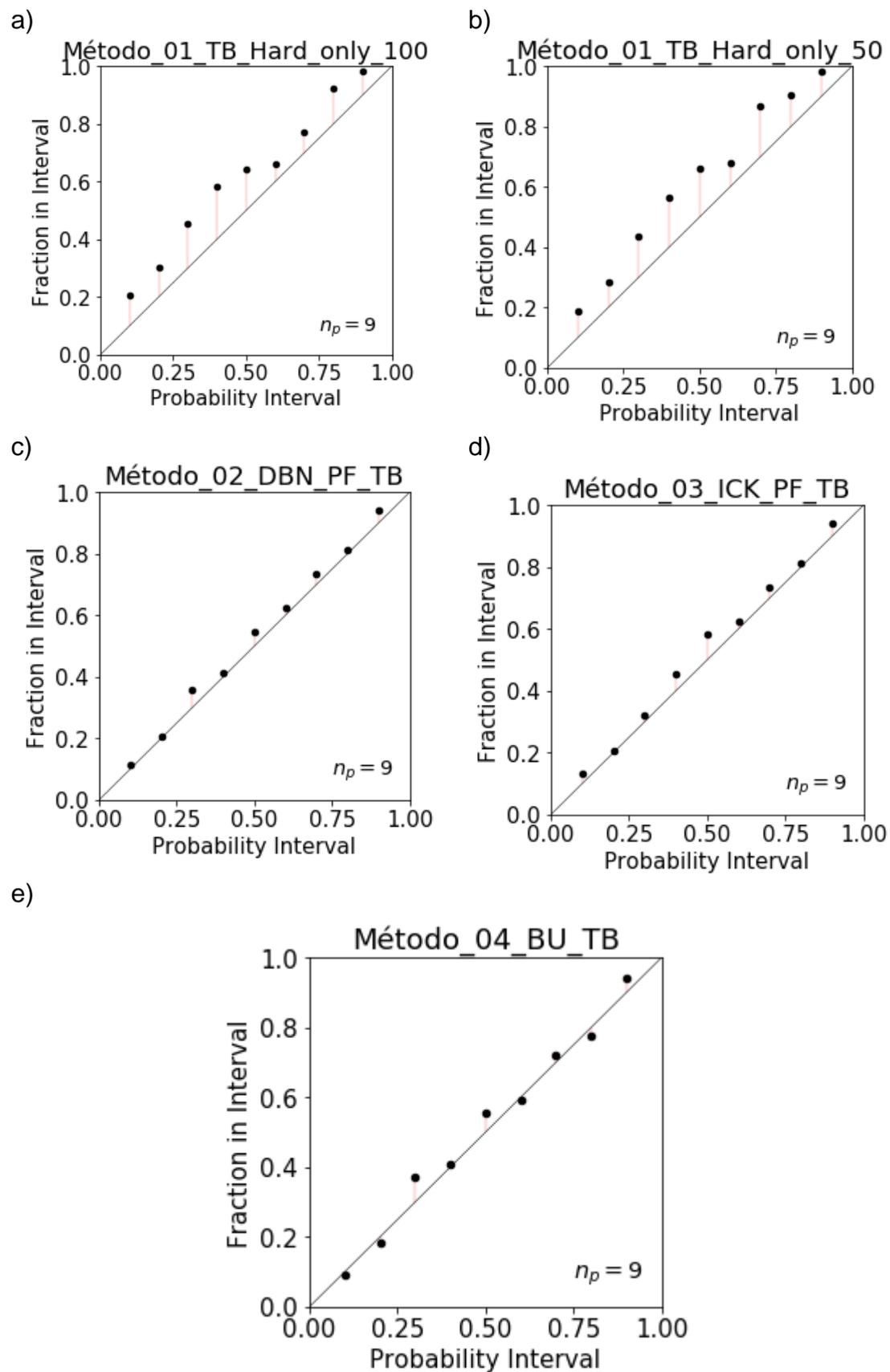


Figura 3. 17 - Gráfico de acuracidade das metodologias a) Método 01_TB_Hard_only com 100 realizações, b) Método 01_TB_Hard_only com 50 realizações, c) Método 02_DBN_PF_TB, d) Método 03_ICK_PF_TB e e) Método 04_BU_TB.

3.7 Considerações

As metodologias abordadas propõem uma abordagem para incorporar dados secundários como distribuição local utilizando algoritmos de simulação estocástica. Ao comparar os modelos obtidos por simulação estocástica utilizando apenas a variável primária, a metodologia proposta mostra que a incorporação da variável secundária como distribuição local melhora significativamente os modelos finais. Sobre a construção da distribuição de probabilidade local, foram investigadas 03 abordagens. Na distribuição bivariada normal, foi utilizada somente a variável secundária local. Na distribuição construída por cokrigagem usando o modelo de correionalização intrínseca, foram utilizados todos os dados da variável secundária dentro da vizinhança de busca e as variáveis primárias, ou seja, mais amostras para condicionar a distribuição local. Na abordagem por atualização bayesiana, a distribuição é uma combinação entre a distribuição da variável primária e secundária. Nas situações em que foram utilizadas mais amostras, para condicionar as distribuições locais, os modelos foram mais acurados e precisos. Sobre a validação dos modelos finais, todas as metodologias investigadas reproduziram o histograma e a continuidade espacial da variável primária. Ao comparar a variância condicional dos modelos, a metodologia 03_ICK_PF_TB teve maior redução (20%), quando comparado a metodologia 01_TB_Hard_Only

Sobre a validação dos *pseudohards* gerados, todas as metodologias reproduziram o histograma da variável primária e sua continuidade espacial. A metodologia 04 houve reprodução melhor e com menores flutuações nos modelos do *pseudohards* quando comparados com o modelo da variável primária. Isto ocorreu porque foram utilizadas duas abordagens para amostrar a distribuição de probabilidade local. A primeira abordagem, que foi utilizada nas metodologias 02_DBN_PF_TB e 03_ICK_PF_TB, utilizou-se o algoritmo de simulação por campos de probabilidade (*p-field*), para garantir que os *pseudohards* seriam correlacionados com a variável primária e a continuidade espacial reproduzida. A limitação teórica abordada por Pyrcz e Deustch (2001), não foi verificada neste estudo de caso. Acredita-se, que como o algoritmo foi utilizado somente para amostrar a distribuição nos locais da variável secundária

e o para o restante do grid foi utilizado o algoritmo de simulação por bandas rotativas (*Turning Bands*), com os dados com a covariância correta, os artefatos foram suavizados no modelo final. Sobre a metodologia 02_DBN_PF_TB que utiliza somente a variável secundária colocada para construir a distribuição, este método pode ser considerado mais rápido e simples. Os resultados mostraram que a reprodução da continuidade dos pseudohards foi deteriorada. Na metodologia 04_BU_TB, a distribuição é amostrada de forma sequencial, o valor amostrado é utilizado para condicionar a amostragem posterior. Ao comparar estas duas abordagens, a abordagem sequencial reproduziu melhor a continuidade espacial da variável primária. É importante ressaltar que o método de atualização bayesiana (metodologia 04_BU_TB), a construção da distribuição a *priori* e *likelihood* são realizadas a partir da estimativa por krigagem simples das variáveis primária e secundária no espaço Gaussiano, dando origem a média e variância da distribuição multiGaussiana a ser amostrada. Portanto, não leva em consideração a incerteza dos dados.

Sobre a precisão e acuracidade, este estudo de caso demonstrou que na metodologia 03_ICK_PF_TB, a correlação entre o valor considerado como verdadeiro (Cu_DDH) e o valor médio (Cu_E-type) é maior, os erros absolutos são menores, a média do erro relativo é menor. Na estatística *Goodness*, que avalia a discrepância entre o valor estimado e o valor real, os parâmetros: coeficiente de *Goodness* e a acurácia são maiores e mais próximos a 1 que nas demais metodologias investigadas. Isto indica que esta metodologia é mais acurada e precisa. Nos gráficos de acuracidade, conhecidos como *Accuracy Plot*, os pontos estão próximos a linha de 45°. Portanto, os resultados apresentados por esta metodologia proposta foram mais precisos e acurados

O fluxo de trabalho proposto na metodologia 03_ICK_PF_TB pode ser uma solução alternativa fácil para incorporar dados secundários na simulação, desde que os dados estejam no mesmo suporte, uma vez que precisa de poucas etapas e a modelagem da continuidade espacial apenas para os dados da variável primária

Em situações que a variável primária não estão no mesmo suporte, deverá ser utilizada as metodologias investigadas no capítulo 2.

4 CONSIDERAÇÕES FINAIS

A tese abordou três assuntos: (1) integração das amostras com suporte diferente ou igual nos algoritmos simulação em geoestatística (2) integração da variável secundária como distribuição de probabilidade, (3) incorporação de informação secundária sem utilizar o modelo linear de correionalização. A seção 4.1 mostra o comparativo entre as metodologias investigadas. A seção 4.2 contribuições da tese. A seção 4.3 alerta sobre as limitações dos estudos desenvolvidos. Por último, a seção 4.4 lista sugestões para trabalhos futuros.

4.1 Comparativo entre as metodologias investigadas

A tabela 4.1 mostra o comparativo entre as metodologias investigadas nesta tese

a) Sobre os algoritmos de simulação estocástica

Distribuição dos dados da variável primária e secundária	Imposição teórica	Vantagens	Algoritmos de simulação estocástica disponíveis
Dados normalizados ou no espaço gaussiano	Dados com mesmo suporte	<ul style="list-style-type: none"> -É necessário a transformação <i>a priori</i> dos dados, -Deve garantir que os dados tenham média zero e variância igual a um 	<ul style="list-style-type: none"> -Simulação sequencial Gaussiana (sgsim) -Simulação por bandas rotativas (TBSim)
Dados no espaço original	Dados podem ter suporte diferente	<ul style="list-style-type: none"> -Não é realizada nenhuma transformação nos dados, - Honra as relações lineares entre as amostras nas diversas escalas. -Reproduz melhor a continuidade espacial dos dados 	<ul style="list-style-type: none"> -Simulação sequencial direta (DSS)

b) Sobre as metodologias para incorporar dados secundários com heterotopia total

Metodologias	Imposição teórica	Vantagens	Desvantagens	Metodologias disponíveis/investigadas
Metodologias clássicas	- Existir correlação entre os dados primários e secundários	- Incorpora a informação secundária nos modelos, a partir do modelo linear de correionalização (MLC). -Utiliza a variável secundária como ponto.	- Depende da configuração dos dados para definir qual a metodologia adequada. -É necessário modelar todas as variáveis, o que torna a metodologia laborioso computacionalmente. - Na maioria dos algoritmos, está implementada o algoritmo de cokrigagem colocada que exige a variável secundária exaustiva. - Simula a variável secundária em todo o grid e gera maior esforço computacional.	-Algoritmos de co-simulação, cokrigagem e suas derivações. - Metodologia para inferência da variável primária para que o banco de dados se torne isotópicos, a partir de regressão linear, atualização bayesiana, etc.

Técnicas de inserção múltiplas		<ul style="list-style-type: none"> - Transforma problema multivariado em univariado. - Utiliza a variável secundária como uma distribuição de probabilidade. - Eficiência computacional, simula a variável secundária somente no local. 	<ul style="list-style-type: none"> - A construção da distribuição de probabilidade deve ser avaliada e investigada. 	<ul style="list-style-type: none"> - Atualização bayesiana combinada com algoritmos de simulação estocástica - Simulação sequencial direta com distribuição pontual (DSS) utilizando krigagem e cokrigagem simples. - Metodologias que inferem a incerteza nos dados a partir de distribuições de probabilidade (metodologias investigadas no capítulo 03)
--------------------------------	--	--	--	---

c) Sobre a construção da distribuição de probabilidade no local da secundária

Metodologias	Amostras consideradas na construção da distribuição	Vantagens
Distribuição bivariada normal	- Utiliza somente a variável secundária local para construir a distribuição.	- A incerteza local é incorporada na distribuição de probabilidade. -Evita a utilização do modelo linear de correção regionalização (MLC).
Distribuição construída por atualização bayesiana (BU)		
Distribuição construída a partir da cokrigagem simples	- Utiliza a variável secundária no local em que será construída a distribuição, as secundárias localizadas na vizinhança e a variável primária para construir a distribuição.	

d) Sobre a amostragem da distribuição de probabilidade no local da secundária (*pseudohards*)

Metodologias	Vantagens	Desvantagens
Amostragem sequencial por simulação de Monte Carlo (<i>MCS Simulation</i>)	<ul style="list-style-type: none"> - Reproduz melhor a distribuição e continuidade espacial dos <i>pseudohards</i> gerados, já que o nó anterior é utilizado para condicionar a próxima distribuição a ser amostrada. 	<ul style="list-style-type: none"> - Maior esforço computacional
Amostragem por simulação campos de probabilidade (<i>p-field</i>)	<ul style="list-style-type: none"> - Rápido - Os dados amostrados são correlacionados espacialmente -Flexibilidade e fácil implementação em vários workflows de simulação 	<ul style="list-style-type: none"> -Os variogramas dos <i>pseudohards</i> possuem maior flutuações ergódicas quando comparadas aos modelos dos dados.

4.2 Contribuições da tese

A seguir são apresentadas as considerações referentes as contribuições propostas nesta tese.

4.2.1 Incorporação da incerteza na variável secundária considerada como distribuição de probabilidade (*pdf*)

Esta tese propôs uma nova abordagem para incorporar a variável secundária nos algoritmos de simulação estocástica. No contexto de geoestatística, a variável secundária possui informações quanti e qualitativas que auxiliam a compreender o comportamento geológico da variável de interesse. Nas abordagens tradicionais, assumimos que esta segunda variável é um ponto com um único valor de medida. Nesta tese, assumimos uma abordagem probabilística para essa variável secundária (no caso dela ter uma medida imprecisa no seu valor), que passa a ser considerada como uma distribuição de probabilidade. A incerteza é adicionada *a priori* nos dados. Portanto, admitimos que as medidas possuem flutuações e incerteza/erros e que são incorporados na amostra.

Dentre as metodologias disponíveis em geoestatística, somente o algoritmo de krigagem de indicadores (*soft indicator kriging*), proposto por Alabert (1987), utiliza esta abordagem de considerar os dados secundários com uma distribuição local (*cdf's*). A limitação desta proposta é que necessita que se modele múltiplos variogramas para k indicadores tornando-a laboriosa. A vantagem da metodologia proposta nesta tese é que para a sua execução é necessário somente o variograma da variável primária, tornando fácil sua execução e implementação.

4.2.2 Metodologias que não utilizam o modelo linear de correionalização (MLC)

O uso adicional de informação, denominada secundária, para a construção dos modelos de teores tende a produzir modelos mais precisos e acurados. A escolha da metodologia para incorporar esses dados secundários pode ser considerada uma tarefa difícil. Em situações em que o banco de dados

possui heterotopia total, duas metodologias são recomendadas: i) isotopização dos dados a partir de regressões multivariadas ou inferir a variável primária com metodologias de inserção múltiplas (Barnett e Deustch, 2015) ou ii) inferência da correlação entre a variável primária e secundária para $h = 0$, a partir da projeção do correlograma cruzado entre as variáveis (Minnitt e Deustch, 2014) e posterior modelagem do modelo linear de correionalização (MLC). Esta abordagem é de difícil implementação quando o número de variáveis secundárias é superior a dois. Softwares podem ser utilizados para que sejam feitos ajustes automáticos das contribuições das variâncias na modelagem simultânea dos variogramas diretos e cruzado para satisfazer as condições de definição impostas para o modelo que a matriz de covariância seja positiva definida.

As metodologias investigadas na tese propõem uma nova alternativa que consiste em quatro etapas: i) construir a distribuição de probabilidade no local da variável secundária, que representa a incerteza local da variável primária; ii) amostrar a distribuição de probabilidade para inferir os *pseudohards* correlacionados espacialmente; iii) criar um novo banco de dados que contenha a variável primária e os n -conjuntos de *pseudohards* amostrados e iv) executar o algoritmo de simulação estocástica com o modelo de continuidade espacial da variável primária. Portanto, o desafio das metodologias é avaliar qual a forma adequada para construir as distribuições de probabilidade no local da variável secundária.

Outro benefício das metodologias investigadas nos capítulos 2 e 3, é que a informação secundária imprecisa e enviesada foi incorporada sem que o erro de viés e imprecisão fosse transferido para o modelo final.

As metodologias investigadas, no capítulo 2, podem ser aplicados a dados com suporte igual ou diferente e utilizou os algoritmos de krigagem e cokrigagem para inferir os *pseudohards*. A partir da validação por *Jackknife*, verificou-se que a utilização do algoritmo de cokrigagem (metodologia 03), produziu modelos mais acurados e precisos.

No capítulo 3, verificou-se que a construção das distribuições de probabilidade nos locais da secundária quando utilizadas as variáveis primárias e secundárias no local e na vizinhança (metodologia 03_DBN_PF_TB) produziu

modelos mais acurados e precisos. Além disto, as metodologias investigadas são recomendadas para dados com mesmo suporte e podem ser utilizadas em qualquer algoritmo de simulação estocástica. Para isto, é necessário somente o modelo de continuidade espacial da variável de interesse.

4.3 Limitações das metodologias propostas

Sobre os algoritmos de simulação estocástica investigados, apesar de não identificados, podemos destacar que as metodologias possuem as seguintes limitações:

i) Simulação por campos de probabilidade (*p-field*): Apesar deste algoritmo ser utilizado apenas para amostrar a distribuição de probabilidade no local da secundária (nos demais nós do grid usa-se *tbsim* ou *sgsim*), este método possui duas limitações: o aumento da continuidade espacial dos valores simulados quando comparados ao modelo da variável primária e gerar artefatos quando os nós a serem simulados estão próximos ao local dos dados condicionantes da variável primária. Portanto, é necessário avaliar estas situações.

ii) Simulação por bandas rotativas (*Turning Bands*): Apesar da eficiência computacional dos algoritmos espectrais. As metodologias investigadas poderão ser executadas no algoritmo de simulação sequencial Gaussiana (*SGS*). Mesmo assim, a metodologia proposta é recomendada somente para dados primários e secundários com o mesmo suporte, já que utiliza a transformação não linear para gerar dados normalizados.

iii) Simulação sequencial direta (*DSS*): O algoritmo permite utilizar dados de diferentes suportes e reproduzir melhor a continuidade espacial dos dados, mas a reprodução dos histogramas dos modelos obtidos deve ser avaliada e na literatura ferramentas já existe soluções implementadas.

Sobre a incorporação da informação secundária no processo de estimativa e simulação, a maioria das metodologias disponíveis baseiam-se no uso do modelo linear de correlogramização (*LMC*) que é laborioso. Em geral, os softwares comerciais disponíveis possuem as simplificações dos modelos de Markov (*MM1* e *MM2*), que impõe que a variável secundária seja exaustiva. Os modelos obtidos por estas metodologias algumas vezes possuem aumento da

variância. Portanto, integrar a informação secundária nos modelos pode ser um desafio. Assim, o uso das metodologias investigadas que sugere o uso apenas do modelo de continuidade espacial da variável primária, reproduz o histograma e variograma dos dados da variável de interesse pode ser relevante.

4.4 Sugestões para trabalhos Futuros

A partir da implementação das metodologias utilizadas nesta tese e dos resultados obtidos, surgem aspectos que podem ser analisados em trabalhos futuros.

A metodologia apresentada nesta tese, foi testada apenas em um banco de dados com variáveis contínuas, com apenas uma variável secundária com coeficiente de correlação forte (0,90) com a variável primária. Sugere-se avaliar esta metodologia para dados com coeficiente de correlação diferente e verificar a acuracidade e precisão dos modelos.

Em casos, que existem mais de uma variável secundária verificar a acuracidade e precisão destas metodologias:

- i) Verificar a possibilidade de construir as distribuições a partir de modelos misturas gaussianas (*Gaussian Mixture Models*),
- ii) Utilizar o conceito de supersecundária, que consiste em combinar todas as variáveis secundárias em apenas uma variável para construir a distribuição de probabilidade no local da variável secundária.

Sobre as metodologias investigadas no capítulo 02, as sugestões para trabalhos futuros são:

- i) Investigar outros algoritmos para inferir os *pseudohards*.
- ii) Implementar e avaliar as metodologias propostas no capítulo 03 no algoritmo de simulação sequencial direta.

As sugestões para trabalhos futuros para as metodologias investigadas no capítulo 03 são:

- i) Investigar os resultados das metodologias propostas utilizando os dados das variáveis primárias e secundárias originais, sem realizar a transformação dos dados.
- ii) Analisar os resultados do Método 04- Simulação estocástica utilizando inserção múltipla, utilizando o algoritmo de cokrigagem no local da secundária para construir a distribuição no algoritmo de atualização bayesiana (BU).

REFERÊNCIAS BIBLIOGRÁFICAS

ALABERT, F. G. **Stochastic imaging of spatial distributions using hard and soft information**. Master of Science, Department of Applied Earth Sciences, Stanford University, Stanford, California. 1987. 416 p.

ALMEIDA, A. S.; JOURNEL, A. G. Joint simulation of multiple variables with a Markov-type coregionalization model. **Mathematical Geology**, v. 26, n. 5, p. 565–588, 1994.

ARAÚJO, C. P.; COSTA, J. F. C. Integration of different-quality data in short-term mining planning. **Rem: Revista Escola de Minas**, v. 68, n. 2, p. 221–227, 2015.

ARAÚJO, C. P.; COSTA, J. F. C. L. KOPPE, V. C. Improving short-term grade block models: alternative for correcting soft data. **REM:International Engineering Journal**, v. 71, n. 1, p. 117-122, 2018.

AZEVEDO, L. “**Reunião sobre desenvolvimento do algoritmo de simulação sequencial direta com distribuições pontuais**”. Reunião realizada com o autor em outubro de 2018.

BABAK, O. **Theoretical and practical improvements to geostatistical estimation and simulation**. Doctor of Philosophy. Department of Civil and Environmental Engineering. University of Alberta, Edmonton, Canada. 2008. 212 p.

BABAK, O.; DEUSTCH, C.V. An intrinsic model of coregionalization that solves variance inflation in collocated cokriging. **Computers & Geosciences**, v.35, n 3, p.603-614,2009.

BAHAR, A.; KELKAR, M. Integrated lithofacies and petrophysical properties simulation. In: **SPE Western Regional Meeting**. Society of Petroleum Engineers.

BARNETT, R. M.; DEUTSCH, C. V. Multivariate imputation of unequally sampled geological variables. **Mathematical Geosciences**. V.47, n. 7, p. 791-817, 2015.

BEHRENS, R.A.; MACLEOD, M.K.; TRAN, T.T; ALIMÍ ,A.C. Incorporating seismic attribute maps in 3D reservoir models. In:**SPE Annual Technical Conference and Exhibition**. Society of Petroleum Engineers, v. 1, n. 1, p. 122-126,1998.

BOURGAULT, G. Using non-Gaussian distributions in geostatistical simulations. **Mathematical Geology**, v. 29, n. 3, p. 315–334, 1997.

CAERS, J. Adding local accuracy to direct sequential simulation. **Mathematical Geology**, v. 32, n. 7, p. 815–850, 2000a.

CAERS, J. Direct sequential indicator simulation. In: Kleingeld, W. J.; Krige, D. G. (Eds.) **Proceedings of 6th International Geostatistics Congress**, Cape Town, South Africa. 2000b. pp. 39-48.

CASSIRAGA, E. F. **Incorporacion de informacion blanda para la cuantificacion de la incertidumbre: aplicacion a la hidrogeologia** Tese de doutorado, Universidade Técnica de Valência, Espanha, 331 p, 1999.

CHILES, J. P.; DELFINER, P. **Geostatistics: modeling spatial uncertainty**. John Wiley e Sons, 1999. 659 p.

CORNAH, A.; MACHAKA, E. Integration of imprecise and biased data into mineral resource estimates. **Journal of the Southern African Institute of Mining and Metallurgy**, v.115, n. 6, p. 523-530, 2015.

DAVIS, B. M.; GREENES, K. A. Estimation using spatially distributed multivariate data: an example with coal quality. **Journal of the International Association for Mathematical Geology**, v. 15, n. 2, p. 287-300, 1983.

DERAISME, J.; STRYDOM, M. Estimation of iron ore resources integrating diamond and percussion drillholes. **Proceedings, 34th International Symposium on Computer Applications in the Mineral Industries (APCOM-2009)**, Vancouver, Canada. 2009. pp.1-8. CD-ROM.

DESBARATS, A. J.; DIMITRAKOPOULOS, R. Geostatistical simulation of regionalized pore-size distributions using min/max autocorrelation factors. **Mathematical Geology**, v. 32, n. 8, p. 919-942, 2000.

DEUTSCH, C. V.; JOURNEL, A.G. **GSLIB: Geostatistical software library and user's guide**, 2. ed.: Oxford University Press, New York, 1998.369 p.

DEUSTCH, C.V. Direct assessment of local accuracy and precision. **Geostatistics Wollongong**, v.96, p.115-125,1997.

DEUTSCH, C. V; TRAN, T;XIE,Y. A Preliminary report on: An Approach to Ensure Histogram reproduction in direct Sequential Simulation. **Proceedings, Report 03–2001**, Centre for Computational Geoestatics, University of Alberta, Edmonton, Canadá, Paper 2001-103. 103 (1) - 103 (19) p.

DONOVAN, P.N.; DEUTSCH, C.V. Cokriging with Multiple Data Types – Theory. **Proceedings, Report 16–2014**, Centre for Computational Geoestatics, University of Alberta, Edmonton, Canada, Paper 2014-313. 2014. p.313 (1)-313(11).

DOYEN, P.M.; DEN BOER, L.D.; PILLET, W.R. Seismic porosity mapping in the Ekofisk field using a new form of collocated cokriging. In: **SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers**, 1996.

EFRON, B. **The jackknife, the bootstrap, and other resampling plans** Society for Industrial and applied Mathematics Philadelphia, 1982, 91p.

EMERY, X. Disjunctive kriging with hard and imprecise data. **Mathematical Geology**, v. 35, n. 6, p. 699–718, 2003.

EMERY, X. Geology Resource and reserve evaluation in the presence of imprecise data. **CIM Bulletin**, v. 90, n. 1089, p. 366–377, 2005.

EMERY, X.; LANTUÉJOU, C. Tbsim: A computer program for conditional simulation of three-dimensional Gaussian random fields via the turning bands method. **Computers & Geosciences**, v. 32, n. 10, p. 1615-1628, 2006.

FICHTL, P.; FOURNIER, F.; ROYER, J. J. Cosimulations of lithofacies and associated reservoir properties using well and seismic data. In: **SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers**, 1997.

FROIDEVAUX R. Probability field simulation. **Proceedings, Geostatistics Troia'92**. Springer Netherlands, 1992. p. 73-83.

GY, P. M. **Sampling of heterogeneous and dynamic material systems: theories of heterogeneity, sampling and homogenizing**. Vol. 10. Elsevier, 1992. 642 p.

GODOY, M. The effective management of geological risk in long-term production scheduling of open pit mines. Tese de Doutorado, Escola de Engenharia, Universidade Queensland, Brisbane, Austrália. 2002.

GÓMEZ-HERNÁNDEZ, J. J.; JOURNEL, A. G. Joint sequential simulation of multigaussian fields. **Proceedings, Geostatistics Troia'92**. Springer Netherlands, 1992. p. 85-94.

GOOVAERTS, P. **Geostatistics for Natural Resources Evaluation**. New York: Oxford University Press. 1997. 483 p.

GOOVAERTS, P. Geostatistical modelling of uncertainty in soil science. **Geoderma**, v.103, n (1-2), p.3-26, 2001.

GRAHAM, J.W.; OLCZOWSKI, E.A.; TAMIKA D.; GILREATH, T.D. How many imputations are really needed? Some practical clarifications of multiple imputation theory. **Prevention science**, v.8, n. 3, p. 206-213, 2007.

HORTA, A.; SOARES, A. Direct sequential Co-simulation with joint probability distributions. **Mathematical Geosciences**, v. 42, n. 3, p. 269–292, 2010.

ISAAKS, E. H.; SRIVASTAVA, R. M. **An Introduction to Applied Geostatistics**. New York: Oxford University Press, 1990.561 p.

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

JOHANN, P.; FOURNIER, F.; SOUZA, O.; JR.; DOLIGEZ, B. 3-D stochastic reservoir modeling: A non-stationary approach guided by seismic data: SPE Paper 38997, 1997.

JOHNSON, R. A. *et al.* **Applied multivariate statistical analysis**. Upper Saddle River, NJ: Prentice hall, 2002.

JOURNEL, A. G. Geostatistics for conditional simulation of orebodies: **Economic Geology**, v. 69, p. 673–680, 1974.

JOURNEL, A. G.; HUIJBREGTS, C. H. **Mining Geostatistics**. London: Academic Press, 1978.600 p.

JOURNEL, A. G. The indicator approach to estimation of spatial distributions. **Proceedings of the 17th APCOM International Symposium, New York**. 1982. p. 793-806.

JOURNEL, A. G.; ISAAKS, E.H. Conditional indicator simulation: application to a Saskatchewan uranium deposit. **Journal of the International Association for Mathematical Geology**,v.16, n.7, p.685-718, 1984.

JOURNEL A. G. Geostatistics: models and tools for the earth sciences. **Mathematical geology**, v. 18, n. 1, p. 119-140, 1986.

JOURNEL, A. G. Modeling uncertainty: some conceptual thoughts. **Proceedings: Geostatistics for the next century**. Springer Netherlands, 1994. p. 30-43.

JOURNEL, A. G. Markov models for cross-covariances. **Mathematical Geology**, v. 31, n. 8, p. 955-964, 1999.

JOURNEL, A.G.; Z, YING. "The theoretical links between sequential Gaussian simulation, Gaussian truncated simulation, and probability field simulation." **Mathematical Geology**, v.33, n. 1, p. 31-40, 2001.

KUMAR, A.; DEUTSCH, C. V. Optimal correction of indefinite correlation matrices. **Proceedings, Report 11–2009**, Centre for Computational

Geostatistics, University of Alberta, Edmonton, Canada, Paper 2009-401. 2009. p.401 (1)-401(8).

LIU, Y.; JOURNEL, A.G. A package for geostatistical integration of coarse and fine scale data. **Computers & Geosciences**, vol.35, n. 3, p.527-547, 2009.

MARECHAL, A. **Cokrigeage et regression en correlation intrinsique**. France, Fontainebleau: Centre de Geostatistique de Fontainebleau, 1970. 40p.

MATHERON, G. Principles of geostatistics, economic geology. **Economic Geology**, v. 58, n. 8, p. 1246-1266, 1963.

MATHERON, G. Les fonctions de transfert des petits panneaux. **Note du Centre de Géostatistique, Ecole des Mines de Paris**, 1974.

MATHERON, G. The intrinsic random functions and their applications. **The Advances in Applied Probability**, v. 5, n. 3, p. 439–468, 1973.

MATHERON, G. La destructure des hautes teneurs et le krigeage des indicatrices. **Ecole des Mines de Paris**, 1982.

MATHERON, G.; BEUCHER, H.; DE FOUQUET, H.; GALLI, A.; GUERILLOT, D.; RAVENNE, C., Conditional simulation of the geometry of fluvio-deltaic reservoirs: SPE Paper 16753, 1987.

MINNITT, R. C. A.; DEUTSCH, C. V. Cokriging for optimal mineral resource estimates in mining operations. **The Journal of The Southern African Institute of Mining and Metallurgy**, v. 114, n. 3, p. 189–203, 2014.

NEVES, J.; PEREIRA, M. J.; PACHECO, N.; SOARES, A. Updating Mining Resources with Uncertain Data. **Mathematical Geosciences**, Special Issue, p. 1–20, 2018.

NOWAK, M. S.; SRIVASTAVA, R. M. A geostatistical conditional simulation algorithm that exactly honours a predefined grade-tonnage curve. **Geostatistics Wollongong**, v. 96, p. 669-682, 1997.

OWEN, M. L.; MEYER, L. NI 43-101 technical report for Neves-Corvo Mine and Semblana Deposit, Portugal. Wardell Armstrong International Ltd., Truro. 2013.

OSTERHOLT, V.; BENNDORF, J. Real-time mining process flow analysis. **Published document of European Programme Horizon 2020 funded project Real-Time Mining**. Published under www.realtimemining.eu on 05/11/2015

OZ, B.; DEUTSCH, C. V.; TRAN, T. T.; XIE, Y. DSSIM-HR: A FORTRAN 90 program for direct sequential simulation with histogram reproduction. **Computers & Geosciences**, v. 29, n. 1, p. 39-51, 2003.

REN, W. **Exact downscaling in reservoir modeling**. Tese de Doutorado, University of Alberta, 2007, 213p

PYRCZ, M. J.; DEUTSCH, C. V. "Two artifacts of probability field simulation. **Mathematical Geology**, v.33, n. 7, p.775-799, 2001.

RIBEIRO, D.; MONTEIRO, C.; CUNHA, E.; CATARINO, M. Correlation study between reverse circulation and diamond drilling in iron ore deposits. **Proceedings of the 5 th WCBS Conference on Sampling and Blending**, Santiago, Chile. 2011 p. 277-285. (CD-ROM).

ROSSI , M. E.; DEUTSCH, C. V. **Mineral resource estimation**. New York: Springer, 2014.332 p.

SILVA, C. Z.. **A descorrelação de variáveis com fatorização maf em estimativas de teores**. Dissertação de Mestrado, Programa de Pós-Graduação em Engenharia de Minas, Metalúrgica e Materiais, Universidade Federal do Rio Grande do Sul, Brasil.2013. 123 p.

SILVA, D.S.; DEUTSCH, C.V. Multivariate data imputation using gaussian mixture models. **Spatial Statistics**, v.27, pp.74-90, 2018.

SOARES, A. Direct sequential simulation and cosimulation. **Mathematical Geology**, v.33, n.8, p. 911-926, 2001.

SOARES, A. Classification of mining reserves using direct sequential simulation. **Proceedings Geostatistics Banff 2004**. Springer Netherlands, 2005. p. 511-522.

SOARES, A.; NUNES, R.; AZEVEDO, L. Integration of uncertain data in geostatistical modelling. **Mathematical Geosciences**, v.49, n.2, p. 253-273, 2017.

SOUZA, L. **Proposição geoestatística para quantificação do erro em estimativas de tonelagens e teores**. Tese de Doutorado, Programa de Pós-Graduação em Engenharia de Minas, Metalúrgica e Materiais, Universidade Federal do Rio Grande do Sul, Brasil. 2007. 193 p.

SRIVASTAVA, R. M. Reservoir characterization with probability field simulation. **Proceedings of SPE Annual Technical Conference and Exhibition 1992 Janeiro**. Society of Petroleum Engineers.

TRAN, T.; DEUTSCH, C. V.; XIE, Y.T. Direct geostatistical simulation with multiscale well, seismic, and production data. In: **SPE Annual Technical Conference and Exhibition**. Society of Petroleum Engineers, 2001.

VAN DER CORPUT, J.G. Verteilungsfunktionen (Erste Mitteilung), **Proceedings of the Koninklijke Akademie van Wetenschappen te Amsterdam (in German)**: p.813–821,1935.

VERLY, G. W. **Estimation of spatial point and block distributions: The Multigaussian model**. Tese de Doutorado, Stanford University, 416p, 1984

VERLY, G. W. Sequential Gaussian cosimulation: a simulation method integrating several types of information. In: **Geostatistics Troia'92**. Springer Netherlands, 1992. p. 543-554

WACKERNAGEL, H. **Multivariate geostatistics an introduction with applications**. 2 ed Springer-Verlag Berlin Heidelberg GmbH, 1996.291 p.

WACKERNAGEL, H. **Multivariate geostatistics an introduction with applications**. 3 ed Springer-Verlag Berlin Heidelberg GmbH, 2003.387 p.

XU, W. TRAN, T.T; SRIVASTAVA, R.M; JOURNEL, A.G. Integrating seismic data in reservoir modeling: the collocated cokriging alternative. **Proceedings: SPE annual technical conference and exhibition**. Society of Petroleum Engineers, 1992. 833-842 p.

YAO, T.; JOURNEL, A.G. 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**. v. 28, p. 65-79,2000.

ZANON, S.D; DEUTSCH, C.V. Direct prediction of reservoir performance with Bayesian updating under a multivariate Gaussian model. In: **Canadian International Petroleum Conference**. Petroleum Society of Canada, 2004.

ZHU, H. **Modeling Mixture of Spatial Distributions with Integration of Soft Data**, Tese de doutorado, Department of Applied Earth Sciences, Stanford University, 1991, p.310

ZHU, H.; JOURNEL, A. G. Formating and integrating soft data: Stochastic imaging via the Markov-Bayes Algorithm. **Geostatistics-Troia**, v. 1, p. 1-12, 1992.

XU, W., JOURNEL, A.G., GTSIM: Gaussian truncated simulations of reservoir units in a West Texas carbonate field. paper SPE,27412, 1993