

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE MINAS,
METALÚRGICA E DE MATERIAIS (PPGE3M)**

RAFAEL MONIZ CAIXETA

**SIMULAÇÃO GEOESTATÍSTICA UTILIZANDO MÚLTIPLOS PASSEIOS
ALEATÓRIOS**

Porto Alegre

2015

RAFAEL MONIZ CAIXETA

**SIMULAÇÃO GEOESTATÍSTICA UTILIZANDO MÚLTIPLOS PASSEIOS
ALEATÓRIOS**

Dissertação submetida ao Programa de Pós-Graduação em Engenharia de Minas, Metalúrgica e de Materiais da Universidade Federal do Rio Grande do Sul, como requisito parcial à obtenção do título de Mestre em Engenharia, modalidade Acadêmica. Área de Concentração: Tecnologia Mineral

Orientador: Prof. Dr. João Felipe Coimbra Leite Costa

Porto Alegre

2015

[Verso]

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Carlos Alexandre Netto

Vice-Reitor: Rui Vicente Oppermann

ESCOLA DE ENGENHARIA

Diretor: Luiz Carlos Pinto da Silva filho

Vice-Diretora: Carla Schwengber ten Caten

PROGRAMA DE PÓS-GRADUAÇÃO

Coordenador: Carlos Pérez Bergmann

[Ficha Catalográfica]

RAFAEL MONIZ CAIXETA

**SIMULAÇÃO GEOESTATÍSTICA UTILIZANDO MÚLTIPLOS PASSEIOS
ALEATÓRIOS**

Esta dissertação foi analisada e julgada adequada para a obtenção do título de Mestre em Engenharia, área de concentração de Tecnologia Mineral, e aprovada em sua forma final pelo Orientador e pela Banca Examinadora designada pelo Programa de Pós-Graduação em Engenharia de Minas, Metalúrgica e de Materiais da Universidade Federal do Rio Grande do Sul.

Prof. Dr. João Felipe Coimbra Leite Costa

Orientador

Prof. Dr. Carlos Pérez Bergmann

Coordenador do PPGE3M

Aprovado em: 13/11/2015

BANCA EXAMINADORA:

Dr. Diniz Tamantini Ribeiro – Vale S/A

Prof. Dr. Rodrigo de Lemos Peroni – DEMIN/UFRGS

Prof^a. Dr^a. Vanessa Cerqueira Koppe – DEMIN/UFRGS

À minha família

AGRADECIMENTOS

Ao Prof. Dr. João Felipe C.L. Costa, pela recepção no laboratório e por toda a orientação e formação passada ao longo desse período de mestrado.

Ao Dr. Diniz T. Ribeiro, pelas conversas e toda contribuição no desenvolvimento dessa dissertação.

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

À Vale S/A, pelo apoio e fornecimento dos dados para essa dissertação

A todos os amigos e colegas que fiz nesse período em Porto Alegre, pela ajuda, companheirismo e bons momentos.

À minha família, pelo apoio incondicional, seja perto ou à distância.

*To see a world in a grain of sand
And a heaven in a wild flower
Hold infinity in the palm of your hand
And eternity in an hour.*

William Blake

RESUMO

Simulação geoestatística compreende uma variedade de técnicas que permitem gerar cenários que reproduzem a continuidade espacial e o histograma do fenômeno de interesse. Essas técnicas podem ser usadas para ajudar nas tomadas de decisões, permitindo um acesso à incerteza nas funções de resposta (que dependem dos cenários simulados), geralmente por meio de uma relação não-linear (retorno financeiro, recuperação geometalúrgica...). No entanto, uma de suas limitações é que as simulações podem demandar um tempo considerável para serem executadas em grandes depósitos. E a motivação dessa dissertação se concentra justamente nesse fato, buscando uma alternativa para acelerar o processamento computacional dessas simulações. A opção escolhida para isso foi desenvolver a Simulação via Múltiplos Passeios Aleatórios, que é uma nova abordagem para se realizar simulações geoestatística. Ela combina a krigagem com a simulação de passeios aleatórios independentes, de modo a gerar cenários simulados de uma maneira mais rápida que os algoritmos tradicionais. Essa dissertação apresenta detalhes do método e importantes contribuições desenvolvidas para melhorar o desempenho e a qualidade dos resultados gerados com o método. Foi também desenvolvido um software específico para possibilitar um uso simples, prático e rápido da técnica em qualquer situação (2D ou 3D). Estudos de caso foram feitos para checar a validade das simulações, que demonstraram boa reprodução dos histogramas e variogramas, além de um ganho de velocidade considerável, alcançando uma aceleração de até 5,65 x (em comparação com a Simulação por Bandas Rotativas) na simulação de um depósito de ferro em 3D, desempenho que pode ser melhor ainda à medida que mais dados amostrais estão disponíveis.

Palavras-chave: Geoestatística. Simulação condicional. Mineração. Paralelização.

ABSTRACT

Geostatistical simulation comprises a variety of techniques, which allow the generation of multiple scenarios reproducing the spatial continuity and the histogram of the desired phenomenon (grades for instance). These methods can be used in the decision making process, allowing the assess to the uncertainty of functions responses (which depend on the simulated inputs) commonly through a non-linear relationship (net present value, interest tax return, ore geometallurgical recovery...). However, one of their limitations is that running simulation can take a considerable processing time to be executed in large deposits or large grids. Therefore, the motivation of this dissertation focuses on this fact, leading to the main goal, i.e. investigating an alternative to accelerate the simulation process. The option chosen is based on the development and adaptation of the Multiple Random Walk Simulation, which is algorithm to build geostatistical simulations. It combines kriging with the simulation of independent random walks in order to generate simulated scenarios faster than via traditional simulation algorithms. This dissertation presents details of the method and new important contributions developed to improve its performance and statistics reproduction. An algorithm and software was also presented in order to allow a simple, practical and fast use of the method in any condition (2D or 3D). Case studies were developed to check the validity of the simulations, which showed good reproduction of histogram and variograms, in addition to a considerable speed gain, achieving an acceleration up to 5.65 x (in comparison with Turning Bands Simulation) in the simulation of a 3D iron deposit, performance that can be enhanced as more conditioning samples are available.

Keywords: Geostatistics. Conditional simulation. Mining. Parallelization.

LISTA DE ILUSTRAÇÕES

Figura 1 - Transformação dos dados via <i>normal scores</i> de uma distribuição acumulada F dos dados originais para uma distribuição gaussiana padrão acumulada G.....	18
Figura 2 - Duas simulações utilizando Simulação por Bandas Rotativas. À esquerda, com apenas 15 linhas e, à direita, com 1000 linhas.....	18
Figura 3 - Exemplo ilustrativo de como se obter o valor simulado do ponto $Y_s(u)$ a partir de 8 linhas simuladas ($D_1...D_8$).....	20
Figura 4 - Processo de obtenção de um valor simulado a partir da geração de um número aleatório entre $[0,1]$	21
Figura 5 - Exemplo de um gráfico de acuracidade.....	25
Figura 6 - Ilustração dos passeios aleatórios e suas propriedades Gaussianas. Começando em 0, um incremento positivo ou negativo é adicionado randomicamente a cada separação. Para um conjunto de passeios aleatórios, seus valores a qualquer distância h apresentam uma distribuição gaussiana de média 0 e cuja variância cresce proporcionalmente com o aumento de h.	28
Figura 7 - Exemplo ilustrativo de duas realizações simuladas via múltiplos passeios aleatórios. Foi simulado o ponto $Y_s^{(1)}$ a partir de duas amostras y_1 e y_2 . O resultado da primeira realização é apresentado na parte superior da figura e o da segunda realização logo abaixo.	30
Figura 8 - Calibragem do valor do incremento f para que realizações apresentem variância unitária.....	31
Figura 9 - Um exemplo de como um banco de dados (vermelho) é populado com pontos auxiliares (azul), considerando um raio mínimo de 1/3 da distância amostral original.	33
Figura 10 - Interface do plugin para Simulação via Múltiplos Passeios Aleatórios desenvolvido para SGeMS.....	36
Figura 11 - À esquerda, o mapa do modelo de referência. À direita, o semivariograma desses dados.....	39
Figura 12 - Detalhes visuais em uma janela de 500 por 500 m do modelo simulado por SMPA em comparação com o modelo de referência.	41
Figura 13 - Reprodução do semivariograma (à esquerda) e do histograma (à direita) para as simulações obtidas por SMPA.....	41
Figura 14 - Curvas teor-tonelagem das simulações via SMPA e do modelo de referência.	41

Figura 15 - Gráfico de acuracidade para cada conjunto de simulações comparados aos dados de referência. Os intervalos de probabilidade são mais precisos quanto mais próximos os pontos estiverem da reta de 45° e acurados quando sobre ou acima da mesma reta.	42
Figura 16 - Detalhes visuais em uma janela de 500 por 500 m dos dois modelos simulados por SMPA em comparação com o modelo de referência.	43
Figura 17 - Reprodução do semivariograma (à esquerda) e do histograma (à direita) para as simulações obtidas por SMPA(1).....	44
Figura 18 - Curvas teor-tonelagem das simulações via SMPA(1) e do modelo de referência.	44
Figura 19 - Reprodução do semivariograma (à esquerda) e do histograma (à direita) para as simulações obtidas por SMPA(2).	45
Figura 20 - Curvas teor-tonelagem das simulações via SMPA(2) e do modelo de referência.	45
Figura 21 - Gráfico de acuracidade para cada conjunto de simulações comparados aos dados de referência. SMPA(1) e (2) geraram os mesmos resultados e foram representados unicamente como SMPA. Os intervalos de probabilidade são mais precisos quanto mais próximos os pontos estiverem da reta de 45° e acurados quando sobre ou acima da mesma reta	46
Figura 22 - Mapa regional do Quadrilátero Ferrífero.....	47
Figura 23 - Coluna estratigráfica mostrando, em vermelho, a localização do minério de ferro do Quadrilátero Ferrífero.....	48
Figura 24 - Mapa de locação da boca dos furos de sondagem e uma vista isométrica.	50
Figura 25 - Comprimentos das amostras regularizadas.....	51
Figura 26 - Histogramas de cada um dos quatro domínios após o desagrupamento amostral.	52
Figura 27 - Uma realização simulada do modelo completo, exibido em perspectiva.	54
Figura 28 - Seção horizontal (na cota de 875 m) de uma realização simulada, destacando as partes correspondentes a cada um dos domínios.	55
Figura 29 - Seções verticais de uma realização simulada, destacando as partes correspondentes a cada um dos domínios. Vista em perspectiva isométrica.	55
Figura 30 - Histogramas e estatísticas básicas das simulações de cada domínio.	56
Figura 31 - Variogramas horizontais das simulações de cada domínio.	57
Figura 32 - Variogramas verticais das simulações de cada domínio.....	57
Figura 33 - Gráficos de acuracidade referentes às simulações de cada domínio.	58
Figura 34 - Curvas de teor-tonelagem referentes às simulações de cada domínio.....	59

LISTA DE TABELAS

Tabela 1 - Detalhes do modelo de referência simulado.....	39
Tabela 2 - Tempo decorrido para realizar 100 simulações via cada um dos métodos. Utilizado 20 processadores de um Intel(R) Xeon(R) CPU E5-2630 0 @ 2.30GHz 2x Hexacore.	43
Tabela 3 - Tempo decorrido para realizar 100 simulações via cada um dos métodos. Utilizado 20 processadores de um Intel(R) Xeon(R) CPU E5-2630 0 @ 2.30GHz 2x Hexacore	46
Tabela 4 - Detalhes da quantidade de pontos auxiliares simulados e do total de nós simulados.	53
Tabela 5 - Detalhes dos parâmetros variográficos utilizados para as simulações.	54
Tabela 6 - Tempo decorrido para realizar 100 simulações via cada um dos métodos. Utilizados 20 processadores de um SGI H2106 Server (4x AMD Opteron 6376 16-core 2.3GHZ, 256GB RAM Mem, 2x 1TB Disk).....	60
Tabela 7 - Tempo decorrido para realizar 100 simulações via cada um dos métodos. Utilizados 50 processadores de um SGI H2106 Server (4x AMD Opteron 6376 16-core 2.3GHZ, 256GB RAM Mem, 2x 1TB Disk).....	60

LISTA DE ABREVIATURAS E SIGLAS

1D	Uma dimensão
2D	Duas dimensões
3D	Três dimensões
GWN	Ruído branco gaussiano (do inglês: <i>Gaussian White Noise</i>)
RAM	Memória de acesso aleatório (do inglês: <i>Random Access Memory</i>)
RW	Passeio aleatório (do inglês: <i>Random Walk</i>)
SBR	Simulação por Bandas Rotativas
SMPA	Simulação via Múltiplos Passeios Aleatórios
SSG	Simulação Sequencial Gaussiana

SUMÁRIO

1.	INTRODUÇÃO	13
1.1.	META DA DISSERTAÇÃO	14
1.2.	OBJETIVOS ESPECÍFICOS DA DISSERTAÇÃO	14
1.3.	METODOLOGIA	15
1.4.	ORGANIZAÇÃO DA DISSERTAÇÃO	15
2.	MÉTODOS CONSOLIDADOS DE SIMULAÇÃO GEOESTATÍSTICA	17
2.1.	SIMULAÇÃO POR BANDAS ROTATIVAS	17
2.2.	SIMULAÇÃO SEQUENCIAL GAUSSIANA.....	20
2.3.	VALIDAÇÕES DAS SIMULAÇÕES	24
3.	SIMULAÇÃO VIA MÚLTIPLOS PASSEIOS ALEATÓRIOS	26
3.1.	TRABALHOS ANTERIORES E NOVAS CONTRIBUIÇÕES	26
3.2.	INTRODUÇÃO TEÓRICA.....	26
3.3.	PASSEIOS ALEATÓRIOS	27
3.4.	UTILIZANDO PASSEIOS ALEATÓRIOS PARA GERAR SIMULAÇÕES GEOESTATÍSTICA CONDICIONAIS	28
3.5.	REPRODUÇÃO DO EFEITO PEPITA E CALIBRAGEM DA VARIÂNCIA	30
3.6.	SIMULANDO EM CASOS COM POUCOS DADOS AMOSTRAIS.....	32
3.7.	CONDIÇÕES MÍNIMAS DE ADENSAMENTO AMOSTRAL	33
3.8.	ALGORITMO E PLUGIN DESENVOLVIDOS	34
4.	APLICAÇÃO EM ESTUDO DE CASO CONTROLADO	39
4.1.	1ª ETAPA.....	40
4.2.	2ª ETAPA.....	43
5.	APLICAÇÃO EM ESTUDO DE CASO REAL	47
5.1.	GEOLOGIA REGIONAL	47
5.2.	ANÁLISE PRELIMINAR DOS DADOS	49
5.3.	SIMULAÇÃO DO DEPÓSITO.....	53
6.	CONSIDERAÇÕES FINAIS	62
6.1.	CONCLUSÕES	62
6.2.	RECOMENDAÇÕES DE TRABALHOS FUTUROS	62
	REFERÊNCIAS	64
	ANEXO A – ANÁLISE DA SENSIBILIDADE DA SIMULAÇÃO VIA MÚLTIPLOS PASSEIOS ALEATÓRIOS ÀS DISTÂNCIAS DAS AMOSTRAS AOS PONTOS SIMULADOS	67

1. INTRODUÇÃO

O aparecimento de métodos de simulação geoestatística datam do final do século passado. As primeiras publicações sobre o assunto são dos anos 1970 na França, com a introdução da Simulação de Bandas Rotativas (MATHERON 1973, JOURNAL 1974). Entretanto, a aplicação desse e outros métodos sempre foi restrita a modelos menores por problemas técnicos referentes ao desempenho computacional. Por exemplo, há alguns anos as simulações de grandes depósitos minerais com milhões de nós de grid poderiam demandar vários dias de processamento, além das subsequentes validações do modelo, que também demandam um tempo considerável.

Este cenário favoreceu o uso de simulações inicialmente em tarefas de modelagem geológica e planejamento de lavra de longo prazo, propiciando diversos benefícios práticos. Além do acesso ao valor médio local (aproximadamente equivalente à estimativa via krigagem), as simulações permitem um acesso à incerteza em todos os locais simulados, ao mesmo tempo que respeita a distribuição global e a reprodução da continuidade espacial do fenômeno.

Dessa maneira, é possível aplicar funções de transferências não-lineares sobre os cenários simulados para derivar qualquer outro parâmetro de interesse: resposta geometalúrgica (COWARD E DOWD, 2015); retorno financeiro após a simulação de teores (MELLO, 2015; CHERCHENEVSKI, 2015); produção de óleo e gás após simulação da permeabilidade (PYRCZ E DEUSTCH, 2015, p. 9), entre outros. E, em posse dessas informações, é possível realizar um planejamento muito mais fiel sobre o empreendimento mineiro, petrolífero, ou seja qual for.

Porém, o uso das simulações em tarefas mais rotineiras na indústria mineral ainda é mínimo. No planejamento mineiro de curto-prazo poderia, por exemplo, ser usado para se obter melhor conhecimento sobre a variabilidade dos teores dos blocos e melhorar a classificação de recursos (ROSSI E DEUTSCH, 2014). E entre os principais motivos para os métodos de simulação não serem usados nessas tarefas mais rotineiras é a questão do tempo de processamento dos modelos usando os diversos métodos disponíveis. Embora hoje em dia as máquinas de alta performance sejam mais acessíveis e uma boa infraestrutura computacional possa tornar as simulações significativamente mais rápidas ao utilizar múltiplos processadores, são escassos os algoritmos de simulação escritos para ter vantagem com essa arquitetura de *hardware*.

Simulação Sequencial Gaussiana (ISAAKS, 1990) e Simulação Sequencial Direta (SOARES, 2001) são pouco paralelizáveis devido à natureza sequencial desses algoritmos. A Simulação por Bandas Rotativas tende a apresentar um melhor desempenho em ambientes de multiprocessadores quando bem implementado – Isatis® (GEOVARIANCES, 2014). Métodos de simulação espectral (PARDO-IGUZQUIZA E CHICA-OLMO, 1993; YAO, 1998) prometem rápidos resultados nesses ambientes, mas não se sabe de um software disponível com essa função paralelizada para casos 3D.

Essa escassez de algoritmos, combinada com o alto custo de licenças comerciais de *softwares* e também com uma deficiência na formação de recursos humanos para utilização dessas técnicas, impediram uma maior propagação do uso de simulações em tarefas de médio a curto-prazo. Esse contexto motivou novos desenvolvimentos para melhorar o desempenho de algoritmos de simulação geoestatística. Mariethoz (2010) discute algumas estratégias de como paralelizar algoritmos sequenciais e Rasera *et al.* (2015) apresentam uma eficiente estratégia para paralelizar algoritmos sequenciais a nível de caminho.

Em busca de algoritmos mais rápidos para facilitar seu uso na indústria mineral foi proposto um novo método designado por Simulação via Múltiplos Passeios Aleatórios (SMPA). É uma nova alternativa desenvolvida para tentar tornar simulações condicionais 3D mais viáveis em aplicações rotineiras. Ribeiro *et al.* (2012) apresentaram a ideia do método e novos aprimoramentos foram feitos desde então (RIBEIRO *ET AL.*, 2013; CAIXETA *ET AL.*, 2015).

A maior potencialidade dessa nova técnica é que pode ser altamente paralelizada em termos de processamento computacional, permitindo que os processos sejam individualizados em nível de nó. Em termos práticos, ela apresenta também um grande potencial para atualizar localmente o *grid* simulado após incorporação de novas amostras, sem ter que processar o *grid* completo do depósito.

1.1. META DA DISSERTAÇÃO

A meta dessa dissertação de mestrado é desenvolver e operacionalizar a Simulação via Múltiplos Passeios Aleatórios para permitir seu uso na indústria mineira de maneira simples e rápida.

1.2. OBJETIVOS ESPECÍFICOS DA DISSERTAÇÃO

Para atingir tal meta, alguns objetivos específicos foram traçados.

- (i) Pretende-se estudar alternativas para melhorar a reprodução espacial da variabilidade à curta distância das simulações (efeito pepita) e criar uma maneira de evitar a geração de artefatos visuais e reprodução espacial deficiente das simulações em situações com pouca densidade amostral.
- (ii) Desenvolver um algoritmo que funcione de maneira rápida e eficiente para grandes depósitos e que, além disso, seja de uso simples e intuitivo para qualquer tipo de usuário.
- (iii) Por fim, realizar estudos de casos para verificar e mensurar a qualidade geoestatística e eficiência computacional do método em comparação com métodos já consagrados de simulação geoestatística.

1.3. METODOLOGIA

Foi feita inicialmente uma revisão bibliográfica sobre todos os principais métodos de simulação geoestatística de variáveis contínuas de modo a buscar ideias e soluções para aprimorar a parte teórica do método, segundo os objetivos específicos apresentados anteriormente.

Para tornar o método funcional, um algoritmo foi escrito em C++ para operar eficientemente em um ambiente de múltiplos processadores. Ele foi criado para ser utilizado como *plugin* do SGeMS (REMY ET AL., 2009), um software geoestatístico de código aberto, permitindo assim o uso de todas suas funcionalidades gráficas e (geo)estatísticas.

Para analisar a aplicação do método em condições controladas e reais, dois estudos de casos foram desenvolvidos. O primeiro com um banco de dados sintético e o segundo com um depósito real de ferro. Em ambos os casos, foi feita a aplicação da SMPA e seus resultados foram comparados em termos de qualidade e velocidade com os resultados obtidos via os outros dois principais métodos de simulação disponíveis: Simulação por Bandas Rotativas e Simulação Sequencial Gaussiana.

1.4. ORGANIZAÇÃO DA DISSERTAÇÃO

Os capítulos seguintes estão organizados de modo a melhor apresentar os resultados obtidos com esta dissertação.

O capítulo 2 revisa os dois métodos mais tradicionais de simulação geoestatística: Simulação por Bandas Rotativas e Simulação Sequencial Gaussiana; além das maneiras de como se avaliar a qualidade das simulações.

O capítulo 3 apresenta a Simulação via Múltiplos Passeios Aleatórios, com o contexto teórico baseado nos trabalhos iniciais sobre o método e aprimorado com os resultados dessa dissertação. Neste capítulo, foi também detalhado o algoritmo desenvolvido para testes e aplicações complexas.

O capítulo 4 expõe o estudo de caso em ambiente controlado a partir de um banco de dados sintético em 2D e demonstra o funcionamento do método em condições mais simples.

O capítulo 5 inclui o estudo de caso em ambiente real, aplicado em um depósito de ferro com informações de sondagens de longo-prazo e de curto-prazo, totalizando quase 10.000 amostras.

O capítulo 6 encerra a dissertação reunindo as conclusões obtidas ao longo do trabalho e apresentando o que pode ser ainda desenvolvido como trabalhos futuros relacionados ao tópico.

2. MÉTODOS CONSOLIDADOS DE SIMULAÇÃO GEOESTATÍSTICA

Os métodos de estimativa baseados no teorema de krigagem (MATHERON, 1963) são adequados e eficientes para se determinar o valor médio esperado dos blocos utilizando dados amostrais e seu modelo de continuidade espacial. No entanto, quando o objetivo é reproduzir as estatísticas globais e espaciais do depósito, a krigagem não é uma boa ferramenta devido ao seu inerente efeito de suavização.

Para o caso de ser necessário a reprodução dessas características globais, os métodos de simulação são os mais indicados. A partir deles são geradas uma variedade de imagens equiprováveis da variável desejada, das quais os valores e incertezas associadas podem ser utilizados para diversos fins, como otimização de planos de produção, sequenciamento de lavra, classificação de recursos, estratégias de blendagem, entre outros.

Há diversos métodos de simulação geoestatística desenvolvidos na literatura, cada qual com uma abordagem diferente – métodos paramétricos gaussianos ou não paramétricos; condicionado sequencialmente ou posteriormente adicionando um resíduo a um valor krigado. Uma lista com diversos métodos de simulação geoestatística pode ser encontrada em Chilès e Delfiner (1999) e também em Lantuéjoul (2002). Aqui são apresentadas as duas técnicas mais populares: Simulação por Bandas Rotativas (MATHERON, 1973; JOURNEL, 1974) e Simulação Sequencial Gaussiana (ISAAKS, 1990).

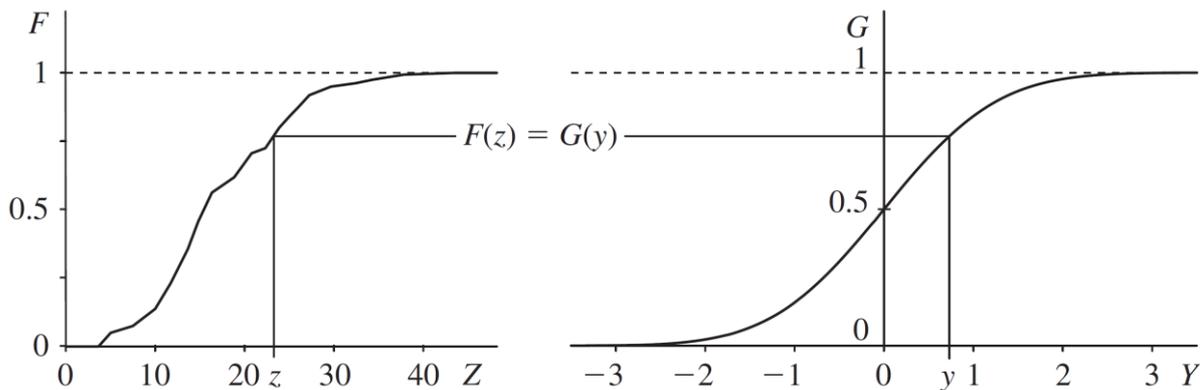
2.1.SIMULAÇÃO POR BANDAS ROTATIVAS

A Simulação por Bandas Rotativas foi o primeiro método desenvolvido para simulação geoestatística em 3D. Ela se baseia na simulação de diversas linhas 1D com a covariância estipulada e intrapola os valores dessas linhas no espaço 2D ou 3D para gerar simulações não-condicionais pontuais. O condicionamento é feito posteriormente via krigagem simples dos resíduos.

O primeiro passo do processo é a normalização dos dados. Isso é necessário tanto para viabilizar as simulações 1D, como para reproduzir o histograma de referência ao final do processo. A forma mais simples de se realizá-la é por *normal scores*, que se trata de uma transformação não-linear que associa os valores referentes dos quantis da distribuição dos dados originais com os quantis da distribuição gaussiana padrão de média 0 e variância 1 (Figura 1). Essa transformação deve levar em conta a distribuição desagrupada e, quando houver amostras

com valores idênticos, é necessário transformá-las adequadamente com as técnicas de *despiking* ou desempate (DEUTSCH E JOURNAL, 1998; VERLY, 1984) para evitar artefatos na distribuição gaussiana gerada.

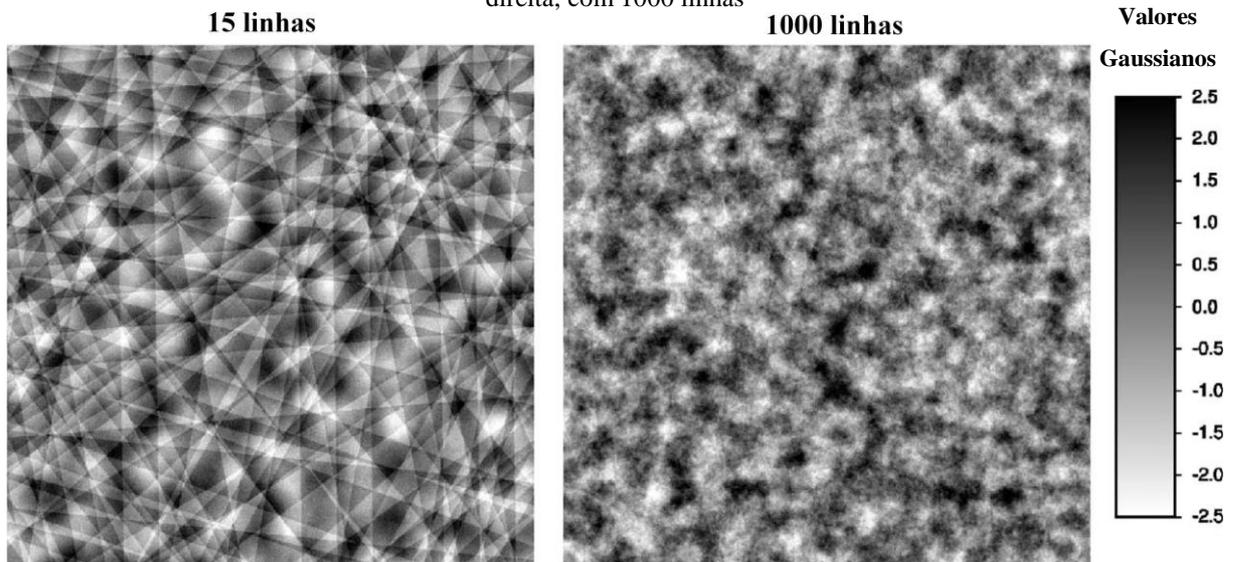
Figura 1 - Transformação dos dados via *normal scores* de uma distribuição acumulada F dos dados originais para uma distribuição gaussiana padrão acumulada G



Fonte: Chilès e Delfiner (1999, p. 391)

A seguir, é definido um número N de linhas independentes a serem simuladas. O número de linhas é importante para evitar feições espaciais indesejadas nas simulações (Figura 2). Emery e Lantuéjoul (2006) alegam que, na prática, 1000 linhas geram bons resultados na maioria dos casos, mas um número maior pode ser eventualmente necessário dependendo do espaçamento entre pontos simulados e alcance dos variogramas.

Figura 2 - Duas simulações utilizando Simulação por Bandas Rotativas. À esquerda, com apenas 15 linhas e, à direita, com 1000 linhas



Fonte: Emery e Lantuéjoul (2006)

A simulação das linhas pode ser feita de diversas maneiras. Os métodos mais tradicionais são por médias móveis (BLACK E FREYBERG, 1990) ou via métodos espectrais contínuos (SHINOZUKA E JAN, 1972). Emery e Lantuéjoul (2006) apresentam, para cada tipo do modelo de covariância (esférico, exponencial, gaussiano e outros), os métodos mais eficientes para simular as linhas.

Cada uma das linhas é simulada repetidamente com sementes aleatórias diferentes, até atingir o número de realizações estipulado pelo usuário. O próximo passo é alocar as N linhas no espaço. Elas podem ser distribuídas uniformemente, mas as implementações de Lantuéjoul (1994) e Emery e Lantuéjoul (2006) fazem essa alocação utilizando uma sequência de *van der Corput*, justificando que desse modo há uma melhor representação das propriedades ergódicas.

Com as linhas espacialmente distribuídas, a simulação de um ponto qualquer no espaço é dada pela soma dos valores das projeções desse ponto nas N linhas:

$$Y_s(u) = \frac{1}{\sqrt{N}} \sum_{i=1}^N u'_i \quad (1)$$

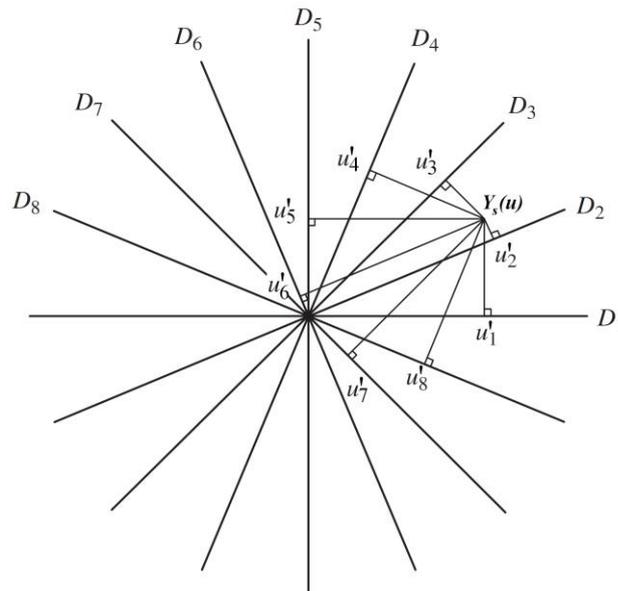
no qual $Y_s(u)$ é o ponto simulado no local u ; N é o número de linhas simuladas independentemente e u'_i é o valor da projeção do ponto u na linha i .

Um exemplo em 2D com 8 linhas ($D_1...D_8$) é apresentado na Figura 2.3 para exemplificar o processo. Para simular o ponto $Y_s(u)$, primeiramente são identificadas as projeções $u'_1...u'_8$ desse ponto nas oito linhas. Como os valores de $u'_1...u'_8$ são conhecidos, o valor simulado de $Y_s(u)$ é obtido por meio da equação (1).

Journel (1974) demonstra que esse procedimento produz realizações com valor esperado igual a zero, estacionaridade de segunda ordem e covariância que tende à das linhas simuladas à medida que se aumenta o número de linhas, seja para casos 2D ou 3D.

É importante ressaltar que o modelo de covariância utilizado para simular as linhas é isotrópico e representa apenas uma estrutura. A anisotropia geométrica é imposta multiplicando os valores simulados por fatores referentes à orientação das linhas no espaço. No caso de múltiplas estruturas, cada parte do modelo é simulada independentemente e, ao fim, somadas para gerar o modelo final. O efeito pepita também é posteriormente adicionado ao modelo e é simulado aplicando Monte Carlo em uma distribuição gaussiana de média zero e variância igual ao valor do efeito pepita desejado (EMERY E LANTUÉJOUL, 2006).

Figura 3 - Exemplo ilustrativo de como se obter o valor simulado do ponto $Y_s(u)$ a partir de 8 linhas simuladas ($D_1 \dots D_8$)



Fonte: Modificado de Chilès e Delfiner (1999, p. 503)

Com as simulações não condicionais prontas, o próximo passo é condicioná-las aos valores amostrais conhecidos. O processo para isso é apresentado em Journel (1974) e consiste em, primeiramente, calcular o resíduo entre os valores conhecidos e os simulados no mesmo local. Isso é feito para cada uma das realizações, gerando um conjunto de resíduos para cada uma delas. Em seguida, é realizada a krigagem simples de cada um desses conjuntos de resíduos para média igual a zero. O modelo de simulação final condicionado aos dados é então obtido adicionando os valores de cada resíduo krigado à sua realização não condicional.

Com os modelos simulados prontos, basta retro-transformar os valores gaussianos para as unidades originais. Isso pode ser feito pelo processo inverso ao *normal score*, associando os quantis da distribuição simulada gaussiana de média zero e variância unitária para a distribuição de referência nas unidades originais.

2.2.SIMULAÇÃO SEQUENCIAL GAUSSIANA

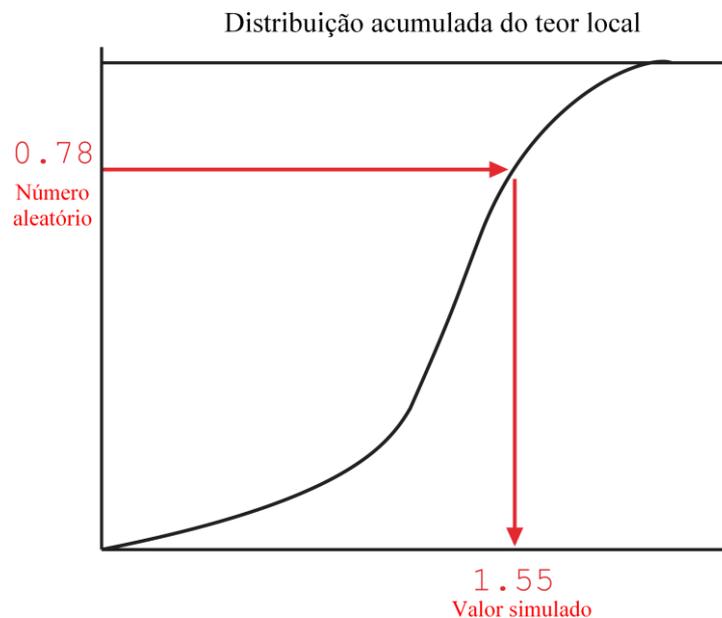
Os algoritmos sequenciais se baseiam na aplicação recursiva do axioma Bayesiano para expressar que qualquer distribuição acumulada condicional de dois pontos é igual ao produto das distribuições univariadas de cada ponto. Considerando N variáveis aleatórias a serem simuladas, esse teorema é desenvolvido conforme mostrado em (2):

$$\begin{aligned}
P(A_1, \dots, A_N) &= P(A_N | A_1, \dots, A_{N-1}) \cdot P(A_1, \dots, A_{N-1}) \\
P(A_1, \dots, A_N) &= P(A_N | A_1, \dots, A_{N-1}) \cdot P(A_{N-1} | A_1, \dots, A_{N-2}) \cdot P(A_1, \dots, A_{N-2}) \\
&\vdots \\
P(A_1, \dots, A_N) &= P(A_N | A_1, \dots, A_{N-1}) \cdot P(A_{N-1} | A_1, \dots, A_{N-2}) \cdots P(A_2 | A_1) \cdot P(A_1)
\end{aligned} \tag{2}$$

onde A_1, \dots, A_N representam um conjunto de N variáveis aleatórias e P representa a probabilidade conjunta, individual ou condicional ($()$) das variáveis.

Em outras palavras, a reprodução da probabilidade conjunta dos dados pode ser atingida estimando as distribuições locais sequencialmente. Uma distribuição é estimada e dela é simulado aleatoriamente um valor por meio de um algoritmo de geração de números aleatórios (exemplo na Figura 4). Esse valor é então utilizado como condicionante para a simulação dos pontos subsequentes.

Figura 4 - Processo de obtenção de um valor simulado a partir da geração de um número aleatório entre [0,1]



Fonte: Modificado de Rossi e Deutsch (2014, p. 170)

As estimativas das distribuições são feitas por métodos de estimativa não-linear. Quando essas distribuições são estimadas parametricamente, via krigagem multigaussiana, o método de simulação é o chamado SSG.

A aplicação da SSG se inicia com a transformação dos dados para o espaço gaussiano (da mesma maneira realizada no método anterior), assumindo a hipótese de multigaussianidade

dos dados, de modo que todas as distribuições locais sejam gaussianas, podendo ser expressas assim parametricamente por meio das respectivas médias e variâncias.

O processo sequencial é então iniciado. É definido o caminho aleatório das visitas a serem feitas nos locais do *grid* e, no primeiro desses pontos, é estimada sua distribuição via krigagem multigaussiana. Isso é feito a partir do estimador de krigagem simples, do qual o valor estimado e a variância da krigagem são utilizados, respectivamente, como média e variância da distribuição gaussiana local. Dessa distribuição, é amostrado aleatoriamente um valor, que é incorporado ao banco de dados para condicionar as simulações dos pontos subsequentes. O processo é repetido até se completar a simulação de todo o *grid*.

A reprodução da covariância entre dado simulado e dado condicionante é provada no teorema abaixo. Essa reprodução também é válida entre dados simulados entre si, uma vez que a metodologia sequencial utiliza os pontos simulados como condicionantes no processo. A partir do estimador de krigagem simples $Y^*(u)$, o valor simulado $Y_s(u)$ pode ser expresso como $Y^*(u)$ somado a um resíduo $R(u)$, de média zero e variância igual à da variância da krigagem simples:

$$Y_s(u) = Y^*(u) + R(u) \quad (3)$$

A fórmula de covariância pode então ser desenvolvida como demonstrado em (4), de tal maneira que, ao fim, a covariância entre dado simulado $Y_s(u)$ e dado condicionante $Y(u_\alpha)$ se mostra a mesma que a do modelo de covariância considerado no sistema de krigagem simples:

$$\begin{aligned} Cov\{Y_s(u), Y(u_\alpha)\} &= E\{Y_s(u), Y(u_\alpha)\} \\ Cov\{Y_s(u), Y(u_\alpha)\} &= E\left\{\left[\sum_{\beta=1}^n \lambda_\beta \cdot Y(u_\beta) + R(u)\right] \cdot Y(u_\alpha)\right\} \\ Cov\{Y_s(u), Y(u_\alpha)\} &= \sum_{\beta=1}^n \lambda_\beta \cdot E\{Y(u_\beta) \cdot Y(u_\alpha)\} + E\{\cancel{R(u)} \cdot Y(u_\alpha)\} \\ Cov\{Y_s(u), Y(u_\alpha)\} &= \sum_{\beta=1}^n \lambda_\beta \cdot E\{Y(u_\beta) \cdot Y(u_\alpha)\} \\ Cov\{Y_s(u), Y(u_\alpha)\} &= \sum_{\beta=1}^n \lambda_\beta \cdot C(u_\alpha - u_\beta) \\ Cov\{Y_s(u), Y(u_\alpha)\} &= C(u - u_\alpha) \end{aligned} \quad (4)$$

↪ A partir do sistema de krigagem simples

onde $Y(u_\alpha)$ é um dado amostral condicionante, λ_β é o peso da krigagem simples para a amostra $Y(u_\beta)$ – uma das n amostras da vizinhança de $Y(u)$; C consiste no modelo de covariância interpretado para o fenômeno espacial.

Ao fim, as simulações são retro-transformadas reestabelecendo assim as unidades originais e reproduzindo o histograma de referência. Após a retro-transformação é possível realizar a reblockagem, quando desejado.

Quanto às questões práticas do método, Rossi e Deustch (2014) expõem alguns comentários relevantes:

- Os valores gerados por SSG apresentam menor conectividade que os dados originais pois a distribuição gaussiana maximiza a entropia dos dados. Esse efeito é menor quanto melhor forem definidos os domínios, uma vez que a SSG é bastante sensível aos desvios na estacionaridade;
- Para melhor desempenho computacional, os dados podem ser associados às células do grid, o que acelera consideravelmente as simulações. No entanto, essa aproximação deve ser aplicada com cautela, pois pode gerar significativas perdas de informação dependendo do adensamento amostral e das dimensões das células do grid;
- O número de dados utilizados para simulação gera diferentes consequências. A utilização de um maior número de dados previamente simulados reproduz melhor a variabilidade local, enquanto a utilização de um maior número de dados amostrais possibilita uma estimativa mais acurada das médias e variâncias condicionais, além de melhor reproduzir o variograma.

Para implementação desses métodos, há algumas ressalvas a serem feitas (GOOVAERTS, 1997, p. 378-379). Para *grids* grandes, em determinado momento do processo a utilização de todos os valores previamente simulados no sistema de krigagem se torna inviável computacionalmente. Para contornar isso, as estimativas das distribuições são restringidas às informações da vizinhança estabelecida e as buscas são separadas em dados amostrais e dados previamente simulados. Outra questão é quanto à visita dos locais a serem simulados. Caso se siga uma determinada ordem espacial, corre-se o risco da geração de eventuais “continuidades artificiais” devido à maior concentração de informações previamente simuladas em determinada região. Por causa disso, a sequência de pontos visitados é aleatorizada para evitar esses problemas.

Outra questão implementacional é para casos em que se trabalha com modelos variográficos de *ranges* longos. No decorrer das visitas aos locais a serem simulados, tende-se a diminuir rapidamente a influência dos dados mais distantes devido a presença de pontos previamente simulados nas proximidades, o que pode desfavorecer a reprodução das estruturas dos variogramas a longas distâncias. A maneira proposta por Tran (1994) para melhor reproduzir os variogramas nesses casos foi utilizar múltiplos *grids*. Primeiro, são simulados valores em um *grid* mais esparsos e esses valores são depois utilizados em *grids* cada vez mais densos, até completar todas as localidades a serem simuladas.

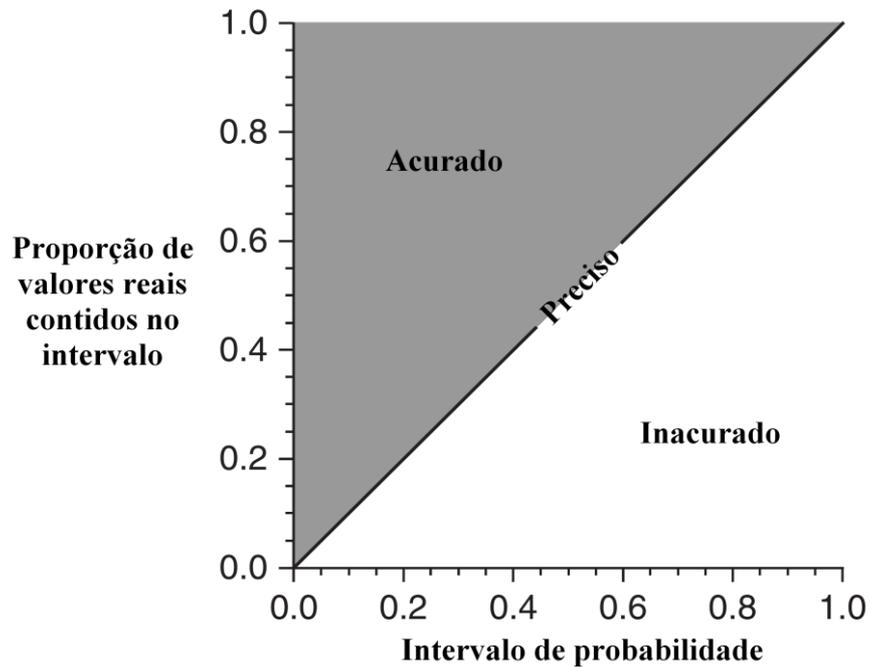
2.3. VALIDAÇÕES DAS SIMULAÇÕES

Checar as simulações é uma parte crítica do processo pois elas são bastante sensíveis aos parâmetros utilizados, nem sempre reproduzindo as estatísticas desejadas. As validações básicas são a reprodução do histograma e dos variogramas de referência. Para melhorar essas reproduções, diferentes estratégias de buscas são testadas, assim como diferentes números de dados para o condicionamento das distribuições.

Quando há um bom número de dados condicionantes, podem ser também cheçadas a acuracidade e precisão dos dados via gráficos de acuracidade. Para isso, é omitido um valor de um ponto amostral e neste local são geradas diversas simulações utilizando apenas os outros dados amostrais como condicionantes, e assim sucessivamente, tal como nas validações cruzadas realizadas para estimativas. Os valores simulados nos pontos são então comparados com o valor amostral do local por meio da reprodução dos intervalos de probabilidades. Ou seja, para a probabilidade de 90% (representada pelo intervalo entre os quantis 0.05 e 0.95 dos valores simulados), o valor amostral deve estar contido neste intervalo em 90% dos casos, e assim por diante para os outros valores de probabilidades.

Esses resultados expressos num gráfico de intervalos de probabilidade *versus* proporção de valores reais dentro dos correspondentes intervalos (Figura 5), representa acuracidade dos dados quando os pontos plotados estão acima da reta de 45° e precisão quanto mais próximo estiverem da mesma reta.

Figura 5 - Exemplo de um gráfico de acuracidade



Fonte: Modificado de Pycrz e Deutsch (2014)

Outras validações também podem ser feitas quando há outras informações à disposição. Por exemplo, quando há dados confiáveis de produção, é interessante que as curvas de teor-tonelagem desses dados também sejam adequadamente reproduzidas pelas simulações.

3. SIMULAÇÃO VIA MÚLTIPLOS PASSEIOS ALEATÓRIOS

3.1. TRABALHOS ANTERIORES E NOVAS CONTRIBUIÇÕES

Passeio aleatório é um tema que foi difundido dentro de diversas áreas nas últimas décadas. Por exemplo, para traçar caminhos de moléculas em fluidos e inferir flutuações de preço de ações em bolsas de valores. Por ser abordado em trabalhos desde o início do século passado (PEARSON, 1905), o passeio aleatório apresenta um formalismo matemático já bem desenvolvido que pode ser encontrado em diversos livros (LAWLER E LIMIC, 2010).

Combinar os passeios aleatórios com krigagem para gerar simulações geoestatísticas é uma abordagem nova com trabalhos iniciais de Chu (1993) e Costa (1997). Os conceitos de passeios aleatórios foram revisitados e adaptados com os trabalhos de Ribeiro *et al.* (2012) e Ribeiro *et al.* (2013). Neles, são apresentados toda a base da nova proposta, determinando como é possível gerar realizações espaciais por meio desses componentes aleatórios combinados com krigagem. Todo esse processo é abordado ao longo desse capítulo junto com as novas complementações.

Quanto às diferenças entre a metodologia dos primeiros trabalhos e a apresentada aqui, não há mudanças na base do método, mas há duas complementações a serem destacadas: a simulação independente do efeito pepita e a simulação de pontos auxiliares quando necessário. Essas contribuições também são abordadas ao longo do capítulo e foram essenciais para resolver problemas anteriores referentes à reprodução da covariância à curtas distâncias e quanto ao aparecimento de artefatos visuais.

Na parte prática, há também contribuições significativas dessa dissertação. Nos primeiros trabalhos, o método era executado via protótipos que eram restritos a casos isotrópicos e estruturas esféricas. Com o novo *plugin* apresentado ao fim desse capítulo, foi possível operacionalizar todo o processo em um único aplicativo, utilizar mais modelos variográficos, incorporar anisotropia, paralelizar o processo eficientemente em múltiplos processadores, calibrar automaticamente as simulações, utilizar toda a interface gráfica do SGeMS entre outras vantagens.

3.2. INTRODUÇÃO TEÓRICA

Considere a introdução à Simulação Sequencial Gaussiana apresentada no capítulo anterior. Segundo Pyrcz e Deutsch (2014, p. 123), o valor simulado pode ser descrito como o valor estimado por krigagem simples adicionado a um resíduo com covariância correta, conforme já apresentado na equação (3) – vide Capítulo 2, seção 2.2.

Segundo Journel e Huijbregts (1978, p. 495), esse resíduo simulado deve ser isomórfico ao resíduo real (desconhecido) e independente do valor krigado – condição que é atendida ao aplicar o procedimento de krigagem sobre um resíduo com a mesma configuração dos dados amostrais.

A abordagem da Simulação via Múltiplos Passeios Aleatórios se baseia nesses dois pilares. Sua principal diferença em relação aos outros métodos é a forma do resíduo, derivada de passeios aleatórios.

3.3. PASSEIOS ALEATÓRIOS

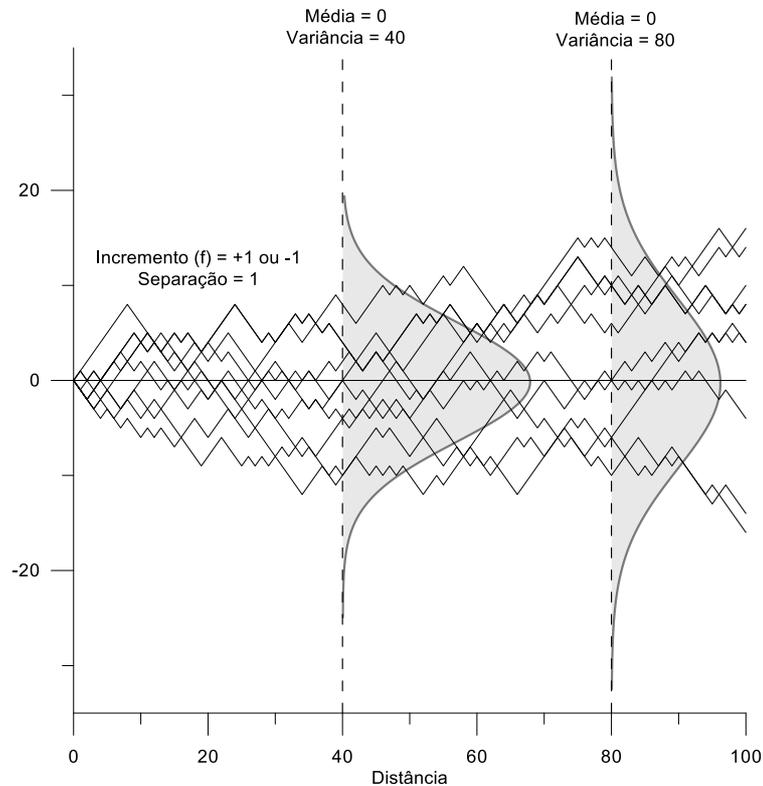
Passeios aleatórios são caminhos randômicos unidimensionais definidos por separações δh , um valor de incremento f e um número de separações n . O caminho inicia com o valor $RW(0) = 0$ e a cada separação recebe um incremento independente de $+f$ ou $-f$, com 50% de chance cada. O caminho continua até o número de separações estipuladas, resultando em caminhos tais quais os da Figura 6.

Em um conjunto de vários passeios aleatórios, o conjunto dos valores $RW(h)$ para uma mesma distância h compõem uma distribuição gaussiana de média 0 e variância definida por:

$$\sigma^2 = \frac{h}{\delta h} \cdot f^2 \quad (5)$$

na qual σ^2 representa a variância dos valores $RW(h)$ na distância h para a separação δh e incremento f .

Figura 6 - Ilustração dos passeios aleatórios e suas propriedades Gaussianas. Começando em 0, um incremento positivo ou negativo é adicionado randomicamente a cada separação. Para um conjunto de passeios aleatórios, seus valores a qualquer distância h apresentam uma distribuição gaussiana de média 0 e cuja variância cresce proporcionalmente com o aumento de h .



Fonte: Caixeta et al. (2015)

3.4.UTILIZANDO PASSEIOS ALEATÓRIOS PARA GERAR SIMULAÇÕES GEOESTATÍSTICA CONDICIONAIS

A ideia do método é assumir que o resíduo condicionado em cada nó do grid simulado seja resultante dos passeios aleatórios krigados, o que permite utilizar as suas propriedades apresentadas anteriormente, de modo que:

- O resíduo seja gaussiano, uma vez que é uma combinação linear (krigagem simples) de valores de distribuições gaussianas;
- O resíduo apresente média nula, sem implicar assim em viés;
- A variância do resíduo possa ser linearmente ajustada ao alterar o valor de f .

Para utilizar os passeios aleatórios a fim de realizar simulações que levem em consideração tanto o condicionamento pelos dados amostrais como a continuidade espacial do fenômeno, esses passeios são associados às localidades onde há amostras. O processo ocorre no espaço gaussiano e segue os seguintes passos:

- (i) Normalizar os dados para unidades gaussianas;
- (ii) Simular um conjunto de passeios aleatórios a partir dos pontos onde existem amostras e associar para cada amostra um passeio do conjunto simulado;
- (iii) Resolver o sistema de krigagem simples para o nó a ser simulado e usar os pesos obtidos para obter o valor simulado tal que:

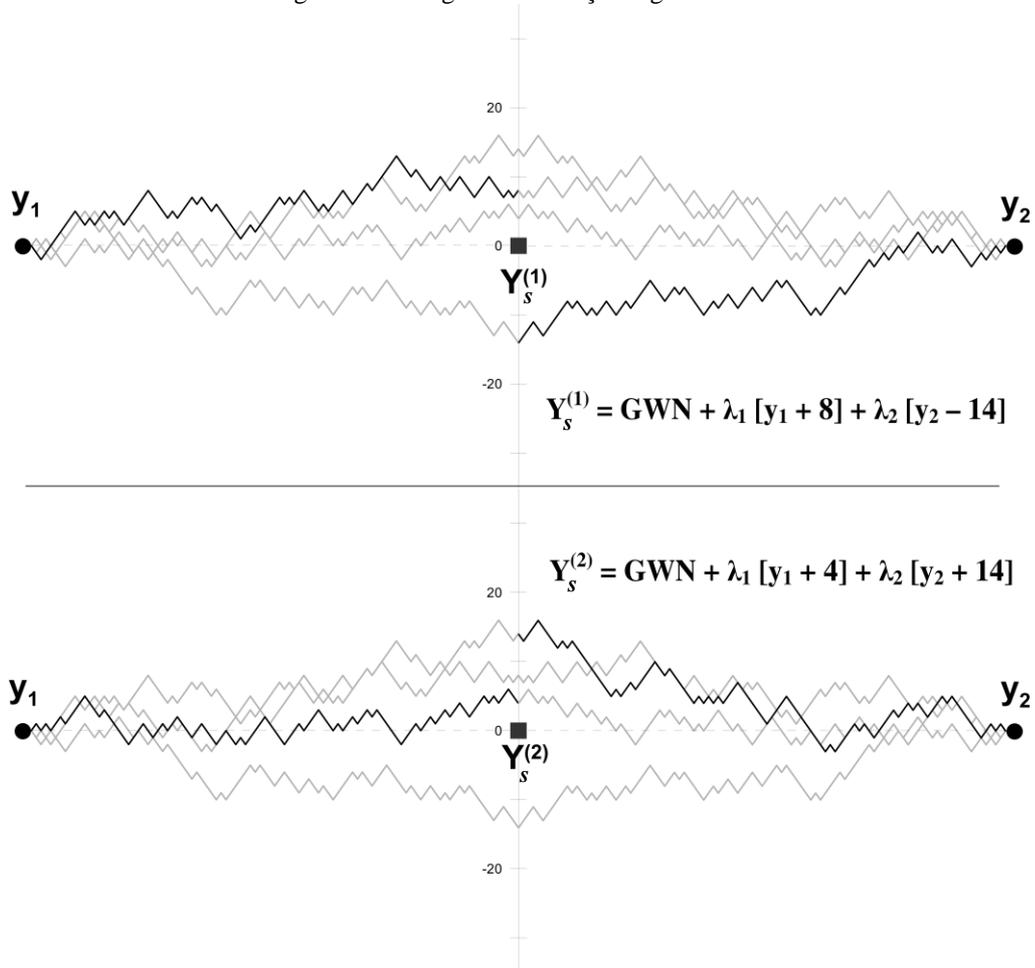
$$Y_s^{(l)}(u) = \left[\sum_{i=0}^n \lambda_i [Y_{ui}] \right]_{Y^*(u)} + \left[GWN + \sum_{i=0}^n \lambda_i [RW_{ui}(h)] \right]_{R(u)} = \sum_{i=0}^n \lambda_i [RW_{ui}(h) + Y_{ui}] + GWN \quad (6)$$

no qual $Y_s^{(l)}$ é o valor simulado para a realização (l) no local u ; n é o número de dados condicionantes na vizinhança de u ; λ_i é o peso calculado para a amostra na localidade u_i ; Y_{ui} é o valor da amostra u_i em unidades gaussianas; $RW_{ui}(h)$ é o valor do passeio aleatório associado à amostra u_i na distância h (distância anisotrópica entre amostra e ponto a ser simulado); GWN é uma componente adicionada para reestabelecer o efeito pepita suavizado pela krigagem (mais detalhes adiante).

- (iv) Simular todos os nós do grid seguindo o passo anterior. As próximas realizações são geradas da mesma maneira, porém usando diferentes passeios aleatórios em cada localidade amostral para cada nova realização.
- (v) Calibrar linearmente a variância do resíduo para que iguale a variância global unitária.
- (vi) Retrotransformar as simulações para as unidades originais.

A Figura 7 ilustra este processo. Inicialmente, uma série de passeios aleatórios independentes são gerados ao longo de linhas a partir das localizações das amostras (em cinza na figura) em um espaço 2D ou 3D. Em seguida, em cada local onde há um dado condicionante é alocado, aleatoriamente, um desses passeios (linhas em preto). Considerando dois dados condicionantes y_1 e y_2 , e o ponto Y_s a ser simulado, o valor do ponto para a primeira realização $Y_s^{(1)}$ é calculado utilizando os valores de $RW^{(1)}y_1 = 8$ e $RW^{(1)}y_2 = -14$.

Figura 7 - Exemplo ilustrativo de duas realizações simuladas via múltiplos passeios aleatório. Foi simulado o ponto $Y_s^{(i)}$ a partir de duas amostras y_1 e y_2 . O resultado da primeira realização é apresentado na parte superior da figura e o da segunda realização logo abaixo.



Fonte: Caixeta et al. (2015)

Para a segunda realização, representada na parte inferior da Figura 7, são utilizados os mesmos pesos e valores amostrais de y_1 e y_2 , mas diferentes passeios aleatórios (novas linhas pretas), alterando os valores referentes aos passeios para $RW^{(2)}y_1 = 4$ e $RW^{(2)}y_2 = 14$ ao se calcular o valor simulado $Y^{(2)}$.

3.5.REPRODUÇÃO DO EFEITO PEPITA E CALIBRAGEM DA VARIÂNCIA

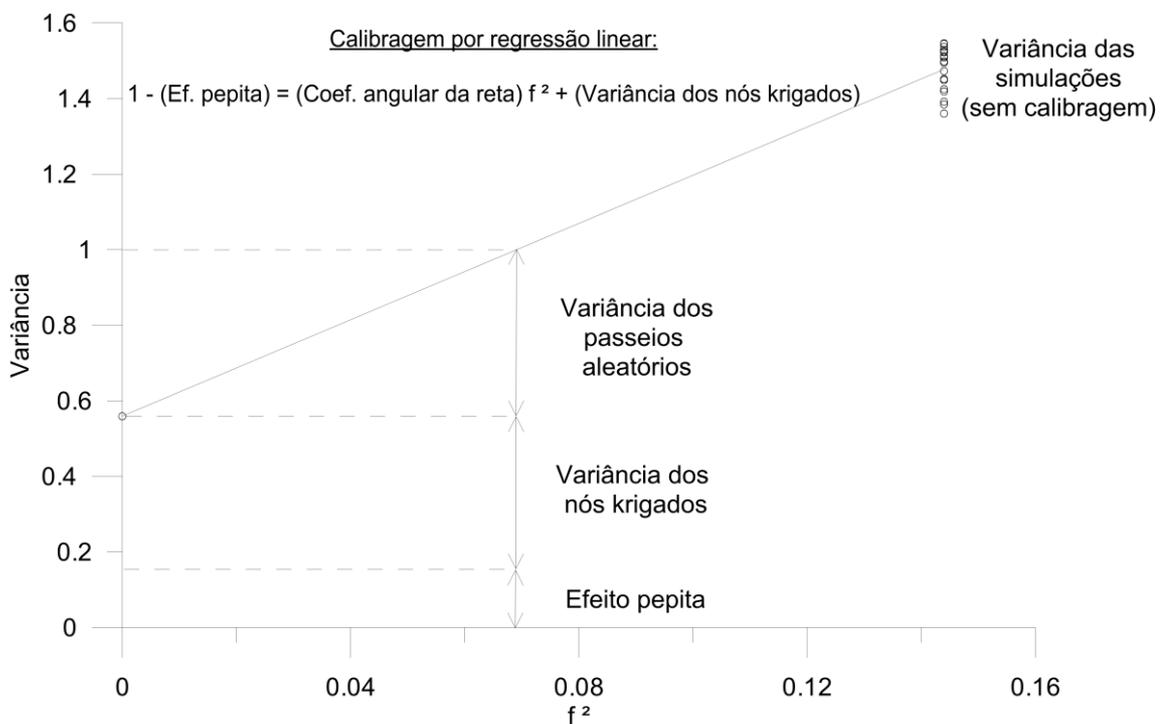
Como mencionado anteriormente, há no processo de simulação a adição de uma componente denominada GWN (sigla proveniente do termo inglês *Gaussian White Noise*). Esse

elemento foi uma contribuição do presente trabalho para resolver o problema de que as simulações não reproduziam corretamente o efeito pepita.

Para corrigir essa questão, foi utilizada uma abordagem semelhante à aplicada na Simulação por Bandas Rotativas, que simula o efeito pepita à parte e o adiciona aos valores simulados (EMERY E LANTUÉJOUL, 2006). A simulação desse efeito pepita é feita amostrando aleatoriamente (para cada localidade simulada) um valor de uma distribuição gaussiana de média zero e variância igual ao efeito pepita.

Após as simulações, ainda é necessária uma calibragem da variância do resíduo para que as simulações apresentem variância unitária. A variância das realizações é composta pela variância dos valores krigados, da componente do efeito pepita e dos passeios aleatórios krigados (Figura 8). Como cada componente é gaussiana e possui média nula, a variância total do conjunto é dada pela soma de suas partes. Após uma ou mais realizações, o valor do incremento f para que o conjunto apresente variância unitária pode ser calculado via uma simples regressão linear, uma vez que a variância dessa componente é linearmente correlacionada ao quadrado do incremento f conforme a equação (5).

Figura 8 - Calibragem do valor do incremento f para que realizações apresentem variância unitária.



3.6.SIMULANDO EM CASOS COM POUCOS DADOS AMOSTRAIS

A Simulação via Múltiplos Passeios Aleatórios garante a reprodução global do histograma dos dados originais uma vez que retrotransforma uma distribuição gaussiana de média nula e variância unitária usando uma função inversa à utilizada para alocar os quantis originais aos Gaussianos. O método também garante a reprodução da covariância entre dado condicionante e ponto do grid, de modo análogo ao que foi apresentado na equação (4) para a Simulação Sequencial Gaussiana.

No entanto, como não é um método sequencial, não é possível garantir a reprodução da covariância entre os nós simulados. Quando há uma boa densidade de dados condicionantes, isso não chega a influenciar a reprodução global do variograma. Porém, quando há poucos dados amostrais, isso afeta a reprodução da continuidade a curtas distâncias, podendo também gerar artefatos visuais.

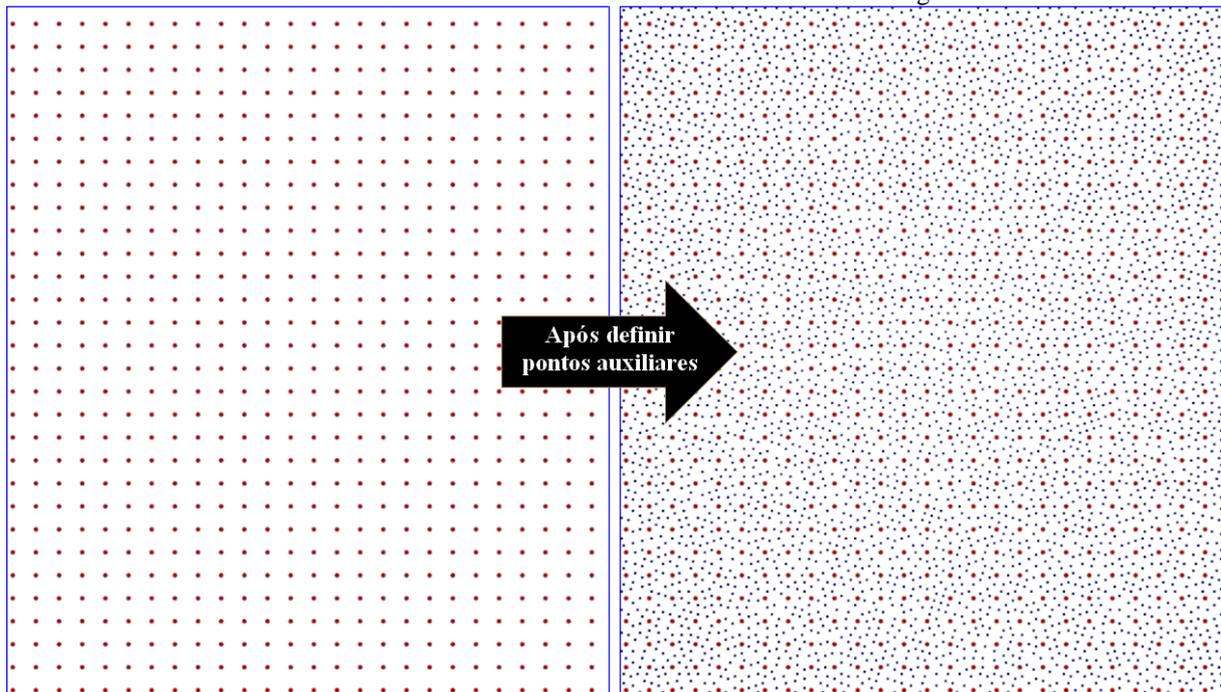
Assim sendo, a Simulação via Múltiplos Passeios Aleatórios é altamente dependente da densidade de dados condicionantes. Para evitar então problemas decorrentes da falta de dados condicionantes e tornar o método mais robusto, foi desenvolvida uma maneira de popular previamente o grid com pontos auxiliares quando necessário.

Inicialmente, são definidos os locais em que os pontos auxiliares são necessários. É estabelecida uma distância mínima entre pontos para que haja uma boa reprodução dos variogramas e não ocorra artefatos (mais detalhes sobre essas condições mínimas de espaçamento na seção 3.5).

Todos os locais do grid são visitados aleatoriamente e sequencialmente para checar se há ao menos uma amostra (ou ponto auxiliar já definido no processo) dentro desse raio mínimo. Se nada for encontrado, este local será definido como ponto auxiliar. Esse procedimento segue até todos os locais do grid serem avaliados. A Figura 9 ilustra o resultado.

O próximo passo consiste em associar valores aos pontos auxiliares. Para isso, são realizadas simulações nesses locais. Qualquer método que gere valores gaussianos condicionados aos dados amostrais pode ser utilizado para isso (por exemplo, Simulação Sequencial Gaussiana ou Simulação por Bandas Rotativas). Assim, será criado um novo banco de dados para cada realização, tal que os dados amostrais são iguais em todos os bancos, mas os valores nos pontos auxiliares são variáveis.

Figura 9 - Um exemplo de como um banco de dados (vermelho) é populado com pontos auxiliares (azul), considerando um raio mínimo de $1/3$ da distância amostral original.



Tendo à disposição um novo banco de dados para cada realização, a simulação é conduzida tal como anteriormente, com o acréscimo que agora haverá diferentes valores krigados para cada realização ao invés de um único – embora os pesos de krigagem sejam os mesmos para todas as realizações, uma vez que as localidades são as mesmas. Em suma, o grid final de simulação é pré-preenchido com valores simulados com outro algoritmo até atingir a densidade de nós compatível com o bom funcionamento da Simulação com Múltiplos Passeios Aleatórios (SMPA).

3.7.CONDIÇÕES MÍNIMAS DE ADENSAMENTO AMOSTRAL

O uso de pontos auxiliares resolve eventuais problemas quanto a reprodução da continuidade a curtas distâncias e ocorrência de artefatos. Mas para definir os pontos auxiliares adequadamente é necessário realizar uma boa escolha quanto à distância mínima entre os pontos, de modo a garantir bons resultados e não sobrecarregar o processo gerando um excessivo número de pontos auxiliares.

Como discutido anteriormente, a principal razão para a ocorrência de artefatos e problemas na reprodução variográfica é que o método não garante a reprodução da covariância

entre pontos do grid, mas apenas entre amostra e ponto do grid. Alguns outros fatores também tendem a contribuir para ocorrência desses problemas:

- Os parâmetros do modelo variográfico, que influenciam diretamente se o sistema de krigagem dará um peso muito mais alto à amostra mais próxima do que às outras da vizinhança;
- O quanto de variância terá que ser restaurada após a krigagem naquela vizinhança. Quanto maior essa adição de variância, mais marcados deverão ser os artefatos;
- A discretização do grid. Quanto mais próximos entre si os pontos a serem simulados, mais provável a presença de artefatos nas proximidades das amostras.

A sensibilidade a essas condições foi avaliada com diversos testes. Foram analisados diferentes espaçamentos amostrais, modelos variográficos e espaçamentos entre pontos a serem simulados. Os principais resultados estão detalhados no Anexo A e mostraram que, em um caso de malha amostral regular, deve haver amostras a uma distância mínima de 3-6 vezes o valor da separação entre os pontos do grid a ser simulado – o que representa uma variância a ser restaurada de 3-15% dependendo do tipo do modelo variográfico. Uma reprodução apropriada dos variogramas também corresponde a essas mesmas condições.

Entre os modelos variográficos, os do tipo esférico e gaussiano são os mais flexíveis, enquanto o exponencial requer amostras (ou pontos auxiliares) bastante próximas entre si para gerar bons resultados. Isso se dá devido à sua característica de alta variabilidade a curtas distâncias, em comparação com os outros dois modelos. Não foram feitos testes utilizando outros modelos além desses três.

Outras condições tendem a ocultar esses artefatos e conseqüentemente diminuir as restrições quanto à distância amostral mínima. Uma delas é a reblockagem dos pontos simulados para um maior suporte. Outra é o efeito pepita, que tende a ocultar os artefatos à medida que seu valor aumenta.

3.8.ALGORITMO E PLUGIN DESENVOLVIDOS

A característica mais importante da Simulação via Múltiplos Passeios Aleatórios é a de que pode ser paralelizada tanto em escala de grid como de realizações, uma vez que não é sequencial e que os pesos serão os mesmos para todas as realizações. Os subprocessos, tais

quais as simulações dos passeios e calibragem do ruído, também podem ser eficientemente paralelizados. Dessa maneira, é possível obter resultados com um processamento computacional bastante rápido quando a técnica é executada em máquinas com vários processadores.

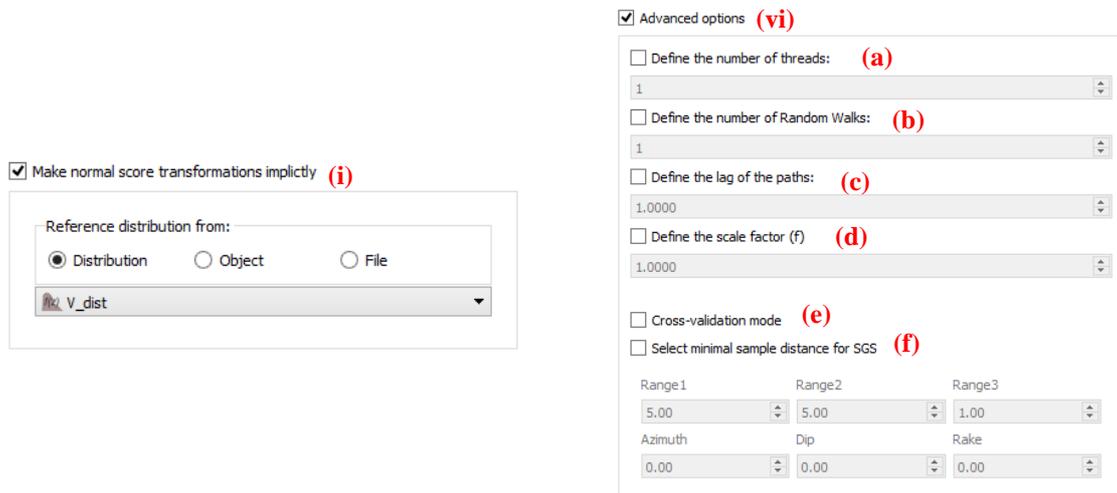
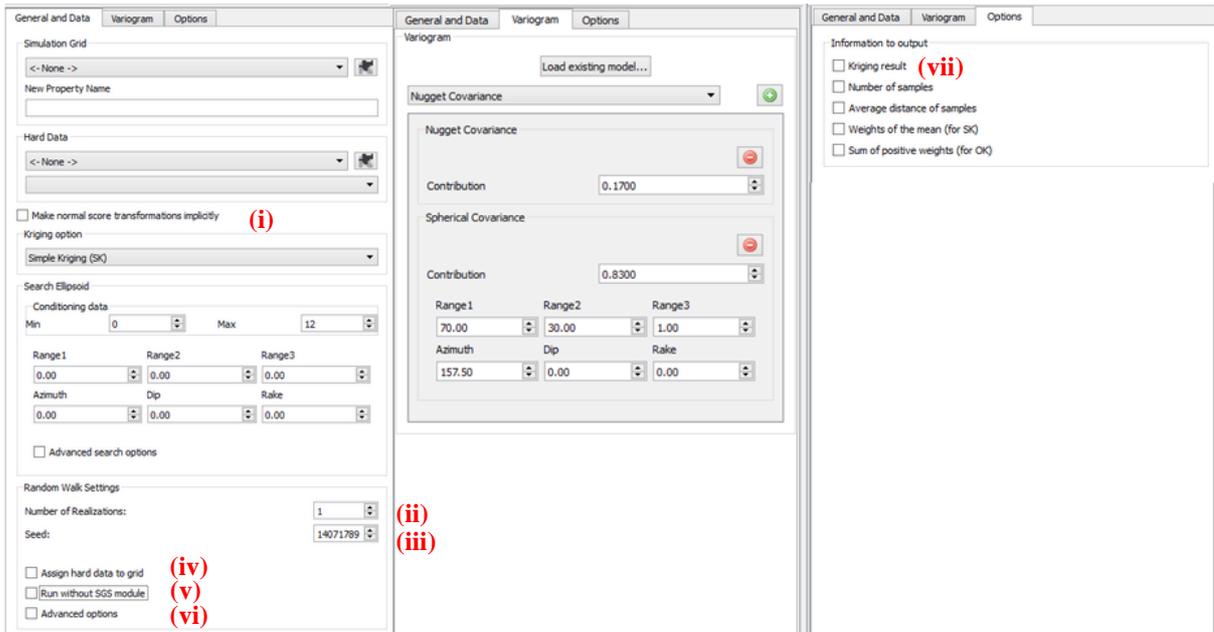
Para poder utilizar o método em sua plenitude, foi desenvolvido em C++ um plugin para ser utilizado junto ao SGeMS (REMY *ET AL.*, 2009). Ele permite simulações 3D paralelamente e aproveita toda a interface e funcionalidades já incorporadas ao software.

Ele foi inicialmente baseado no plugin de krigagem do SGeMS e foi estruturado para que possa ser executado apenas com os próprios parâmetros utilizados comumente na krigagem, com o número de realizações a serem simuladas e com a distribuição de referência para realizar as transformações gaussianas iniciais e finais. Parâmetros mais específicos são inicializados automaticamente ou podem ser customizados pelo usuário em uma aba de opções avançadas. A rotina do algoritmo segue a sequência abaixo:

- (i) É realizada a visita em todos os locais do grid (tal como explicado na seção 3.4) para definir as localidades que serão simuladas como pontos auxiliares. O elipsóide de busca pode ser automaticamente ou manualmente definido. Os pontos auxiliares são simulados via Simulação Sequencial Gaussiana (paralelizada a nível de realização);
- (ii) São gerados um número específico de passeios aleatórios usando um conjunto de especificações automaticamente definidas, tais como separação, incremento f e número de incrementos;
- (iii) Para cada realização, um passeio aleatório diferente é randomicamente alocado para cada dado condicionante;
- (iv) Os dados são normalizados implicitamente via *normal score*, com base na distribuição de referência informada;
- (v) Os sistemas de krigagem são resolvidos em cada local do grid com as amostras encontradas na vizinhança. Seus pesos são utilizados para calcular o valor simulado para cada realização através da equação (6);
- (vi) Ao finalizar as simulações, todos os pontos são novamente visitados para calibrar automaticamente a variância do ruído e retrotransformar o valor simulado final;

A interface do plugin é mostrada na Figura 10. Os novos parâmetros implementados foram brevemente explicados abaixo. Os parâmetros em comum com o plugin de krigagem não foram aqui abordados e podem ser consultados em Remy *et al.* (2009).

Figura 10 - Interface do plugin para Simulação via Múltiplos Passeios Aleatórios desenvolvido para SGeMS.



(i) *Make normal score transformations implicitly*: caso selecionada, o algoritmo irá normalizar os dados implicitamente e retrotransformar o resultado final; caso contrário, os

dados de entrada já devem estar normalizados e o resultado final será dado em unidades gaussianas;

- (ii) *Number of realizations*: deve ser informado o número desejado de realizações;
- (iii) *Seed*: o valor da semente aleatória;
- (iv) *Assign hard data to grid*: atribui o valor de cada dado amostral à célula do grid mais próxima;
- (v) *Run without SGS module*: caso selecionado, não será feita uma varredura para checar se há necessidade de pontos auxiliares;
- (vi) *Advanced options*:
 - (a) *Number of threads*: número de processadores utilizados para simulação;
 - (b) *Number of Random Walks*: número de passeios aleatórios simulados;
 - (c) *Lag of the paths*: espaçamento entre cada incremento do passeio aleatório – δh na equação (5);
 - (d) *Scale factor*: é o valor do incremento f ; se for pré-estabelecido aqui, o resultado final será dado usando esse f e não será feita calibragem automática ao fim;
 - (e) *Cross-validation mode*: caso selecionado, o algoritmo realizará a validação cruzada das simulações (apenas nas localidades com pontos amostrais);
 - (f) *Select minimum sample distance for SGS*: neste campo podem ser atribuídas as dimensões do elipsoide de busca para ser utilizado na varredura por pontos auxiliares;
- (vii) *Kriging result*: caso selecionado, é gerado um mapa com apenas os valores krigados durante o processo (em unidades gaussianas).

Alguns desses parâmetros são inicializados automaticamente para tornar o algoritmo mais prático. Esses valores padrões são apresentados abaixo.

- *Number of threads*: toda a capacidade do computador;
- *Number of Random Walks*: o número de dados amostrais mais 2.000 (é adicionado também o número de pontos auxiliares definidos, se for o caso); o valor de 2.000 é adicionado para evitar que em casos com poucas amostras sejam gerados passeios aleatórios insuficientes;

- *Lag of the paths*: a dimensão mais curta da célula do grid (X, Y ou Z) dividida por 50; isso é feito para melhor discretizar os valores das distâncias e não gerar muitos arredondamentos;
- *Scale factor*: valor inicial de 0,5 e realiza a calibragem automaticamente;
- *Minimum sample distance for SGS*: o elipsoide é definido pelo comprimento de cada eixo da célula do grid multiplicado por 3.

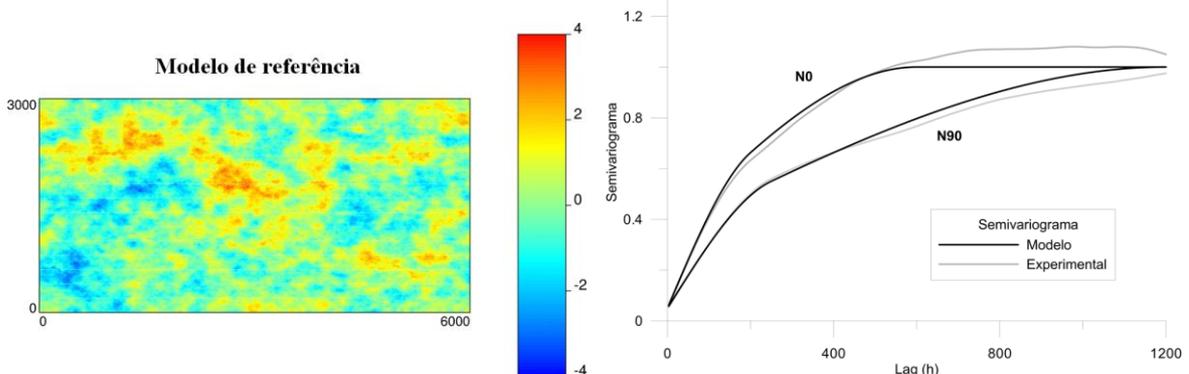
4. APLICAÇÃO EM ESTUDO DE CASO CONTROLADO

O primeiro estudo de caso para ilustrar a técnica foi feito em 2D e se baseou em um modelo de referência previamente simulado, de modo a poder melhor avaliar os resultados obtidos pelas simulações. Os detalhes desse modelo de referência estão na Tabela 1 e na Figura 11. Esse modelo de referência foi gerado via simulação espectral não-condicional (SILVA E CUBA, 2013), de modo a gerar valores multigaussianos com a continuidade espacial especificada.

Tabela 1 - Detalhes do modelo de referência simulado

Dimensões do grid	6000 x 3000 m
Dimensões de cada célula	3,125 x 3,125 m
Variograma	$\gamma = 0,05 + 0,3Sph\left(\frac{EW}{250m}, \frac{NS}{200m}\right) + 0,65Sph\left(\frac{EW}{1200m}, \frac{NS}{600m}\right)$

Figura 11 - À esquerda, o mapa do modelo de referência. À direita, o semivariograma desses dados.



Este modelo foi amostrado regularmente de duas maneiras: a cada 25 m e a cada 100 m. Foram assim criados dois bancos de dados, cada qual utilizado em uma das etapas desse estudo de caso, apresentadas a seguir. A partir desse capítulo, serão utilizadas as siglas de SMPA em referência à Simulação via Múltiplos Passeios Aleatórios, SBR para Simulação por Bandas Rotativas e SSG para Simulação Sequencial Gaussiana.

- 1ª ETAPA: foi utilizado o banco de dados com amostras a cada 25 m e a SMPA foi executada sem utilizar pontos auxiliares; resultados via SBR e SSG foram gerados para realizar algumas comparações.
- 2ª ETAPA: foi utilizado o banco de dados com amostras a cada 100 m e a SMPA foi executada utilizando pontos auxiliares; resultados via SBR e SSG também foram gerados para algumas comparações.

Foram simuladas 100 realizações via cada método com 20 processadores. A SMPA foi aplicada utilizando o algoritmo apresentado no capítulo anterior. Para a SBR, o software Isatis® foi utilizado com o mesmo número de processadores. A SSG foi aplicada no SGeMS, paralelizada à nível de realização (5 realizações por processador).

4.1. 1ª ETAPA

Nesta primeira etapa, foi emulado um cenário de curto-prazo, com bastante amostras próximas entre si. Assim, o modelo completo de 1.843.200 pontos foi simulado utilizando as 29.161 amostras a cada 25 metros.

O primeiro aspecto aqui apresentado é quanto à parte visual. Foi ampliada uma região de 500 por 500 m do grid para melhor visualização dos detalhes (Figura 12). Nota-se que as feições visuais de uma realização obtida por SMPA são similares às feições do modelo exaustivo e que não há artefatos visuais. A variância restaurada pelos passeios aleatórios nesse caso foi de 11,5% do total.

A reprodução dos variogramas e histogramas também foram boas, conforme mostrado na Figura 13. Além disso, as curvas teor-tonelagem das simulações foram semelhantes à curva referente ao modelo exaustivo, como mostrado na Figura 14.

Figura 12 - Detalhes visuais em uma janela de 500 por 500 m do modelo simulado por SMPA em comparação com o modelo de referência.

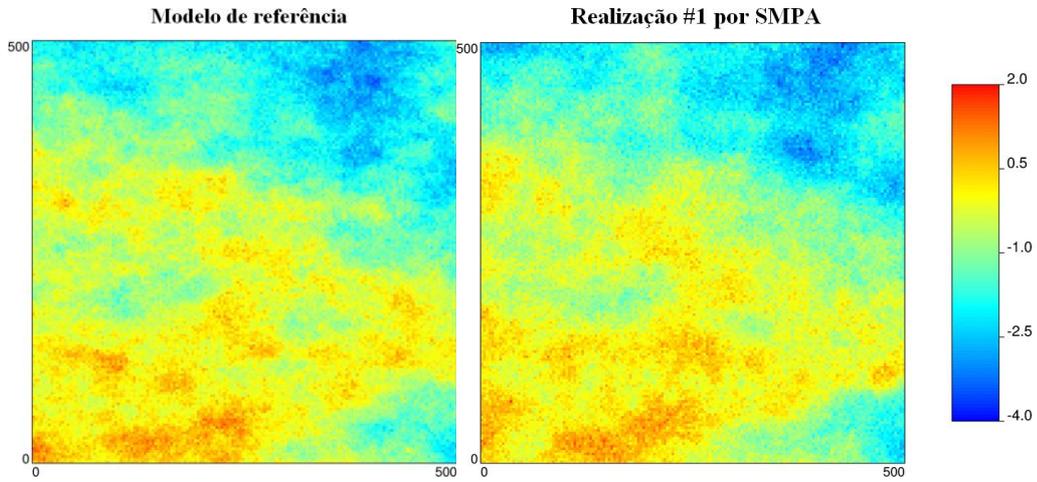


Figura 13 - Reprodução do semivariograma (à esquerda) e do histograma (à direita) para as simulações obtidas por SMPA.

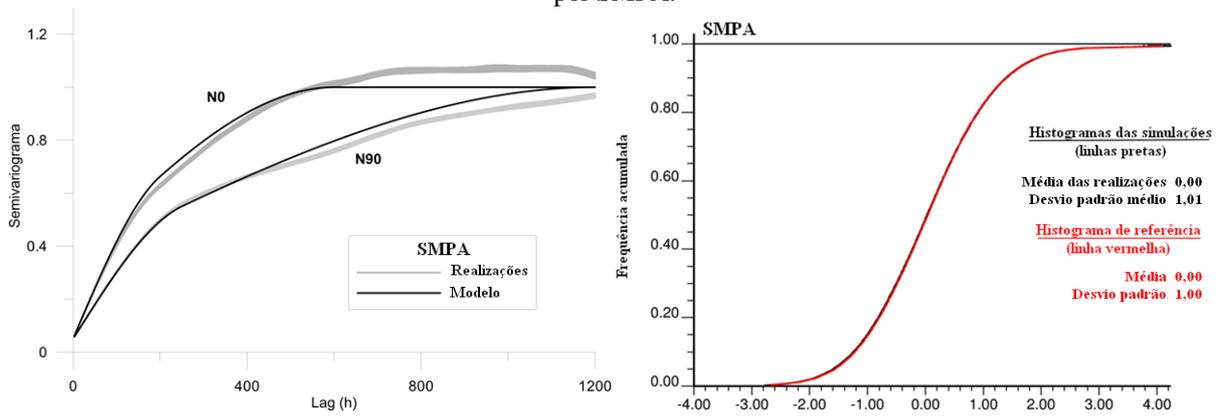
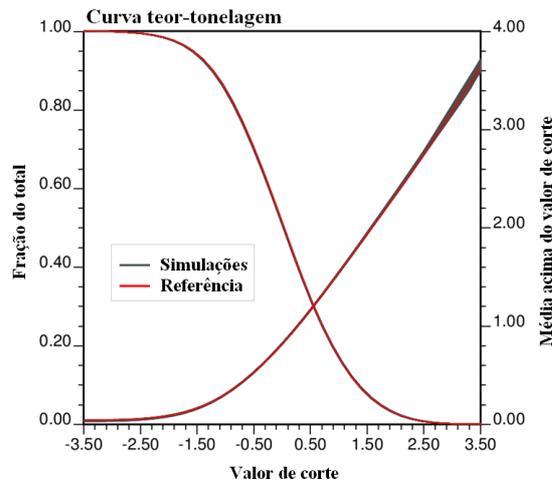
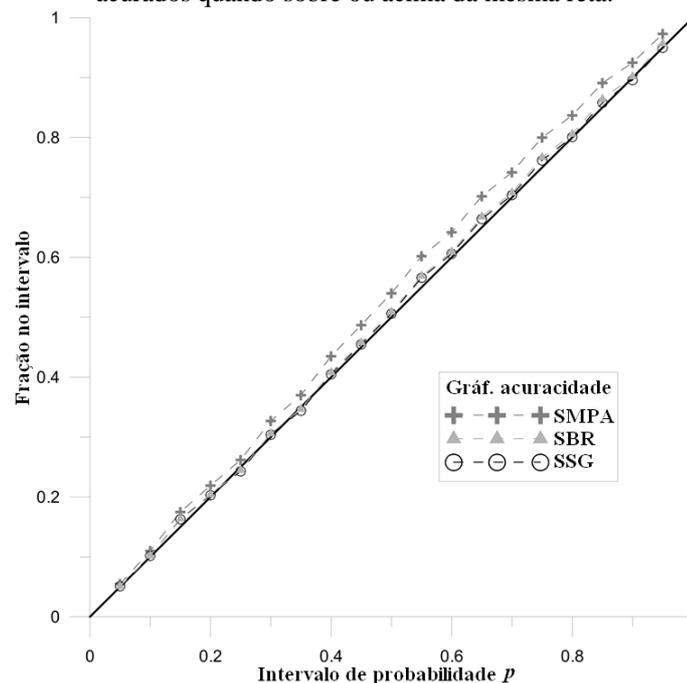


Figura 14 - Curvas teor-tonelagem das simulações via SMPA e do modelo de referência.



Foi analisada também a reprodução do espaço de incerteza das simulações no gráfico de acuracidade (Figura 15). É possível observar que, para a SMPA, o espaço de incerteza abrange os valores de referência acuradamente, mas é menos preciso que os espaços de incerteza da SBR e SSG – embora a diferença seja bem pequena para os intervalos de probabilidade mais altos.

Figura 15 - Gráfico de acuracidade para cada conjunto de simulações comparados aos dados de referência. Os intervalos de probabilidade são mais precisos quanto mais próximos os pontos estiverem da reta de 45° e acurados quando sobre ou acima da mesma reta.



O maior diferencial da SMPA é quanto à performance computacional em um ambiente de multiprocessadores. A Tabela 2 mostra o tempo decorrido para 100 simulações utilizando 20 processadores. O ganho de tempo da SMPA em comparação com os outros métodos foi de 7-9 vezes, o que é uma aceleração bastante considerável. Era esperado que a SBR apresentasse melhores resultados nesse caso, mas mostrou uma perda de eficiência ao lidar com um banco de dados com mais de 25.000 amostras.

Tabela 2 - Tempo decorrido para realizar 100 simulações via cada um dos métodos. Utilizado 20 processadores de um Intel(R) Xeon(R) CPU E5-2630 0 @ 2.30GHz 2x Hexacore.

Método	Tempo decorrido	Aumento da performance utilizando SMPA
SMPA	63,541 s	1,00x
SBR	575,313 s	9,05x
SSG	449,139 s	7,07x

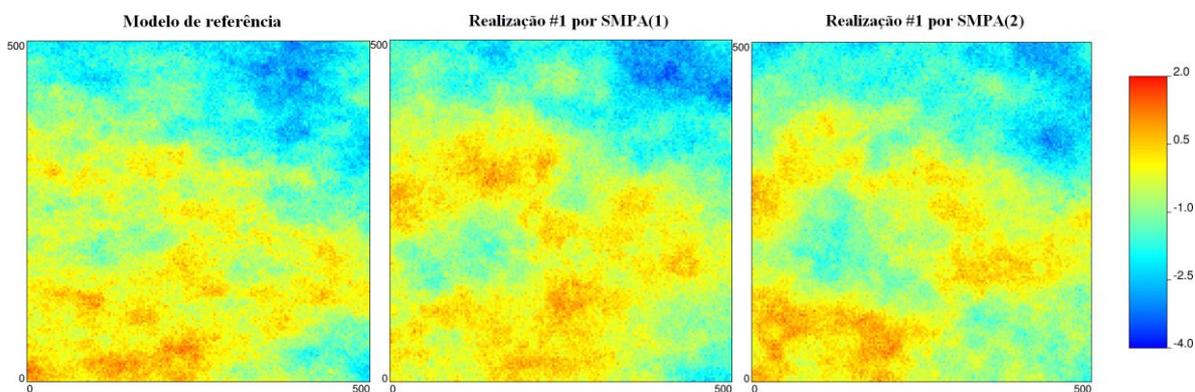
4.2. 2ª ETAPA

Nessa segunda parte do estudo, foi emulado um cenário de médio a longo-prazo, com menos dados à disposição (1.891 amostras). Nesse caso, foi necessário utilizar pontos auxiliares no processo de SMPA para evitar os artefatos e produzir modelos com a qualidade necessária.

Foram simulados dois cenários via SMPA, cada um com diferente número de pontos auxiliares. O chamado de SMPA(1) utiliza de pontos auxiliares gerados para cada 20 m, quando necessários, totalizando 25.825 pontos auxiliares por realização. Já o SMPA(2) utiliza de pontos auxiliares a cada 30 m, num total de 11.449 pontos auxiliares por realização.

Sem pontos auxiliares, os passeios aleatórios teriam que restaurar 25% da variância total. O SMPA(1) teve que restaurar apenas 13%, enquanto o SMPA(2) cerca de 15%. Dessa maneira, os resultados apresentaram aspecto visual coerente, tal como representado na Figura 16.

Figura 16 - Detalhes visuais em uma janela de 500 por 500 m dos dois modelos simulados por SMPA em comparação com o modelo de referência.



As reproduções do histograma e semivariograma nos dois casos foi boa, assim também como a reprodução da curva de teor-tonelagem dos dados de referência. Esses resultados estão representados nas Figuras 17-20.

Figura 17 - Reprodução do semivariograma (à esquerda) e do histograma (à direita) para as simulações obtidas por SMPA(1)

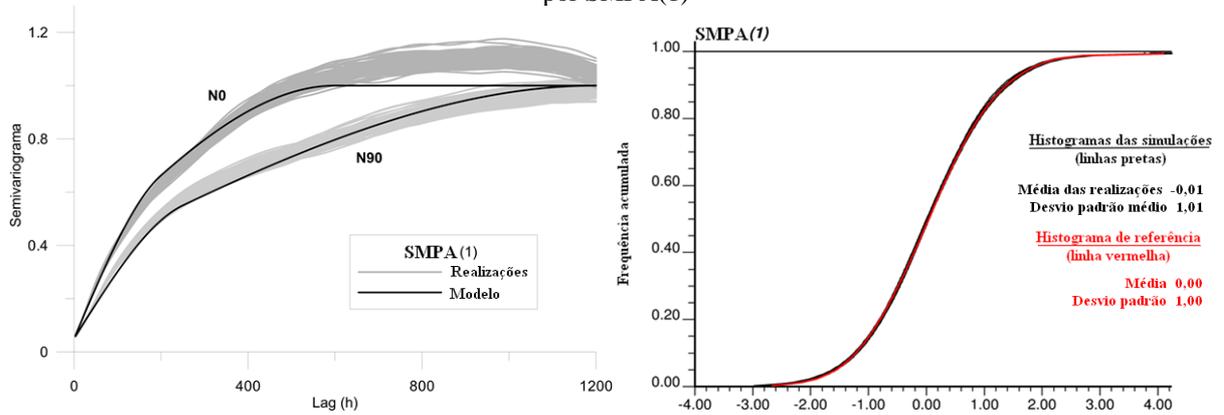


Figura 18 - Curvas teor-tonelagem das simulações via SMPA(1) e do modelo de referência.

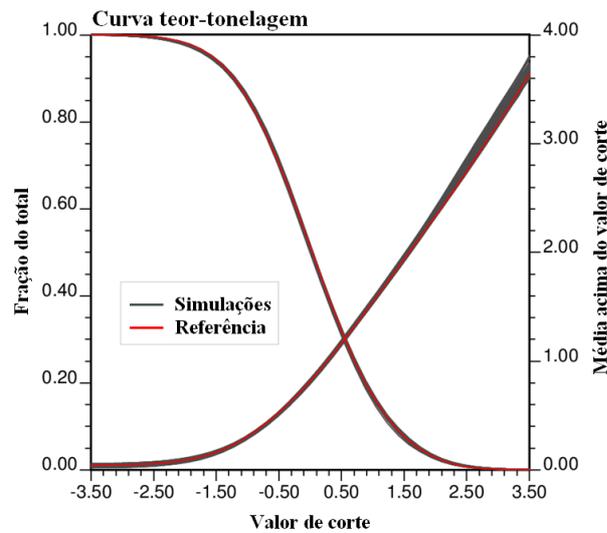


Figura 19 - Reprodução do semivariograma (à esquerda) e do histograma (à direita) para as simulações obtidas por SMPA(2).

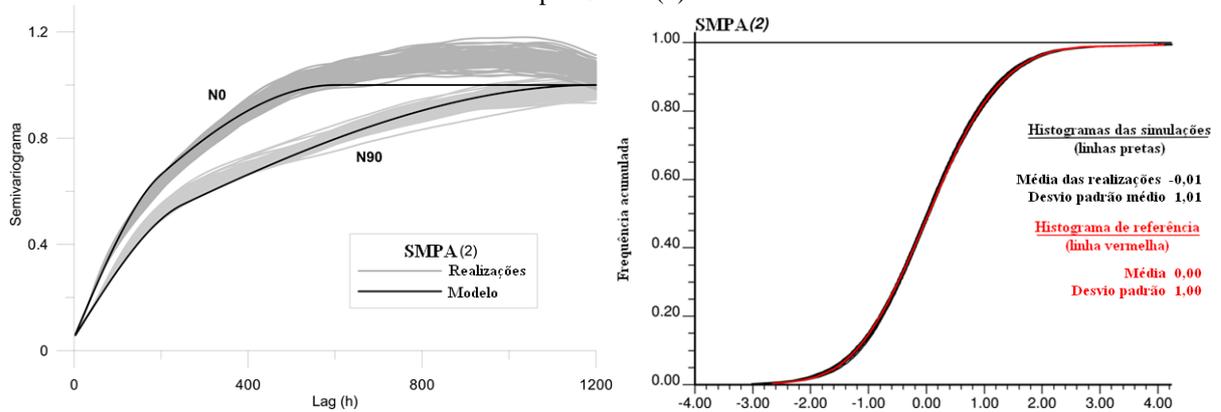
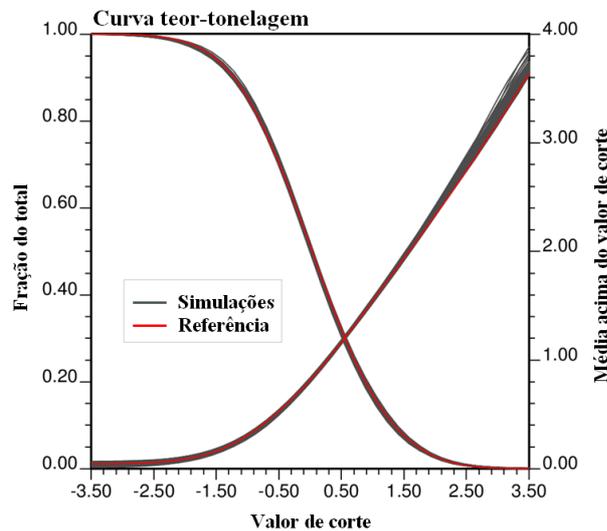


Figura 20 - Curvas teor-tonelagem das simulações via SMPA(2) e do modelo de referência



Quanto à reprodução do espaço de incerteza das simulações, é possível observar pelo gráfico de acuracidade (Figura 21), que os resultados são similares ao da etapa anterior, com uma ligeira perda de precisão em comparação com os outros métodos. Uma observação é que os resultados de SMPA(1) e SMPA(2) foram os mesmos nesse teste e foram representado unicamente como SMPA.

O ganho de tempo com SMPA nessas condições foi menor em relação ao caso anterior (Tabela 3). Ainda assim, gerou uma economia de tempo considerável, entre 1,81x a 2,69x em comparação com SBR e SSG, respectivamente. Há partir dessas informações foi possível observar como a quantidade de pontos auxiliares simulados influencia a velocidade da SMPA. Uma diferença de cerca de 15.000 pontos auxiliares causou uma diferença de tempo de 1,57x

entre SMPA(1) e SMPA(2). Uma maneira de melhorar essa performance seria utilizar um método mais rápido que a SSG para simular os pontos auxiliares.

Figura 21 - Gráfico de acuracidade para cada conjunto de simulações comparados aos dados de referência. SMPA(1) e (2) geraram os mesmos resultados e foram representados unicamente como SMPA. Os intervalos de probabilidade são mais precisos quanto mais próximos os pontos estiverem da reta de 45° e acurados quando sobre ou acima da mesma reta

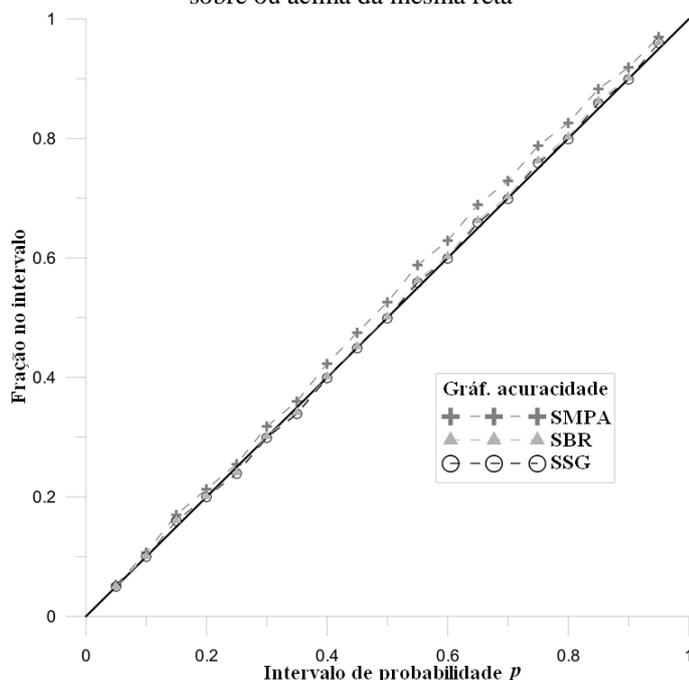


Tabela 3 - Tempo decorrido para realizar 100 simulações via cada um dos métodos. Utilizado 20 processadores de um Intel(R) Xeon(R) CPU E5-2630 0 @ 2.30GHz 2x Hexacore

Método	Tempo decorrido	Aumento da performance	
		utilizando SMPA(1)	utilizando SMPA(2)
SMPA(1)	267,698 s	1,00x	1,57x
SMPA(2)	170,494 s	0,64x	1,00x
SBR	308,121 s	1,15x	1,81x
SSG	458,427 s	1,71x	2,69x

Em suma, esse estudo de caso demonstrou que as simulações geradas por SMPA são satisfatórias, tanto no aspecto gráfica como estatístico. E que a sua utilização proporcionou ganhos consideráveis de velocidade, especialmente nos casos que há maior adensamento amostral.

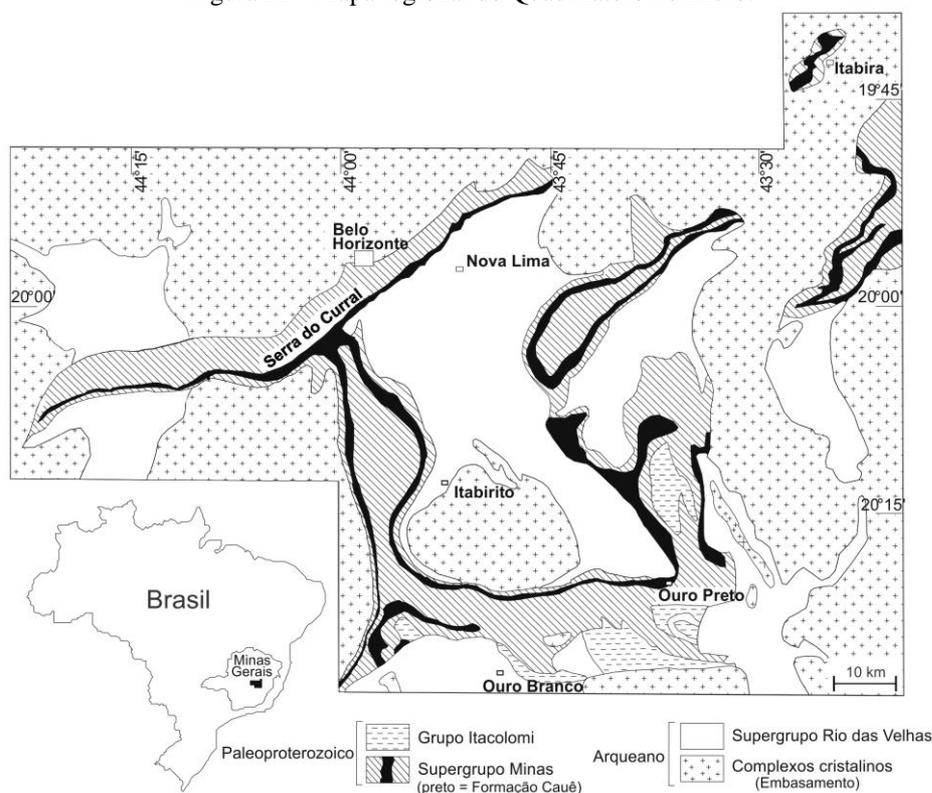
5. APLICAÇÃO EM ESTUDO DE CASO REAL

O segundo estudo de caso se trata de uma mina de ferro localizada no estado de Minas Gerais. Informações detalhadas da mina são omitidos por questão de sigilo solicitado pela Vale S/A, detentora dos direitos minerários. Sendo assim, será feita apenas uma apresentação do contexto geológico regional, seguido da análise dos dados e simulação do depósito.

5.1. GEOLOGIA REGIONAL

O depósito está localizado na região do Quadrilátero Ferrífero no estado de Minas Gerais (Figura 22), um dos maiores complexos ferríferos do mundo. O minério de ferro está localizado estratigraficamente na Formação Cauê, do Grupo Itabira, de idade Proterozóica (~2,52 Ga – BABINSKI *ET AL.*, 1991), nos denominados itabiritos e suas rochas derivadas. Itabirito é o nome dado à formação ferrífera bandada local, caracterizada por bandas de quartzitos intercaladas por bandas de minerais ferríferos – principalmente hematita (Fe_2O_3), goethita ($\text{FeO}(\text{OH})$) e magnetita (Fe_3O_4).

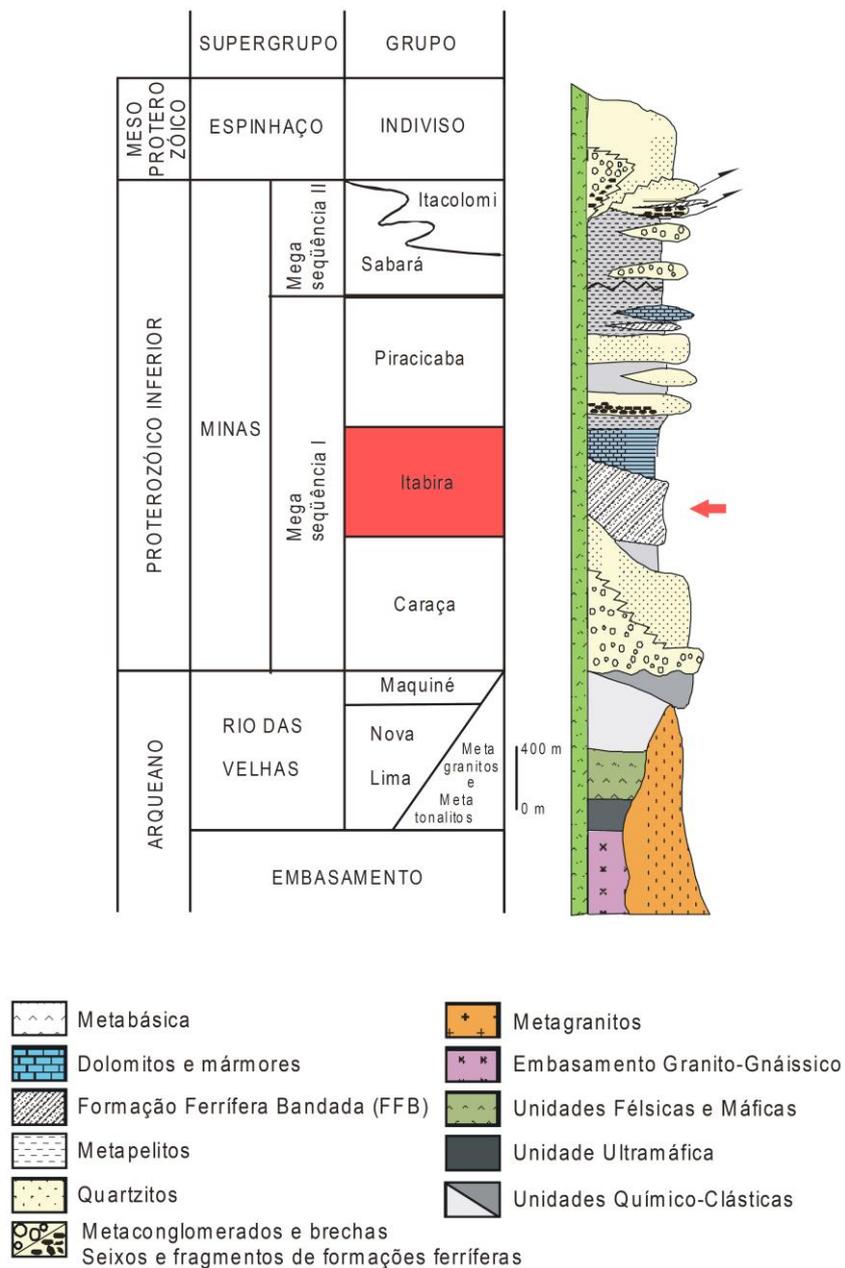
Figura 22 - Mapa regional do Quadrilátero Ferrífero.



Fonte: Modificado de Sanglard *et al.* (2014).

O Grupo Itabira (Figura 23) consiste em uma unidade basal onde ocorrem os itabiritos (Formação Cauê, podendo apresentar até 300 m de espessura) e uma unidade superior (Formação Gandarela), composta principalmente por dolomitos. O contato entre essas formações frequentemente é brusco e intercalado, tanto verticalmente como lateralmente (ROSIÈRE E CHEMALE JR, 2000).

Figura 23 - Coluna estratigráfica mostrando, em vermelho, a localização do minério de ferro do Quadrilátero Ferrífero.



Fonte: Modificado de Rosière e Chemale Jr (2000).

As formações ferríferas bandadas do Proterozóico consistem nas maiores concentrações ferríferas já identificadas na superfície terrestre. São póstumos a um período em que o oxigênio era praticamente ausente e os oceanos mais quentes e ácidos (ARNDT E GANINO, 2012). Essas condições favoreceram uma maior solubilidade de Si e Fe nos oceanos, provenientes tanto de erosão dos continentes como de fontes exalativas vulcânicas. Acredita-se que o crescente aumento da concentração de oxigênio no planeta alterou as condições físico-químicas desses oceanos primitivos, favorecendo a precipitação de uma enorme quantidade de óxidos desses elementos em plataformas continentais no período entre 2,6-2,0 Ga.

No caso do Quadrilátero Ferrífero, essas formações passaram por dois eventos de deformação e metamorfismo regional: o Transamazônico (entre 2,06-2,03 Ga) e o Brasileiro (entre 0,6-0,5 Ga), sofrendo metamorfismo sob fácies xisto verde baixa na parte oeste e fácies anfíbolito/granulito na parte leste. Além disso, os itabiritos estão em grande parte marcados por feições de laterização, o que favoreceu um maior enriquecimento supérgeno do ferro em vários locais.

5.2. ANÁLISE PRELIMINAR DOS DADOS

Foram disponibilizados dados de 1.378 furos de sondagem (Figura 24), dos quais 490 são provenientes da exploração de longo prazo e 888 do curto prazo. Os dados dos furos foram fornecidos já regularizados a cada 10 metros e foram utilizadas todas as compostas entre 3-10 m (histograma dos comprimentos das compostas utilizadas na Figura 25), totalizando 5.117 amostras dos furos de longo-prazo e 3.193 amostras dos furos de curto-prazo. Como ambos os dados de longo e de curto prazo aqui utilizados são provenientes de sondagens diamantadas e igualmente validados, todas essas informações foram reunidas em um único banco de dados para as posteriores simulações.

Figura 24 - Mapa de localização da boca dos furos de sondagem e uma vista isométrica.

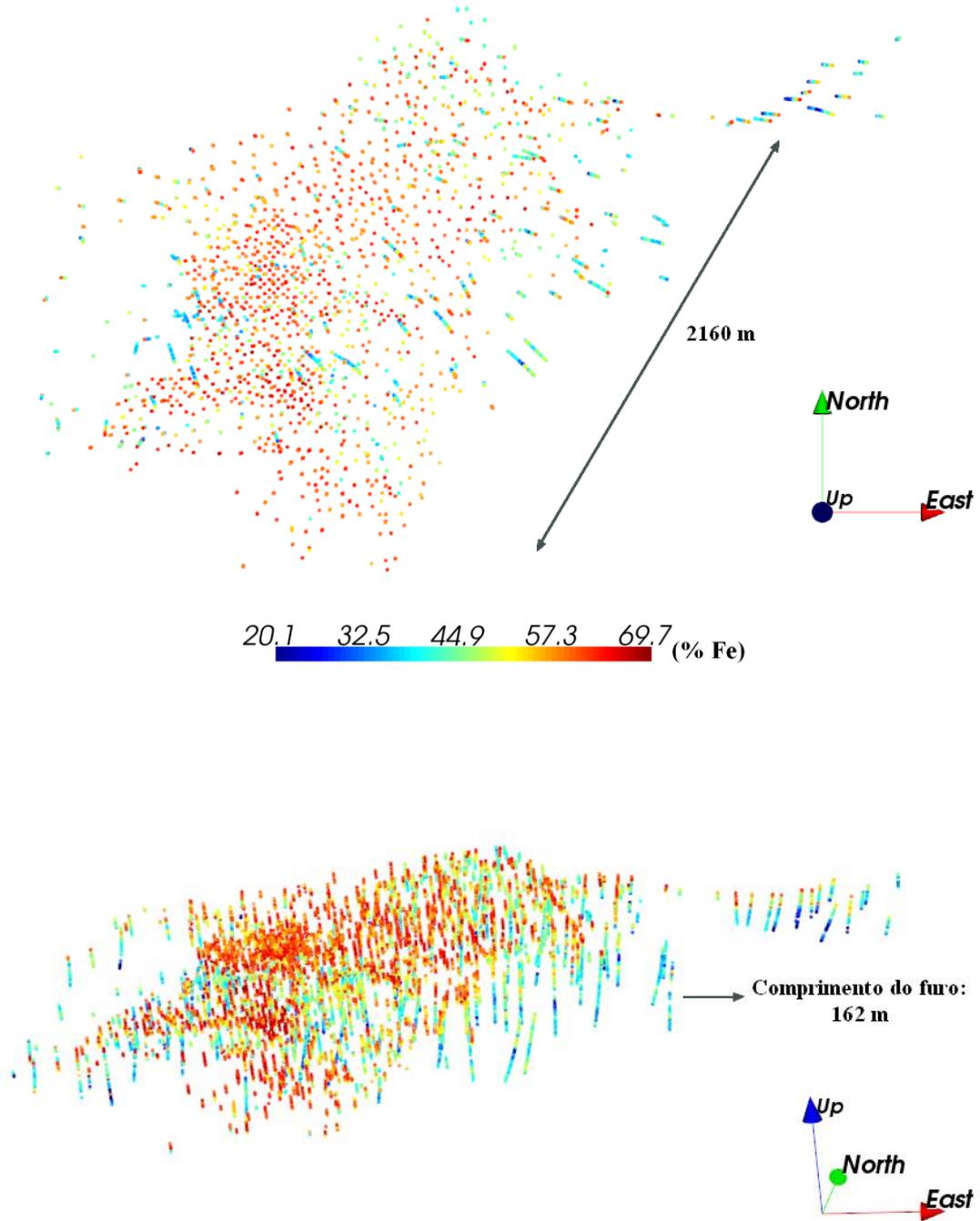
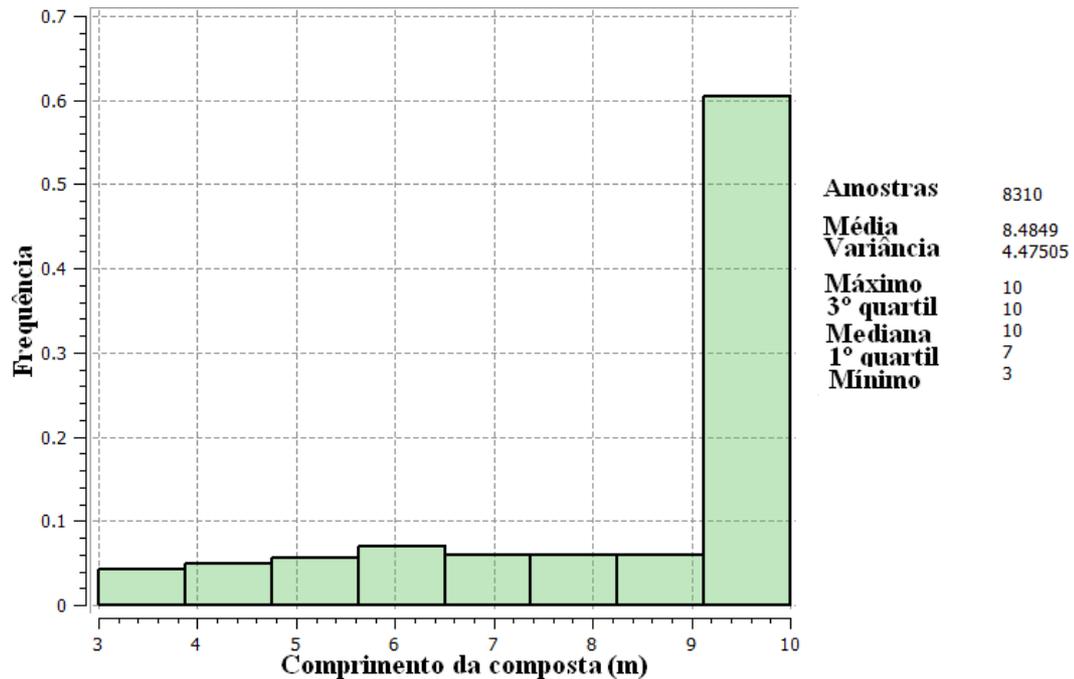


Figura 25 - Comprimentos das amostras regularizadas.

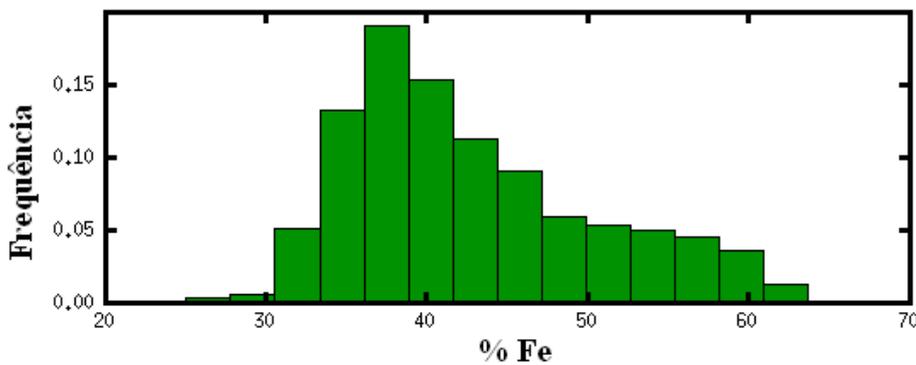
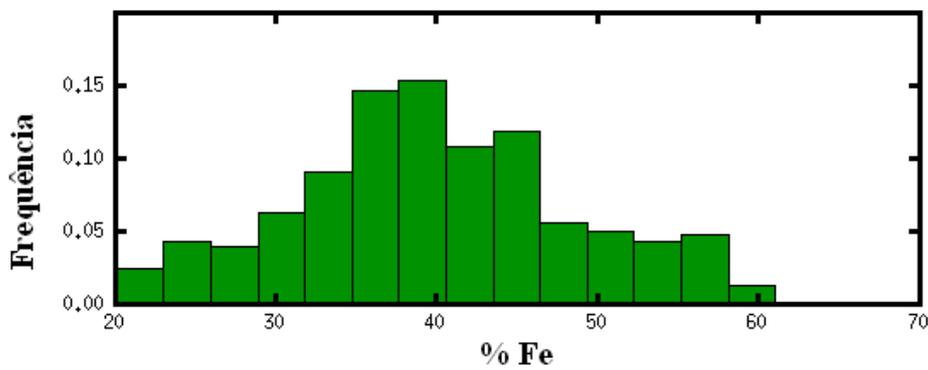
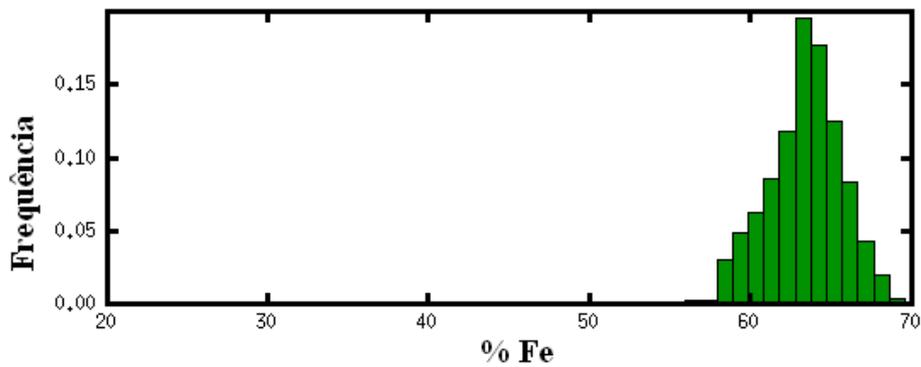
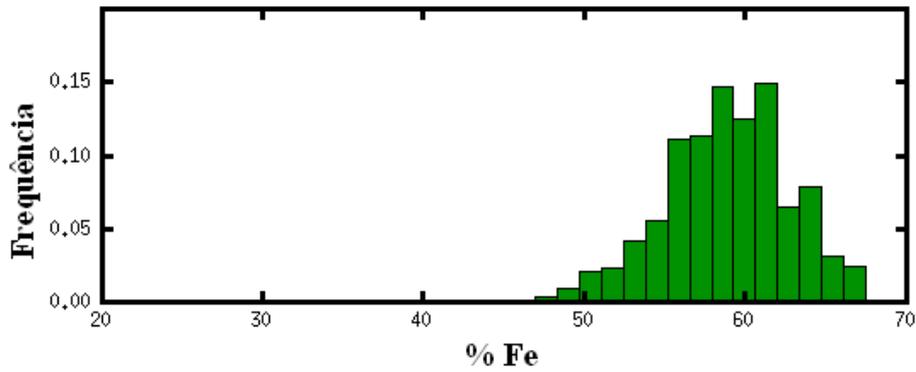


Essas diversas amostras compreendem diferentes litologias. Como forma de criar sub-regiões estacionárias quanto à variável de interesse (que no caso é apenas o teor de ferro global), os dados foram separados em quatro domínios:

- Domínio 1: Canga (484 amostras)
- Domínio 2: Hematitito (1.823 amostras)
- Domínio 3: Itabirito contaminado (1.336 amostras)
- Domínio 4: Itabirito silicoso (4.667 amostras)

Foi realizado o desagrupamento dos dados de cada domínio via o método de áreas/volumes de influência. As distribuições obtidas após esse processo estão apresentadas na Figura 26. Foi fixada a mesma escala e mesmo número de classes dos histogramas para ressaltar as diferenças nas distribuições do teor de ferro em cada domínio. A canga e o hematitito apresentam baixa variância dos dados e predominantemente altos teores, enquanto os itabiritos apresentam médias menores e variâncias mais altas. As distribuições apresentam uma simetria visível, com exceção do itabirito silicoso que apresenta uma ligeira assimetria.

Figura 26 - Histogramas de cada um dos quatro domínios após o desagrupamento amostral.



O modelo geológico do depósito foi também fornecido pela Vale S/A e utilizado para definir os blocos pertencentes a cada domínio. Primeiramente, foi definida uma envoltória próxima aos dados para evitar grandes extrapolações além da fronteira definida pelas amostras. Em seguida, os blocos pertencentes a cada domínio foram separados nas respectivas regiões de acordo com as litologias definidas no modelo geológico.

5.3. SIMULAÇÃO DO DEPÓSITO

O tamanho de bloco utilizado na mina é de 12,5m x 12,5m x 10m. Para as simulações eles foram discretizados em duas partes por eixo, totalizando 8 pontos por bloco. Foram realizadas 100 simulações para cada um dos quatro domínios utilizando a SMPA. Pontos auxiliares foram utilizados nos locais do grid sem amostras a um raio de 20 m na horizontal e 10 m na vertical. Os detalhes estão apresentados na Tabela 4.

Tabela 4 - Detalhes da quantidade de pontos auxiliares simulados e do total de nós simulados.

Domínio	Nº de pontos auxiliares simulados	Total de nós simulados
Canga	5.669	154.176
Hematítico	6.465	201.512
Itab. Contaminado	27.114	840.568
Itab. Silicoso	65.441	2.954.864
TOTAL	104.689	4.151.120

Foram utilizados os variogramas dos dados normalizados de cada domínio para realizar a simulação. A busca foi realizada sem divisão por octantes, selecionando um máximo de 40 amostras mais próximas. As dimensões do elipsóide de busca foram baseadas nos modelos variográficos interpretados. Detalhes dos parâmetros estão apresentados na Tabela 5.

Tabela 5 - Detalhes dos parâmetros variográficos utilizados para as simulações.

Variogramas utilizados nas simulações

$$\gamma_{CANGA} = 0,05 + 0,364Sph\left(\frac{N60}{56m}, \frac{N150}{90m}, \frac{Vert}{15m}\right) + 0,586Sph\left(\frac{N60}{805m}, \frac{N150}{605m}, \frac{Vert}{20m}\right)$$

$$\gamma_{HEMAT} = 0,05 + 0,45Sph\left(\frac{N60}{45m}, \frac{N150}{45m}, \frac{Vert}{5m}\right) + 0,2Sph\left(\frac{N60}{195m}, \frac{N150}{55m}, \frac{Vert}{25m}\right) + 0,3Sph\left(\frac{N60}{1300m}, \frac{N150}{900m}, \frac{Vert}{100m}\right)$$

$$\gamma_{IT.CONT.} = 0,05 + 0,37Sph\left(\frac{N60}{200m}, \frac{N150}{70m}, \frac{Vert}{120m}\right) + 0,56Sph\left(\frac{N60}{690m}, \frac{N150}{400m}, \frac{Vert}{120m}\right)$$

$$\gamma_{IT.SIL.} = 0,05 + 0,27Sph\left(\frac{N60}{50m}, \frac{N150}{80m}, \frac{Vert}{45m}\right) + 0,29Sph\left(\frac{N60}{175m}, \frac{N150}{190m}, \frac{Vert}{152m}\right) + 0,39Sph\left(\frac{N60}{1106m}, \frac{N150}{190m}, \frac{Vert}{157m}\right)$$

Dessa maneira, foram simulados 100 modelos que foram validados e reblocados para o suporte padrão da mina. Os resultados gráficos do modelo estão apresentados nas Figuras 27-29. Nota-se uma boa representação visual, sem artefatos, respeitando as características de cada domínio, com uma concentração dos altos teores na região superior do depósito.

Figura 27 - Uma realização simulada do modelo completo, exibido em perspectiva.

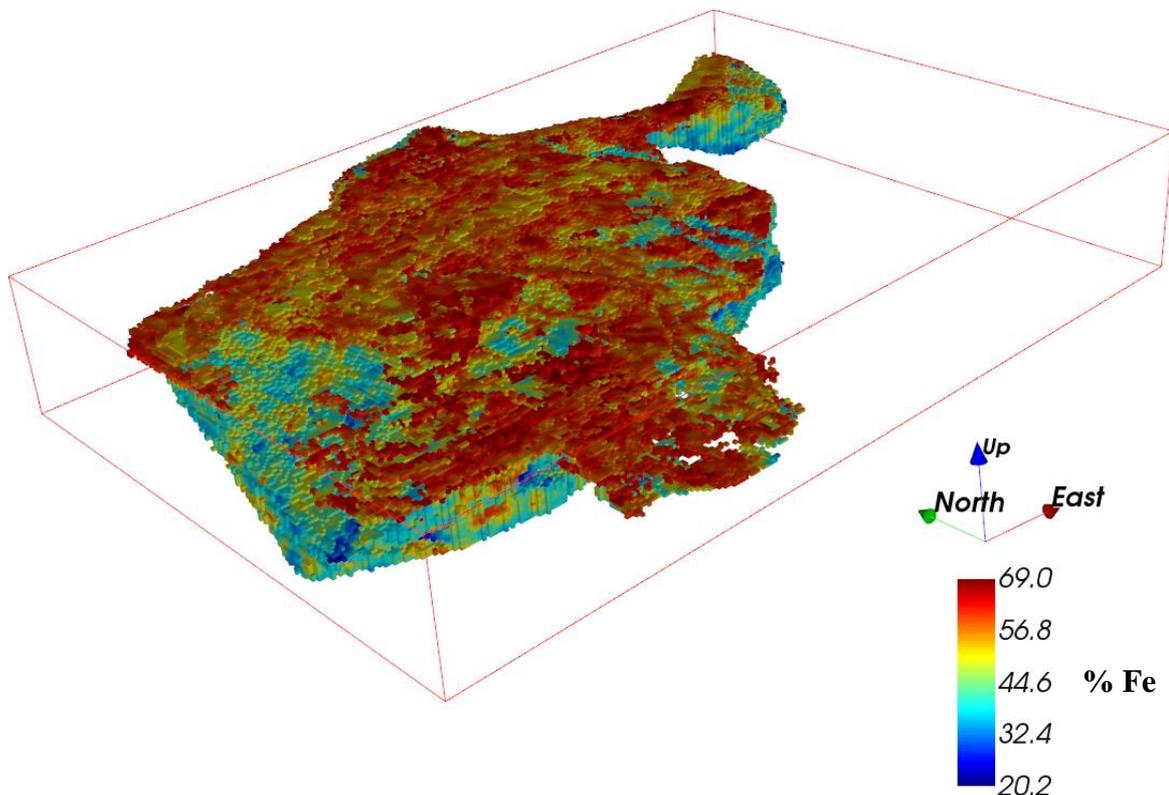


Figura 28 - Seção horizontal (na cota de 875 m) de uma realização simulada, destacando as partes correspondentes a cada um dos domínios.

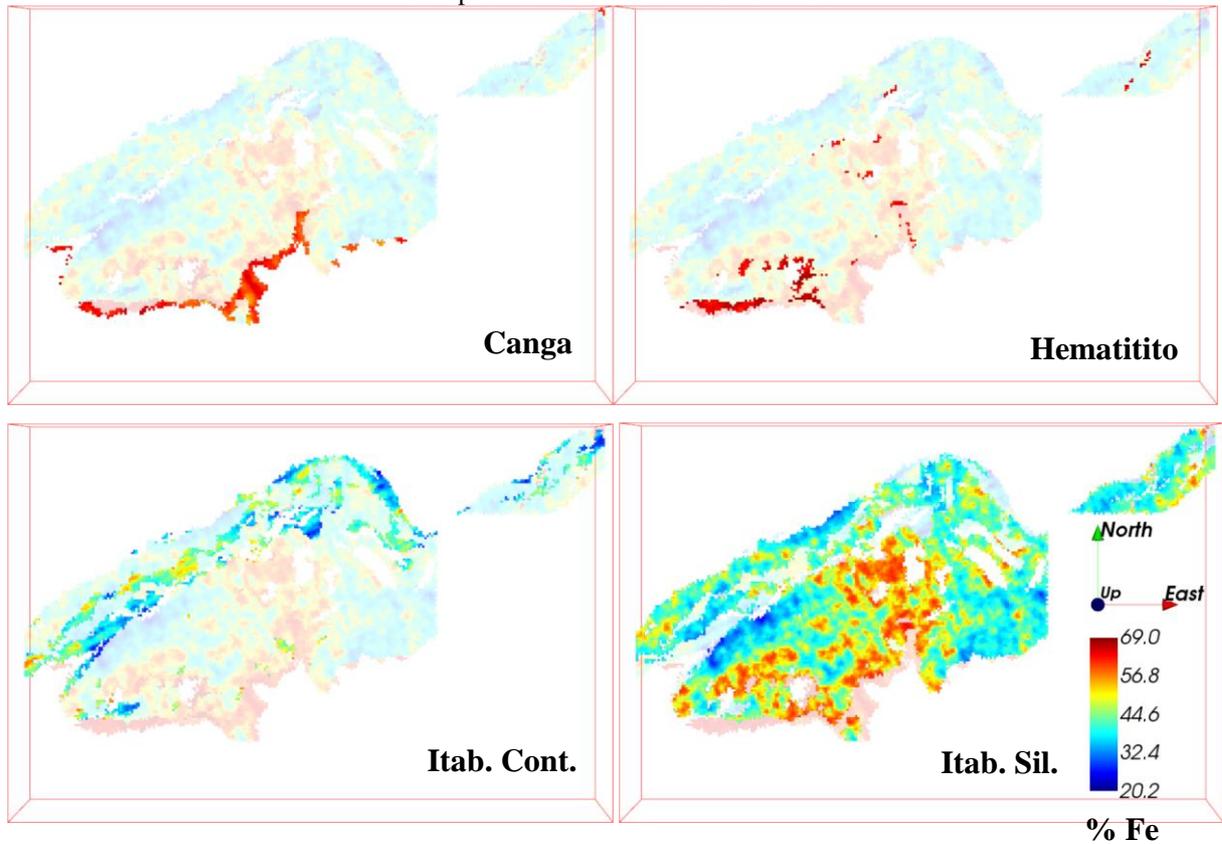
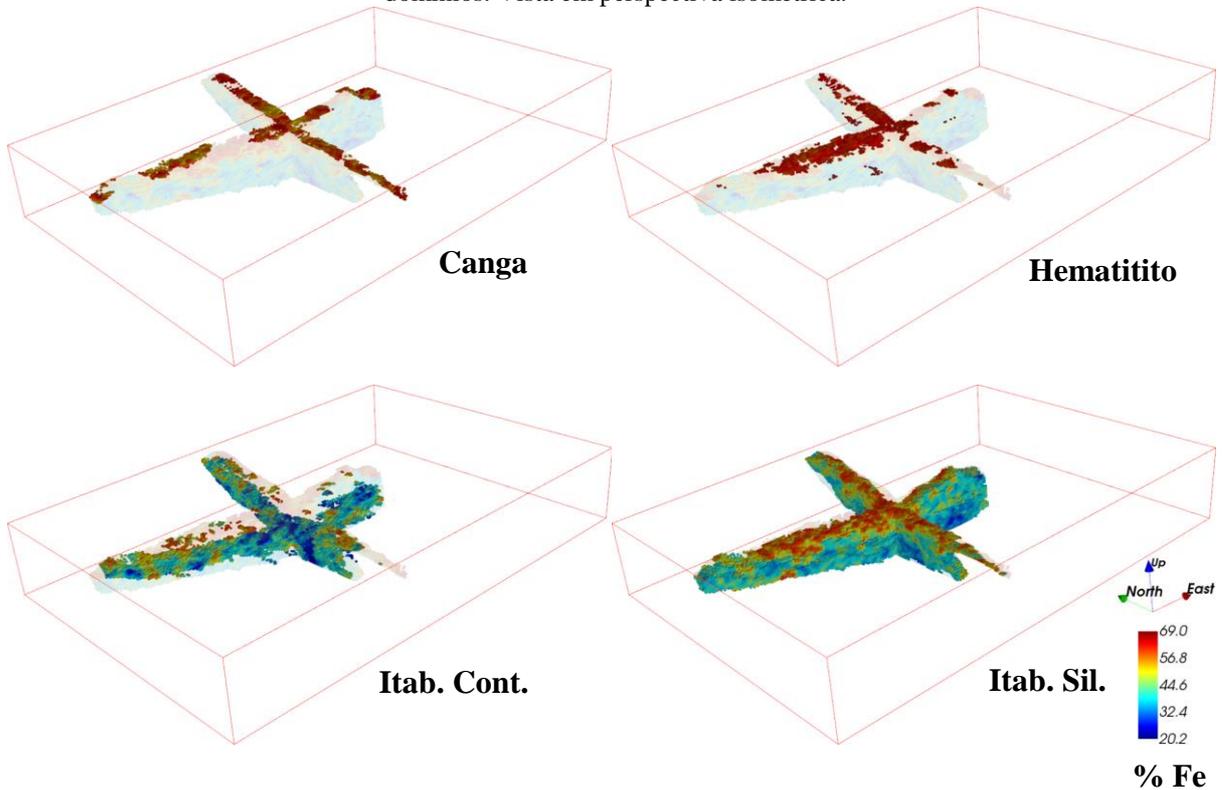
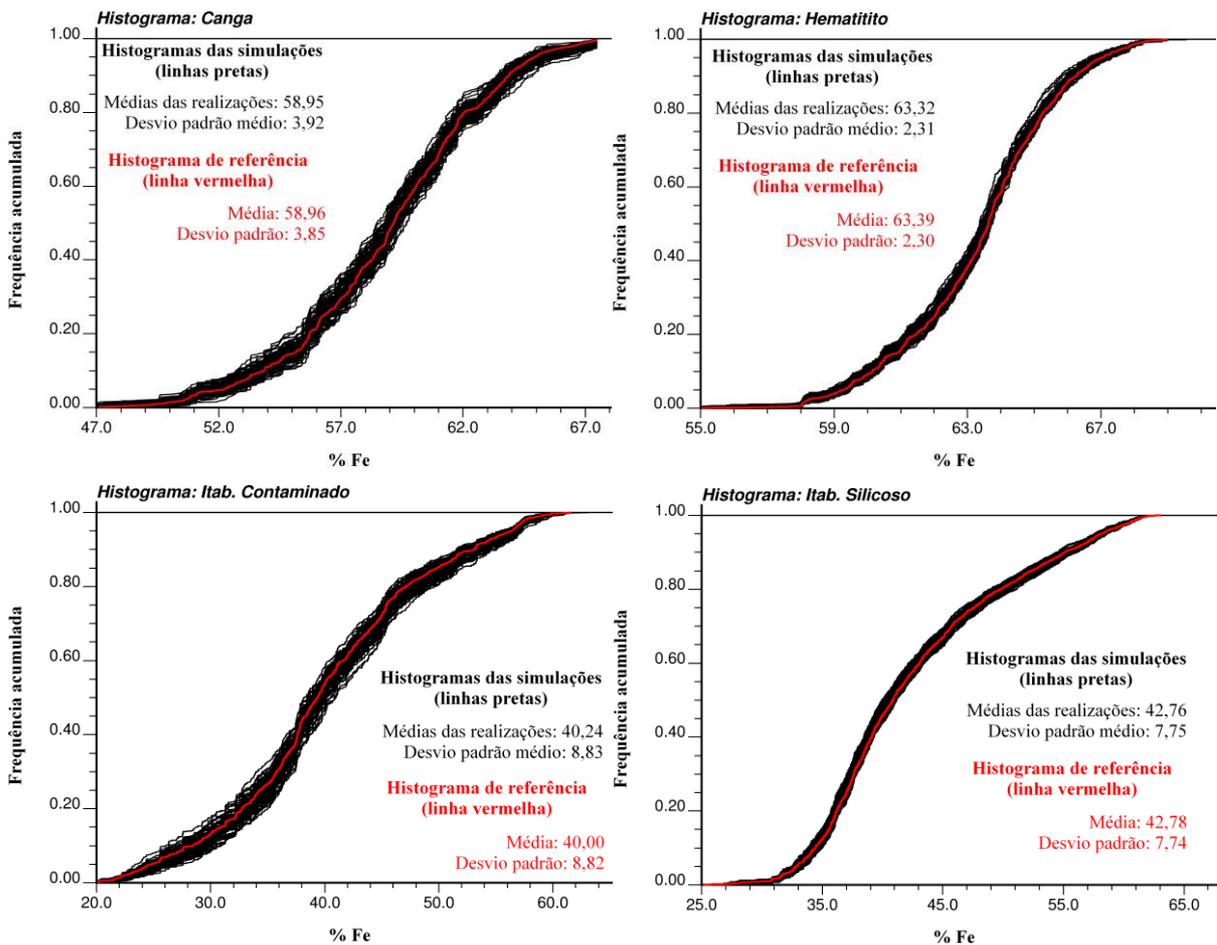


Figura 29 - Seções verticais de uma realização simulada, destacando as partes correspondentes a cada um dos domínios. Vista em perspectiva isométrica.



A validação foi feita por domínios. Os histogramas e estatísticas básicas dos modelos simulados apresentaram reprodução muito boa em comparação às distribuições dos dados originais desagrupados, tal como apresentado na Figura 30.

Figura 30 - Histogramas e estatísticas básicas das simulações de cada domínio.



Os variogramas dos domínios simulados também reproduziram bem os variogramas desagrupados experimentais dos dados, conforme representado nas Figuras 31 e 32. Uma observação quanto ao variograma experimental da canga na vertical: está apresentado apenas um ponto pois é o único mais representativo (91 pares) enquanto todos os outros apresentam menos de 20 pares.

Figura 31 - Variogramas horizontais das simulações de cada domínio.

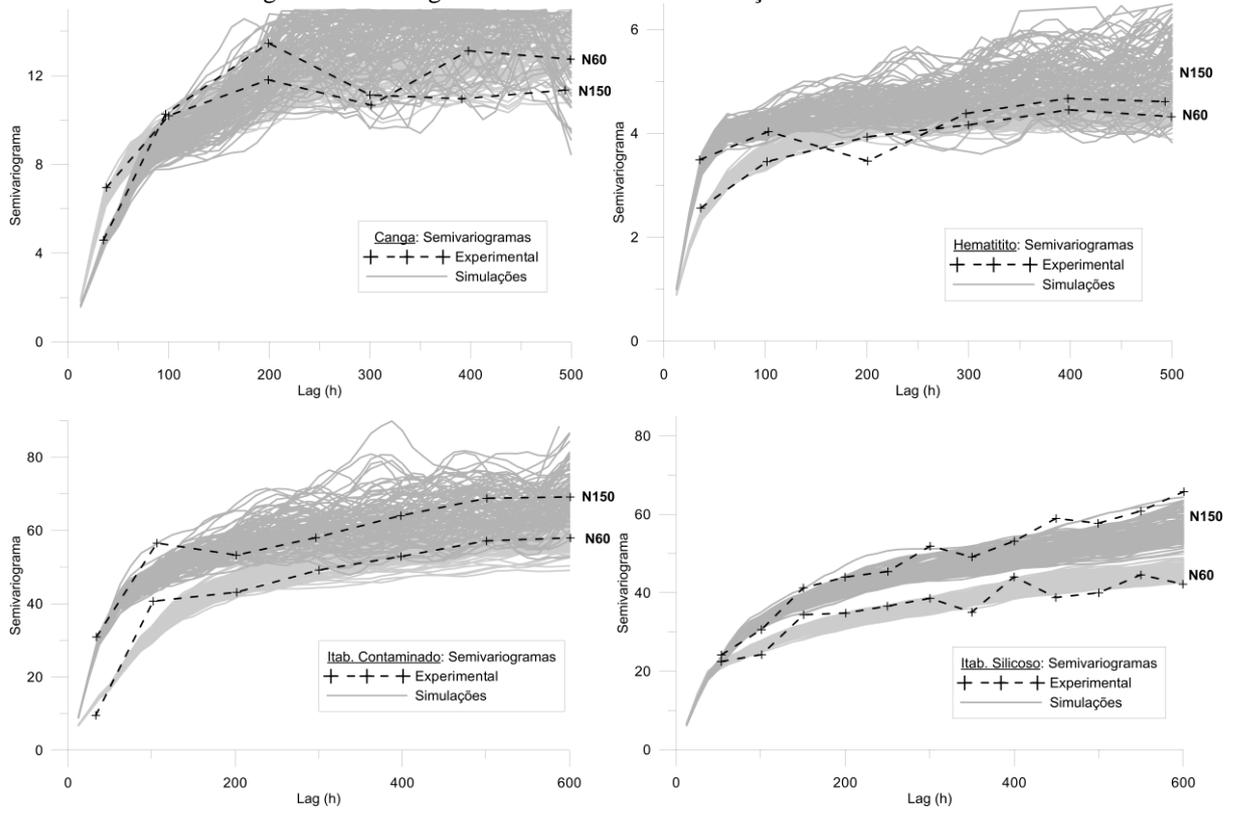
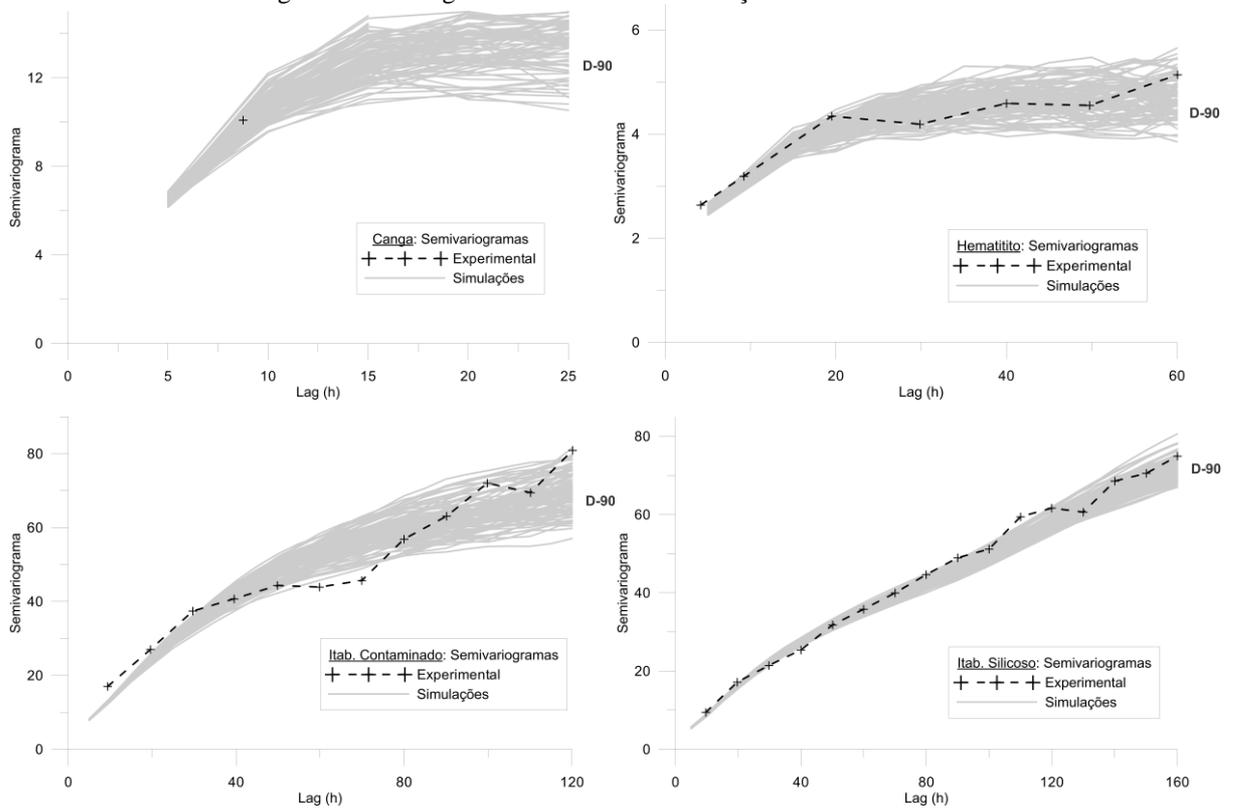
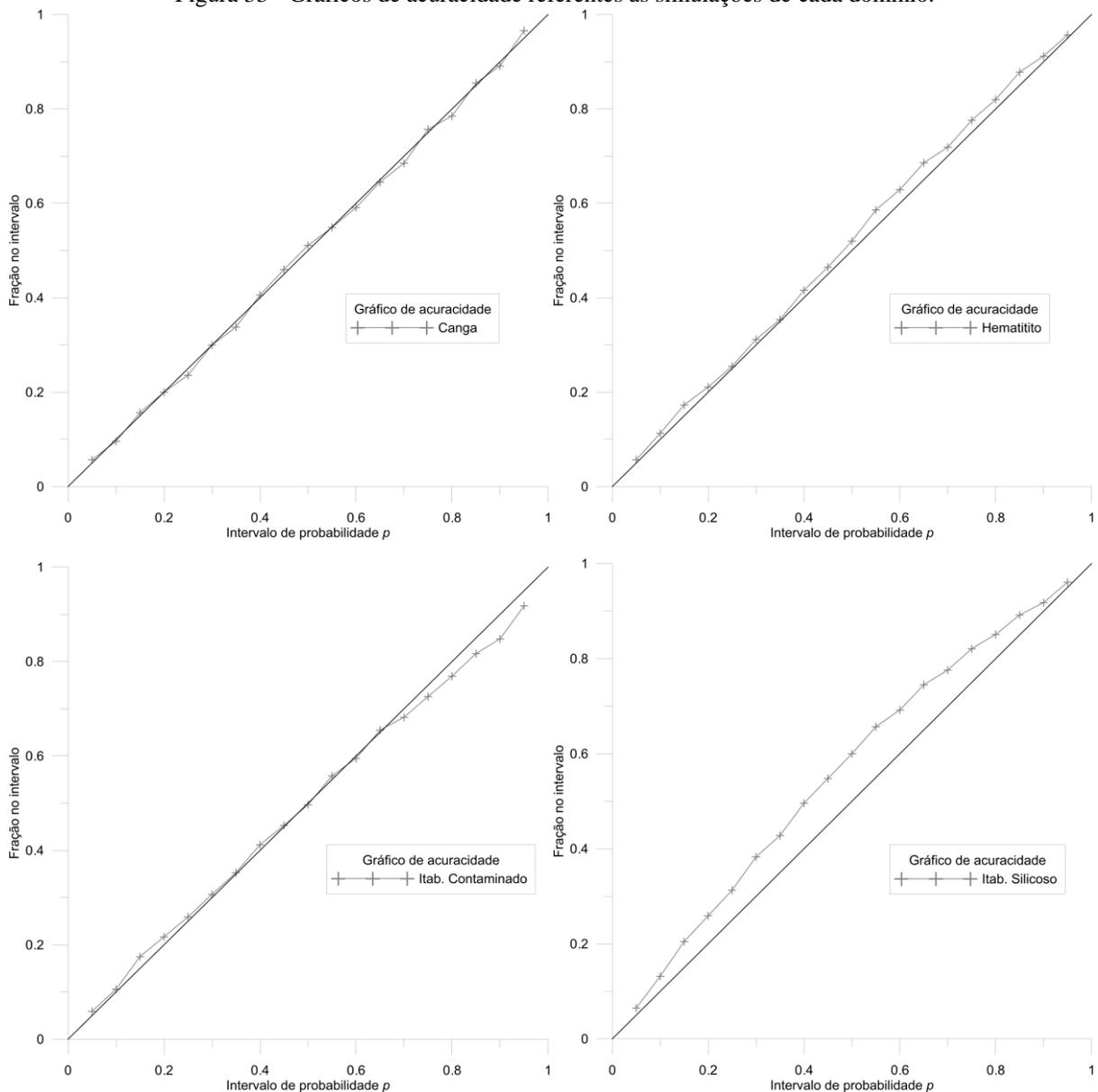


Figura 32 - Variogramas verticais das simulações de cada domínio.



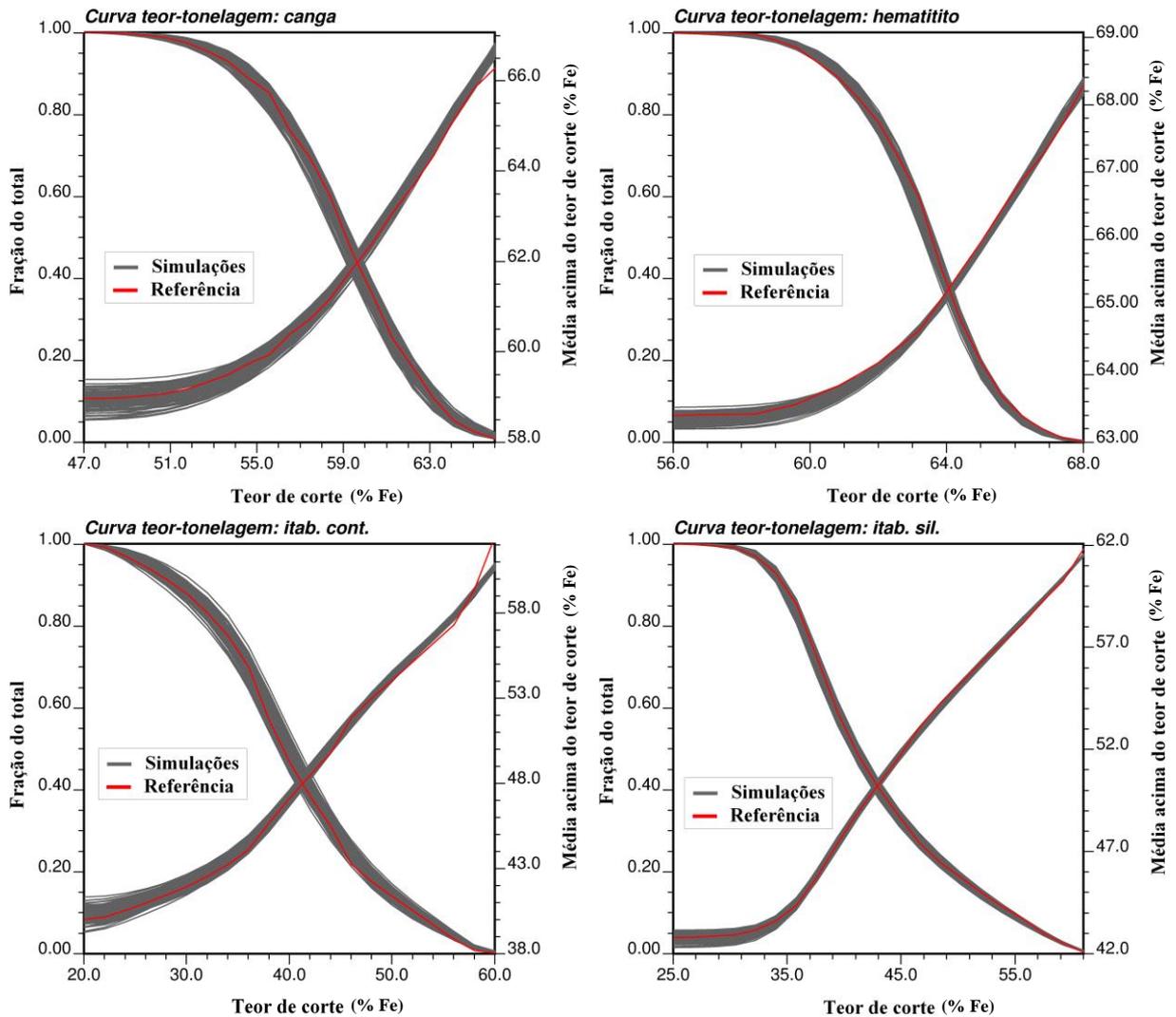
Foi avaliado também o espaço de incerteza por meio dos gráficos de acuracidade (Figura 33). Eles foram construídos a partir da comparação dos resultados de validação cruzada das simulações contra os valores reais nos pontos. A canga e o hematitito apresentaram resultados muito bons. O itabirito contaminado apresentou uma ligeira imprecisão e inacuracidade para os intervalos de probabilidade mais altos, enquanto o itabirito silicoso apresentou resultados sempre acurados, mas menos precisos para os intervalos de probabilidade centrais. Vale ressaltar que a validação cruzada não leva em conta o uso dos pontos auxiliares, o que possivelmente proporcionaria melhores resultados na acuracidade e precisão das probabilidades das simulações (espaço de incerteza).

Figura 33 - Gráficos de acuracidade referentes às simulações de cada domínio.



Foram avaliadas também a reprodução das curvas de teor tonelagem de cada domínio (Figura 34). Ela foi feita no suporte dos blocos simulados e comparada contra a distribuição de referência dos dados originais, após correção da variância para o suporte blocos. A reprodução foi bastante satisfatória, com boa aderência das curvas.

Figura 34 - Curvas de teor-tonelagem referentes às simulações de cada domínio.



Por fim, foi também avaliado o tempo de processamento da SMPA em comparação com a SSG e SBR. As simulações via SSG e SBR foram geradas apenas para medir o tempo de processamento e serem comparadas a SMPA. Não estão sob análise as reproduções dos momentos estatísticos com esses métodos, já que se tratam de técnicas consagradas. Nesse caso, foram utilizados parâmetros semelhantes aos da SMPA. A SSG foi executada no SGEMS, utilizando processamento paralelo em nível de realização. A SBR foi realizada no Isatis®,

utilizando 1.000 linhas para simulações não condicionais. Os resultados estão apresentados nas Tabelas 6 e 7.

Tabela 6 - Tempo decorrido para realizar 100 simulações via cada um dos métodos. Utilizados 20 processadores de um SGI H2106 Server (4x AMD Opteron 6376 16-core 2.3GHZ, 256GB RAM Mem, 2x 1TB Disk).

20 processadores	SMPA	SSG	SBR
Canga	99 s	294 s	2807 s
Hematitito	150 s	491 s	2935 s
Itab. Cont.	350 s	1510 s	2997 s
Itab. Sil.	2159 s	7062 s	6833 s
Tempo total	45min 58s	2h 35min 57s	4h 19min 32s
<i>Speed-up</i> SMPA	1,00	3,39	5,65

Tabela 7 - Tempo decorrido para realizar 100 simulações via cada um dos métodos. Utilizados 50 processadores de um SGI H2106 Server (4x AMD Opteron 6376 16-core 2.3GHZ, 256GB RAM Mem, 2x 1TB Disk).

50 processadores	SMPA	SSG	SBR
Canga	87 s	164 s	1586 s
Hematitito	118 s	267 s	3368 s
Itab. Cont.	280 s	823 s	2379 s
Itab. Sil.	1873 s	3830 s	4443 s
Tempo total	39min 18s	1h 24min 44s	3h 16min 16s
<i>Speed-up</i> SMPA	1,00	2,16	4,99

Esses valores demonstram uma vantagem clara da SMPA nas condições apresentadas, alcançando um *speed-up* de não menos que 2x, chegando até 5,65x no caso mais favorável. Em outras palavras, há uma economia de tempo de ao menos 45 min no caso mais próximo e de 3h 33 min no caso mais discrepante dentre os testados.

Vale ressaltar que diferentes quantidades de memória RAM disponível, número de realizações, número de processadores, estruturação do algoritmo e outros fatores influenciam fortemente nesses resultados. Havendo memória suficiente, a SSG paralelizada à nível de realização se torna mais eficiente à medida que o número de processadores se iguala ao número de realizações.

A comparação com a SBR do Isatis® é difícil de ser feita uma vez que os softwares são diferentes e o código não é fornecido. Mas o desempenho do algoritmo parece depender bastante do tamanho dos domínios. Tanto o SGeMS como o Isatis® trabalham inicialmente em

um grid em forma de um paralelepípedo 3D, dentro do qual são definidas as regiões pertencentes a cada domínio. Por algum motivo, parece que quando o tamanho do domínio é muito pequeno em relação ao paralelepípedo inicial, isso desfavorece o desempenho do algoritmo. Tomemos o caso da Tabela 6. Para a canga (domínio que corresponde a 0,6% do todo), o *speed-up* do SMPA em comparação à SBR é de 28,47x, enquanto para o itabirito silicoso (domínio que corresponde a 12% do todo) o *speed-up* é de 3,16x, diferença bastante considerável. Porém, para poder explicar melhor as razões dessa perda de performance seria necessário acesso ao código do programa.

Em suma, esse estudo de caso demonstra que a SMPA é também satisfatória numa aplicação real em depósito de ferro, reproduzindo bem os histogramas, variogramas e espaços de incerteza de cada domínio. O ganho de velocidade também foi bastante considerável nesse caso 3D, mesmo utilizando uma quantidade razoável de pontos auxiliares no processo.

6. CONSIDERAÇÕES FINAIS

6.1. CONCLUSÕES

Após todos as análises apresentadas nos capítulos anteriores, foi possível concluir que a Simulação via Múltiplos Passeios Aleatórios é um método de rápido e simples uso. E, ao se recapitular os objetivos colocados inicialmente para essa dissertação, todos foram cumpridos com êxito.

Dentre as novas contribuições à parte teórica do método, se destacam o uso de pontos auxiliares e a simulação independente do efeito pepita, que resolveram os problemas que ocorriam previamente na reprodução da variabilidade a curtas distâncias.

O desenvolvimento do algoritmo foi também bastante satisfatório, resultando em uma ferramenta de uso simples e resultados bastante eficientes. Utilizando essa nova ferramenta foi possível validar o uso da técnica nos dois estudos de casos apresentados, demonstrando um ganho de velocidade considerável em comparação com técnicas consagradas e boa reprodução dos momentos estatísticos. Em contrapartida, houve apenas uma ligeira perda de precisão global.

O ganho de velocidade se mostrou maior quanto menor a necessidade de pontos auxiliares. Em casos que não houve necessidade de pontos auxiliares, o método chegou a ser de 7 a 9 vezes mais rápido que os métodos tradicionais. No estudo de caso real, mesmo com a simulação de um número considerável de pontos auxiliares simulados, houve uma aceleração de ao menos 2x no caso mais desfavorável, chegando até 5,65x de aceleração no caso mais favorável.

6.2. RECOMENDAÇÕES DE TRABALHOS FUTUROS

Apesar dos diversos desenvolvimentos feitos sobre a Simulação via Múltiplos Passeios, o método ainda tem bastante a ser explorado. Como trabalhos futuros, seria interessante checar a validade dos resultados em cenários de outras commodities minerais (Au, Cu, Ni, Zn...). Também testar e desenvolver o uso do método em aplicações multivariadas, seja combinado com técnicas de fatorização (PCA e MAF) ou com modelos lineares de correionalização.

Outra linha de estudo com grande importância prática seria validar extensivamente o uso do método para atualizar os modelos de teores localmente (em uma janela do depósito

representando um dado volume a ser minerado a curto prazo) após inclusão de novas amostras ao banco de dados. Alguns testes feitos indicam que, após uma simulação global (devidamente validada), pode-se simular localmente uma região utilizando os mesmos parâmetros globais (inclusive o fator de calibragem do modelo), de modo que seria possível atualizar o modelo local a ser minerado de uma maneira bastante simples e rápida, utilizando os mesmos parâmetros globais mas sem ter que simular novamente o depósito todo. Esse processo de simulação local favoreceria o uso corriqueiro em planejamento de lavra a curto prazo.

REFERÊNCIAS

ARNDT, N.; GANINO, C. L. M. **Metals and society: An introduction to economic geology**. Heidelberg: Springer, 2012.

BABINSKI, M.; CHEMALE, F.; VAN SCHMUS, W. R. The Pb/Pb age of the Minas Supergroup carbonate rocks, Quadrilátero Ferrífero, Brazil. **Precambrian Research**, Amsterdam, v. 72, n. 3, p. 235-245, Apr. 1995.

BLACK, T. C.; FREYBERG, D. L. Simulation of one-dimensional correlated fields using a matrix-factorization moving average approach. **Mathematical geology**, New York, v. 22, n. 1, p. 39-62, Jan. 1990.

CAIXETA, R. M.; COSTA, J. F. C. L.; MACHADO, P. L.; RIBEIRO, D. T.; DIAS, D. H.; PIRES, R. O. How precise and accurate is the space of uncertainty derived from simulations? In: INTERNATIONAL SYMPOSIUM ON THE APPLICATION OF COMPUTERS AND OPERATIONS RESEARCH IN THE MINERAL INDUSTRY, 37., 2015, Fairbanks. **Proceedings**... Englewood: SME, 2015. p. 553-563.

CHERCHENEVSKI, P. K. **Influência da incerteza dos teores no planejamento de lavra aplicado ao sequenciamento de longo prazo**. 2015. 104 f. Dissertação (Mestrado em Engenharia) – Programa de Pós-Graduação em Engenharia de Minas, Metalúrgica e de Materiais, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2015.

CHILES, J. P.; DELFINER, P. **Geostatistics: modeling spatial uncertainty**. 2. ed. New Jersey: John Wiley & Sons, 1999.

CHU, J. **Conditional fractal simulation, fast sequential indicator simulation, and interactive variogram modeling**. 1993. 294 f. Tese (PhD) – Department of Applied Earth Sciences, Stanford University, Stanford, 1993.

COSTA, J. F. C. L. **Developments in recoverable reserves and ore body modeling**. 1997. 333 f. Tese (PhD) – WH Bryan Mining Geology Research Centre, University of Queensland, Queensland, 1997.

COWARD, S.; DOWD, P. Geometallurgical models for the quantification of uncertainty in mining project value chains. In: INTERNATIONAL SYMPOSIUM ON THE APPLICATION OF COMPUTERS AND OPERATIONS RESEARCH IN THE MINERAL INDUSTRY, 37., 2015, Fairbanks. **Proceedings**... Englewood: SME, 2015. p. 360-369.

DEUTSCH, C. V.; JOURNAL, A. G. **GSLib: Geostatistical software library and user's guide**, 2. ed. New York: Oxford University Press, 1998.

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

GEOVARIANCES. **Isatis technical references**. Paris: Geovariances & Ecole des Mines de Paris, 2014.

GOOVAERTS, P. **Geostatistics for natural resources evaluation**. New York: Oxford University Press, 1997.

ISAAKS, E.H. **The application of Monte Carlo methods to the analysis of spatially correlated data**. 1990. 213 f. Tese (PhD) – Department of Applied Earth Sciences, Stanford University, Stanford, 1990.

JOURNEL, A. G. Geostatistics for conditional simulation of ore bodies. **Economic Geology**, Littleton, v. 69, n. 5, p. 673-687, Aug. 1974.

JOURNEL, A. G.; HUIJBREGTS, C. J. **Mining geostatistics**. London: Academic press, 1978.

LANTUÉJOUL, C. **Geostatistical simulation: models and algorithms**. Heidelberg: Springer, 2002.

LAWLER, G. F.; LIMIC, V. **Random walk: a modern introduction**. New York: Cambridge University Press, 2010.

MARIETHOZ, G. A general parallelization strategy for random path based geostatistical simulation methods. **Computers & Geosciences**, Amsterdam, v. 36, n. 7, p. 953-958, Jul. 2010.

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

MATHERON, G. The intrinsic random functions and their applications. **Advances in applied probability**, Jerusalem, v. 5, n. 3, p. 439-468, Dec. 1973.

MELLO, P. R. G. **Planejamento de lavra sob a incerteza de teores aplicado à lavra subterrânea**. 2015. 162 f. Dissertação (Mestrado em Engenharia) – Programa de Pós-Graduação em Engenharia de Minas, Metalúrgica e de Materiais, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2015.

PARDO-IGUZQUIZA, E.; CHICA-OLMO, M. The Fourier integral method: an efficient spectral method for simulation of random fields. **Mathematical Geology**, New York, v. 25, n. 2, p. 177-217, Feb. 1993.

PEARSON, K. The Problem of the Random Walk. **Nature**, London, v. 72, n. 1865, p. 294, Jul. 1905.

PYRCZ, M. J.; DEUTSCH, C. V. **Geostatistical reservoir modeling**. New York: Oxford University Press, 2014.

RASERA, L. G.; MACHADO, P. L.; COSTA, J. F. C. L. A conflict-free, path-level parallelization approach for sequential simulation algorithms. **Computers & Geosciences**, Amsterdam, v. 80, p. 49-61, Jul. 2015.

REMY, N.; BOUCHER, A.; WU, J. **Applied geostatistics with SGeMS: a user's guide**. New York: Cambridge University Press, 2009.

RIBEIRO, D. T.; CUNHA-FILHO, E. M.; COSTA, J. F. C. L.; ROLDÃO, D. G.; TAMANTINI, T. B. F. Multidirectional random walk: an alternative to local simulation of Gaussian data. In: INTERNATIONAL GEOSTATISTICS CONGRESS, 9., 2012, Oslo. **Digital Proceedings**... Dordrecht: Springer, 2012. Disponível em: <<http://geostats2012.nr.no/pdfs/1747874.pdf>>. Acesso em: 29 set. 2015.

RIBEIRO, D. T.; COSTA, J. F. C. L.; TAMANTINI, T. B. F.; CUNHA-FILHO, E. M.; ROLDÃO, D. G.; FILHO, C. M. Local geostatistical simulation based on multidirectional random walk. In: INTERNATIONAL SYMPOSIUM ON THE APPLICATION OF COMPUTERS AND OPERATIONS RESEARCH IN THE MINERAL INDUSTRY, 36., 2013, Fairbanks. **Proceedings**... Porto Alegre: FLE, 2013. p. 178-188.

ROSIÈRE, C. A.; CHEMALE JR, F. Itabiritos e minérios de ferro de alto teor do Quadrilátero Ferrífero—uma visão geral e discussão. **Revista Geonomos**, Belo Horizonte, v. 8, n. 2, p. 27-43, Jul. 2000.

ROSSI, M. E.; DEUTSCH, C. V. **Mineral resource estimation**. Dordrecht: Springer, 2014.

SANGLARD, J. C. D.; ROSIERE, C. A.; SANTOS, J. O. S. A estrutura do segmento oeste da Serra do Curral, Quadrilátero Ferrífero, e o controle tectônico das acumulações compactas de alto teor em Fe. **Geologia USP, Série Científica**, São Paulo, v. 14, n. 1, p. 81-95, mar. 2014.

SHINOZUKA, M.; JAN, C. M. Digital simulation of random processes and its applications. **Journal of sound and vibration**, London, v. 25, n. 1, p. 111-128, Nov. 1972.

SILVA, D. S. F.; CUBA, M. A. SpectralSim – A program for unconditional spectral simulation. **Centre for Computational Geostatistics Annual Report**, Alberta, v. 15, n. 410, p. 1-8, Sept. 2013.

TRAN, T. T. Improving variogram reproduction on dense simulation grids. **Computers & Geosciences**, Amsterdam, v. 20, n. 7, p. 1161-1168, Out. 1994.

VERLY, G. **Estimation of spatial point and block distributions: the multigaussian model**. 1984. 367 f. Tese (PhD) – Department of Applied Earth Sciences, Stanford University, Stanford, 1984.

YAO, T. Conditional spectral simulation with phase identification. **Mathematical Geology**, New York, v. 30, n. 3, p. 285-308, Apr. 1998.

ANEXO A – ANÁLISE DA SENSIBILIDADE DA SIMULAÇÃO VIA MÚLTIPLOS PASSEIOS ALEATÓRIOS ÀS DISTÂNCIAS DAS AMOSTRAS AOS PONTOS SIMULADOS

Foram gerados três modelos sintéticos 2D de 600 por 600 m via simulações não-condicionais (Tabela A1), cada qual com um diferente tipo de modelo de covariância – esférico, exponencial e gaussiano. Os modelos apresentam estrutura única isotrópica de 60 m de alcance (sem efeito pepita).

Destes modelos, foram feitas diversas amostragens regulares a cada 40, 30, 12 e 6 metros. Utilizando esses bancos de dados, foram realizadas simulações em grids de 600 x 600 m com diferentes tamanhos de células: 1,5x1,5, 3x3, 5x5 e 10x10 m (Tabela A2)

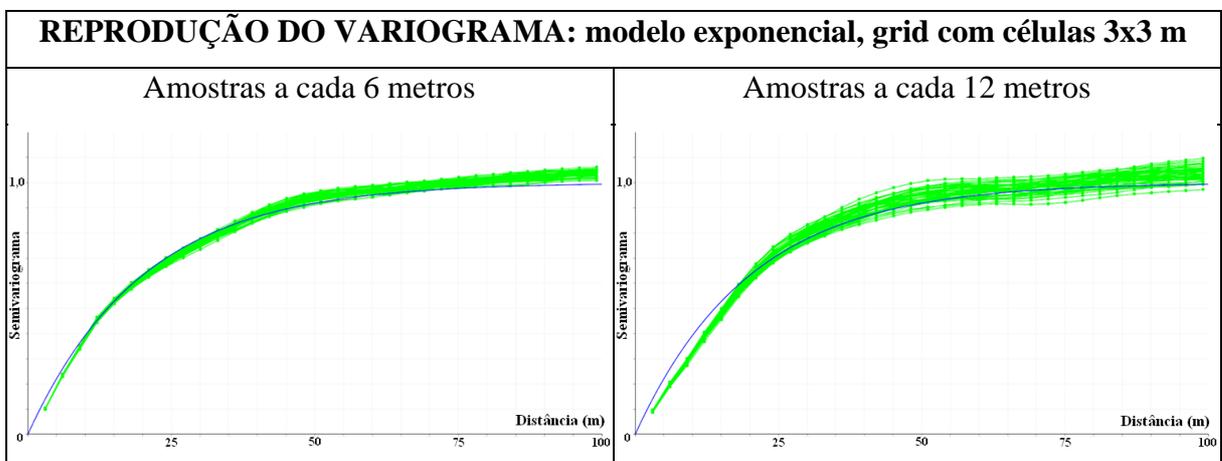
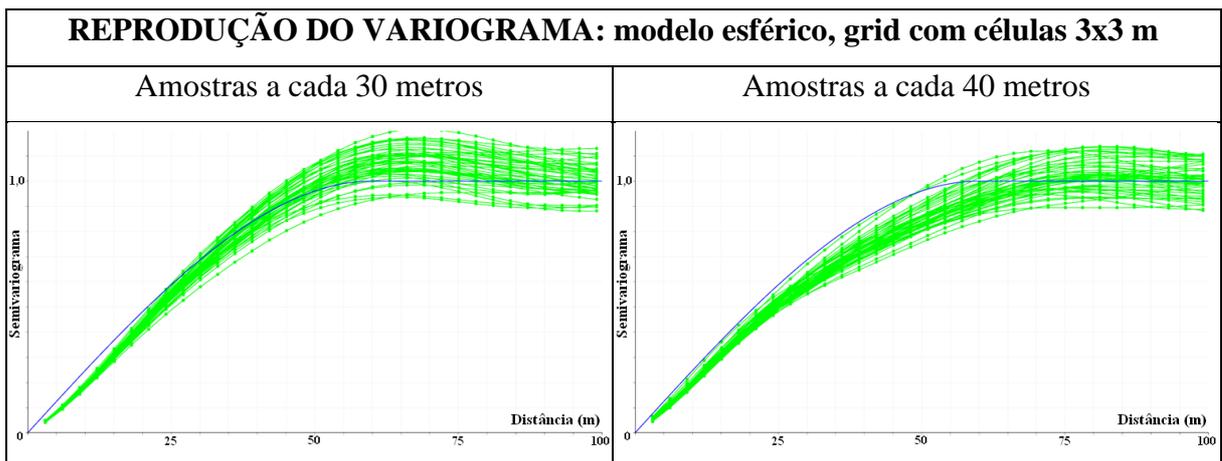
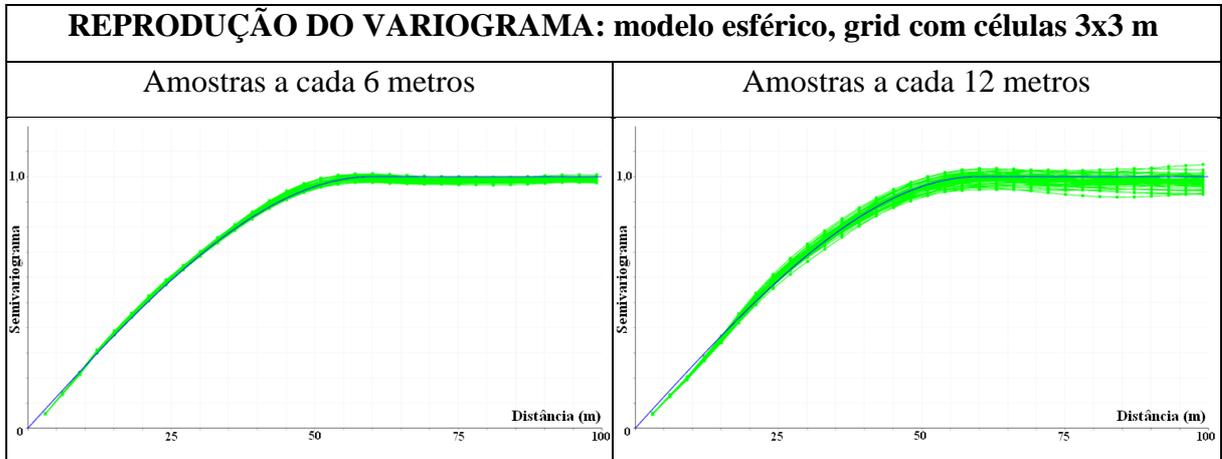
Os principais resultados estão representados abaixo. Como resumido no Capítulo 3, para gerar boa reprodução da continuidade e evitar artefatos, deve haver amostras a uma distância mínima de 3-6 vezes o valor da separação entre os pontos do grid a ser simulado – o que representa uma variância a ser restaurada de 3-15% dependendo do tipo do modelo variográfico – sendo os do tipo esférico e gaussiano mais robustos, enquanto o exponencial requer amostras (ou pontos auxiliares) bastante próximos entre si para gerar bons resultados.

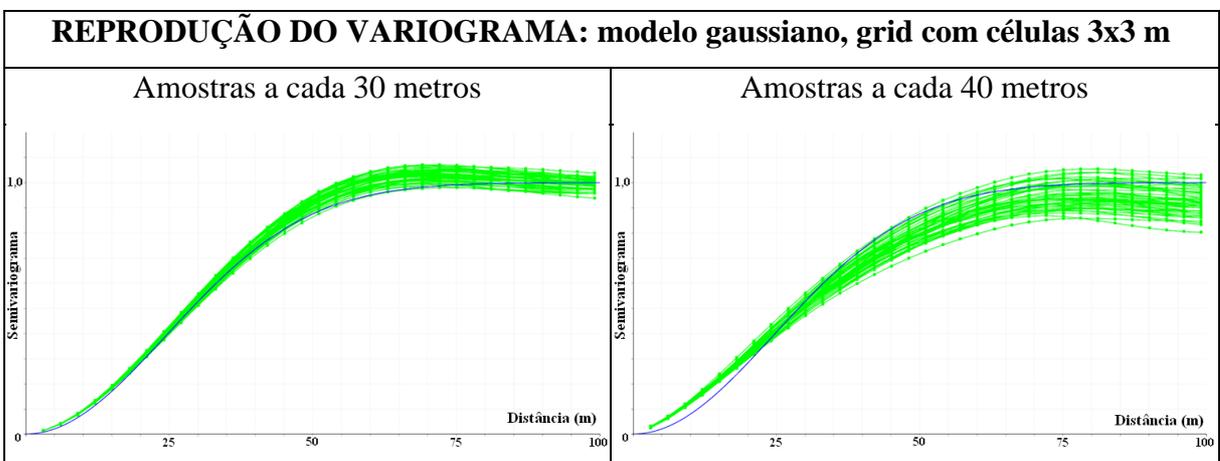
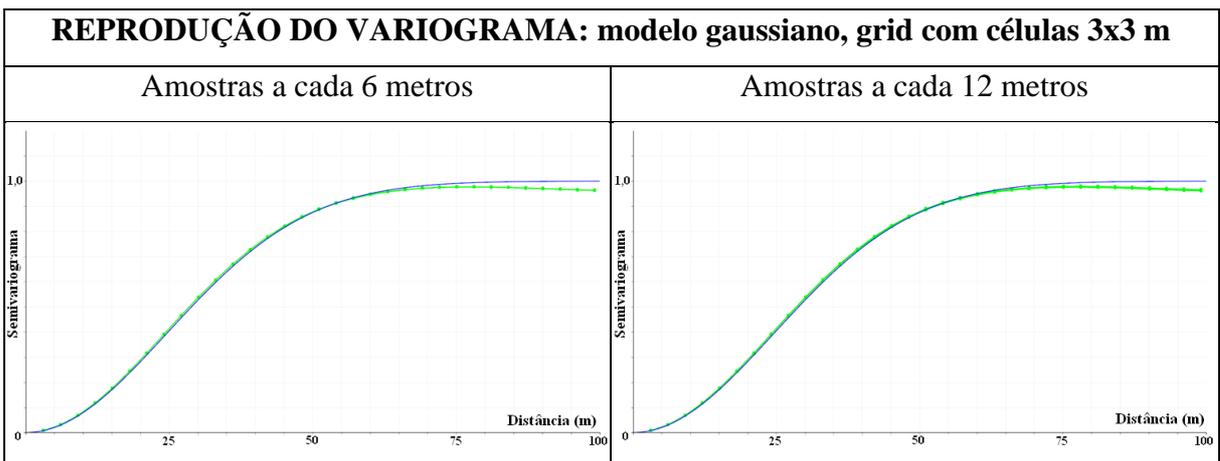
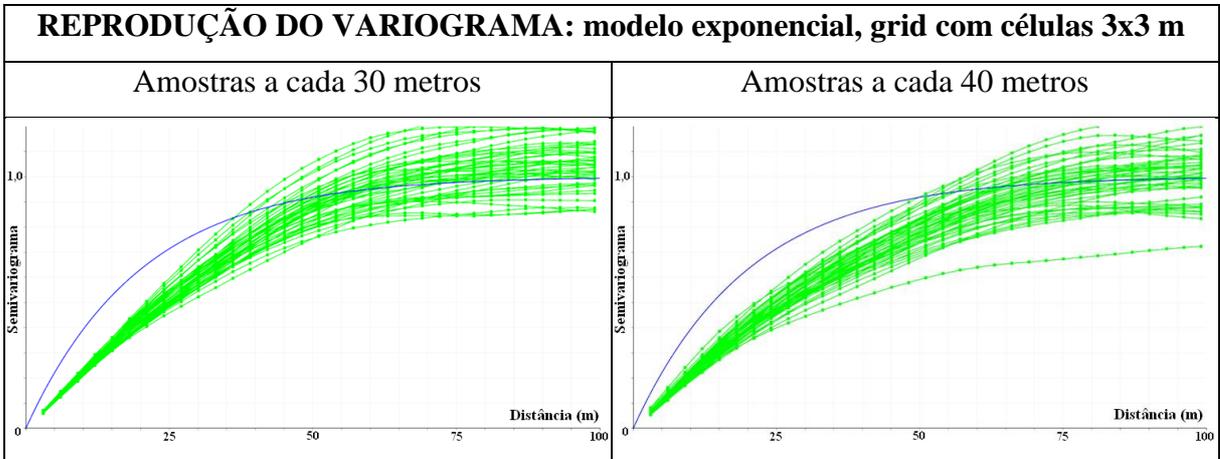
Tabela A1 - Detalhes dos três modelos de referência simulados

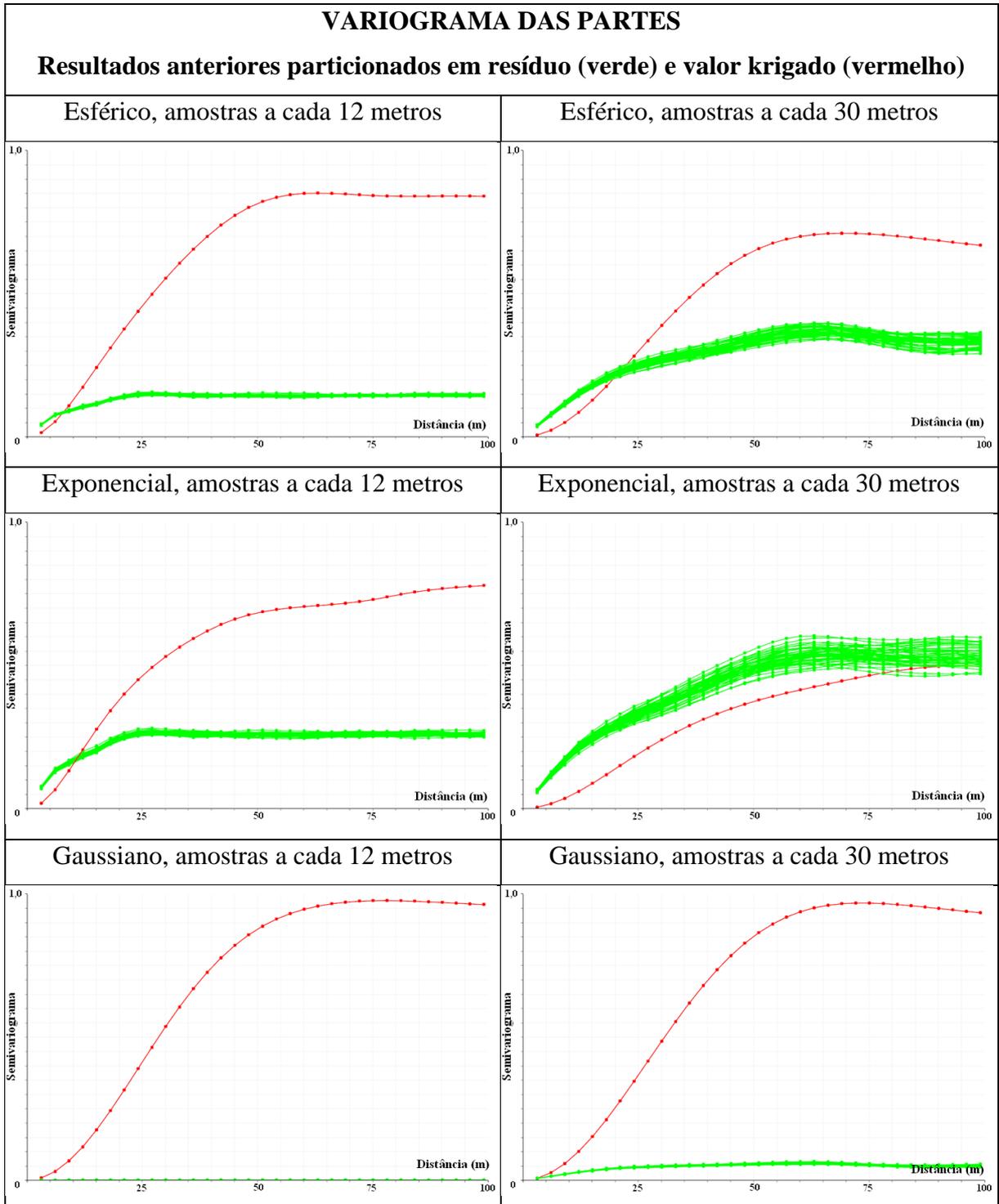
Dimensões do grid	600 x 600 m
Variogramas	$\gamma_1 = 1,0Exp\left(\frac{Omni}{60m}\right), \gamma_2 = 1,0Sph\left(\frac{Omni}{60m}\right), \gamma_3 = 1,0Gau\left(\frac{Omni}{60m}\right)$

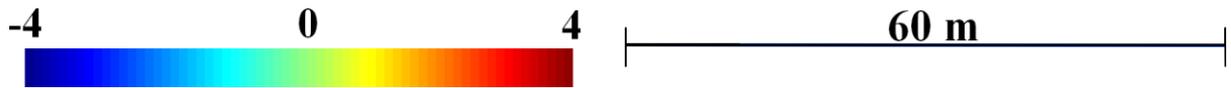
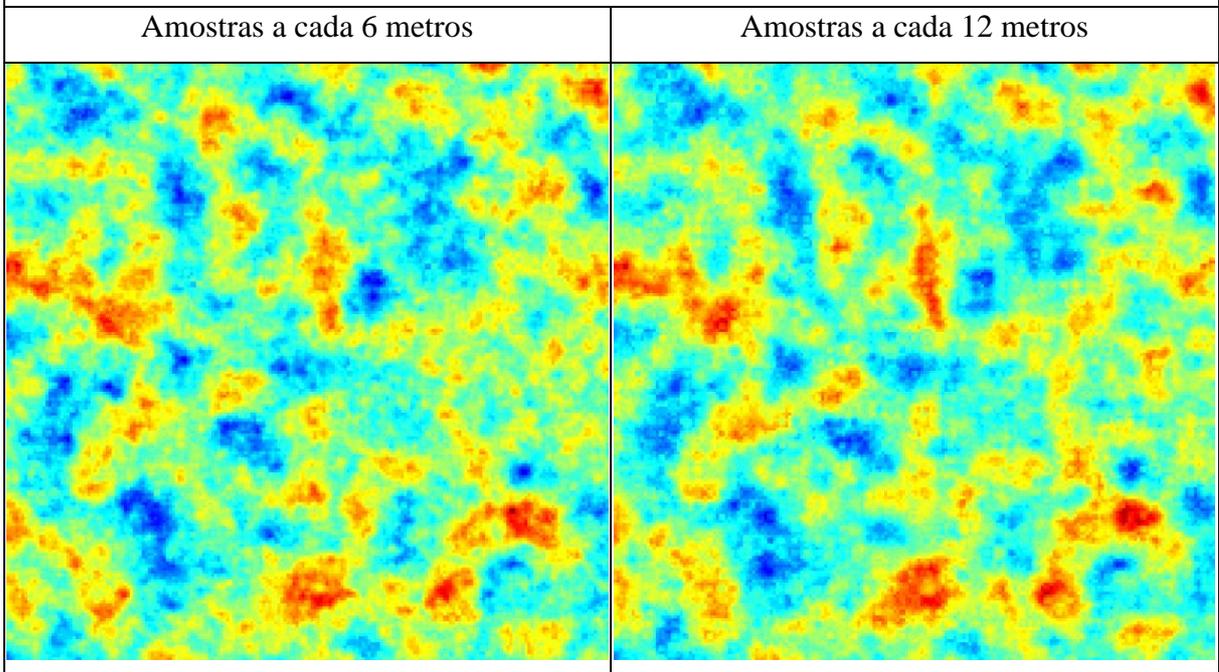
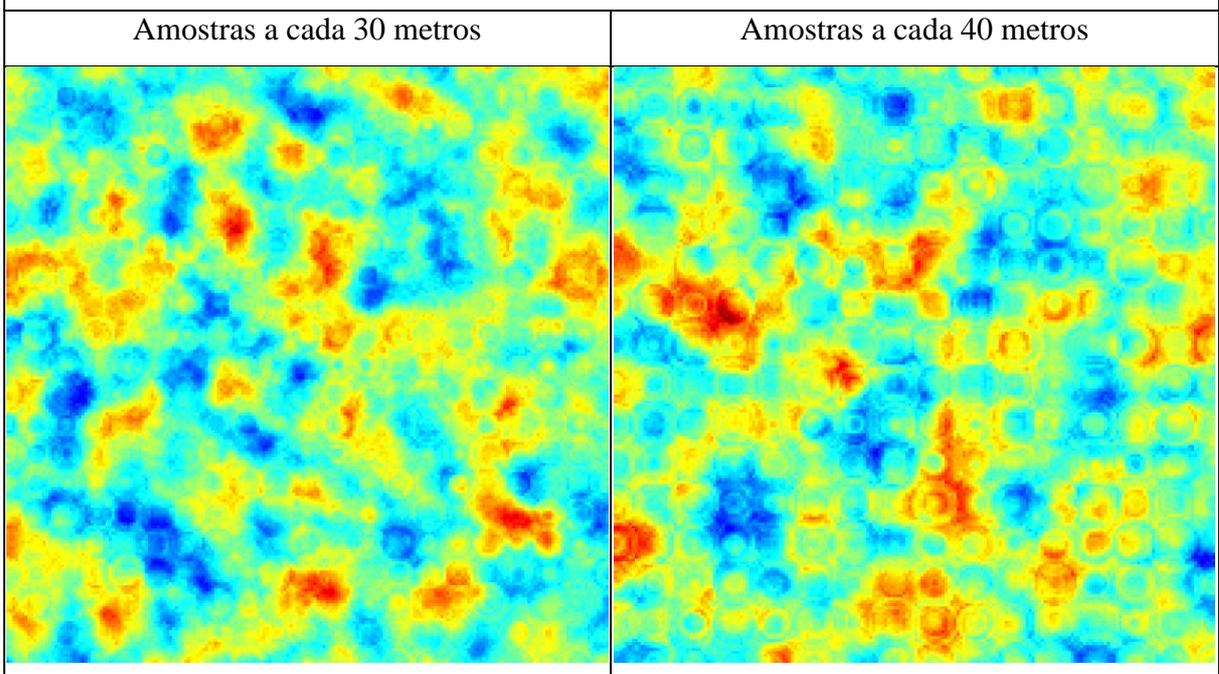
Tabela A2 – Parâmetros utilizados para os testes de Simulação via Múltiplos Passeios Aleatórios

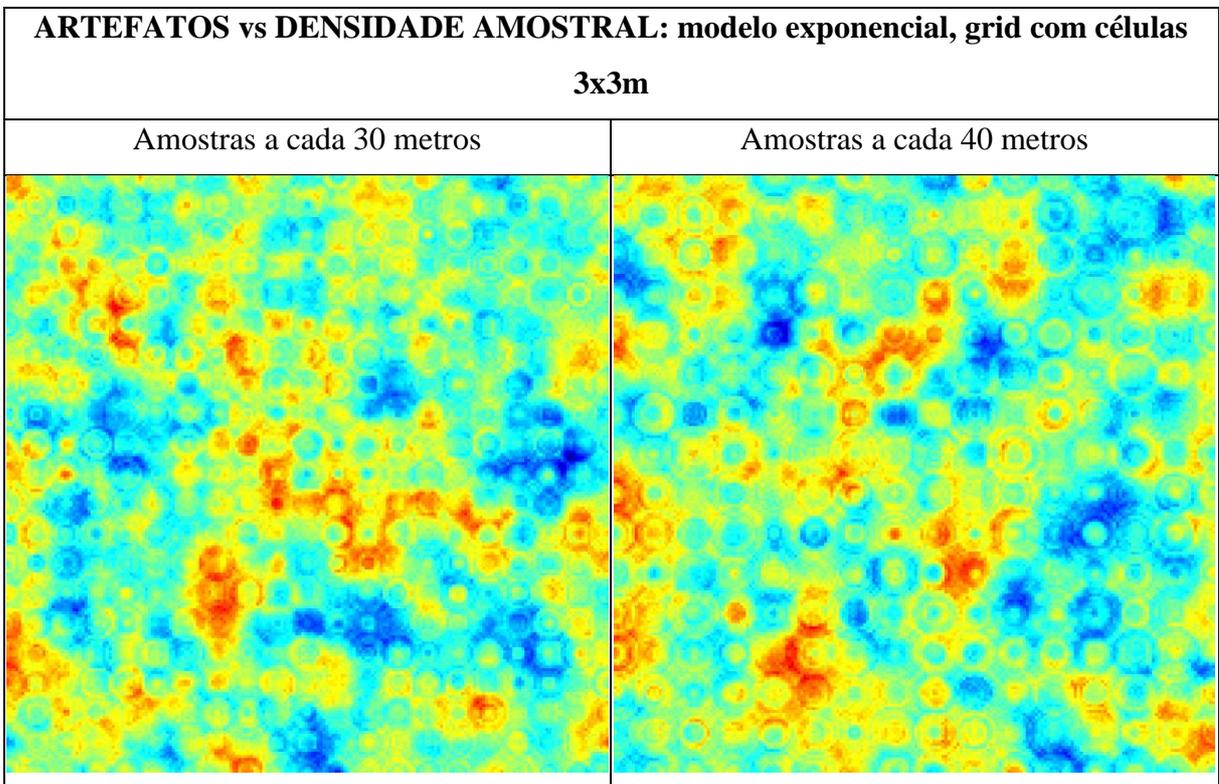
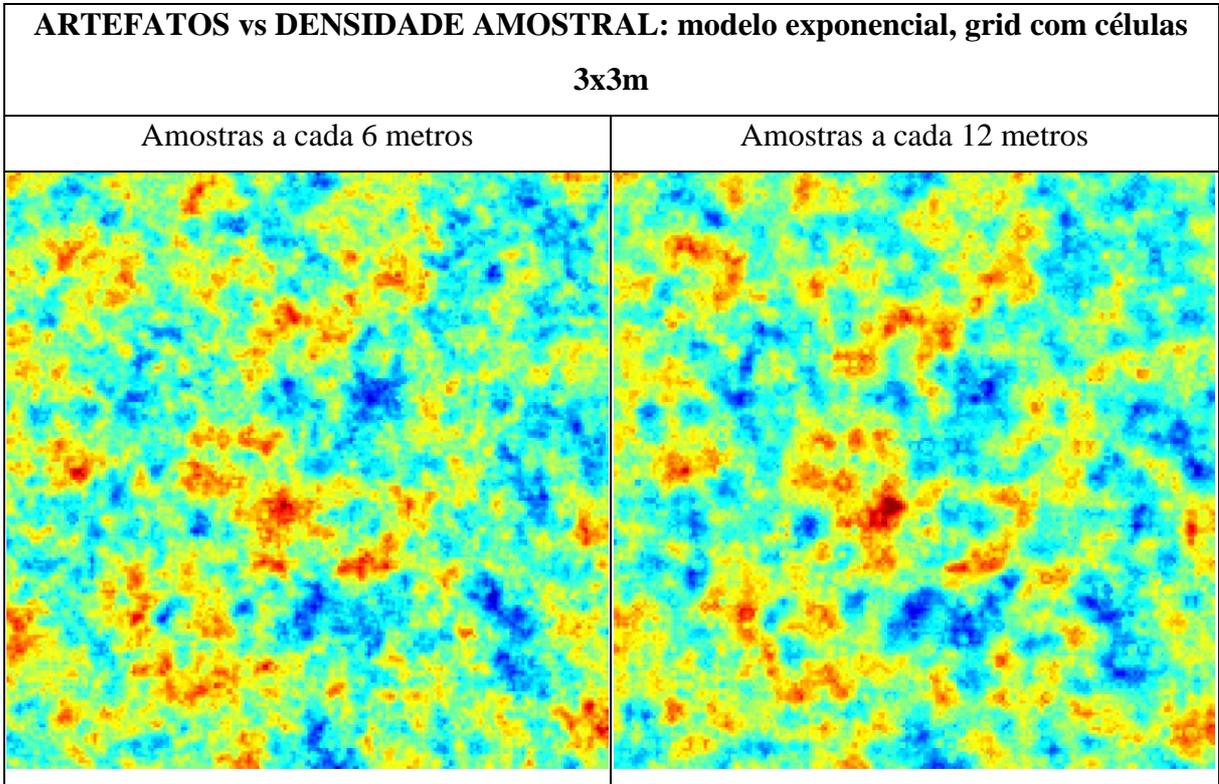
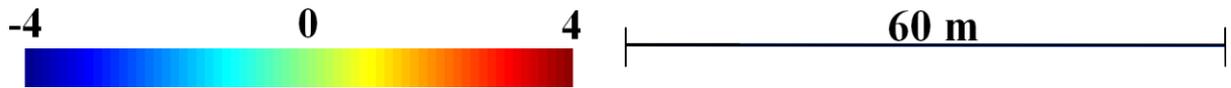
Dimensões do grid	600 x 600 m
Dimensões das células do grid	Testes em 1,5x1,5, 3x3, 5x5 e 10x10 m
Amostragem	Testes com amostras a cada 40, 30, 12 e 6 m
Variogramas	Mesmos dos modelos de referência

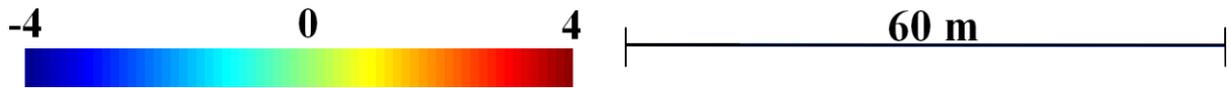
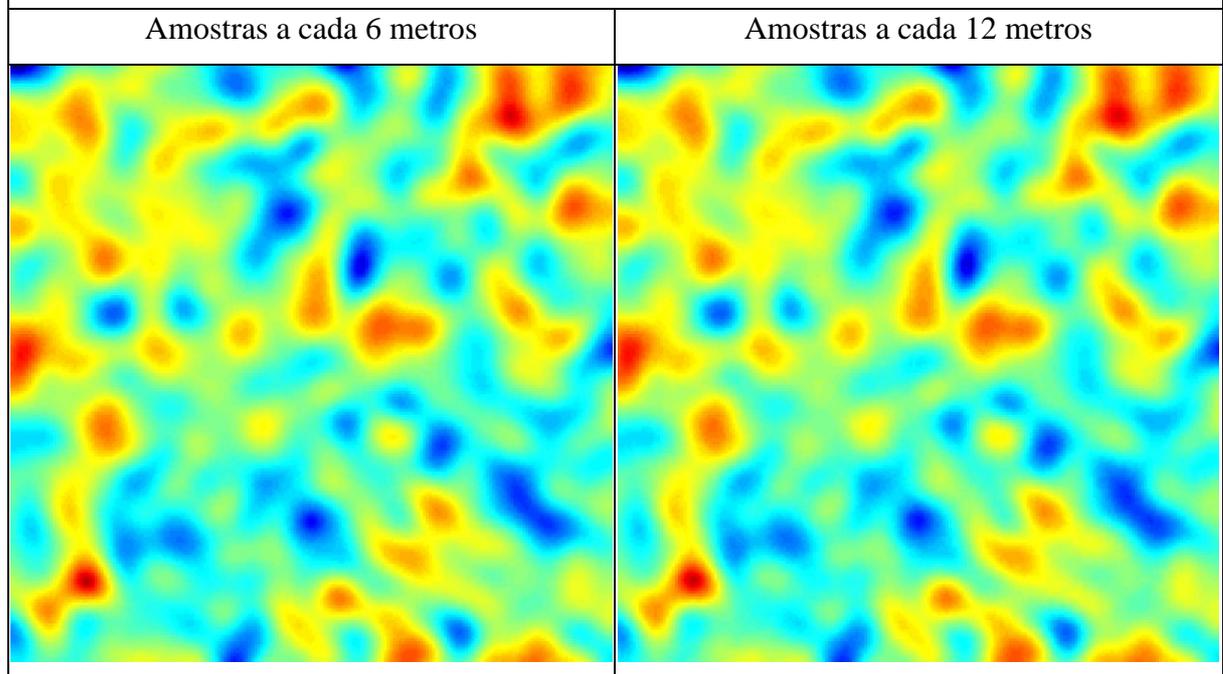






**ARTEFATOS vs DENSIDADE AMOSTRAL: modelo esférico, grid com células 3x3 m****ARTEFATOS vs DENSIDADE AMOSTRAL: modelo esférico, grid com células 3x3 m**



**ARTEFATOS vs DENSIDADE AMOSTRAL: modelo gaussiano, grid com células 3x3m****ARTEFATOS vs DENSIDADE AMOSTRAL: modelo gaussiano, grid com células 3x3m**