

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
Escola de Engenharia
Programa de Pós-Graduação em Engenharia de Minas, Metalúrgica e de
Materiais – PPGEM.

**UTILIZAÇÃO DE SIMULAÇÃO CONJUNTA COLOCADA
COM VARIÁVEL SUPERSECUNDÁRIA PARA
CONSTRUÇÃO DE MODELO GEOMETALÚRGICO DE
NIÓBIO
ARAXÁ – MG**

JOSÉ MARQUES BRAGA JÚNIOR

Dissertação para obtenção do título de mestre em Engenharia.

Porto Alegre

Abril de 2017

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
Escola de Engenharia
Programa de Pós-Graduação em Engenharia de Minas, Metalúrgica e de
Materiais – PPGEM.

**UTILIZAÇÃO DE SIMULAÇÃO CONJUNTA COLOCADA
COM VARIÁVEL SUPERSECUNDÁRIA PARA
CONSTRUÇÃO DE MODELO GEOMETALÚRGICO DE
NIÓBIO**

ARAXÁ – MG

JOSÉ MARQUES BRAGA JÚNIOR

Dissertação realizada no Laboratório de Pesquisa Mineral e Planejamento Mineiro da Escola de Engenharia da UFRGS, dentro do programa de Pós-Graduação em Engenharia de Minas, Metalurgia e de Materiais (PPGEM), como parte dos requisitos para obtenção do Título de Mestre em Engenharia.

Área de concentração: Metalurgia Extrativa e Tecnologia Mineral.

Porto Alegre

Abril de 2017

Folha de rosto constando a ficha catalográfica

JOSÉ MARQUES BRAGA JÚNIOR

UTILIZAÇÃO DE SIMULAÇÃO CONJUNTA COLOCADA COM VARIÁVEL
SUPERSECUNDÁRIA PARA CONSTRUÇÃO DE MODELO GEOMETALÚRGICO DE
NIÓBIO ARAXÁ – MG

**Esta dissertação foi julgada adequada para a
obtenção do Título de Mestre em Engenharia,
área de concentração de Metalurgia Extrativa
e Tecnologia Mineral e aprovada em sua forma
final, pelo Orientador e pela Banca
Examinadora do curso de Pós-graduação.**

Orientador: Prof. Dr. João Felipe Coimbra Leite Costa

Coordenador: Prof. Dr. Carlos Pérez Bergmann

Aprovado em: ____/____/____

BANCA EXAMINADORA

Ana Carolina Chierregati (Doutora) – USP _____

Francisco Gregianin Testa (Doutor) – CBMM/USP _____

Vanessa Cerqueira Koppe (Doutora) – UFRGS _____

Dedico esse trabalho aos meus pais: José Marques e Sirlene Correia, aos meus irmãos, à minha esposa Iara, aos meus colegas de trabalho e ao meu orientador João Felipe. Por terem me apoiado durante toda minha formação como ser humano. Obrigado pelo amor, carinho e amizade.

AGRADECIMENTOS

Ao Professor João Felipe Coimbra Leite Costa pelos ensinamentos e pela oportunidade de compartilhar comigo conhecimentos indispensáveis para a minha formação. Agradeço a ele também pela orientação na elaboração desse trabalho.

Aos bolsistas de iniciação científica do PPGEM que me auxiliaram diversas vezes durante a elaboração desse trabalho.

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

A todos colegas e amigos do LPM que fazem deste um grupo de trabalho competente e qualificado.

À CBMM por ter me concedido a oportunidade de realizar esse mestrado conciliado com meu trabalho na empresa.

Aos colegas de trabalho da CBMM que me ajudaram durante o desenvolvimento desse trabalho, principalmente à Amanda Santos.

“Os ventos que às vezes tiram algo que amamos, são os mesmos que trazem algo que aprendemos a amar. Por isso não devemos chorar pelo que nos foi tirado e sim, aprender a amar o que nos foi dado. Pois tudo aquilo que é realmente nosso, nunca se vai para sempre...”

Bob Marley

Resumo

Modelo de blocos para teor é um recurso comumente utilizado pelo planejamento de lavra na indústria mineira. Na maioria dos casos o conhecimento sobre os teores das variáveis químicas não é suficiente para prever o desempenho geometalúrgico do minério quando submetido ao processo de concentração. A geometalurgia engloba um conjunto de testes de comportamento metalúrgico do minério e seus resultados são incorporados ao modelo de bloco, ajudando a tornar o planejamento da lavra mais preciso quanto à capacidade de produção, melhorando os ganhos financeiros e reduzindo os riscos associados à lavra e a tomada de decisões. A recuperação metalúrgica de nióbio mede o quanto do conteúdo metálico de interesse no minério é recuperado no concentrado após o processamento mineral. Esta informação é muitas vezes subutilizada no modelo de bloco devido à baixa quantidade de dados primários, o que dificulta a construção de um modelo de bloco confiável. No entanto, para complementar a variável de interesse, informações secundárias de outros atributos podem ser utilizadas. A cossimulação de informações não aditivas em depósitos multivariados com mais de duas variáveis secundárias envolvidas é extremamente trabalhosa e normalmente seus resultados precisam ser ajustados posteriormente. A necessidade de ajustes posteriores, aliada a falta de praticidade da maioria dos métodos de cossimulação, motiva a busca por soluções alternativas que gerem resultados tão ou mais precisos e que sejam de fácil aplicação na rotina de modelamento geológico. É comum que os programas utilizados para a cossimulação se baseiem em uma única variável secundária, porém, o fenômeno analisado pode estar sendo influenciado por vários fatores, neste caso, o uso combinado de todos fatores relevantes pode melhorar a predição da variável de interesse. O uso de múltiplas variáveis secundárias pode ser gerenciado criando-se uma variável supersecundária. Neste caso, a quimiometria pode ser aplicada, resolvendo problemas preditivos e modelando propriedades de sistemas químicos visando prever a recuperação metalúrgica. Nesse trabalho, após a combinação de múltiplas variáveis em um preditivo supersecundário, a cossimulação sequencial gaussiana foi aplicada para gerar o modelo geometalúrgico. A simulação conjunta colocada permite a simulação conjunta do dado supersecundário com o dado primário, integrando mais informações para melhorar a predição da recuperação metalúrgica do nióbio. A cossimulação foi realizada com base no modelo de correionalização de Markov para simplificar a modelagem da covariância cruzada. O modelo probabilístico geometalúrgico obtido se mostrou eficiente, mantendo uma precisão adequada na previsão da variável de interesse.

Abstract

Grade block models are a standard input in mine planning throughout the mining industry. In most cases, the ore grades knowledge is not enough to predict the behavior of the ore at the processing plant. Geometallurgy comprises a set of ore metallurgical behavior tests and their results incorporated into the block model, helping in making mine planning more precise when it comes to the production capacity, improving financial earnings and reducing risks. Niobium Metallurgical Recovery is a very important variable to be controlled, measuring how much of the metal content in the ore is recovered in the concentrate after mineral processing. This information is often underused in the block model due to the low quantity of primary data, which makes the construction of a reliable block model difficult. However, to supplement the variable of interest, secondary information from other attributes can be used. Cosimulation of non-additive information in multivariate deposits with more than two secondary variables involved is extremely labor-intensive and its results usually need to be later adjusted. The need for subsequent adjustments, combined with the lack of practicality of most cossimulation methods, motivates the search for alternative solutions that generate results that are as accurate and easy to apply in the routine of geological modeling in the mineral industry. In multivariate geostatistics most programs used for cosimulation are based on one secondary variables. Frequently the analyzed phenomenon is influenced by several factors. In this case, the use of them combined can improve the prediction of the variable of interest. The use of multiple secondary variables can be managed by creating a super-secondary variable. In this case, chemometrics can be applied, solving predictive problems, modeling properties of chemical systems aiming at predicting the metallurgical recovery. After combining multiple variables into a super-secondary predictive, Sequential Gaussian Cosimulation was applied in this study to generate a geometallurgical model. The collocated joint simulation allows the joint simulation of a super-secondary data with the primary data, integrating more information to improve the cosimulation of the niobium metallurgical recovery. The cosimulation was run based on the Markov coregionalization model to simplify the cross-covariance modeling. The result is a representative probabilistic geometallurgical model, which proved to be efficient maintaining an adequate precision in forecasting the predicted variable.

Índice de figuras

Figura 1- Fluxograma das etapas dos trabalhos realizados.	9
Figura 2- Fluxograma do teste de simulação da concentração.	39
Figura 3- Estatística descritiva da variável primária (DCCG).	41
Figura 4- Histograma da variável DCCG desagrupada utilizando o método dos polígonos de Voronoi.	42
Figura 5- Gráfico de probabilidade normal da variável primária (DCCG). Eixo x corresponde a variável DCCG e eixo y corresponde à porcentagem.	42
Figura 6- Mapa com a distribuição dos furos contendo os dados primários.	43
Figura 7- Fluxograma com os principais tipos de regressão.	44
Figura 8- Gráfico de carga fatorial considerando nove variáveis químicas e a recuperação metalúrgica analisada em laboratório. Os primeiros componentes estão no eixo x e os segundos componentes no eixo y.	46
Figura 9- Histograma dos dados das variáveis dependentes e da variável independente.	47
Figura 10- Scatter plots das cinco variáveis consideradas na regressão multivariada analisadas individualmente contra a resposta (SEC-REG). No eixo Y, tem-se os valores calculados da recuperação metalúrgica (SEC-REG) e no eixo X estão os valores das variáveis independentes utilizadas na regressão.	49
Figura 11- Scatter plot entre a recuperação metalúrgica analisada em laboratório (DCCG) comparada com o resultado do cálculo da recuperação metalúrgica através da regressão multivariada de cinco óxidos (SEC-REG).	50
Figura 12- Gráfico para análise de resíduos da regressão múltipla.	52
Figura 13- Resultado da regressão obtido com o software MINITAB.	53
Figura 14- Gráficos de efeitos entre as variáveis dependentes e a resposta.	54
Figura 15- Histograma dos valores da variável supersecundária.	55
Figura 16- Mapa com a distribuição da boca das sondagens com os dados supersecundário s.	56
Figura 17- Histograma dos dados primário transformado.	59
Figura 18- Variograma na direção vertical.	60
Figura 19- Variogramas horizontais modelados.	61
Figura 20- Variogramas modelados para a direção de maior e menor continuidade na horizontal.	62
Figura 21- Variograma na direção (N0, D10) modelado.	63
Figura 22- Histograma da variável supersecundária transformada (NSCORE).	64
Figura 23- Variograma na direção vertical (Down the hole)	65

Figura 24- Variogramas horizontais modelados.	66
Figura 25- Variogramas experimentais na direção N0 variando o dip a cada 10 graus.	67
Figura 26- (A) Variograma modelado da variável DCCG na direção (N90, D0) (B) Flutuações ergódicas das 100 cossimulações da variável recuperação metalúrgica de nióbio (DCCG) com a supersecundária colocada (SEC-REC) na direção (N90, D0).	70
Figura 27- (A) Variograma modelado da variável DCCG na direção (N0, D0) (B) Flutuações ergódicas das 100 cossimulações da variável recuperação metalúrgica de nióbio (DCCG) com a supersecundária colocada (SEC-REC) na direção (N0, D0).	71
Figura 28- (A) Variograma modelado da variável DCCG na direção (N0, D90) (B) Flutuações ergódicas das 100 cossimulações da variável recuperação metalúrgica de nióbio (DCCG) com a supersecundária colocada (SEC-REC) na direção (N0, D90).	72
Figura 29- Histograma da média das 100 cossimulações e histograma da cokrigagem da variável geometalúrgica.	73
Figura 30- Histograma acumulado das 100 cossimulações juntamente com o histograma do dado primário desagrupado.	74
Figura 31- (A) Histograma do erro relativo no intervalo de 90% de confiança dos dados cossimulados. (B) Mapa dos erros relativos para os intervalos de 90% de confiança.	75
Figura 32- Mapa de uma das 100 cossimulações realizadas.	78
Figura 33- (A) Scatter plot entre a cokrigagem e a medias das cossimulações. (B) Swath plot (N0, D0) do resultado das 100 cossimulações (linhas pretas) e do resultado da cokrigagem (linha vermelha)...	79
Figura 34- Curvas de cutoff de recuperação metalúrgica por tonelagem obtidos pelos diferentes resultados da cossimulação da variável DCCG. No eixo vertical tem-se a tonelagem e no eixo horizontal tem-se os diferentes valores de corte para a recuperação metalúrgica.	80

LISTA DE SIGLAS

CBMM	Companhia Brasileira de Metalurgia e Mineração
CODEMIG	Companhia de Desenvolvimento Econômico de Minas Gerais
BDMG	Banco de Desenvolvimento de Minas Gerais
CEMIG	Companhia Energética de Minas Gerais
GASMIG	Companhia de Gás de Minas Gerais
JUCEMG	Junta Comercial de Minas Gerais
INDI	Instituto de Desenvolvimento Integrado de Minas Gerais
COMIPA	Companhia Mineradora do Pirocloro de Araxá
FSA	Furos de sonda <i>Air Core</i> – Campanha de sondagem para amostragem
PCA	Análise por componentes principais
NSCORE	Algoritmo para transformação de uma distribuição qualquer para uma distribuição normal.
HCA	Análise por agrupamento hierárquico
MSR	Metodologia de superfície de resposta
SSE	Soma dos quadrados do erro residual
SGS	Simulação Sequencial Gaussiana
SSD	Simulação Sequencial Direta
SIS	Simulação Indicadora Sequencial
FDAC	Função de Distribuição Acumulada Condicional
sgcosim	Algoritmo de Cossimulação Sequencial Gaussiana
DCCG ou DCCA	Recuperação metalúrgica global de nióbio (variável primária)
QAQC	Garantia da Qualidade e Controle da Qualidade
SEC-REC	Variável supersecundária (obtida por regressão múltipla)
GSLIB	Geostatistical Software Library

SGEMS	Stanford Geostatistical Modeling Software
3D	3 Dimensões
ICs	Intervalos de confiança

LISTA DE SÍMBOLOS

N-	Azimute
D-	Dip
R^2	Coefficiente de determinação
t/h	Toneladas por hora
m	Metros
%	Porcentagem
ε	Erro aleatório
φ	Função de transformação para a distribuição normal

Sumário

1	INTRODUÇÃO	1
1.1	ÁREA DE ESTUDO E ASPECTOS GERAIS DO EMPREENDIMENTO.....	4
1.2	META.....	5
1.3	OBJETIVOS.....	5
1.4	METODOLOGIA.....	6
	Fonte: Autoria própria.....	9
1.5	ORGANIZAÇÃO DA DISSERTAÇÃO.....	10
2	REVISÃO BIBLIOGRÁFICA	11
2.1	ESTATÍSTICA MULTIVARIADA.....	11
2.1.1	Conceitos Gerais.....	11
2.1.1	Coeficiente de correlação linear de Pearson.....	12
2.1.2	Análise de componente principal	12
2.1.3	Superfície de resposta.....	15
2.1.3.1	Modelo de Primeira Ordem.....	17
2.1.3.2	Modelo de Segunda Ordem.....	19
2.1.4	Regressão Não-Linear	20
2.2	SIMULAÇÃO E COSSIMULAÇÃO GEOESTATÍSTICA.....	22
2.2.1	Métodos de simulação sequencial	26
2.2.2	Simulação Sequencial Gaussiana - SGS.....	27
2.2.3	Simulação Sequencial Direta - SSD	28
2.2.4	Simulação indicadora sequencial – SIS	30
2.2.5	Cossimulação Geoestatística	32
2.2.6	O método de cossimulação utilizado	36
3	ESTUDO DE CASO	37
3.1	VARIÁVEL PRIMÁRIA (HARD DATA)	37
3.1.1	Obtenção do dado primário (experimental).....	37
3.1.2	Estatística descritiva do dado primário.....	39
3.2	DADO SECUNDÁRIO (SOFT DATA)	43

3.2.1 Obtenção da variável supersecundária	43
3.3 SIMULAÇÃO CONJUNTA COLOCADA	56
3.3.1 Variografia da variável primária.....	58
3.3.2 Variografia da variável supersecundária	63
3.3.3 Cossimulação.....	68
3.4 VALIDAÇÃO DA COSSIMULAÇÃO	69
4 CONSIDERAÇÕES FINAIS	76
5 CONCLUSÕES	81
REFERÊNCIAS	83

1 INTRODUÇÃO

Durante o desenvolvimento de um projeto de mineração, o qual haja verticalização da produção do bem mineral ou metal, é importante prever a resposta geometalúrgica do minério ao processo de concentração ao qual será submetido. Essa informação ajuda a maximizar o benefício econômico do projeto através da integração do conhecimento geológico com as atividades de mina, processamento mineral, meio ambiente e mercado.

Procura-se um método preciso e acurado para previsão da variável geometalúrgica no modelo de blocos a ser utilizado pelo planejamento de lavra, na avaliação de recurso e reserva, bem como no planejamento estratégico. Essa é uma das formas de maximizar o aproveitamento do recurso mineral, contribuindo com a sustentabilidade do empreendimento mineiro.

A informação “recuperação metalúrgica” pode ser obtida de forma experimental. Para isso, é necessário submeter uma amostra representativa de minério a ensaios laboratoriais que simulem as etapas de concentração. Esses experimentos normalmente demandam elevado tempo de resposta e alto custo. Uma forma de minimizar o elevado custo decorrente dos experimentos de simulação da resposta do minério ao processo de concentração seria identificar ou criar uma ou mais variáveis, que sejam de fácil aquisição e possuam uma forte correlação com a recuperação metalúrgica. Essas variáveis poderiam ser utilizadas, conjuntamente com o dado experimental, na construção de um modelo probabilístico de distribuição da variável de interesse.

A separação dos tipos de minério de acordo com suas principais características físicas, químicas, mineralógicas e texturais (exemplo: recobrimento do mineral minério por aluminofosfatos ou hidróxido de ferro; grau de liberação, inclusões de Pirocloro em cristais maiores de outros minerais e outros) facilitam o entendimento do comportamento da resposta de cada um dos tipos de minério quando submetidos a um mesmo processo de concentração. Essa individualização dos domínios litológicos melhora a correlação entre as variáveis, uma vez que as amostras dentro de um mesmo domínio tendem a produzir respostas no processo mais parecidas entre si. O desenvolvimento prático desse estudo ficou restrito a litologia denominada “Alterito Laranja”, por se tratar da atual fonte de recurso de nióbio da empresa (CBMM).

Para se construir e empregar modelos para prever e explicar fenômenos com precisão é necessária uma etapa de seleção cuidadosa das variáveis mais significativas que podem explicar em parte ou na totalidade o comportamento do sistema. A razão é que, embora seja necessário um grande número de variáveis e/ou parâmetros para prever um fenômeno com exatidão, um

pequeno número de variáveis, em geral, explica grande parte dele. Deste modo, o ponto inicial de um processo de modelagem consiste em identificar as variáveis certas e as relações entre elas (GAVIRA, 2003).

Estudos recentes mostraram que o comportamento da recuperação metalúrgica do minério de nióbio não é função somente de uma variável. Ainda não se sabe ao certo quais são todas as variáveis que contribuem para a explicação do fenômeno “recuperação metalúrgica”. No entanto, utilizando uma técnica combinada de análise estatística multivariada com regressões múltiplas pelo método de superfície de resposta, foi possível obter funções envolvendo alguns óxidos presentes no minério, cuja resposta seria a própria recuperação metalúrgica com um erro associado.

Os métodos de estatística multivariada utilizados nesse trabalho foram escolhidos de acordo com os objetivos dessa pesquisa. Sabe-se que a análise multivariada é uma análise exploratória de dados, prestando-se a gerar hipóteses, e não tecer confirmações a respeito dos mesmos, o que seria uma técnica confirmatória, como nos testes de hipótese, nos quais se tem uma afirmação a respeito da amostra em estudo.

Para a criação de um modelo probabilístico, medidas de um atributo de interesse, em alguns casos, podem ser complementadas por informação secundária proveniente de outros atributos relacionados. A informação secundária pode estar ou não amostrada mais densamente que a variável primária. Quando a informação secundária é mais adensada, levá-la em consideração pode melhorar as estimativas da variável primária, principalmente quando a informação da variável de interesse é escassa ou pobremente correlacionada no espaço.

A maioria dos programas que aplicam cokrigagem e cossimulação se baseiam na utilização de uma única variável secundária. No entanto, é comum que o fenômeno analisado seja influenciado por diversos fatores, sendo, nesse caso, sugerido por esse trabalho, utilizar a combinação destas variáveis com métodos estatísticos multivariados para obtenção de uma variável que represente da melhor forma possível a influência de todas as variáveis secundárias sobre a variável principal.

Quando a dependência espacial entre as variáveis é considerada importante no modelo, deve-se considerar a correlação espacial cruzada entre elas e realizar uma simulação condicional de mais de uma variável ou uma simulação conjunta. De acordo com os autores Deutsch e Journel (1998) e Goovaerts (1997), o problema, neste caso, reside na inferência e modelagem da matriz de covariância cruzada. Devido aos problemas implementados, a

cossimulação é raramente realizada na sua forma completa, sendo possível realizar aproximações, tais como:

(a) substituir a simulação de N variáveis dependentes por N fatores independentes, a partir dos quais as variáveis originais possam ser reconstituídas;

(b) utilizar a variável mais importante, chamada variável primária, para ser simulada primeiro; então, todas as outras variáveis correlacionadas (variáveis secundárias) são simuladas por realizações obtidas de distribuições condicionais específicas. As autocorrelações das variáveis secundárias são indiretamente reproduzidas a partir daquela da variável primária. Portanto, esta aproximação da cossimulação não é recomendada quando é importante reproduzir acuradamente as covariâncias de variáveis secundárias que diferem acentuadamente daquela da variável primária.

O método de cossimulação escolhido para ser utilizado nesse trabalho é conhecido como Cossimulação Sequencial Gaussiana, algoritmo *sgcosim*. Ele permite a simulação conjunta de diversas variáveis integrando diferentes informações dos N dados secundários.

Esse algoritmo se baseia em duas ideias centrais, sendo elas:

- (a) A ideia de cokrigagem colocada para reduzir o esforço computacional envolvido no processo de simulação, e para deixar a matriz de cokrigagem mais estável. Essa ideia implica em manter o dado da variável secundária próximo do local onde a variável primária está sendo estimada (dado secundário colocado).
- (b) O modelo de correionalização de Markov I para simplificar a inferência e o modelamento dos variogramas cruzados. Esse modelo se baseia no seguinte argumento: a variável secundária mais próxima, em particular, do dado colocado, minimiza a influência dos dados semelhantes mais distantes.

A simplificação do procedimento de cossimulação utilizando o modelo de Markov I pode ser visto como uma generalização do tradicional algoritmo da Cossimulação Sequencial Gaussiana para lidar simultaneamente com diversas variáveis primárias e diversas variáveis secundárias. O algoritmo se baseia no axioma de Bayes de probabilidade condicional, adequado para a cossimulação sequencial, onde a simulação conjunta de vários eventos dependentes é obtida a partir da elaboração de uma sequência univariada de distribuição condicional.

Para facilitar a determinação dessa distribuição condicional, um modelo Gaussiano multivariado é assumido. Assim, a ideia de cokrigagem colocada é incorporada para reduzir o

esforço computacional. Finalizando, o modelo de correionalização de Markov é introduzido para simplificar o modelamento e a covariância cruzada.

O método de cossimulação empregado foi validado através da verificação do seu desempenho em reproduzir um conjunto de dados reais, tais como a reprodução de variogramas e histogramas. Foi verificada também a incerteza da cossimulação para o intervalo de 90% de confiança. Além disso, foi avaliada a correlação do resultado da média da cossimulação com a cokrigagem entre a variável primária (DCCG) e a variável supersecundária (SEC-REC). A variável supersecundária foi obtida através da regressão múltipla pelo método de superfície de resposta a partir de cinco óxidos e possui uma alta correlação com a variável primária (DCCG).

Esse estudo foi desenvolvido em um caso real, com o objetivo de modelar a recuperação metalúrgica do minério de nióbio em uma mina ativa.

1.1 ÁREA DE ESTUDO E ASPECTOS GERAIS DO EMPREENDIMENTO

A CBMM – Companhia Brasileira de Metalurgia e Mineração – é uma empresa dedicada à lavra, ao processamento, à industrialização e à comercialização de produtos de nióbio. A CODEMIG é uma empresa de economia mista, cujo acionista majoritário é o Estado de Minas Gerais. Ao lado do BDMG, CEMIG, GASMIG, JUCEMG e do INDI, ela integra o sistema liderado pela Secretaria de Desenvolvimento Econômico de Minas Gerais. Atua na realização de projetos, obras, serviços e empreendimentos, com destaque para o setor de infraestrutura. Na área de mineração, realiza investimentos em busca de oportunidades que viabilizem novos empreendimentos no Estado, em parceria com o setor privado.

A lavra do minério de nióbio é realizada pela COMIPA – Companhia Mineradora do Pirocloro de Araxá – sob forma de arrendamento, em dois decretos de lavra: um da CBMM e outro da CODEMIG.

A mineração é realizada a céu aberto, em bancadas de 10 m de altura com a operação sendo executada por desmonte mecânico, carga e transporte. O minério, após ser lavrado, é conduzido, via caminhões, até o ponto de alimentação na mina que o direciona ao pátio de homogeneização para posterior retomada. A homogeneização do minério é feita através do empilhamento em sequência de camadas do minério no padrão Chevron, sendo posteriormente retomado por um retomador de caçamba e direcionado à usina de concentração. Atualmente, a capacidade nominal do retomador é de 1.820 t/h (base úmida).

As principais etapas do processo de beneficiamento e tratamento do minério consistem em: britagem primária e secundária, separação magnética, classificação, deslamagem, condicionamento, flotação, espessamento e filtração.

O produto da usina de concentração é um concentrado de nióbio com cerca de 55,0% de Nb_2O_5 , sendo esse utilizado exclusivamente pela CBMM como insumo nas fases subsequentes de sua verticalização, tais como: sinterização, desfosforação, metalurgia, química para obtenção de óxidos e forno de feixe de elétrons para obtenção do nióbio metálico. Os produtos finais industrializados são: Ferro-Nióbio *Standard*, Óxidos de Nióbio, Nióbio Metálico e Nióbio-Zircônio, além de ligas especiais: Ferro-Nióbio e Níquel-Nióbio.

1.2 META

A meta desse trabalho é melhorar a estimativa da variável geometalúrgica: Recuperação Metalúrgica de nióbio, incorporando informações adicionais a partir de dados secundários utilizando técnicas avançadas de estatística multivariada e geoestatística.

1.3 OBJETIVOS

Uma vez contextualizada a relevância da modelagem de variáveis geometalúrgicas propõe-se investigar relações, encontrar e propor técnicas que expliquem o comportamento do minério no atual processo de concentração sem precisar realizar testes laboratoriais. Para isso, é necessário controlar, manipular e medir as variáveis que são consideradas relevantes ao entendimento do fenômeno analisado. Muitas são as dificuldades em traduzir as informações obtidas em conhecimento, principalmente quando se trata da avaliação estatística dessas informações.

Objetiva-se também com essa pesquisa empregar um método de cossimulação geoestatística utilizando uma variável supersecundária, sem a inferência e o modelamento completo da matriz de covariância cruzada. Essa abordagem pode ser investigada com uso do modelo de Markov I (ALMEIDA, 1993).

Dispõe-se de duas fontes de informações da recuperação metalúrgica, sendo elas: testes de laboratoriais de bancada ou em uma planta piloto (dados primários) e por regressão múltipla (dados secundários) com distinta precisão e acurácia. Essas duas fontes de informações para a mesma variável foram combinadas utilizando métodos estatísticos multivariados e geoestatísticos, dando origem a um único modelo probabilístico.

Também se almeja com esse trabalho construir um modelo geometalúrgico da recuperação mássica de nióbio, correspondente com o da produção real na usina de processamento do concentrado de nióbio, com níveis de precisão aceitáveis e mensuráveis.

Com a aplicação dos resultados desse trabalho foi possível reduzir a quantidade de testes de processo de bancada para obtenção da recuperação metalúrgica de nióbio, contribuindo de forma significativa para a redução de custos com a caracterização metalúrgica das amostras de minério.

Objetiva-se também fomentar a atual demanda do planejamento de lavra com a confecção de um modelo probabilístico representativo da distribuição da recuperação metalúrgica de nióbio dentro dos limites de desenvolvimento da mina.

A utilização da equação de regressão para o cálculo da recuperação metalúrgica de nióbio também poderá ser aplicada em outros segmentos da produção de nióbio, como por exemplo, na previsão da recuperação metalúrgica das amostras de pilhas formadas no pátio de homogeneização e/ou em furos de sondagem obtidos pelo método de circulação reversa, sendo esses utilizados para o controle de teor através do modelo de blocos de curto prazo.

1.4 METODOLOGIA

Para criar o modelo probabilístico representativo da recuperação metalúrgica de nióbio utilizando a simulação conjunta das variáveis recuperação metalúrgica analisada e recuperação metalúrgica calculada, foi necessário primeiramente classificar os dados de acordo com suas características, em diferentes domínios geológicos.

Através da descrição geológica dos testemunhos de sondagem da campanha FSA foi possível caracterizar e individualizar os grandes domínios geológicos, sendo classificados em: Solo Argiloso; Alterito Laranja; Alterito Marrom; Saprolito e Rocha Fresca.

Dentre os cinco domínios geológicos citados anteriormente, apenas um possui robustez de dados e informações suficientes para ser submetido a uma análise de variância espacial e, conseqüentemente, ser submetido a uma simulação geoestatística. O domínio geológico em questão é o Alterito Laranja.

Para a obtenção da variável supersecundária, utilizou-se do método de regressão por superfície de resposta. Essa metodologia estatística é comumente utilizada para a modelagem e análise de problemas nos quais a variável resposta é influenciada por vários fatores, cujo objetivo é a otimização dessa resposta.

O banco de dados considerado para a obtenção da equação de regressão foi o correspondente aos furos de sondagem da campanha FSA, os quais tiveram amostras submetidas aos ensaios de bancada para avaliação do desempenho metalúrgico do minério, além das análises químicas dos principais óxidos de interesse.

Antes de utilizar a técnica de superfície de resposta, foi necessário definir quais variáveis independentes contribuem para a explicação do fenômeno de interesse, no caso a recuperação metalúrgica de nióbio. Para tanto, foi utilizada a técnica de análise de componentes principais (PCA).

Para as regressões multivariadas e para a análise de componentes principais foi utilizado o software MINITAB® (FROST, 2014), o qual possui diversas ferramentas estatísticas tanto para análises univariadas quanto para análises multivariadas.

Para a utilização do método de cossimulação geoestatística pelo modelo de Markov I, foi necessário que os dados, tanto da variável primária quanto da variável secundária, seguissem uma distribuição Gaussiana, sendo necessário realizar a transformação das variáveis primária e supersecundária utilizando a transformação NSCORE (software GSLIB®).

A variografia da variável supersecundária normalizada foi realizada utilizando o software SGems. A variografia foi realizada para ser utilizada na simulação conjunta, conforme requisita o método Markov I.

A simulação geoestatística da variável supersecundária transformada foi realizada no software SGems e para tanto foi necessário definir os parâmetros de simulação, tais como raio de busca, número mínimo e máximo de amostras, quantidade máxima de amostra por octante, etc.

Após a normalização dos dados primários, foi realizada a variografia dos dados transformados, sendo esse dado posteriormente utilizado como um dos parâmetros para proceder com a cossimulação.

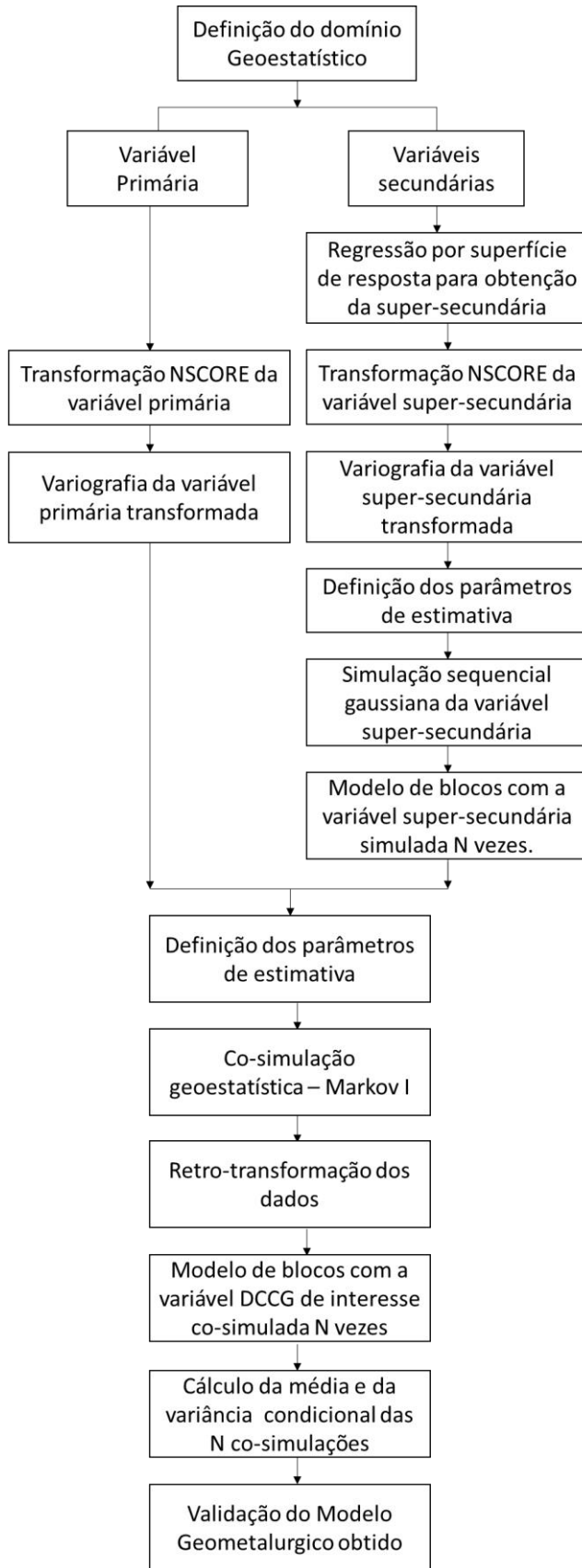
Os parâmetros de estimativa para realizar a cossimulação foram definidos e utilizados na rotina de simulação conjunta do SGems, o qual gerou 100 resultados de valores prováveis de acordo com os dados inseridos.

Após a cossimulação, os 100 resultados obtidos foram retro-transformados e submetidos a um processamento pós-simulação, com o objetivo de se obter a média dos valores, bem como a variância condicional dos dados cossimulados e retro-transformados.

O resultado da cossimulação foi validado com base na reprodução dos variogramas e histogramas. Além da verificação da incerteza da cossimulação para o intervalo de 90% de confiança. Outra técnica utilizada foi comparação entre o resultado da média da cossimulação e uma cokrigagem realizada com o mesmo conjunto de dados.

A figura 1 apresenta o fluxograma esquemático proposto para geração do modelo geometalúrgico cossimulado.

Figura 1- Fluxograma das etapas dos trabalhos realizados.



Fonte: Autoria própria.

1.5 ORGANIZAÇÃO DA DISSERTAÇÃO

Os capítulos dessa dissertação foram organizados de forma a ter uma sequência temporal lógica dos trabalhos desenvolvidos até chegar na conclusão.

O Capítulo 2 apresenta uma breve revisão bibliográfica sobre os principais métodos de sondagem para coleta de amostras em profundidade. Esse capítulo também resume os métodos de estatística multivariada, dando enfoque à análise de componentes principais, superfície de resposta e regressões não-lineares, utilizados nesse trabalho. O Capítulo 2 apresenta também uma síntese sobre a geoestatística, destacando a geoestatística multivariada e a cossimulação geoestatística.

O Capítulo 3 trata sobre os dados utilizados na construção do modelo geometalúrgico. Descreve os métodos de coleta, preparação e caracterização das amostras para a obtenção do dado primário (*hard data*), além de detalhar também a lógica matemática por de traz do cálculo da variável supersecundária (*soft data*), em que se baseia na regressão múltipla de óxidos pelo método de superfície de resposta de segunda ordem.

O Capítulo 4 descreve como foi realizado o tratamento geoestatístico dos dados. Detalhando os procedimentos utilizados para a realização da cossimulação geoestatística, com o intuito de gerar o modelo geometalúrgico.

O Capítulo 5 foi dedicado à validação do modelo geometalúrgico, realizando análises qualitativas para verificar a confiabilidade e a qualidade das informações dispostas no modelo geometalúrgico. Nessa etapa, foram utilizadas técnicas de validação da cossimulação. Entre as técnicas utilizadas destacam-se: verificação do desempenho do resultado em reproduzir um conjunto de dados reais através da reprodução de variogramas e histogramas e verificação da incerteza da simulação para o intervalo de 90% de probabilidade.

O Capítulo 6 apresenta uma conclusão dos trabalhos realizados.

No final da dissertação tem-se ainda a descrição das referências bibliográficas utilizadas nessa obra.

2 REVISÃO BIBLIOGRÁFICA

2.1 ESTATÍSTICA MULTIVARIADA

Nas decisões tomadas no dia a dia, normalmente leva-se em conta um grande número de fatores e, obviamente, existem fatores mais relevantes do que outros, sendo necessário considerar o grau de importância de cada um deles antes de tomar uma decisão. Muitas das vezes, decisões tomadas por intuição mascaram estes fatores. Ou seja, não são identificadas quais variáveis realmente afetaram as decisões tomadas e quais são as mais relevantes.

Frequentemente analisam-se somente as variáveis isoladamente e a partir desta análise fazem-se inferências sobre a realidade. Esta simplificação tem vantagens e desvantagens. Quando um fenômeno depende de muitas variáveis, a análise univariada passa a ser insuficiente, pois não basta conhecer informações estatísticas isoladas, é também necessário conhecer a totalidade destas informações fornecida pelo conjunto das variáveis, podendo a interação entre as variáveis ser inclusive, mais importante do que a análise direta entre a variável dependente e as demais variáveis.

As relações existentes entre as variáveis, na maioria das vezes, não são percebidas facilmente e, assim, efeitos antagônicos ou sinérgicos de efeito mútuo entre variáveis podem complicar a interpretação do fenômeno a partir das variáveis consideradas. Porém, no caso restrito de variáveis independentes entre si, é possível, com razoável segurança, interpretar um fenômeno complexo usando as informações estatísticas de poucas variáveis. As informações estatísticas mais relevantes neste tipo de análise são as medidas de tendência central e de dispersão dos dados.

2.1.1 Conceitos Gerais

A denominação Análise Estatística Multivariada corresponde a um grande número de métodos e técnicas estatísticas que permitem utilizar simultaneamente as variáveis importantes para a interpretação teórica de um determinado conjunto de dados.

A objetividade da pesquisa científica só começa depois da escolha das variáveis e das metodologias de análise, antes disto a atividade científica é completamente subjetiva.

Existem vários métodos de análise multivariada com finalidades bem diversas entre si. Por isso, é importante saber o que se pretende gerar antes de escolher algum método. De acordo com Vicini (2005), quando o interesse é verificar como as amostras se relacionam, ou seja, o quanto estas são semelhantes segundo as variáveis utilizadas no trabalho, destacam-se dois

métodos: a análise por agrupamento hierárquico (HCA) e a análise por componentes principais (PCA). Quando a finalidade principal é fazer previsão, por exemplo, quando temos muitas variáveis independentes e queremos encontrar uma variável dependente, a regressão linear múltipla e as redes neurais são métodos muito indicados.

2.1.1 Coeficiente de correlação linear de Pearson

O coeficiente de correlação de Pearson é uma medida do grau de relação linear entre duas variáveis quantitativas. Este coeficiente varia entre os valores -1 e 1. O valor 0 (zero) significa que não há relação linear, o valor 1 indica uma relação linear perfeita e o valor -1 também indica uma relação linear perfeita, mas inversa, ou seja, quando uma das variáveis aumenta a outra diminui. Quanto mais próximo estiver de 1 ou -1, mais forte é a associação linear entre as duas variáveis (PEARSON, 1901).

O coeficiente de correlação linear de Person é calculado segundo a seguinte fórmula:

$$R^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

onde x_1, x_2, \dots, x_n e y_1, y_2, \dots, y_n são os valores medidos de ambas as variáveis. E \bar{x} é a média aritmética dos n valores de x e \bar{y} é a média aritmética dos n valores de y .

2.1.2 Análise de componente principal

A Análise de Componentes Principais (PCA) é uma técnica estatística poderosa que pode ser utilizada para redução do número de variáveis e para fornecer uma visão estatisticamente privilegiada do conjunto de dados. A análise de componentes principais fornece as ferramentas adequadas para identificar as variáveis mais importantes no espaço das componentes principais.

Análise de Componentes principais é um dos métodos multivariados mais simples. O objetivo da análise é tomar p variáveis X_1, X_2, \dots, X_p e encontrar combinações destas para produzir índices Z_1, Z_2, \dots, Z_p que sejam não correlacionados na ordem de sua importância, e que descrevam a variação nos dados. A falta de correlação significa que os índices estão medindo diferentes "dimensões" dos dados, e a ordem é tal que $Var(Z_1) \geq Var(Z_2) \geq \dots \geq Var(Z_p)$, em que $Var(Z_i)$ denota a variância de (Z_i) . Os índices Z são então os componentes principais.

Análise de Componentes é um modelo fatorial no qual os fatores são baseados na decomposição da variância total. Na análise de componentes, unidades (1s) são usadas na

diagonal da matriz de correlação. Esse procedimento implica computacionalmente que toda a variância é comum ou compartilhada.

Uma análise de componentes principais começa com dados de p variáveis para n indivíduos. O primeiro componente principal é então a combinação linear das variáveis X_1, X_2, \dots, X_p :

$$Z_1 = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \quad (2)$$

que varia tanto quanto possível para os indivíduos, sujeitos à condição de que:

$$Z_2 = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p \quad (3)$$

Assim $Var(Z_1)$, a variância de Z_1 , é tão grande quanto possível dada esta restrição sobre as constantes a_{ij} . A restrição é introduzida porque se isto não for feito, então $Var(Z_1)$ poderia ser aumentada fazendo simplesmente crescer qualquer um dos valores a_{ij} .

O segundo componente principal:

$$Z_2 = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p \quad (4)$$

é escolhido de modo que $Var(Z_2)$ seja tão grande quanto possível sujeito à restrição de que:

$$a_{21}^2 + a_{22}^2 + \dots + a_{2p}^2 = 1 \quad (5)$$

e também à condição de que Z_1 e Z_2 tenham correlação zero para os dados. O terceiro componente principal:

$$Z_3 = a_{31}X_1 + a_{32}X_2 + \dots + a_{3p}X_p \quad (6)$$

é tal que a $Var(Z_3)$ seja tão grande quanto possível sujeita à restrição de que:

$$a_{31}^2 + a_{32}^2 + \dots + a_{3p}^2 = 1 \quad (7)$$

e também que Z_3 seja não correlacionada com ambas Z_1 e Z_2 . Posteriormente, componentes principais são definidos continuando da mesma maneira. Se existirem p variáveis, então existirão no máximo p componentes principais.

Para se usarem os resultados de uma análise de componentes principais, não é necessário saber como as equações, para os componentes principais, são obtidas. Entretanto, é útil entender a natureza das equações. De fato, uma análise de componentes principais envolve encontrar os autovalores de uma matriz de covariâncias amostral.

A matriz de covariâncias é simétrica e tem a forma:

$$C = \begin{bmatrix} C_{11} & \cdots & C_{p1} \\ \vdots & \ddots & \vdots \\ C_{1p} & \cdots & C_{pp} \end{bmatrix} \quad (8)$$

em que o elemento c_{ii} na diagonal é a variância de X_i e o termo fora da diagonal c_{ij} é a covariância entre as variáveis $X_i X_j$.

As variâncias dos componentes principais são os autovalores da matriz C . Existem p destes autovalores, alguns dos quais podem ser zero. Autovalores negativos não são possíveis para uma matriz de covariância. Assumindo que os autovalores estão ordenados como $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$, então, λ_i corresponde ao i -ésimo componente principal:

$$Z_i = a_{i1}X_1 + a_{i2}X_2 + \cdots + a_{ip}X_p \quad (9)$$

Em particular, $Var(Z_i) = \lambda_i$, e as constantes $a_{i1}, a_{i2}, \dots, a_{ip}$ são os elementos do correspondente autovetor, escalonado de modo que:

$$a_{i1}^2 + a_{i2}^2 + \cdots + a_{ip}^2 = 1 \quad (10)$$

Uma propriedade importante dos autovalores é que a soma deles é igual à soma dos elementos da diagonal (o traço) da matriz C . Isto é:

$$\lambda_1 + \lambda_2 + \cdots + \lambda_p = c_{11} + c_{22} + \cdots + c_{pp} \quad (11)$$

Porque c_{ii} é a variância de X_i e λ_i é a variância de Z_i , isto significa que a soma das variâncias dos componentes principais é igual à soma das variâncias das variáveis originais. Portanto, em certo sentido, os componentes principais contam com toda a variação nos dados originais.

A fim de evitar que uma ou duas variáveis tenham uma indevida influência nos componentes principais, é usual normalizar as variáveis X_1, X_2, \dots, X_p para terem médias zero e variâncias um no início de uma análise. A matriz C então toma forma:

$$C = \begin{bmatrix} 1 & C_{12} & \cdots & C_{1p} \\ C_{21} & 1 & \cdots & C_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1p} & C_{p2} & \cdots & 1 \end{bmatrix} \quad (12)$$

em que $c_{ij} = c_{ji}$ é a correlação entre X_i e X_j . Em outras palavras, a análise de componentes principais é feita sobre a matriz de correlação. Neste caso, a soma dos termos da diagonal, e, portanto, a soma dos autovalores, é igual a p , o número de variáveis X .

De acordo com Vicini (2005) os passos em uma análise de componentes principais podem ser estabelecidos como:

- (a) Comece normalizando as variáveis X_1, X_2, \dots, X_p para terem médias zero e variâncias unitárias. Isto é usual, mas é omitido em alguns casos em que se assume que a importância das variáveis é refletida em suas variâncias.
- (b) Calcule a matriz de covariâncias C . Esta é uma matriz de correlações se o passo 1 foi feito.
- (c) Encontre os autovalores $\lambda_1, \lambda_2, \dots, \lambda_p$ e os correspondentes autovetores a_1, a_2, \dots, a_p . Os coeficientes do i -ésimo componente principal são então os elementos de a_i , enquanto que λ_i é sua variância.
- (d) Descarte quaisquer componentes que explicam somente uma pequena proporção da variação nos dados. Por exemplo, começando com 20 variáveis, pode ser obtido que os primeiros três componentes expliquem 90% da variância total. Com base nisto, os outros 17 componentes podem ser razoavelmente ignorados (VICINI, 2005).

2.1.3 Superfície de resposta

No contexto da estatística experimental, há constante interesse em caracterizar a possível relação entre uma ou mais variáveis resposta e um conjunto de fatores de interesse. Isso pode ser executado através da construção de um modelo que descreva a variável resposta em função dos níveis aplicáveis desses fatores.

Certos tipos de problemas científicos envolvem a expressão de uma variável resposta, tal como o rendimento de um produto, como uma função empírica de um ou mais fatores quantitativos, tais como a temperatura de reação e a pressão. Isso pode ser efetuado utilizando-se uma metodologia que permita modelar a relação: rendimento em função de temperatura de reação e pressão. O conhecimento da forma funcional de f , frequentemente obtida com a modelagem de dados provenientes de experimentos planejados, permite tanto sumarizar os resultados do experimento quanto predizer a resposta para níveis dos fatores quantitativos. Assim, a função f define a superfície de resposta, que em sua essência, consiste em estimar coeficientes da regressão polinomial para a geração de um modelo empírico, por meio do qual é possível aproximar uma relação (inicialmente desconhecida ou conhecida) entre os fatores e as respostas do processo.

Podemos então definir a Superfície de Resposta como sendo a representação geométrica obtida quando uma variável resposta é plotada como uma função de dois ou mais fatores quantitativos. A função pode ser assim definida:

$$Y = f(x_1, x_2, \dots, x_k) + \varepsilon \quad (13)$$

em que Y é a resposta (variável dependente); x_1, x_2, \dots, x_k são os fatores (variáveis independentes); e ε é o erro aleatório.

Denota-se a resposta esperada por:

$$E(Y) = f(x_1, x_2, \dots, x_k) = \eta \quad (14)$$

então,

$$\eta = f(x_1, x_2, \dots, x_k) \quad (15)$$

é chamada de superfície de resposta.

A metodologia de superfície de resposta, ou MSR, é um conjunto de técnicas matemáticas e estatísticas que são úteis para modelagem e análise nas aplicações em que a resposta de interesse seja influenciada por várias variáveis e o objetivo seja otimizar essa resposta. Por exemplo, suponha que um engenheiro químico deseje encontrar os níveis de temperatura (X_1) e concentração da alimentação (X_2) que maximizem o rendimento (Y) de um processo. O rendimento de um processo é uma função dos níveis de temperatura e concentração de alimentação, como $Y = f(x_1, x_2) + \varepsilon$, em que ε representa o erro observado na resposta esperada $E(Y) = f(x_1, x_2) = \eta$, então a superfície representada por $\eta = f(x_1, x_2)$ é chamada de superfície de resposta (MONTGOMERY, 2008).

Segundo Montgomery et al. (2001), as equações definidas de superfície de resposta podem ser representadas graficamente e utilizadas de três formas:

- (a) Descrever como as variáveis em teste afetam as respostas;
- (b) Determinar as inter-relações entre as variáveis em teste; e
- (c) Para descrever efeitos combinados de todas as variáveis em teste sobre a resposta.

Ainda segundo Montgomery et al. (2001), algumas considerações devem ser feitas quando utilizamos superfície de resposta, a saber:

- (a) Os fatores que são críticos ao processo são conhecidos;
- (b) A região em que os fatores influem o processo é conhecida;
- (c) Os fatores variam continuamente ao longo da faixa experimental escolhida;

- (d) Existe uma função matemática que relaciona os fatores à resposta medida;
- (e) A resposta que é definida por essa função é uma superfície lisa.

Essa técnica também apresenta algumas limitações que devem ser consideradas e, entre as mais importantes, podemos citar:

- (a) Pode apresentar grandes variações dos fatores, o que pode resultar em conclusões falsas;
- (b) Se os fatores críticos não foram especificados corretamente, o resultado obtido pode estar incorreto;
- (c) Em alguns casos a região de ótimo pode não ser determinada devido ao uso de uma faixa muito estreita ou ampla dos dados;
- (d) Como em qualquer experimento, resultados distorcidos podem ser obtidos se os princípios clássicos da experimentação não forem seguidos (casualização, repetição e controle local); e
- (e) Superestimar a computação: o pesquisador deve utilizar de bom senso e seu conhecimento sobre o processo para chegar a conclusões apropriadas a seus dados.

A aplicação dessa metodologia foi realizada inicialmente na indústria química, tendo sido seus fundamentos formalizados por Box e Draper (1987). No campo agrônomo, o uso concentrou-se no estudo do rendimento de cultivares, como efeito de níveis de nutrientes aplicados ao solo, incluindo-se outros fatores, como: densidade de plantio e lâminas de irrigação.

Dentre as vantagens da metodologia de superfície de resposta, a principal é que seus resultados são resistentes aos impactos de condições não ideais, como erros aleatórios e pontos influentes, isso porque a metodologia é robusta. Outra vantagem é a simplicidade analítica da superfície de resposta obtida, pois a metodologia gera polinômios. Em geral, polinômios de duas ou mais variáveis são funções contínuas.

Após o ajuste do modelo aos dados, é possível estimar a sensibilidade da resposta aos fatores, além de determinar os níveis dos fatores nos quais a resposta é ótima, por exemplo: máxima ou mínima.

2.1.3.1 Modelo de Primeira Ordem

Na maioria dos problemas em superfície de resposta, a forma do relacionamento entre as variáveis dependentes e independentes é desconhecida. Assim, o primeiro passo é encontrar uma aproximação para o verdadeiro relacionamento entre a variável resposta e as variáveis

independentes (fatores). Geralmente, utiliza-se de uma regressão polinomial de baixo grau em alguma região das variáveis independentes. A forma geral para o modelo de primeira ordem (ou modelo de grau um), contendo k variáveis independentes de entrada, pode ser representado por:

$$Y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \varepsilon \quad (16)$$

onde Y é a variável resposta observada; $\beta_0, \beta_1, \dots, \beta_k$ são parâmetros desconhecidos; e ε é o termo do erro aleatório. Se ε tem média zero, então a porção não aleatória do modelo geral de primeira ordem representa a verdadeira resposta média, η , que é:

$$\eta = E(Y) = \beta_0 + \sum_{i=1}^k \beta_i x_i \quad (17)$$

e ε é tido como erro experimental. Se, entretanto, o modelo descrito é inadequado para representar a verdadeira resposta média, então ε contém, adicionalmente ao erro experimental, um erro não aleatório (sistemático). Este último erro é atribuído a omissão de termos em X_1, X_2, \dots, X_k de grau superior a um que podem ser entendidos como outras variáveis as quais têm alguma influência sobre a variável resposta. Este erro adicional (excluindo o erro experimental) é chamado de falta de ajuste.

Escrevendo o modelo em notação matricial, considerando N observações, temos:

$$Y = X\beta + \varepsilon \quad (18)$$

em que Y é um vetor de N ($N \geq k+1$) observações: $\beta' = [\beta_0 \beta_1 \dots \beta_k]$ é um vetor $(k+1) \times 1$ de parâmetros desconhecidos, $\varepsilon' = [\varepsilon_1 \varepsilon_2 \dots \varepsilon_N]$ é um vetor $N \times 1$ de erros, e X é uma matriz $N \times (k+1)$ de coeficientes dos parâmetros que compreendem os níveis das variáveis independentes. Assume-se que os erros aleatórios são $NID(0, \sigma^2)$, isto é, independentemente distribuídos com distribuição normal de média zero e variância comum. Quando a matriz X é de posto coluna completo, então o estimador de mínimos quadrados ordinários de β pode ser obtido por:

$$\hat{\beta} = (X'X)^{-1} X'Y \quad (19)$$

e a matriz de variância e covariância é dada por:

$$V(\hat{\beta}) = (X'X)^{-1} \sigma^2 \quad (20)$$

Na maioria dos casos, os cálculos para estimativa dos parâmetros podem ser simplificados codificando os níveis das k variáveis independentes X_i por meio de:

$$x_{iu} = \frac{2(X_{iu} - \bar{X}_i)}{R_i} \quad (21)$$

onde $i = 1, 2, \dots, k$ e $u = 1, 2, \dots, N$. Em que $\bar{X}_i = \frac{\sum_{u=1}^N X_{iu}}{N}$ e R_i é a diferença entre o maior e menor valor dos níveis. A codificação apresenta a característica: $\sum_{u=1}^N x_{iu} = 0$; onde $i = 1, 2, \dots, k$

2.1.3.2 Modelo de Segunda Ordem

Na falta de conhecimento suficiente acerca da forma da verdadeira superfície de resposta, geralmente o experimentador tenta a aproximação pelo modelo de primeira ordem. Quando, entretanto, o modelo de primeira ordem apresenta falta de ajuste para a superfície, incorpora-se termos de ordem superior.

Quando o experimentador está relativamente próximo do “ótimo”, um modelo que incorpora a curvatura é usualmente requerido para aproximar a resposta. A forma geral para o modelo de segunda ordem (ou modelo de grau dois) contendo k variáveis independentes de entrada, pode ser representado por:

$$Y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \beta_{ii} x_i^2 + \sum_{j=2}^{k-1} \sum_{i=1}^k \beta_{ij} x_i x_j + \varepsilon \quad (22)$$

Em que X_1, X_2, \dots, X_k são as variáveis independentes que tem influência na resposta Y ; $\beta_0, \beta_i (i = 1, 2, \dots, k), \beta_{ij} (j = 1, 2, \dots, k)$ são parâmetros desconhecidos e ε é um erro aleatório.

Utilizando variáveis codificadas, x_i , obtidas por: $x_{iu} = \frac{X_{iu} - \bar{X}_i}{S_{X_i}}$ $\{u = 1, 2, \dots, N; i = 1, 2, \dots, k\}$ em que X_{iu} é o u -ésimo nível da i -ésima variável independente, $\bar{X}_i = \frac{\sum_{u=1}^N X_{iu}}{N}$ é a média dos valores X_{iu} , $S_{X_i} = [\sum_{u=1}^N (X_{iu} - \bar{X}_i)^2 / N]^{1/2}$ é o desvio padrão, e N é o número de observações. Sem perda de generalidade podemos considerar os valores de X_i sendo substituídos pelos correspondentes valores $x_i (i = 1, 2, \dots, k)$. Os valores da variável resposta obtidos com as variáveis codificadas podem, então, ser representados como:

$$Y_u = \beta_0 + \sum_{i=1}^k \beta_i X_{iu} + \sum_{i=1}^k \beta_{ii} X_{iu}^2 + \sum_{i=1}^{k-1} \sum_{j=2}^k \beta_{ij} X_{iu} X_{ju} + \varepsilon_u \quad (23)$$

Em que ε_u é o erro experimental em Y_u . Assume-se que os valores de ε_u sejam independentemente distribuídos como variáveis aleatórias com média zero e variância σ^2 .

Na forma de matriz esse modelo pode ser escrito da seguinte maneira:

$$Y = X\beta + \varepsilon \quad (24)$$

Em que $Y = (Y_1 Y_2 \dots Y_N)$, X é uma matriz $N \times P$ de coeficientes dos parâmetros que compreende os níveis das variáveis independentes; $p = \frac{(k+1)(K+2)}{2}$; β é um vetor $p \times 1$ de parâmetros desconhecidos e $\varepsilon' = (\varepsilon_1 \varepsilon_2 \dots \varepsilon_N)$.

2.1.4 Regressão Não-Linear

Até o início da década de 70, as principais técnicas desenvolvidas para os modelos de Regressão Não-Linear se restringiam à suposição de normalidade para a variável resposta e, mesmo após a extensão da distribuição da variável resposta para a família exponencial de distribuições, quando os modelos lineares generalizados foram desenvolvidos por Nelder e Baker (1972), os modelos normais não-lineares continuaram recebendo um tratamento especial, surgindo diversos artigos científicos na mesma década e em décadas posteriores, destacando-se o livro de Ratkowsky (1983), que descreve vários modelos normais não-lineares, segundo diversos aspectos.

Alguns trabalhos importantes retratam sobre os principais aspectos dos modelos não-lineares, dentre eles tem-se: Cordeiro et al. (2000) propuseram uma correção das estimativas de máxima verossimilhança na classe simétrica de modelos de regressão não-linear. Cysneiros e Vanegas (2008) fizeram um estudo analítico e empírico das propriedades estatísticas dos resíduos nos modelos simétricos de regressão não-linear. Vanegas e Cysneiros (2010) propuseram procedimentos de diagnóstico baseados no método de eliminação de casos para modelos de regressão não-linear simétricos. Cancho et al. (2011) introduziram o modelo de regressão não-linear skew-normal e apresentaram uma análise inferencial completa.

É aconselhável a utilização da regressão não-linear no lugar de regressão de mínimos quadrados comum quando não for possível modelar adequadamente a relação com parâmetros lineares. Parâmetros são lineares quando todos os termos no modelo são aditivos e contêm somente um parâmetro que multiplica o termo.

Modelos de regressão não-linear baseiam-se na observação de dados de uma variável resposta descritos por uma função de uma ou mais variáveis explicativas com seus parâmetros de forma não-linear. O objetivo da aplicação desse modelo é identificar e estabelecer a relação entre as variáveis independentes (explicativas) com a resposta. Enquanto os modelos lineares definem, em geral, relações empíricas, os modelos não-lineares podem ir além, motivados pelo conhecimento do tipo de relação entre as variáveis. Sendo assim, podem ter aplicação nas diversas áreas onde relações físicas, biológicas, químicas, mineralógicas, dentre outras, são

estabelecidas por funções não-lineares que devem ter parâmetros estimados a partir de dados observados ou experimentais.

O modelo de regressão não-linear pode ser escrito como:

$$y_i = f(x_i, \theta) + \varepsilon_i \quad i = 1, 2, \dots, N \quad (25)$$

onde y_i é a variável resposta, $f(.)$ é uma função não-linear contínua, com forma conhecida do vetor de variáveis explicativas x_i , e os parâmetros desconhecidos θ . Os erros aleatórios ε_i podem ser considerados independentes e identicamente distribuídos, normais com média zero e variância (σ^2) constante.

Segundo Bates e Watts (1988) um modelo é classificado como não-linear se pelo menos uma das derivadas parciais da função não-linear em relação aos parâmetros depende de pelo menos um dos parâmetros do modelo. Uma das principais características desses modelos é que os mesmos em geral são deduzidos a partir de suposições teóricas e os parâmetros resultantes são interpretáveis.

Existem várias classes de modelos de regressão não-linear, dentre eles, os que historicamente têm proporcionado uma boa motivação para o desenvolvimento estatístico em modelos de regressão não linear são: Modelos de crescimento; Modelos de Rendimento e Modelos Compartimentados.

A obtenção das estimativas dos parâmetros pode ser realizada de várias maneiras, dentre elas, por meio do método dos mínimos quadrados ou pelo método da máxima verossimilhança. Para os modelos de regressão não-lineares, o sistema de equações normais não é resolvido facilmente, sendo necessários métodos iterativos na obtenção dessas estimativas.

A escolha da função muitas vezes depende de conhecimento prévio sobre a forma da curva de resposta ou o comportamento de propriedades físicas ou químicas no sistema de interesse. Formas não-lineares potenciais incluem crescimento ou declínio côncavo, convexo ou exponencial e curvas sigmoidais e assintóticas. É necessário especificar a função que satisfaz aos dois requisitos de conhecimento prévio e pressuposições de regressão não-lineares.

Para especificar uma nova função, é necessário que ela contenha pelo menos um dos três componentes básicos, sendo eles: parâmetros, preditores ou operações e funções matemáticas.

O software utilizado para fazer a regressão não-linear nesse trabalho (MINITAB®) estima parâmetros ajustando a função expectativa aos dados usando um algoritmo alternativo que minimiza a Soma dos Quadrados do Erro residual (SSE).

Os preditores são as variáveis independentes a serem utilizadas na construção da equação para obtenção da resposta.

As operações e funções matemáticas são obtidas pela relação matemática entre os parâmetros e os preditores, os quais produzem o valor esperado da variável de resposta.

A regressão não linear minimiza a soma dos quadrados do erro residual (SSE) para estimar os parâmetros, porém, no caso da regressão não-linear, não existe solução direta para minimizar a soma dos quadrados do erro residual. O MINITAB® utiliza um algoritmo iterativo para estimar os parâmetros ajustando-os sistematicamente às estimativas dos parâmetros para reduzir a soma dos quadrados do erro residual. Depois de escolher o modelo de regressão, define-se o algoritmo e o valor inicial para cada parâmetro. O algoritmo baseia-se nesses valores iniciais para calcular a soma inicial dos quadrados do erro residual.

Para cada iteração, o algoritmo ajusta as estimativas dos parâmetros de forma a reduzir a soma dos quadrados do erro residual em relação à iteração anterior. Algoritmos diferentes usam abordagens diferentes para determinar os ajustes em cada iteração. As iterações continuam até que o algoritmo converge na menor soma dos quadrados do erro residual, até que um problema impeça a iteração subsequente ou ainda se o número máximo de interações seja extrapolado. Se o algoritmo não convergir, é aconselhável alterar os valores dos parâmetros iniciais ou então tentar com outro algoritmo.

Para algumas funções de expectativas e conjuntos de dados, os valores iniciais podem afetar significativamente os resultados. Alguns valores iniciais podem impossibilitar a convergência ou convergir para uma soma de quadrados local ao invés de global. Algumas vezes é necessário muito trabalho para se desenvolver bons valores iniciais.

2.2 SIMULAÇÃO E COSSIMULAÇÃO GEOESTATÍSTICA

A krigagem proporciona a estimativa em um ponto não amostrado com base nas informações dos pontos vizinhos amostrados. Essa estimativa é feita minimizando-se a variância do erro de estimativa. A minimização da variância do erro envolve a suavização das dispersões reais. Essa suavização ocorre mesmo que as estimativas sejam condicionais aos

pontos amostrais. Entretanto, esse condicionamento não garante que a estimativa resultante não esteja suavizada.

Todos os algoritmos de interpolação tendem a suavizar a variabilidade do atributo. Essa suavização se caracteriza pela subestimativa de valores altos e superestimativa de valores baixos (GOOVAERTS, 1997). Segundo Goovaerts (1997), a suavização não é uniforme, pois é “zero” nos pontos amostrais e vai aumentando à medida que se distancia dos pontos de dados.

O efeito de suavização da estimativa pode ser comparado com o histograma entre amostra e pontos estimados; Diagrama P-P (probabilidade x probabilidade) e variograma da estimativa x amostras, entre outros.

A consequência da suavização de krigagem é que o efeito de suavização não reproduz adequadamente as características das amostras usadas para fazer as estimativas em pontos não amostrados. Dessa forma, o processo de inferência do fenômeno espacial em estudo não pode ser realizado com exatidão, pois não permite concluir corretamente sobre a distribuição e variabilidade espacial da variável de interesse.

Segundo Olea (2012), a simulação estocástica foi a solução adotada pela geoestatística para resolver o problema da suavização de krigagem. Entretanto, segundo ele, a simulação ainda não é a solução perfeita, pois ganha-se em precisão global em detrimento da precisão local.

Deutsch (2002) salienta que a simulação estocástica também foi a solução adotada pelos geoestatísticos para modelar a incerteza associada à estimativa, uma vez que a variância de krigagem foi reconhecida apenas como um índice da qualidade da configuração espacial dos pontos vizinhos próximos.

De toda forma, os métodos de simulação disponíveis podem ser aplicados, devendo os resultados ser analisados com muita atenção, pois algumas realizações podem mostrar cenários muito distintos da realidade. A simulação estocástica é uma ferramenta poderosa para descrever fenômenos dos quais suas complexidades não podem ser descritas de forma determinística.

Em algumas aplicações, as variáveis de interesse podem ser simuladas independentemente. Na maioria das aplicações, no entanto, é necessário utilizar método estocástico de variabilidade espacial conjunta de vários atributos condicionados por um modelo prévio de covariâncias e covariâncias cruzadas. Em mineração, por exemplo, os diversos teores de óxidos de um depósito polimetálico devem ser simulados conjuntamente, desde que as variáveis sejam geneticamente correlacionadas.

Nas décadas de 70, 80 e 90, foi destinada uma quantidade significativa de esforços para o desenvolvimento de algoritmos para simular conjuntamente variáveis aleatórias interdependentes. Journel e Huijbregts (1978) usaram um modelo linear de correionalização para simular conjuntamente material estéril e uma camada mineralizada em níquel. Carr e Myers (1985) apresentaram uma técnica baseada na generalização do algoritmo de “*turning bands*” para a simulação de uma função aleatória vetorial $Z(u) = \{Z_1(u), Z_2(u), \dots, Z_k(u)\}$. Luster (1985) apresentou um estudo de caso de simulação conjunta do CaCO_3 e MgCO_3 em cimento do tipo Portland restringido a relações estequiométricas. Suro-Perez e Journel (1991) utilizaram simulação conjunta de diversas variáveis categóricas através da simulação dos seus componentes principais. Verly (1992) propôs a generalização da técnica de simulação sequencial gaussiana por simulação de uma função aleatória vetorial. Apesar de todas essas técnicas produzirem resultados razoáveis, elas possuem em comum a desvantagem de exigir a inferência e modelamento da covariância e covariância cruzada, complicando um pouco o caso dos métodos vetoriais, os quais requerem uma solução numérica intensiva, exigindo muito processamento computacional dos sistemas de krigagem. Consequentemente, faz-se necessária uma solução adicional para a indústria, a qual seja de fácil aplicação e continue sendo acurada.

A cossimulação colocada pode ser utilizada como uma solução alternativa para reduzir o esforço computacional envolvido no processo de simulação, além de deixar a matriz de cokrigagem mais estável. Essa ideia implica em manter o dado da variável secundária próximo do local onde a variável primária está sendo estimada (dado secundário colocado).

Nesse trabalho optou-se por usar o modelo de correionalização de Markov para simplificar a inferência e o ajuste dos variogramas cruzados. Esse modelo se baseia no argumento de que a variável secundária mais próxima, em particular, do dado colocado, minimiza a influência dos dados semelhantes mais distantes.

Métodos de simulação estocástica são procedimentos que envolvem a geração de números aleatórios com o objectivo de explorar o espaço de incerteza ou campo de possibilidades de um dado fenômeno físico ou qualquer outro tipo de variável de estudo cujo comportamento possa ser quantificado matematicamente. Devido à aleatoriedade envolvida, os métodos de simulação estocásticos são também conhecidos como métodos de Monte Carlo. O nome Monte Carlo é uma referência ao famoso cassino em Mônaco. A aleatoriedade e a repetição são as principais características dos métodos de Monte Carlo, que são análogas às atividades praticadas num cassino.

A simulação estocástica visa imitar ou replicar o comportamento de sistemas complexos explorando a sua aleatoriedade para obter cenários das possíveis saídas desses sistemas. Os dois principais requisitos para os métodos de simulação estocástica são: possuir o conhecimento da distribuição de probabilidade das variáveis de entrada do sistema e possuir um gerador de números aleatórios para gerar cenários das variáveis de entrada do sistema.

A geração de números aleatórios é o alicerce de qualquer sistema de simulação estocástica. Porém, nos programas computacionais as conhecidas funções que geram números aleatórios não são efetivamente aleatórias. Na prática, o que se usa em simulação estocástica são os geradores de números pseudo-aleatórios. Esses geradores produzem uma sequência determinística de números inteiros ou em ponto flutuante na precisão do computador, que imita uma sequência de variáveis aleatórias independentes e uniformemente distribuídas entre 0 e 1.

A essência de uma sequência de números pseudo-aleatórios é a sua imprevisibilidade, no sentido de que ninguém é capaz de, ao vê-la, dizer qual é a regra determinística que a produz e conseguir prever qual é o próximo número da sequência.

Segundo Deutsch (2002), o método de simulação gaussiana sequencial (SGS) é o mais utilizado na modelagem de reservatórios de petróleo e depósitos minerais por sua simplicidade, flexibilidade e razoável eficiência. Existem outros algoritmos de simulação estocástica, os quais não são extensivamente utilizados por apresentarem restrições e problemas nos resultados, dentre eles se encontram:

- (a) Decomposição LU da matriz dos coeficientes, que tem a restrição da ordem N para sua resolução;
- (b) Métodos espectrais com base na transformada rápida de Fourier;
- (c) Fractais também não tem sido muito empregados por causa da suposição restritiva da autossimilaridade.
- (d) Métodos de médias móveis são raramente usados por causa do elevado tempo de processamento.

O algoritmo de simulação “*turning bands*” também seria uma boa opção para realização desse trabalho, tendo em vista os avanços tecnológicos que promoveram a redução de artefatos gerados por esse método, o que, até então, era tido como uma das principais vantagens desse método.

O método da simulação sequencial gaussiana pertence à classe de métodos sequenciais, na qual se incluem a simulação indicadora sequencial (DEUSCH e JOURNAL, 1998) e a simulação sequencial direta (SOARES, 2001).

2.2.1 Métodos de simulação sequencial

Por causa da simplicidade de execução, os algoritmos dos métodos sequenciais de simulação são os mais comuns e populares para a reprodução da distribuição espacial e da incerteza de diferentes variáveis em dados geológicos (SOARES, 2001).

Seja uma distribuição com N variáveis aleatórias $\{Z_i, i = 1, N\}$, em que N é muito grande e pode ser:

- (a) Os nós de uma malha densa, considerando-se as variáveis Z_i como medidas do mesmo atributo;
- (b) Os N atributos medidos na mesma localização (x);
- (c) A representação de uma combinação de K diferentes atributos definidos sobre os N^2 nós de uma malha com $N = Kn^2$.

Segundo Goovaerts (1997), o objetivo da simulação sequencial é a geração das várias realizações conjuntas dessas N variáveis aleatórias:

$$\{Z^l(x_j), j = 1, N, l = 1, L\} \quad (26)$$

condicionadas ao conjunto de dados $\{z(x_\alpha), \alpha = 1, n\}$.

É importante ressaltar que a simulação estocástica pode ser condicional, quando passa exatamente pelos pontos amostrais ou condicionantes, ou não condicional.

Deve-se considerar, em seguida, a simulação conjunta de z em somente dois pontos x_1 e x_2 , da qual se obtém um conjunto de pares de realizações geradas por amostragem da função de distribuição acumulada condicional bivariada (GOOVAERTS, 1997):

$$F(x_1, x_2; z_1, z_2 | (n)) = Pr\{Z(x_1) \leq z_1, Z(x_2) \leq z_2 | (n)\} \quad (27)$$

ou seja, o valor $z^l(x_1)$ é simulado com base na função de distribuição acumulada condicional $F(x_1; z_1 | (n))$, a qual é posteriormente atualizada pelo valor previamente simulado $z^l(x_1)$, além dos n pontos de dados (GOOVAERTS, 1997).

Ainda segundo Goovaerts, a equação anterior pode ser generalizada para N variáveis:

$$F_N(x_1, \dots, x_N; z_1, \dots, z_N | (n)) = Pr\{Z(x_i) \leq z_i, i = 1, N | (n)\} \quad (28)$$

a qual, de acordo com ele, pode ser aproximada como produto de N funções de distribuição acumulada condicional, que são determinadas sequencialmente.

$$\begin{aligned}
 F_N(x_1, \dots, x_N; z_1, \dots, z_N | (n)) &= F\{x_1; z_1 | (n)\} \\
 &XF\{x_2; z_2 | (n + 1)\} \dots \\
 &XF\{x_{N-1}; z_{N-1} | (n + N - 2)\} \\
 &XF\{x_N; z_N | (n + N - 1)\}
 \end{aligned} \tag{29}$$

Essa é a fundamentação teórica dos métodos sequenciais de simulação estocástica. Cada novo ponto simulado é usado para atualizar a função de distribuição acumulada condicional, da qual o valor simulado é extraído por Monte Carlo.

Estes processos de simulação sequencial não determinam qualquer critério na ordem escolhida para o caminho aleatório sem repetição (*random path* ou *random walk*), embora seja possível utilizar os valores já simulados no cálculo de nós ainda a simular, implicando, necessariamente, que esta ordem tenha influência no modelo simulado final.

Esta influência é aceita pelo fato de já se ter concluído que o caminho aleatório é o processo estocástico com o menor efeito no modelo final conforme maior for o número de simulações feitas.

2.2.2 Simulação Sequencial Gaussiana - SGS

Simulação Sequencial Gaussiana (SGS) é um tipo de simulação sequencial habitualmente utilizada no ramo da geoestatística e, conseqüentemente, como processo de simulação ou estimação sobre os nós de uma malha no qual cada um deles está condicionado aos restantes simulados anteriormente.

O procedimento foi apresentado por Isaaks (1990), descrito também por Deustch e Journel (1992), e ganhou popularidade na década de 90 devido à sua simplicidade e razoável baixo tempo de processamento computacional. Todo o processo deste tipo de simulação é desenvolvido em ambiente gaussiano e admite a hipótese de multi-gaussianidade para a variável que se pretende simular (SOARES, 2001).

Considerando a simulação de N variáveis aleatórias $\{Z(x_i), i = 1, N\}$, localizados sobre os nós de uma malha regular e condicionadas ao conjunto de n pontos de dados $\{z(x_\alpha), \alpha = 1, n\}$, uma realização SGS é obtida conforme os seguintes passos (GOOVAERTS, 1997):

- (a) Inicialmente, a distribuição da variável $Z(x)$ é transformada para uma distribuição normal por meio de $Y(x) = \varphi(Z(x))$ (em que φ é a função de transformação para os *scores* da distribuição normal), com média nula, $E[Y(x)] = 0$, e variância unitária, $Var[Y(x)] = 1$.
- (b) Em seguida, o variograma experimental da variável transformada $Y(x)$ é calculado e obtém-se o modelo de correlação espacial $\gamma_Y(h)$ que é usado na SGS.
- (c) Em seguida, a simulação sequencial é feita para a variável $Y(x)$;
- (d) Ao final da SGS obtém-se o conjunto de valores simulados $\{y^l(x_i), i = 1, N\}$ que estão no domínio da distribuição Gaussiano. Desse modo, esses valores devem ser transformados de volta para a escala original da variável:

$$Z^l(x_i) = \varphi^{-1}(y^l(x_i)), i = 1, N \quad (30)$$

De acordo com Deutsch e Journel (1998), a decisão prévia de estacionaridade requer que seja usada krigagem simples com média zero para a estimativa do nó a ser simulado.

O método da simulação sequencial gaussiana necessita de uma transformação dos dados originais para ambiente gaussiano. Isto pode ser problemático nos casos em que o histograma da variável a simular é muito assimétrico tendo como consequência a dificuldade na reprodução do variograma (SOARES, 2001).

A SGS pode ser utilizada em cossimulação a partir do método de cossimulação sequencial gaussiana que utiliza a cokrigagem com suporte ao cálculo da média e variância em cada nó da malha de simulação.

O método de simulação gaussiana sequencial (SGS) apresentado por Isaaks e Srivastava (1989) é, segundo Koltermann e Gorelick (1996), o mais poderoso dos algoritmos gaussianos de geração de campos aleatórios. Os referidos autores destacam o fato de que a SGS, por definição, gera simulações condicionais (ou condicionadas). De Marsily et al. (1998) destacam a SGS como sendo provavelmente o método mais simples de ser utilizado.

Uma descrição do método de simulação sequencial gaussiana (SGS) pode ser encontrada, por exemplo, em Goovaerts (1997) e Deutsch e Journel (1998), bem como em Yamamoto e Landim (2013).

2.2.3 Simulação Sequencial Direta - SSD

Simulação Sequencial Direta (SSD) é um tipo de simulação sequencial habitualmente utilizada no ramo da geoestatística. É utilizada para estimativa e especialmente simulação

estocástica dos nós de uma malha na qual cada um deles está condicionado aos restantes simulados anteriormente.

Essa simulação é chamada de direta porque, ao contrário de outras metodologias, como é o caso da simulação sequencial gaussiana ou simulação sequencial da indicatriz, não necessita de qualquer transformação da variável original (SOARES, 2001). Foi originalmente descrita por Andre Journel em 1994 mas só em 2001 com Amílcar Soares foi desenvolvida uma implementação que correspondesse aos critérios de uma simulação sequencial.

A SSD utiliza as médias e variâncias locais para uma re-amostragem da Função de Distribuição Cumulativa (FDC) original, de modo a construir uma nova função centrada na média local e com amplitude derivada da variância local, sendo estes parâmetros locais resultados da krigagem simples.

$$z_{KS}^*(x_0) = m + \sum_{i=1}^n \lambda_i [z(x_i) - m] \quad (31)$$

O método utiliza uma lei de distribuição gaussiana como função auxiliar, porém não se trata de uma transformação dos dados originais para ambiente gaussiano como a SGS, para simular uma variável $Z(x)$ em N nós sobre uma região.

Ao contrário da SGS, que usa a média e a variância condicionais para a função de distribuição acumulada condicional, essas estatísticas determinam o intervalo de valores da função de distribuição acumulada global $F_z(z)$, dentro do qual o valor simulado é amostrado por Monte Carlo (SOARES, 2001). Para definir esse intervalo, esse autor propõe selecionar um conjunto de valores condicionantes $\{z(x_i), i=1, n\}$, de tal modo que a média e a variância desses n pontos de dados sejam iguais à estimativa por krigagem simples e à variância de krigagem simples, respectivamente (YAMAMOTO e LANDIM, 2013):

$$\frac{1}{n} \sum_{i=1}^n z(x_i) = z_{KS}^*(x_0) \quad \text{e} \quad \frac{1}{n} \sum_{i=1}^n [z(x_i) - z_{KS}^*(x_0)]^2 = \sigma_{KS}^2(x_0) \quad (32)$$

Contudo, é muito difícil definir o intervalo de valores centrado em $z_{KS}^*(x_0)$, pois os dados em $Z(x)$ não estão igualmente espaçados. Assim, Soares (2001) propõe usar a função de distribuição acumulada Gaussiana, que pode ser caracterizada pela média e variância condicionais.

$$y_{KS}^*(x_0) = \varphi(z_{KS}^*(x_0)) \quad \text{e} \quad \sigma_{KS}^2(x_0) = \frac{1}{n} \sum_{i=1}^n [y(x_i) - y_{KS}^*(x_0)]^2 \quad (33)$$

em que φ é a função de transformação para os scores da distribuição normal.

A função de distribuição acumulada Gaussiana é, então, amostrada por Monte Carlo. Mas, como o valor amostrado está no domínio gaussiano, o valor simulado é obtido aplicando-se a transformação reversa (SOARES, 2001):

$$z_{SSD}^l(x_0) = \varphi^{-1}(y_{SSD}^l(x_0)) \quad (34)$$

Trata-se, portanto, de um método misto entre uma SGS e uma SSD, esta sem a necessidade de transformação dos valores originais para os “scores” da distribuição normal. Essa transformação é usada apenas para localizar o intervalo de valores para a construção da função de distribuição acumulada Gaussiana, da qual o valor simulado é extraído por Monte Carlo. De qualquer modo, a simulação sequencial direta é uma aproximação perfeitamente válida e teoricamente correta, que, dependendo da aplicação, deve ser considerada como alternativa viável.

2.2.4 Simulação indicadora sequencial – SIS

Os métodos sequenciais envolvem a obtenção, em cada ponto a ser simulado, de uma função de distribuição acumulada condicional, a qual é amostrada por Monte Carlo e resulta no valor simulado. O que diferencia as duas aproximações descritas (SGS ou, SSD e SIS) é a forma pela qual são definidas essas funções de distribuição acumulada condicional.

No caso da SGS, essa função é determinada pela média e a variância condicionais resultantes da krigagem simples dos valores transformados para os escores da distribuição normal. A SSD calcula a média e a variância condicionais dos valores originais, as quais definem o intervalo de valores para a obtenção de uma função de distribuição acumulada Gaussiana, da qual o valor simulado é extraído aleatoriamente.

Nesse caso, a média e a variância condicionais devem ser tais que identifiquem a estimativa e a variância obtidas por krigagem simples.

A simulação indicadora sequencial (SIS) também faz a amostragem da função de distribuição acumulada condicional por Monte Carlo, mas com a diferença de que essa função é obtida por meio da krigagem indicadora. Esse método apresenta grande vantagem de ser aplicado a variáveis aleatórias contínuas ou categóricas.

Segundo Deutsch e Journel (1998), a principal contribuição do método da indicadora é a avaliação direta das probabilidades condicionais, as quais são usadas pelo método de simulação sequencial. A simulação indicadora sequencial é a técnica não gaussiana mais comumente empregada (GOOVAERTS, 1997).

Para a obtenção da função de distribuição acumulada condicional pela krigagem indicadora, procede-se primeiro a transformação para funções indicadoras. Se a variável aleatória for contínua, o intervalo de variação de $Z(x)$ é discretizado por K teores de corte, e as funções indicadoras são assim obtidas:

$$\begin{cases} i(x; z_k) = 1 \text{ se } z(x) \leq z_k \\ i(x; z_k) = 0 \text{ se } z(x) > z_k \end{cases} \quad (35)$$

Por outro lado, uma variável aleatória categórica é definida pelos tipos que a compõem. Assim, as funções indicadoras são determinadas como:

$$\begin{cases} i(x; k) = 1 \text{ se } x \in \text{tipo } k \\ i(x; k) = 0 \text{ se } x \notin \text{tipo } k \end{cases} \quad (36)$$

Após a transformação, tem-se sempre K vetores binários. Para cada ponto a ser simulado, os valores da vizinhança (originais e simulados) são localizados e as k proporções estimadas pela técnica da krigagem indicadora.

Entretanto, ao invés de estimar as funções indicadoras propriamente ditas, Deutsch (2002) propõe usar os resíduos das indicadoras (krigagem simples dos indicadores). Para variáveis aleatórias contínuas discretizadas em K teores de corte, tem-se, para esse autor:

$$F_{KI}(x_0; z_k) = \sum_{\alpha=1}^n \lambda_{\alpha} [i(x_{\alpha}; z_k) - F_{zk}] + F_{zk} \quad (37)$$

em que $F_{zk} = \frac{1}{N} \sum_{\alpha=1}^N i(x_{\alpha}; z_k)$ é a média para o teor de corte z_k .

Para variáveis aleatórias categóricas com K tipos, tem-se, ainda segundo ele:

$$P_{KI}(x_0; k) = \sum_{\alpha=1}^N \lambda_{\alpha} [i(x_{\alpha}) - p_k] + p_k \quad (38)$$

em que $p_k = \frac{1}{N} \sum_{\alpha=1}^N i(x_{\alpha}; k)$ é a proporção do tipo k , de tal modo que $\sum_{k=1}^N p_k = 1$.

Para o caso de variáveis contínuas, o valor simulado é extraído diretamente da Função de Distribuição Acumulada Condicional FDAC, conforme o número aleatório p (entre 0 e 1), obtendo-se os valores em função dos teores de corte z_k . No caso de variáveis categóricas, em que a função de distribuição acumulada condicional é discreta, o tipo simulado é obtido verificando-se a que classe pertence o número aleatório p , conforme Soares (2001):

$$x_0 \in \text{tipo } k \text{ se } p \in [P_{KI}(x_0; k-1), P_{KI}(x_0; K)] \quad (39)$$

Como o procedimento da simulação indicadora sequencial é baseado na krigagem indicadora, isso implica o cálculo e modelagem de K variogramas experimentais. Assim, haverá

K modelos de variogramas diferentes, pois as funções indicadoras poderão apresentar correlações espaciais distintas.

Para variáveis aleatórias contínuas, os teores de corte situados nas caudas da distribuição de frequência de $Z(x)$ tenderão a apresentar poucos pares possíveis no cálculo do variograma experimental, pois somente pares com indicadoras iguais a 1 e 0, ou vice-versa, podem ser acumulados para o cálculo da função variograma. Isso significa que os variogramas correspondentes aos teores de corte nas caudas da distribuição terão poucos pares e, conseqüentemente, serão menos confiáveis por causa das flutuações estatísticas. Além da dificuldade no cálculo e modelagem de K variogramas experimentais quando modelos diferentes são usados, a krigagem indicadora não garante que as probabilidades calculadas para teores de corte crescentes não apresentem problemas na relação de ordem.

Dessa forma, a solução é o cálculo de modelagem de um único variograma experimental correspondente ao teor de corte igual à mediana da distribuição $Z(x)$. O procedimento que utiliza um único modelo de variograma da mediana chama-se krigagem indicadora da mediana (DEUTSCH e JOURNAL, 1998).

No caso de variáveis aleatórias discretas, a obtenção de K modelos de variogramas é uma tarefa muito difícil, pois os tipos tendem a estar agrupados espacialmente, conforme os tipos que compõem a variável categórica.

2.2.5 Cossimulação Geoestatística

Na prática, muito frequentemente, a estimativa de um conjunto de N atributos correlacionados é feita separadamente, uma variável de cada vez, resultando em duas conseqüências principais:

- (a) A estimativa de N variáveis separadamente aumenta o tempo de processamento computacional em N vezes.
- (b) Necessidade de realizar correlações ou variogramas cruzados para cada variável condicionada.

O básico para qualquer algoritmo de estimativa conjunta é cokrigagem (MATHERON, 1971) e (JOURNAL e HUIJBREGTS, 1978). Myers (1982 e 1984) introduziu a formulação da matriz de cokrigagem. A partir dessa formulação a cokrigagem vem sendo utilizada para estimativa conjunta de um vetor de variáveis. Myers (1982) também introduziu a análise de componente principal e a análise canônica para reduzir o tamanho do sistema de equações,

sugerindo um método interativo para resolvê-los. Carr e Myers (1985) desenvolveram um software computacional correspondente à proposta de Myer.

Isaaks e Srivastava (1989) propuseram um sistema de cokrigagem com uma única condição de viés, o qual possui a média das variáveis secundárias normalizadas para a média da variável primária. Dessa forma, a condição tradicional $\sum_{\alpha'} v_{\alpha'} = 0$ para o peso das variáveis secundárias é evitada. Esta condição tende a reduzir a influência das variáveis secundárias, conduzindo a uma solução mais simples.

Journel e Srivastava (1991) apresentaram a ideia de cokrigagem colocada (*collocated cokriging*). Eles desenvolveram esse método baseados na seguinte observação prática:

- (a) quando o dado secundário é demasiadamente amostrado, comparado com o dado primário, o lado esquerdo da matriz do sistema de cokrigagem fica perto da singularidade; e
- (b) considerando que o que interessa é a estimativa da variável primária, não tem por que utilizar variáveis secundárias com informações redundantes. A ideia deles consiste em manter apenas as variáveis secundárias mais próximas da localização da variável primária desconhecida a ser estimada, ou seja, da variável secundária colocada.

Carr e Myers (1985) apresentaram um algoritmo e o software correspondente aos estudos de Xu et al. (1992) para executar a simulação condicional conjunta de diversas funções aleatórias correlacionadas. Para isso foi utilizado o procedimento de simulação condicional intitulado bandas rotativas (*turning bands method*). A etapa de condicionamento para utilização desse método exige um vetor de cokrigagem. As principais desvantagens dessa abordagem são:

- (c) a inferência e modelagem da matriz de covariância e covariância cruzada;
- (d) a limitação do modelo de correionalização linear necessitar de assegurar termos positivos na matriz de cokrigagem; e
- (e) a falta de flexibilidade do algoritmo de bandas rotativas.

Perez e Journel (1991) sugeriram transformar as variáveis indicadoras originais em suas componentes principais lineares, dessa maneira, encurtaria o modelamento da matriz de covariância cruzada do indicador. Os componentes principais são definidos a partir da amostragem da matriz de covariância cruzada normalmente definida em $h_0 = 0$. Esse algoritmo é rápido e não necessita dos modelos de covariância cruzada. No entanto, rigorosamente as correlações cruzadas são reproduzidas apenas para o vetor de separação especificado h_0 . Outra questão é que as variáveis do componente principal tendem a ofuscar a variabilidade espacial

dos valores extremos quando pequenos pesos são dados para seus indicadores extremos correspondentes. Então, se o propósito da simulação é capturar o comportamento espacial de valores extremos, o uso de componente principal pode mascarar o resultado final. Essa técnica é um pouco complicada, pois realiza simulação de variáveis contínuas, necessitando decompor uma matriz grande de covariância.

Verly (1992) apresentou um algoritmo para unir a simulação de diversas variáveis considerando a simulação de função vetorial aleatória. Esse algoritmo é uma generalização da abordagem do vetor da simulação sequencial gaussiana.

As técnicas disponíveis de simulação conjunta podem ser classificadas em três categorias principais:

Simulação vetorial (CARR e MYERS, 1985); (GÓMEZ-HERNÁNDEZ e JOURNAL, 1992) e (VERLY, 1992). Uma função de vetor aleatório $Z(u) = \{Z_1(u), Z_2(u), \dots, Z_k(u)\}$ é considerada quando cada elemento $Z_1(u), Z_2(u), \dots, Z_k(u)$ representa a variável específica, por exemplo, porosidade de um reservatório, permeabilidade horizontal, permeabilidade vertical, etc.

A simulação de um vetor $Z(u)$ fornece diretamente a simulação das N variáveis $Z_1(u), Z_2(u), \dots, Z_k(u)$ respeitando suas matrizes de correlação espacial.

Essa abordagem possui dois inconvenientes principais:

- (a) Ela normalmente requer a inferência do modelamento da matriz de covariância e de variância cruzada, descrito na equação 40.

$$\Sigma = [C_{k,k'}(h) = Cov\{Z_k(u), Z_{k'}(u+h), k, k' = 1, \dots, K\}] \quad (40)$$

- (b) Elevado consumo de processamento computacional para resolver o sistema de equações.
- (c) Além disso, em cokrigagem completa (full cokriging) quando a variável secundária é densamente amostrada, a autocorrelação da variável secundária pode ser muito maior do que a auto correlação da variável primária, por possuir normalmente menor quantidade de dados, podendo resultar em uma matriz instável.

Transformação em fatores independentes (LUSTER, 1985) e (SURO PEREZ, 1991). Essa abordagem consiste em uma transformação linear de N variáveis independentes em N fatores ortogonais. Por exemplo, os N componentes principais da matriz de covariância por qualquer distância h_0 .

A hipótese crítica nesse caso é ter que assumir que a independência obtida do lag h_0 se estende para todas outras distâncias h .

Essa abordagem consiste em realizar a simulação dos fatores independentes. Então, os valores simulados são retro-transformados para os valores originais usando a função inversa da transformação, que garante o grau de dependência com a variável original.

Essa técnica tem duas vantagens principais:

- (a) é rápida;
- (b) informações a respeito da correlação cruzada espacial são introduzidas sem que se tenha nenhum modelo de variância cruzada, caso a separação do intervalo selecionado para obtenção do componente principal seja igual a zero.

As desvantagens principais são:

- (a) a grosso modo, a correlação espacial cruzada é reproduzida somente no intervalo selecionado;
- (b) a falta de significado físico para modelar os variogramas dos fatores independentes. Informações secundárias sobre continuidade ou anisotropia dos dados originais não auxiliam no modelamento do variograma dos fatores independentes.

Sorteio para a distribuição condicional específica (ALMEIDA et al., 1994). Essa abordagem é comumente utilizada na indústria do petróleo, e consiste em simular primeiramente a variável mais importante ou a melhor autocorrelacionada, chamada de primeira variável e denotada por $Z_1(u)$. Então, as outras covariáveis $Z_k(u)$, $k > 1$ são sucessivamente simuladas por sorteio da distribuição condicional específica dos $Z_k(u)$ dado o primeiro valor colocado simulado $z_1(u)$.

No modelamento da variável geometalúrgica, por exemplo, o dado supersecundário (regressão múltipla de alguns óxidos) pode ser simulado primeiramente desde que ele esteja disponível e sua variabilidade espacial seja razoavelmente suave. Então, em qualquer outra localização, a recuperação metalúrgica analisada (DCCG) poderá ser simulada por tiragem aleatória da distribuição condicional da recuperação metalúrgica utilizando o valor da simulação da variável secundária no mesmo local.

Essa abordagem é rápida e simples, mas baseia-se em uma forte simplificação. Na realidade, ela assume que o condicionamento singular da variável secundária colocada seja

suficiente para reproduzir toda importante correlação cruzada e toda autocovariância secundária.

2.2.6 O método de cossimulação utilizado

O método de cossimulação escolhido para ser utilizado nesse trabalho é conhecido como Cossimulação Sequencial Gaussiana, algoritmo **sgcosim**. Ele permite a simulação conjunta de diversas variáveis integrando diferentes informações dos N dados secundários.

Esse algoritmo se baseia em duas ideias centrais, sendo elas:

- (a) A ideia de cokrigagem colocada para reduzir o esforço computacional envolvido no processo de simulação, e para deixar a matriz de cokrigagem mais estável. Essa ideia implica em manter o dado da variável secundária próximo do local onde a variável primária está sendo simulada (dado secundário colocado).
- (b) O modelo de correionalização de Markov para simplificar a inferência e o ajuste dos variogramas cruzados. Esse modelo se baseia no seguinte argumento: a variável secundária mais próxima, em particular, do dado colocado, minimiza a influência dos dados semelhantes mais distantes.

Esse algoritmo pode ser visto como uma generalização do tradicional algoritmo da Simulação Sequencial Gaussiana para lidar simultaneamente com diversas variáveis primárias e diversas variáveis secundárias. O algoritmo se baseia no axioma de Bayes de probabilidade condicional, o qual decompõe uma distribuição qualquer K em um produto de $(K-1)$ distribuições univariadas condicionais. O algoritmo utilizado segue o método de simulação sequencial, onde a simulação conjunta de K eventos dependentes é obtida a partir da elaboração de uma sequência univariada de distribuição condicional K . Para facilitar a determinação dessa distribuição condicional, um modelo Gaussiano multivariado é assumido. Assim, a ideia de cokrigagem colocada é incorporada para reduzir o esforço computacional. Finalizando, o modelo de correionalização de Markov é introduzido para simplificar o modelamento e a covariância cruzada.

3 ESTUDO DE CASO

3.1 VARIÁVEL PRIMÁRIA (*HARD DATA*)

A recuperação metalúrgica global (DCCG ou DCCA) é a variável que mede o quanto de nióbio contido no minério (recuperação mássica em porcentagem) foi separado na fração do concentrado após as etapas de beneficiamento. A DCCG é obtida através de testes laboratoriais feitos com as amostras do minério de nióbio extraídos da mina da CBMM em Araxá-MG.

Os dados de recuperação metalúrgica utilizados nesse trabalho foram obtidos a partir da coleta de amostras de sondagem, realizadas pelo método *Air Core*, o qual não utiliza fluido de perfuração na fase líquida durante a perfuração. As amostras coletadas foram preparadas no laboratório de preparação física da CBMM, seguindo um protocolo que inclui etapas de homogeneização, quarreamento e cominuição. As alíquotas separadas por esse processo foram enviadas ao Laboratório de Tratamento de Minérios da própria CBMM. Essas amostras foram, então, submetidas a testes de bancada que simulam o processo de concentração do pirocloro.

3.1.1 Obtenção do dado primário (experimental)

Para realizar os testes de caracterização, são usadas duas alíquotas de cada amostra, sendo elas: uma alíquota de 400 g para análise granulométrica e análise química e uma alíquota de 3000 g para os ensaios de concentração mineral.

As alíquotas de 3000 g foram submetidas ao teste de avaliação de desempenho industrial na fase de concentração, passando pelas etapas de: peneiramento a úmido para individualização das faixas granulométricas; homogeneização para posterior individualização das alíquotas necessárias a cada teste e secagem na estufa durante 2 horas a 150°C. Para a realização dos testes de concentração mineral, faz-se necessário utilizar diversos insumos, entre eles: reagentes coletores; reagentes umectantes; reagentes antiespumantes; ácidos para estabilização do pH e floculantes.

As amostras foram então submetidas à seguinte sistemática de caracterização para flotação:

- (a) Peneiramento da alíquota de 3.000 g para obtenção das frações granulométricas que seguem fluxos de processos diferentes por devido às suas características;
- (b) Peneiramento da alíquota de 400 g para determinações da partição mássica entre o material já liberado e o que ainda deve ser cominuído até atingir a granulometria de liberação adequada. Essa alíquota é também submetida à análise química do minério;

- (c) Moagem a úmido da fração grossa em um moinho de bolas;
- (d) Após a moagem, o material é retirado do moinho com auxílio de água e colocado em um balde. É feita então, a separação da fração magnética do material com o auxílio de um imã de mão. A fração não magnética é homogeneizada e retira-se uma alíquota dela para o teste de flotação. A fração magnética é seca, pesada e uma alíquota é retirada para análise química;
- (e) A fração fina é homogeneizada e retira-se dela uma alíquota para o teste de flotação;
- (f) As alíquotas da fração grossa, agora moída, e da fração fina são colocadas individualmente em cubas de acrílico ou metal e encaminhadas à célula de flotação em bancada onde são adicionados os reagentes para os testes. São realizadas quatro etapas de flotação, sendo que o rejeito da primeira etapa é o rejeito final do teste (Rejeito Sujo) e o concentrado de cada etapa é a alimentação da etapa seguinte. Os rejeitos da segunda, terceira e quarta etapas são acumulados simulando a carga circulante do processo (Rejeito Limpo). O concentrado da quarta etapa é o concentrado final.

A adição dos reagentes coletor e umectante e também o ácido fluossilícico é fixa e ocorre somente no condicionamento da primeira etapa de cada teste. A dosagem do reagente antiespumante ocorre conforme característica do minério sendo sua dosagem variável em cada teste.

Após a realização do teste, são gerados diversos produtos, sendo eles:

Teste de flotação da fração grossa:

- (a) Rejeito magnético (etapa de separação magnética);
- (b) Rejeito sujo;
- (c) Rejeito limpo;
- (d) Concentrado final.

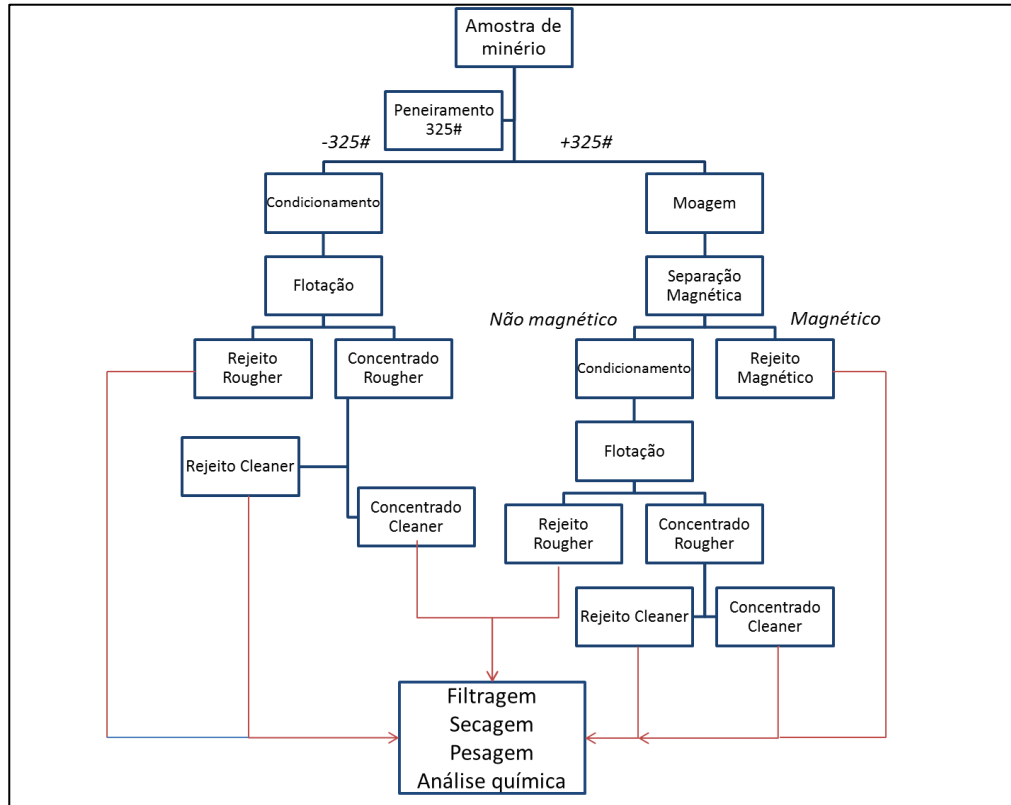
Teste de flotação da fração fina:

- (a) Rejeito sujo;
- (b) Rejeito limpo;
- (c) Concentrado final.

Os produtos dos testes, após secagem, são pesados e encaminhados para análise química.

A figura 2 apresenta um fluxograma esquemático aplicado ao teste de verificação do processo de concentração das amostras de minério de nióbio.

Figura 2- Fluxograma do teste de simulação da concentração.



Fonte: Autoria própria.

Conhecendo as massas e os teores de nióbio de cada um dos produtos obtidos após o teste de flotação, é possível calcular a porcentagem de metal contido nas frações do concentrado *cleaner* para cada uma das frações granulométricas (fino e grosso). A média das recuperações ponderadas pela partição mássica das frações granulométricas resulta na recuperação metalúrgica global.

Os resultados das amostras submetidas ao teste de beneficiamento mineral são validados através de uma rotina de controle de qualidade. A recuperação metalúrgica final obtida por este teste é a variável primária utilizada na cossimulação geoestatística para obtenção de um modelo de distribuição geometalúrgico representativo do minério de nióbio da referida empresa.

3.1.2 Estatística descritiva do dado primário

Com o objetivo de promover um detalhamento amostral do horizonte Alterito Laranja, foi realizado, entre os anos de 2011 e 2012, uma campanha de sondagem utilizando o método *Air Core*, tendo sido realizados 170 furos em diâmetro HQ (4”).

Essa campanha de sondagem, intitulada FSA, foi dimensionada em malha regular e quadrada, com espaçamento de 150 m entre os furos nas direções norte e leste. Essa malha foi confeccionada de forma a adensar a malha pré-existente.

A campanha de sondagem FSA foi realizada pela empresa GEOSOL e a análise química realizada no laboratório da empresa SGS/GEOSOL. O monitoramento dos resultados das análises químicas pelo laboratório externo foi realizado através da inserção de amostras de controle, estabelecendo assim um programa de QAQC para essa campanha. As amostras provenientes dessa campanha tiveram sua preparação física realizada no laboratório da CBMM, sendo posteriormente submetidas aos testes de caracterização metalúrgica e análise química.

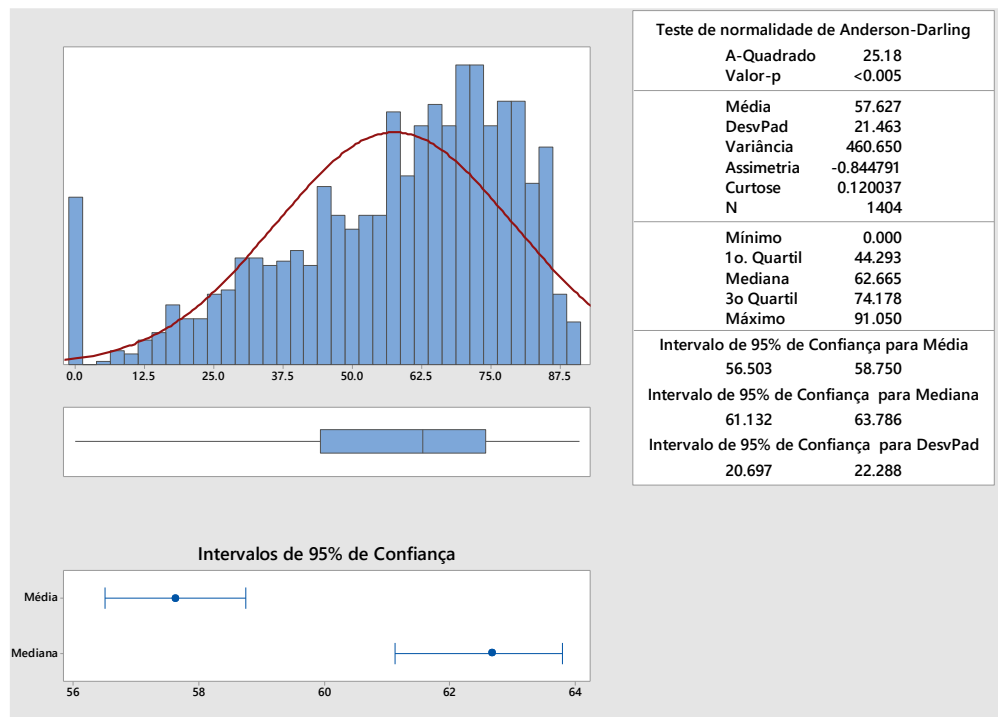
A média dos valores de recuperação metalúrgica de nióbio na litologia Alterito Laranja analisados durante a campanha de sondagem FSA foi de 57,63%, em um total de 1404 amostras. O valor mínimo encontrado foi 0,0%, enquanto o valor máximo foi 91,05%. A mediana, é 62,67%. Já o quartil inferior foi 44,29%, enquanto o quartil superior foi de 74,18%.

A variância absoluta (a priori) das amostras é 460,65%, lembrando que a variância é uma medida de dispersão estatística, o qual indica a distância média quadrática que os valores se encontram do valor esperado (esperança matemática).

O histograma da figura 3 mostra a distribuição de probabilidade da variável primária. É possível observar um pico de valores igual a zero, que correspondendo principalmente a amostras coletas na região de contato entre o Solo Argiloso (estéril) e o Alterito Laranja (zona mineralizada), desconsiderando esse pico, a distribuição dos valores passa a ser ascendente até as proximidades do quartil superior, tendo um comportamento decrescente posteriormente.

A distribuição dos dados primários não se enquadra em uma distribuição normal de probabilidade. A assimetria negativa (-0,844) dessa distribuição revela uma distribuição assimétrica, sendo, nesse caso, a média menor do que a mediana, que por sua vez é menor do que a moda.

Figura 3- Estatística descritiva da variável primária (DCCG).

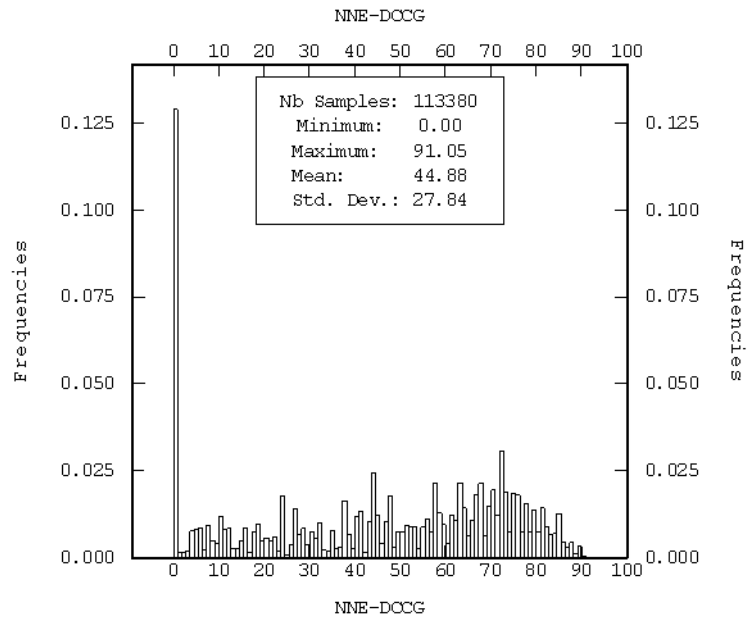


Fonte: Autoria própria.

A obtenção da estatística descritiva da variável DCCG com os dados desagrupados foi feita com base no método dos polígonos de Voronoi. Esse é um método de desagrupamento em que os pesos atribuídos às amostras são diretamente proporcionais à área do Polígono de Voronoi ao seu redor. Em zonas de dados agrupados, as áreas dos polígonos tendem a ser pequenas, recebendo, então, pesos menores. Para realizar esse desagrupamento, foi utilizado o artifício da estimativa pelo método do vizinho mais próximo em um grid de células pequenas, sendo utilizado o mesmo grid da cossimulação.

A figura 4 apresenta o histograma do resultado da interpolação da variável DCCG pelo método do vizinho mais próximo no grid da cossimulação. Os raios de busca utilizados para a interpolação foram de 250 m na direção (N0, D0), 250 m na direção (N90, D0) e 50 m na direção (N0, D90).

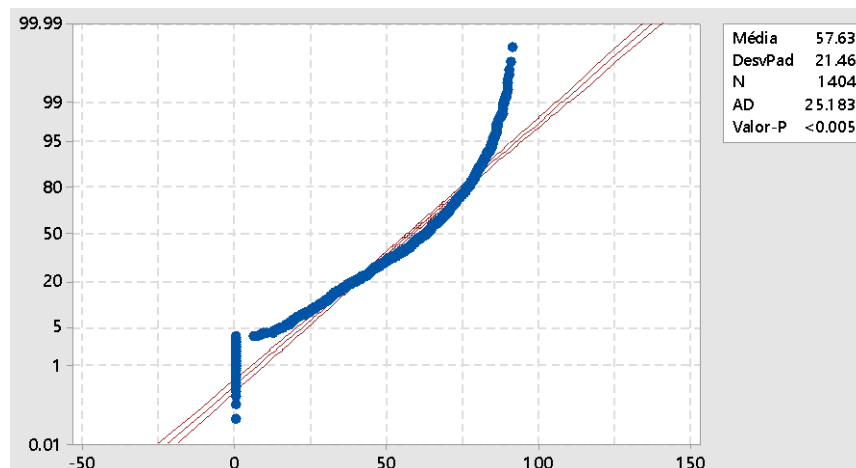
Figura 4- Histograma da variável DCCG desagrupada utilizando o método dos polígonos de Voronoi.



Fonte: Autoria própria.

A partir do gráfico de probabilidade acumulada em escala Gaussiana da figura 5, é possível avaliar o ajuste da distribuição dos dados de recuperação metalúrgica das amostras à distribuição normal. Como as observações não acompanham muito bem as linhas ajustadas, pode-se dizer que a distribuição não se aproxima de uma distribuição Gaussiana.

Figura 5- Gráfico de probabilidade normal da variável primária (DCCG). Eixo x corresponde a variável DCCG e eixo y corresponde à percentagem.

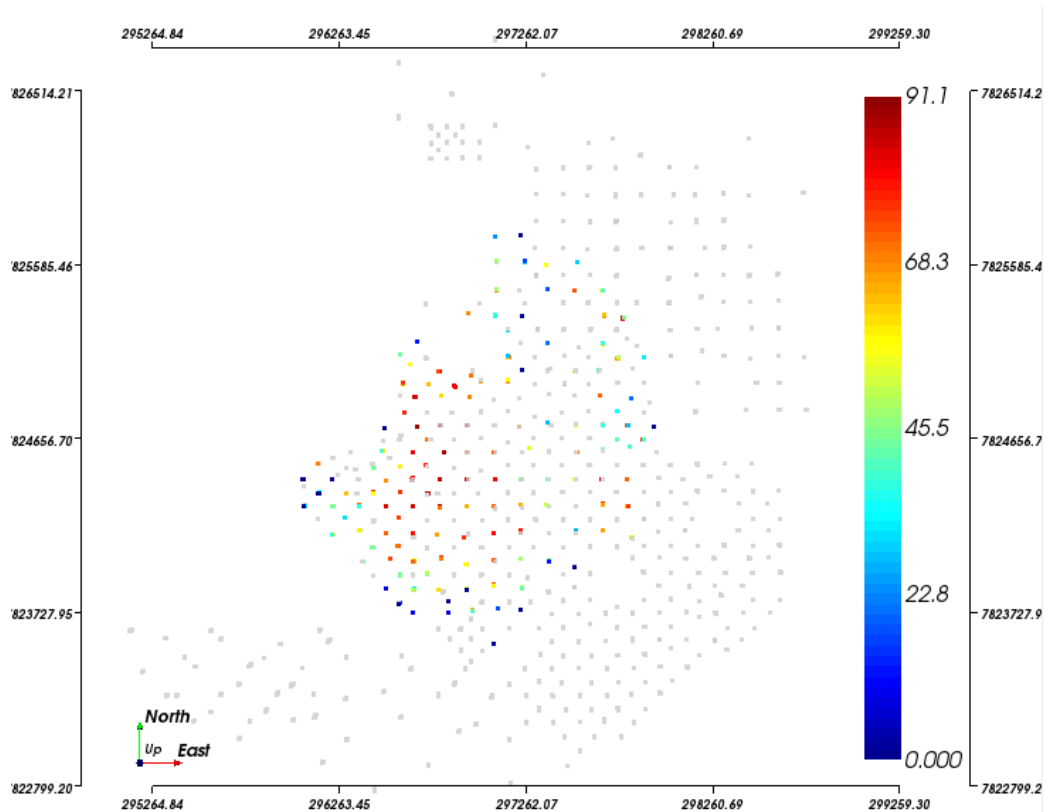


Fonte: Autoria própria.

O mapa da figura 6 mostra a localização dos furos de sondagem. Os pontos cinza representam os furos que não possuem dados de recuperação metalúrgica. Os pontos coloridos

representam os furos da campanha de sondagem FSA, sendo essa utilizada para análise do desempenho geometalúrgico do minério, bem como análise química.

Figura 6- Mapa com a distribuição dos furos contendo os dados primários.



Fonte: Autoria própria.

O dado supersecundário a ser criado visa gerar informações com alta correlação com a variável recuperação metalúrgica e que possa ser usado de forma exaustiva. O modelo resultante incorporando as duas informações aumenta o nível de conhecimento geometalúrgico no depósito mineral de nióbio da CBMM.

3.2 DADO SECUNDÁRIO (*SOFT DATA*)

É possível calcular a recuperação metalúrgica do minério de nióbio a partir da regressão feita com a análise química de alguns óxidos. No entanto, existe um erro associado a esse cálculo, uma vez que as variáveis químicas não explicam na íntegra a resposta metalúrgica analisada. Esse erro pode ser reduzido aplicando-se um método de regressão mais adequado a este tipo de análise.

3.2.1 Obtenção da variável supersecundária

Apesar de as correlações individuais de cada um dos óxidos analisados com a

recuperação metalúrgica experimental ser muito baixa, a interação entre esses óxidos, bem como a contribuição quadrática de cada um deles, demonstrou que a “recuperação metalúrgica calculada” apresentasse uma forte correlação com a “recuperação metalúrgica experimental”.

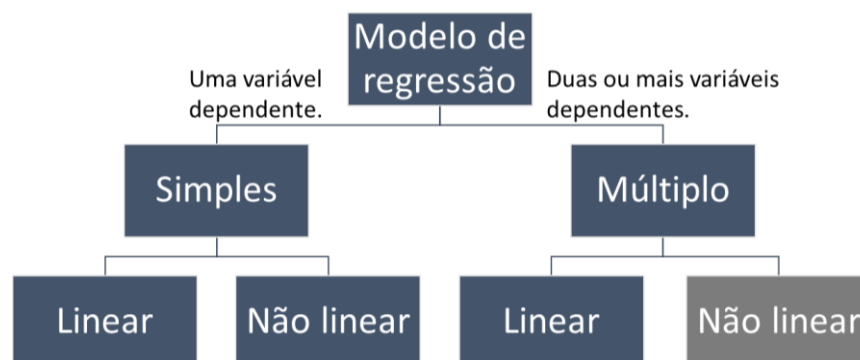
No contexto da estatística experimental, há constante interesse em caracterizar a possível relação entre uma ou mais variáveis resposta e um conjunto de fatores de interesse. Isso pode ser executado através da construção de um modelo que descreva a variável resposta em função dos níveis aplicáveis desses fatores.

Dessa forma, nossa variável supersecundária foi criada a partir da regressão múltipla de cinco óxidos, sendo esses previamente definidos com base na correlação individual entre cada um dos óxidos com a resposta (DCCG). Essa análise foi feita utilizando o método de análise de componentes principais.

Com o objetivo de criar um modelo que explicasse o comportamento da recuperação metalúrgica, foram testados diversos tipos de regressões, sendo posteriormente escolhido o modelo que melhor explicou o comportamento da variável dependente. O método de regressão é uma técnica que permite explorar e inferir a relação de uma variável dependente com variáveis independentes. A regressão designa uma equação matemática que descreva a relação entre duas ou mais variáveis.

As primeiras regressões realizadas foram simples, ou seja, foram analisados os coeficientes de correlação linear de Pearson (PEARSON, 1901) obtidos entre a variável dependente (DCCG) e as variáveis independentes uma a uma. O resultado com essa análise mostrou que, para todas variáveis independentes analisadas, a correlação linear com a variável dependente é muito baixa. A figura 7 apresenta um fluxograma simplificado dos principais modelos de regressão.

Figura 7- Fluxograma com os principais tipos de regressão.



Fonte: Autoria própria

Após serem testados os modelos de regressão linear simples, testaram-se os modelos de regressão mais complexos, dentre eles, o que se gerou um resultado melhor foi o modelo de regressão múltipla não-linear, através do método de superfície de resposta.

A Metodologia de Superfície de Resposta (MSR) é uma técnica estatística utilizada para o modelamento e análise de problemas nos quais a variável resposta é influenciada por vários fatores, cujo objetivo é a otimização dessa resposta. (MONTGOMERY e MYERS, 1995).

Antes de utilizar a técnica de superfície de resposta, é importante definir quais variáveis independentes contribuem para a explicação do fenômeno de interesse, no caso a recuperação metalúrgica de nióbio.

Quando se pretende verificar como as amostras se relacionam, ou seja, o quanto estas são semelhantes, destacam-se dois métodos que podem ser utilizados: a análise de agrupamento hierárquico e a análise fatorial com análise de componentes principais. Nesse trabalho foi adotada a técnica de Análise de Componentes Principais (PCA).

O PCA foi utilizado para definir quais óxidos deveriam ser considerados na regressão multivariada. Esse procedimento matemático utiliza uma transformação ortogonal para converter um conjunto de observações de variáveis possivelmente correlacionadas a um conjunto de valores de variáveis linearmente descorrelacionadas chamadas componentes principais. Esta transformação é definida de forma que o primeiro componente principal tem a maior variância possível e cada componente seguinte, por sua vez, tem a máxima variância sob a restrição de ser ortogonal aos componentes anteriores (VICINI, 2005).

Além de ser possível identificar quais variáveis mais contribuem para explicar um determinado fenômeno, com o PCA também é possível identificar quais variáveis são “redundantes”, ou seja, possuem uma correlação linear muito alta entre si.

O critério utilizado para fazer a afirmação de que as variáveis X_1 , X_2 , X_3 , X_4 e X_5 foram as que mais contribuíram para explicação do fenômeno analisado se baseou na análise do gráfico de carga fatorial, que revela as relações entre as variáveis consideradas no espaço dos dois primeiros componentes principais. Cada variável é um ponto com suas coordenadas dadas pelas cargas sobre as componentes principais 1 e 2. Neste caso, as variáveis X_6 , X_7 , X_8 e X_9 têm cargas baixas e semelhantes para o primeiro componente.

A correlação entre um componente e uma variável na estrutura PCA é chamada de carga e mede a informação que compartilham, dessa forma, as cargas fatoriais representam o quanto um fator explica uma variável na análise fatorial. Algumas variáveis podem ter cargas

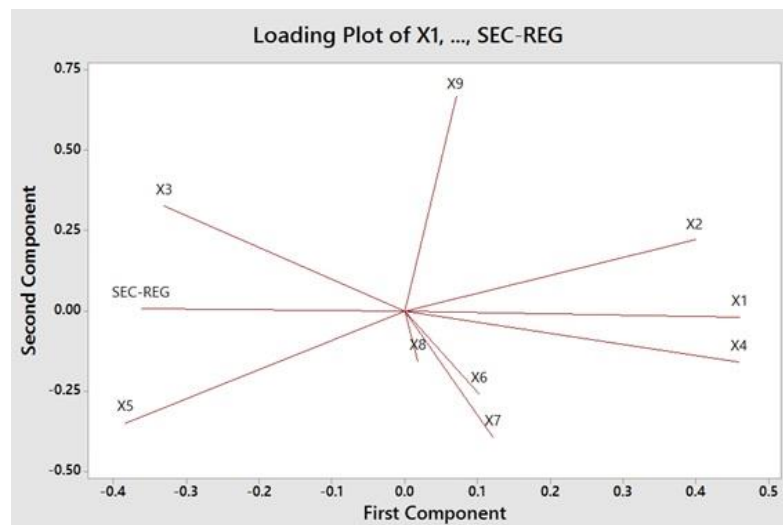
elevadas em múltiplos fatores. As cargas de fator podem variar de -1 a 1. Cargas próximas de -1 ou 1 indicam que o fator afeta fortemente a variável. Cargas próximas de zero indicam que o fator tem um efeito fraco sobre uma variável específica.

O primeiro componente é mais importante do que o segundo componente. Analisando o gráfico de cargas fatoriais da figura 8, constata-se que variáveis (X_1 e X_4) são as variáveis que mais fornecem informações. Sendo que essas duas variáveis também estão fortemente correlacionadas entre si. As variáveis X_6 , X_7 e X_8 estão intimamente associadas e fornecem pouca informação, pois estão próximas de zero com relação ao primeiro componente. A variável X_9 também tem baixa carga no primeiro componente, porém, diferentemente das variáveis X_6 , X_7 e X_8 , a variável X_9 tem alta carga no segundo componente.

De acordo com Abdi e Williams (2010), em geral, significados diferentes de cargas fatoriais levam a interpretações equivalentes dos componentes. Isso acontece porque os diferentes tipos de cargas fatoriais diferem principalmente pelo seu tipo de normalização.

Essa análise de redundância pode ser feita através do gráfico de carga fatorial, conforme ilustrado da figura 8. Analisado esse gráfico de análise fatorial, observa-se que a recuperação metalúrgica “calculada” (SEC-REC), a variável X_3 e a variável X_5 estão no mesmo extremo do gráfico (extremo oeste), o que significa que essas variáveis são positivamente correlacionadas. Já as variáveis que estão no extremo oposto da variável SEC-REC (variáveis X_1 , X_2 e X_4) estão negativamente correlacionadas com a variável SEC-REC.

Figura 8- Gráfico de carga fatorial considerando nove variáveis químicas e a recuperação metalúrgica analisada em laboratório. Os primeiros componentes estão no eixo x e os segundos componentes no eixo y.



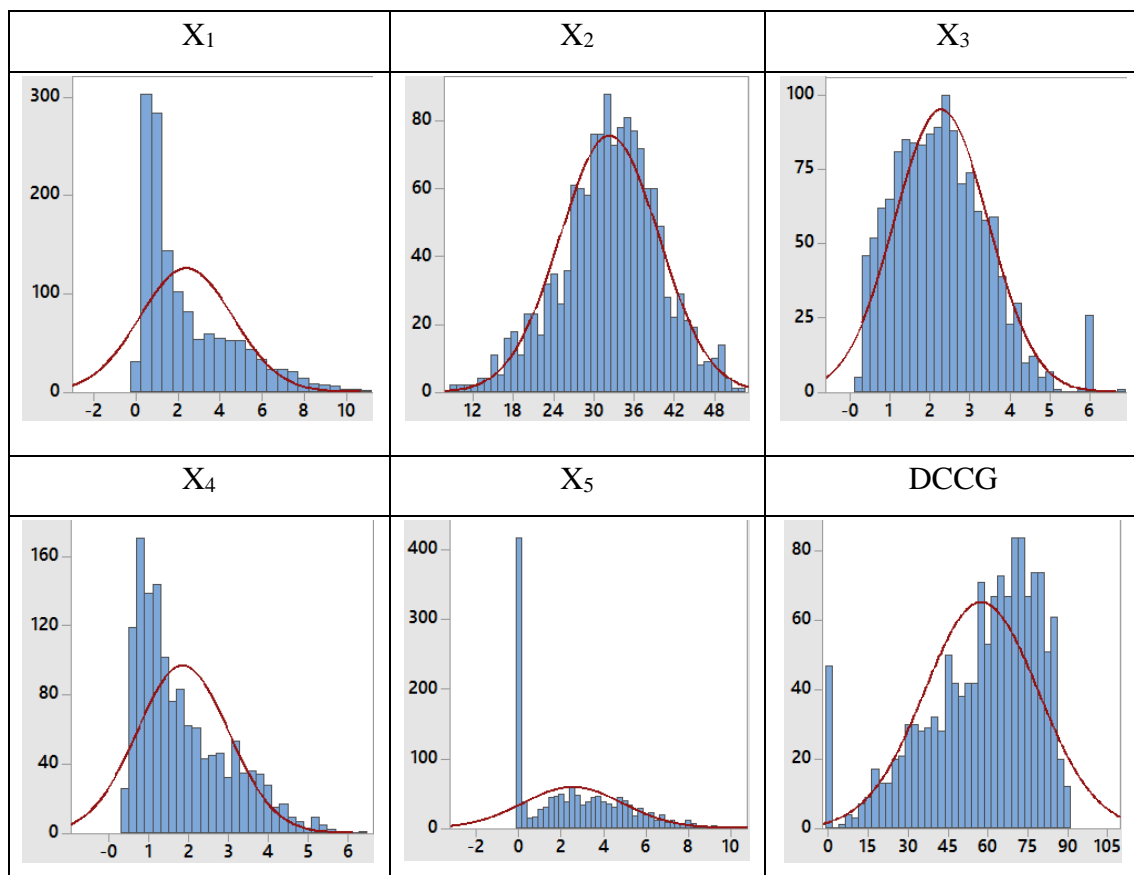
Fonte: Autoria própria.

As variáveis que mais contribuíram para explicar o fenômeno “recuperação metalúrgica” foram: X_1 ; X_2 ; X_3 ; X_4 e X_5 . O software utilizado para fazer as regressões (MINITAB®) possui uma limitação de variáveis a serem utilizadas na regressão, sendo essa de no máximo cinco variáveis. Definidas quais variáveis são importantes para explicação da recuperação metalúrgica, iniciaram-se os trabalhos de modelagem das superfícies de respostas.

As cargas fatoriais quadráticas também poderiam ser usadas para interpretar a relação entre as variáveis. A soma dos coeficientes quadrados de correlação entre uma variável e todos os componentes é igual a 1. Isto se dá pelo fato de que as cargas quadráticas dão a soma da proporção da variância das variáveis explicativas de cada componente principal.

A figura 9 apresenta o histograma de cada uma das variáveis independentes (X_1 ; X_2 ; X_3 ; X_4 e X_5), as quais são óxidos analisados pelo método de Fluorescência de Raio-X, bem como da variável resposta recuperação metalúrgica (DCCG). É possível notar claramente a diferença de escala entre as variáveis, fazendo-se necessário padronizá-las para evitar resultados tendenciosos.

Figura 9- Histograma dos dados das variáveis dependentes e da variável independente.



Fonte: Autoria própria.

Pelo fato de os parâmetros se apresentarem em escalas diferentes, os dados foram “padronizados”, tornando-se adimensionais, ou seja, cada medida foi apresentada como um valor que representa o quanto ela se afasta da média no respectivo parâmetro. As equações de parametrização serão apresentadas posteriormente.

Os métodos adimensionais são aqueles que não possuem unidades convencionais de quantificação para expressarem o resultado obtido e podem ser subdivididos em dois tipos: os que atribuem valores numéricos ou pontos a determinados graus de amplitude de movimentos articulares e os que apenas dicotomizam uma resposta em sim ou não, ou ainda, em positiva ou negativa. Tipicamente, eles não dependem de equipamentos, utilizando-se unicamente de critérios ou mapas de referência preestabelecidos para comparação.

A primeira etapa para a adimensionalização consiste no exame das correlações entre as variáveis observáveis, procedimento realizado pela obtenção da matriz de correlações. Por meio dessa matriz, é possível identificar subconjuntos de variáveis que estão muito correlacionadas entre si no interior de cada subconjunto, mas pouco associadas a variáveis de outros subconjuntos. O uso da matriz de correlações, em detrimento da matriz de covariâncias, procura minimizar a influência da magnitude das distintas unidades em que as variáveis foram aferidas. Desta forma, se for preferido utilizar a matriz de covariâncias e sendo as variáveis adimensionais, deve-se realizar a padronização das mesmas, pois este procedimento neutraliza o efeito das distintas unidades, colocando todas elas na mesma escala.

O intervalo estipulado para a adimensionalização ficou entre -2,0 e 2,0 (ou valores próximos, dependendo da necessidade).

As adimensionalizações se basearam nos valores mínimos e máximos das variáveis contidas no domínio geológico considerado para esse estudo (Alterito Laranja). Foram observados também os histogramas com a distribuição dos dados de cada variável.

A adimensionalização das variáveis foi feita através da subtração do valor amostral pela média de sua distribuição e posterior divisão do resultado pelo desvio padrão da mesma distribuição. Abaixo, têm-se as equações (40 a 44) utilizadas para adimensionalizar cada uma das variáveis consideradas:

$$X_1' = \frac{X_1 - 6}{4,5} \quad (41)$$

$$X_2' = \frac{X_2 - 42,5}{17,5} \quad (42)$$

$$X_3' = \frac{X_3 - 6}{4} \quad (43)$$

$$X_4' = \frac{X_4 - 3,25}{2} \quad (44)$$

$$X_5' = \frac{X_5 - 3,5}{2} \quad (45)$$

onde X_1' ; X_2' ; X_3' ; X_4' e X_5' correspondem aos valores adimensionais normalizados das respectivas variáveis originais X_1 ; X_2 ; X_3 ; X_4 e X_5 .

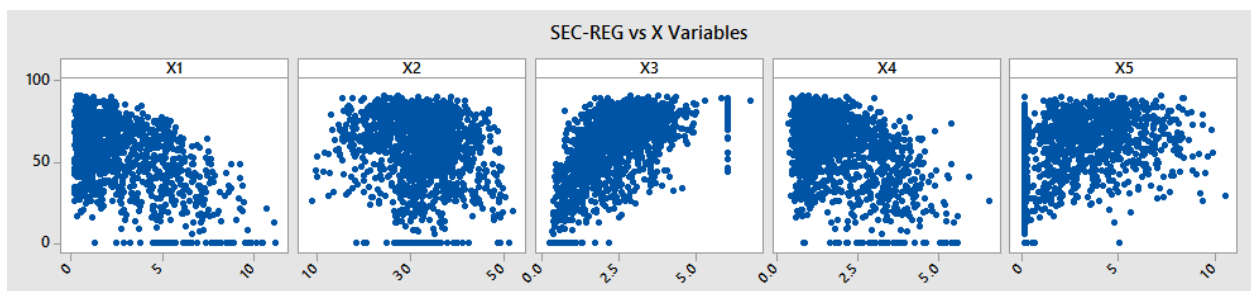
O resultado da primeira regressão multivariada, sem o tratamento dos resíduos, apresentou um coeficiente de correlação igual a 65,27%. Tratando os resíduos de uma forma adequada pode conduzir a resultados melhores, otimizando a resposta da variável dependente sem influenciar na representatividade do domínio geológico considerado.

O tratamento dos resíduos foi feito primeiramente excluindo as amostras que não possuíam o resultado das cinco variáveis químicas consideradas para a regressão, ou seja, criando um banco de dados isotópico. A falta de resultado analítico de alguma das variáveis consideradas na equação tende a gerar resultados incorretos. Posteriormente, excluíram-se as amostras que apresentavam teores anômalos para o domínio geológico considerado (Alterito Laranja). Em uma última etapa, foi realizada uma análise dos resíduos visando melhorar o ajuste (R^2) do modelo proposto aos dados.

Foram realizadas exclusões de dados *outliers* com base em critérios pré-definidos. Dos 1608 (mil seiscentos e oito) dados iniciais, foram excluídos 202 (duzentos e dois), correspondendo a 13% (treze por cento) do total, restando 1406 (mil quatrocentos e seis) dados.

A figura 10 mostra o resultado da regressão sem a análise dos *outliers*.

Figura 10- Scatter plots das cinco variáveis consideradas na regressão multivariada analisadas individualmente contra a resposta (SEC-REG). No eixo Y, tem-se os valores calculados da recuperação metalúrgica (SEC-REG) e no eixo X estão os valores das variáveis independentes utilizadas na regressão.



Fonte: Autoria própria.

Após a exclusão dos dados considerados espúrios, o coeficiente de regressão linear entre a recuperação metalúrgica calculada e a recuperação metalúrgica analisada subiu para 82,59%.

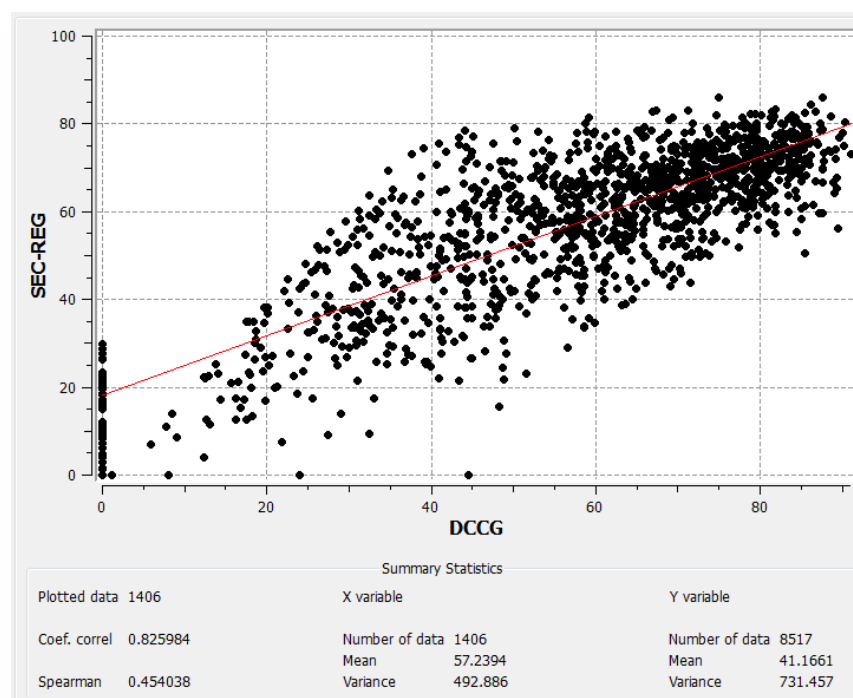
A correlação de 82,59% entre as variáveis pode ser considerada satisfatória, sendo possível utilizar o método de simulação conjunta entre a variável analisada (DCCG) com a variável supersecundária calculada (SEC-REG) para otimizar a estimativa da recuperação metalúrgica de nióbio.

A equação 46 foi obtida através da regressão múltipla de cinco variáveis pelo método de superfície de resposta de 2° ordem. Os cinco óxidos utilizados foram aqueles que apresentaram maior relevância para o cálculo da recuperação metalúrgica. Para fazer essa regressão foi utilizado o software MINTAB®. Essa equação será usada para fazer a previsão da recuperação metalúrgica de nióbio a partir dessas cinco variáveis químicas.

$$\text{SEC-REG} = -33,1 - 15,1X_1' - 0,09X_2' + 39,08X_3' + 41,91X_4' + 13,53X_5' - 2,51X_3'^2 - 5,66X_4'^2 - 0,62X_5'^2 + 2,57X_1'X_3' + 1,49X_1'X_4' - 0,23X_2'X_3' + 0,23X_2'X_4' - 5,16X_3'X_4' - 1,63X_3'X_5' - 2,78X_4'X_5' \quad (46)$$

A figura 11 apresenta o *scatter plot* entre as variáveis: recuperação metalúrgica analisada e a variável supersecundária calculada a partir da regressão multivariada de óxidos.

Figura 11- Scatter plot entre a recuperação metalúrgica analisada em laboratório (DCCG) comparada com o resultado do cálculo da recuperação metalúrgica através da regressão multivariada de cinco óxidos (SEC-REG).



Fonte: Autoria própria.

O *scatter plot* da figura 11 mostra que existem problemas de predição da variável SEC-REC quando os testes de bancada apresentam valor de recuperação metalúrgica igual a zero. Isso se deve principalmente por problemas relacionados à cinética de flotação, os quais não podem ser explicados pelos óxidos considerados na análise. Nesses casos, onde as recuperações metalúrgicas analisadas nos testes de bancada são iguais a zero, tem-se a presença de contaminantes minerais (argilas), de difícil quantificação pela associação dos óxidos considerados. Por isso, ao fazer a regressão múltipla dos óxidos, o resultado da recuperação metalúrgica “calculada” apresenta um viés quando comparado com a sua correspondente analisado como zero.

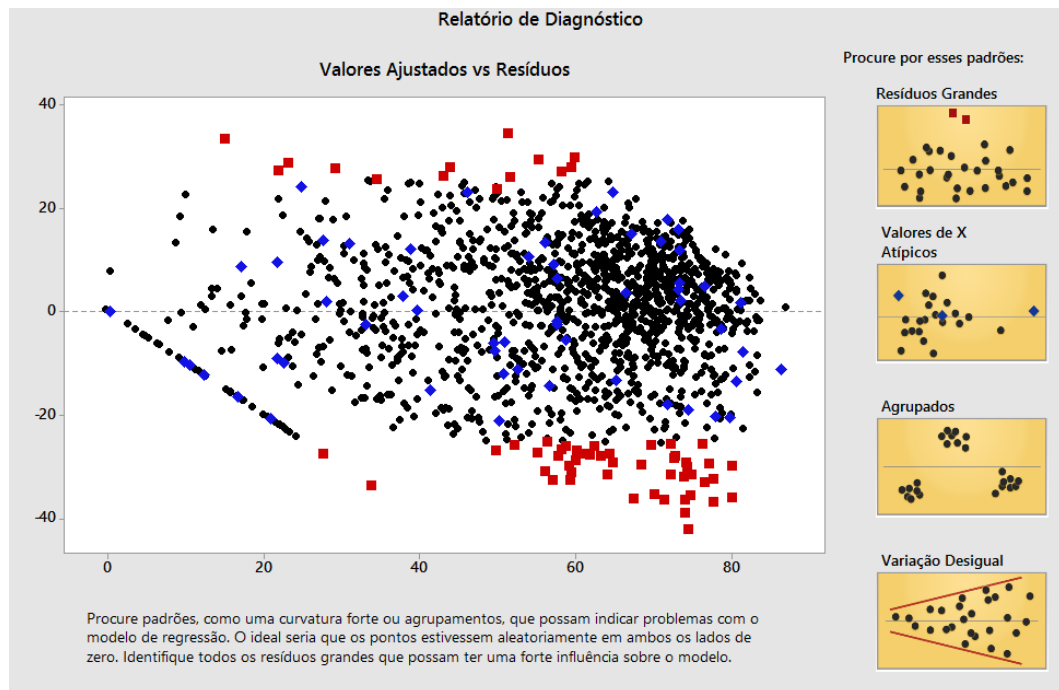
O contrário do exposto no parágrafo anterior também ocorre, porém com uma frequência bem menor. Resultados iguais a zero na regressão múltipla dos óxidos ocorrem em situações onde o teste de bancada foi diferente de zero. Essa situação ocorre nos casos onde a soma dos pesos negativos da equação foram muito altos, o que significa que, em alguns casos, mesmo os elementos deletérios indicando que a probabilidade de o Pirocloro ser recuperado em determinados minérios é muito baixa, acontece de ele ser recuperado no teste de bancada. O fator mineralógico pode ser o motivo pelo qual isso ocorre. Por exemplo: embora o Al_2O_3 esteja normalmente associado aos argilo-minerais, os quais são contaminantes, em alguns casos pode aparecer na estrutura de outros minerais, tais como a monazita e a gorceixita, que não necessariamente são contaminantes.

Essas considerações com problemas de predição do modelo de regressão são nas faixas extremas e deixam uma janela de oportunidade para a otimização desse modelo. O qual certamente seria mais acurado se levasse em consideração fatores mineralógicos e texturais como componentes de entrada.

Após a exclusão dos principais resíduos, a função de ajuste para regressão múltipla foi rodada novamente. Mesmo ainda apresentando 61 pontos de dados com resíduos grandes, optou-se por não excluir nenhum outro dado para não criar uma falsa impressão de alta correlação.

A figura 12 mostra os principais pontos de dados com resíduos grandes (pontos em vermelho no gráfico), bem como os valores atípicos da variável resposta DCCG (pontos em azul no gráfico). É possível também observar, no gráfico, um alinhamento no quadrante sudoeste, esse alinhamento diz respeito aos valores iguais a zero da variável resposta DCCG.

Figura 12- Gráfico para análise de resíduos da regressão múltipla.



Fonte: Autoria própria.

A equação 47 é a equação de regressão para o cálculo da variável supersecundária, após o tratamento dos resíduos, escrita na forma de equação matricial:

$$RM = -33,1 + \tilde{x}'\tilde{b} + \tilde{x}'B\tilde{x} \quad (47)$$

$$\tilde{x}': \begin{bmatrix} X_1' \\ X_2' \\ X_3' \\ X_4' \\ X_5' \end{bmatrix} \quad \tilde{b}_0: \begin{bmatrix} -33,1 \\ -33,1 \\ -33,1 \\ -33,1 \\ -33,1 \end{bmatrix} \quad \tilde{b}: \begin{bmatrix} -15,1 \\ -0,09 \\ 39,08 \\ 41,91 \\ 13,53 \end{bmatrix} \quad B: \begin{bmatrix} 0 & 0 & 1,28 & 0,75 & 0 \\ 0 & 0 & -0,11 & 0,11 & 0 \\ 1,28 & -0,11 & -2,51 & -2,58 & -0,82 \\ 0,75 & 0,11 & -2,58 & -5,66 & -1,39 \\ 0 & 0 & -0,82 & -1,39 & -0,62 \end{bmatrix}$$

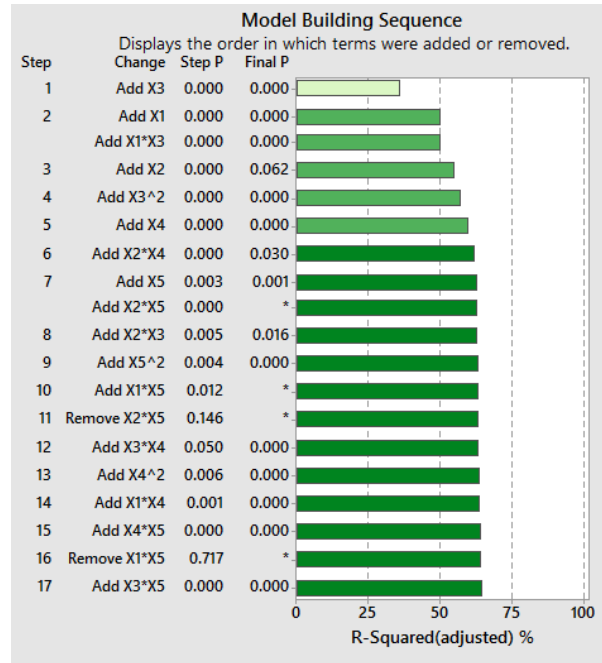
onde a b_0 corresponde à matriz do termo independente, a matriz b é a matriz dos coeficientes lineares de regressão de cada variável com a resposta (DCCG) e a matriz B é a matriz dos coeficientes de interação entre as variáveis com a resposta (DCCG).

Essa equação matricial pode ser escrita na forma algébrica, sendo mais adequada para sua posterior utilização para o cálculo da variável supersecundária. A equação 45 é a equação de regressão da variável supersecundária na forma algébrica.

A figura 13 apresenta o resultado da regressão múltipla obtida com o software MINITAB[®]. Nele é possível observar a sequência da construção do modelo de regressão com a contribuição de cada variável, bem como a contribuição dos pares de variáveis em ordem

crescente. A variável que mais contribui para a explicação do modelo é a X_3 , seguida da variável X_1 e assim sucessivamente.

Figura 13- Resultado da regressão obtido com o software MINITAB.



Fonte: Autoria própria

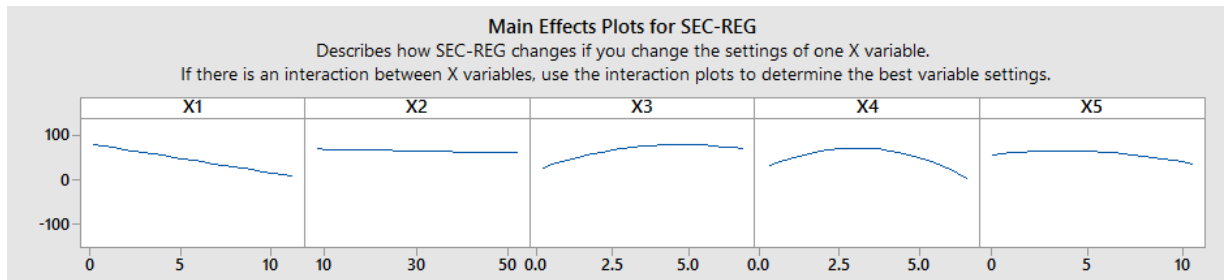
Uma das vantagens da utilização do método de regressão múltipla de segunda ordem é a contribuição da interação entre as variáveis para a explicação do modelo de regressão. Sem essa contribuição, o coeficiente de correlação linear (R^2) entre a variável supersecundária e a variável primária (DCCG) seria bem menor.

Entre as interações dos pares de variáveis, a interação entre o X_1 e X_3 é uma das que se destaca para explicação do modelo, além dela a interação entre o X_2 e X_4 também apresentam coeficientes de covariância altos.

A contribuição quadrática de cada variável também ajuda a definir o modelo de regressão. A contribuição quadrática de X_4 é a mais expressiva, seguida da contribuição quadrática de X_3 . A contribuição quadrática das demais variáveis é muito baixa.

A figura 14 apresenta os gráficos mostrando os efeitos de cada variável independente sobre a resposta (DCCG). Ele descreve como a variável SEC-REC se comporta com a mudança de cada variável. Por exemplo, a variável X_1 possui uma correlação próximo da linear inversa.

Figura 14- Gráficos de efeitos entre as variáveis dependentes e a resposta.



Fonte: Autoria própria

Certos tipos de problemas científicos envolvem a expressão de uma variável resposta, tal como a recuperação metalúrgica de nióbio, podendo ser obtida a partir de uma função empírica de um ou mais fatores quantitativos, como por exemplo os teores dos óxidos obtidos pela análise química via fluorescência de raio x. Isso pode ser efetuado utilizando-se uma metodologia que permita modelar a relação: recuperação metalúrgica em função dos teores de óxidos dos diversos elementos analisados. O conhecimento da forma funcional, frequentemente obtida com a modelagem de dados provenientes de experimentos planejados, permite tanto sumarizar os resultados do experimento quanto prever a resposta para níveis dos fatores quantitativos. Assim, a função f define a superfície de resposta, que, em sua essência, consiste em estimar coeficientes da regressão polinomial para a geração de um modelo empírico, por meio do qual é possível aproximar uma relação (inicialmente desconhecida ou conhecida) entre os fatores e as respostas do processo.

As interações entre as variáveis químicas consideradas, bem como a interação quadrática de cada uma dessas variáveis, podem contribuir demasiadamente para explicar a recuperação metalúrgica do minério de nióbio. A metodologia de superfície de resposta, ou MSR, é um conjunto de técnicas matemáticas e estatísticas que são úteis para modelagem e análise nas aplicações em que a resposta de interesse seja influenciada por várias variáveis e o objetivo seja otimizar o erro quadrático dessa resposta (MONTGOMERY, 2008).

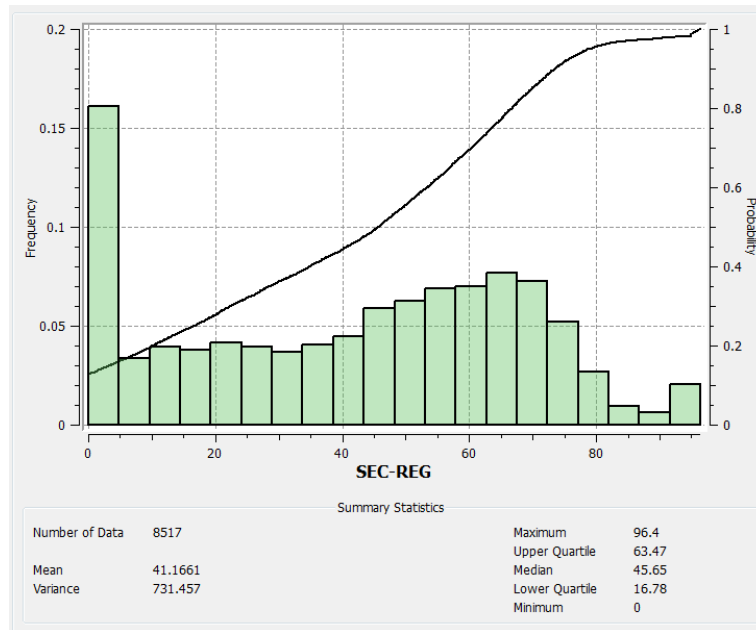
De posse da equação de regressão multivariada para o cálculo da recuperação metalúrgica, que será a nossa variável supersecundária, calculou-se a recuperação metalúrgica para todas as amostras que possuíam as variáveis independentes necessárias para o cálculo.

O cálculo da variável supersecundária ficou restrito às amostras contidas dentro do domínio geológico Alterito Laranja. As amostras contendo as análises químicas dos óxidos requeridos na fórmula de regressão foram obtidas por furos de sondagem rotativa diamantada,

não sendo possível submetê-las aos ensaios geometalúrgicos devido à contaminação das mesmas por polímeros e betonita utilizados nos fluidos de perfuração.

Ao todo, foram calculados 8.517 dados supersecundários, os quais possuem sua distribuição de probabilidade conforme ilustrado pelo histograma da figura 15.

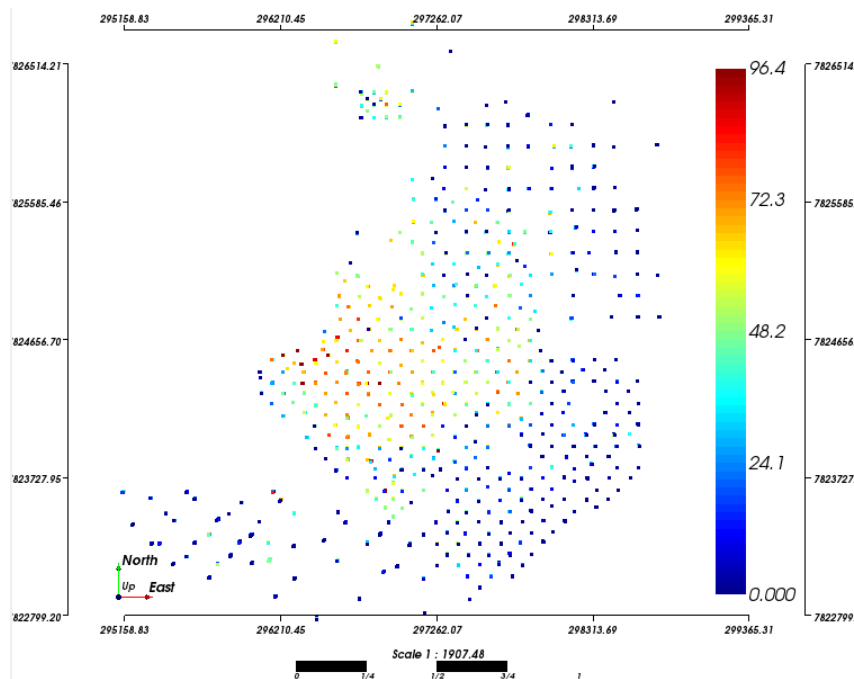
Figura 15- Histograma dos valores da variável supersecundária.



Fonte: Autoria própria.

A figura 16 mostra o mapa com a distribuição em planta da boca dos furos de sondagem de onde foram consideradas as amostras para o cálculo da variável supersecundária. A escala de cores está configurada para a própria variável supersecundária, sendo possível reconhecer a região de maior recuperação metalúrgica na posição central do mapa (cores quentes).

Figura 16- Mapa com a distribuição da boca das sondagens com os dados supersecundário s.



Fonte: Autorial própria.

A simulação conjunta da variável supersecundária com a primária (DCCG) visa melhorar a predição da resposta geometalúrgica do minério de nióbio quando submetido às etapas de concentração mineral.

3.3 SIMULAÇÃO CONJUNTA COLOCADA

Trabalhando com dados não aditivos, como a recuperação metalúrgica de nióbio, é extremamente recomendável a utilização de simulação geoestatística para obter um modelo com a previsibilidade da variável nos locais não amostrados. A variável supersecundária utilizada nesse trabalho foi obtida a partir de dados aditivos (regressão múltipla de óxidos), e, neste caso, é possível considerar o resultado da regressão também aditiva, sendo teoricamente compatível a utilização de krigagem para interpolá-la. Sendo assim, pode-se optar por realizar krigagem ordinária ou simulação estocástica para predizer os fatores explicativos que permitiram “colocar” a variável secundária nos nós da rede de cossimulação para que esta seja posteriormente utilizada como dado supersecundário colocado.

Diferentes implementações de simulações sequenciais podem ser usadas para diferentes propósitos. Para este trabalho, foi selecionada a Cossimulação Sequencial Gaussiana com a simplificação proposta pelo modelo de Markov I. O fato de não ser preciso estabelecer a correlação cruzada pelo modelamento do correlograma, embora mantendo a capacidade de

produzir resultados realistas, é uma vantagem deste método diante de pontos de dados limitados.

A cossimulação sequencial gaussiana permite a simulação de uma variável gaussiana enquanto contabiliza a informação secundária a que se correlaciona. Devido à natureza desse método de simulação, as variáveis a serem cossimuladas devem possuir uma distribuição Gaussiana ou, caso não a tenham naturalmente, devem ser transformadas através de uma função de anamorfose.

A cossimulação colocada usando o modelo de Markov requer que a informação secundária esteja presente em todos os nós a serem simulados. Para satisfazer esta condição, os dados secundários podem ser estimados ou simulados em um grid com as mesmas dimensões do grid onde a cossimulação será realizada.

O modelo de Markov considera que a dependência da variável secundária sob o dado primário é limitada à variável primária colocada. A covariância cruzada é proporcional à autocovariância da variável primária (REMY et al., 2008), podendo ser descrita como na equação 47.

$$C_{12}(h) = \frac{C_{12}(0)}{C_{11}} C_{11}(h) \quad (48)$$

onde h é o vetor de distância, C_{12} é a covariância cruzada entre as duas variáveis e C_{11} é a covariância da variável primária. A resolução do algoritmo de cokrigagem com o modelo de Markov requer o conhecimento da correlação entre as variáveis primárias e secundárias, bem como o semivariograma modelado da variável primária.

A implementação da rotina de cossimulação envolveu os seguintes procedimentos:

- (a) Definição da hierarquia entre as variáveis primárias começando da mais importante $Z_I(u)$ para a menos importante $Z_K(u)$. Como nesse trabalho só existe uma variável primária DCCG, não se fez necessário definir níveis de hierarquia para as variáveis primárias.
- (b) Transformação das variáveis em suas respectivas distribuições normais (NSCORE): de $Y_I(u)$ para $Y_K(u)$ para as variáveis primárias e de $B_I(u)$ para $B_L(u)$ para as variáveis secundárias. É importante transformar a variável secundária antes de submetê-la à cossimulação.
- (c) Variografia dos dados primário e secundário após a normalização para o domínio Gaussiano.

- (d) Definição do caminho aleatório da simulação visando simular um nó de cada vez u, u', u'', \dots
- (e) No nó u , determinar a função de distribuição acumulada normal de $Y_l(u)$ dando n dados primários vizinhos do mesmo tipo $y_l = (u_\alpha)$, $\alpha = 1, \dots, n$ e apenas o dado colocado secundário $b_l(u)$, $l = 1, \dots, L$.
- (f) Retro-transformar os dados da cossimulação para o domínio da variável real. Essa transformação reversa é feita com a função de anamorfoses inversa ao da normalização dos dados.
- (g) Pós processamento da cossimulação para obtenção de dados como: média, desvio padrão, intervalo de confiança e outros.
- (h) Validação da cossimulação.

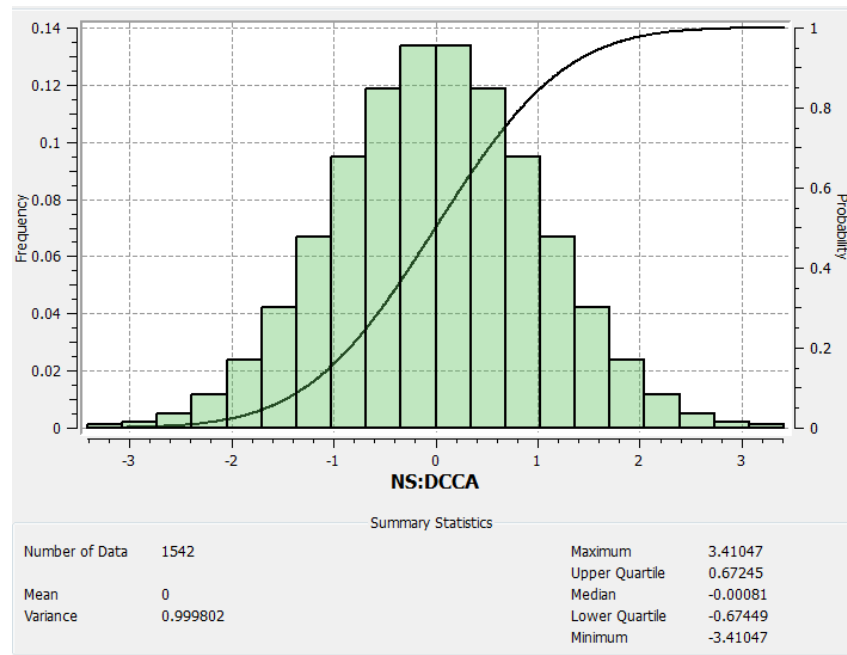
3.3.1 Variografia da variável primária

Para utilização da variável primária (DCCG) em simulação conjunta colocada, considerando o modelo Markov I, é preciso que os dados satisfaçam a condição de distribuição normal. A transformada Gaussiana dos dados foi o método escolhido para fazer a normalização do dado primário. Para tanto, o software GSLIB com o procedimento NSCORE foi utilizado.

Os dados transformados da variável primária foram então importados para o SGEMS, onde serão posteriormente utilizados para a simulação conjunta colocada.

A figura 17 apresenta o histograma do dado primário transformado. Observa-se que a variância dos dados dessa distribuição é igual a 0,9998 e a média igual a 0, satisfazendo dessa forma a condição de distribuição normal.

Figura 17- Histograma dos dados primário transformado.



Fonte: Aatoria própria.

De posse do dado primário transformado, já importado para o Sgems, foi realizada a variografia do mesmo.

O primeiro variograma foi realizado com o intuito de avaliar o efeito pepita, além de avaliar também a variabilidade dos dados ao longo dos furos de sondagem. O variograma com *azimute* 0° e *dip* 90°, também conhecido como “*down the hole*”, foi modelado com duas estruturas esféricas.

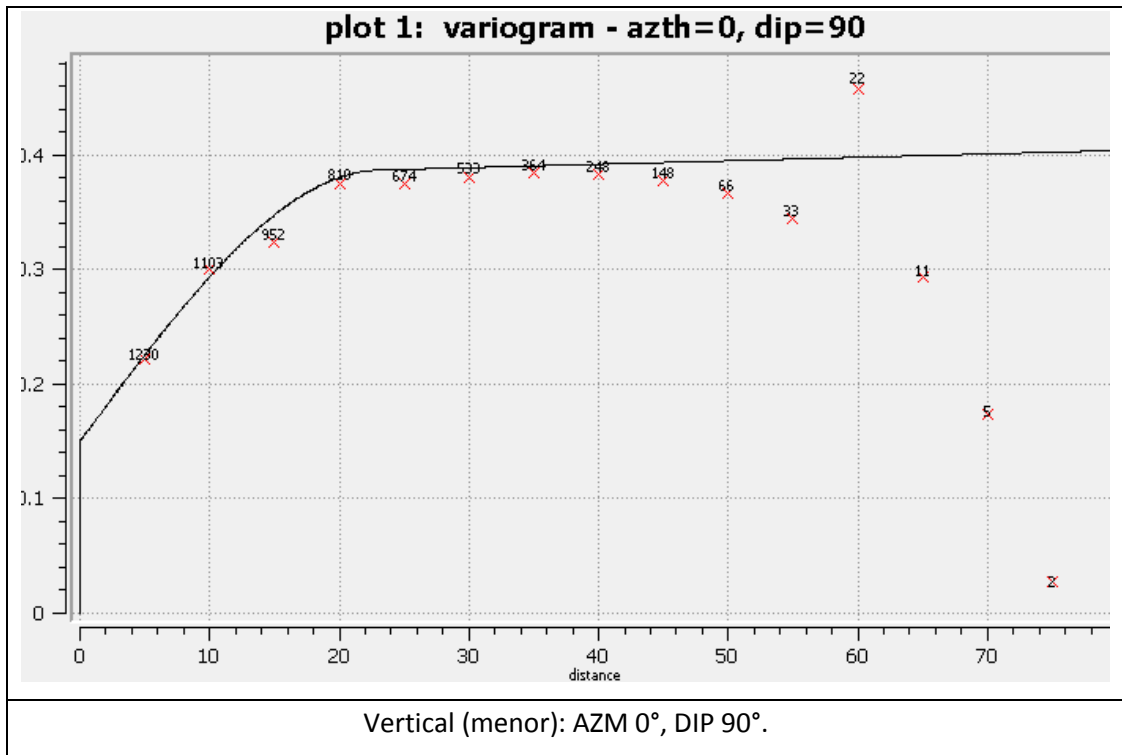
Os parâmetros utilizados para confecção do variograma experimental vertical foram: 20 *lags*, com 5 metros de separação entre eles e tolerância linear de 2,5 metros. Além disso, a tolerância angular considerada foi de 10 graus e o *Bandwidth* igual a 10 metros.

A curva modelada desse variograma intercepta o eixo y do gráfico no ponto 0,15, sendo esse o valor considerado para o efeito pepita.

O patamar considerado para a primeira estrutura foi igual à 0,23 e o *range* dessa estrutura igual a 23 m. A segunda estrutura foi definida com patamar igual à 0,62. Devido a anisotropia zonal observada nessa direção, o *range* da terceira estrutura foi definido como um valor muito alto (3.000 m) para se modelar o variograma.

A figura 18 mostra o variograma vertical modelado.

Figura 18- Variograma na direção vertical



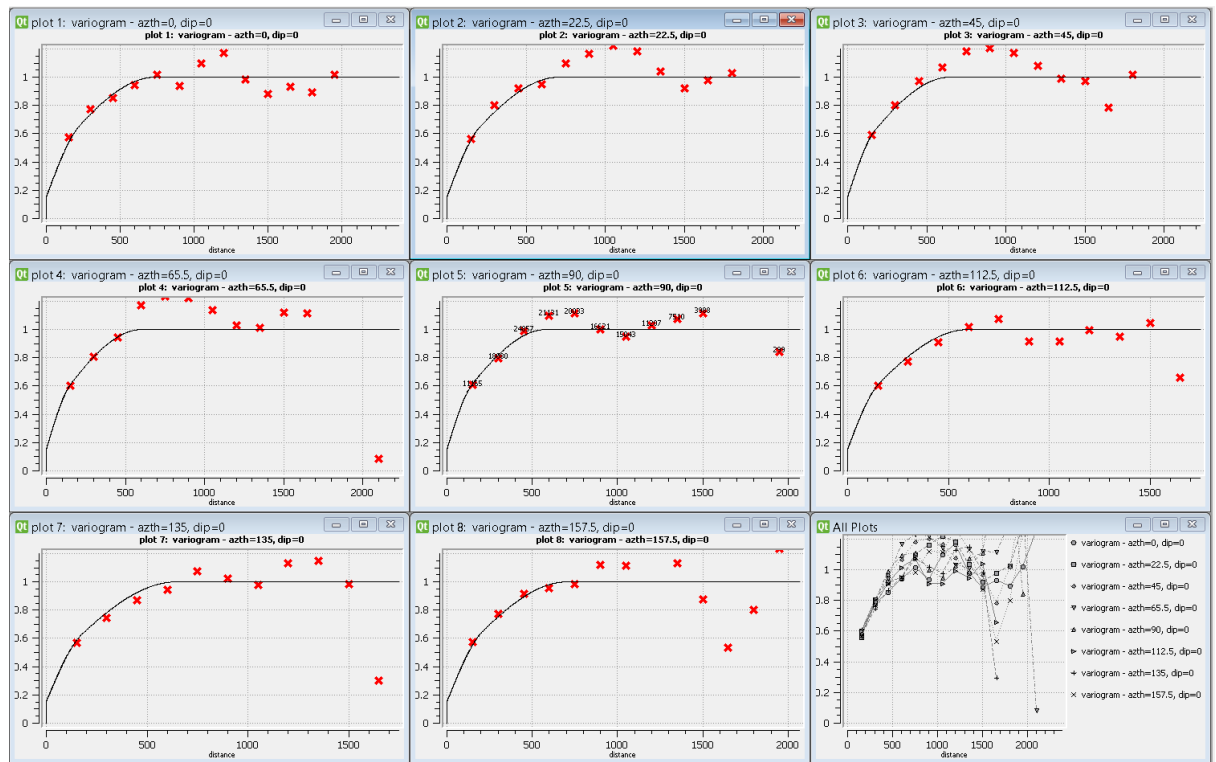
Fonte: Autoria própria.

O resumo dos parâmetros do variograma vertical modelado estão descritos na equação (49).

$$\gamma(h) = 0,15 + \left[0,23 \cdot Sph_1 \left[\frac{N0, D90}{23m} \right] + 0,62 \cdot Sph_2 \left[\frac{N0, D90}{3000m} \right] \right] \quad (49)$$

Após o modelamento do variograma vertical, foram modelados os variogramas horizontais. Para construção dos variogramas experimentais, foram utilizados os seguintes parâmetros: número de *lags* igual a 20, separação entre os *lags* igual a 150 m, com tolerância de 75 m. Além disso, a tolerância angular considerada foi de 22,5° e o *Bandwith* igual a 150 m. Nessa etapa, foram modelados oito variogramas com as seguintes direções: (N0, D0), (N22,5, D0), (N45, D0), (N67,5, D0), (N90, D0), (N112,5, D0), (N135, D0) e (N157,5, D0). A figura 19 mostra todos os oito variogramas considerados na direção horizontal.

Figura 19- Variogramas horizontais modelados.



Fonte: Autoria própria.

Dentre as oito direções horizontais variografadas, a que apresentou maior continuidade espacial foi a direção (N0, D0). Por consequência, a direção (N90, D0) apresentou baixa continuidade espacial.

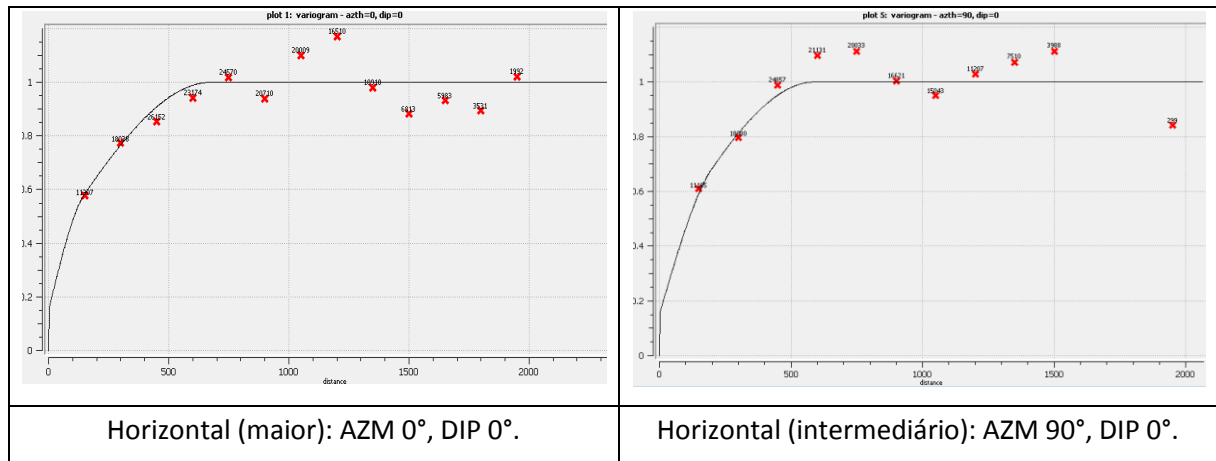
Para modelar as estruturas nos variogramas horizontais foi considerado o efeito pepita igual a 0,15, sendo esse valor obtido a partir do modelamento do variograma vertical.

O variograma na direção (N0, D0) foi então modelado com duas estruturas esféricas. A primeira estrutura apresentando patamar igual a 0,23 e *range* igual a 250 m e a segunda estrutura com patamar igual a 0,62 e *range* igual a 735 m.

O variograma na direção (N90, D0) também foi modelado, os parâmetros para o modelamento desse variograma foram os mesmo do variograma na direção (N0, D0), sendo que os ranges da primeira e segunda estrutura sendo iguais a respectivamente: 220 m e 580 m.

A figura 20 apresenta os variogramas nas direções (N0, D0) e (N90, D0) modelados.

Figura 20- Variogramas modelados para a direção de maior e menor continuidade na horizontal.



Fonte: Autoria própria.

O resumo dos parâmetros dos variogramas horizontais modelado nas direções (N0, D0) e (N90, D0) estão descritos na equação 50.

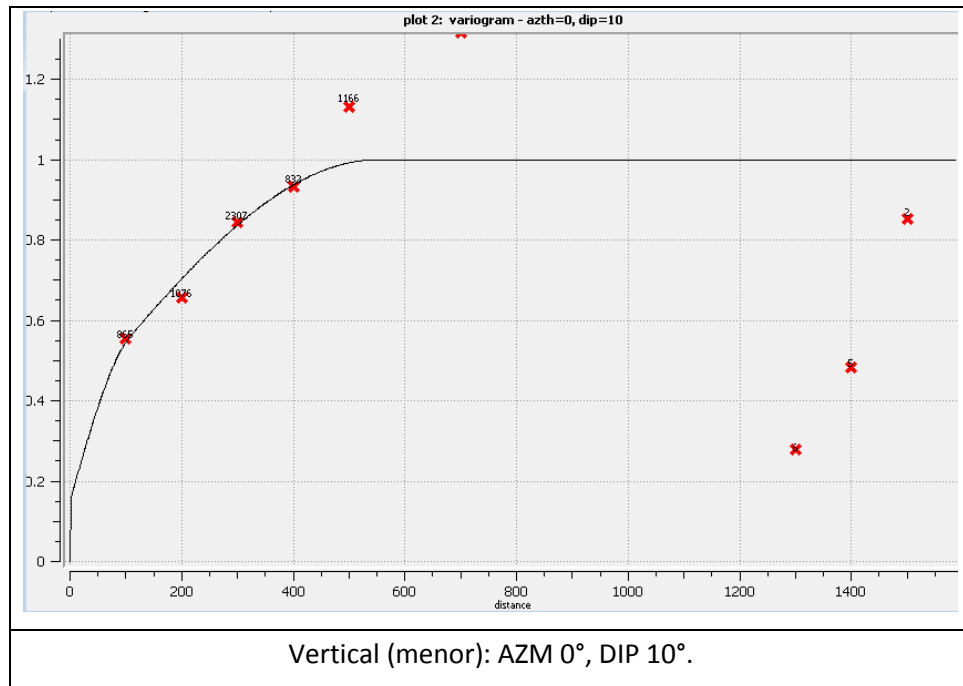
$$\gamma(h) = 0,15 + \left[0,23 \cdot Sph_1 \left[\frac{N0,D0}{250m}, \frac{N90,D0}{220m} \right] + 0,62 \cdot Sph_2 \left[\frac{N0,D0}{735m}, \frac{N90,D0}{580m} \right] \right] \quad (50)$$

Após a finalização do modelamento dos variogramas horizontais, foi avaliada a continuidade espacial dos dados na direção N0 variando os *Dips*. Para isso foram consideradas as direções: (N0, D0), (N0, D10), (N0, D20), (N0, D30), (N0, D40), (N0, D50), (N0, D60), (N0, D70) e (N0, D80). Além do variograma na direção (N0, D0) que já havia sido modelado anteriormente, o único outro variograma que pôde de ser modelado nessa variação foi o de direção (N0, D10). Os demais variogramas ficaram muito erráticos ou, como no caso dos de direção (N0, D50), (N0, D60), (N0, D70) e (N0, D80) não foi encontrado nenhum par de amostras com os parâmetros utilizados.

O variograma na direção (N0, D10) foi então modelado considerando o efeito pepita igual a 0,15. Possuindo duas estruturas esféricas, sendo a primeira modelada com um patamar de 0,23 e *range* igual a 150 m e a segunda estrutura com patamar igual a 0,62 e *range* igual a 550 m.

A figura 21 apresenta o variograma na direção (N0, D10) modelado.

Figura 21- Variograma na direção (N0, D10) modelado.



Fonte: Autoria própria.

O resumo dos parâmetros do variograma na direção (N0, D10) estão descritos na equação 50.

$$\gamma(h) = 0,15 + \left[0,23 \cdot Sph_1 \left[\frac{N0,D10}{150m} \right] + 0,62 \cdot Sph_2 \left[\frac{N0,D10}{550m} \right] \right] \quad (51)$$

Como a direção de maior continuidade está na horizontal (N0, D0), a direção com continuidade intermediária fica sendo a perpendicular a ela, sendo nesse caso a direção (N90, D0). A direção de menor continuidade passa a ser a direção vertical (N0, D90). Dessa forma, têm-se os três eixos do elipsoide variográfico para a variável primária transformada.

A equação 52 representa o elipsoide variográfico para a variável primária transformada (NSCORE).

$$\gamma(h) = 0,15 + \left[0,23 \cdot Sph_1 \left[\frac{N0,D0}{250m}, \frac{N90,D0}{220m}, \frac{N0,D90}{23m} \right] + 0,62 \cdot Sph_2 \left[\frac{N0,D0}{735m}, \frac{N90,D0}{580m}, \frac{N0,D90}{3000m} \right] \right] \quad (52)$$

Os parâmetros desse variograma modelado foi introduzido no software Isatis (GEOVARIANCES, 1994) para a realização da simulação conjunta da variável DCCG com a variável SEC-REG.

3.3.2 Variografia da variável supersecundária

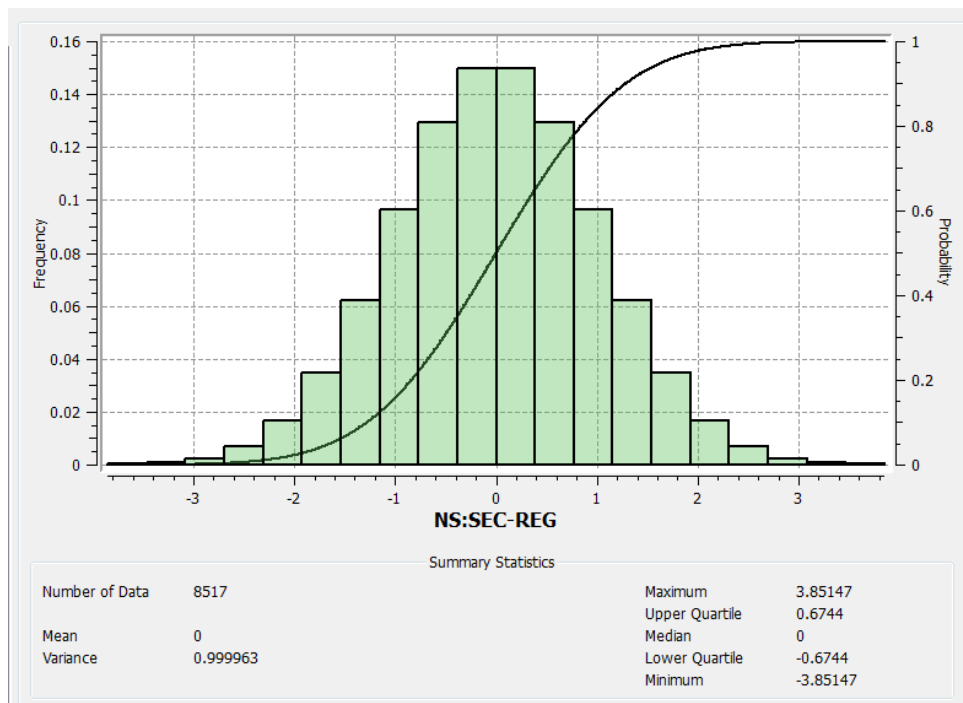
Para utilização da variável supersecundária (SEC-REC) na simulação conjunta colocada, considerando o modelo Markov I, é preciso que os dados satisfaçam a hipótese de

multiGaussianidade. A transformada Gaussiana dos dados foi o método escolhido para fazer a normalização do dado supersecundário. Para tanto, o software GSLIB com o procedimento NSCORE foi utilizado.

Os dados transformados da variável supersecundária foram então importados para o software SGEMS.

A figura 22 apresenta o histograma do dado supersecundário transformado. Observa-se que a variância dos dados dessa distribuição é igual a 0,9999 e a média igual a 0, satisfazendo dessa forma a condição de distribuição normal.

Figura 22- Histograma da variável supersecundária transformada (NSCORE).



Fonte: Autoria própria.

De posse do dado supersecundário transformado, já importado para o Sgems, foi realizada a variografia do mesmo.

O primeiro variograma foi realizado com o intuito de avaliar o efeito pepita, além de avaliar também a variabilidade dos dados ao longo dos furos de sondagem.

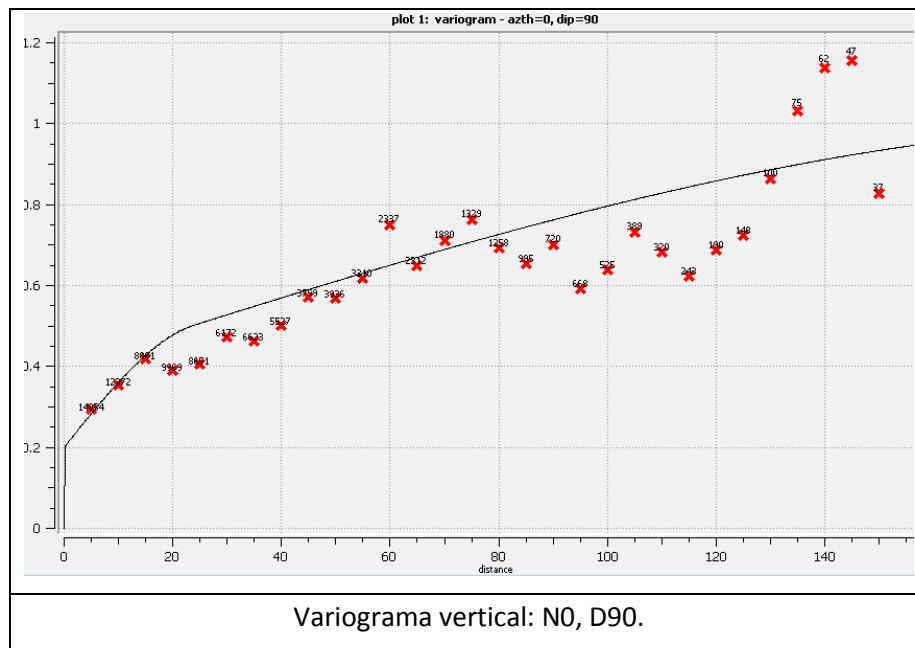
Os parâmetros utilizados para confecção do variograma experimental foram: 30 *lags*, com 5 m de separação entre eles e tolerância de 2,5 m. Além disso, a tolerância angular considerada foi de 10° e 10 m de *Bandwidth*.

O prolongamento da curva modelada desse variograma intercepta o eixo y do gráfico no ponto 0,2, sendo esse o valor considerado para o efeito pepita.

O patamar considerado para a primeira estrutura foi igual à 0,2 e o *range* dessa estrutura igual a 25 m. A segunda estrutura foi definida com patamar igual à 0,6 e o *range* igual a 210 m.

A figura 23 mostra o variograma vertical modelado.

Figura 23- Variograma na direção vertical (Down the hole)



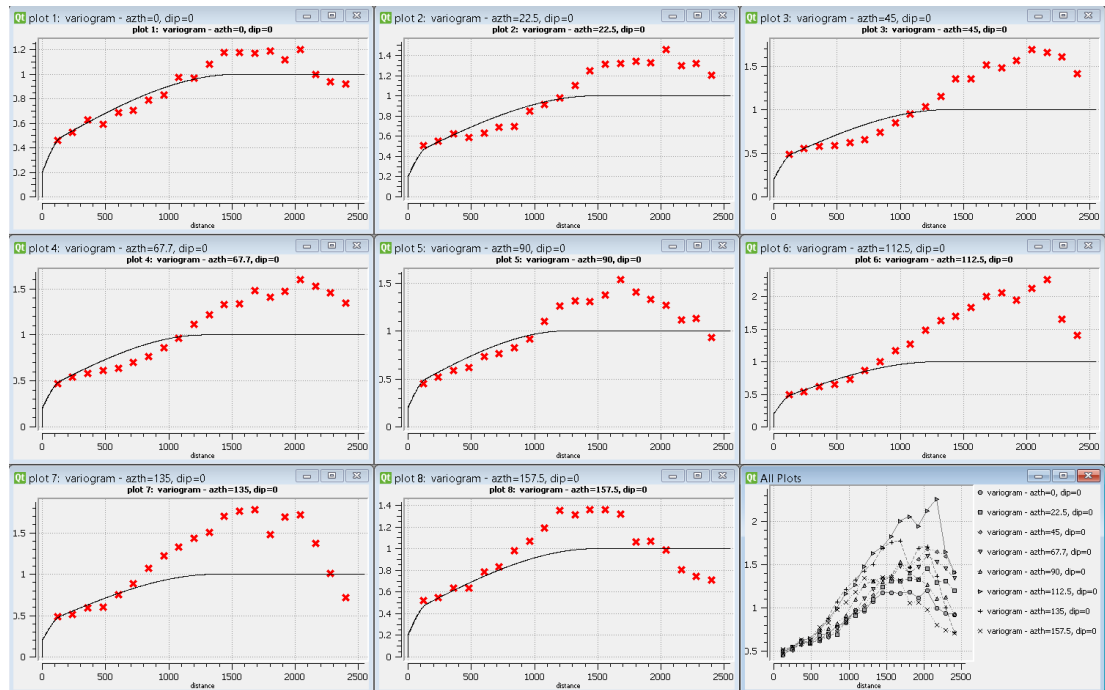
Fonte: Autoria própria.

O resumo dos parâmetros do variograma vertical modelado estão descritos na equação 53.

$$\gamma(h) = 0,20 + \left[0,20.Sph_1 \left[\frac{N0,D90}{25m} \right] + 0,60.Sph_2 \left[\frac{N0,D90}{210m} \right] \right] \quad (53)$$

Após o modelamento do variograma vertical, foram modelados os variogramas horizontais. Para construção dos variogramas experimentais, foram utilizados os seguintes parâmetros: número de *lags* igual a 20, separação entre os *lags* igual a 150 m, com tolerância de 75 m. Além disso, a tolerância angular considerada foi de 22,5° e o *Bandwidth* igual a 150 m. Nessa etapa, foram modelados oito variogramas com as seguintes direções: (N0, D0), (N22,5, D0), (N45, D0), (N67,5, D0), (N90, D0), (N112,5, D0), (N135, D0) e (N157,5, D0). A figura 24 mostra todos os oito variogramas considerados na direção horizontal.

Figura 24- Variogramas horizontais modelados.



Fonte: Autoria própria.

Dentre as oito direções horizontais variografadas, a que apresentou maior continuidade espacial foi a de direção (N0, D0). Por consequência, a direção (N90, D0) apresentou baixa continuidade espacial.

Para modelar as estruturas nos variogramas horizontais foi considerado o efeito pepita igual a 0,2, sendo esse valor obtido a partir do modelamento do variograma vertical.

O variograma na direção (N0, D0) foi então modelado com duas estruturas esféricas. A primeira estrutura apresentando patamar igual a 0,2 e *range* igual a 250 m e a segunda estrutura com patamar igual 0,6 e *range* igual a 1530 m.

O variograma na direção (N90, D0) também foi modelado. Os parâmetros para o modelamento desse variograma estão descritos na equação 54, assim como o resumo dos parâmetros do variograma horizontal modelado na direção (N0, D0).

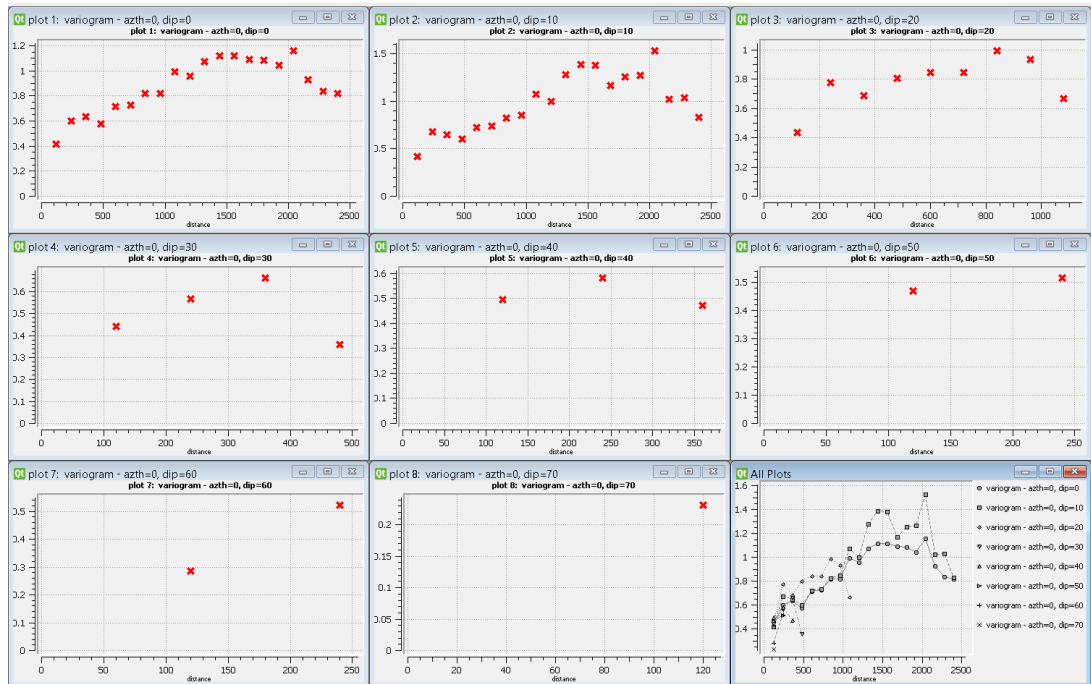
$$\gamma(h) = 0,2 + \left[0,2 \cdot Sph_1 \left[\frac{N0,D0}{250m}, \frac{N90,D0}{220m} \right] + 0,6 \cdot Sph_2 \left[\frac{N0,D0}{1530m}, \frac{N90,D0}{1250m} \right] \right] \quad (54)$$

Após a finalização do modelamento dos variograma horizontais, foi avaliada a continuidade espacial dos dados na direção N0 variando os *Dips*. Para isso foram consideradas as direções: (N0, D0), (N0, D10), (N0, D20), (N0, D30), (N0, D40), (N0, D50), (N0, D60), (N0, D70) e (N0, D80). Os únicos variogramas possíveis de serem modelados com essa

variação do *dip* foram o de direção (N0, D0) e (N0, D10). Os demais variogramas ficaram muito erráticos ou, como no caso dos de direção (N0, D40), (N0, D50), (N0, D60) e (N0, D80) não foi encontrado nenhum par de amostras com os parâmetros utilizados.

A figura 25 apresenta todos os variogramas experimentais considerados na direção N0 variando o *dip* de dez em dez graus.

Figura 25- Variogramas experimentais na direção N0 variando o *dip* a cada 10 graus.



Fonte: Autoria própria.

O variograma na direção (N0, D10) foi então modelado considerando o efeito pepita igual a 0,2, modelado com duas estruturas esféricas: a primeira estrutura esférica com patamar igual a 0,2 e *range* igual a 240 m e a segunda estrutura (gaussiana) com patamar igual a 0,6 e *range* igual 1220 m.

Como a direção de maior continuidade está na horizontal (N0, D0), a direção com continuidade intermediária será a perpendicular a ela, sendo nesse caso a direção (N90, D0). Nesse caso, a direção de menor continuidade é a direção vertical (N0, D90). Dessa forma, têm-se os três eixos do elipsoide variográfico para a variável supersecundária transformada.

A equação 55 representa o elipsoide variográfico para a variável supersecundária transformada (NSCORE).

$$\gamma(h) = 0,2 + \left[0,2 \cdot Sph_1 \left[\frac{N0,D0}{250m}, \frac{N90,D0}{220m}, \frac{N0,D90}{25m} \right] + 0,6 \cdot Sph_2 \left[\frac{N0,D0}{1530m}, \frac{N90,D0}{1250m}, \frac{N0,D90}{210m} \right] \right] \quad (55)$$

Os parâmetros desse variograma modelado foram introduzidos no software Isatis para a realização da simulação conjunta da variável DCCG com a variável SEC-REG.

3.3.3 Cossimulação

Para fazer a simulação conjunta utilizando o modelo de Markov I é preciso que a informação secundária esteja presente em todos os nós a serem simulados. Para satisfazer essa condição, o dado secundário pode ser krigado ou simulado em um grid com as mesmas dimensões do grid onde será realizada a simulação conjunta colocada.

Em dados não aditivos, como a recuperação metalúrgica do minério de nióbio, é extremamente recomendado utilizar a simulação geoestatística para simular e, assim, obter a previsibilidade da variável nos locais não amostrados. A variável secundária advém de dados aditivos (óxidos), porém, a regressão utilizada possui componentes de 2ª ordem, conferindo um caráter não aditivo à essa variável. Sendo assim, a estimativa dessa variável por simulação é extremamente recomendada, sendo essa a técnica utilizada para criar o dado supersecundário colocado.

O grid construído para fazer a simulação da variável secundária a ser colocada em cada nó a ser cossimulado possui 462 células em X, 468 células em Y e 150 células em Z. As dimensões das células utilizadas na simulação foram de 5 m em X, 5 m em Y e 1 m em Z. Esse grid construído possui um total de 32.432.400 células, sendo esse grid muito extenso para o trabalho pretendido e tendo, nesse caso, uma restrição computacional para viabilizar a cossimulação. Esse problema foi resolvido delimitando o grid de cossimulação ao domínio geológico “Alterito Laranja” previamente modelado.

A simulação para criação da variável colocada a ser simulada conjuntamente com a variável primária foi realizada no Isatis. Para o elipsoide de busca, foi considerado um máximo de 25 dados primários e 25 dados previamente simulados. Foi considerado um raio de busca igual a 250 m x 250 m x 15 m.

O dado de entrada para essa operação foi o dado supersecundário transformado (NSCORE) e a distribuição considerada foi igual a distribuição do dado supersecundário desagrupado.

Para fazer o desagrupamento do dado supersecundário transformado e construir a referência de distribuição, foi utilizado o método dos polígonos de Voronoi. Para isso foi

realizada a estimativa pelo método do vizinho mais próximo com raio de busca igual a 150 m x 150 m x 10 m.

Foi utilizado o variograma definido na secção 3.2 para simular a variável supersecundária transformada.

Foi utilizado o caminho aleatório para realização da cossimulação. A semente inicial da cossimulação foi definida como 717161 com incremento de 2. Já a semente para o caminho foi iniciada com 1117111, também com incremento de 2. O coeficiente linear de correlação entre a variável primária e a supersecundária também foi introduzido como parâmetro para cossimulação, sendo esse igual a 0,826.

3.4 VALIDAÇÃO DA COSSIMULAÇÃO

Aqui serão apresentadas as incertezas referentes à cossimulação da variável recuperação metalúrgica (DCCG), ou seja, temas tais como: flutuações ergódicas, equiprobabilidade das realizações, incertezas nos parâmetros e no modelo.

Em simulação estocástica, as flutuações ergódicas são utilizadas para se referir às variações nas estatísticas das várias realizações em relação à estatística do modelo de variograma e distribuição escolhidos. Os variogramas e as FDAC (Funções de Distribuição Acumulada Condicionais) das realizações apresentam diferenças em relação ao variograma e à CDF (Função de Distribuição Acumulada) que foram utilizados para gerá-las. As flutuações ergódicas identificam o espaço de incertezas. Realizações que honrem exatamente os modelos de cdf e variograma podem ser consideradas como um subconjunto de todas as realizações.

Algumas flutuações ergódicas são aceitáveis desde que o modelo estatístico esteja afetado por flutuações de amostragem. Em algumas aplicações, os modelos estatísticos são inferidos de amostras esparsas, não podendo ser considerados como perfeitamente conhecidos. Portanto, desvios em relação ao modelo estatístico podem ser aceitáveis, pois eles se relacionam com o aspecto da incerteza inerente do banco de dados utilizado para gerá-los.

A partir das formulações apresentadas por Matheron (1989), pode-se verificar que o tamanho de um domínio, suficiente para verificar a ergodicidade, depende do alcance do modelo do semi-variograma. Quanto mais contínuo for o semi-variograma na origem e maior seu alcance, maior será o domínio necessário para se verificar a ergodicidade. A ergodicidade é uma propriedade do modelo probabilístico e não característica dos dados ou do fenômeno de

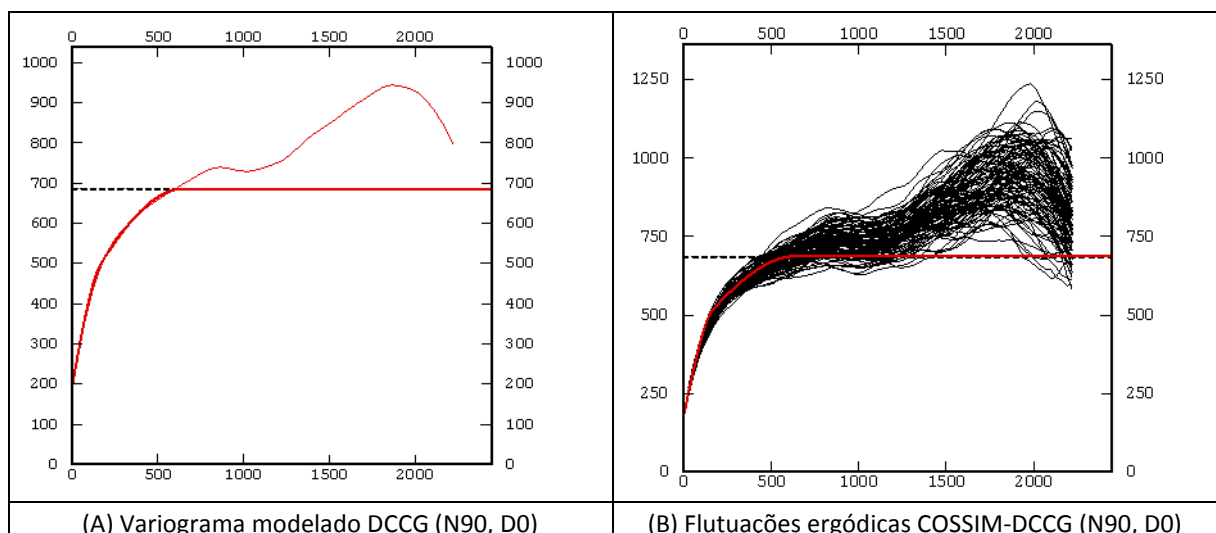
estudo. Através da propriedade da ergodicidade, é possível inferir a distribuição e momentos do modelo de cada uma das realizações.

A ocorrência das flutuações ergódicas, amplamente reportadas na literatura, poucas vezes tem sua importância relacionada à incerteza. A figura 26(A) apresenta o semi-variograma modelado da variável primária (DCCG) na direção (N90, D0). Observa-se que o variograma é muito bem modelado com duas estruturas esféricas e o efeito pepita. Esse modelo de variograma não é o mesmo do utilizado como parâmetro de entrada para rodar a cossimulação, como visto anteriormente, foi preciso modelar-se o variograma após a normalização gaussiana dos dados originais, porém, é fundamental que a reprodução do variograma da variável original corresponda ao resultado da cossimulação após os dados serem retro-transformados.

A figura 26(B) mostra as flutuações ergódicas do resultado da cossimulação (linhas pretas) em conjunto com o variograma modelado da variável primária de entrada (linha vermelha). Nota-se que as flutuações ergódicas acompanham o modelo variográfico da variável primária até atingir o patamar (variância a priori das amostras), mostrando que os variogramas das 100 realizações da cossimulação reproduziram o modelo de variograma esperado.

A direção (N90, D0) foi modelada como sendo a direção de continuidade intermediária.

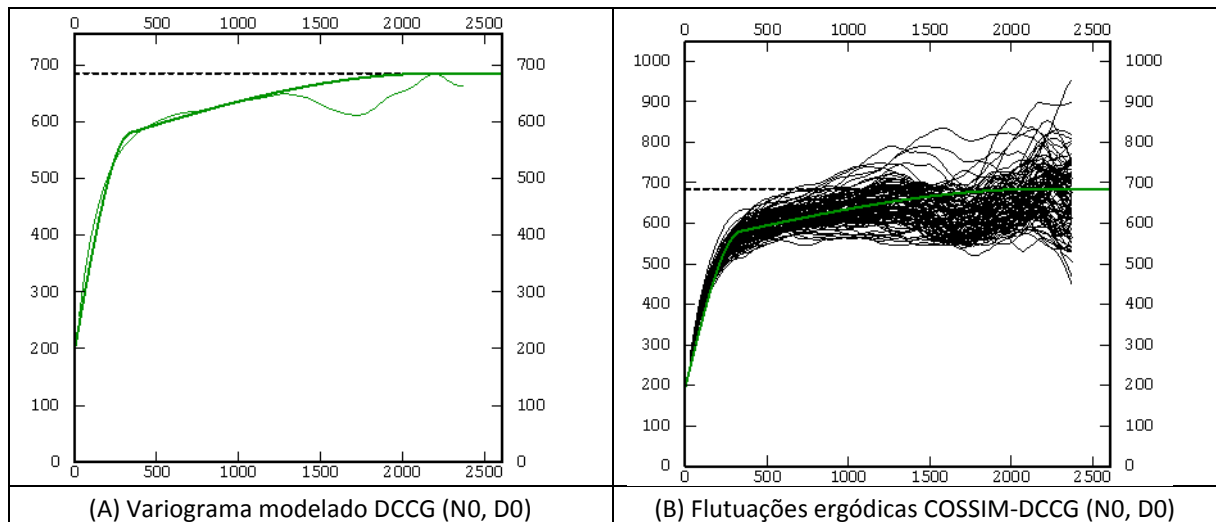
Figura 26- (A) Variograma modelado da variável DCCG na direção (N90, D0) (B) Flutuações ergódicas das 100 cossimulações da variável recuperação metalúrgica de nióbio (DCCG) com a supersecundária colocada (SEC-REC) na direção (N90, D0).



Fonte: Autoria própria.

A direção (N0, D0) foi modelada como sendo a direção de maior continuidade. A figura 27(A) apresenta o semi-variograma modelado da variável primária (DCCG) na direção (N0, D0). Assim como o variograma da direção de continuidade intermediária (N90, 0), o variograma na direção (N0, D0) também é facilmente modelado com duas estruturas esféricas e o efeito pepita. A figura 27(B) mostra as flutuações ergódicas do resultado da cossimulação (linhas pretas) em conjunto com o variograma modelado da variável primária de entrada (linha verde) na direção (N0, D0). As flutuações ergódicas acompanham perfeitamente o modelo variográfico da variável primária até atingir o seu patamar. Isso mostra que a reprodução dos variogramas das 100 cossimulações realizadas reproduzem o modelo de variograma esperado.

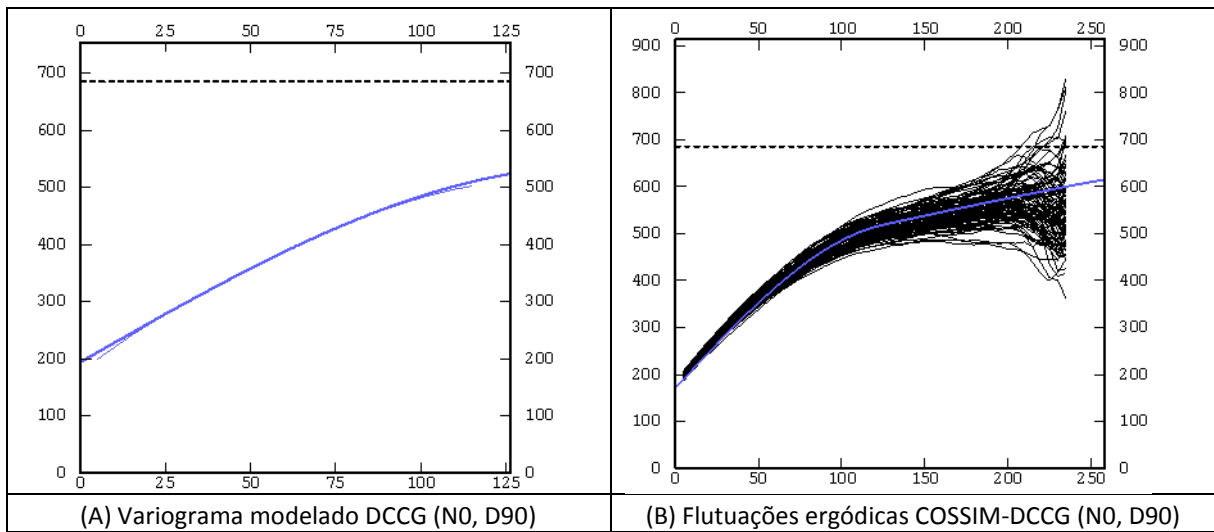
Figura 27- (A) Variograma modelado da variável DCCG na direção (N0, D0) (B) Flutuações ergódicas das 100 cossimulações da variável recuperação metalúrgica de nióbio (DCCG) com a supersecundária colocada (SEC-REC) na direção (N0, D0).



Fonte: Autoria própria.

A direção de menor continuidade da variável primária (DCCG) foi a (N0, D90). A figura 28(A) apresenta o semi-variograma modelado da variável primária (DCCG) na direção (N0, D90). Esse variograma foi modelado com duas estruturas esféricas mais o efeito pepita. A figura 28(B) mostra as flutuações ergódicas do resultado da cossimulação (linhas pretas) em conjunto com o variograma modelado da variável primária de entrada (linha azul) na direção (N0, D90). As flutuações ergódicas acompanham o modelo variográfico da variável primária até atingir o seu patamar nessa direção. Isso mostra que a reprodução dos variogramas das 100 cossimulações realizadas reproduzem o modelo de variograma esperado na direção vertical.

Figura 28- (A) Variograma modelado da variável DCCG na direção (N0, D90) (B) Flutuações ergódicas das 100 cossimulações da variável recuperação metalúrgica de nióbio (DCCG) com a supersecundária colocada (SEC-REC) na direção (N0, D90).



Fonte: Autoria própria.

Um critério que pode ser adotado para a seleção de imagens baseia-se na verificação do semi-variograma da imagem cossimulada, não podendo esse se distanciar muito do considerado representativo. As flutuações ergódicas podem ser utilizadas no sentido de cobrir alguma incerteza que se tem no modelo de semi-variograma escolhido. Pode-se afirmar que: quanto maior a confiabilidade no modelo e nos parâmetros do semi-variograma escolhido, menores deverão ser as flutuações ergódicas aceitas.

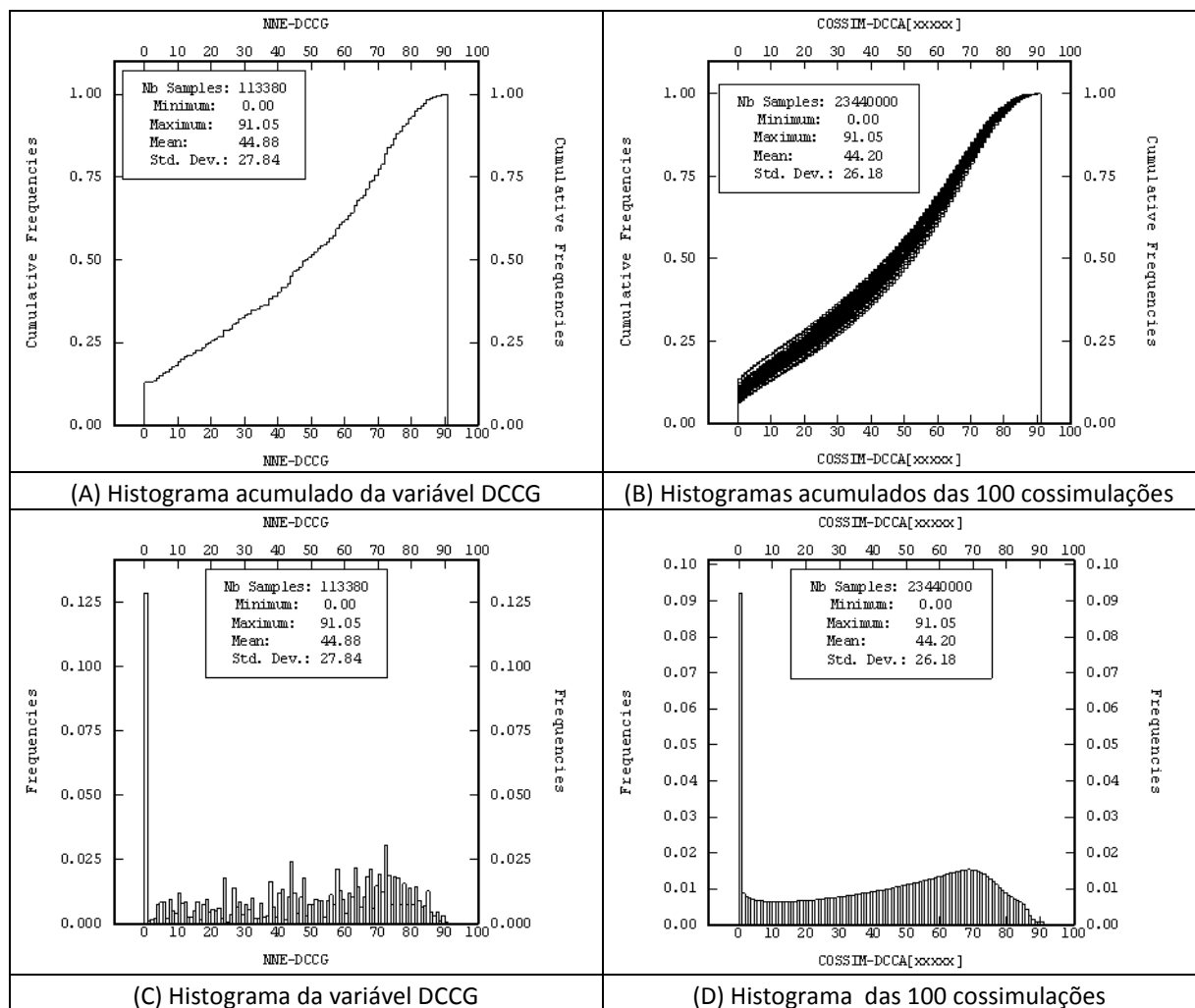
Quando não se observa ergodicidade da média, a ergodicidade do semi-variograma ou de qualquer outro momento de ordem superior também não será observada. Na modelagem da recuperação metalúrgica de nióbio, a ocorrência das flutuações ergódicas tanto da média quanto do semi-variograma é bastante normal, uma vez que o range dos semi-variogramas modelados são muitas vezes da mesma ordem de grandeza do tamanho do domínio cossimulado.

Outra forma muito comum de validar modelos cossimulados é através da capacidade de reprodução do histograma da variável primária (DCCG) pela cossimulação. Para fazer essa comparação é importante que a variável primária esteja desagrupada, caso contrário, a tendência amostral irá gerar uma estatística descritiva do dado primário que não representa a realizada do depósito analisado. O desagrupamento da variável primária foi realizado utilizando o conceito dos polígonos de Voronoi. Esse é um método de desagrupamento em que os pesos atribuídos às amostras são diretamente proporcionais à área do Polígono de Voronoi ao seu redor. Em zonas de dados agrupados, as áreas dos polígonos tendem a ser pequenas, recebendo, então, pesos menores. Para realizar esse desagrupamento, foi utilizado o artifício da estimativa pelo

método do vizinho mais próximo da variável primária (DCCG) em um grid de células pequenas, sendo utilizado o mesmo grid da cossimulação.

A figura 29(A) apresenta o histograma acumulado da variável primária (DCCG) após ser desagrupada. A figura 29(A) mostra o histograma acumulado das 100 cossimulações. A figura 29(C) mostra o histograma de probabilidade da variável primária (DCCG) após o desagrupamento e a figura 29(D) apresenta o histograma de probabilidade das 100 cossimulações realizadas. A figura 4 da (pág. 41) apresentou o histograma do resultado da interpolação da variável DCCG pelo método do vizinho mais próximo no grid da cossimulação.

Figura 29- Histograma da média das 100 cossimulações e histograma da cokrigagem da variável geometalúrgica.



Fonte: Autoria própria.

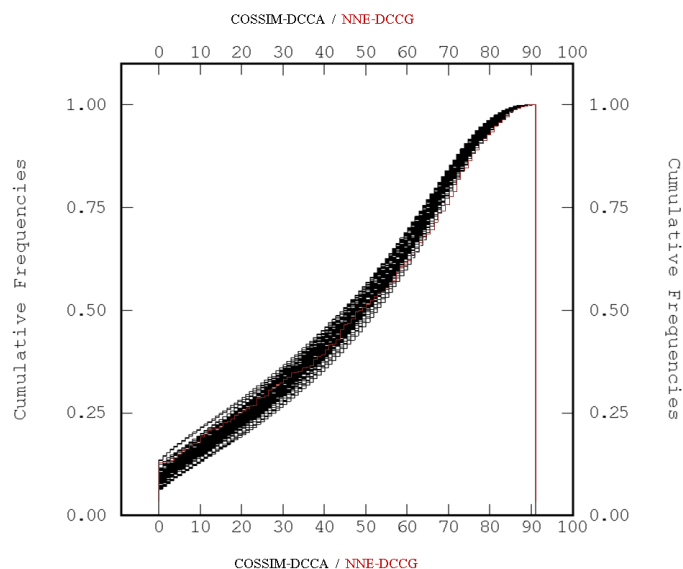
Observa-se que os histogramas da variável primária, tanto o cdf quanto o pdf, são muito bem reproduzidos pela nuvem de histogramas da cossimulação. A média dos dados originais

desagrupados é igual a 44,88% com desvio padrão de 27,84, enquanto a média dos resultados das 100 cossimulações é igual a 44,20%, com desvio padrão absoluto de 26,18.

Pela análise dos histogramas e variogramas pode-se dizer que a cossimulação realizada gerou cenários com valores reproduzindo adequadamente a covariância e a distribuição de frequências do dado primário original (DCCG). Além disso, por ter sido realizada uma cossimulação condicional, os dados também honram os valores nos locais onde a variável DCCG é conhecida.

O método apresentou bons resultados, tendo em vista que os histogramas e variogramas foram reproduzidos. A figura 30 apresenta a sobreposição dos histogramas acumulados das 100 cossimulações realizadas (linhas pretas), sobrepostos pelo histograma acumulado da variável primária (DCCG) desagrupada (linha vermelha).

Figura 30- Histograma acumulado das 100 cossimulações juntamente com o histograma do dado primário desagrupado.



Fonte: Autoria própria.

Nos textos sobre incerteza na estimativa geostatística, o primeiro parâmetro que geralmente aparece é a variância de krigagem e, conseqüentemente, os intervalos de confiança derivados da mesma. A variância de krigagem fornece uma indicação da precisão do valor estimado e blocos tendo variâncias de estimativa similares são muito bem ou pessimamente estimados. Contudo, a variância de estimativa não leva em conta a distribuição assimétrica dos dados. Por esta razão, quando se trabalha com dados assimétricos, os intervalos de confiança (ICs) permitem uma maior apreciação da confiabilidade da estimativa do que a real incerteza local.

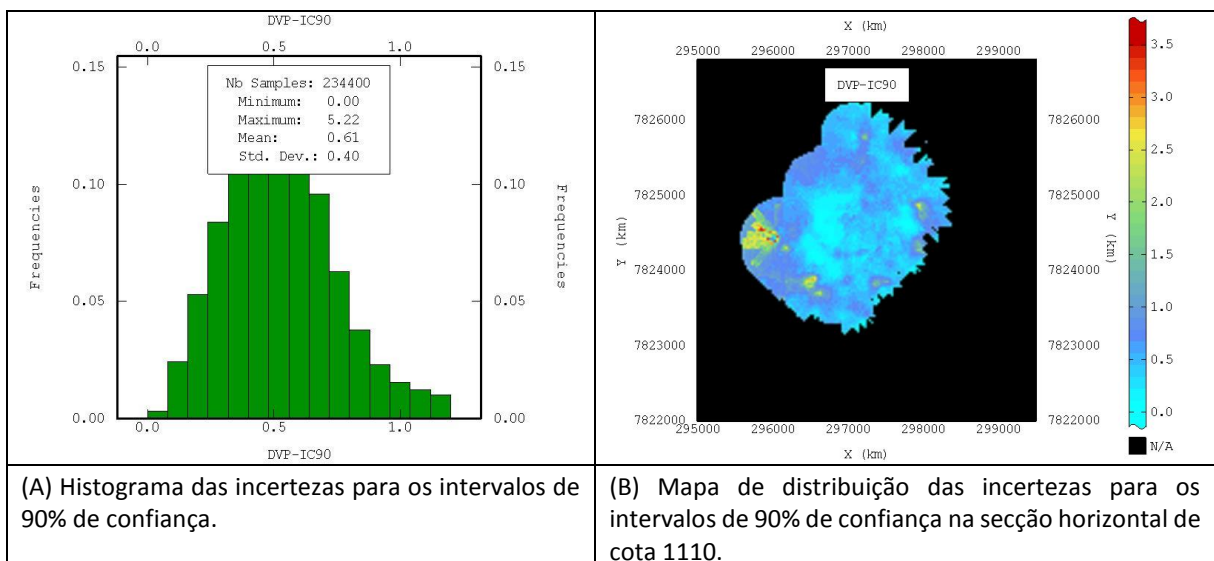
A predição espacial feita pela krigagem, embora válida, não permite determinar o intervalo de confiança das previsões realizadas pela simulação/cossimulação, pois, no caso da krigagem, gera-se apenas um evento probabilístico. A utilização de cossimulações condicionais permitiu cossimular vários cenários equiprováveis de recuperação metalúrgica (DCCG) para o depósito, sendo esses modelos baseados nos parâmetros e dados de entrada previamente definidos.

A distribuição dos valores cossimulados em torno da sua média representa experimentalmente a incerteza da estimativa. O erro relativo para o intervalo de 90% de confiança foi calculado pela subtração: $Q(95) - Q(5)$ da distribuição dos valores equiprováveis da cossimulação bloco a bloco, sendo esse valor posteriormente dividido pela média dos valores equiprováveis do bloco (KOPPE, 2017). Esse valor obtido foi ainda dividido por dois para obter o erro relativo com 90% de probabilidade.

Pelo fato de ter sido utilizada a cossimulação condicional nesse trabalho, espera-se que o intervalo de confiança nos nós cossimulados sejam iguais ou próximos de zero.

A figura 31 apresenta o histograma e o mapa de distribuição do erro relativo para o intervalo de 90% de confiança dos 100 possíveis cenários de recuperação metalúrgica cossimulados.

Figura 31- (A) Histograma do erro relativo no intervalo de 90% de confiança dos dados cossimulados. (B) Mapa dos erros relativos para os intervalos de 90% de confiança.



Fonte: Autoria própria.

Quanto menor for o erro relativo maior será a confiança no resultado cossimulado. No histograma da figura 31 observa-se que aproximadamente 50% dos dados têm erro relativo inferior a 0,5%.

4 CONSIDERAÇÕES FINAIS

Montgomery et al. (2001) escreveram que as equações de superfície de resposta podem ser representadas graficamente e usadas de três maneiras, sendo elas:

- (a) Descrever como as variáveis independentes de teste afetam a resposta;
- (b) Determinar as interações entre as variáveis independentes; e
- (c) Descrever o efeito combinado de todas as variáveis de teste na resposta.

Para este trabalho específico, a utilização do método de superfície de resposta teve como objetivo principal utilizar o efeito combinado de cinco variáveis (X_1 , X_2 , X_3 , X_4 e X_5) para descrever o fenômeno recuperação metalúrgica do nióbio.

Ainda, de acordo com Montgomery et al. (2001), cinco premissas devem ser respeitadas para o uso efetivo da superfície de resposta, sendo elas:

- (a) Os fatores críticos para o processo devem ser conhecidos;
- (b) A região onde os fatores influenciam o processo deve ser conhecida;
- (c) Os fatores devem variar continuamente ao longo do grupo experimental escolhido;
- (d) Deve existir uma função matemática que se relaciona com os fatores de resposta medida;
- (e) A resposta que é definida pela função matemática é uma superfície suavizada.

Antes de construir o modelo de regressão pela metodologia da superfície de resposta, foi necessário realizar uma análise exploratória criteriosa de dados estatísticos multivariados. Os resultados demonstram que é possível aplicar esta metodologia para descobrir os efeitos da combinação das variáveis químicas na resposta de recuperação metalúrgica durante o processo de concentração.

De acordo com Montgomery et al. (2001), na ausência de conhecimento suficiente sobre a verdadeira superfície de resposta, o usuário geralmente tenta o modelo de primeira ordem. No entanto, quando o modelo de primeira ordem não é suficiente para ajustar a superfície, é necessário incorporar termos de ordem superior para melhorar o modelo de regressão.

Utilizando modelos de regressão de primeira ordem, com o mesmo conjunto de dados deste estudo, verificou-se que o modelo obtido foi insuficiente para explicar o comportamento da recuperação metalúrgica a partir das mesmas cinco variáveis químicas. Utilizando este modelo de regressão de primeira ordem, a melhor correlação obtida foi de 52% (R^2), sendo essa correlação já otimizada pela análise residual. Um modelo de segunda ordem foi então assumido,

melhorando significativamente a correlação entre o dado calculado pela regressão (SEC-REC) com o dado experimental (DCCG).

Dentre todas as interações entre os pares de variáveis, a interação entre X_4 e X_3 se destacou para ajudar na explicação da função respostas, além da interação entre X_3 e X_1 que também exibiu alto coeficiente de correlação.

A contribuição quadrática de cada variável também ajuda a definir o modelo de regressão. A contribuição quadrática da variável X_4 e X_3 foram as mais importantes no caso analisado. A contribuição quadrática das demais variáveis foram muito baixas.

Sem a análise de resíduos, o coeficiente de regressão entre a recuperação metalúrgica analisada (DCCG) e o resultado da regressão dos óxidos para o cálculo da recuperação metalúrgica (SEC-REC) foi de 65,04% (R^2). Isso significa que as cinco variáveis químicas consideradas explicariam em parte (65,04%) a resposta para recuperação metalúrgica. Com o tratamento dos resíduos foram excluídos alguns *outliers* e dados incoerentes. Após o tratamento de resíduos, a correlação entre as duas variáveis (DCCG e SEC-REC) passou de 65,04% para 82,59%, tendo um aumento significativo.

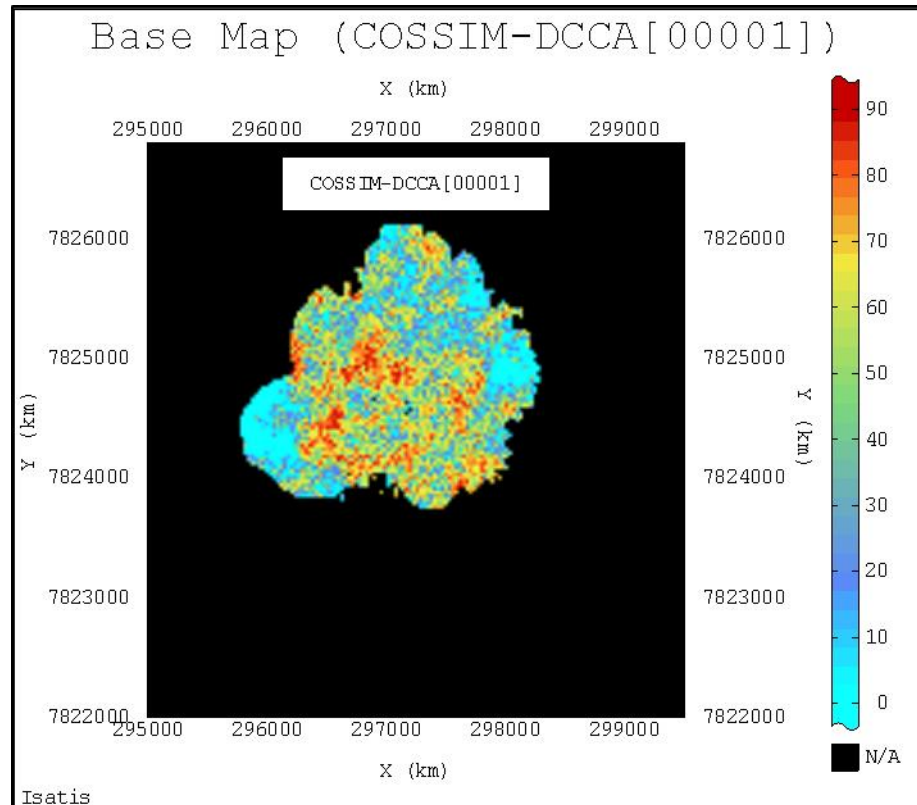
O algoritmo sequencial para a simulação conjunta de diversas variáveis aleatórias dependentes pode ser aplicado desde que as Funções de Distribuições Acumuladas Condicionais (CCDF) necessárias possam ser estabelecidas. A maneira mais fácil e mais comum de derivar essas distribuições condicionais é começar assumindo um modelo para a distribuição multivariada, sendo o modelo mais simples o modelo multigaussiano. Este modelo é extremamente simples e bem compreendido, e tem sido aplicado com sucesso para modelagem de variáveis contínuas com pouca ou nenhuma continuidade de valores extremos (Almeida, 1993).

O algoritmo utilizado requer uma distribuição gaussiana que implica uma normalidade univariada. Infelizmente, em Ciências da Terra, a maioria dos dados não apresentam uma distribuição univariada normal. Nesse caso, antes de utilizar o algoritmo, é preciso primeiramente transformar os dados originais em valores univariados normalmente distribuídos. A cossimulação é então realizada nesse domínio Gaussiano, sendo o seu resultado posteriormente retrotransformado para os valores do atributo original através da função de anamorfose inversa ao da transformação Gaussiana.

Um interessante e curioso aspecto na determinação do número de simulações estocásticas necessárias para cobrir o espaço de incertezas é a pouca referência às flutuações

ergódicas. Lantuéjoul (1994) reporta que, quanto maior for o alcance total do variograma, maior será este número, isto porque maior será a magnitude das flutuações ergódicas. A figura 32 mostra o mapa com a distribuição espacial de uma das 100 cossimulações realizadas.

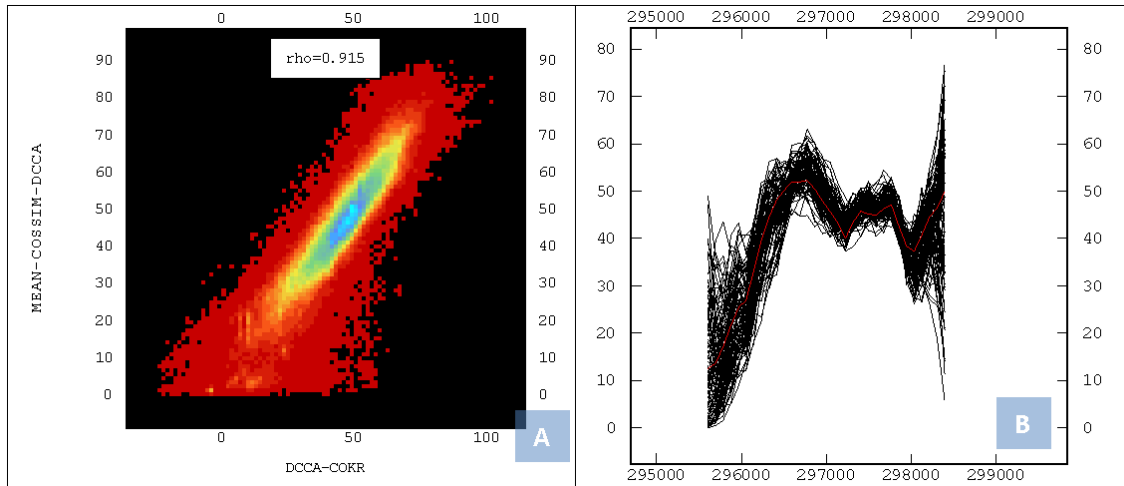
Figura 32- Mapa de uma das 100 cossimulações realizadas.



Fonte: Autoria própria.

Outra maneira de validar a cossimulação é comparar a média de toda cossimulação com o resultado da cokrigagem. Na teoria, esses dois resultados devem ser muito semelhantes. A figura 33 apresenta o *scatter plot* (33A) da média da cossimulação versus o resultado da cokrigagem e o *swath plot* (33B) de uma seção com o resultado das 100 cossimulações (Linhas pretas) e a cokrigagem (linha vermelha).

Figura 33- (A) Scatter plot entre a cokrigagem e a medias das cossimulações. (B) Swath plot (N0, D0) do resultado das 100 cossimulações (linhas pretas) e do resultado da cokrigagem (linha vermelha).

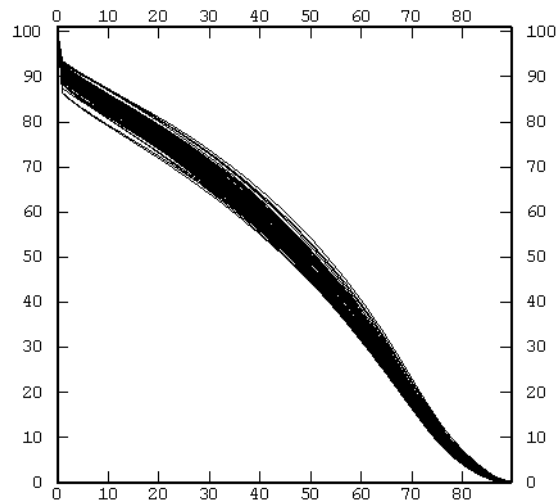


Fonte: Autoria própria.

Curvas simuladas de teor de corte e tonelagem de corte, onde cada realização constitui um cenário plausível para o depósito de nióbio foram construídas de modo que as verdadeiras curvas de tonelagem e metal estejam dentro do conjunto dessas curvas simuladas. Assim, tem-se uma imagem da incerteza sobre a quantidade total de recursos minerais. Uma análise de risco pode ser realizada com as curvas simuladas para avaliar os piores e os melhores cenários, sem necessidade de classificação. Além disso, pode-se avaliar a incerteza sobre os recursos minerais após a aplicação de um teor de corte sem gerar conclusões tendenciosas, enquanto, na maioria dos casos, utiliza-se a abordagem limitada aos recursos totais devido ao efeito da suavização dos estimadores com base na krigagem ou média das realizações. A abordagem de simulação é bastante consistente para quantificar a incerteza sobre os recursos, mas requer alguns cuidados.

A figura 34 apresenta as curvas de *cutoff* de recuperação metalúrgica por tonelagem obtidos pelos diferentes resultados da cossimulação da variável DCCG. No eixo vertical tem-se a tonelagem e no eixo horizontal tem-se os diferentes valores de corte para a recuperação metalúrgica.

Figura 34- Curvas de cutoff de recuperação metalúrgica por tonelagem obtidos pelos diferentes resultados da cossimulação da variável DCCG. No eixo vertical tem-se a tonelagem e no eixo horizontal tem-se os diferentes valores de corte para a recuperação metalúrgica.



Fonte: Autoria própria.

O modelo geoestatístico utilizado para cossimular as variáveis devem descrever com precisão a sua distribuição espacial. Em particular, o modelo deve incluir características estruturais como anisotropias, tendências, efeito proporcional, desestruturação dos graus extremos ou, ao contrário, conectividade dos valores de alto grau. Se necessário, o depósito deve ser dividido em vários domínios com uma distribuição homogênea de acordo com características geológicas tais como tipo de rocha, tipo mineral ou alteração.

Os limites da zona de minério para realizar uma boa cossimulação devem ser precisamente definidos levando em conta as considerações geológicas, e de processo: de fato, se as cossimulações forem realizadas fora da área mineralizada (isto é, em áreas de estéril), a quantidade de recursos aumentará irrealisticamente (o princípio da simulação estocástica é desenhar valores de grau de acordo com um modelo especificado, de modo que "cria" recursos não existentes se o domínio simulado aumenta demais). Na opinião do autor, a delimitação correta do depósito é a questão mais crucial para que a abordagem de cossimulação seja confiável. Na verdade, essa delimitação é inerente ao problema em questão: a definição dos recursos se refere implicitamente a um domínio espacial específico, e uma modificação desse domínio espacial específico pode alterar os valores dos recursos minerais.

5 CONCLUSÕES

Para criar um modelo que explique o comportamento da recuperação metalúrgica, foram testadas diferentes regressões matemáticas. A regressão multivariada de segunda ordem (superfície de resposta) mostrou-se adequada para uso na modelagem e análise do desempenho geometalúrgico do minério de nióbio, sendo essa uma variável influenciada por vários fatores.

A aplicação da equação de regressão multivariada foi restringida às amostras dispostas no domínio geológico conhecido como Alterito Laranja.

Antes de utilizar a metodologia de regressão por superfície de resposta, foi necessário definir quais variáveis independentes (óxidos obtidos por análise química via FRX) mais contribuem para explicar o fenômeno de interesse, neste caso, a recuperação metalúrgica do nióbio. Durante essa análise constatou-se que cinco variáveis químicas (X_1 , X_2 , X_3 , X_4 e X_5) são suficientes para explicar 82% do fenômeno “recuperação metalúrgica de nióbio”.

As interações entre as variáveis químicas consideradas, bem como suas interações quadráticas, contribuíram significativamente para explicar o fenômeno “recuperação metalúrgica do nióbio”. Utilizando a equação de regressão multivariada obtida, foi possível calcular a recuperação metalúrgica para todas as amostras ou locais onde as cinco variáveis químicas consideradas são conhecidas.

A metodologia de regressão por superfície de resposta, utilizada para prever a recuperação metalúrgica do nióbio, gerou informações com alta correlação da variável dependente (DCCG) com a supersecundária calculada (SEC-REC). A variável supersecundária foi então incorporada, através de uma rotina de cossimulação geoestatística, para melhorar a previsibilidade e o conhecimento do dado geometalúrgico de interesse.

A união de vários dados secundários, os quais isoladamente possuem baixa correlação com a variável de interesse, em uma única variável supersecundária, a qual possui uma alta correlação com a variável de interesse, viabilizou a implementação de uma rotina de simulação conjunta colocada utilizando o modelo simplificado de Markov.

A abordagem de cossimulação utilizada neste estudo foi baseada em um algoritmo de cossimulação gaussiana sequencial, o qual permitiu a simulação conjunta da variável supersecundária (SEC-REC) para integrar diferentes fontes de informação correlacionadas com a variável principal. Esse método de cossimulação incorpora a ideia de cokrigagem colocada para reduzir o esforço computacional envolvido no processo de simulação e torna a matriz de cokrigagem mais estável. Essa implica em manter a informação advinda da variável

supersecundária no mesmo local onde a variável primária foi estimada, dando origem ao conceito de variável supersecundária colocada.

O modelo de correionalização de Markov também foi absorvido para simplificar a inferência e modelagem das covariâncias cruzadas, utilizando o dado supersecundário colocado, reduzindo a influência dos dados secundários distantes do nó cossimulado.

Os resultados da cossimulação da variável geometalúrgica foram validados através da capacidade de reprodução dos variogramas e histogramas originais pela cossimulação. Também foi realizada uma análise por *swat plot* com o resultado da cossimulação. O erro relativo no intervalo de 90% de confiança foi calculado para avaliar o grau de confiabilidade do modelo gerado bloco a bloco. Além disso, foi feita uma comparação entre a média da cossimulação e o resultado de uma cokrigagem realizada com o mesmo conjunto de dados amostrais e parâmetros de estimativa.

De acordo com o resultado dos métodos de validação adotados, o modelo cossimulado gerado mostrou-se eficiente para a previsão da variável geometalúrgica com nível de precisão satisfatório, sendo adequado para prever o resultado da variável geometalúrgica considerada na escala de planejamento de longo prazo da mina de nióbio de Araxá, bem como no planejamento estratégico da empresa.

REFERÊNCIAS

- ABDI, Hervé; WILLIAMS, Lynne J. Principal component analysis. **Wiley interdisciplinary reviews: computational statistics**, v. 2, n. 4, p. 433-459, 2010.
- ALMEIDA, Alberto S. Joint Simulation of Multiple Variables with a Markov-Type Coregionalization Model, **Ph.D. Thesis**, Stanford University, 1993.
- ALMEIDA, Alberto S.; JOURNEL, Andre G. Joint simulation of multiple variables with a Markov-type coregionalization model. **Mathematical Geology**, v. 26, n. 5, p. 565-588, 1994.
- ANNELS, A. E.; DOMINY, S. C. Core recovery and quality: Important factors in mineral resource estimation. **Applied Earth Science**, v. 112, n. 3, p. 305-312, 2003.
- BATES, Douglas M.; WATTS, Donald G. **Nonlinear regression analysis and its applications**. Wiley, 1988.
- BOX, George EP; DRAPER, Norman R. **Empirical model-building and response surfaces**. New York: Wiley, 1987.
- CANCHO, Vicente G.; LOUZADA-NETO, Francisco; BARRIGA, Gladys DC. The Poisson-exponential lifetime distribution. **Computational Statistics & Data Analysis**, v. 55, n. 1, p. 677-686, 2011.
- CARR, James R.; MYERS, Donald E. COSIM: a FORTRAN IV program for conditional simulation. **Computers & Geosciences**, v. 11, n. 6, p. 675-705, 1985.
- CYSNEIROS, Francisco José A.; VANEGAS, Luis Hernando. Residuals and their statistical properties in symmetrical nonlinear models. **Statistics & Probability Letters**, v. 78, n. 18, p. 3269-3273, 2008.
- CORDEIRO, Gauss M. et al. Corrected maximum-likelihood estimation in a class of symmetric nonlinear regression models. **Statistics & probability letters**, v. 46, n. 4, p. 317-328, 2000.
- DEUTSCH, Clayton V. **Geostatistical Reservoir Modeling**. **Applied Geostatistics Series**, New York, Oxford University Press, 376 pp, 2002.
- DEUTSCH, Clayton V.; JOURNEL, Andre G. GSLib. **Geostatistical software library and user's guide**, v. 369, 1998.
- EMERY, Xavier; ORTIZ, Julián M. Shortcomings of multiple indicator kriging for assessing local distributions. **Applied Earth Science**, v. 113, n. 4, p. 249-259, 2004.
- FROST, J. How to interpret a regression model with low R-squared and low p values. **Minitab Inc. (ed) Getting Started with Minitab**, v. 17, 2014.

- GAVIRA, Muriel O. Simulação computacional como uma ferramenta de aquisição de conhecimento. **São Carlos: USP**, 2003.
- GEOVARIANCES. Manuais do Software ISATIS, versão 2.2. Fontainebleau, France: Geovariances, 1994.
- GÓMEZ-HERNÁNDEZ, J. Jaime; JOURNEL, André G. Joint sequential simulation of multigaussian fields. In: **Geostatistics Troia'92**. Springer Netherlands, p. 85-94, 1993.
- GOOVAERTS, Pierre. **Geostatistics for natural resources evaluation**. Oxford University Press on Demand, 1997.
- ISAAKS, Edward H. The Application of Monte Carlo Methods to the Analysis of Spatial Correlated Data: **PhD thesis**, Stanford University, CA, 1990.
- ISAAKS, Edward H.; SRIVASTAVA, R. Mohan. **Applied geostatistics**. 1989.
- JOURNEL, Andre G.; HUIJBREGTS, Ch J. **Mining geostatistics**. Academic press, 1978.
- KOLTERMANN, Christine E.; GORELICK, Steven M. Heterogeneity in sedimentary deposits: A review of structure-imitating, process-imitating, and descriptive approaches. **Water Resources Research**, v. 32, n. 9, p. 2617-2658, 1996.
- YAMAMOTO, J. K.; LANDIM, P. M. B. Geoestatística: conceitos e aplicações. **Oficina de Textos, São Paulo, (214 pp.)**, 2013.
- LANTUÉJOUL, Christian. Non conditional simulation of stationary isotropic multigaussian random functions. In: **Geostatistical Simulations**. Springer Netherlands, p. 147-177, 1994.
- LUSTER, Gordon Ray. Raw materials for Portland cement: **Applications of conditional simulation of coregionalization**. Stanford University, 1985.
- DE MARSILY, G. et al. Some current methods to represent the heterogeneity of natural media in hydrogeology. **Hydrogeology Journal**, v. 6, n. 1, p. 115-130, 1998.
- MATHERON, Georges. **The theory of regionalized variables and its applications**. École nationale supérieure des mines, 1971.
- MATHERON, Georges. The internal consistency of models in geostatistics. In: **Geostatistics**. Springer Netherlands, p. 21-38, 1989.
- MYERS, Donald E. Matrix formulation of co-kriging. **Mathematical Geology**, v. 14, n. 3, p. 249-257, 1982.

- MYERS, Donald E. Cokriging: new developments. **Geostatistics for natural resources characterization**, p. 295-305, 1984.
- MONTGOMERY, Douglas C., et al. Design and analysis of experiments, ed 3. New York, NY: **John Wiley and sons**, 2001.
- MONTGOMERY, Douglas C.; MYERS, Raymond H. Response surface methodology: process and product optimization using designed experiments. **Raymond H. Meyers and Douglas C. Montgomery. A Wiley-Interscience Publications**, 1995.
- MONTGOMERY, Douglas C. **Design and analysis of experiments**. John Wiley & Sons, 2008.
- NELDER, John A.; BAKER, R. Jacob. Generalized linear models. **Encyclopedia of statistical sciences**, 1972.
- OLEA, Ricardo A. **Geostatistics for engineers and earth scientists**. Springer Science & Business Media, 2012.
- PEARSON, Karl. LIII. On lines and planes of closest fit to systems of points in space. **The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science**, v. 2, n. 11, p. 559-572, 1901.
- RATKOWSKY, David A. Nonlinear Regression Modeling: Unified Practical Approach. Marcel Dekker, New York, 1983.
- REMY, Nicolas; BOUCHER, Alexandre; WU, Jianbing. **Applied geostatistics with SGeMS: A user's guide**. Cambridge University Press, 264p, 2008.
- SOARES, Amilcar. Direct sequential simulation and cosimulation. **Mathematical Geology**, v. 33, n. 8, p. 911-926, 2001.
- SURO-PEREZ, V.; JOURNEL, A. G. Indicator principal component kriging. **Mathematical Geology**, v. 23, n. 5, p. 759-788, 1991.
- VANEGAS, Luis Hernando; CYSNEIROS, Francisco José A. Assessment of diagnostic procedures in symmetrical nonlinear regression models. **Computational Statistics & Data Analysis**, v. 54, n. 4, p. 1002-1016, 2010.
- KOPPE, Vanessa Cerqueira; RUBIO, Ricardo Hundelshausen; COSTA, João Felipe Coimbra Leite. A Chart for Judging Optimal Sample Spacing for Ore Grade Estimation. **Natural Resources Research**, p. 1-9, 2017.

VERLY, G. Sequential Gaussian Cosimulation: **A Simulation Method Integrating Several Types of Information**, in Amilcar Soares (Ed.), Geostatistics Troia: Kluwer Academic Publishers, Dordrecht, Holland, p. 85–94, 1992.

VICINI, Lorena; SOUZA, Adriano Mendonça. Análise multivariada da teoria à prática. **Santa Maria: UFSM, CCNE**, 2005.

XU, Wenlong et al. Integrating seismic data in reservoir modeling: the collocated cokriging alternative. In: **SPE annual technical conference and exhibition**. Society of Petroleum Engineers, 1992.