
Contributions of computing in geosciences: a case study of geometallurgical forecasting

Contribuições da computação em geociências: um estudo de caso de previsão geometalúrgica

Received: 21-07-2024 | Accepted: 25-08-2024 | Published: 31-08-2024

Silvânia A Braga de Castro

ORCID: <https://orcid.org/0000-0002-1343-660X>

Centro Federal de Educação Tecnológica de Minas Gerais, Brasil

E-mail: silvaniabraga@cefetmg.br

Elenice Maria Schons Silva

ORCID: <https://orcid.org/0000-0003-1360-6450>

Universidade Federal de Catalão, Brasil

E-mail: eschons@ufcat.edu.br

André Carlos Silva

ORCID: <https://orcid.org/0000-0002-9760-0728>

Universidade Federal de Catalão, Brasil

E-mail: ancarsil@ufcat.edu.br

ABSTRACT

Geometallurgical modeling has been improved over the years with the incorporation of mineralogical, chemical, and metallurgical information. However, the diversity of the database brings challenges since the information is not unified in procedures, sample support, or quantities. Characteristics such as nonlinearity and nonadditivity of the geometallurgical database impose even more difficulties. The objective of this article is to demonstrate the feasibility of geometallurgical models from multivariate techniques combined with computational tools. The algorithms used in this study were developed in Python programming language. The multivariate analysis and machine learning techniques allowed the grouping of transitional typological domains and the quantification of uncertainty; the selection and reduction of the dimensionality of the heterotopic database; and the prediction of metallurgical behavior from exhaustive primary variables using a transfer function. The results were verified and proved to be satisfactory. Other benefits of the computational structure were: reduction of subjectivity; automation; and dynamic updates of models using new data.

Keywords: Geosciences; Computing; Geometallurgy.

RESUMO

A modelagem geometalúrgica tem sido aprimorada ao longo dos anos com a incorporação de informações geológicas, químicas e metalúrgicas. Contudo, a diversidade do banco de dados traz desafios já que as informações não são unificadas nos procedimentos, no suporte amostral e nem nas quantidades. Características como não-linearidade e não aditividade do banco de dados metalúrgicos impõem ainda mais dificuldades. O objetivo deste artigo é demonstrar a viabilidade dos modelos geometalúrgicos a partir de técnicas multivariadas combinadas com ferramentas computacionais. Os algoritmos utilizados nesse estudo foram desenvolvidos em linguagem de programação Python. As técnicas de análise multivariada e aprendizado de máquina, permitiram realizar o agrupamento de domínios tipológicos transitórios e quantificar a incerteza; reduzir a dimensionalidade do banco de dados heterotópico e prever o comportamento metalúrgico a partir das variáveis primárias exaustivas usando uma função de transferência. Os resultados foram verificados e se mostraram satisfatórios. Outros benefícios da estrutura computacional foram: redução da subjetividade; automação e atualizações dinâmicas de modelos usando novos dados.

Palavras-chave: Geociências; Computação; Geometalurgia.

INTRODUÇÃO

Os depósitos minerais de rocha fosfática no Brasil são formados por rochas ígneas (GIBSON et al., 1995). A complexidade genética das intrusões alcalinas e as ações do intemperismo explicam a alta variabilidade dos minérios fosfáticos (BRAGA & COSTA 2016) e com isso as dificuldades em obter respostas satisfatórias nas unidades de processamento mineral. As definições das tipologias de minérios e a incorporações dos dados de processos devem compor o modelo geometalúrgico para subsidiar o planejamento de lavra garantindo uma alimentação homogênea para a planta de processamento (DEUTSCH, 2013). A integração entre essas áreas resulta numa diversidade de informações dispendiosa para a modelagem 3D, já que, na maioria das vezes, os dados são amostrados por métodos distintos, em diferentes volumes e normalmente com informações metalúrgicas em quantidades muito inferiores em relação às amostras de teores e litologias (LOZANO & BENNET, 2003). Esses fatores impedem a integração das respostas metalúrgicas no modelo de recursos usando métodos geoestatísticos.

Como consequência uma busca por técnicas computacionais para tratamento de dados tem aumentado. Avanços recentes nas ciências da computação estão revolucionando a geometalurgia, fornecendo ferramentas que podem gerenciar a complexidade e a variabilidade dos dados de geociências. Nesse estudo de caso, a interface multidisciplinar entre geociências, estatística e ciências da computação utilizando diferentes algoritmos viabilizou a proposta do modelo geometalúrgico para a mina de Catalão, sem que informações adicionais fossem coletadas, permitindo o aproveitamento máximo dos recursos materiais e do banco de dados disponível pela empresa.

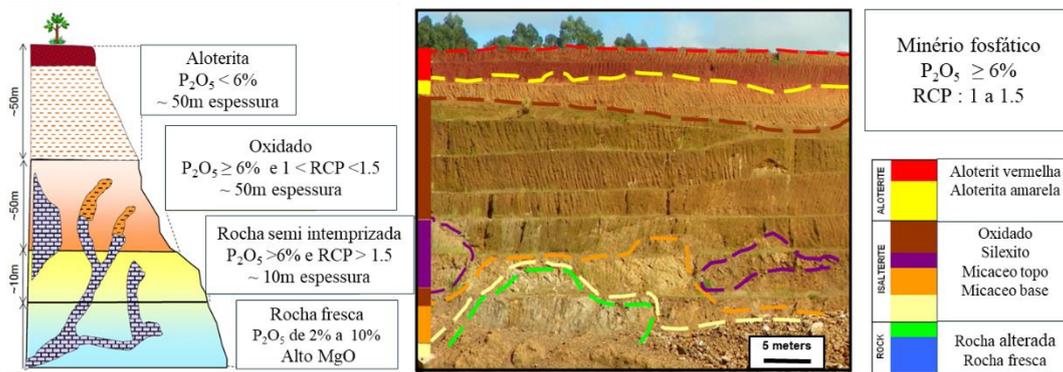
ÁREA DE ESTUDO, GEOLOGIA E MINERALIZAÇÃO

A Mina em estudo está posicionada no Complexo Catalão I, no estado de Goiás-GO, tem estrutura dômica com aproximadamente 6 km de diâmetro e é constituído por um conjunto de magmas neocretácios alcalinos intrusivos (GIBSON et al., 1995). Conforme Brod et al. (2004), três séries petrogenéticas distintas constituem o complexo Catalão I: bebedourítica, foscorítica e carbonatítica. A série bebedourítica representa as rochas produzidas por cristalização fracionada de magma silicático ultrapotássico (flogopita picrito) e inclui dunitos, clinopiroxenitos, bebedouritos, sienitos e são constituídos por minerais: olivina, clinopiroxênio, flogopita, perovskita, magnetita e

apatita. Estão frequentemente intrudidas por stockworks de carbonatito e foscorito, o que ocasiona um metassomatismo potássico e transforma as rochas ultramáficas da série bebedourítica em flogopititos (RIBEIRO, 2008). Os flogopititos metassomáticos são potencialmente mineralizados em apatita, já os de origem magmática não demonstram a mesma potencialidade. A série foscorítica é formada por rochas cristalizadas a partir de resfriamento de magmas fosfáticos representadas por foscoritos, nelsonitos, magnetititos e apatititos; e são constituídas pelos minerais magnetita, olivina e apatita. Os foscoritos ocorrem em veios, diques, pequenos bolsões e plugs cortando diversos tipos de rochas. A série foscorítica se faz mais importante do ponto de vista econômico, pois são responsáveis por grande parte da jazida em exploração (RIBEIRO, 2008). A série carbonatítica consiste em rochas compostas por mais de 50% em volume de carbonatos, essencialmente dolomita, calcita, calcita mais dolomita, ou dolomita /ankerita (BROD, 2004). Os grandes volumes de carbonatito presente no complexo é sempre dolomita carbonatito, sendo que os calcita carbonatitos ocorrem em bolsões carbonatíticos que alimentam veios e diques espalhados por todo o complexo (RIBEIRO, 2008). Nas rochas calcita carbonatitos, a apatita ocorre como cristais prismáticos curtos, arredondadas, disseminados na rocha. Nos carbonatitos dolomíticos, a apatita ocorre em bolsões monominerálicos, sugerindo tratar-se de um produto de imiscibilidade de um líquido fosfático a partir do magma carbonatítico.

Sobre essas rochas desenvolveu-se um espesso perfil de alteração laterítica, com aproximadamente 100m. Oliveira e Imbernon (1998), com base em características petrológicas e químicas, definem quatro horizontes distintos, a saber, rocha alterada, saprólito isalterítico, saprólito aloterítico e solo. A ação intempérica provoca modificações de ordem física e química nas rochas e minerais. Jazidas e zonas mineralizadas são, por vezes, assim formadas por concentração residual de minerais econômicos. A figura 1 mostra a relação entre os teores de P_2O_5 % com as classes de minério utilizadas e o perfil intempérico. Segundo Brod et al. (2004), o minério supergênico de Catalão I é classificado em três tipos básicos, do topo para a base, coincidindo com intensidade decrescente de intemperismo: minério oxidado, minério micáceo e rocha alterada (minério sílico carbonatado). Os dois primeiros tipos correspondem ao horizonte isalterítico. O minério micáceo pode ser subdividido em micáceo de topo e micáceo de base; a diferença principal entre eles é a presença de vermiculita e flogopita na base, e ausência de flogopita no micáceo de topo.

Figura 1 – Modelo do Perfil de Intemperismo de Catalão I proposto por Ribeiro, (2008).



Fonte: Adaptado de Ribeiro, 2008.

O minério oxidado corresponde às porções mais evoluídas do saprolito isalterítico possuindo uma coloração variando do ocre, róseo a marrom escuro e consiste fundamentalmente em apatita e óxidos e hidróxidos de ferro (magnetita, hematita e goethita). A presença da apatita é responsável pelo aumento dos teores de P_2O_5 acima de 6%. A relação CaO/P_2O_5 , conhecida como RCP se mantém baixa (menor que 1.5%). O horizonte aloterita é marcado por baixos teores (em torno de 1%) P_2O_5 , correspondendo à zona de estéril, em função da lixiviação e alteração dos minerais da rocha original.

METHODOLOGIA

A figura 2 resume as etapas de um programa geometalúrgico. A fase I é marcada pela obtenção do banco de dados usado na fase II de confecção do modelo geometalúrgico. A meta desse estudo, inserida na fase II, é: definir os domínios de estimativa (zonas mineralizadas e homogêneas), selecionar variáveis significativas e testar técnicas para predição das respostas de processo.

Figura 2. Exemplo das etapas de um programa geometalúrgico. Adaptado de P. Lamberg, 2011.



Disponibilidade dos dados

O repositório de códigos usado na metodologia pode ser encontrado no GitHub através do link <https://github.com/madamsilte/classfuzzy.git>.

Banco de dados

O banco de dados utilizado consta de um mapa geológico das frentes de lavra na escala 1:1000 e amostras de furos de sondagem até a data 21/03/2022. Nas amostras de furo de sondagem, 18.583 apresentam descrição geológica de 32 litotipos; 44.803 informações químicas de até nove óxidos analisados e 1.548 testes de bancada de dez variáveis de processo.

Uma avaliação exploratória envolvendo composição das amostras, verificação de litologia e análise estatística das variáveis quantitativas (histograma, matriz de correlação, box plot, etc.) em todo o banco de dados foi realizada previamente à aplicação de qualquer técnica de análise multivariada. Três pontos relacionados ao banco de dados merecem destaques: o primeiro é sobre as descrições litológicas ambíguas (mais de uma litologia reconhecida por amostras, resultando em 32 combinações tipológicas) presentes em 53% do banco; o que dificulta o direcionamento das amostras para compor os domínios de modelagem. O segundo é a quantidade reduzida da informação de processo e a falta de isotopia dos atributos medidos, ou seja, nem todas as litologias foram ensaiadas, e quando ensaiadas as quantidades de testes não eram igualmente distribuídas. O terceiro está relacionado ao tratamento dos dados de processo e a não aditividade de algumas variáveis. A aditividade é a propriedade que permite que a média de algumas variáveis seja calculada por meio de uma simples média linear (CARRASCO et al. 2008). Em atributos que apresentam características de não aditividade, a estimativa linear não deve ser aplicada, pois pode resultar em médias tendenciosas. Esses fatores dificultam a modelagem direta das variáveis de processo.

Como tentativa de solucionar as limitações relatadas e reproduzir o modelo geometalúrgico, alguns algoritmos da inteligência computacional foram testados e serão discutidos a seguir.

Definição dos domínios tipológicos

A elaboração dos modelos 3D dos depósitos minerais para subsidiar a lavra ocorre após as etapas de amostragem e testes de laboratórios (LOZANO & BENNET, 2003). Identificar zonas ou domínios mineralizados é uma das primeiras tarefas na modelagem (BRAGA & COSTA, 2016) e pode ser realizada por técnicas como interpretação manual, modelagem implícita e métodos geoestatísticos avançados. Na prática, essa atividade é frequentemente manual e laboriosa para o geomedelador, sendo propensa a erros de subjetividade e ignorando a incerteza associada à interpretação. Além disso, a modelagem

dos domínios tem sido restrita aos grupos geológicos com base na litologia e teor. Os domínios metalúrgicos são modelados com menos frequência, dada a quantidade inferior de informação, como acontece no banco de dados em estudo. Uma adaptação pode ser obtida para suprir essa demanda quando as relações entre os grupos geológicos e as respostas de processo são conduzidas pelo conhecimento especialista para definir os domínios de estimativa (WILLIAMS & RICHARDSON, 2004).

Com esse propósito, a equipe especialista da mina pré-definiu com base no conhecimento do depósito oito domínios tipológicos predominantes: dolomita carbonatito (CB), foscoritos (FO), flogopititos (FL), picrito brechado (PI), calcita carbonatito (CBMS) e foscorito com carbonatito (CBMG). Porém, a modelagem desses domínios é uma etapa crítica já que as descrições litológicas são ambíguas. As rochas do depósito ocorrem na forma de diques com espessura centimétrica a decimétrica, que se recortam em diferentes direções, formando uma estrutura conhecida como *stockwork* (BROD *et al.*, 2004), os contatos são transicionais e sobrepostos o que dificulta a descrição precisa na escala dos furos, podendo levar a uma classificação tipológica distinta dos grupos predominantes.

Para tentar lidar com a complexidade do fenômeno estudado e reduzir o tempo de análise manual pelo geomodelador para agrupar as tipologias transicionais e imprecisas, o algoritmo fuzzy foi testado (Figura 3). A lógica fuzzy foi escolhida em função do conceito dos conjuntos nebulosos cujo limites não são precisos. Mais aplicações da lógica fuzzy em geociências pode ser encontrada em Demicco, 2003; Orozco-del-Castillo, 2011; Moreira, 2002. Para a construção do algoritmo fuzzy um subconjunto do banco de dados original (47%) foi gerado contendo somente as amostras com uma única descrição litológica, consideradas mais uniformes ou precisas. Os teores foram analisados sobre critérios estatísticos a fim de criar uma identidade química para cada domínio. O processo de fuzzificação, transformou os valores das variáveis de entrada para a linguagem usual, ou seja, as características químicas de cada tipologia foram classificadas qualitativamente em termos linguísticos (baixo, médio e alto), utilizando o conhecimento especialista e, quantitativamente, atribuindo-se um grau de compatibilidade aos termos linguísticos através da função de pertinência trapezoidal. A base de regras foi montada para os oito domínios de estimativa e o método de inferência fuzzy uniu cada uma das proposições da base de regras, por meio de operações lógicas (AND), para simular o processo de dedução humana e produzir a variável fuzzy de saída. O processo se completou com a desfuzzificação do conjunto fuzzy de saída [0,1], em um número real nesse caso, uma

estimativa da classe pertencente. A metodologia completa do modelo de inferência fuzzy adotado pode ser encontrada em Castro et al. (2024).

As amostras classificadas pelo algoritmo fuzzy foram validadas e os domínios interpolados no modelo 3D por krigagem dos indicadores. O resultado do agrupamento fuzzy foi comparado com duas técnicas de aprendizado de máquinas tradicionais e bem conhecidas, k-means e hierárquico.

Figura 3. Arquitetura básica do algoritmo fuzzy.

```

from openpyxl import load_workbook
from collections import namedtuple, defaultdict
import enum
import tqdm

# =====
# Arquitetura básica
# =====

Regra = namedtuple('Regra', 'proposicoes consequente')

class Niveis(enum.IntEnum):
    BAIXO = 0
    MEDIO = 1
    ALTO = 2

def fn_trapezio(x, a, b, c, d, alargamento=0.25):
    if x < a:
        return 0
    elif x < b:
        return (x - a) / (b - a)
    elif x < c:
        return 1
    elif x <= d:
        return (d - x) / (d - c)
    else:
        return 0

def classificar(teores, regras, t_norm=min, fn_implicacao=min, fn_agregacao=max):
    # Nomenclatura conforme exemplo 2.5
    UFC = {}
    U = defaultdict(list)
    FR = defaultdict(list)

    # Passo 1: Fuzzificação, aplicação da T-Norma e construção dos conjuntos de saída
    for regra in regras:
        print(f'Litologia {regra.consequente[0]}')
        mu_proposicoes = []
        for elemento, conjunto in regra.proposicoes:
            entrada = teores[elemento]
            fn = fn_implicacao[elemento][conjunto]
            print(f'\t {elemento}={entrada} é {Niveis(conjunto).name}: {fn(entrada)}')
            mu_proposicoes.append(fn(entrada))
        mu_r = t_norm(mu_proposicoes)
        litologia, conjunto = regra.consequente
        fn = fn_saida[litologia][conjunto]
        fr_valores = []
        for i in range(1001):
            x = i * 0.1
            mu_saida = fn(x)
            #Print(fn_implicacao(mu_r, mu_saida))
            fr_valores.append(fn_implicacao(mu_r, mu_saida))
        FR[litologia].append(fr_valores)

    # Passo 2: Agregação
    for litologia, conjuntos in FR.items():
        for i in range(1001):
            U[litologia].append(fn_agregacao([conjunto[i] for conjunto in conjuntos]))

    # Passo 3: Centro de Massa
    for litologia, conjunto in U.items():
        num = 0
        den = 0
        for i, y in enumerate(conjunto):
            x = i * 0.1
            num += x * y
            den += y
        UFC[litologia] = num/den if den != 0 else 0

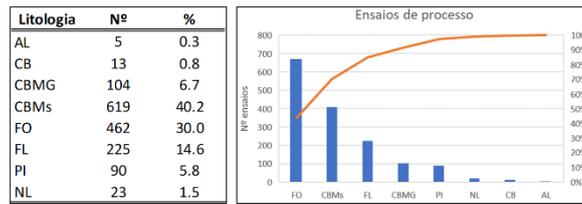
```

Seleção de variáveis e definição dos domínios de processo

A análise das variáveis de processo nos domínios modelados revelou que somente quatro dos oito grupos, correspondem as zonas de minérios. Ou seja, apesar da apatita estar presente nos outros grupos, eles não apresentam potencialidade e isso explica a quantidade baixa de ensaios de bancadas nessas litologias, pois, os ensaios são realizados para subsidiar a lavra de minérios determinada pelos teores de P_2O_5 e RCP. Portanto, a predição das informações de processo nas etapas seguintes será realizada nos grupos FO, FL, CBMs e CBMG (Figura 4).

Os testes de processo não são suficientes para a interpolação direta desses atributos, e para contornar esse problema, uma opção é testar uma função de transferência que adeque as relações entre as variáveis primárias, aditivas e exaustivas com as variáveis de respostas. O banco de dados dos teores é aditivo e exaustivo, contendo mais de 40 mil amostras, porém, não é isotópico para todos os elementos analisados. Os dados metalúrgicos não chegam a 4% das informações dos teores. Por esta razão, a seleção das variáveis é fundamental tanto para preservar o maior número de amostras possíveis, quanto para otimizar as etapas futuras de simulação das múltiplas variáveis.

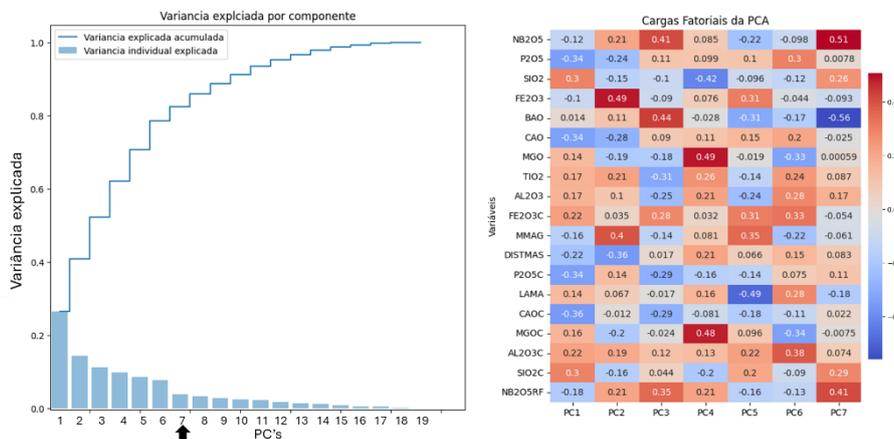
Figura 4. Quantitativo dos ensaios de processo por domínio.



A análise de componentes principais (PCA) é uma técnica de análise multivariada muito utilizada na seleção de atributos e redução da dimensão dos dados. Os problemas de PCA envolvem a realização de uma transformação linear de dados para obter autovalores (λ) e autovetores (α), maximizando a variância da transformação para uma matriz de covariância Σ . Os autovalores λ são ordenados segundo a porcentagem de covariância. Os primeiros autovalores fornecem a porcentagem máxima de covariância dos dados, sendo esta informação utilizada como critério de seleção de atributos (Al KANDARI & JOLLIFFIE 2005).

O algoritmo PCA foi aplicado no banco de dados e foram avaliadas três abordagens para indicar o número de variáveis selecionadas, nomeadamente métodos B2 e B4 propostos por Jolliffe (1972), e um índice de importância de variáveis (IIV) apresentado por Cervo (2015). Os dados foram normalizados e o critério utilizado para descarte das variáveis foi o recomendado por Jolliffe (1972), ou seja, o número de variáveis descartadas deve ser igual ao número de componentes em que a variância explicada (autovalor) é inferior a 0,7. Com o alcance dos componentes principais (PC's) foram obtidos os autovalores e os autovetores, revelando que um número reduzido de sete componentes principais explica mais de 80% da variabilidade dos dados (Figura 5). Informações sobre a metodologia completa podem ser vistas em Castro & Silva (2024).

Figura 5. Variância explicada (esquerda) e carga fatorial das sete primeiras PC's (direita).



Modelos de predição das variáveis primárias

Algoritmos de aprendizado de máquina são amplamente utilizados para conectar as propriedades da rocha com as variáveis de processo (DEUTSCH, 2013). Regressão linear (RL), Random forest (RF) e support-vector machine (SVM) são ferramentas de predição com resultados promissores (DEUTSCH et al., 2013; LISHCHUK et al., 2019). A regressão linear é um dos métodos mais intuitivos e utilizados para essa finalidade. Na regressão linear, uma variável dependente pode ser explicada por uma ou mais variáveis independentes mediante uma função ($f(x)$) (MONTGOMERY et al., 2012). A técnica random forest consiste na construção de diversas árvores de decisão que têm seus resultados combinados para fornecer a estimativa de uma variável dependente. Uma árvore de decisão é uma estrutura binária que segmenta em partes o banco de dados, a partir de critérios de ganho de informação e entropia. A divisão entre os ramos das árvores de decisão é realizada escolhendo um critério de otimização utilizando funções que meçam o máximo espalhamento dos dados, tal como a função de Gini ou a função entropia. O SVM, em sua forma básica, é um classificador binário linear, que identifica um único limite entre duas classes. O SVM assume que os dados multidimensionais são linearmente separáveis no espaço de entrada. Em particular, as SVMs determinam um hiperplano ótimo para separar o conjunto de dados em um número discreto de classes predefinidas usando os dados de treinamento. Para maximizar a separação ou margem, SVMs usam uma parte da amostra de treinamento que fica mais próxima no espaço de recursos do limite de decisão ótimo, agindo como vetores de suporte (SCHOLKOPF & SMOLA, 2002). Existem casos em que não é possível separar bem as instâncias de forma linear e, neste caso, podemos mapear os dados para um espaço dimensional maior (onde eles serão linearmente separáveis por estarem mais dispersos) e criar um classificador linear nesse espaço.

Os três algoritmos foram avaliados para prever as informações metalúrgicas: recuperação mássica, teor de P_2O_5 no concentrado e massa de material magnético, a partir das variáveis selecionadas na PCA para cada domínio tipológico. Os parâmetros dos algoritmos seguem na tabela 1. Os resultados foram verificados por validação cruzada através do algoritmo K-fold. A validação cruzada é um método simples e muito utilizado para verificar os modelos de predição. A ideia da validação cruzada é dividir o banco de dados em k partes iguais e utilizar cada uma das partes como conjunto de teste em uma das rodadas e como conjunto de treinamento nas demais. Para cada subdivisão, vamos utilizar dados diferentes para treino e teste. Ou seja, vamos ter resultados diferentes de

acordo com cada iteração K do nosso modelo. Ao final das k iterações, calcula-se a acurácia sobre os erros encontrados e obtém-se uma medida mais confiável sobre a capacidade do modelo de representar o processo gerador dos dados.

Tabela 1. Parâmetros dos algoritmos usados para os modelos de predição.

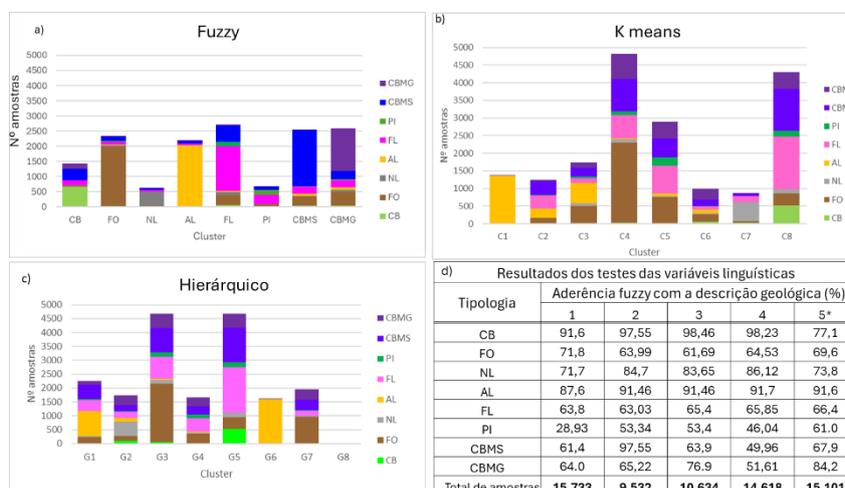
Algoritmo	Parâmetros
Regressão linear	Estimativa dos fatores= mínimos quadrados ordinários
Floresta aleatória	Número de árvores na floresta= 1000; Profundidade máxima da árvore = (até que o número mínimo de amostras para dividir um nó seja 2); O número máximo de recursos é = (número de recursos);
SVM	Número mínimo de amostras para uma folha = (2, 3, 4, 5, 6, 7, 8, 9, 10); Número mínimo de amostras para uma divisão = (2), Critério= (entropia)
K fold	C= (0.01, 1, 10, 35, 50, 100); ϵ = 0.001; Kernel= (linear, RBF)
	Número de folds= 5

RESULTADOS E DISCUSSÕES

Algoritmo Fuzzy

Vinte simulações foram realizadas, alterando os atributos e os valores da função pertinência conforme as análises estatísticas. As amostras classificadas foram comparadas com as descrições geológicas, admitindo que, mesmo imprecisa, essa informação é correta. A porcentagem de acerto na simulação 5 (escolhida) foi aproximadamente 70% (Figura 6d), com menores índices para as tipologias com características químicas muito semelhantes (flogopititos e picrito com xenólitos). Os grupos fuzzy foram verificados com outros métodos de agrupamento. Na figura 6, é possível observar que o algoritmo fuzzy mostrou-se mais eficiente para separar os domínios, com menos misturas de tipologias nos grupos gerados quando comparado aos demais. Como informação adicional à classificação dos domínios, o algoritmo difuso atribuiu um grau de pertinência das tipologias a todas as amostras que pode ser usado como uma medida de probabilidade, ou seja, uma maneira simples de quantificar a incerteza do agrupamento.

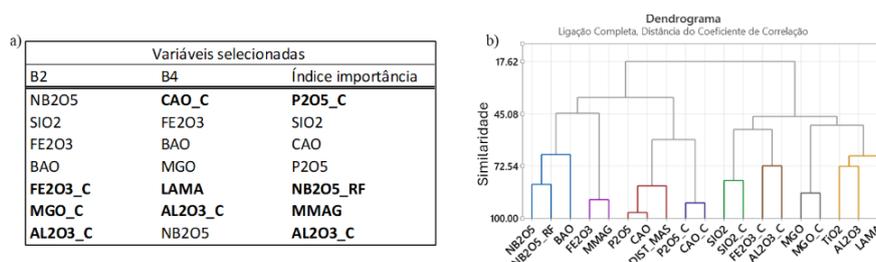
Figura 6. Resultados dos agrupamentos. Em a) fuzzy; b) k-means e c) hierárquico.



Algoritmo PCA

Doze dos dezanove componentes principais apresentaram variância inferior a 0,7, isso significa que é possível descartar 12 atributos, restando sete variáveis significativas, uma redução de 63% no número de variáveis. Vale ressaltar que em todas as metodologias, quatro das sete variáveis selecionadas são químicas e, as três variáveis de processo (em negrito na figura 7a) apresentam valores elevados de correlação com as outras variáveis químicas (Figura 7b), sendo consideradas redundantes. O Fe_2O_3 apresenta alta correlação MMAG consequentemente, os valores altos de um induzem os valores altos do outro e por isso não precisam coexistir simultaneamente. O mesmo acontece com Nb_2O_5 e $Nb_2O_5_RF$; P_2O_5 , CaO , $P_2O_5_C$ e CaO_C ; MgO e MgO_C . Assim, podemos concluir que as variáveis químicas exaustivas são as mais relevantes para explicar a variabilidade dos dados.

Figura 7. Em a) variáveis selecionadas na PCA, em b) similaridade entre as variáveis.



Algoritmos regressão linear, random forest e SVM

Das variáveis selecionadas na PCA, três (P_2O_5 , Fe_2O_3 e MgO) estão presentes em todas as amostras do banco de dados. Além disso, essas variáveis são utilizadas para classificação do minério e controle de lavra, portanto o rigor analítico desses atributos tende a ser maior, ponto positivo a considerar nos modelos de predição. Para verificar se haveria impacto em trabalhar com menos variáveis, os modelos de predição foram testados utilizando todas as variáveis químicas selecionadas e somente essas três. Não foi observado uma queda significativa nos coeficientes de correlação quando se utilizou as três variáveis (Figura 9a). Os modelos de predição apresentaram de maneira geral índices satisfatórios para previsão das respostas metalúrgicas. O Random forest gerou os maiores coeficiente de correlação, porém, os melhores resultados para as métricas: erro médio absoluto (MAE) e erro quadrático médio (MSE) foram obtidos no modelo SVM e RL. Os valores de MAE e MSE (Figura 9b) entre os três modelos se mantiveram muito próximos, variando, na maioria das vezes, em casas decimais. O coeficiente de correlação mostrou

maior diferença do RF para RL e SVM, sendo o RF o método escolhido para predição das variáveis de processo. Esses resultados são uma importante contribuição das ciências da computação para previsibilidade da usina, pois permitirão em etapas futuras a simulação dos teores e aplicação da função de transferências nos dados simulados para compor o modelo de recursos final.

Figura 9. Em a) coeficiente de correlação dos modelos usando todas e três variáveis. Em b) Validação cruzada dos modelos preditivos e métricas de avaliação.

a) Comparação dos modelos alterando o número de variáveis							b) Validação cruzada dos algoritmos testados com as variáveis P_2O_5 , Fe_2O_3 , MGO											
Variável	Domínio	Todas variáveis			P_2O_5 , Fe_2O_3 , MGO			Variável	Domínio	RL			RF			SVM		
		RL	RF	SVM	RL	RF	SVM			MAE	MSE	CORRELAÇÃO	MAE	MSE	CORRELAÇÃO	MAE	MSE	CORRELAÇÃO
Recup. Mássica	FO	0.48	0.86	0.63	0.46	0.83	0.52	FO	4.05	27.60	0.46	4.26	30.47	0.83	4.05	28.10	0.52	
	FLO	0.49	0.86	0.72	0.37	0.77	0.42	FLO	4.89	44.47	0.37	5.19	48.67	0.77	5.10	46.51	0.42	
	CBMS	0.68	0.9	0.78	0.67	0.87	0.7	CBMS	4.78	36.76	0.67	5.13	42.30	0.87	4.75	36.76	0.7	
	CBMG	0.64	0.88	0.77	0.62	0.85	0.68	CBMG	4.90	41.12	0.62	5.00	42.44	0.85	4.87	41.54	0.68	
Massa magnético	FO	0.82	0.94	0.88	0.79	0.92	0.82	FO	5.48	57.53	0.79	5.61	60.42	0.92	5.21	55.65	0.82	
	FLO	0.88	0.96	0.94	0.85	0.94	0.88	FLO	6.63	71.09	0.85	6.82	75.15	0.94	6.23	68.59	0.88	
	CBMS	0.66	0.9	0.79	0.57	0.87	0.65	CBMS	4.44	35.93	0.57	4.58	37.54	0.87	4.38	34.85	0.65	
	CBMG	0.83	0.95	0.89	0.79	0.94	0.86	CBMG	5.28	50.30	0.79	4.60	45.72	0.94	4.64	41.52	0.86	
P205_C	FO	0.62	0.87	0.81	0.43	0.81	0.51	FO	2.08	8.02	0.43	2.08	8.07	0.81	2.09	8.07	0.51	
	FLO	0.67	0.9	0.86	0.53	0.86	0.59	FLO	3.45	18.20	0.53	3.44	19.08	0.86	3.46	19.17	0.59	
	CBMS	0.73	0.93	0.9	0.56	0.85	0.6	CBMS	2.05	7.95	0.56	2.10	8.83	0.85	1.99	8.37	0.6	
	CBMG	0.78	0.92	0.91	0.62	0.87	0.65	CBMG	2.17	8.65	0.62	2.17	8.97	0.87	2.09	9.04	0.65	

CONCLUSÕES

A geometalurgia é um ramo multidisciplinar da engenharia e cada vez mais as ferramentas computacionais se fazem necessárias quando se pretende obter respostas de processo. Nas principais fases da construção do modelo geometalúrgico vimos que a integração entre o conhecimento geológico e as ciências da computação foi fundamental para a viabilidade do modelo. A capacidade de definir classes consistentes usando o algoritmo fuzzy reduziu os riscos da etapa crítica de modelagem dos domínios. Além de garantir a replicabilidade do procedimento, diminuiu a subjetividade e o tempo gasto na interpretação dos dados pelo geomodelador, e introduziu uma medida de incerteza dada pelo grau de pertencimento da classificação fuzzy. A seleção dos atributos por PCA otimizou todas as análises subsequentes relacionadas aos modelos preditivos, não só por eliminar recursos redundantes, garantido maior acurácia e agilidade dos algoritmos e procedimentos geoestatísticos, mas também por evidenciar que as variáveis aditivas eram suficientes para captar a variabilidade do fenômeno neutralizando questões relacionadas não aditividade das variáveis de processo.

Em última análise, vimos que obter funções de transferência, que associem as características das propriedades primárias do depósito com o desempenho da planta é plausível quando o banco de dados metalúrgico não consegue suprir as premissas para

uma boa modelagem desses atributos. Essa alternativa tem possibilitado a construção de modelos preditivos cujo objetivo final é otimizar recursos minerais, custos e receitas.

É fato que integração entre geociências, estatística e computação é um pilar fundamental para os estudos geometalúrgicos.

REFERENCIAS

AL KANDARI, N. M.; JOLLIFFE, I. T., 2005. Variable selection and interpretation of correlation principal component. *Environmetrics*, Published online 28 April 2005 in Wiley InterScience (www.interscience.wiley.com) 16: 659-672 2005.

BRAGA, S. A., COSTA, J. F. C. L., 2016. Krigagem dos indicadores aplicada a modelagem das tipologias de minério fosfatados da mina F4. *HOLOS*, 1, 394–403. <https://doi.org/10.15628/holos.2016.3870>.

BROD, J.A., RIBEIRO, C.C., GASPAR, J.C., BROD, T.C., BARBOSA, E.S.R., RIFFEL, B.F., SILVA, J.F., CHABAN, N., FERRARI, A. J. D., 2004. Excursão Geologia e mineralizações dos complexos alcalino-carbonatíticos Congresso Brasileiro de Geologia, XLII, Araxá-MG, pp. 129.

CASTRO, S. A. B. D.; SILVA, A. C., 2024. Evaluation of PCA with variable selection for cluster typological domains. *REM-International Engineering Journal*, 77(2). <https://doi.org/10.1590/0370-44672023770071>.

CASTRO, S. A. B., SILVA, E., MATTIOLI, L.; SILVA, A., 2024. Use of the fuzzy algorithm in samples with imprecise descriptions to define typological domains. *Concilium*, 24. <https://doi.org/10.53660/CLM-3723-23P20>.

CERVO, V. L.; ANZANELLO, M. J., 2015. Seleção de variáveis para clusterização de bateladas produtivas através de PCA e remapeamento kernel. *Production*, v. 25, n. 4, p. 826-833 2015. <https://doi.org/10.1590/0103-6513.14361>.

DEMICCO, R. V, KLIR, G. J., 2003. *Fuzzy Logic in Geology*. Academic Press. An imprint of Elsevier Science. 352P.

DEUTSCH, C. V., 2013. Geostatistical modelling of geometallurgical variables – problems and solutions. In: *Proceedings of the International Geometallurgy Conference*, Brisbane, Australia, v. 30.

GIBSON, S.A., THOMPSON R.N., DICKIN A.P., LEONARDOS O.H., 1995a. High-Ti and low-Ti mafic potassic magmas: Key to plume-lithosphere interactions and continental flood-basalt genesis. *Earth Planet Sci. Lett.* 136: 149-165.

JOLLIFFE, I. T. Discarding variables in a principal component analysis - I: artificial data. *Journal of the Royal Statistical Society*, n. 2, p.160-173, 1972.

LAMBERG, P., 2011. Particles – the bridge between geology and metallurgy. In: *Conference in Mineral Engineering*, 2011, Luleå. Proceeding, Luleå, p. 1-16.

- LISHCHUK, V.; LUND, C.; GHORBANI, Y., 2019. Evaluation and comparison of different machine-learning methods to integrate sparse process data into a spatial model in geometallurgy. *Minerals Engineering*, 134, 156-165. <https://doi.org/10.1016/j.mineng.2019.01.032>
- LOZANO, C.; BENNET, C., 2003. Geometallurgical modeling applied to production forecasting, plant design and optimisation. Universidad de Chile – Departamento Ingeniería de Minas Taller PROCEMIN, 11p.
- MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G., 2012. Introduction to linear regression analysis. [S.l.: s.n.], v. 821. 645 p.
- MOREIRA, F. R. D. S., ALMEIDA FILHO, R. D., & CÂMARA, G., 2002. Aplicação da abordagem de Importação Semântica (IS) para caracterização de contatos litológicos em modelos de inferência espacial. *Rem: Revista Escola de Minas*, 55, 301-306.
- NIQUINI, F. G. F.; BRANCHES, A. M. B.; COSTA, J. F. C. L.; MOREIRA, G. D. C.; SCHNEIDER, C. L.; ARAÚJO, F. C. D.; CAPPONI, L. N., 2023. Recursive Feature Elimination and Neural Networks Applied to the Forecast of Mass and Metallurgical Recoveries in A Brazilian Phosphate Mine. *Minerals*, 13, 748. <https://doi.org/10.3390/min13060748>.
- OLIEIRA, S. M. B.; IMBERNON, R. A. L., 1998. Weathering alteration and related REE concentration in the Catalão I carbonatite complex, central Brazil. *Journal of South American Earth Sciences*, 11(4), 379-388.
- OROZCO-DEL-CASTILLO, M. G; ORTIZ-ALEMÁN, C.; URRUTIA-FUCUGAUCHI, J.; RODRÍGUEZ-CASTELLANOS, A., 2011. Fuzzy logic and image processing techniques for the interpretation of seismic data. *Journal of Geophysics and Engineering*, 8(2), 185-194. <https://doi.org/10.1088/1742-2132/8/2/006>
- RIBEIRO, C. C., 2008. Geologia, geometalurgia, controles e gênese dos depósitos de fósforo, terras raras e titânio do complexo carbonatítico Catalão I, GO. 508 f. Tese de doutorado em Geologia-Universidade de Brasília, Brasília
- SCHOLKOPF, B; SMOLA, A. J., 2002. *Learning With Kernels: Support Vector Machines Regularization Optimization and Beyond*, Cambridge, MA, USA: MIT Press.
- WILLIAMS, S. R. E RICHARDSON, J. M., 2004. Geometallurgical Mapping: A New Approach That Reduces Technical Risk. *Proceedings 36th Annual Meeting of the Canadian Mineral Processors*, Paper 16, 28p.