

Received:
September 30, 2020

Accepted:
October 12, 2020

Published:
October 31, 2020

A comparative analysis between the Geostatistics and Machine Learning methods for mineral resource estimation

Ricardo Alves da Silva¹ 

¹ Universidade Federal de Pernambuco, Recife, Brasil.

Email address

ricardoalvesdasilva@yahoo.com.br (Ricardo A. Silva)

Abstract

It is presented in this article two methods for estimating minerals resources. The first one is well known in the literature of mining engineering, it is called Ordinary Kriging, which is one of most used algorithms in Geostatistics to deal with regionalized variables. The second one is not yet so popular to be used for estimating minerals resources, although its applicability to be very effective for this task. This method is called Random Forests, which is an algorithm of Machine Learning. Thus, its analysis will focus on the construction procedure of block models of a copper mineral deposit using both methodologies.

Keywords: Ordinary Kriging, Geostatistics, Random Forests, Machine Learning.

1. Introdução

A estimativa de recursos minerais é uma etapa muito importante para o planejamento de uma mina, tendo em vista que o inventário mineral será dado em função da quantidade e qualidade do minério avaliado. Dentro desse contexto percebe-se que uma previsão não condizente com a realidade existente no subsolo, pode acarretar em sérios prejuízos ou até mesmo impossibilitando a continuidade dos trabalhos de lavra e beneficiamento do minério.

De acordo com Curi (2014), os métodos existentes para a avaliação de recursos minerais são: Os métodos clássicos, estatísticos e geoestatísticos. Dentre os métodos citados, os mais aplicados são aqueles provenientes da Geoestatística, cujo estimador de maior sucesso, Segundo Yamamoto e Landim (2015, p.73), é o de Krigagem Ordinária.

Dado o significativo avanço nas metodologias computacionais para a análise de dados, e em particular dos métodos de aprendizagem de máquina, que vem ganhando bastante espaço em todos os seguimentos acadêmicos, propõe-se neste trabalho realizar uma análise comparativa entre os preditores de krigagem Ordinária e Floresta Aleatória – *Random Forest* – sendo este último um estimador da categoria dos *Ensembles*, que em

outras palavras pode ser visto como um algoritmo que consegue combinar diversos classificadores ou regressores ao mesmo tempo.

Portanto a finalidade desse estudo é mostrar a importância dos métodos de Aprendizagem de Máquina para o setor mineral, onde neste caso particular a aplicação se dará para a estimativa dos teores de uma jazida mineral de cobre.

1.1. Banco de Dados

Os bancos de dados, proveniente das campanhas de sondagem, geralmente, são dados sigilosos das empresas, os quais, na maioria das vezes, só são disponibilizados ao público após o término da lavra na região onde foi realizado os furos de sondagem. Portanto, partindo do princípio da confidencialidade dos dados, e tendo como meta mostrar com o máximo de fidelidade os resultados; observou-se para esta análise um banco de dados que faz parte do livro *Practical Geostatistics 2000*.

A matriz de dados, analisada neste trabalho, possuem, além das especificações de controle de cada furo, as seguintes variáveis: $X_{co-ordinate}$, $Y_{co-ordinate}$, $Z_{co-ordinate}$, $Cu\%$ e $Mo\%$, onde as três primeiras representam os eixos cartesianos e as duas últimas o teor, em percentual, de cobre e de

molibdênio, respectivamente. Informações mais específicas sobre esse banco de dados encontram-se no seguinte endereço:

“http://www.kriging.com/books/Chapter1_PG2000_CS.pdf”.

Portanto, dado a matriz de dados, objeto de análise desse trabalho, será avaliada a distribuição de teores de minério de cobre dentro do volume mineralizado, tendo como principal objetivo medir a acurácia das duas metodologias aplicadas à predição dos teores.

A seguir é apresentado uma tabela com as dimensões, em metros, dos eixos cartesianos, seguido do mapa de localização dos furos de sondagem gerado por meio do software, de código aberto, *SGeMS*:

Tabela 1 – Domínio cartesiano do corpo de minério.

$X_{m\acute{a}x}$	4301	$Y_{m\acute{a}x}$	5400	$Z_{m\acute{a}x}$	5467
$X_{m\grave{m}n}$	1899	$Y_{m\grave{m}n}$	2117	$Z_{m\grave{m}n}$	4330
ΔX	2402	ΔY	3283	ΔZ	1137

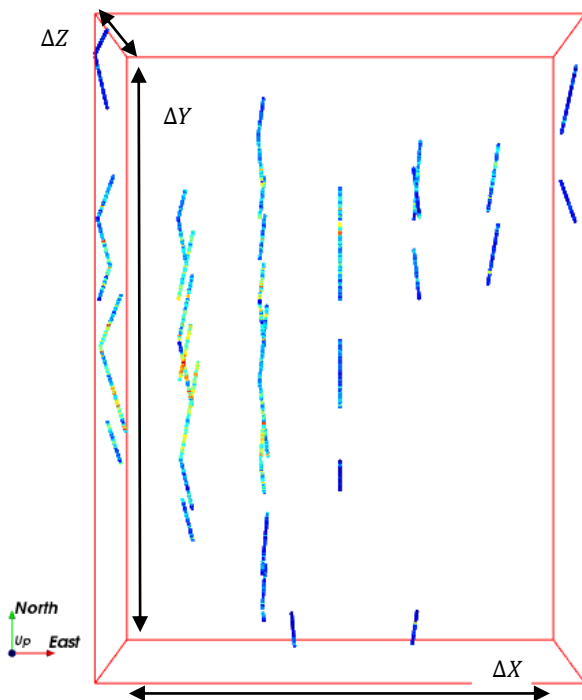


Figura 1 – Mapa de localização dos furos de sondagem da variável cobre.

As linhas oblíquas representam os furos de sondagem, as quais variam de acordo com a intensidade das cores, estas obedecem a uma ordem crescente com o grau de teor, que varia do azul para o vermelho, conforme ilustra a Figura (1).

Os principais dados descritivos do banco de dados para a variável *Cu*, aproximados por três casas decimais, são mostrados a seguir:

Tabela 2 – Principais valores descritivos dos dados de *Cu*.

Amostra de <i>Cu</i>	Valor
Média μ	0.664
Mediana M_d	0.650
Moda M_o	0.550
Variância σ^2	0.008
Desv. Padrão σ	0.087

1.1.1. Conceitos Fundamentais

Nesta subseção o objetivo é apresentar os principais conceitos pertinentes às disciplinas de Geoestatística e Aprendizagem de Máquina, para o melhor entendimento da base teórica aplicada neste trabalho.

- Variável Regionalizada

O elemento básico da Geoestatística é a variável regionalizada, cuja variação espacial caracteriza o fenômeno regionalizado que a originou. Estas variáveis possuem as componentes casuais e estruturais, ou seja, podem assumir localmente qualquer valor segundo uma função de probabilidade e globalmente possuem uma estruturação que pode ser tratada por uma função espacial, Journel e Huijbregts, (1978).

- Variograma

De acordo com Yamamoto e Landim (2015, p.41), a função variograma mede a variância entre pontos separados por uma distância *h*, ou seja, ele mede a esperança matemática do quadrado da diferença entre os valores de pontos no espaço, separados por uma distância *h*.

- Krigagem

A krigagem é um processo geoestatístico de estimativa de valores de variáveis distribuídas no espaço e/ou no tempo, com base em valores adjacentes, quando considerados interdependentes pela análise variográfica, Yamamoto (2015, p.55).

Uma vez apresentados, de forma bastante objetiva, os entes que compõem o elenco teórico geoestatístico, dar-se-á início as definições que integram o corpo teórico da Aprendizagem de Máquina.

- Atributo

São as variáveis independentes do banco de dados em análise. Neste caso específico as variáveis independentes são as componentes espaciais: *X*_co-ordinate, *Y*_co-ordinate, *Z*_co-ordinate.

- Rótulos

Representam as variáveis dependentes. Que neste caso, em particular, é simbolizada pelo teor de cobre, Cu%.

- Aprendizagem de Máquina

De acordo com Arthur Samuel (1959), o aprendizado de máquina é o “campo de estudo que dá aos computadores a habilidade de aprender sem serem explicitamente programados”. Em outras palavras isso quer dizer que não se faz necessário implementar todas as ações de resposta para um código computacional.

- Aprendizagem Supervisionada

O supervisionamento é feito quando a partir de um conjunto de dados rotulados, previamente definido, deseja-se encontrar uma função que seja capaz de prever rótulos desconhecidos. Existem outros tipos de aprendizagem tais como: aprendizagem semi-supervisionada, aprendizagem não supervisionada e aprendizagem por reforço.

Este trabalho, em particular, está fundamentado no treinamento supervisionado, uma vez que tem-se o conjunto de dados composto de atributos e rótulos.

2. Metodologia

Em uma análise geoestatística, faz-se necessário em primeiro lugar, avaliar o comportamento da distribuição de probabilidades da variável regionalizada, e isto é equacionado por meio do histograma, o qual é mostrado a seguir:

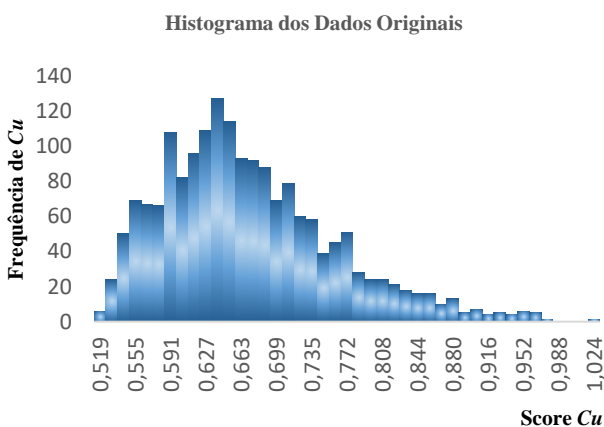


Figura 2 – Distribuição de Frequência dos dados originais de Cu.

É possível observar uma assimetria positiva, o que caracteriza uma distribuição lognormal. Para

quantificar essa assimetria aplica-se o coeficiente de assimetria de Pearson, o qual é dado pela equação a seguir:

$$A = \frac{3(\mu - M_d)}{\sigma} = \frac{3(0,664 - 0,650)}{0,087} = 0,483 \quad (1)$$

As três classificações para o coeficiente de assimetria, são as seguintes:

$$0 < |A| < 0,5 \Rightarrow \text{Assimetria muito fraca}$$

$$0,5 < |A| < 1,5 \Rightarrow \text{Assimetria fraca}$$

$$|A| > 1,5 \Rightarrow \text{Assimetria forte}$$

Como o valor obtido por meio do coeficiente de assimetria de Pearson está entre 0 e 0,5, conclui-se que a influência dos valores mais altos da variável Cu é muito fraca, ou seja, esses valores mais altos não exercem influência significativa nos valores menores, isso possibilita a interpretação dos dados por meio de uma distribuição normal, a qual é mostrada, juntamente com seu histograma, a seguir:

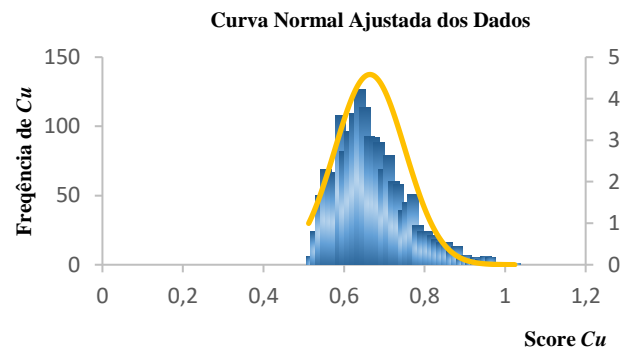


Figura 3 – Curva normal ajustada a distribuição de frequência dos dados de Cu.

Portanto, a variável em análise, que neste caso representa o teor de cobre, é dada por uma distribuição gaussiana de probabilidade do tipo: $Cu \sim N(0,664; 0,008)$.

2.1. Método Geoestatístico

Uma vez determinado o tipo de distribuição de probabilidade que governa os dados, aplica-se o estimador geoestatístico mais indicado para a distribuição. Neste caso específico, o estimador será o de krigagem ordinária, visto que não houve necessidade de transformação dos dados para se

obter uma curva gaussiana. Caso houvesse a necessidade de transformação dos dados, ela seria dada pela aplicação da transformada logarítmica e consequentemente o estimador seria o de krigagem lognormal, que é um caso particular de krigagem ordinária.

Para o cálculo da estimativa por krigagem é necessário obter o modelo de continuidade espacial, da variável regionalizada, o qual é obtido por meio da análise variográfica.

Essa análise é composta de três fases: Elaboração do variograma experimental, avaliação da anisotropia e criação do modelo teórico que governa a variabilidade espacial da jazida.

- Variograma

O cálculo do variograma experimental do corpo mineralizado, que neste caso é o minério de cobre, é realizado por meio da função semivariograma. Sua expressão matemática é definida, de como a seguir:

$$\gamma_{x_i+h, x_i} = \frac{1}{2n} \sum_{i=1}^n [z(x_{i+h}) - z(x_i)]^2 \quad (2)$$

Onde:

- $z(x_{i+h}), z(x_i)$ = São os valores de Cu nas posições espaciais x_{i+h} e x_i , respectivamente.
- n = Número de pares de pontos obtidos para cada distância h (lag) avaliada.

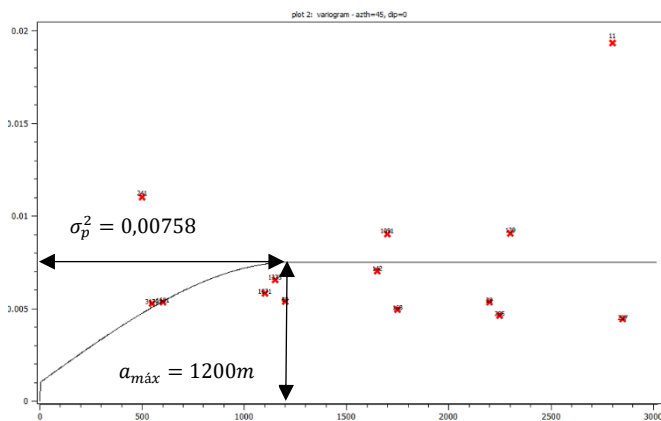


Figura 4 – Semivariograma para a direção de 45°; 0°.

Uma vez encontrado o variograma experimental que melhor representa a estrutura mineralizada, ajusta-se um modelo teórico que será utilizado para realizar a interpolação por meio do método da krigagem. Isso é mostrado a seguir:

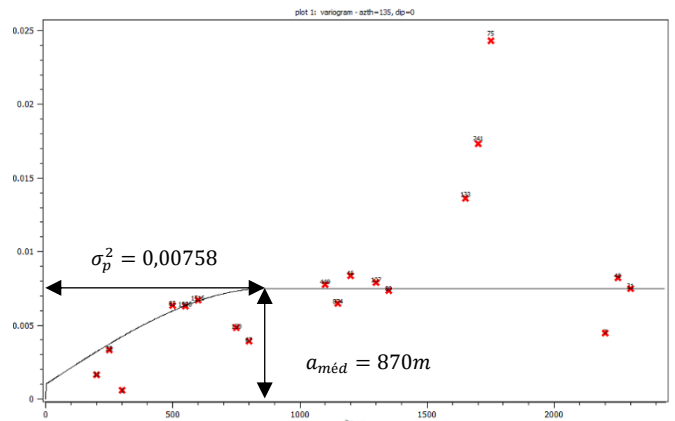


Figura 5 – Semivariograma para a direção de 135°; 0°.

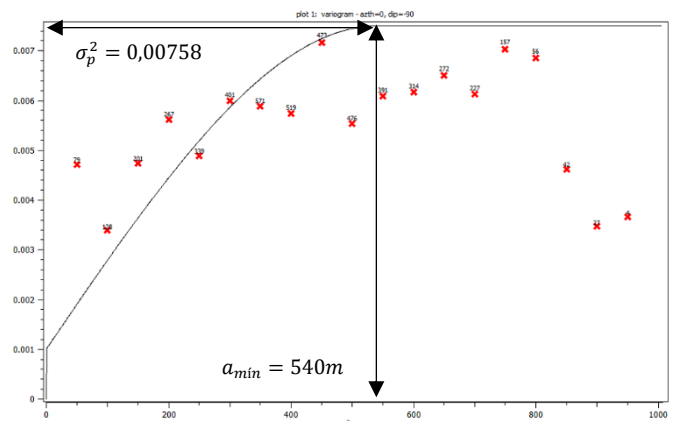


Figura 5 – Semivariograma para a direção vertical 0°; -90°.

De acordo com os resultados obtidos nessa análise variográfica, observa-se que existe uma anisotropia geométrica, o que caracteriza uma direção de maior, intermediária e menor continuidades nas direções 45°, 135° e 0°. Dentre as várias combinações de parâmetros para encontrar os variogramas experimentais, chega-se àquela que melhor representou o fenômeno estruturado:

Tabela 3 – Parâmetros utilizados para a construção dos variogramas experimentais de Cu .

Az	Dip	Tol	Bw	N° Lg	Lg	LgTol
45	0	12,5	50	60	50	25
135	0	12,5	50	60	50	25
0	-90	12,5	50	60	50	25

Onde:

- Az: Azimute, direção de maior continuidade;
- Dip: Inclinação em relação a vertical;
- Tol: Tolerância angular;
- Bw: Largura de banda;
- N° Lg: Quantidade de passos que o vetor h terá que caminhar em uma determinada direção;
- Lg: Quantidade de passos;
- LgTol: Tolerância ao passo.

O modelo teórico ajustado aos variogramas experimentais é o gaussiano, cujas expressões analíticas, para as três direções são mostradas a seguir:

- Variograma esférico para a direção de 45°; 0°:

$$\gamma(h) = 0,001 + 0,00658 \left[1,5 \frac{h}{1200} - 0,5 \left(\frac{h}{1200} \right)^3 \right] \quad (3)$$

$$\gamma(h) = 0,00758, \forall h \geq 1200$$

- Variograma esférico para a direção de 135°; 0°:

$$\gamma(h) = 0,001 + 0,00658 \left[1,5 \frac{h}{870} - 0,5 \left(\frac{h}{870} \right)^3 \right] \quad (4)$$

$$\gamma(h) = 0,00758, \forall h \geq 870$$

- Variograma esférico para a direção de 0°; -90°:

$$\gamma(h) = 0,001 + 0,00658 \left[1,5 \frac{h}{540} - 0,5 \left(\frac{h}{540} \right)^3 \right] \quad (5)$$

$$\gamma(h) = 0,00758, \forall h \geq 540$$

O efeito pepita e o patamar são, respectivamente: $C_0 = 0,001$ e $C = 0,00758$, este corresponde a variância a priori dos dados. O efeito pepita foi obtido considerando-se o comportamento linear próximo a origem. A base para essa consideração está fundamentada na escolha do variograma teórico, segundo Issaks e Srivastava (1989).

Dado que as variações mais significativas ocorreram em relação as amplitudes, $a_{m\acute{a}x} < a_{m\acute{e}d} < a_{m\acute{i}n}$, conclui-se que essa jazida possui uma anisotropia geométrica.

O modelo geométrico ou elipsoide de busca, que representa as três direções de continuidades do teor Cu , é mostrado na figura a seguir:

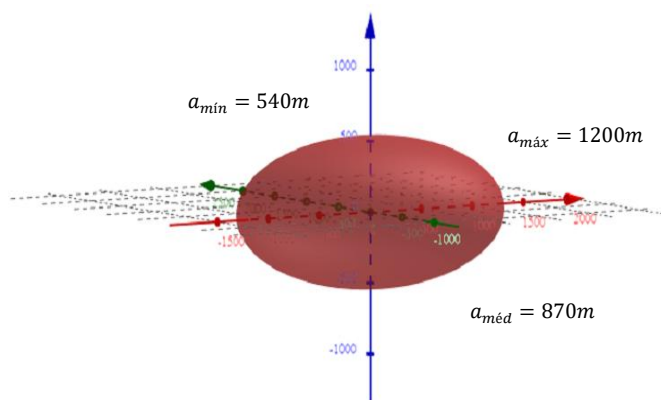


Figura 6 – Elipsoide de busca para a variável Cu .

O elipsoide de busca é o elemento mais importante em uma estimativa por krigagem, uma vez que, todos os pontos cujos valores do teor de

cobre é desconhecido, serão estimados ou interpolados mediante os valores de contorno estabelecidos pelo elipsoide.

O modelo analítico que possibilita calcular, pelo método da krigagem qualquer ponto dentro do supracitado elipsoide, dado a latitude λ , longitude β e o vetor h que separa cada par de pontos, é dado pela seguinte equação:

$$\gamma(h) = 0,001 + 0,00658 \left[1,5 \frac{|h|}{a(\lambda, \beta)} - 0,5 \left(\frac{|h|}{a(\lambda, \beta)} \right)^3 \right]$$

$$\gamma(h) = 0,00758, \forall |h| \geq 1200 \quad (6)$$

Onde:

$$a(\lambda, \beta) = \text{sqrt}[(1200 \cos(\beta) \cos(\lambda))^2 + (870 \cos(\beta) \sin(\lambda))^2 + (540 \sin(\beta))^2]$$

• Krigagem

A krigagem ordinária é um método de estimativa que está fundamentado no conceito de combinação linear das variáveis regionalizadas com os seus respectivos ponderadores. E isto é feito mediante o cálculo, dado pela seguinte equação:

$$Z_{ko}^*(x_0) = \sum_{i=1}^n \lambda_i z(x_i) \quad (7)$$

Onde:

$Z_{ko}^*(x_0)$ = Valor da variável regionalizada no ponto desconhecido x_0 ;

λ_i = Valor dos i -ésimos ponderadores ótimos;

$z(x_i)$ = Valor da variável regionalizada nos i -ésimos pontos conhecidos.

O cálculo dos ponderadores ótimos é realizado mediante a solução da matriz de covariância, a qual pode ser escrita, em termos da função semivariograma de acordo com Goovaerts (1997), como a seguir:

$$\begin{bmatrix} \gamma(x_1 - x_1) & \gamma(x_1 - x_2) & \dots & \gamma(x_1 - x_n) & 1 \\ \gamma(x_2 - x_1) & \gamma(x_2 - x_2) & \dots & \gamma(x_2 - x_n) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma(x_n - x_1) & \gamma(x_n - x_2) & \dots & \gamma(x_n - x_n) & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma(x_0 - x_1) \\ \gamma(x_0 - x_2) \\ \vdots \\ \gamma(x_0 - x_n) \\ 1 \end{bmatrix} \quad (8)$$

Portanto, uma vez conhecido a função analítica de variabilidade espacial, que governa a jazida mineral, que neste estudo é dada pela Equação (6), pode-se estimar qualquer ponto dentro

do volume limitado pelo corpo de minério, aplicando-se as Equações (7) e (8), respectivamente.

A estratégia de busca que resultou na variância mínima é mostrada na sequência abaixo:

1. Grid de Estimativa

O grid é composto de 1.946.208 blocos, com dimensão de 25x25x7,5 cada um, em metros. Esse grid possui a forma de um paralelepípedo retângulo.

2. Discretização dos Blocos

Cada um dos blocos foi subdividido em 32 blocos de dimensões 6,25x6,25x3,75 cada um, em metros. Esse procedimento é realizado em krigagem para garantir um melhor resultado para cada bloco estimado.

3. Condições de Contorno do Elipsoide de Busca

- Quantidade mínima e máxima de amostras dentro do elipsoide: 4 e 24 respectivamente.
- Alcance máximo, intermediário e mínimo: 1200, 870 e 540, em metros, respectivamente.
- Posicionamento do elipsoide dentro da jazida mineral: 45° azimuth, 0° mergulho e 0° rotação em torno do eixo maior (*Rake*).
- Origem do elipsoide de busca: 1900, 2160 e 4340, em metros, respectivamente.
- Fracionamento dos octantes do elipsoide: 5 octantes não vazios, pelo menos 1 amostra em cada octante e no máximo 3 amostras em cada um dos octantes.

Cada um dos blocos definido nessa estratégia de busca atendeu as condições de contorno estabelecidas desse elipsoide de busca. Portanto, através desse procedimento encontra-se o modelo de blocos para essa jazida de Cu.

2.2. Método de Aprendizagem de Máquina

A análise feita pelo método de aprendizagem de máquina consiste, primeiramente, em organizar ou processar o banco de dados para, em seguida, aplicar o algoritmo que irá treinar os dados. Esse pré-processamento, também conhecido como análise de dados, não se fará necessário, visto que o

banco de dados precisa ser exatamente o mesmo para evitar distorções na comparação de performance dos dois métodos aplicados na estimativa do minério de cobre.

O algoritmo utilizado para estimar o teor de minério de cobre foi o de Florestas Aleatórias – *Random Forest*. A linguagem de programação utilizada para implementação do modelo utilizado na estimativa de minério de cobre foi Python 3, e a plataforma utilizada para desenvolver o código de programação foi o *Jupyter Notebook*. O conteúdo matemático desse algoritmo pode ser encontrado, de forma detalhada, no livro: *The Elements of Statistical Learning*.

• Módulo 1

```
import pandas as pd
import numpy as np
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import train_test_split
from sklearn.model_selection import KFold
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import GridSearchCV
from sklearn.ensemble import RandomForestRegressor
import math
```

O código implementado acima mostra as bibliotecas do Python utilizadas para resolver o problema de estimativa do minério de cobre – Cu.

• Módulo 2

```
dataset = pd.read_csv('BM.csv')
print(dataset.shape)
print(dataset.dtypes)
```

O módulo (2) representa apenas a leitura do banco de dados, ou seja, os dados de sondagem do Cu.

• Módulo 3

```
array = dataset.values
X = array[:,0:3]
Y = array[:,3]
validation_size = 0.20
seed = 7
num_folds = 10
X_train, X_validation, Y_train, Y_validation =
train_test_split(X, Y, test_size=validation_size,
random_state=seed)
```

O código acima representa os vetores de atributo, X, e rótulo, Y, respectivamente, como também os conjuntos de treinamento e validação.

• Módulo 4

```

scaler = StandardScaler().fit(X_train)
rescaledX = scaler.transform(X_train)

n_estimators = [50,60,70,80,90,100,110]
min_samples_split = [2,3,4,5,6,7,8]
min_samples_leaf = [1,2,3,4,5,6]
max_features = ['auto','sqrt','log2']

param_grid = dict(n_estimators=n_estimators,
                  min_samples_split=min_samples_split,
                  min_samples_leaf=min_samples_leaf,
                  max_features=max_features)

model = RandomForestRegressor()
kfold = KFold(n_splits=num_folds,
             random_state=seed)
grid = GridSearchCV(estimator=model,
                   param_grid=param_grid, scoring=scoring, cv=kfold)
grid_result = grid.fit(rescaledX, Y_train)

print('Best: %f using %s' % (grid_result.best_score_,
                             grid_result.best_params_))
means = grid_result.cv_results_['mean_test_score']
stds = grid_result.cv_results_['std_test_score']
params = grid_result.cv_results_['params']
for mean, stdev, param in zip(means, stds, params):
    print('%f (%f) with: %r' % (mean, stdev, param))
    
```

• Módulo 5

```

from sklearn.metrics import mean_squared_error
scaler = StandardScaler().fit(X_train)
rescaledX = scaler.transform(X_train)
model =
RandomForestRegressor(max_features='log2',
min_samples_leaf=2, min_samples_split=3,
n_estimators=90)
model.fit(rescaledX, Y_train)
    
```

• Módulo 6

```

rescaledValidationX =
scaler.transform(X_validation)
predictions = model.predict(rescaledValidationX)
print(mean_squared_error(Y_validation,
predictions))
    
```

O quarto módulo representa o treinamento do algoritmo Florestas Aleatórias (*Random Forest*). O quinto módulo representa o modelo finalizado, ou seja, a criação da melhor máquina de aprendizagem para a jazida de Cu analisada. Já o sexto módulo é apenas a aplicação dessa máquina aos dados de validação, e com isto comparar o Erro Médio Quadrático – *Mean Squared Error* (MSE) entre os

dois métodos aplicados para a estimativa do teor de minério de cobre.

A explicação detalhada de linha por linha de cada um dos módulos utilizados para implementar o modelo fugiria do objetivo desse artigo, uma vez que, este visa mostrar, de forma objetiva, a utilização de ferramentas de aprendizagem de máquina para a estimativa de recursos minerais. Mas desde já me coloco a disposição para quaisquer questionamentos.

3. Resultados e Discussão

O resultado do MSE para as duas metodologias aplicadas é mostrado na tabela a seguir:

Tabela 4 – Cálculo de MSE para o conjunto de validação de Cu.

	MSE
Geoestatística	0.003747
Aprendizagem de Máquina	0.002798

Portanto, o método que apresentou o menor MSE para os dados de validação de cobre, foi o de aprendizagem de máquina, 0.002798, este representa uma diminuição da ordem de 25,33% no erro de estimativa para a jazida em análise.

A seguir é mostrado os modelos de blocos gerados a partir das duas metodologias aplicadas:

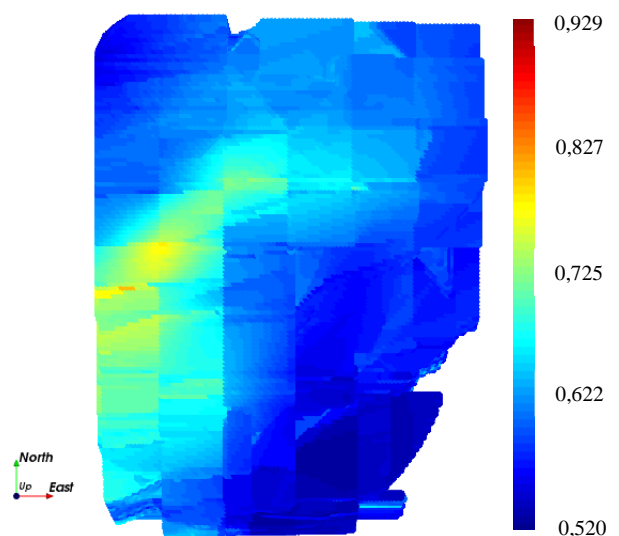


Figura 7 – Modelo de bloco da jazida de Cu gerado pelo método geoestatístico – Krigagem Ordinária.

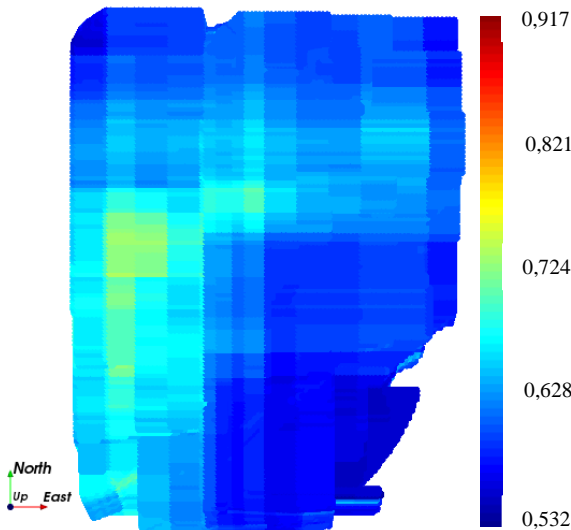


Figura 8 – Modelo de bloco da jazida de Cu gerado pelo método de Aprendizagem de Máquina – Florestas Aleatórias.

Cada um dos modelos contém um total de 1.402.944 blocos de dimensões 25x25x7,5 o que dá aproximadamente 6,5 bilhões de metros cúbicos de minério de cobre. Observa-se bastante semelhança nas escalas de teores de ambos os modelos, ou seja, as duas metodologias aplicadas conseguiram gerar regiões de estimativas consonantes com a realidade da jazida mineral, isso também pode ser observado por meio do mapa de localização, Figura (1), onde estão os dados originais, pois, observa-se através desse mapa que as regiões com alto e baixo teores de cobre foram identificadas corretamente.

Para quantificar a precisão da estimativa por pode-se observar a variância do modelo gerado. Portanto a tabela a seguir mostra os valores encontrados para cada um dos modelos:

Tabela 5 – Cálculo das variâncias dos dados estimados pelos métodos de krigagem ordinária e florestas aleatórias.

	δ^2
Krigagem Ordinária	0.002923
Florestas Aleatórias	0.002798

Através desses resultados observa-se um aspecto muito importante, no que concerne a capacidade de generalização dos algoritmos aplicados, pois, quanto menor a variância dos dados estimado mais acurácia possui o modelo que gerou os dados. Dentro desse contexto foi constatado que o método de florestas aleatórias apresentou uma precisão 4,3% melhor que o de krigagem ordinária.

De uma forma geral os dois métodos conseguiram generalizar de forma satisfatória os teores de cobre para a jazida analisada. Contudo, verifica-se, por meio da Tabela 4, que o método de aprendizagem de máquina apresentou viés sistemático menor que o método geoestatístico, isso se deve, basicamente, a quantidade de árvores que o algoritmo cria para conseguir generalizar bem o modelo sem aumentar o viés sistemático. Nesse modelo, o número de árvores que representa a melhor estimativa para os dados é 90, dado no Modulo 5 pelo termo: `n_estimators=90`. Portanto, isso possibilita uma ótima capacidade de generalização sem gerar um alto viés sistemático.

4. Conclusão

A técnica de aprendizagem de máquina, de acordo com os resultados obtidos, atende aos critérios de qualidade no estudo de estimativas de recursos minerais. Isso foi mostrado por meio das métricas MSE e δ^2 aplicadas aos dados de validação e os valores estimados nos respectivos modelos de blocos. Portanto, dentro do contexto da estimativa de recursos minerais o algoritmo de Florestas Aleatórias, que está classificado na categoria dos Ensembles, apresenta-se com uma alternativa bastante satisfatória para a análise de teor em jazidas minerais, isso vem para corroborar em todo o processo produtivo da mineração, desde a campanha de sondagem da área que será lavrada até o beneficiamento do minério. Logo, as técnicas de aprendizagem de máquina estarão cada vez mais presente para ajudar nos trabalhos da mineração.

Referências

- Curi, A., 2014. *Minas a céu aberto: planejamento de lavra*. Oficina de Textos, pp.87-104.
- Goovaerts, P., 1997. *Geostatistics for natural resources evaluation*. Oxford University Press on Demand, pp.131-137.
- Hastie, T., Tibshirani, R., and Friedman, J., 2009. *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media, pp.587-603. <https://doi.org/10.1007/978-0-387-84858-7>
- Houlding, S., 2000. *Practical geostatistics: modeling and spatial analysis. Manual*. Springer Science & Business Media.

Isaaks, E.H., and Srivastava, M.R., 1989. *Applied geostatistics* (No. 551.72 ISA), pp.370-397.

Journel, A.G., and Huijbregts, C.J., 1978. *Mining geostatistics*. London: Academic press.

Samuel, A.L., 1959. Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, 3(3), pp.210-229.
<https://doi.org/10.1147/rd.33.0210>

Yamamoto, J.K., and Landim, P.M.B., 2015. *Geoestatística: conceitos e aplicações*. Oficina de textos.