

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 (PPGE3M)

PREDIÇÃO SIMULTÂNEA DE PRODUTOS E REJEITOS EM PLANTAS DE
PROCESSAMENTO DE ZINCO E OURO A PARTIR DAS CARACTERÍSTICAS DO
MINÉRIO

FERNANDA GONTIJO FERNANDES NIQUINI

Mestre em Engenharia

Tese para obtenção do título de Doutora em Engenharia

PORTO ALEGRE, RS

2020

Fernanda Gontijo Fernandes Niquini

PREDIÇÃO SIMULTÂNEA DE PRODUTOS E REJEITOS EM PLANTAS DE
PROCESSAMENTO DE ZINCO E OURO A PARTIR DAS CARACTERÍSTICAS DO
MINÉRIO

Tese submetida ao Programa de Pós-graduação em Engenharia de Minas, Metalúrgica e
Materiais (PPGE3M) da Universidade Federal do Rio Grande do Sul como requisito parcial à
obtenção do título de Doutor em Engenharia

Orientador: Prof. PhD. João Felipe Coimbra Leite Costa

Porto Alegre, RS

2020

Fernanda Gontijo Fernandes Niquini

PREDIÇÃO SIMULTÂNEA DE PRODUTOS E REJEITOS EM PLANTAS DE
PROCESSAMENTO DE ZINCO E OURO A PARTIR DAS CARACTERÍSTICAS DO
MINÉRIO

Esta tese foi julgada adequada para a obtenção do título de Doutora em Engenharia, na área de Tecnologia Mineral e Metalurgia Extrativa e aprovada em sua forma final pelo Orientador e pela Banca Examinadora designada pelo Programa de Pós-Graduação em Engenharia de Minas, Metalúrgica e Materiais (PPGE3M) da Universidade Federal do Rio Grande do Sul

Prof. Dr. João Felipe Coimbra Leite Costa

PhD. pela University of Queensland

Orientador

Prof. Dr. Carlos Pérez Bergmann

Dr. pela Rheinisch Westfälische Technische Hochschule Aachen

Coordenador do PPGE3M

BANCA EXAMINADORA

Dr. Áttila Leães Rodrigues – UFRGS _____

Dr. Luciano Nunes Capponi – Mosaic Fertilizantes _____

Dra. Vanessa Cerqueira Koppe – UFRGS _____

Dedico esta vitória à minha filha, Maria Luíza,
luz dos meus dias e alegria da minha vida,
e ao bebê que carrego em meu ventre,
cujo rosto ainda não conheço, mas já amo de todo coração.

AGRADECIMENTOS

Agradeço primeiramente à Deus, por me conceder força e sabedoria para seguir em frente e lutar pelos meus sonhos, e à Nossa Senhora Aparecida, pela intercessão em todos os momentos.

Ao meu orientador, Prof. João Felipe, exemplo de profissional e educador, capaz de ver além de um diploma e enxergar que a diversidade de conhecimentos pode levar qualquer equipe mais longe. Obrigada por sempre acreditar no meu potencial!

À Nexa Resources que forneceu os dados para esta pesquisa e aos engenheiros Jorge Lucas Bechir e Breno Valente, disponíveis para esclarecer dúvidas e auxiliar no que era necessário.

À mineradora de ouro que apoiou este trabalho, em especial à geóloga Mariana Gazire Lemos, que acreditou sempre no potencial deste estudo.

Ao CNPq e à Fundação Luiz Englert, que financiaram esta tese.

À minha filha, Maria Luíza, que veio ao mundo no meio do doutorado para me mostrar que o amor nos torna capaz de fazer o impossível.

Aos meus pais, Célio e Celeste, por cuidarem de minha filha em meus momentos de estudo e por serem para mim um porto seguro em meio às dificuldades. Sem vocês essa vitória não existiria.

Ao meu marido Thiago, pela compreensão nos momentos de ausência, por se dispor a me acompanhar no APCOM e por cuidar de nossa filha para que eu pudesse realizar meu sonho.

Aos colegas do LPM pelo constante apoio.

Às amigas que a mineração meu deu, Evangelina e Carolina, obrigada pelo incentivo em todos os momentos.

“Em Tuas mãos seguras minha vida guardarás...”

(Eliana Ribeiro)

RESUMO

A geometalurgia vem se mostrando uma forte aliada da indústria mineral, por ser uma disciplina capaz de aumentar a eficiência e os lucros da operação, minimizar a ocorrência de eventos indesejados no beneficiamento mineral e levar a um melhor aproveitamento do minério. Todos benefícios que impactam diretamente no lado financeiro das operações. Apesar de serem raramente elaborados, modelos geometalúrgicos que incorporam variáveis que impactam diretamente no meio ambiente geram conhecimento a respeito do impacto ambiental da operação mineira, em termos de consumo de água e energia e até mesmo no volume de rejeitos gerados e suas características químicas. Neste contexto, este estudo propõe desenvolver uma metodologia capaz de elaborar, com dados já existentes, um modelo geometalúrgico que contemple variáveis com impacto financeiro e ambiental, visando fornecer insumos para que o planejamento de mina considere ambos fatores em seus estudos. Para atingir este objetivo, as recuperações mássica e metalúrgica de todos os pontos de saída de massa e metal no beneficiamento foram modeladas, sendo elas relacionadas à produtos vendáveis ou rejeitos. Ao final do estudo, quando todas as previsões relacionadas à recuperação mássica forem somadas, o fechamento mássico deve ser honrado. O mesmo deve ocorrer com o fechamento metalúrgico ao somar todas as previsões das variáveis de recuperação metalúrgica. Tendo em vista a necessidade do fechamento, optou-se por estudar uma técnica capaz de prever todas as variáveis simultaneamente, o que possivelmente levaria à melhores resultados. Identificou-se a técnica de redes neurais artificiais como capaz de atender à tal requisito e os resultados gerados através desta técnica foram confrontados com os obtidos por outras três técnicas de estatística multivariada e *machine learning*. Ao avaliar as previsões, observou-se que a técnica de redes neurais apresentava resultados superiores às demais, mostrando-se a mais adequada para o propósito do estudo. Finalizada a elaboração do modelo, o passo seguinte foi reconciliar suas previsões com as respostas obtidas pela usina de processamento em uma operação industrial, visando comprovar sua eficiência. Para isso, 24 pilhas de minério da mina de zinco de Vazante tiveram sua passagem pela usina monitorada e as previsões feitas pelo modelo foram confrontadas com as respostas reais da planta. Os resultados mostraram-se excelentes para as variáveis de maior impacto financeiro e ambiental na operação, provando que a metodologia elaborada é adequada e pode ser utilizada na prática.

Palavras-chave: Geometalurgia; Redes Neurais; Machine Learning; Recuperação Mássica; Recuperação Metalúrgica.

ABSTRACT

Geometallurgy has proved to be beneficial to mining chain helping to improve operation earnings, minimize the occurrence of undesirable events during mineral beneficiation and to lead to a better ore utilization and recovery. All benefits described above impacts directly on the operation cash flow. Although rarely elaborated, geometallurgical models incorporating variables that directly impacts on the environment bring knowledge about the operation environmental impact, in terms of water and energy consumption and even about the quantity of tailings generated and its chemical characteristics. The focus of this study is to propose a methodology able to create, using the data available, a geometallurgical model that includes variables with financial and environmental impact, aiming at providing the needed *inputs* for the mining planning to consider both factors while planning. To achieve this goal, the mass and metallurgical recoveries of all mass and metal *output* points were modelled, being them related to saleable products or tailings. At the end of the study, when all the forecasts related to the mass recovery were added, the mass closure needs to be honored. The same needs to happen with the metallurgical closure when adding all the forecasts related to the metallurgical recovery variables. Bearing in mind this closure need, it was decided to investigate a technique able to forecast all variables simultaneously, which would probably bring better results. The neural networks technique was identified as able to meet this requirement and the results generated through this technique were compared against other three multivariate statistical and machine learning techniques. By evaluating the predictions, it was observed that the neural networks presented superior results than the other techniques, proving itself more appropriated to the study purposes. Finalized the model elaboration, the next step was to reconcile its forecasts against the actual beneficiation plant responses, aiming at checking its operational efficiency. Thus, 24 ore piles from Vazante zinc mine were monitored during their passage through the plant and the model forecasts were compared against the actual results. The results were excellent for the variables with highest financial and environmental impact, demonstrating that the elaborated methodology is adequate and can be used in practice.

Keywords: Geometallurgy; Neural Networks; Machine Learning; Mass Recovery; Metallurgical Recovery.

LISTA DE FIGURAS

Figura 1 - Exemplo de rede com uma camada oculta.....	25
Figura 2 - Exemplo de rede com duas camadas ocultas.....	26
Figura 3 - Cálculos tradicionais de recuperação mássica considerando cada etapa do processamento mineral individualmente.	30
Figura 4 - Cálculos de recuperação para realização do balanço mássico da usina.....	31
Figura 5 - Quadro ilustrativo da conversão de recursos para reservas. Fonte: JORC (2012) ..	42
Figura 6 - Curva Geometalúrgica. Fonte: Vann <i>et al.</i> (2011).....	46
Figura 7 - Demonstração gráfica da Desigualdade de Jensen - Função Côncava $\ln(x)$	51
Figura 8 - Demonstração gráfica da Desigualdade de Jensen - Função Convexa x^2	52
Figura 9 - Verificação da suposição de variância constante.....	74
Figura 10 - Exemplo de árvore de decisão para classificação de litologias	77
Figura 11 - Representação do SVM para um conjunto de dados linearmente separáveis com duas categorias. Fonte: Hastie, Tibshirani e Friedman (2009).....	91
Figura 12 - Representação do SVM para um conjunto de dados não linearmente separáveis com duas categorias.....	94
Figura 13 - Exemplo de diferentes valores de C em um <i>support vector machine</i> para classificação.....	95
Figura 14 - Exemplo de valores dos parâmetros γ e C em um SVM de classificação .	100
Figura 15 - Support Vector Machine para regressão.....	102
Figura 16 - Ilustração de um neurônio humano.....	107
Figura 17 - Representação de um neurônio artificial	108
Figura 18 - Exemplo de rede com camada oculta	111
Figura 19 - Exemplo de RNA construída para prever variáveis geometalúrgicas	112
Figura 20 - Gráficos ilustrativos das funções de ativação	116
Figura 21 - Localização geográfica do município de Vazante (MG).....	142
Figura 22 - Fluxograma simplificado da usina de Vazante.....	143
Figura 23 - Histogramas dos teores do ROM – Mina de Zinco	145
Figura 24 - Box-plots dos teores de zinco e chumbo separados por alvo	146
Figura 25 - Histogramas das variáveis de processo – Mina de Zinco	147
Figura 26 - Box-plots das variáveis de Recuperação Mássica por alvo – Mina de Zinco.....	149
Figura 27 - Box-plots das variáveis de Recuperação Metalúrgica por alvo – Mina de Zinco	150

Figura 28 - Dados do modelo de regressão construído para Rec. Mássica do Conc. de Zn ..	153
Figura 29 - Dados do modelo de regressão construído para Rec. Mássica do Conc. de Pb...	154
Figura 30 - Dados do modelo de regressão construído para Rec. Mássica do Rejeito Final .	155
Figura 31 - Dados do modelo de regressão construído para Rec. Met. de Zn no Conc. de Zn	156
Figura 32 - Dados do modelo de regressão construído para Rec. Met. de Zn no Conc. de Pb	157
Figura 33 - Dados do modelo de regressão construído para Rec. Met. de Zn no Rej. Final..	158
Figura 34 - Histograma dos resíduos – Mina de Zinco	159
Figura 35 - Gráficos dos valores preditos x resíduos – Mina de Zinco.....	160
Figura 36 - Gráfico dos valores preditos x observados – Mina de Zinco.....	162
Figura 37 - Valores preditos x observados - Random Forest – Mina de Zinco.....	165
Figura 38 - Predito x Observado - <i>Support vector regression</i> linear – Mina de Zinco	168
Figura 39 - Predito x Observado - <i>Support vector regression</i> não linear – Mina de Zinco ...	170
Figura 40 - Rede Neural utilizada para prever as variáveis geometalúrgicas da mina de Vazante	172
Figura 41 - Predito x Observado - Redes Neurais Artificiais – Mina de Zinco	173
Figura 42 - Exemplo de funcionamento da técnica de validação cruzada com $k = 5$	174
Figura 43 - Fluxograma simplificado das usinas 1 e 2	194
Figura 44 - Box-plots dos teores de Au e S separados por mina.....	201
Figura 45 - Box-plots das variáveis metalúrgicas – Minas de Ouro	201
Figura 46 - Histograma da variável calcinação	203
Figura 47 - Gráfico de dispersão entre teor de enxofre no concentrado da flotação e a massa que sobrou após etapa de calcinação	203
Figura 48 - Histograma da calcinação com dados imputados	204
Figura 49 - Dados do modelo de regressão construído para RMAS Conc. Gravimétrico	205
Figura 50 - Dados do modelo de regressão construído para RMAS Rej. Flotação Final.....	206
Figura 51 - Dados do modelo de regressão construído para RMAS Rej. Lixiviação Final ...	207
Figura 52 - Dados do modelo de regressão construído para RMAS Gás Final.....	208
Figura 53 - Dados do modelo de regressão construído para RMET Conc. Gravimétrico.....	209
Figura 54 - Dados do modelo de regressão construído para RMET Rej. Flotação Final.....	210
Figura 55 - Dados do modelo de regressão construído para RMET Conc. Lixiviação Final.	210
Figura 56 - Dados do modelo de regressão construído para RMET Rej. Lixiviação Final ...	211
Figura 57 - Histograma dos resíduos – Minas de Ouro.....	212

Figura 58 - Gráficos dos valores preditos x resíduos – Minas de Ouro	213
Figura 59 - Gráfico dos valores preditos x observados – Análise de regressão - Minas de Ouro	214
Figura 60 - Valores preditos versus observados - Random Forest – Minas de Ouro	216
Figura 61 - Valores preditos versus observados - SVR linear – Minas de Ouro.....	218
Figura 62 - Valores preditos versus observados - SVR não linear - Minas de Ouro	220
Figura 63 - Configuração da rede neural construída para as minas de ouro	222
Figura 64 - Valores preditos versus observados - Redes Neurais - Minas de Ouro	223

LISTA DE TABELAS

Tabela 1 - Tabela ANOVA para modelos de regressão simples	67
Tabela 2 - Tabela ANOVA para modelos de regressão múltipla.....	71
Tabela 3 – Exemplo de matriz de confusão.....	82
Tabela 4 - Estatística descritiva dos teores no ROM – Mina de Zinco	145
Tabela 5 - Estatística descritiva das variáveis de processo – Mina de Zinco.....	147
Tabela 6 - Correlações entre valores preditos x observados - Análise de regressão – Mina de Zinco.....	161
Tabela 7 - Número mínimo de amostras por folha - Mina de Zinco	163
Tabela 8 - Importância das variáveis independentes nas árvores elaboradas – Mina de Zinco	164
Tabela 9 - Correlações entre valores preditos x observados - Random Forest – Mina de Zinco	165
Tabela 10 - Valores de C escolhidos para construção do SVR linear – Mina de Zinco.....	166
Tabela 11 - Correlações entre valores preditos x observados - SVR Linear – Mina de Zinco	167
Tabela 12 - Valores dos parâmetros C e <i>gamma</i> para o SVR não linear – Mina de Zinco....	169
Tabela 13 - Correlações entre valores preditos x observados - SVR não linear – Mina de Zinco	170
Tabela 14 - Correlações entre valores preditos x observados – Redes Neurais – Mina de Zinco	172
Tabela 15 - Avaliação dos resultados da validação cruzada <i>5-fold</i> – Mina de Zinco	176

Tabela 16 - Análise do fechamento mássico e metalúrgico – Mina de Zinco.....	177
Tabela 17 - Métricas observadas ao aplicar os modelos elaborados nas 25 amostras separadas para validação final.....	178
Tabela 18 – Resultados da validação cruzada <i>5-fold</i> para o modelo quadrático utilizado para prever a rec. metalúrgica de zinco no concentrado de zinco	179
Tabela 19 - Pilhas escolhidas para reconciliação	181
Tabela 20 - Predições geradas pela rede neural para as seis variáveis geometalúrgicas estudadas	182
Tabela 21 - Respostas da usina para cada uma das pilhas em estudo	185
Tabela 22 - Divisão das pilhas em dados de treinamento, teste e dados futuros.....	186
Tabela 23 – Diferença relativa entre predito pela rede neural versus realizado pela usina para as pilhas de treinamento do primeiro conjunto de dados.....	187
Tabela 24 - Diferenças relativas entre valores preditos via redes neurais com fator de escala e respostas observadas na usina para as pilhas de teste do primeiro conjunto de dados.....	188
Tabela 25 - Fatores de escala encontrados para as três partições de dados estudadas	188
Tabela 26 - Fatores de escala aplicados em seus respectivos bancos de teste.....	189
Tabela 27 – Diferença relativa observada quando o fator de escala foi aplicado nas previsões da rede neural para a quarta partição dos dados	190
Tabela 28 - MAE, RMSE e correlação entre predito com fator de escala e observado na usina para a quarta partição dos dados.....	190
Tabela 29 - MAE, RMSE e correlação entre predito com fator de escala e observado na usina para a quarta partição dos dados – desconsiderando a pilha 4773	191
Tabela 30 - Estatísticas descritivas – Minas de Ouro.....	200
Tabela 31 - Correlações obtidas via Análise de Regressão – Minas de Ouro	212
Tabela 32 - Número mínimo de amostras por folha – Minas de Ouro	215
Tabela 33 - Coeficiente de correlação - Random Forest – Minas de Ouro	216
Tabela 34 - Valores escolhidos do parâmetro C para o SVR linear – Minas de Ouro	217
Tabela 35 - Coeficientes de correlação - SVR linear – Minas de Ouro	218
Tabela 36 - Valores escolhidos dos parâmetros C e <i>gamma</i> - SVR não linear – Minas de Ouro	219
Tabela 37 - Correlação entre valores preditos e observados - SVR não linear - Minas de Ouro	220
Tabela 38 – Correlações obtidas utilizando Redes Neurais – Minas de Ouro	222
Tabela 39 - Avaliação dos resultados da validação cruzada <i>5-fold</i> – Minas de Ouro.....	224

Tabela 40 - Fechamento mássico e metalúrgico - Minas de Ouro	226
--	-----

SUMÁRIO

1. INTRODUÇÃO.....	18
1.1. Geometalurgia.....	20
1.2. Técnicas de estatística multivariada e <i>machine learning</i>	21
1.2.1. Análise de regressão	21
1.2.2. <i>Random forest</i>	22
1.2.3. <i>Support vector regression</i>	22
1.2.4. Aplicação de técnicas de estatística multivariada e <i>machine learning</i> em mineração.....	23
1.3. Redes neurais artificiais	24
1.3.1. Redes neurais artificiais aplicadas em mineração	27
1.4. Meta	29
1.5. Objetivos	32
1.6. Metodologia	33
1.7. Contribuições da tese	34
1.8. Organização da tese	35
2. GEOMETALURGIA	36
2.1. Definição.....	36
2.2. Histórico da geometalurgia	37
2.3. Objetivo de um modelo geometalúrgico.....	38
2.4. Benefícios de um modelo geometalúrgico.....	38
2.5. Ferramentas que auxiliam na elaboração de um modelo geometalúrgico	43
2.5.1. Matriz geometalúrgica	43
2.5.2. Análise de componentes principais e análise de grupos (clusters).....	44
2.5.3. Curva geometalúrgica.....	46
2.6. Desafios durante a elaboração de um modelo geometalúrgico.....	49
2.6.1. Não-aditividade do atributo estudado.....	49

2.6.2.	Trabalho em equipe	58
2.6.3.	Amostragem	59
2.6.5.	Softwares para armazenamento e processamento de grandes bases de dados ...	62
3.	TÉCNICAS DE ESTATÍSTICA CLÁSSICA E MACHINE LEARNING	64
3.1.	Análise de regressão	64
3.1.1.	Coeficiente de correlação	64
3.1.2.	Regressão simples	66
3.1.3.	Regressão múltipla	69
3.1.4.	Suposições dos modelos de regressão	73
3.1.5.	Validação do modelo construído	75
3.2.	Árvore de decisão	76
3.2.1.	Árvore de decisão para classificação.....	76
3.2.2.	Árvore de decisão para regressão	83
3.3.	Random forest.....	85
3.3.1.	Bagging.....	85
3.3.2.	Random forest	87
3.4.	Support vector machine	89
3.4.1.	Dados linearmente separáveis	89
3.4.2.	Dados não linearmente separáveis.....	94
3.4.3.	Support vector machine não linear	98
3.4.4.	Support vector machine para regressão	101
4.	REDES NEURAIIS ARTIFICIAIS	106
4.1.	Funcionamento do cérebro humano e compreensão do processo de aprendizagem	107
4.2.	Redes Neurais: o começo da história.....	109
4.3.	Função de ativação.....	113
4.4.	Algoritmo de aprendizado	116
4.4.1.	Backpropagation.....	117

4.4.2.	Rprop	122
4.4.3.	Adam	124
4.4.4.	Função de base radial aplicada a redes neurais	126
4.5.	Considerações sobre os dados utilizados	131
4.5.1.	Divisão da base de dados.....	131
4.5.2.	Alteração da escala das variáveis	132
4.6.	Considerações sobre a rede construída	132
4.6.1.	Número de nós na camada oculta	132
5.	IMPUTAÇÃO DE DADOS FALTANTES	134
6.	DESENVOLVIMENTO.....	142
6.1.	Mina de Vazante	142
6.1.1.	Apresentação do depósito.....	142
6.1.2.	Apresentação dos dados e estatísticas descritivas	144
6.1.3.	Modelos construídos com dados de treinamento e teste	152
6.1.4.	Validação cruzada k-fold.....	174
6.1.6.	Análise dos modelos em dados não utilizados nos bancos de treinamento e teste	177
6.1.7.	Comparação dos resultados obtidos com o modelo de referência utilizado na mina de Vazante	178
6.1.8.	Reconciliação.....	180
6.1.9.	Discussão dos resultados e sugestões de trabalhos futuros	191
6.2.	Minas de ouro	193
6.2.1.	Apresentação dos depósitos.....	193
6.2.2.	Apresentação dos dados e estatísticas descritivas	193
6.2.3.	Modelos construídos com todos os dados	204
6.2.4.	Validação cruzada k-fold.....	224
6.2.5.	Análise do fechamento mássico e metalúrgico.....	225
6.2.6.	Reconciliação.....	226

6.2.7. Discussão dos resultados e sugestão de trabalhos futuros.....	226
7. CONCLUSÕES.....	228
7.1. Conclusões iniciais.....	228
7.2. Benefícios relacionados à utilização da metodologia proposta.....	229
7.3. Trabalhos futuros.....	230
7.4. Considerações finais.....	230
8. REFERÊNCIAS BIBLIOGRÁFICAS.....	232

1. INTRODUÇÃO

Durante o processo de avaliação de um depósito mineral, as etapas de coleta, preparação física e análise química de amostras são realizadas com o objetivo de quantificar as espécies químicas existentes em uma área de interesse. Ao contrário do que era feito nas décadas passadas, a pesquisa mineral e os estudos geológicos atuais raramente são elaborados considerando apenas os teores da variável de interesse econômico do depósito. Os ganhos obtidos ao conhecer e estimar as principais variáveis químicas vão desde uma melhor definição do modelo geológico até um planejamento de lavra mais eficiente, fatores que vêm convencendo engenheiros e geólogos da importância de se analisar e estimar múltiplas variáveis. Aliado a esta conscientização, o avanço da tecnologia dos equipamentos de análise química permitiu a quantificação de diversos elementos com qualidade e preço acessível, o que facilitou o investimento das mineradoras.

Com o passar do tempo, observou-se que um modelo com múltiplos teores, estimados ou simulados, não era suficiente para compreender o comportamento do minério na usina de beneficiamento. Em diversos depósitos minerais, como os de fosfato, zinco e ouro, por exemplo, é comum observar materiais com alto teor do elemento de interesse apresentando respostas medíocres no processo, enquanto por vezes se observam materiais com teores medianos e respostas melhores, uma vez que a presença de contaminantes afeta, de forma considerável, as recuperações mássica e metalúrgica, além da qualidade do concentrado. Nasce então a geometurgia, com o objetivo de quantificar o impacto da geologia e da mineralogia na resposta metalúrgica, tornando possível compreender o comportamento das variáveis de processo ao longo do depósito. O acréscimo de informações geometúrgicas no modelo de blocos leva a uma melhor aderência financeira do projeto, a um melhor aproveitamento do minério e a uma melhor gestão dos impactos ambientais causados pela operação mineira.

Pouco se vê a respeito da elaboração de modelos geometúrgicos focados em fatores ambientais, seja por falta de investimento das mineradoras neste aspecto (por preferir focar o orçamento na compreensão de fatores que afetam a quantidade e qualidade do produto gerado) ou pela ausência de cobrança dos órgãos governamentais para compreender a fundo como as operações impactam o meio ambiente. O lado ambiental da geometurgia pode ajudar a reverter o pensamento popular de que a indústria mineira é prejudicial em todos os níveis

ambientais e que atuar de forma sustentável é algo incompatível com a exploração mineral. Acidentes como o derramamento de cianeto em Baia Mare na Romênia, o rompimento de barragens de rejeito como em Los Frailes na Espanha e mais recentemente da barragem de Fundão, no município de Bento Rodrigues e da barragem da mina Córrego do Feijão, em Brumadinho, ambas situadas em Minas Gerais, são incidentes que reforçam a opinião pública contrária à mineração e aumentam preocupação das comunidades com a gestão dos rejeitos. O primeiro passo para provar para a sociedade que a indústria mineral tem controle sobre os rejeitos gerados é provando que há total conhecimento do volume e das características químicas do material descartado, o que pode ser obtido com auxílio da geometurgia. Ainda que realização de estudos geometúrgicos com foco ambiental nos dias atuais só tenham seus reflexos percebidos à médio e longo prazo, é necessário iniciar os trabalhos algum dia para colher os frutos de uma melhor reputação no futuro.

Como dito anteriormente, o foco dos investimentos financeiros realizados pelas mineradoras está na obtenção de informações precisas e acuradas a respeito da quantidade e qualidade dos produtos por elas gerados, deixando de avaliar os impactos ambientais produzidos por não existirem recursos financeiros e/ou mão de obra disponíveis para ambas atividades. A proposta deste trabalho é elaborar uma metodologia capaz de informar no modelo de blocos a recuperação mássica e metalúrgica dos produtos e dos rejeitos que serão obtidos após o processamento do material na usina de beneficiamento, resolvendo de forma simultânea o lado financeiro e ambiental das operações, sem que haja necessidade de investimentos adicionais aos já praticados pela companhia.

Para desenvolver e testar a metodologia em questão, dados geometúrgicos foram coletados na mina de zinco de Vazante (propriedade da Nexa Resources) e nas duas minas de ouro no estado de Minas Gerais, visando construir um modelo com previsão da recuperação mássica dos produtos gerados e dos rejeitos enviados para barragem e da recuperação metalúrgica destes materiais. Para avaliar a eficácia da metodologia proposta, foram realizados estudos de reconciliação entre as predições realizadas e os valores reais observados na usina de beneficiamento de Vazante, comprovando o funcionamento da metodologia em questão. Ao aplicar a metodologia proposta será possível avaliar e otimizar os ganhos financeiros além de minimizar e controlar os impactos ambientais provenientes das operações mineiras, caminhando assim para a elaboração de estudos mineiros mais sustentáveis.

1.1. GEOMETALURGIA

A geometalurgia é uma disciplina que busca explicar a performance de um minério no processo de beneficiamento através de suas características físico-químicas. Esta performance pode ser medida através da recuperação mássica e metalúrgica obtida após um processo de concentração (como gravimetria, flotação e lixiviação), através dos teores do metal de interesse no concentrado, através do consumo de energia na etapa de moagem e diversas outras variáveis utilizadas para controlar o processo de beneficiamento. As características físico-químicas, por sua vez, são os teores de diversos elementos no *Run-of-Mine* (ROM), a densidade, a litologia, informações de textura, dentre outras abundantemente disponíveis nos bancos de dados das mineradoras.

Obter informações de variáveis de processo é algo oneroso e, por isso, o número de dados de testes de processo é sempre muito reduzido em relação ao número de informações relacionadas às características físico-químicas. O depósito de zinco em estudo neste trabalho, por exemplo, contém milhares de informações a respeito do teor de zinco e chumbo em amostras coletadas a partir de testemunhos de sondagem, o que torna possível estimar/simular tais teores no ROM através de técnicas geoestatísticas, possibilitando o conhecimento das características químicas de todo o depósito. Entretanto, somente 140 dados estão disponíveis com informações a respeito das características do minério de zinco durante o beneficiamento em ensaios laboratoriais gerados em escala de bancada, o que impossibilita o uso do mesmo conjunto de técnicas para criar um modelo de blocos 3D com previsão das variáveis de processo. Uma alternativa para a utilização de técnicas geoestatísticas na previsão de variáveis geometalúrgicas é a criação de uma função de transferência: um modelo numérico capaz de conectar as propriedades da rocha com as variáveis de processo (Deutsch, 2013).

Diversas técnicas já foram utilizadas como o objetivo de estabelecer uma função de transferência para geometalurgia. A mais comumente utilizada é a análise de regressão, uni ou multivariada, seguida pelas técnicas de árvore de decisão, *random forest*, *support vector regression*, *alternate conditional expectation (ACE)*, entre outras. A seção seguinte contém uma breve explicação das principais técnicas empregadas na elaboração de modelos geometalúrgicos e de modelos para prever outras variáveis relacionadas à mineração.

1.2. TÉCNICAS DE ESTATÍSTICA MULTIVARIADA E *MACHINE LEARNING*

Técnicas de estatística multivariada e *machine learning* são amplamente utilizadas para resolver problemas de previsão de variáveis no universo da mineração. Tendo em vista que as técnicas de *random forest* e *support vector regression* apresentaram resultados promissores em outros estudos relacionados em engenharia, em especial em dados na mineração, e que a técnica de análise de regressão é a mais empregada até o momento para elaborar modelos que relacionam variáveis geometalúrgicas, optou-se por explorar o comportamento de tais técnicas na previsão da recuperação mássica e metalúrgica de produtos e rejeitos.

1.2.1. Análise de regressão

A técnica de análise de regressão é de longe a técnica de dependência mais amplamente utilizada para construir modelos de dependência entre variáveis (Hair *et al.*, 2009). O objetivo desta técnica é encontrar uma equação que descreva algebricamente a relação entre uma variável dependente (que se deseja prever) e uma ou mais variáveis independentes (variáveis utilizadas para estimar a variável dependente de interesse), de modo que a soma dos quadrados dos resíduos (resíduo = valor observado – valor predito) seja a menor possível.

Ao ajustar um modelo de regressão linear, é necessário que uma série de suposições sejam satisfeitas para que o modelo construído seja estatisticamente válido. Entre estas suposições estão a normalidade e homocedasticidade dos resíduos, a independência entre os dados utilizados para ajustar o modelo e a linearidade da relação entre as variáveis. Além disso, é necessário verificar se os coeficientes do modelo são estatisticamente significativos, levando a conclusão de que a variável independente inserida realmente tem capacidade de prever o comportamento da variável dependente. A necessidade de satisfazer todas as condições supracitadas é algo que impulsionou a busca de outras técnicas para tratar problemas de previsão, uma vez que nem sempre é possível atender à todas exigências, principalmente no contexto da geometalurgia onde muitas das variáveis tem relação não linear com a variável dependente.

1.2.2. *Random forest*

A técnica de *random forest* consiste na construção de diversas árvores de decisão que tem seus resultados combinados para fornecer a estimativa de uma variável dependente. A variável dependente pode ser do tipo categórica ou contínua, sendo que neste trabalho somente variáveis contínuas foram utilizadas como variável resposta.

Uma árvore de decisão é formada por um nó inicial, que contém todas as amostras do banco, e estas amostras são divididas com base em uma pergunta com resposta dicotômica (sim ou não), associada a uma variável independente. Exemplo: os teores de Fe_2O_3 são maiores do que 30%? Assim, os dados são divididos em dois conjuntos, um no qual os dados respondem positivamente à pergunta e outro com os dados que respondem negativamente. Uma série de perguntas é realizada considerando as diferentes variáveis independentes disponíveis, de modo que cada nó final da árvore contenha amostras com características muito similares entre si, permitindo a determinação de uma categoria associada ao dado nó ou a determinação de um valor contínuo, a depender da característica da variável dependente que se deseja prever.

Como dito anteriormente, a técnica de *random forest* combina diversas árvores de decisão, o que aumenta sua capacidade de prever corretamente o valor da variável dependente. Cabe ao pesquisador determinar quantas árvores devem ser geradas, o número mínimo de dados em um nó final e o número de variáveis independentes a serem consideradas durante a elaboração de uma árvore. O algoritmo determina qual variável será utilizada para realizar o corte do nó e qual o valor de referência desta variável para dividi-lo, visando maximizar as classificações corretas (para variáveis dependentes categóricas) ou minimizar os resíduos (para variáveis dependentes contínuas).

Além de não necessitar que nenhuma suposição estatística seja satisfeita, a técnica de *random forest* é capaz de prever adequadamente relações não-lineares com a variável dependente, tornando-se assim uma boa opção para o ajuste de modelos geometalúrgicos.

1.2.3. *Support vector regression*

A técnica de *support vector regression* (SVR) é uma adaptação da técnica de *support vector machine* para casos nos quais é necessário prever uma variável contínua. O ponto

central desta técnica consiste em encontrar uma equação capaz de prever uma variável dependente considerando uma tolerância de erro (entre o valor real e o valor predito) pré-estabelecida pelo usuário. A técnica de SVR pode utilizar uma função linear ou um kernel não linear para transformar as variáveis independentes em questão, sendo necessário, no primeiro caso, a calibração do parâmetro C e no segundo, a calibração dos parâmetros C e *gamma*.

Pela possibilidade de utilizar kernels não lineares no ajuste do modelo, a técnica de SVR é capaz de tratar dados não aditivos, tornando-se assim uma opção viável para ajustar variáveis com esta característica, que é comumente observada em variáveis geometalúrgicas.

1.2.4. Aplicação de técnicas de estatística multivariada e *machine learning* em mineração

Diversas técnicas de estatística multivariada e *machine learning* foram aplicadas no modelamento de variáveis geometalúrgicas e alguns exemplos destas aplicações encontram-se apresentados a seguir.

Sepulveda *et al.* (2016) utilizaram a técnica de *projection pursuit regression* (PPR) para prever a recuperação metalúrgica de ouro e cobre e quatro índices de cominuição. Quando os resultados foram comparados com os obtidos através de modelos de regressão multivariados, observou-se uma melhor resposta ao utilizar PPR.

Exemplos da aplicação da técnica de análise de regressão na modelagem de variáveis geometalúrgicas podem ser vistos em MacMillan *et al.* (2011), Montoya *et al.* (2011), Fernandes (2013), Motta (2014) e Vieira (2016).

Lishchuk, Lund e Ghorbani (2019) testaram as técnicas de *random forest*, SMO (uma variação de *support vector regression*), regressão linear, M5 e M5P (variações da técnica de árvore de decisão), além de outras técnicas de *machine learning* para preverem 6 variáveis geometalúrgicas, entre elas a recuperação metalúrgica de ferro. Os autores concluíram que melhores previsões são encontradas quando técnicas que utilizam de árvores de decisão são aplicadas para prever variáveis não-aditivas.

Prades e Deutsch (2016) testaram a performance de modelos de regressão linear, modelos de regressão quadráticos, ACE (*alternating conditional expectations*), *ridge regression*, *random forest* e *gradient boosted models* (GBM) na previsão de consumo de ácido,

recuperação e a impureza em um depósito de cobre de Antucoya. *Random forest* e GBM apresentaram os melhores resultados em termos de R^2 .

Drumond (2019) utilizou as técnicas de regressão logística, análise discriminante, *support vector machine*, *random forest*, redes neurais artificiais, *k-nearest neighbours*, árvore de decisão e *naive bayes* para prever partículas minerais como minério ou estéril em um processo de pré-concentração utilizando *ore-sorting*, tendo como variáveis independentes medidas oriundas de raios-X. O autor considerou que as técnicas de *support vector machine*, *k-nearest neighbours*, *random forest* e redes neurais apresentaram os melhores resultados.

Para o conhecimento de outros trabalhos que utilizaram técnicas de *machine learning* na solução de problemas relacionados ao processamento de minérios, não entrando no mérito de trabalhos de geometurgia, recomenda-se a leitura de McCoy e Auret (2018).

1.3. REDES NEURAIAS ARTIFICIAIS

A técnica de redes neurais tem como objetivo criar uma “máquina” capaz de reproduzir a capacidade cerebral humana de resolver determinados problemas e, tendo em vista que os neurônios são o ponto central da capacidade humana de aprender, qualquer estudo neste sentido deveria se iniciar com a criação de um neurônio artificial. Em 1943, McCulloch e Pitts construíram o primeiro protótipo do que seria a atividade de um neurônio humano. A partir deste momento, muitas adaptações ao primeiro estudo realizado foram feitas até chegarmos ao que conhecemos hoje como a técnica de redes neurais artificiais.

Percebe-se, que muito tempo passou entre a elaboração do primeiro neurônio artificial e a utilização da técnica de redes neurais para resolver problemas do cotidiano, a qual ganhou força e maior visibilidade a partir dos anos 90. Esta demora está intimamente ligada à ausência de capacidade computacional nas décadas de 50 a 80 para suportar os cálculos necessários para se ajustar uma rede neural precisa e acurada. Com o avanço da tecnologia nos anos 90, computadores pessoais tornaram-se capazes de criar redes neurais, levando à um rápido avanço na utilização desta técnica.

Redes neurais são compostas de três estruturas fundamentais: camada de *input*, camada oculta e camada de *output*. Vários estudos utilizam de duas ou mais camadas ocultas, dado à complexidade do problema que se deseja tratar, sendo esta uma decisão do pesquisador que está

utilizando a técnica. Na camada de *input*, são inseridas as variáveis relevantes para o problema, conhecidas em estatística clássica como variáveis independentes. O número de nós na camada de *input* será igual ao número de variáveis independentes inseridas. Na camada oculta, ocorre o processamento matemático dos *inputs*. O número de nós na(s) camada(s) oculta(s) também é uma decisão a ser tomada pelo usuário. Na camada de *output*, estão as variáveis a serem preditas, sendo que o número de nós desta camada é igual ao número de variáveis dependentes do problema.

No caso de uma rede com uma única camada oculta, todos os nós da camada de *input* estão conectados aos nós da camada oculta e estes, conectados ao nós de *output*, como ilustrado na Figura 1, que não está considerando os bias para efeito de simplificação (maiores detalhes serão dados no capítulo 4). Esta conexão entre os nós é chamada de peso e estes são determinados por um algoritmo de aprendizado.

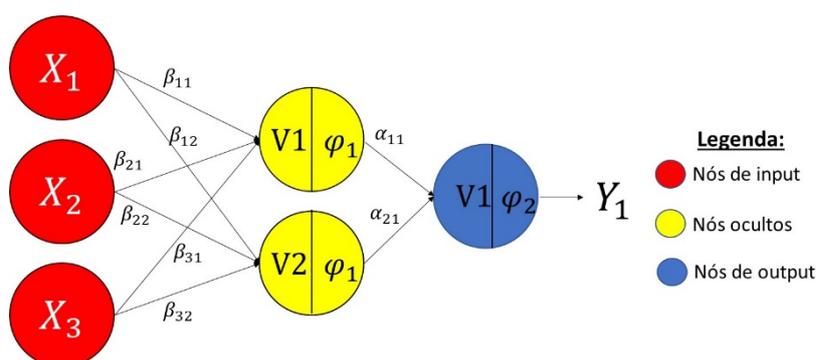


Figura 1 - Exemplo de rede com uma camada oculta

Para redes com duas ou mais camadas ocultas, os nós de *input* se conectam aos nós da primeira camada oculta e os nós da primeira camada oculta se conectam com os nós da segunda camada oculta, e assim sucessivamente, até que os nós da última camada oculta se conectem aos nós de *output*. A Figura 2 ilustra uma rede com duas camadas ocultas, desconsiderando os bias, para efeito de simplificação.

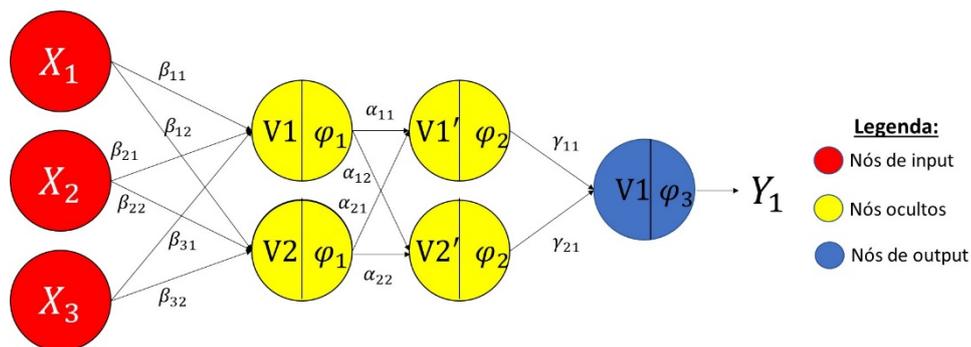


Figura 2 - Exemplo de rede com duas camadas ocultas

Os valores dos nós de *input* são multiplicados aos valores dos pesos que ligam estes nós aos nós ocultos. Todos os valores que alimentam o nó oculto são então somados e armazenados em uma variável denominada contador. Este contador, que está presente em cada nó oculto (representado nas Figuras 1 e 2 pela letra “V”) alimenta uma função de ativação que deve ser determinada pelo usuário. Após calculado o valor de saída da função de ativação para cada nó oculto, deve-se multiplicar o valor de saída da função pelo peso que conecta este nó ao nó da camada seguinte. Da mesma maneira, todas as entradas daquele nó são somadas, adicionadas ao contador e alimentadas na função de ativação, que pode ser igual ou diferente da função adotada na primeira camada. Quando a função de ativação está na camada de *output*, significa que o valor de saída da função é igual a previsão realizada pela rede para a variável em questão.

Em resumo, para ajustar uma rede neural é preciso que se determine:

- O número de nós na camada de *input*;
- O número de camadas ocultas e de nós em cada uma dessas camadas;
- O número de nós de *output*;
- As funções de ativação que estarão presentes em cada nó oculto ou nó de *output*;
- O algoritmo de aprendizado que determinará os pesos de conexão da rede.

No capítulo 4, serão dados todos os detalhes necessários para compreender a fundo o funcionamento de uma rede neural e como ajustá-la na prática.

1.3.1. Redes neurais artificiais aplicadas em mineração

Nesta seção, encontra-se a revisão bibliográfica de trabalhos nos quais a técnica de redes neurais foi utilizada para prever variáveis relacionadas à problemas da indústria mineral.

Nakhaei *et al.* (2012) construíram uma rede para prever os teores de cobre e molibdênio no concentrado e suas recuperações para a planta piloto de Sarcheshmeh. Segundo os autores, estas variáveis são mais difíceis de se obter do que as variáveis de *input* utilizadas na rede, que foram o teor de Cu e Mo na alimentação da etapa *rougher* e na alimentação da coluna de flotação, as dosagens de coletor e espumante, as taxas de água de lavagem e de fluxo de ar, a altura da espuma e o *air holdup*. Os autores criaram o modelo para auxiliar na otimização da planta. As redes neurais construídas produziram correlações entre os valores reais e as previsões entre 0.89 e 0.92, no banco de teste.

Jorjani *et al.* (2008) elaboraram uma rede utilizando como variáveis de *input* o tempo de lixiviação, a densidade da polpa, as concentrações de ácido e as taxas de agitação, com o objetivo de prever as recuperações de lantânio, cério, ítrio e neodímio no concentrado da lixiviação de apatita do depósito de Chadormalu, no Irã. A rede por eles construída tinha duas camadas ocultas com cinco nós ocultos em cada e uma variável de *output* por vez, logo, quatro redes foram criadas. As correlações obtidas entre os valores reais e as previsões feitas pela rede variaram entre 0.952 e 0.985, em um banco de teste contendo 6 amostras. Os resultados foram satisfatórios, segundo os autores, significando que a rede poderia ser utilizada para reduzir custos relacionados à análise de elementos terras raras na planta de lixiviação e auxiliaria também na determinação do melhor cenário de *input*.

Srivastava *et al.* (2017) criaram uma rede neural para prever a energia de uma moagem semi-autógena (SAG) utilizando como variáveis de *input* a taxa de alimentação e a energia gasta no moinho em intervalos de tempo anteriores. A rede utilizou um modelo do tipo NARX (*non-linear auto-regressive model with exogenous inputs*) para capturar a estrutura temporal, que neste caso apresentou os melhores resultados considerando um atraso (retardo) de quatro minutos como *input*.

A taxa de penetração para sondagem percussiva foi predita por Kahraman (2016) através de modelos de regressão e de redes neurais, utilizando a pressão operacional, a pressão de alimentação, a dureza e a densidade como variáveis de *input*. Os resultados mostraram que as

redes neurais apresentaram melhores correlações do que os modelos de regressão elaborados para todos os cenários testados (0.99 contra 0.92, no melhor cenário).

Wu e Zhou (1992) fizeram uma estimativa de teores de cobre utilizando uma rede neural que considerou as coordenadas como variáveis de *input* e possuía duas camadas ocultas com 28 nós ocultos em cada uma. Os autores compararam as estimativas da rede com as obtidas através da utilização dos métodos do polígono de influência, do inverso do quadrado da distância e da utilização de variografia e krigagem, e concluíram que a rede neural apresentava resultados promissores.

Tawo e Al-Alawi (1999) compararam as estimativas de uma rede neural contra as realizadas utilizando krigagem ordinária para um depósito de bauxita. As variáveis estimadas foram a espessura da mineralização e os teores de Al_2O_3 e SiO_2 . A rede neural utilizou as coordenadas como variáveis de *input* e apresentou uma camada oculta com sete nós ocultos, enquanto a krigagem ordinária foi baseada em um variograma esférico para representar a continuidade do depósito. Os valores de R^2 encontrados quando a rede neural foi aplicada no banco de teste foram 0.876 para a variável espessura, 0.755 para Al_2O_3 e 0.792 para SiO_2 . Os resultados mostraram que, a medida em que os dados vão se tornando esparsos, o método geoestatístico gera melhores resultados, mostrando que nas bordas do depósito as redes neurais tiveram uma performance pior.

Samanta *et al.* (2005) apresentaram uma comparação entre uma rede neural simples e múltiplas redes neurais geradas através do algoritmo Adaboost para prever o teor de ouro em Nome, no Alaska. As coordenadas foram utilizadas como variáveis de *input* juntamente com a profundidade do lençol freático. Os autores mostraram não existir nenhuma vantagem em se utilizar o algoritmo Adaboost (que é mais complicado na prática) ao invés de utilizar uma rede neural simples, tendo em vista que a primeira técnica citada apresentou um R^2 de 0.20 no banco de teste contra um R^2 de 0.19 obtido quando uma rede neural simples foi utilizada. Dutta *et al.* (2010) conduziram um estudo no mesmo depósito, comparando as estimativas obtidas com redes neurais com as obtidas através de krigagem e support vector machine (SVM). Nenhum dos métodos mostrou resultados excepcionais, mas o SVM obteve melhor performance, apresentando um R^2 de 0.234.

Panda e Tripathy (2014) ajustaram uma rede neural para prever o teor e a recuperação metalúrgica de cromita utilizando uma mesa vibratória, tendo como variáveis de *input* a taxa de água de lavagem, o ângulo de inclinação do deck e a taxa de lama na alimentação da mesa.

A rede ajustada possuía uma única camada oculta com 11 nós e as duas variáveis de *output* eram preditas simultaneamente. O R^2 obtido para o banco de teste foi de 0.96 para a previsão de teor e de 0.99 para a previsão realizada para recuperação.

Exemplos da aplicação de redes neurais para prever variáveis relacionadas ao transporte e ao desmonte de minérios encontram-se na revisão bibliográfica realizada por Franco-Sepúlveda, Del Rio-Cuervo e Pachón-Hernández (2018).

1.4. META

A meta deste estudo é elaborar uma metodologia para estimativa de variáveis geometalúrgicas, capaz de prever simultaneamente a recuperação mássica e metalúrgica de todos os produtos e rejeitos gerados pela usina de beneficiamento, utilizando como variáveis independentes informações amplamente disponíveis no banco de dados, como teores e descrições geológicas. Através destas estimativas será possível prever o balanço mássico e metalúrgico da usina ao beneficiar cada tipo de material que nela é alimentado, conforme será explicado a seguir.

De maneira geral, as operações mineiras calculam a recuperação mássica e metalúrgica de cada processo individualmente, o que não permite avaliar o balanço mássico e metalúrgico da usina. A Figura 3 apresenta um exemplo de um teste de bancada que imita um processo de beneficiamento com três etapas de concentração, onde os cálculos de recuperação são realizados para cada processo de maneira individual: a recuperação mássica do primeiro processo (em vermelho) é calculada em relação ao alimentado no teste de bancada, ou seja, 800 g de material foram concentradas em relação aos 2000 g alimentados, gerando uma recuperação mássica de 40%. O segundo processo, representado em verde, concentra 300 g de material em relação aos 1200 g que alimentaram o processo 2, gerando uma recuperação mássica de 25% nesta etapa. O último processo (representado em azul), por sua vez, recuperou 200 g de material em relação aos 900 g que alimentaram o processo 3, levando a uma recuperação mássica de 22% neste processo individualmente. O mesmo raciocínio é utilizado para calcular a recuperação metalúrgica de cada etapa.

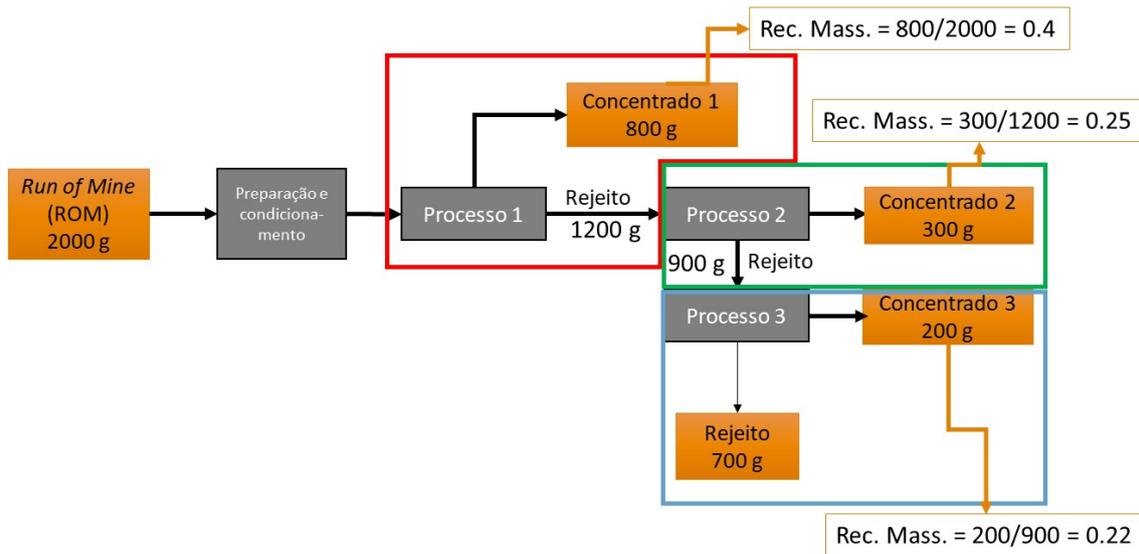


Figura 3 - Cálculos tradicionais de recuperação mássica considerando cada etapa do processamento mineral individualmente.

Para realizar o balanço mássico e metalúrgico, devemos considerar os cálculos realizados em relação ao material alimentado no teste (ou alimentado na usina, caso os cálculos estejam sendo realizados em escala industrial), conforme ilustrado na Figura 4. A recuperação mássica do processo 1 é calculada da mesma maneira que foi citada no exemplo anterior, considerando que 800 g de material foram destinadas ao concentrado 1 em relação aos 2000 g de material alimentado no teste de bancada, indicando que 40% de todo material alimentado tem como destino o concentrado 1. Ao calcular a recuperação mássica do processo 2 vemos que dos 2000 g de material alimentados no teste, 300 g tiveram como destino o concentrado 2, indicando que 15% do material alimentado tem como destino o concentrado do segundo processo de beneficiamento realizado para tratar este minério. Por fim, conclui-se que 200 g de material tiveram como destino o concentrado 3, indicando que dos 2000 g de material alimentado no teste de bancada, 10% foram destinados ao terceiro concentrado gerado e que 35% de todo material que alimentou o teste é destinado ao rejeito, uma vez que 700 g de material não foram recuperados em nenhuma das etapas anteriores.

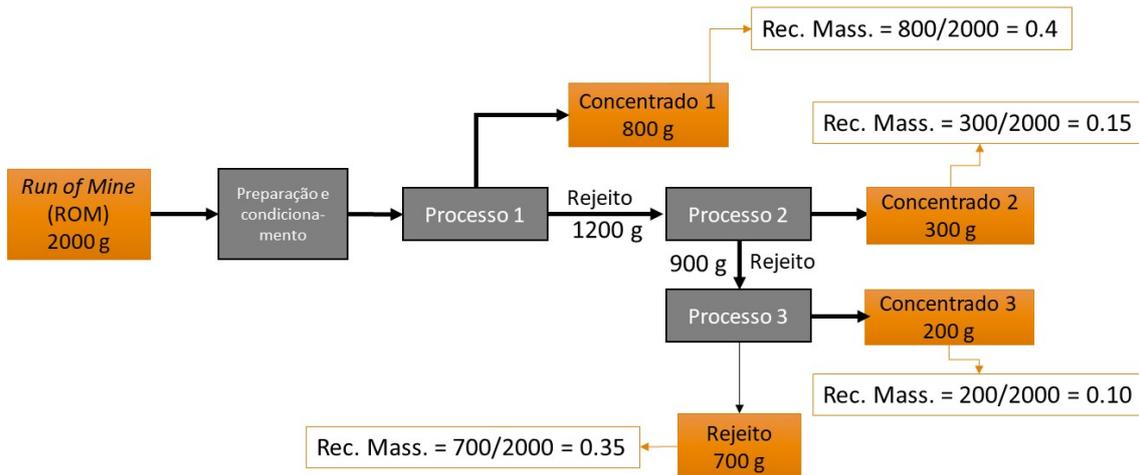


Figura 4 - Cálculos de recuperação para realização do balanço mássico da usina

Tendo em vista que a meta deste estudo é prever a recuperação mássica e metalúrgica de todos os produtos e rejeitos através das características físico-químicas do material alimentado, para obter uma previsão do balanço mássico e metalúrgico da usina, utilizou-se o método de cálculo acima exemplificado. Desta forma, todos os pontos das usinas estudadas onde há saída de massa e metal terão uma previsão da recuperação mássica e metalúrgica em relação ao ROM, de modo que ao somar as previsões de recuperação mássica de todos os produtos e rejeitos gerados o resultado observado seja igual a 100%, honrando assim o fechamento mássico. O mesmo deverá ocorrer ao somar as previsões de recuperação metalúrgica de todos os produtos e rejeitos, para honrar o fechamento metalúrgico da planta de beneficiamento.

Declaração de tese:

É possível prever simultaneamente, em pilhas ou no modelo de blocos do depósito, a recuperação mássica e metalúrgica de todos os produtos e rejeitos oriundos de plantas de beneficiamento, com base nas características físico-químicas do minério que irá ser alimentado?

Para testar essa hipótese investigativa será necessário encontrar uma técnica capaz de prever simultaneamente duas ou mais variáveis dependentes, mantendo o fechamento mássico e metalúrgico o mais próximo possível da realidade. As previsões geradas através da técnica escolhida deverão ser reconciliadas com os resultados de produção reportados pela usina de beneficiamento para comprovar a validade da metodologia proposta no depósito estudado.

1.5. OBJETIVOS

Para atingir a meta estabelecida será necessário desenvolver os seguintes objetivos específicos:

- Estudar técnicas de estatística multivariada (análise de regressão) e *machine learning* (*random forest*, *support vector regression* e redes neurais artificiais) para determinar qual fornece os melhores resultados (em termos de correlação, erro médio absoluto e raiz quadrada do erro quadrático médio) na predição das variáveis geometalúrgicas. Para construir os modelos utilizando tais técnicas foram utilizadas bibliotecas e pacotes disponíveis no Python®;
- Avaliar a qualidade do fechamento mássico e metalúrgico obtido ao utilizar as técnicas supracitadas e determinar qual é capaz de gerar resultados mais aderentes à realidade;
- Verificar se a técnica de redes neurais, capaz de prever mais de uma variável simultaneamente, gera melhores resultados que as demais técnicas em termos de fechamento e das demais métricas avaliadas;
- Determinada a melhor técnica de previsão, aplicá-la em pilhas de minério que alimentarão a unidade de processamento cujas variáveis independentes são conhecidas e reconciliar as previsões do modelo com a realidade observada na usina de beneficiamento.

1.6. METODOLOGIA

A primeira etapa do estudo estava relacionada à identificação e aplicação de uma técnica capaz de prever simultaneamente duas ou mais variáveis de processo distintas. É comum observar modelos construídos para prever uma variável dependente por vez, geralmente por meio das técnicas de análise de regressão (AR), *random forest (RF)* ou *support vector regression (SVR)*, sendo que a previsão simultânea de todos os produtos e rejeitos da usina era algo não tentado até então. A técnica de redes neurais artificiais (RNA) foi identificada como uma técnica capaz de realizar previsões de mais de uma variável simultaneamente, ainda que a maioria dos problemas de previsão simultânea, nos quais tal técnica é empregada, estejam relacionados a problemas de classificação. Desta maneira, iniciou-se o estudo de vários algoritmos de aprendizado, das funções de ativação existentes e de softwares capazes de construir um modelo de redes neurais para os casos em estudo.

O passo seguinte foi encontrar modelos de previsão robustos com uma quantidade escassa de variáveis independentes (4 para o estudo da mina de Vazante e 3 para as minas de ouro) para prever um grande número de variáveis dependentes (6 para a mina de zinco e 8 para as minas de ouro). Para as minas de ouro, houve um dificultador: a quantidade reduzida de testes de processo e um banco de dados heterotópico. Para superar tal dificuldade, foi necessário estudar e aplicar uma técnica de imputação de dados faltantes, caso contrário, não seria possível prosseguir o estudo com as 22 amostras isotópicas disponíveis. Para realizar a imputação, foi escolhida a técnicas de Máxima Verossimilhança e *Expectation Maximization* e um script em Python® foi desenvolvido para efetuar os cálculos.

Após elaborados os melhores modelos possíveis utilizando as técnicas em estudo (AR, RF, SVR e RNA), a terceira etapa foi determinar qual técnica apresentava os melhores resultados em termos de correlação, erro médio absoluto, raiz quadrada do erro quadrático médio e em termos de fechamento mássico e metalúrgico das previsões. Para a mina de Vazante (que dispunha de um modelo geometalúrgico prévio), a melhor técnica teve seus resultados confrontados também com os obtidos através do modelo utilizado atualmente pela operação para prever a recuperação metalúrgica de zinco no concentrado de zinco.

Comprovada a superioridade da técnica escolhida em relação às demais, o passo seguinte consistiu em reconciliar os resultados das previsões com os produtos e rejeitos gerados pela usina. Neste ponto, surgiu uma dificuldade, uma vez que muitas mineradoras ainda não

possuem capacidade de acompanhar perfeitamente o percurso do minério lavrado nos stopes, que irá compor as pilhas que alimentarão a usina e, conseqüentemente, os produtos e rejeitos gerados a partir de tais pilhas. Ao analisar os dados disponíveis das duas usinas de beneficiamento em estudo, concluiu-se não ser possível realizar uma reconciliação adequada dos modelos elaborados para as minas de ouro, devido à falta de informações. Logo, somente a usina de Vazante participou desta etapa do estudo e da etapa citada a seguir.

A última etapa esteve relacionada à determinação do fator de escala entre os testes de bancada e a usina de beneficiamento. Sabe-se, que os testes de bancada consideraram um circuito de flotação aberto, no qual a recuperação metalúrgica é inferior à observada na usina, pelo fato desta última considerar um circuito de flotação fechado, que permite a recirculação dos rejeitos, aumentando assim a recuperação metalúrgica de zinco no concentrado de zinco e reduzindo a recuperação metalúrgica de zinco no rejeito. Uma vez que o modelo elaborado para a previsão das variáveis geometalúrgicas foi criado considerando amostras geradas através de testes de bancada, foi necessário encontrar um fator de aumento de escala (laboratório para industrial) que ao multiplicar as previsões as tornam comparáveis aos resultados observados na usina de beneficiamento, permitindo assim a reconciliação adequada dos resultados.

Através da reconciliação dos resultados da usina com as previsões, usando o fator de escala, foi possível determinar os erros aos quais tal metodologia está submetida e avaliar a eficácia desta na previsão do balanço mássico e metalúrgico da planta de beneficiamento de zinco em estudo.

1.7. CONTRIBUIÇÕES DA TESE

As contribuições desta tese são:

- A elaboração de uma metodologia que permita a criação de um modelo geometalúrgico capaz de prever, simultaneamente, variáveis com impacto financeiro e ambiental, sem que informações adicionais tenham que ser coletadas e sem a necessidade de dispendir mais tempo no tratamento de tais variáveis;

- A utilização da técnica de redes neurais artificiais para modelar variáveis geometalúrgicas que necessitem de fechamento mássico e metalúrgico, uma vez que esta técnica apresentou resultados superiores às demais neste aspecto;
- Permitir que as empresas que utilizem da metodologia proposta controlem um maior número de variáveis no planejamento de mina, passando a considerar variáveis que afetem o meio ambiente (como a quantidade de rejeitos) dentro da função benefício, para que o impacto ambiental das operações mineiras seja conhecido previamente e minimizado.

1.8. ORGANIZAÇÃO DA TESE

Os capítulos seguintes deste documento estão estruturados da seguinte maneira: o capítulo 2 apresenta a revisão bibliográfica de geometalurgia, além de apresentar os benefícios e desafios que envolvem esta disciplina. O capítulo 3 revisa as técnicas de estatística clássica e de *machine learning* utilizadas na elaboração deste estudo. O capítulo 4 apresenta a revisão bibliográfica da técnica de redes neurais artificiais. O capítulo 5 apresenta os conceitos de imputação de dados faltantes, necessária para elaboração do modelo geometalúrgico das minas de ouro aqui estudadas. O capítulo 6 apresenta o desenvolvimento do trabalho, passando por toda parte de análise estatística, elaboração de modelos preditivos utilizando as técnicas apresentadas nos capítulos 4 e 5 e a reconciliação das previsões do modelo escolhido com os resultados reportados pela usina de beneficiamento. O capítulo 7 apresenta as conclusões do estudo. Encerra-se com as referências bibliográficas utilizadas.

2. GEOMETALURGIA

2.1. DEFINIÇÃO

De acordo com Williams (2013), a geometalurgia é a ciência que nos permite compreender porque as rochas e minerais apresentam tais performances quando passam pelos processos metalúrgicos. Segundo David (2013), a geometalurgia pode ser definida como: “disciplina científica na qual dados geológicos, de mina e de processo são analisados conjuntamente para gerar informações úteis e conhecimento para otimizar a rentabilidade dos recursos”. Fernandes (2013) define a geometalurgia como uma disciplina cujo objetivo é “descobrir o elo existente entre a geologia e a resposta do minério no processo de beneficiamento”. Para Dunham e Vann (2007), “a geometalurgia é uma abordagem multidisciplinar com o objetivo de estudar algumas das complexidades associadas com a determinação de valor de um recurso e, portanto, se é econômico explorá-lo. Através da integração da geologia, planejamento de mina, design operacional, processamento mineral e metalurgia, a geometalurgia tem como objetivo melhorar a compreensão fundamental dos recursos econômicos”. Deutsch (2013), por sua vez, vê a geometalurgia como “um ramo aplicado da metalurgia e da análise numérica que tem ênfase no controle geológico do processamento mineral e na predição da performance do processo a partir de múltiplas propriedades de rocha”.

Percebe-se imediatamente que não existe uma definição única para a palavra geometalurgia, mas todas as definições encontradas na literatura passam pelo estudo do minério e suas características intrínsecas e de como essas características afetam o beneficiamento e os recursos minerais. É importante ressaltar que o prefixo “geo” em geometalurgia significa “terra”, o que faz com que a palavra como um todo tenha o significado de estudo da metalurgia da terra, ou como proposto por Kittler *et al.* (2011), “estudo da metalurgia de pelo menos parte da terra que contenha os minerais de interesse”.

Pensando em uma definição abrangente, podemos definir a geometalurgia como sendo a disciplina que analisa dados geológicos, mineralógicos, químicos, físicos e ambientais coletados em um depósito mineral, e busca compreender quais são os fatores que afetam a performance do minério no processo de beneficiamento, tanto em relação a parâmetros

econômicos, como recuperação mássica e metalúrgica, quanto em relação a parâmetros ambientais, como consumo de água, energia e geração de rejeitos.

2.2. HISTÓRICO DA GEOMETALURGIA

De acordo com Jackson *et al.* (2011), a primeira menção na literatura da palavra “geometalúrgico” foi feita por McQuiston e Bechaud (1968): “ (...) muito pouco tem aparecido na literatura sobre amostragem de um depósito virgem sobre o ponto de vista metalúrgico. Talvez, seria mais descritivo designar este ponto de vista como ‘geometalúrgico’, uma vez que a geologia é inseparavelmente entrelaçada com a metalurgia na compreensão das complexidades de um depósito, eventualmente levando a definição de reservas lavráveis, com o desenvolvimento de um fluxograma e critérios de engenharia para o planejamento de uma operação bem-sucedida e rentável”. Esta citação mostra que na década de 60 já existia uma preocupação em estudar a relação entre a geologia e a metalurgia, e mais do que isto, já existia a consciência de que fazê-lo traria benefícios econômicos para a operação mineira.

Conforme apresentado no artigo de Williams (2013), alguns trabalhos das décadas de 70 já tratavam de problemas que hoje são considerados como desafios da geometalurgia. Ele cita em seu artigo os trabalhos de: Palagi e Stillar (1976), que abordava problemas relacionados a falta de homogeneidade do minério e o impacto no processo de flotação; Sharp (1976) que descreveu uma enorme variabilidade dos minerais de sulfeto na mina de Magmont (Estados Unidos) e citou quais foram as estratégias encontradas para minimizar os problemas encontrados; e Hinckfuss (1976), que descreveu o impacto do processo de moagem na eficiência operacional da mina de Bougainville Copper (Papua Nova Guiné), apresentou relações entre a granulometria da moagem e a recuperação de cobre para diferentes tipos de minério, entre outras associações da mineralogia com o processo de beneficiamento.

De acordo com a revisão realizada por Williams (2013), é possível perceber que apesar da geometalurgia ter tido seu espaço reconhecido como ciência por volta dos anos 2000, que ela está presente no mundo mineiro há muitas décadas, apesar de não ser identificada com este nome. Segundo Jackson *et al.* (2011), “o atual nível de interesse na geometalurgia talvez tenha sido impulsionado pela revolução na capacidade computacional, a qual permitiu aquisição, armazenamento e análise de dados em uma escala não praticável a 20 ou 30 anos atrás”. Deutsch (2016), por sua vez, considera que “o aumento do conhecimento da geologia, da mineração, da

metalurgia e de processos de modelagem, combinados com uma diminuição nos teores médios dos minérios fez com que esta abordagem se tornasse particularmente atrativa atualmente”.

2.3. OBJETIVO DE UM MODELO GEOMETALÚRGICO

O grande objetivo de um modelo geometalúrgico é ser capaz de responder a seguinte pergunta colocada por Kittler *et al.* (2011): “Como será a performance do bloco X quando ele for processado?”.

A performance, colocada na pergunta anterior, pode ser mensurada de diferentes maneiras e através de diferentes variáveis. Em minas de cobre, por exemplo, cujo problema geralmente é a resistência mecânica e a influência na cominuição do material, a performance de um bloco X pode ser medida através do tempo de moagem necessário para que o material apresente uma granulometria específica. Em minas localizadas em regiões com escassez de recursos hídricos, a performance pode ser medida através da quantidade de água necessária para processar um dado bloco de minério. Outro exemplo pode ser dado através de depósitos de fosfato, os quais devem produzir um concentrado com teor dentro das especificações do cliente, logo, a performance pode ser medida através do teor de determinados elementos no produto final. Ou ainda, a performance pode ser medida através da recuperação mássica e metalúrgica de um minério após o processo de beneficiamento, como é comum em minas de ferro, zinco e muitas outras commodities minerais.

É importante perceber que a pergunta que deve ser respondida trata cada bloco individualmente e não a mina como um todo. Foi e ainda é muito comum observar operações mineiras que adotam um valor fixo das variáveis de processo, desconsiderando a variabilidade existente no minério. Um modelo geometalúrgico robusto possui, para cada bloco X do depósito em estudo, sua própria estimativa da variável de interesse, de preferência com uma medida da incerteza associada a ela.

2.4. BENEFÍCIOS DE UM MODELO GEOMETALÚRGICO

De uma forma geral, o maior ganho que um modelo geometalúrgico pode trazer para uma mina é a redução das incertezas associadas ao minério e seu beneficiamento, obtida através

de um maior conhecimento da variabilidade do material que será processado. Podem ser vistos como três, os grandes benefícios que podem ser adquiridos ao conduzir um estudo geometalúrgico, os quais serão abordados com maior detalhe a seguir: melhor aproveitamento do minério devido a uma planta adaptada as características do mesmo, melhor gestão dos impactos ambientais causados pela operação mineira e melhor aderência financeira do planejado para o projeto.

O primeiro benefício é uma melhor operação da planta de beneficiamento, que pode ser obtida através do conhecimento prévio do minério que será alimentado e de suas características de processo. O modelo geometalúrgico apresentado por Bye (2011) mostra uma diferença de 50 milhões de dólares por ano quando são comparados o planejamento utilizando o modelo geometalúrgico e o planejamento do caso base. Ao identificar períodos de alto risco no beneficiamento de um dado minério, é possível planejar estratégias para minimizá-lo e otimizar o planejamento, agregando valor graças ao conhecimento prévio de fatores que sem o estudo da geometalurgia passariam despercebidos. Dunham e Vann (2007) também sugerem que é possível tomar medidas durante a etapa de planejamento para administrar as consequências de um material com resposta metalúrgica inferior à esperada, como por exemplo adotando a blendagem do minério em casos nos quais há conhecimento de que este não atingiria a qualidade do concentrado desejada. Sugerem ainda a possibilidade de adaptar o circuito de beneficiamento, como por exemplo inserindo uma etapa *cleaner* no circuito de flotação. Percebe-se, portanto, que é possível planejar a alimentação de um material que atenda não somente as especificações de teor da alimentação da usina, como também as especificações do produto final.

Em situações nas quais o projeto será implementado, é possível tirar vantagem da geometalurgia para elaborar uma planta de beneficiamento que atenda todas as necessidades da operação mineira em estudo. Como abordado por McQuiston e Bechaud (1968), ao conhecer-se bem as características geológicas e texturais do depósito e os efeitos dessas no comportamento geometalúrgico, é possível construir uma planta concentradora ou lavadora, adaptada para lidar com as diferenças que serão observadas localmente, no curto ou no longo prazo. Ao planejar a usina com todas as possibilidades de adaptação no início do projeto, evita-se surpresas de ter que adaptá-la ao longo da vida da mina, o que certamente causará maiores transtornos operacionais e financeiros. Uma planta flexível permite atingir melhores recuperações, levando a um melhor aproveitamento dos recursos minerais.

O outro benefício na elaboração de modelos geometalúrgicos está associado a um melhor aproveitamento dos recursos naturais, colaborando para um menor impacto no meio ambiente. Dunham, Vann e Coward (2011) citam que a geometalurgia é capaz de melhorar a sustentabilidade do projeto através de medidas para minimização proativa da poluição, tanto nas operações mineiras, quanto no processamento mineral, além de permitir uma otimização do consumo energético e até das emissões de carbono na atmosfera. McCullough *et al.* (2013) também colocam a gestão dos aspectos ambientais como um dos benefícios de programas geometalúrgicos, citando principalmente os benefícios relacionados a redução da poluição.

Para se ter uma ideia do impacto que a mineração possui em países com tradição mineira, Pimentel (2004) apud Dunham, Vann e Coward (2011) apontam que cerca de 30% da energia elétrica do Chile (maior produtor mundial de cobre) é consumida pela mineração, processamento, fundição e refinamento do cobre. Na Austrália, é percebido um aumento anual no consumo de energia referente a 3.7%, sem que um aumento na produção mineral seja percebido, ou seja, se consome mais energia para minerar a mesma quantidade de material, fato relacionado à queda da qualidade dos depósitos, o aumento das profundidades das mineralizações e ao acesso do minério em áreas remotas (Sandu e Syed, 2008 apud Bye, 2011). Vê-se, portanto, a importância de elaborar modelos para prever o consumo energético e tentar otimizá-lo cada vez mais, uma vez que um consumo maior leva a uma maior emissão de gases poluentes na atmosfera, principalmente quando combustíveis fósseis são utilizados na geração de energia.

É de conhecimento geral que corpos de minério são recursos não renováveis, por isso é necessário tratá-los da forma mais sustentável possível. Conforme definido por Brundtland (1987), apud Dunham, Vann e Coward (2011), desenvolvimento sustentável é “o desenvolvimento que atende a necessidade do presente sem comprometer a possibilidade das futuras gerações de atenderem suas próprias necessidades”. Por isso, é necessário maximizar a geração de produto, o que está relacionado com as recuperações metalúrgicas obtidas durante o beneficiamento. Um processo ineficiente enviará minério para pilhas de estéril ou barragens de rejeito, causando um desperdício muitas vezes irreversível. É necessário adaptar o processo para que a recuperação seja sempre a maior possível e os recursos minerais aproveitados de forma máxima. Além disso, faz-se necessária uma gestão responsável do estéril e do rejeito gerado, para que esses não comprometam as próximas gerações causando danos ambientais como a contaminação do solo e de lençóis freáticos, por exemplo. Sabe-se que a mineração gera grandes impactos ao meio ambiente e é o dever de todos reduzi-los ao máximo.

Ao pensar em operações mineiras com baixa eficiência no processo de beneficiamento, pode-se chegar mesma conclusão que Wates *et al.* (2010) apud Louwrens, Napier-Munn e Keeney (2015) chegaram: “O rejeito de hoje é o recurso de amanhã”. Por isso, os autores sugerem uma gestão responsável do rejeito, que além de ser um potencial problema ambiental, pode também se tornar fonte de renda no futuro. Esta preocupação levou os autores a elaborar o primeiro modelo de blocos 3D que se tem conhecimento de uma barragem de rejeitos (Louwrens, Napier-Munn e Keeney, 2015), no qual encontram-se estimadas as recuperações metalúrgicas de ouro, magnetita, cobalto e cobre.

O tratamento do estéril e dos rejeitos, assim como os fatores ambientais, vêm ganhando mais importância na mineração com o passar do tempo, com a tendência de esses serem cada vez mais estudados pela geometurgia, como afirmam Dunham, Vann e Coward (2011): “a geometurgia vai muito além do link entre as propriedades do minério e sua performance no processamento mineral, devendo englobar também fatores ambientais relacionados a sustentabilidade das operações”. Knight, Olson Hoal e Abraham (2011) também colocam que a tendência atual é a incorporação de metas e objetivos com desenvolvimento sustentável, tanto nos projetos de exploração quanto nas minas em operação. Um exemplo de modelo geometúrgico com preocupação ambiental foi elaborado por esses autores, os quais modelaram regiões onde o teor de arsênio no rejeito tem grandes possibilidades de estar acima dos limites toleráveis quando o material for processado, indicando regiões onde a lavra deve ser evitada ou pensada com cautela.

O terceiro benefício da geometurgia está relacionado ao aumento do conhecimento do corpo de minério e de sua variabilidade, o que refletirá em ganhos financeiros para o projeto. De acordo com Lamberg (2011), programas geometúrgicos levam a uma melhor definição dos limites do que é minério e do que não é, considerando a performance metalúrgica como fator auxiliar para tomar esta decisão. Graças a geometurgia, mineralizações que já haviam sido descobertas e classificadas como não-econômicas poderiam passar a ser, além de ser possível decidir com maior rapidez e confiança sobre abandonar um depósito visto como não rentável, evitando gastos desnecessários com sondagens e análises (Dunham, Vann e Coward, 2011). Projetos e operações que contam com o apoio da geometurgia tem maior velocidade de resposta e melhores oportunidades de reduzir a duração de eventos não planejados (Dunham e Vann, 2007), o que por si só já gera uma vantagem financeira em relação às demais operações que não contam com esta ferramenta, aumentando a competitividade no mercado.

É importante ressaltar que nem sempre o produto da geometalurgia será um recurso maior do que quando comparado com o recurso no caso base, onde o modelo geometalúrgico não é utilizado. O que existirá é uma maior certeza relacionada ao valor reportado, o que permite um planejamento econômico mais aderente a realidade do depósito, como colocado por Keeney e Walters (2011). Dunham e Vann (2007) colocam que quanto maior o conhecimento, menor é a incerteza na produção e menor é a probabilidade de uma nova operação falhar ao ignorar fatores que afetam o desempenho do processo. É possível enxergar esses fatores de desempenho como parte dos fatores modificadores apontados pelo código JORC, cuja definição é “fatores modificadores são considerações utilizadas para converter Recursos Minerais em Reservas de Minério. Esses fatores incluem, mas não se restringem a mineração, processo, metalurgia, infraestrutura, economia, marketing, aspectos jurídicos, ambientais, sociais e fatores governamentais” (JORC, 2012).

Analisando a lista de fatores apontada pelo JORC e suas correspondentes disciplinas, é possível perceber que a geometalurgia é capaz de contribuir para aumentar o conhecimento relacionado a vários deles. A Figura 5 ilustra o quadro de conversão de recursos para reservas apresentado no código em questão.

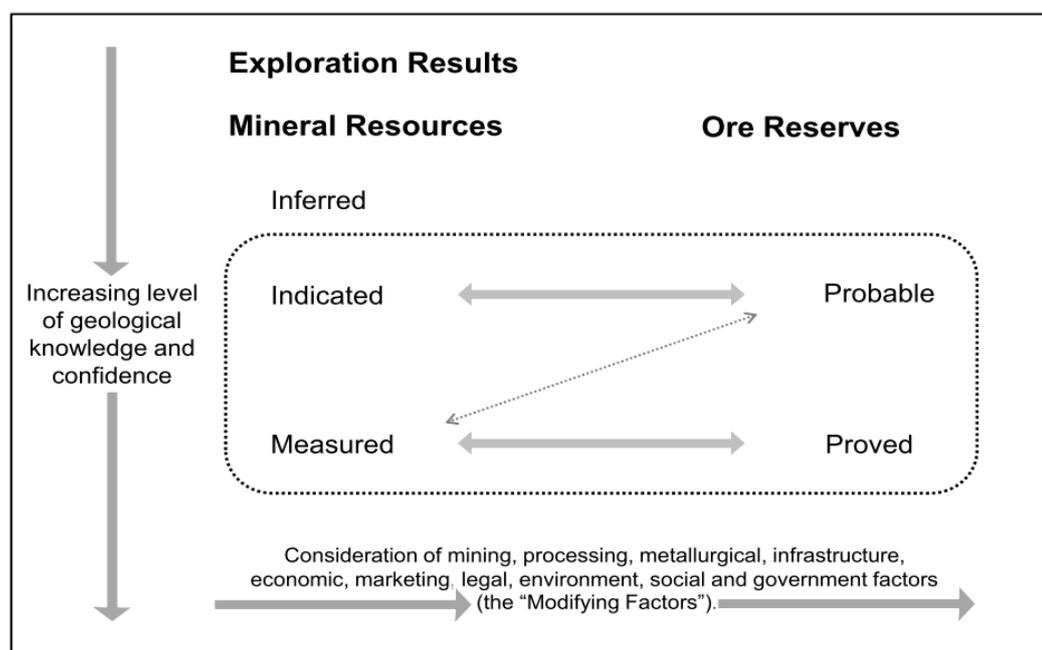


Figura 5 - Quadro ilustrativo da conversão de recursos para reservas. Fonte: JORC (2012)

Analisando a imagem, é possível perceber que os recursos minerais classificados como indicados são diretamente atribuídos a classe de reservas prováveis, enquanto os recursos classificados como medidos podem pertencer tanto às reservas provadas quanto às prováveis. Essa atribuição dos recursos medidos a uma das duas categorias de reservas está associada a influência dos fatores modificadores. Caso haja uma incerteza relacionada a um ou mais fatores modificadores, o profissional competente poderá classificar os recursos medidos como reservas prováveis, a fim de adotar uma postura mais conservadora na classificação.

De acordo com o próprio código, os fatores modificadores devem, na maioria das vezes, ser estimados com a contribuição de conhecimento vindo de uma variedade de disciplinas, o que é pré-requisito para elaboração de um bom estudo geometalúrgico. A geometalurgia, portanto, pode ser vista como a ciência que ajuda a mensurar, de forma mais acurada, a influência de alguns fatores modificadores na jazida mineral, auxiliando na redução de riscos de implantação e operação. Uma vez que a declaração de reservas na bolsa de valores está associada à captação de investimentos do mercado, percebe-se mais uma vez os benefícios financeiros da implantação de um modelo geometalúrgico. Conclui-se, portanto, que a geometalurgia é uma prática que traz inúmeros benefícios para a operação mineira na qual é implementada, de modo que sua presença deveria ser considerada indispensável.

2.5. FERRAMENTAS QUE AUXILIAM NA ELABORAÇÃO DE UM MODELO GEOMETALÚRGICO

2.5.1. Matriz geometalúrgica

A matriz geometalúrgica, proposta por Williams e Richardson (2004), é uma ferramenta muito útil para auxiliar na etapa de seleção de amostras geometalúrgicas, que segundo os próprios autores, é o primeiro e mais crítico passo do mapeamento geometalúrgico. Um exemplo da utilização desta metodologia pode ser visto em Fernandes (2013).

Dois eixos desta matriz, que é tridimensional, carregam fatores geológicos: o tipo de rocha e a alteração. O terceiro eixo está relacionado a parâmetros críticos que podem facilitar ou dificultar o processamento do minério. Exemplos de parâmetros críticos podem ser: a presença de um mineral que dificulta a flotação; uma região do depósito conhecida por

demandar maior energia para atingir o grau de liberação; a presença abundante de um mineral que traz bônus financeiro ao produto final comercializado; ou ainda, a presença de minerais que podem contaminar o meio ambiente se forem descartados no rejeito. É importante que toda a equipe envolvida esteja reunida durante a elaboração da matriz geometalúrgica, pois ela deve contemplar todas as características relevantes do depósito.

Dependendo da complexidade geológica do depósito, a matriz geometalúrgica pode facilmente apresentar mais de 50 cruzamentos, o que na prática é inviável de ser amostrado, analisado e submetido a testes de processo, ainda mais quando consideramos que cada cruzamento deve conter pelo menos 10 amostras distintas. Ao considerar que dificilmente todos os cruzamentos são realmente observados no depósito, inicia-se a redução do número de cruzamentos que realmente devem ser estudados. Além disso, Williams e Richardson (2004) sugerem que categorias economicamente insignificantes ou geograficamente muito pequenas possam ser desprezadas, fazendo com que o estudo seja focado nas zonas mais críticas e representativas do depósito. Outra abordagem é o agrupamento de cruzamentos com características similares. Desta forma, o orçamento disponível para a realização do estudo geometalúrgico pode ser melhor distribuído entre os cruzamentos relevantes para o depósito, tanto no âmbito geológico quanto no econômico.

2.5.2. Análise de componentes principais e análise de grupos (clusters)

Após a amostragem e execução dos testes de processo nas amostras de cada um dos cruzamentos relevantes da matriz geometalúrgica, o próximo passo é a definição dos domínios geometalúrgicos, nos quais serão baseados todos os cálculos estatísticos das variáveis de processo.

Os domínios geometalúrgicos nem sempre estão relacionados com o domínio geológico (Mc Quiston e Bechaud, 1968; Bye, 2011; Lamberg, 2011; Keeney e Walters, 2011; David, 2013; Walters, 2008), seja porque um mesmo tipo geológico apresenta comportamentos metalúrgicos distintos, seja porque tipos geológicos distintos apresentam comportamentos metalúrgicos similares. Várias técnicas de estatística multivariada podem ser utilizadas para auxiliar o pesquisador na definição dos domínios em questão, sendo as mais utilizadas para este fim a análise de componentes principais (Pearson, 1901) e a análise de cluster (Mingoti, 2005). Ambas metodologias admitem como variáveis de *input* análises químicas e mineralógicas,

medidas de variáveis físicas (como densidade e RQD, por exemplo), além de variáveis de processo.

Tendo em vista que a análise de componentes principais permite a redução de dimensionalidade do problema estudado para p componentes, é facilitada a identificação de grupos de amostras com comportamento similar entre si, desde que as componentes retidas para análise carreguem consigo boa parte da variabilidade total das variáveis em estudo. Exemplos da utilização de componentes principais na identificação de domínios geometalúrgicos podem ser lidos em Lopera (2014), que elaborou componentes principais a partir da mineralogia e da densidade, visando determinar grupos para construção de modelos de regressão para predição de $BMWi$ e $A*b$. Keeney e Walters (2011) apresentam um exemplo de componentes principais calculadas com base na mineralogia com o objetivo de identificar as principais tendências mineralógicas e associações não evidentes no banco de dados original, sendo que estas mesmas componentes foram utilizadas no estudo de Newton e Graham (2011) para auxiliar na determinação de domínios para predição de $A*b$ e $BMWi$.

A análise de cluster, por sua vez, agrupa amostras similares através da análise de múltiplas variáveis simultaneamente, formando grupos estatisticamente homogêneos em seu interior e heterogêneos entre si. A aplicação da análise de clusters neste contexto pode ser vista em Louwrens, Napier-Munn e Keeney (2015), onde os autores utilizaram esta técnica combinada com componentes principais para determinar domínios para elaboração de um modelo geometalúrgico para a barragem de rejeitos de uma mineradora australiana.

É importante lembrar que nenhuma das técnicas supracitadas tem como imposição a formação de grupos espacialmente contínuos, o que deve ser verificado e validado posteriormente (Newton e Graham, 2011). Outro ponto que deve ser observado durante a elaboração de domínios é a necessidade de que cada um tenha dados suficientes para permitir que as distribuições estatísticas e variogramas sejam estimados de forma razoável (Deutsch *et al.*, 2015). Caso o número de amostras seja insuficiente em um ou mais domínios, é prudente avaliar a possibilidade de novos agrupamentos, tentando manter ao máximo a similaridade entre os domínios agrupados.

Finalizada a definição dos domínios geometalúrgicos, o próximo passo é modelá-los espacialmente. Newton e Graham (2011) apresentam quatro possíveis métodos para modelar domínios: interpretando seções e construindo sólidos; utilizando vizinho mais próximo; utilizando krigagem da indicatriz; e utilizando análise de tendência estocástica, método baseado

na análise de séries temporais cujos detalhes encontram-se no texto citado. De acordo com os testes realizados pelos autores nos dados por eles avaliados, a krigagem da indicatriz apresentou melhores resultados quando comparada com a técnica de vizinho mais próximo. A análise de tendência estocástica apresentou bons resultados, mas segundo eles necessita de mais investigação, enquanto a interpretação de seções mostrou-se uma tarefa difícil e laboriosa. Keeney e Walters (2011) e Bye (2011) também sugerem a krigagem da indicatriz como uma abordagem simples e objetiva para estimar domínios geometalúrgicos no modelo de blocos.

2.5.3. Curva geometalúrgica

A curva geometalúrgica, idealizada por Vann *et al.* (2011) e ilustrada na Figura 6, apresenta um guia do que deve estar contemplado em cada fase do estudo geometalúrgico e divide o mesmo em 7 etapas fundamentais e sequenciais. As 7 etapas, por sua vez, são divididas em 4 estágios de maturidade geometalúrgica. O avanço em cada uma das etapas implica em um aumento do conhecimento e uma redução do risco, associado a um aumento de custo financeiro.

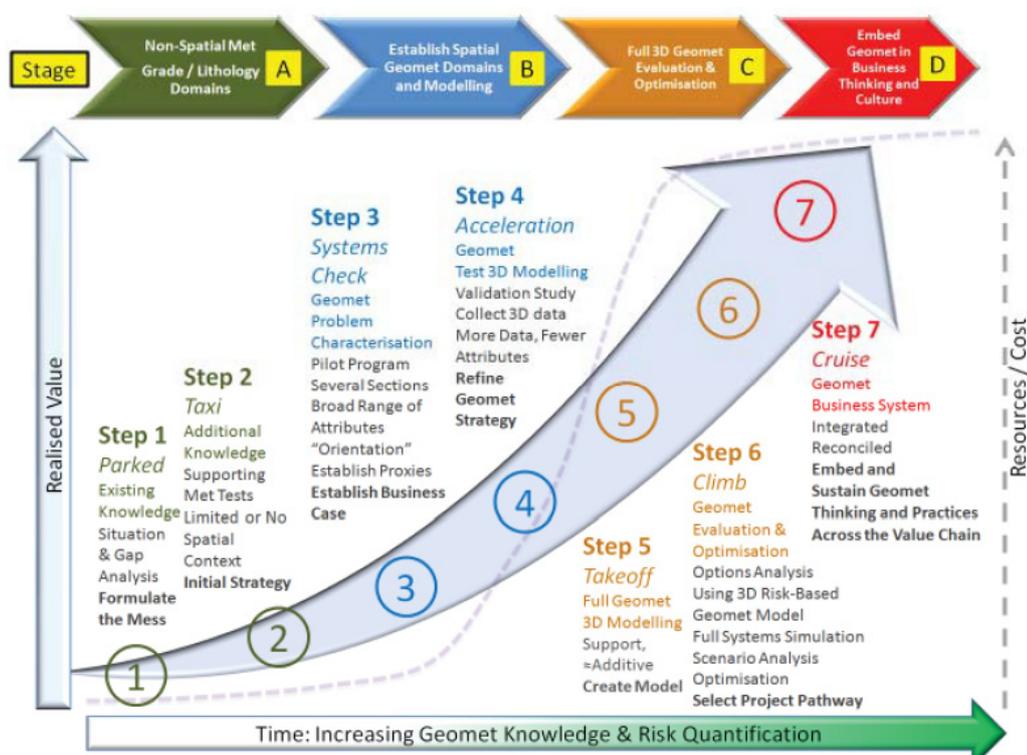


Figura 6 - Curva Geometalúrgica. Fonte: Vann *et al.* (2011)

O estágio de maturidade A, corresponde ao cenário que normalmente é visto nos depósitos, ou seja, existe um modelo de recursos com domínios geológicos definidos e teores estimados, mas não há uma associação destes com respostas metalúrgicas.

A primeira etapa da curva geometalúrgica, a qual está inserida no estágio A, consiste em acessar e organizar toda a informação e conhecimento existente sobre o depósito em estudo. Neste momento, inicia-se a tentativa de compreender quais fatores afetam o processo de beneficiamento.

Na segunda etapa, novas amostras são coletadas e testes de processo são realizados, mas em um contexto espacial limitado, visando descobrir de uma forma geral quais são os fatores que impactam no processo. Tendo uma ideia geral da resposta, é possível programar testes considerando uma maior abrangência espacial.

No estágio B, são determinados: a distribuição espacial dos dados geometalúrgicos; os domínios geometalúrgicos existentes; e como os dados serão utilizados no modelamento 3D.

Durante a etapa 3, englobada no estágio B, é estabelecido um programa piloto de geometalurgia, no qual busca-se identificar correlações entre as variáveis de processo (cuja obtenção é onerosa e trabalhosa) e as variáveis cuja obtenção da informação é mais fácil e acessível financeiramente (como a análise química, por exemplo). O ideal é que diversas variáveis sejam analisadas nesta fase do estudo, para que estas correlações possam ser percebidas. O principal objetivo desta etapa é desenvolver uma estratégia de amostragem e de execução de testes padrão. É possível que nessa etapa domínios geometalúrgicos já sejam identificados.

A quarta etapa tem como objetivo estabelecer a construção de um banco de dados geometalúrgico 3D, com base no programa estabelecido no passo anterior. Os custos dessa fase são altos devido a grande quantidade de dados coletados, analisados e submetidos aos testes de processo. As variáveis consideradas importantes no passo 3 deverão ser coletadas em vários locais, tentando cobrir toda a mineralização. As variáveis que não mostraram nenhuma relação com a resposta do minério no processo não são coletadas. Nesta etapa, são definidos os domínios geometalúrgicos espaciais.

O estágio de maturidade C envolve a elaboração do modelo geometalúrgico 3D com todos os dados coletados até o momento, sua utilização e otimização.

Na etapa 5, os dados coletados até então são utilizados para elaborar o modelo geometalúrgico 3D que será utilizado pelo planejamento de mina e de processo. A metodologia que deve ser utilizada para estimar a resposta metalúrgica bloco a bloco não é determinada pelos autores, cabendo ao pesquisador decidir sobre tal questão.

O estágio D, envolve a conscientização do corpo técnico e dos gestores da companhia sobre a importância da geometalurgia, de modo que ela faça parte da cultura da empresa como uma ferramenta que agrega conhecimento e valor para o negócio.

Na sétima e última etapa, a geometalurgia passa a ser considerada uma abordagem padrão e os líderes passam a enxergar sua importância na redução de incertezas. O modelo geometalúrgico deve ser reconciliado com a performance da usina, permitindo assim a avaliação de possíveis vieses e dando a oportunidade de novas melhorias serem implementadas, de forma que ele se torne cada vez mais robusto.

O modelo geometalúrgico elaborado na mina de Batu Hijau, apresentado por Wirfiyata e McCaffery (2011), mostra a importância da atualização rotineira do mesmo, seja por motivos operacionais, como mudança de equipamentos ou ajustes no processo, seja devido a mudança geológica do material lavrado com o avanço da mina. Nessa mina, a sala de controle recebe uma previsão do *throughput* e da recuperação do minério 3 horas antes dele ser alimentado na planta, o que permite aos operadores e supervisores realizarem ajustes na estratégia de processamento do material, levando a uma melhor resposta do mesmo. Essa sintonia entre mina e processo só é possível graças a constante atualização do modelo geometalúrgico, fazendo com que ele esteja sempre aderente a resposta do minério no processo de beneficiamento.

É importante ressaltar que o conhecimento e a redução de riscos não estão associados a elaboração do modelo geometalúrgico, mas sim a sua utilização no dia a dia da mina. Sendo assim, o real valor da geometalurgia só é atingido na sexta etapa, de modo que todo estudo de geometalurgia deve ser programado para atingir ao menos este nível. Completar as sete etapas da curva geometalúrgica traz como benefício o apoio da liderança, que é fundamental para que o modelo continue em evolução constante, se tornando cada vez mais aderente a realidade da mina estudada.

2.6. DESAFIOS DURANTE A ELABORAÇÃO DE UM MODELO GEOMETALÚRGICO

2.6.1. Não-aditividade do atributo estudado

Um dos maiores desafios relacionados ao tratamento de dados geometalúrgicos é a não-aditividade de muitas variáveis inseridas neste contexto (Richmond e Shaw, 2009). De acordo com Coward *et al.* (2009), a aditividade é a propriedade que permite que a média de algumas variáveis seja calculada por meio de uma simples média linear, ou como descrito por Carrasco *et al.* (2008), uma quantidade é dita aditiva se a quantidade média é igual a média das quantidades.

A não-aditividade de um atributo faz com que técnicas que utilizam de médias lineares para realizar suas previsões gerem resultados enviesados ou incoerentes com o fenômeno estudado. Logo, a krigagem que é o método de estimativa espacial mais utilizado no meio geoestatístico é uma das técnicas que não deve ser utilizada para tratar variáveis com esta característica, visto que essa baseia suas previsões em médias ponderadas que assumem aditividade do atributo que está sendo estimado (Dunham e Vann, 2007; Bye, 2011).

Para compreender como a não-aditividade afeta o cálculo dos valores médios, é necessário compreender o artigo publicado pelo matemático dinamarquês Johan Jensen em 1905, intitulado “Sobre as funções convexas e a desigualdade entre os valores médios”. Neste trabalho, são apresentadas várias desigualdades, mas a mais importante para o desenvolvimento matemático da época é a denominada desigualdade de Jensen (Jensen, 1905).

Sabe-se que uma função $\varphi(x)$ é convexa, se sua segunda derivada é maior do que zero e é côncava caso sua segunda derivada é menor do que zero. Seja $\varphi(x)$ uma função convexa. Pela desigualdade de Jensen temos que:

$$\varphi\left(\frac{1}{n}\sum_{i=1}^n x_i\right) \leq \frac{1}{n}\sum_{i=1}^n \varphi(x_i). \quad (1)$$

Interpretando a equação 1 temos que ao aplicar o valor médio de x na função convexa $\varphi(x)$ sempre teremos um valor menor ou igual a média dos valores da função $\varphi(x)$ aplicada em cada valor individual de x . Como o negativo de uma função convexa é uma função côncava, basta inverter o sinal da equação para obter a desigualdade de Jensen para funções côncavas.

Em resumo:

$$\text{Se } \varphi(x) \text{ é côncava: } E(\varphi(x)) \leq \varphi(E(x));$$

$$\text{Se } \varphi(x) \text{ é convexa: } E(\varphi(x)) \geq \varphi(E(x));$$

Se $\varphi(x)$ é linear, a desigualdade se torna uma igualdade, ou seja: $E(\varphi(x)) = \varphi(E(x))$.

Para visualizar gráfica e numericamente a desigualdade de Jensen, basta avaliar o exemplo a seguir. Seja x uma variável com 20 observações variando de 1 a 20 e seja $\varphi(x) = \ln(x)$, que é uma função côncava. Considere os seguintes resultados:

- A média das observações da variável x é 10.50 e o logaritmo natural deste valor médio é 2.35;
- Ao aplicar o logaritmo natural em cada valor de x e retirar a média dos resultados tem-se como resposta 2.12;
- Portanto, a desigualdade de Jensen para funções côncavas foi satisfeita, pois $2.12 \leq 2.35$.

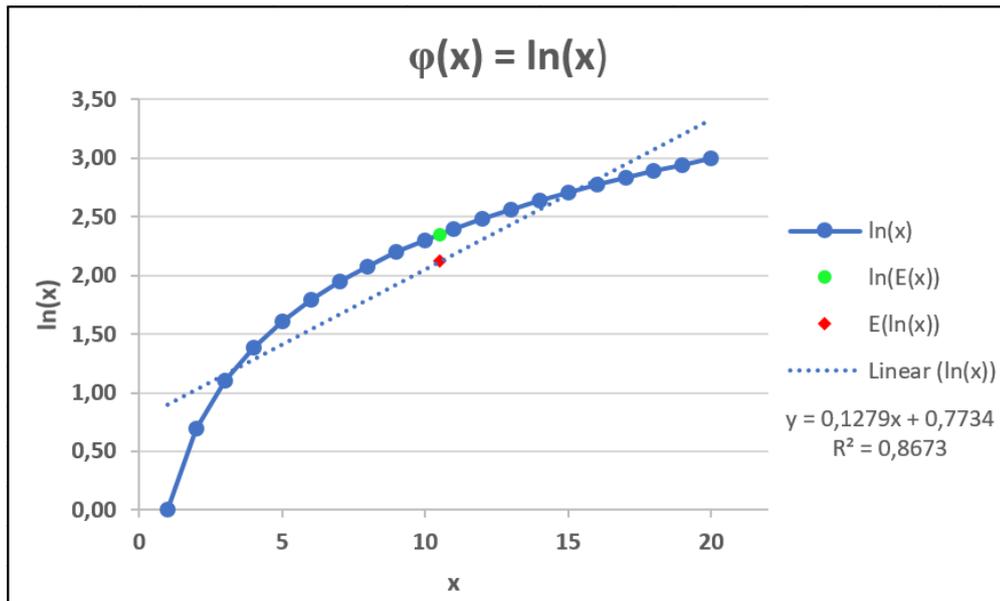


Figura 7 - Demonstração gráfica da Desigualdade de Jensen - Função Côncava $\ln(x)$

Analisando a Figura 7, que reflete os resultados apresentados anteriormente, percebe-se que o valor representado em verde ($\ln(E(x))$) corresponde a um valor médio aderente a função $\ln(x)$. O valor vermelho ($E(\ln(x))$), por sua vez, é aderente a reta azul pontilhada, que corresponde ao ajuste linear das observações x na função $\ln(x)$. Conclui-se, portanto, que aplicar o logaritmo natural em cada valor x e calcular a média dos resultados gera valores médios viesados, relacionados ao ajuste linear dos dados em questão, o que gera uma subestimativa.

Considere agora um exemplo de função convexa. Seja x a mesma variável definida anteriormente e $\varphi(x) = x^2$. Analisando os resultados tem-se que:

- A média das observações da variável x é 10.50 e sua segunda potência é 110.25;
- Ao elevar ao quadrado cada valor de x e retirar a média dos resultados tem-se como resposta 143.50;
- Analisando a desigualdade de Jensen para funções convexas vê-se que essa foi satisfeita, pois $143.50 \geq 110.25$.

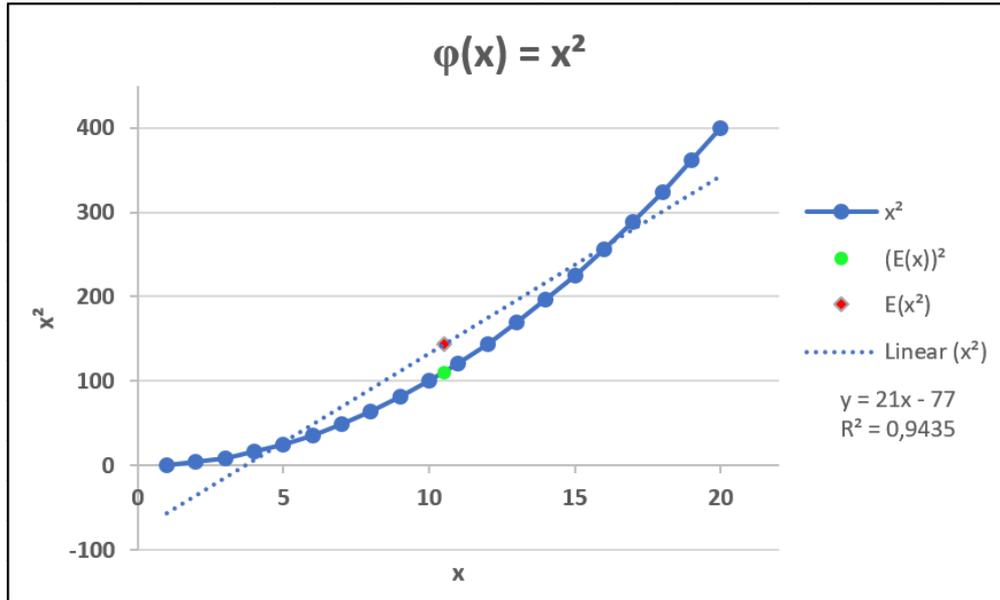


Figura 8 - Demonstração gráfica da Desigualdade de Jensen - Função Convexa x^2

A Figura 8 ilustra os resultados apresentados. Da mesma forma que na função côncava, a função convexa mostra uma aderência do ponto verde ($(E(x))^2$) à função $\varphi(x) = x^2$, enquanto o ponto vermelho ($E(x^2)$) adere ao ajuste linear dos dados. Logo, uma superestimativa do valor médio é obtida quando uma função convexa é aplicada aos dados e sua esperança é calculada a partir dos resultados desta função.

Carrasco *et al.* (2008) apresentam a relação das variáveis metalúrgicas com a aditividade, para demonstrar que a recuperação mássica é aditiva, mas a recuperação metalúrgica não é e por isso deve ser tratada com cautela durante a estimativa. Considere as seguintes equações, retiradas no artigo supracitado:

$$T_H = T_C + T_T, \quad (2)$$

$$Q_H = Q_C + Q_T, \quad (3)$$

onde T_H é a massa total alimentada, T_C é a massa do concentrado e T_T é a massa do rejeito e onde Q_H é o metal contido na alimentação, Q_C é o metal contido no concentrado e Q_T é o metal contido no rejeito.

Seja Z_H o teor da alimentação, Z_C o teor do concentrado e Z_T o teor do rejeito. Uma vez que o metal contido é o produto da massa pelo teor, a equação 3 pode ser reescrita como:

$$Z_H T_H = Z_C T_C + Z_T T_T \quad \text{ou} \quad Z_H = \frac{Z_C T_C + Z_T T_T}{T_H}. \quad (4)$$

A recuperação mássica nada mais é do que a quantidade de massa do concentrado dividido pela massa total alimentada, ou seja:

$$r = \frac{T_C}{T_H}. \quad (5)$$

Substituindo a equação 5 na equação 4 e isolando T_T temos que a recuperação mássica pode ser escrita como função dos teores, conforme apresentado na equação 6:

$$\begin{aligned} Z_H &= Z_C r + \frac{Z_T (T_H - T_C)}{T_H} \\ Z_H &= Z_C r + Z_T (1 - r) \\ Z_H &= Z_C r + Z_T - Z_T r \\ Z_H - Z_T &= r (Z_C - Z_T) \\ r &= \frac{Z_H - Z_T}{Z_C - Z_T}. \end{aligned} \quad (6)$$

A recuperação metalúrgica, por definição, é o metal contido no concentrado em relação ao metal contido total, o que faz com que essa possa ser escrita como:

$$R = \frac{Q_C}{Q_H} = \frac{Z_C T_C}{Z_H T_H}. \quad (7)$$

Uma vez que a recuperação mássica é a razão da massa do concentrado pela massa de alimentação, é possível substituí-la na equação 7, de modo que a recuperação mássica possa ser escrita em função da recuperação metalúrgica e vice-versa, conforme descrito na equação 8:

$$R = \frac{Z_C}{Z_H} r$$

$$R Z_H = Z_C r. \quad (8)$$

O resultado da equação 8 é definido por Carrasco *et al.* (2008) como sendo o teor recuperável *in situ*, denotado por Z_R , que de acordo com os autores, é uma variável aditiva.

$$Z_R = R Z_H = Z_C r. \quad (9)$$

O teor no concentrado também é uma variável não aditiva. Como explicado por Cornah (2013) e Carrasco *et al.* (2008), esta variável seria aditiva caso a amostra coletada para análise química fosse retirada diretamente do concentrado. Como a realidade é outra e a amostra é originária de uma amostra maior, coletada *in situ*, que é submetida a uma rota de processo, gerando massas de concentrado diferentes de um furo para outro, o cálculo direto do teor do médio de concentrado não pode ser efetuado. O mesmo raciocínio apresentado acima é válido para o teor do rejeito.

Após compreender as características de uma variável não-aditiva, o próximo passo é descobrir qual método deve ser utilizado para predizê-las de forma coerente, pois tratar variáveis com esta natureza pode se tornar algo complexo visto que técnicas de estimativa convencionais não se aplicam. Segundo Deutsch *et al.* (2015), “abordagens determinísticas são preferidas para variáveis lineares onde a estimativa de um único valor ótimo é requerida, mas deve ser evitada para variáveis geometalúrgicas e geomecânicas que não podem ter suas médias calculadas linearmente”. O mesmo autor sugere que uma opção para prever propriedades metalúrgicas é a utilização de simulação, visto que suposições de linearidade da variável não são assumidas neste caso.

Bye (2011) também sugere a utilização da técnica de simulação condicional para simular variáveis não aditivas, uma vez que são simulados pontos e que o conjunto de pontos dentro de um dado volume permite o cálculo do histograma e dos quartis da distribuição no dado local, o que nos fornece um valor aproximado do atributo no volume em questão sem que o cálculo de médias seja efetuado. Deutsch (2013) cita três fortes motivos para considerarmos a simulação em estudos geometalúrgicos: o primeiro deles é o fato da simulação conseguir tratar de forma não enviesada as variáveis não-aditivas; o segundo, está relacionado a resultados realistas obtidos com simulação no que diz respeito à variabilidade em pequena escala; e o último é a possibilidade de avaliar a incerteza em grandes volumes a partir da incerteza local determinada através da simulação, o que leva a tomadas de decisão mais conscientes no âmbito técnico e econômico.

Shaw, Khosrowshahi e Weeks (2013), também defendem o uso da simulação na modelagem de variáveis geometalúrgicas, não só pelo fato da não-aditividade relacionada a estas variáveis, mas também porque a simulação condicional nos fornece meios para conhecer a variabilidade do minério que será alimentado, seja na escala diária, semanal, mensal ou anual, o que não é possível de se obter através das técnicas frequentemente aplicadas que suavizam as estimativas, como a krigagem, por exemplo. Conhecer a variabilidade permite que o planejamento de lavra se prepare para eventuais problemas que possam ocorrer e que ações corretivas possam ser traçadas com antecedência e de forma efetiva. Um exemplo desta atuação preventiva pode ser a blendagem do minério de determinada semana para evitar que a massa alimentada tenha uma resposta inferior a esperada pela usina.

Um exemplo da utilização de simulação para prever o comportamento de variáveis geometalúrgicas pode ser visto em: Boisvert *et al.* (2013), que utilizaram simulação sequencial gaussiana combinada com componentes principais, análise de regressão e demais técnicas de geoestatística multivariada para construir um modelo geometalúrgico para prever a recuperação de Cu e U_3O_8 , consumo de ácido, *Net Recovery*, DWT (*Drop Weight Test*) e BMWi (*Bond Mill Work Index*) na mina australiana de Olympic Dam. Outro exemplo encontra-se em Cornah (2013), que utilizou cosimulação sequencial direta para garantir que a correlação entre o teor no concentrado, teor de alimentação e recuperação mássica fosse mantida. Utilizando os dados simulados bloco a bloco, o autor efetuou os cálculos direto e indireto da média para prever a recuperação metalúrgica e conseguiu avaliar a grandeza do viés existente entre as duas formas de cálculo. Com isso, foi possível verificar que calcular a média aritmética de

dados não aditivos pode não ser tão impactante em termos globais, mas pensando localmente é possível chegar a vieses de aproximadamente -10% para o depósito avaliado.

Outra solução apresentada por Deutsch (2013) é compreender a “regra de mistura” da variável em estudo. Isso só pode ser obtido através da observação de várias amostras cuja variável de interesse é conhecida, sendo que estas deverão ser combinadas e submetidas a avaliação da variável de interesse na combinação, de modo que se compreenda como as proporções afetam o resultado final. Compreendendo a regra por trás da variável estudada, pode ser possível realizar uma transformação não linear da mesma que permita o cálculo da média aritmética e por consequência, a utilização de krigagem ordinária. Entretanto, esta abordagem pode ser muito onerosa e pode não gerar os resultados esperados.

Vann *et al.* (2011) afirmam que “se um forte argumento puder ser utilizado de que a variável resposta pode ser considerada em um range de valores para os quais o comportamento é aproximadamente linear, então modelar a variável resposta diretamente pode ser admissível”. Logo, é possível utilizar metodologias tradicionais de estimativa que desconsideram a não-aditividade do atributo, mas é importante ressaltar que os resultados obtidos utilizando esta abordagem podem não ser os melhores. Na visão de Deutsch (2013), variáveis não-aditivas podem ser krigadas para fins de visualização de tendências e para a compreensão de características espaciais, mas os resultados obtidos não devem ser utilizados para fins de previsão, uma vez que existirão vieses provenientes da natureza da variável e da interação desta com um interpolador escolhido.

Richmond e Shaw (2009) sugerem a utilização de krigagem da indicatriz para estimar variáveis geometalúrgicas cujo comportamento não é aditivo, de modo que seja modelada a probabilidade da variável em questão exceder um dado cut-off e não a variável propriamente dita.

Dunham e Vann (2007) colocam uma importante questão durante a elaboração de estudos geometalúrgicos, que é identificar o que realmente se deseja obter com o modelo elaborado, para a partir deste ponto determinar qual a metodologia que trará os melhores resultados. Os autores citam a variável RQD (rock quality designation), que não é geometalúrgica, mas geotécnica e cujo comportamento não-aditivo serve para ilustrar o ponto de vista dos mesmos. Ao elaborar um modelo para essa variável, o mais importante é determinar regiões nas quais baixos valores de RQD são observados, o que faz com que estimativas baseadas na média ponderada gerem resultados insatisfatórios, visto que essas suavizam os

resultados. Em casos assim, a utilização da krigagem da indicatriz pode ser uma melhor abordagem, por apontar a probabilidade de um determinado bloco estar acima ou abaixo de um dado valor e não o valor em si.

Trazendo a situação acima para o caso geometalúrgico, pode-se pensar no impacto que determinados minerais têm no processo de flotação, ainda que a mineralogia seja uma variável aditiva (Coward *et al.*, 2009) e permita estimativa através de krigagem ordinária. Células de flotação são construídas visando um material alimentado com características mineralógicas que pouco variam ao longo do tempo. A presença de minerais que não foram planejados para alimentar a planta pode gerar problemas na recuperação, entupimento de canos, transbordo excessivo e até mesmo corrosão de partes do equipamento. Portanto, uma característica geometalúrgica importante de ser monitorada é a probabilidade de dado bloco exceder uma determinada porcentagem do mineral em questão, o que pode ser obtido através da krigagem da indicatriz.

Outro exemplo pode ser a elaboração de um modelo geometalúrgico que expresse a resistência do material na etapa de cominuição, utilizando como base ensaios de BMWi (*Ball Mill Work Index*). A variável em questão não é aditiva. Neste caso, mais importante do que determinar o BMWi médio de um bloco é determinar regiões onde há uma maior resistência do material, o que pode gerar problemas operacionais no moinho e até quebras do equipamento. Neste exemplo, assim como nos dois exemplos anteriores, utilizar a krigagem da indicatriz para determinar a probabilidade de um bloco possuir resistência maior do que a suportada pelo moinho pode ser uma saída simples, mas funcional.

Todas as soluções acima apresentadas consideram que o número de dados geometalúrgicos disponível é adequado para utilizar técnicas geoestatísticas de simulação ou estimativa. Entretanto, a maioria dos estudos geometalúrgicos conta com um número reduzido de testes de processo, o que impossibilita a utilização de tais técnicas na prática. Uma alternativa é a criação de uma função de transferência: um modelo numérico capaz de conectar as propriedades da rocha com as variáveis de processo (Deutsch, 2013). A literatura mostra a utilização de diversas técnicas com o objetivo de estabelecer uma função de transferência para geometalurgia. A mais comumente utilizada é a análise de regressão, uni ou multivariada, seguida pelas técnicas de árvore de decisão, *random forest*, *support vector regression*, *alternate conditional expectation (ACE)*, entre outras. Devido ao reduzido número de amostras de processo disponíveis em ambos depósitos estudados nesta tese, optou-se por estimar as variáveis geometalúrgicas de interesse através da utilização de uma função de transferência.

Conclui-se, portanto, que além de verificar se a variável geometalúrgica de interesse é aditiva, é importante avaliar o número de dados disponíveis e as limitações que isso traz ao estudo, além de determinar qual a resposta se deseja obter ao elaborar o modelo geometalúrgico, para que as predições sejam acuradas e úteis.

2.6.2. Trabalho em equipe

Um dos fatores primordiais para o sucesso de um modelo geometalúrgico é o trabalho em equipe, uma vez que a elaboração de um modelo robusto passa pela integração da geologia, planejamento de mina, metalurgia e estatística. Como nenhum profissional domina todas as áreas do conhecimento, a união faz-se extremamente necessária para atingir os objetivos. Apesar de parecer algo simples de ser alcançado, a integração entre as diversas disciplinas requer o rompimento de um padrão que a décadas vem sendo praticado nas mineradoras, onde cada profissional é responsável somente por fornecer *inputs* para as demais áreas, sem que uma preocupação com o produto final exista.

Agricola (1556, apud Jackson *et al.*, 2011) comenta em seu trabalho a necessidade de um minerador conhecer outras áreas do conhecimento e estar familiarizado com outras ciências, ainda que seja mais comum encontrar especialistas em cada área isoladamente. A partir desta citação de 1556, vê-se que a integração das disciplinas é percebida como alicerce para obtenção de bons resultados a mais de quatro séculos. De acordo com Jackson *et al.* (2011), foi ao fim da primeira guerra que essa visão holística foi perdida e cada processo começou a ser tratado individualmente, fato devido principalmente ao aumento da especialização dentro da indústria mineira e ao aumento de projetos com orçamentos e tempo de execução cada vez mais apertados.

Walters (2008) elenca uma comunicação efetiva, o compartilhamento de metodologias e uma linguagem técnica comum entre as diversas disciplinas como ponto chave para reagrupar as estruturas organizacionais divididas que impedem que a integração entre as áreas. O trabalho em equipe talvez seja a única parte fundamental para o desenvolvimento de um bom estudo geometalúrgico que não necessite de nenhum investimento financeiro e ainda assim, uma equipe em total sincronia é raridade nas operações. A curva geometalúrgica, apresentada na seção 2.5.3, apresenta diversas etapas na qual muito trabalho é necessário para avançar no conhecimento metalúrgico do depósito. Percebe-se que é difícil engajar toda equipe no

pensamento geometalúrgico quando a última etapa da curva consiste em espalhar a cultura da geometalurgia em todas as áreas da empresa.

O mais importante é conscientizar os envolvidos de que todos têm um objetivo em comum, que é produzir mais e melhor, o que só pode ser alcançado quando o processo produtivo é pensado e executado de forma holística.

2.6.3. Amostragem

Segundo Williams e Richardson (2004), “a seleção de amostras é o primeiro e mais crítico passo do mapeamento geometalúrgico”. Muito cuidado deve existir durante o planejamento amostral e na coleta das amostras que sustentarão todo o estudo, pois uma falha nesta etapa pode pôr a perder toda a representatividade e validade dos modelos elaborados. A amostragem é percebida por Kittler *et al.* (2011) como o problema mais frequente durante a elaboração de um modelo geometalúrgico e as falhas mais comuns, relacionadas a este tema, encontram-se resumidas a seguir:

- Má distribuição espacial das amostras;
- Má distribuição de amostras em relação ao teor, ou seja, foca-se muito em coletar amostras em torno do teor médio do depósito ou amostras de alto teor e esquece-se de avaliar a resposta metalúrgica quando o teor de alimentação muda e atinge valores abaixo do cut-off;
- Pouca massa para realização dos testes, o que impede a repetição, seja para checar resultados duvidosos ou para variar condições de processamento. Pouca massa também impede que alguns testes que requerem maior quantidade de material sejam gerados, como é o caso dos testes de cominuição;
- Pouco número de amostras, o que impede uma correta definição de domínios geometalúrgicos, sendo definido pelos autores o número de 30 amostras por domínio como quantidade mínima a ser trabalhada;
- Blendagem de amostras, que mascara a real variabilidade das amostras e faz com que o posicionamento espacial seja perdido, dificultando o tratamento geoestatístico;
- Poucos elementos analisados visando geralmente economia financeira, o que resulta em um entendimento parcial do impacto da química no processo de beneficiamento;

- Falta de conhecimento da variabilidade inerente ao teste geometalúrgico, ou seja, ausência de replicatas para compreender a magnitude dos erros que afetam a precisão dos resultados. Os autores sugerem um percentual entre 5 e 10% de replicatas em um estudo geometalúrgico.

Dentre os pontos supracitados, um dos mais preocupantes é a composição ou blendagem de amostras para testes metalúrgicos, prática apontada por Keeney e Walters (2011) como de alto risco, uma vez que realizá-la comprometerá um dos principais objetivos da geometurgia que é inferir a variabilidade das variáveis metalúrgicas. Um fator que geralmente leva à composição de amostras é a intenção de coletar amostras mais representativas da usina em termos de massa, ainda que por maior que seja a massa coletada para o teste, essa será dezenas de vezes menor que a massa real alimentada na planta. Outro fator diretamente relacionado com a composição de amostras é o custo e o tempo necessários para execução de testes de processo, ou seja, busca-se ganhar tempo e dinheiro reduzindo o número de amostras através da junção das mesmas, sem pensar nos impactos que isso ocasionará ao projeto no futuro.

A prática de compor amostras para geometurgia gera ruídos estatísticos que fazem com que a definição de populações se torne difícil e compromete o tratamento geoestatístico por não ser possível atribuir os resultados a uma posição correta no depósito. Walters (2008) defende a ideia de que é mais eficiente estatisticamente coletar mais amostras com um pequeno volume do que coletar um pequeno número de amostras “mais precisas” de maior volume. O mesmo ponto de vista é defendido por Bulled e McInnes (2015). Esta linha de pensamento é coerente e facilmente compreendida quando pensamos que amostras de maior volume apresentam naturalmente menor variabilidade do que amostras de menor volume (aditividade de krige) e quando lembramos que o poder estatístico aumenta com o tamanho amostral.

Outro ponto que gera preocupação é a representatividade das amostras coletadas em relação ao depósito. Dunham e Vann (2007) sugerem a coleta de todos os materiais da mina, sejam eles econômicos ou deletérios, enquanto Williams e Richardson (2004) propõem a coleta de amostras em todas as zonas de minério e estilos de alteração, para garantir uma caracterização completa do depósito estudado. Isso é importante pois, materiais que atualmente parecem uma exceção na alimentação da planta podem se tornar o minério principal dentro de alguns anos. Logo, é necessário que a usina tenha ciência das características de processo de todos os materiais da mina, o que só pode ser obtido através da coleta abrangente de amostras

geometalúrgicas. Sabendo de antemão quais podem ser as adaptações necessárias no projeto original da usina, os custos com infraestrutura ficam melhor dimensionados, levando a um fluxo de caixa mais realista.

2.6.4. Teste padrão confiável

Outro problema que faz parte do cotidiano das pesquisas focadas em geometalurgia é a obtenção de um teste padrão que seja confiável e que possa ser utilizado sem alterações ao longo de todo o estudo. O nome padrão está associado a características constantes que são buscadas em todos os testes executados, como por exemplo a dosagem de coletor, tempo de flotação, percentual de sólidos, pH, tempo de condicionamento dos reagentes, etc.

Mwanga, Rosenkranz e Lamberg (2015) alertam que quantificar a acurácia dos testes padrão é uma tarefa complexa, uma vez que toda a cadeia de amostragem, preparação da amostra, o próprio teste e a análise química devem ser considerados como pontos nos quais erros podem ser agregados, comprometendo a qualidade estatística dos resultados. Para auxiliar na tarefa de quantificar a precisão e variabilidade relacionada aos testes geometalúrgicos, é possível adicionar ferramentas de QA/QC como a inserção de duplicatas cegas na etapa de flotação. Além disso, a presença de brancos de limpeza e a geração de duplicatas na etapa de preparação física e a inserção de padrões na etapa de análise química são práticas bem-vindas. Através destas ferramentas pode ser possível identificar pontos nos quais o teste não está sendo padrão, gerando resultados comprometidos.

Ainda que seja uma tarefa difícil, é importante compreender a variância dos testes, pois assim será possível estabelecer quanto da variância total depende de alguma variável independente e quanto está relacionado ao erro do teste (Williams e Richardson, 2004). Adicionar uma fonte de variabilidade aos resultados que não é devida a fatores geológicos e sim a fatores aleatórios e não controlados compromete a identificação de domínios geometalúrgicos e a construção de modelos para predição.

Para Williams e Richardson (2004) é importante que um teste padrão contenha as seguintes características:

- Ser padrão e respeitar um protocolo bem preciso;
- Ser simples para evitar uma alta variabilidade dos resultados devido ao próprio teste;
- Ser barato, pois o orçamento para estudos geometalúrgicos geralmente é curto;
- Deve representar o processo de beneficiamento que está sendo estudado;
- Deve ser compatível com dados já existentes e se não, é necessário determinar uma ponte para associar os resultados novos aos antigos;
- Deve ser prático para não tomar muito tempo da equipe.

Mwanga, Rosenkranz e Lamberg (2015) e Williams e Richardson (2004) diferem em um ponto crucial relacionado aos testes padrão: os primeiros defendem que o teste não dependa de uma única pessoa para que resultados próximos sejam obtidos em uma possível repetição, ou seja, defendem que a reprodutibilidade do teste não esteja relacionada ao operador, enquanto os outros autores defendem a necessidade de manter o mesmo operador na etapa de flotação, pois o fator humano pode adicionar uma fonte de variabilidade aos resultados.

Tendo em vista as diferentes opiniões a respeito do ponto citado, a autora deste texto acredita ser mais coerente com a realidade das operações mineiras considerar uma mudança no operador, uma vez que mão de obra é um fator em constante variação, seja devido à troca de pessoal, férias ou promoções, de modo que conseguir manter um único operador durante todo o projeto seja algo raro. Sendo assim, é necessário que todos os operadores envolvidos trabalhem de modo muito fiel aos procedimentos, visando minimizar a variabilidade relacionada ao fator humano, a qual inevitavelmente será incorporada aos resultados dos testes.

2.6.5. Softwares para armazenamento e processamento de grandes bases de dados

A seguinte citação de McCullough *et al.* (2013) mostra o outro desafio presente na geometalurgia: “com bancos de dados enormes e complexos, descobrir relações manualmente e otimizar os parâmetros de operação de forma rápida se torna inviável”. Produzir um tratamento estatístico eficiente das múltiplas variáveis envolvidas em um estudo geometalúrgico não é uma tarefa simples e exige, além de conhecimento de múltiplas disciplinas e ferramentas, a utilização de softwares capazes de processar e armazenar grandes bases de dados.

Para uma gestão segura das informações, a aquisição de um software de banco de dados é fundamental. Liebezeit *et al.* (2011) apresentam em seu artigo as dificuldades existentes em configurar as ferramentas de banco de dados existentes no Acquire para atender a todas as necessidades de um estudo geometalúrgico. A aquisição de programas para gerenciamento das informações geralmente requer um grande investimento financeiro que nem sempre é fácil de se conseguir nas mineradoras devido à orçamentos cada vez mais restritos. A necessidade de manter um profissional responsável pela organização e manutenção do banco de dados é outro fato que faz com que esta realidade esteja distante da maioria das empresas, que buscam trabalhar com equipes mais enxutas.

Os dados de ambas mineradoras usados nesta tese não estavam consolidados em uma base de dados única, fazendo com que o tempo gasto para organização dos resultados fosse considerável. A organização de tais dados em um *software* de banco de dados permitiria que mais tempo fosse gasto na análise de resultados, possibilitando inclusive o estudo de outras variáveis geometalúrgicas não consideradas neste trabalho.

3. TÉCNICAS DE ESTATÍSTICA CLÁSSICA E MACHINE LEARNING

3.1. ANÁLISE DE REGRESSÃO

Com aplicações que vão da medicina ao marketing, a análise de regressão vem se mostrando uma ferramenta com enorme campo de aplicação e com forte poder preditivo. Segundo Hair *et al.* (2009), ela é de longe a técnica de dependência mais amplamente utilizada e versátil, aplicável em cada faceta da tomada de decisões em negócios. Sua aplicação em geometurgia possibilita a criação de modelos que relacionam características do minério *in situ* com a resposta desse no processo de beneficiamento. Exemplos da aplicação desta técnica na modelagem geometúrgica podem ser vistos em MacMillan *et al.* (2011), Montoya *et al.* (2011), Fernandes (2013), Motta (2014), Vieira (2016), entre outros.

Uma vez que o objetivo deste trabalho é utilizar os resultados obtidos por meio de análise de regressão como referências para avaliar a qualidade das previsões geradas pelas técnicas de redes neurais, *random forest* e *support vector regression*, encontra-se nas seções seguintes uma revisão dos conceitos ligados a esta técnica.

3.1.1. Coeficiente de correlação

A primeira ferramenta estatística utilizada para verificar a relação da variável dependente com as variáveis independentes que compõem um conjunto de dados em estudo é o coeficiente de correlação. Segundo Triola (2008), o coeficiente de correlação linear é uma medida numérica da força da relação linear entre duas variáveis que representam dados quantitativos. O coeficiente de correlação pode ser denotado pela letra “*r*”, quando se trata de uma estatística calculada com base em uma amostra, ou por “*p*”, quando seu valor é calculado com base em todos os dados da população. Como dificilmente trabalha-se com toda a população, neste texto, este coeficiente será denotado pela letra “*r*”.

O valor do coeficiente de correlação permite concluir se existe ou não associação linear entre duas variáveis e qual o sentido desta relação. A correlação pode ser positiva, quando o

valor de uma variável aumenta e o valor da outra também aumenta, ou negativa, quando o valor de uma aumenta e o valor da outra decresce. Correlações positivas estão relacionadas a valores positivos de “r” e correlações negativas a valores negativos deste coeficiente. O coeficiente de correlação pode ser calculado através da fórmula detalhada na equação 10.

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{[\sum(x_i - \bar{x})^2]^{1/2}[\sum(y_i - \bar{y})^2]^{1/2}}, \quad (10)$$

onde n é o tamanho da amostra, x são os dados da primeira variável e y os dados da segunda, sendo que \bar{x} e \bar{y} são respectivamente suas médias.

É importante ressaltar que o coeficiente de correlação é sempre um número entre -1 (correlação negativa perfeita) e 1 (correlação positiva perfeita), sendo que correlação igual zero não indica independência e sim que não existe dependência linear entre as variáveis (Magalhães e Lima, 2007).

Calculado o valor do coeficiente de correlação, é necessário verificar se este é estatisticamente significativo ou não, ou seja, se a correlação existente entre as variáveis avaliadas realmente existe ou foi devida ao acaso. Para isso, utiliza-se o teste de hipótese para correlação. A hipótese nula do teste é de que não existe correlação linear ($\rho=0$) e a hipótese alternativa é de que existe correlação linear ($\rho \neq 0$). A estatística de teste utilizada é apresentada na equação 11.

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}. \quad (11)$$

Quando o valor obtido pela estatística de teste é inferior ao valor crítico da distribuição t com n-2 graus de liberdade para determinado nível de significância, a hipótese nula não deve ser rejeitada (Triola, 2008). Em caso contrário, conclui-se que existe uma correlação linear significativa entre as duas variáveis em estudo.

Quando uma ou mais variáveis apresentam coeficiente de correlação significativo com a variável dependente, significa que estas variáveis independentes podem ser adequadas para construir um modelo de regressão que explique a variável de interesse. Logo, o coeficiente de correlação é um primeiro indicativo de quais variáveis podem ser utilizadas para construir um modelo de regressão explicativo.

Com base no número de variáveis independentes utilizadas na equação, a análise de regressão pode ser classificada como simples, quando existe somente uma variável independente, ou múltipla (multivariada), quando duas ou mais variáveis independentes são utilizadas para prever a variável dependente de interesse. As seções seguintes apresentam com detalhe estas duas situações.

3.1.2. Regressão simples

3.1.2.1. Equação e cálculo dos coeficientes

A equação de regressão que descreve algebricamente a relação entre duas variáveis é dada por:

$$\hat{y} = b_0 + b_1x, \quad (12)$$

onde b_0 é uma estimativa do valor real do intercepto da reta (β_0), b_1 é uma estimativa do valor real da inclinação da reta (β_1) e \hat{y} é a estimativa do valor da variável dependente com base no valor de x (variável independente).

Os coeficientes b_0 e b_1 são estimados através do método de mínimos quadrados (Johnson e Wichern, 2002), o qual tem como objetivo minimizar a soma do quadrado dos resíduos (resíduo = valor observado – valor predito). As equações 13 e 14 apresentam as fórmulas para o cálculo destes coeficientes.

$$b_0 = \bar{y} - b_1\bar{x}. \quad (13)$$

$$b_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}. \quad (14)$$

3.1.2.2. Análise de variância para modelos de regressão linear simples

A análise de variância (ANOVA) é uma técnica utilizada para verificar se a equação de regressão ajustada é uma boa preditora da variável dependente. Seu cálculo é feito a partir da equação 15.

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (15)$$

$$SQT = SQR + SQE. \quad (16)$$

A equação 16 apresenta uma notação simplificada de 15, onde o termo SQT mede a variação dos valores de y na amostra, SQR mede o quanto a variável y é explicada pelo modelo de regressão ajustado e SQE mede o quanto a variável y não é explicada pelo modelo de regressão construído. Os valores desta equação são normalmente apresentados na tabela de análise de variância, ilustrada na Tabela 1.

Tabela 1 - Tabela ANOVA para modelos de regressão simples

Fonte de Variação	Graus de Liberdade	Soma de Quadrados	Quadrado Médio
Regressão	1	SQR	QMR = SQR/1
Residual	n-2	SQE	QME = SQE/(n-2)
Total	n-1	SQT	

O coeficiente de determinação R^2 é a quantidade de variação em y que é explicada pela reta de regressão (Triola, 2008). Seu cálculo pode ser feito a partir dos valores da tabela ANOVA, conforme indicado na equação 17, ou pode ser obtido elevando ao quadrado o valor

do coeficiente de correlação. Quanto mais próximo de 1, mais explicativo é o modelo construído.

$$R^2 = \frac{SQR}{SQT}. \quad (17)$$

3.1.2.3. Teste de significância dos coeficientes da regressão simples

Um fato que deve ser observado ao construir uma equação de regressão simples é a significância dos coeficientes b_0 e b_1 . Os testes de hipótese para verificação da significância destes parâmetros encontram-se apresentados a seguir.

Para testar a hipótese nula de que o intercepto da reta é igual à zero ($\beta_0 = 0$) contra a hipótese alternativa de que este é diferente de zero ($\beta_0 \neq 0$), utiliza-se a estatística de teste apresentada na equação 18.

$$t_{obs} = \frac{b_0}{EP(b_0)}, \quad (18)$$

onde,

$$EP(b_0) = \left[\frac{\sum x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2} \right]^{\frac{1}{2}} s. \quad (19)$$

O valor da estatística de teste deve ser comparado com o valor da distribuição t com n-2 graus de liberdade, em determinado nível de significância. Caso o valor encontrado da estatística de teste seja inferior ao valor tabelado da distribuição t, a hipótese nula deve ser aceita indicando, portanto, que o intercepto não é estatisticamente diferente de zero.

Para testar a hipótese nula de que a inclinação da reta é igual à zero ($\beta_1 = 0$) contra a hipótese alternativa de que esta é diferente de zero ($\beta_1 \neq 0$), a seguinte estatística de teste deve ser utilizada:

$$t_{obs} = \frac{b_1}{EP(b_1)}, \quad (20)$$

onde,

$$EP(b_1) = \frac{s}{[\sum(x_i - \bar{x})^2]^{1/2}}. \quad (21)$$

Assim como foi feito para β_0 , o valor da estatística de teste para β_1 deve ser comparado com o valor da distribuição t com n-2 graus de liberdade, em determinado nível de significância. Caso o valor encontrado seja inferior ao valor tabelado da distribuição t, a hipótese nula deve ser aceita indicando, portanto, que β_1 não é estatisticamente diferente de zero.

Quando os testes de hipótese supracitados apontam para coeficientes estatisticamente iguais a zero significa que estes não devem ser usados para fins de previsão ou explicação (Hair *et al.*, 2009). Logo, um novo modelo deve ser buscado para explicar a variável dependente de interesse.

3.1.3. Regressão múltipla

3.1.3.1. Equação e cálculo dos coeficientes

A equação de regressão que expressa uma relação linear entre uma variável dependente y e duas ou mais variáveis independentes é dada por:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \dots + b_qx_q, \quad (22)$$

sendo que a notação utilizada é a mesma apresentada para a equação de regressão simples.

Devido ao maior número de variáveis no modelo, é conveniente utilizar a notação matricial, que facilita a compreensão das operações algébricas no caso multivariado. Os coeficientes no modelo de regressão múltipla também são estimados através do método de mínimos quadrados. Após algumas operações de álgebra matricial, apresentadas em Reis (2009), conclui-se que a solução da equação 23 é o melhor estimador para os coeficientes da regressão com duas ou mais variáveis independentes.

$$b = (X^T X)^{-1}(X^T Y), \quad (23)$$

onde b é a matriz de coeficientes procurados, X é a matriz das variáveis independentes com dimensão $[n \times (q+1)]$, X^T é a transposta desta matriz, Y é o vetor da variável dependente com dimensão $[n \times 1]$ e o expoente -1 significa a inversa da matriz.

A matriz de variâncias e covariâncias do vetor b é calculada a partir da equação 24. Através da extração da raiz quadrada do j -ésimo elemento de sua diagonal principal é possível obter o EP (b_j), que será utilizado nos testes de hipótese para avaliar a significância dos coeficientes do modelo, que serão apresentados na seção 3.1.3.3.

$$Var(b) = s^2(X'X)^{-1}. \quad (24)$$

3.1.3.2. Análise de variância dos modelos de regressão múltipla

O papel da análise de variância em modelos de regressão múltipla é mais importante que nos modelos de regressão simples, visto que neste caso ela é utilizada para avaliar a significância do modelo construído, como será explicado a seguir.

Os valores da soma de quadrados para modelos de regressão múltipla são calculados através das equações 25, 26 e 27, sendo que a notação é a mesma utilizada nas demais equações deste capítulo.

$$SQT = Y'Y - n\bar{Y}^2, \quad (25)$$

$$SQR = b'X'Y - n\bar{Y}^2, \quad (26)$$

$$SQE = Y'Y - b'X'Y. \quad (27)$$

A tabela ANOVA para esta classe de modelos encontra-se apresentada na Tabela 2.

Tabela 2 - Tabela ANOVA para modelos de regressão múltipla

Fonte de Variação	Graus de Liberdade	Soma de Quadrados	Quadrado Médio
Regressão	q	SQR	QMR = SQR/q
Residual	n-q-1	SQE	QME = SQE/(n-q-1)
Total	n-1	SQT	

Através dos valores calculados na última coluna da tabela ANOVA constrói-se o teste F, utilizado para verificar a significância do modelo de regressão. A hipótese nula é que nenhum dos b_j é significativo e a hipótese alternativa é que ao menos um dos b_j é estatisticamente significativo. A estatística de teste é calculada pela equação 28.

$$F_{obs} = \frac{QMR}{QME}. \quad (28)$$

O valor obtido através da estatística de teste deve ser comparado com o valor tabelado da distribuição F com (q, n-q-1) graus de liberdade. Caso a hipótese nula seja rejeitada, os testes t

individuais devem ser calculados a fim de verificar quais coeficientes são significantes. Em caso contrário, nenhuma das variáveis em teste deverá permanecer no modelo construído, dispensando, portanto, o cálculo dos testes t individuais (Reis, 2009).

3.1.3.3. Teste de significância dos coeficientes de regressão múltipla

A partir das estimativas do vetor b e da matriz de variâncias e covariâncias descritas na seção 3.1.3.1, é possível calcular os testes t de significância individual para cada coeficiente. A hipótese nula do teste t individual é $\beta_j = 0$ e a hipótese alternativa é $\beta_j \neq 0$, onde j é o coeficiente em teste. A estatística de teste utilizada para confrontar as hipóteses é apresentada a seguir.

$$t_j = \frac{b_j}{EP(b_j)}. \quad (29)$$

O valor obtido utilizando a equação 29 deve ser confrontado com o valor crítico da distribuição t com $n-q-1$ graus de liberdade para dado nível de significância. Caso o valor calculado pela estatística de teste seja inferior ao valor crítico tabelado, conclui-se que $\beta_j = 0$ indicando, portanto, que a variável associada a este coeficiente não deve ser utilizada para fins de previsão.

3.1.3.4. Coeficiente de determinação ajustado

O coeficiente de determinação detalhado na seção 3.1.2.2 pode ser utilizado para avaliar a quantidade da variação de Y que é explicada através do modelo de regressão múltipla construído. Entretanto, sabe-se que na medida em que mais variáveis explicativas são incluídas no modelo de regressão maior é o valor de R^2 . Logo, obtêm-se um maior R^2 pela simples inserção de todas variáveis disponíveis, mas a melhor equação de regressão não usa, necessariamente, todas estas variáveis (Triola, 2008).

Sendo assim, para comparar de forma mais efetiva diversos modelos de regressão considerando o valor de R^2 , utiliza-se o R^2 ajustado, o qual leva em consideração o número de variáveis utilizadas no modelo e o tamanho amostral. A equação para o cálculo do coeficiente de determinação ajustado é:

$$R_{ajustado}^2 = 1 - \frac{(n-1)}{[n-(q+1)]} (1 - R^2), \quad (30),$$

onde n é o tamanho amostral e q é o número de variáveis independentes no modelo.

3.1.4. Suposições dos modelos de regressão

Tanto os modelos de regressão linear simples quanto os de regressão linear múltipla possuem suposições que devem ser satisfeitas para que o modelo construído seja estatisticamente válido. Esta seção trata especificamente destas suposições e de como fazer para certificar que elas estão satisfeitas para os modelos estudados.

Após avaliar a significância dos coeficientes da regressão, concluindo que todas as variáveis são estatisticamente significativas, é necessário verificar se o modelo em teste atende a todas as suposições, que são baseadas na análise dos resíduos.

3.1.4.1. Variância constante e linearidade do modelo de regressão

A suposição de variância constante, também conhecida como homocedasticidade, é verificada de forma mais direta através da análise gráfica. Os valores ajustados pelo modelo de regressão avaliado devem ser plotados em um gráfico de dispersão contra os resíduos.

Para que a suposição de homocedasticidade esteja satisfeita é necessário que o gráfico aqui descrito apresente pontos distribuídos aleatoriamente ao redor da reta de resíduo igual a zero, conforme ilustrado no quadro “a” da Figura 9. Pontos distribuídos como nos quadros “b” e “c” mostram a presença de heterocedasticidade no modelo e para corrigir esta violação

existem duas maneiras: utilizar o método de mínimos quadrados ponderados ou aplicar transformações de estabilização da variância (Hair *et al.*, 2009).

O quadro “d” mostra a violação da suposição de linearidade do modelo. Para corrigir este problema deve-se transformar a variável dependente e/ou as variáveis independentes ou inserir um termo quadrático na variável dependente.

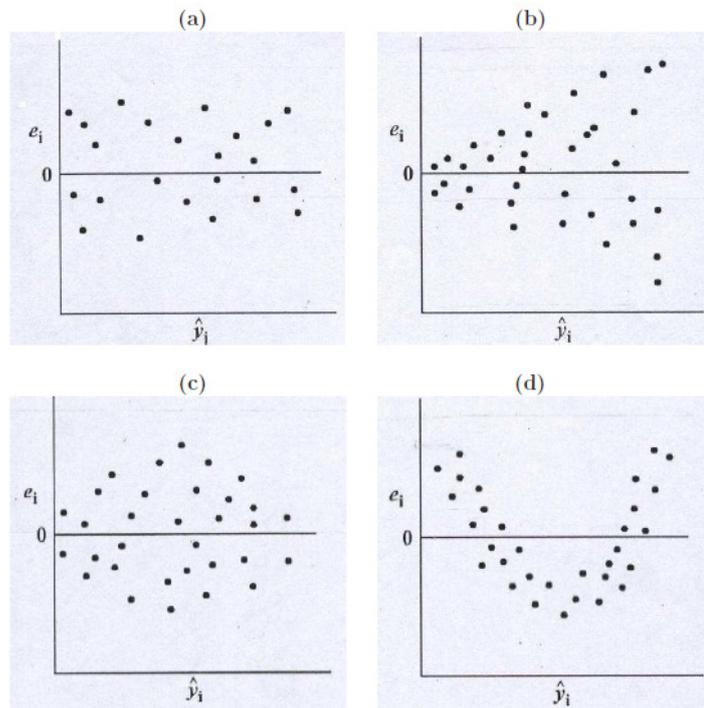


Figura 9 - Verificação da suposição de variância constante

Fonte: Reis (2009)

3.1.4.2. Normalidade dos resíduos

Para avaliar se os resíduos do modelo em estudo são normalmente distribuídos é comum a utilização do gráfico de probabilidade normal associado a um teste de normalidade. Os testes de normalidade têm como hipótese nula que a amostra analisada segue a distribuição normal e como hipótese alternativa que os dados analisados não são normalmente distribuídos. Logo, ao contrário dos demais testes de hipótese aqui apresentados, deseja-se a não rejeição da hipótese nula.

A violação da suposição de normalidade dos resíduos faz com que os testes t e F para avaliação dos coeficientes de regressão e os intervalos de confiança destes coeficientes sejam afetados, a ponto de invalidar os resultados obtidos. Entretanto, para amostras com 200 ou mais observações, esses mesmos efeitos podem ser negligenciados (Hair *et al.*, 2009), uma vez que o tamanho da amostra impacta diretamente no poder dos testes estatísticos.

Quando as suposições de linearidade do modelo e homocedasticidade são violadas, na maioria das vezes, os resíduos não são normalmente distribuídos. Logo, faz-se necessário checar primeiro estas suposições antes de verificar a distribuição dos resíduos.

3.1.4.3. Resíduos correlacionados

Para avaliar graficamente se os resíduos são correlacionados, basta plotar a ordem de coleta das observações ou o tempo em que elas foram coletadas contra os resíduos do modelo construído. Têm-se evidências da violação desta suposição caso o gráfico em questão não apresente pontos aleatoriamente distribuídos ao redor da linha de resíduo igual a zero (assim como o gráfico “a” da Figura 6).

A violação desta suposição faz com que os coeficientes obtidos através do método de mínimos quadrados deixem de ser bons estimadores, afetando a conclusão dos testes de hipótese e seus intervalos de confiança. Uma alternativa caso esta suposição seja violada é utilizar o método de mínimos quadrados ponderados ou generalizados (Reis, 2009).

3.1.5. Validação do modelo construído

Após identificar um modelo de regressão válido e que atenda a todas as suposições, o passo final é garantir que ele represente a população geral (generalização) e seja apropriado às situações nas quais será utilizado (transferibilidade) (Hair *et al.*, 2009). Para validar o modelo construído existem quatro maneiras, as quais são apresentadas a seguir.

A primeira delas é a coleta de uma nova amostra da população. Utilizando a equação construída estimam-se os valores da variável dependente para todas as observações da nova

amostra. A resposta obtida usando o modelo é confrontada com os resultados reais, e desta forma é possível verificar a aderência de novos dados ao modelo.

A segunda maneira também utiliza a nova amostra coletada, que gerará um novo modelo de regressão para explicar a variável dependente de interesse. O modelo inicial e o gerado a partir da nova amostra são confrontados para avaliar se as mesmas variáveis foram incluídas em ambos, se o sinal dos coeficientes é o mesmo e se os coeficientes de determinação são semelhantes.

Quando não é possível coletar uma nova amostra da população original, a terceira opção de validação é utilizada. A amostra é dividida em duas partes: uma delas dará origem ao modelo de regressão e a outra será utilizada para validação através de uma das duas maneiras supracitadas. O banco de dados retirado para construir o modelo em estudo geralmente é conhecido como “training set” ou “banco de treinamento” e o banco retirado para validação denominado de “test set” ou “banco de teste”.

A última maneira de validar a amostra é utilizando a estatística PRESS. Esta técnica consiste em omitir uma observação da amostra original e calcular um novo modelo de regressão. A amostra omissa tem sua previsão realizada através do novo modelo construído. O resíduo entre o valor real e o estimado pelo modelo é calculado. No total, $n-1$ modelos são construídos e os resíduos gerados são somados para fornecer uma medida geral de ajuste preditivo (Hair *et al.*, 2009). A ideia é equivalente ao teste de validação cruzada, comumente utilizado em geoestatística.

3.2. ÁRVORE DE DECISÃO

3.2.1. Árvore de decisão para classificação

3.2.1.1. Conceitos básicos

De acordo com Izenman (2008), uma árvore de decisão para classificação é o resultado de uma sequência ordenada de perguntas e o tipo de pergunta elaborada em cada etapa depende das respostas das perguntas anteriores, de modo que a sequência termina na predição de uma

classe. As perguntas são sempre do tipo $X_i \leq \theta_j$, onde X é uma variável independente e θ é um valor de corte desta variável.

Todos os dados do banco iniciam como pertencentes a um único nó, denominado nó raiz (root node). A partir deste nó, a primeira pergunta é feita, dividindo os dados em dois grupos: o nó da esquerda receberá os dados para os quais a resposta da pergunta foi afirmativa; e o nó da direita receberá os dados cuja resposta foi negativa.

Os nós localizados abaixo do nó raiz podem ser classificados como nó pai (parent node) o qual dá origem a outros dois nós filhos, ou como nó terminal (terminal node), que não gera filhos e, portanto, armazena a resposta final de classificação. Desta forma, os dados que responderam positivamente a primeira pergunta podem ou não ser submetidos a uma nova pergunta, fato que determinará se o dado será um nó terminal ou um nó pai. O mesmo vale para os dados que responderam negativamente a primeira pergunta. O exemplo apresentado na Figura 10 ilustra os conceitos apresentados acima: o nó em marrom é o nó raiz; em laranja estão os chamados nó pai e em verde estão os nós terminais.

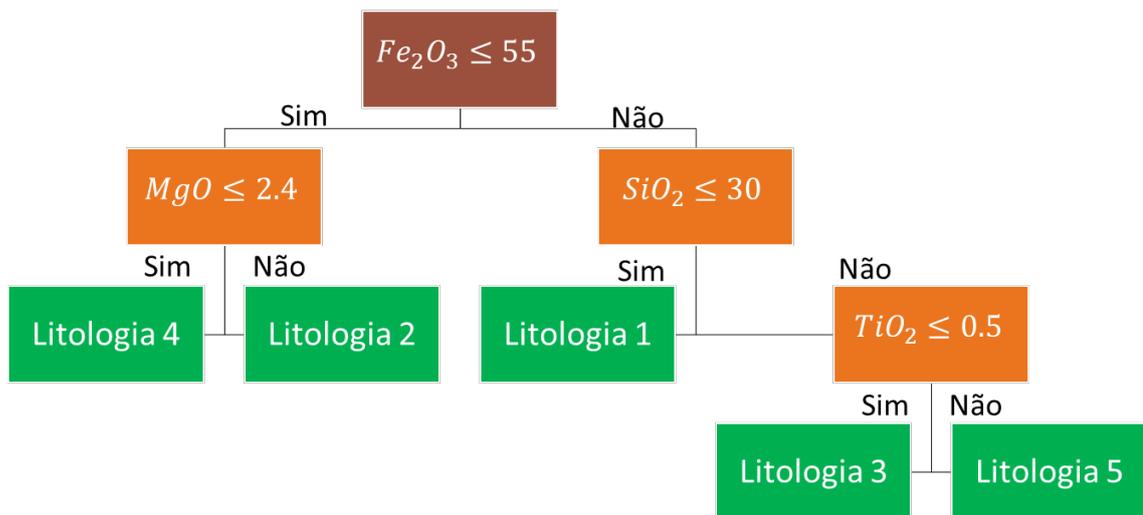


Figura 10 - Exemplo de árvore de decisão para classificação de litologias

3.2.1.2. Elaborando uma árvore de decisão para classificação

A elaboração de uma árvore de decisão passa por três etapas principais, as quais serão detalhadas a seguir.

- 1) Escolhendo a variável e o valor de corte que determinará a divisão do nó

É importante considerar que todas as variáveis independentes incluídas no estudo de árvore de decisão podem ser utilizadas para conduzir os cortes nos dados, sejam essas variáveis categóricas ou contínuas. Para determinar qual variável gera a melhor divisão e qual valor de corte desta variável deve ser aplicado, utiliza-se da função de impureza do nó, definida na equação 31, onde v é o nó avaliado, K é o número de classes da variável dependente analisada e $p(K|v)$ é a probabilidade de uma observação do banco de dados pertencer a classe K dado que estamos analisando o nó v .

$$i(v) = \phi(p(1|v), \dots, p(K|v)). \quad (31)$$

A função ϕ aplicada na equação 39 pode assumir diversas formas, desde que algumas suposições apresentadas em Izenman (2008) sejam satisfeitas, entretanto, as funções mais utilizadas em estudos de árvores de decisão são a função de entropia e o índice de diversidade Gini, definidos pelas equações 32 e 33, respectivamente.

$$i(v) = - \sum_{k=1}^K p(k|v) \log_e p(k|v). \quad (32)$$

$$i(v) = \sum_{k \neq k'} p(k|v)p(k'|v). \quad (33)$$

Calculado o valor da função de impureza para o nó raiz, o passo seguinte é avaliar a função de impureza considerando que um corte $X_i \leq \theta_j$ será aplicado e gerará dois nós: v_E e v_D (esquerdo e direito). Desta forma, deve-se calcular as probabilidades $p(k|v_E)$ e $p(k|v_D)$ e substituí-las nas equações 32 ou 33 para obter os valores de $i(v_E)$ e $i(v_D)$.

Para avaliar a estatística de bondade do corte s realizado no nó v , utiliza-se a equação 34, a qual reflete a redução da impureza ao aplicar um corte no determinado nó em estudo. Tendo em vista que diversos cortes podem ser realizados, principalmente se a variável independente avaliada for contínua, o corte que deverá ser selecionado será aquele que apresentar o maior valor de $\Delta i(s, v)$. Logo, percebe-se a necessidade de calcular a bondade do corte para todos os possíveis cortes da variável X_i .

$$\Delta i(s, v) = i(v) - p(k|v_E)i(v_E) - p(k|v_D)i(v_D). \quad (34)$$

Uma vez que existem diversas variáveis independentes nos bancos de dados avaliados e que qualquer uma delas pode ser escolhida para conduzir o corte do nó, é necessário calcular os valores de $i(v)$ e $\Delta i(s, v)$ para cada uma das variáveis independentes. Deverá ser escolhida aquela que apresentar o maior valor de $\Delta i(s, v)$ dentre todas as variáveis. Após a divisão, o processo deve ser repetido para todos os nós subsequentes, em um processo chamado de partição recursiva.

2) Decidindo que um nó é um nó terminal

É necessário definir um critério para parar os cortes nos nós, ou em caso contrário, o resultado seria uma árvore muito grande e com cada nó terminal associado a um único valor da amostra, o que certamente causaria *overfitting*.

Uma estratégia comumente utilizada é definir um número mínimo de dados (n_{min}) que um nó deve conter, de modo que se $n(v) \leq n_{min}$ declara-se o nó como terminal e nenhuma nova divisão é pesquisada para ele. Outra abordagem é definir um valor mínimo da maior estatística de bondade do corte observada, de modo que se a amostra em estudo fornecesse

valores inferiores a este número seria o momento de parar com as divisões da árvore, definindo assim o nó como terminal.

Uma outra opção é gerar uma “árvore máxima” cujo número de dados em cada nó terminal seja menor que o valor n_{min} e em seguida realizar a “poda” da mesma, até que ela apresente um tamanho adequado, técnica que será apresentada com mais detalhe na seção 3.2.1.4.

3) Associando uma classe aos dados contidos em um nó terminal

Estando em um nó terminal, o último passo do algoritmo de árvore de decisão para classificação consiste em atribuir aos dados uma das k classes da variável dependente em estudo. Esta atribuição é simples e lógica: o nó receberá como classe predita aquela com o maior número de amostras dentro dele.

3.2.1.3. Estimando o erro de classificação de uma árvore

Elaborada a árvore de decisão, deve-se quantificar os erros cometidos ao utilizá-la para predição das classes da variável dependente estudada. A primeira estatística calculada para chegar à medida dos erros de classificação dentro de um nó v é denominada estimativa de resubstituição (*resubstitution estimate*), cuja equação encontra-se apresentada em 35.

$$r(v) = 1 - \max_k p(k|v). \quad (35)$$

Considere T uma árvore de decisão e $\tilde{T} = \{v_1, v_2, \dots, v_L\}$ o conjunto com todos os l nós terminais desta árvore. A estimativa de resubstituição de toda a árvore T é dada pela equação 36:

$$R^{re}(T) = \sum_{l=1}^L r(v_l) p(v_l), \quad (36)$$

onde $p(v_l)$ é a proporção de todas as observações que caíram dentro do nó v_l .

É importante perceber que o valor de $R^{re}(T)$ será sempre menor a medida em que a árvore tem mais nós terminais, pois, quanto maior o número de divisões menor o erro gerado na predição daquele banco de dados, o que não necessariamente implica em uma boa predição quando a árvore é aplicada em outro banco. Logo, uma árvore com número de nós terminais igual ao número de dados apresentaria $R^{re}(T) = 0$, o que indica que esta medida não é tão realista a respeito da qualidade das predições geradas pela árvore em estudo, de modo que outra estatística, apresentada a seguir, deve ser utilizada com a finalidade de decidir qual a melhor árvore a ser considerada, quando existem várias opções.

3.2.1.4. Poda de árvores

Como citado na seção 3.2.1.2, a técnica de poda de árvores de decisão é utilizada para determinar o tamanho final de uma árvore. Para isso, inicia-se elaborando uma árvore de tamanho máximo, com número de dados em cada nó menor do que n_{min} e calcula-se para cada nó o valor da estimativa de ressubstituição, apresentada na equação 35. Inicia-se a poda de baixo para cima. O valor da estimativa de ressubstituição para toda a árvore, no caso em que a técnica de poda é aplicada, não é dado pela equação 36, mas sim por um valor adaptado da mesma, apresentado na equação 37 e denominado medida de complexidade do erro, na qual um parâmetro de complexidade $\alpha \geq 0$ é adicionado ao valor de $R^{re}(T)$:

$$R_{\alpha}(T) = R^{re}(T) + \alpha |\tilde{T}|, \quad (37)$$

onde $|\tilde{T}| = L$ é o número de nós terminais na sub árvore analisada, de modo que a componente $\alpha |\tilde{T}|$ pode ser vista como uma penalidade aplicada ao valor da estimativa de ressubstituição,

que faz com que o valor desta seja maior quando uma árvore com vários nós terminais é avaliada.

Para utilizar a estatística $R_\alpha(T)$ como critério de escolha da melhor sub árvore, basta selecionar aquela que minimiza $R_\alpha(T)$ para um dado valor α . O valor do parâmetro de complexidade α influencia fortemente no tamanho da árvore selecionada, sendo que valores próximos de zero implicarão em árvores maiores e valores grandes indicarão a escolha de árvores menores.

A técnica de validação cruzada ou a utilização de um banco de dados de teste para aplicar a árvore escolhida com determinado α , são estratégias que devem ser utilizadas para garantir que a árvore escolhida gera boas previsões. Maiores detalhes sobre a escolha do valor do parâmetro de complexidade podem ser lidos em Izenman (2008).

3.2.1.5. Matriz de confusão

A representação dos erros de classificação de uma árvore de decisão é geralmente feita através da matriz de confusão. As colunas desta matriz contêm os valores preditos e as linhas os valores observados na amostra trabalhada, sendo que as informações contidas nos cruzamentos das linhas e colunas são utilizadas para calcular o erro cometido pela árvore escolhida. A Tabela 3 ilustra um exemplo de matriz de confusão quando existem somente duas categorias a serem preditas.

Tabela 3 – Exemplo de matriz de confusão

Categoria		Predito	
		VERDADEIRO	FALSO
Observado	VERDADEIRO	n_{11}	n_{12}
	FALSO	n_{21}	n_{22}

Analisando a tabela acima podemos calcular a taxa de erro como sendo a soma de todos os valores que não estão na diagonal principal da matriz dividida pelo número total de dados, conforme equação abaixo. É mais conveniente comparar duas ou mais árvores através do valor

da taxa de erro de cada uma delas do que comparar as matrizes de confusão geradas por cada uma, logo, percebe-se a importância de calcular este valor.

$$\text{Taxa de erro} = \frac{n_{12} + n_{21}}{n_{11} + n_{12} + n_{21} + n_{22}}. \quad (38)$$

A acurácia da árvore em estudo, por sua vez, pode ser calculada pela soma dos valores da diagonal principal dividida pelo número total de dados, conforme representado em 39.

$$\text{Acurácia} = \frac{n_{11} + n_{22}}{n_{11} + n_{12} + n_{21} + n_{22}}. \quad (39)$$

3.2.2. Árvore de decisão para regressão

A ideia por trás da técnica de árvore de decisão para regressão é exatamente a mesma da árvore de decisão para classificação, tendo como diferença básica o fato de que a variável a ser predita é contínua e não categórica. A seguir estão apresentadas as modificações que devem ser realizadas nas fórmulas apresentadas na seção 3.2.1 com o objetivo de construir uma árvore de decisão para regressão.

Como explicado anteriormente, no caso de árvore de decisão para classificação, temos que a categoria predita para um nó terminal é aquela que contém o maior número de observações dentro de um dado nó. Na árvore de decisão para regressão, o valor predito de um dado contido em um nó terminal é o valor médio da variável resposta considerando todas as observações que caíram dentro do nó, conforme apresentado na equação 40.

$$\hat{y}_l = \bar{y}(v_l) = \frac{1}{n(v_l)} \sum_{x_i \in v_l} y_i, \quad (40)$$

onde y_i é o valor da variável resposta para a i -ésima observação e $n(v_l)$ é o número de dados dentro do nó v_l .

De acordo com Izenman (2008) a fórmula de cálculo da estimativa de ressubstituição para um nó pertencente a uma árvore para regressão sofre algumas adaptações quando comparada com a mesma medida para árvores de classificação, resultando na seguinte fórmula para os casos de variável resposta contínua:

$$R^{re}(v_l) = \frac{1}{n} \sum_{x_i \in v_l} (y_i - \bar{y}(v_l))^2. \quad (41)$$

Logo, a estimativa de ressubstituição para toda a árvore para regressão é dada por:

$$R^{re}(T) = \sum_{l=1}^L R^{re}(v_l). \quad (42)$$

A estratégia a ser adotada para dividir um nó pai em dois nós filhos no caso em estudo baseia-se em avaliar qual corte fornece a maior redução no valor de $R^{re}(T)$. Lembrando que todas as variáveis independentes disponíveis na base de dados deverão ser avaliadas com base nesta estatística, de modo que seja identificada qual variável e qual o valor de corte da mesma resultam na maior redução de $R^{re}(T)$.

$$\Delta R^{re}(v) = R^{re}(v) - R^{re}(v_E) - R^{re}(v_D). \quad (43)$$

A ideia utilizada para definir um nó como terminal, apresentada no caso de árvores para classificação, é a mesma a ser adotada para o caso de árvores para regressão: pode-se fixar um valor de n_{min} , definir um valor mínimo que a maior redução no valor da estimativa de ressubstituição deve assumir ou realizar a poda de uma árvore com tamanho máximo, avaliando os valores da estatística $R_\alpha(T)$ para decidir qual tamanho ideal da árvore a ser trabalhada.

Utilizar a técnica de validação cruzada ou um banco de teste são estratégias ideais para auxiliar na determinação da árvore para regressão, assim como foi sugerido para árvores de decisão para classificação.

3.3. RANDOM FOREST

Antes de apresentar a técnica de *random forest* é necessário conhecer um procedimento chamado *bagging*, que combina um conjunto de algoritmos de aprendizado (no caso em estudo um conjunto de árvores de decisão) para melhorar a performance preditiva desse algoritmo. O *bagging* é o primeiro passo para elaboração de uma *random forest*, logo, percebe-se a importância de compreender bem seu funcionamento. Detalhes são apresentados na seção seguinte, que adotou a mesma notação apresentada em Izenman (2008).

3.3.1. Bagging

Inicia-se o *bagging* tendo como base um conjunto de dados \mathcal{L} que contém n informações de diversas variáveis independentes (x) e uma variável dependente (y) que deverá ser estimada, sendo que essa pode ser contínua ou categórica. Logo, $\mathcal{L} = \{(x_i, y_i), i = 1, 2, \dots, n\}$.

O primeiro passo deste procedimento consiste em retirar uma amostra de tamanho n com reposição do conjunto de dados \mathcal{L} . Este passo é repetido B vezes visando criar B conjuntos de dados retirados de uma mesma distribuição, mas com elementos individuais distintos entre uma amostra e outra. Em outras palavras, a técnica de *bootstrap* é aplicada ao banco de dados em estudo, gerando $\mathcal{L}^{*b} = \{(x_i^{*b}, y_i^{*b}), i = 1, 2, \dots, n\}$ e $b = 1, 2, \dots, B$.

Tendo os B conjuntos de dados, o passo seguinte é elaborar B árvores de decisão (de classificação ou regressão) com tamanho máximo, sem que nenhuma poda seja realizada. Esta etapa gerará árvores de decisão distintas entre si em relação aos cortes aplicados em cada nó pai e ao valor predito da variável y .

Quando o estudo é feito visando classificação, a categoria predita para uma amostra do banco \mathcal{L} utilizando *bagging* será a que mais vezes foi atribuída para a dada amostra quando as B árvores são analisadas conjuntamente. No caso de árvores para regressão, o valor predito

através do *bagging* será a média dos valores preditos pelas B árvores construídas, conforme indicado na equação 44.

$$\hat{\mu}_{bag}(x) = \frac{1}{B} \sum_{b=1}^B \hat{\mu}^{*b}(x). \quad (44)$$

O *bagging* permite que cada uma das árvores geradas seja aplicada em um banco de teste distinto, permitindo avaliar o erro de predição. Este procedimento é denominado “out-of-bag” ou somente OOB e é elaborado da seguinte forma: todos os dados que não compuseram o conjunto \mathcal{L}^{*b} são utilizados como dados de teste, ou seja, fazem parte do OOB. Deve-se avaliar quais são as árvores que não utilizaram a amostra x_i em sua etapa de treinamento e em seguida deve-se aplicar nesta amostra os cortes de cada uma dessas árvores, visando obter o valor predito de y_i quando cada uma delas é utilizada. Assim, o valor final predito de y_i será aquele cuja categoria foi mais vezes atribuída à amostra x_i . Logo, o erro de classificação relacionado ao OOB pode ser determinado como:

$$Err_{o_{bag}} = \frac{1}{n} \sum_{i=1}^n I_{\hat{y}_i \neq y_i}, \quad (45)$$

onde I representa a proporção de vezes em que a classe predita foi diferente da classe real para cada um dos i dados que compõem \mathcal{L} .

No caso em que a árvore é elaborada para fins de regressão, a utilização do OOB para avaliar o erro cometido parte de uma lógica diferente da apresentada para os casos de árvore de classificação: primeiro, procura-se quais são as n_i árvores que não utilizaram a amostra x_i em sua construção (sendo n_i um valor menor ou igual a B). Em seguida, calcula-se o valor estimado y_i utilizando cada uma destas árvores. O valor final predito para a amostra pertencente ao OOB é a média das estimativas de todas as árvores que inicialmente não utilizaram a amostra x_i , conforme equacionado em 46.

$$\hat{\mu}_{bag}(x_i) = \frac{1}{n_i} \sum_{b \in \mathcal{N}_i} \hat{\mu}^{*b}(x_i), \quad (46)$$

onde \mathcal{N}_i é o conjunto de árvores que não utilizaram a amostra x_i na etapa de treinamento.

O passo seguinte para calcular o erro no caso de árvore de decisão para regressão utilizando dos resultados da OOB consiste em aplicar a fórmula apresentada em 47, a qual calcula o erro quadrático médio entre os valores reais e os valores estimados, ou seja, $\hat{\mu}_{bag}(x_i)$.

$$Erro_{bag} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\mu}_{bag}(x_i))^2. \quad (47)$$

Desta forma é possível quantificar qual o erro cometido, tanto para árvores de classificação quanto para árvores de regressão, utilizando para isso a técnica de *bagging*.

3.3.2. Random forest

Como apresentado na seção anterior, a técnica de *bagging* consiste em inserir uma componente aleatória na escolha dos dados que serão utilizados para elaborar um conjunto de árvores de decisão. A técnica de *random forest*, além de ter como ponto de partida o *bagging*, utiliza de mais uma componente aleatória: em cada nó da árvore são escolhidas aleatoriamente quais serão as m variáveis avaliadas para determinar o corte a ser aplicado no dado nó, sendo m um número bem menor que o número total r de variáveis disponíveis no banco. Izenman (2008) sugere trabalhar com valores de m iguais a \sqrt{r} , $2\sqrt{r}$ ou $0.5\sqrt{r}$ como ponto de partida. O autor também sugere utilizar pelo menos 1000 amostras *bootstrap* para construir uma *random forest*.

De uma forma geral, uma amostra *bootstrap* é coletada do conjunto de dados e em cada nó da árvore um conjunto m de variáveis é escolhido aleatoriamente para determinar o corte do nó. O processo de escolha das variáveis é repetido até que uma árvore de tamanho máximo seja obtida. Em seguida, seleciona-se outra amostra *bootstrap* e o processo é refeito. Assim, B

árvores serão construídas com base em amostras e variáveis diferentes e este conjunto de árvores é denominado *random forest*, ou floresta aleatória em português.

A classe ou valor predito para cada dado pertencente a \mathcal{L} deve ser calculado considerando as mesmas fórmulas apresentadas na seção anterior para a técnica de *bagging*. O mesmo vale para o erro de classificação ou predição utilizando os dados da amostra OOB.

3.3.2.1. Determinando quais variáveis são importantes para uma boa predição

Uma vez que a técnica de *random forest* considera diferentes variáveis para estudar o melhor corte nó a nó, é importante determinar quais são as variáveis que fornecem os melhores resultados quando utilizadas na elaboração das árvores de decisão. A importância de uma variável para a geração de bons resultados é medida por um escore denominado “*raw permutation accuracy importance*”, que será apresentado a seguir utilizando a notação $imp(j)$.

Seja \mathcal{L}^{*b} uma amostra *bootstrap* e T^{*b} uma árvore relacionada a esta amostra. Para determinar a importância de cada variável, calcula-se primeiramente a taxa de erro da OOB ($Erro_b(OOB)$) relacionada a esta árvore, valor que será utilizado como referência. Em seguida, escolhe-se uma variável j e os valores amostrais para esta variável são trocados entre si, sendo que as demais variáveis permanecem com seus valores amostrais inalterados. A lógica utilizada é: se a variável j for realmente importante para gerar boas predições, a taxa de erro calculada considerando os dados trocados para esta variável será maior do que a taxa de erro original. Desta forma, após a mudança dos valores amostrais da variável j deve-se calcular a taxa de erro da OOB novamente ($Erro_b(OOB_j)$) e assim calcular o valor da estatística $raw_b(j)$, equacionada em 48, a qual dimensiona o impacto da mudança de dados da variável j .

$$raw_b(j) = Erro_b(OOB_j) - Erro_b(OOB). \quad (48)$$

O valor de $raw_b(j)$ deve ser calculado para cada uma das B amostras *bootstrap* e a importância da variável j é finalmente definida como:

$$imp(j) = \frac{1}{B} \sum_{b=1}^B raw_b(j). \quad (49)$$

O escore $imp(j)$ deve ser calculado para todas as variáveis do banco, de modo que as variáveis com os maiores valores desta estatística são as consideradas mais importantes para o estudo em questão.

3.4. SUPPORT VECTOR MACHINE

A técnica de *support vector machine* (SVM) vem sendo amplamente aplicada em estudos de classificação e regressão devido aos bons resultados obtidos através de sua utilização. Para explicá-la, será apresentada inicialmente neste texto a teoria por trás da classificação de um banco de dados linearmente separável que contém somente duas categorias. Em seguida os conceitos serão extrapolados para um banco não separável linearmente, também com variável dependente binária. Até este ponto terão sido apresentados somente SVM lineares, os quais nem sempre geram as melhores respostas. Tendo em vista a necessidade de abordar problemas mais complexos, a seção subsequente tratará dos conceitos por trás de um SVM não linear. Por fim, será apresentada a utilização de *support vector machine* em problemas de regressão.

3.4.1. Dados linearmente separáveis

O conceito básico por trás de SVM é o de encontrar um vetor capaz de dividir corretamente um conjunto de dados nas categorias as quais estes pertencem, de modo que não haja erro na classificação ou que esse seja o menor possível.

Seja, $\mathcal{L} = \{(x_i, y_i), i = 1, 2, \dots, n\}$ um conjunto de dados de treinamento que contém diversas variáveis independentes x e uma variável dependente y que é binária. Logo, o objetivo é encontrar uma função do tipo $f(x) = \beta_0 + x^T \beta = 0$ que seja um hiperplano capaz de classificar corretamente os dados (onde β é um vetor unitário). O sinal desta função $f(x)$ é o

classificador em questão, ou seja, se $f(x) = \beta_0 + x^T \beta \geq 0$ tem-se $C(x) = +1$, e se $f(x) = \beta_0 + x^T \beta < 0$, $C(x) = -1$, conforme apresentado na equação 50.

$$C(x) = \text{senal}(f(x)). \quad (50)$$

Desta forma, para cada dado da variável y atribui-se um valor da função $C(x)$, de modo que seja possível classificar cada amostra em estudo como pertencente a somente uma classe. Se um conjunto de dados é linearmente separável, é possível dizer que existem valores β_0 e β que satisfaçam simultaneamente as equações 51 e 52, as quais podem ser combinadas em uma desigualdade única, representada em 53.

$$\beta_0 + x_i^T \beta \geq +1, \quad \text{se } y_i = +1, \quad (51)$$

$$\beta_0 + x_i^T \beta \leq -1, \quad \text{se } y_i = -1, \quad (52)$$

$$y_i(\beta_0 + x_i^T \beta) \geq +1. \quad (53)$$

Deve-se, portanto, identificar quais valores os parâmetros β_0 e β devem assumir para gerar o hiperplano $f(x) = 0$. É possível que em uma situação estudada haja mais de um hiperplano capaz de separar corretamente os dados, mas o objetivo por trás do SVM é encontrar aquele que maximiza essa separação, o que pode ser obtido através dos cálculos apresentados a seguir.

Considere duas classes, uma representada pelos pontos verdes e outra pelos pontos vermelhos ilustrados na Figura 11. O ponto vermelho e o ponto verde mais próximos da função $f(x)$, os quais na figura em análise estão sobre as linhas pontilhadas azuis, são os denominados “*support vectors*” ou vetores de suporte. O hiperplano a ser escolhido dentre vários possíveis é aquele que maximiza a distância existente entre os vetores de suporte e o hiperplano, também conhecido como “hiperplano de ótimo de separação”. Em Izenman (2008) está demonstrado que a distância perpendicular entre o vetor de suporte e o hiperplano é dada por $1/||\beta||$, logo, a margem que separa os vetores de suporte das duas categorias é dada por $2/||\beta||$, fazendo com

que o objetivo seja maximizar o valor da margem mantendo satisfeita a equação 53 ou minimizar $\frac{1}{2} \|\beta\|^2$ satisfazendo o mesmo critério, o que gera o mesmo resultado.

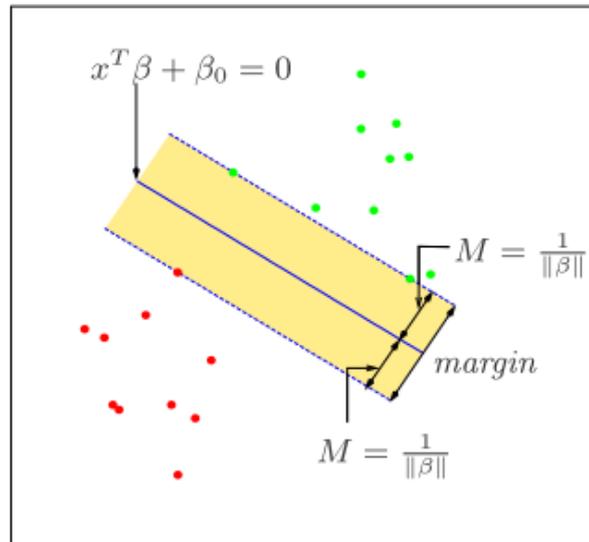


Figura 11 - Representação do SVM para um conjunto de dados linearmente separáveis com duas categorias. Fonte: Hastie, Tibshirani e Friedman (2009)

Para resolver este problema de otimização, utiliza-se de multiplicadores lagrangianos α , que são incorporados na função primal $F_P(\beta_0, \beta, \alpha)$ detalhada na equação 54. Observe que a função objetivo $\frac{1}{2} \|\beta\|^2$ e a condição $y_i(\beta_0 + x_i^T \beta) \geq +1$ são combinadas em uma função única com o objetivo de gerar o resultado ótimo.

$$F_P(\beta_0, \beta, \alpha) = \frac{1}{2} \|\beta\|^2 - \sum_{i=1}^n \alpha_i \{y_i(\beta_0 + x_i^T \beta) - 1\}, \quad (54)$$

onde α é um vetor com n coeficientes lagrangianos que possuem valores maiores ou iguais a zero.

A função $F_P(\beta_0, \beta, \alpha)$ deve ser minimizada em relação às variáveis primais β_0 e β respeitando às seguintes condições, apresentadas em Izenman (2008):

$$\frac{\partial F_p(\beta_0, \beta, \alpha)}{\partial \beta_0} = - \sum_{i=1}^n \alpha_i y_i = 0, \quad (55)$$

$$\frac{\partial F_p(\beta_0, \beta, \alpha)}{\partial \beta} = \beta - \sum_{i=1}^n \alpha_i y_i x_i = 0, \quad (56)$$

$$y_i(\beta_0 + x_i^T \beta) - 1 \geq 0, \quad (57)$$

$$\alpha_i \geq 0, \quad (58)$$

$$\alpha_i \{y_i(\beta_0 + x_i^T \beta) - 1\} = 0. \quad (59)$$

Resolvendo as equações 55 e 56 tem-se, respectivamente, as equações 60 e 61 que quando são substituídas na equação 54 levam ao mínimo da função primal $F_p(\beta_0, \beta, \alpha)$, denominado $F_D(\alpha)$ e equacionado em 62.

$$\sum_{i=1}^n \alpha_i y_i = 0, \quad (60)$$

$$\beta^* = \sum_{i=1}^n \alpha_i y_i x_i, \quad (61)$$

$$F_D(\alpha) = \frac{1}{2} \|\beta^*\|^2 - \sum_{i=1}^n \alpha_i \{y_i(\beta_0^* + x_i^T \beta^*) - 1\}. \quad (62)$$

Aplicando a equação 59 na equação 62, temos como resultado a equação 63, a qual deve ser maximizada em relação ao valor dos multiplicadores lagrangianos, respeitando as condições $\alpha_i \geq 0$ e $\sum_{i=1}^n \alpha_i y_i = 0$.

$$F_D(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (x_i^T x_j) \quad (63)$$

Encontrados os valores $\hat{\alpha}$ que maximizam $F_D(\alpha)$ encontra-se, portanto, quais são os valores ótimos do vetor de pesos através da equação 64.

$$\hat{\beta} = \sum_{i=1}^n \hat{\alpha}_i y_i x_i. \quad (64)$$

O valor de $\hat{\beta}_0$, por sua vez, é determinado após a obtenção do vetor de pesos $\hat{\beta}$, por meio da aplicação da equação 65:

$$\hat{\beta}_0 = -\frac{1}{2} \{x_{-1}^T \hat{\beta} + x_{+1}^T \hat{\beta}\}, \quad (65)$$

onde x_{-1}^T é valor do vetor de suporte da classe vermelha e x_{+1}^T é o valor do vetor de suporte da classe verde, considerando o exemplo em estudo.

Com os valores de $\hat{\beta}$ e $\hat{\beta}_0$ calculados, basta substituí-los na função $f(x)$ para determinar o hiperplano ótimo para classificar os dados em estudo, conforme indicado na equação abaixo:

$$\hat{f}(x) = \hat{\beta}_0 + x^T \hat{\beta}. \quad (66)$$

É possível perceber que somente os vetores de suporte são realmente utilizados na determinação do hiperplano ótimo de separação, sendo que nenhuma outra amostra é considerada nos cálculos aqui apresentados, demonstrando a importância destes vetores na classificação de toda a amostra avaliada.

3.4.2. Dados não linearmente separáveis

A maioria dos problemas de classificação existentes estão relacionados a bancos de dados pertencentes a categoria de dados não linearmente separáveis. A solução encontrada para lidar com situações desta natureza passa pela incorporação de uma variável folga na equação 53, a qual é denotada por ξ e possui um valor ξ_i associado a cada dado (x_i, y_i) , sendo que esta variável só assume valores maiores ou iguais a zero. Tendo em vista a adição de um novo elemento ao problema em estudo, a equação 53 assume a forma da equação 67.

$$y_i(\beta_0 + x_i^T \beta) + \xi_i \geq +1. \quad (67)$$

A Figura 12 representa graficamente a incorporação da variável folga para tratar dados não linearmente separáveis.

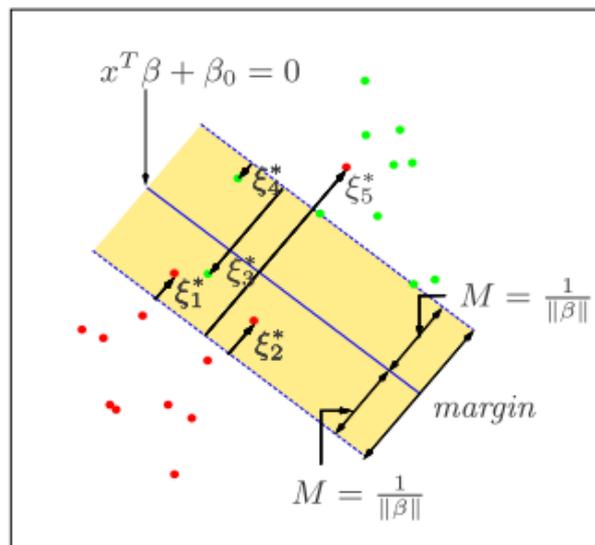


Figura 12 - Representação do SVM para um conjunto de dados não linearmente separáveis com duas categorias

Fonte: Hastie, Tibshirani e Friedman (2009)

Agora, o problema de otimização não passa somente pela minimização de $\frac{1}{2} \|\beta\|^2$, como era feito no caso de dados linearmente separáveis, mas sim pela minimização de toda equação 68, observando as condições $y_i(\beta_0 + x_i^T \beta) \geq +1 - \xi_i$ e $\xi_i \geq 0$.

$$\frac{1}{2} \|\beta\|^2 + C g_\sigma(\xi), \quad (68)$$

onde C é o parâmetro de regularização, que só assume valores maiores que zero e atua como uma constante reguladora do impacto da variável folga na função de minimização. Sabe-se, que valores maiores do parâmetro C penalizam mais as amostras que caem dentro da margem, fazendo com que esta seja menor, o que gera melhores previsões dos dados utilizados para treinar o SVM mas que pode gerar *overfitting*, levando a uma previsão ruim em dados não utilizados na elaboração do SVM. Por outro lado, valores menores de C permitem que mais pontos caiam dentro da margem, fazendo com que ela seja mais larga, o que pode gerar previsões piores nos dados de treinamento, mas evita *overfitting* nos dados. A Figura 13 ilustra diferentes valores de C , 10 e 0.1, escolhidos em um *support vector machine* para classificação. As margens são as linhas pontilhadas na figura.

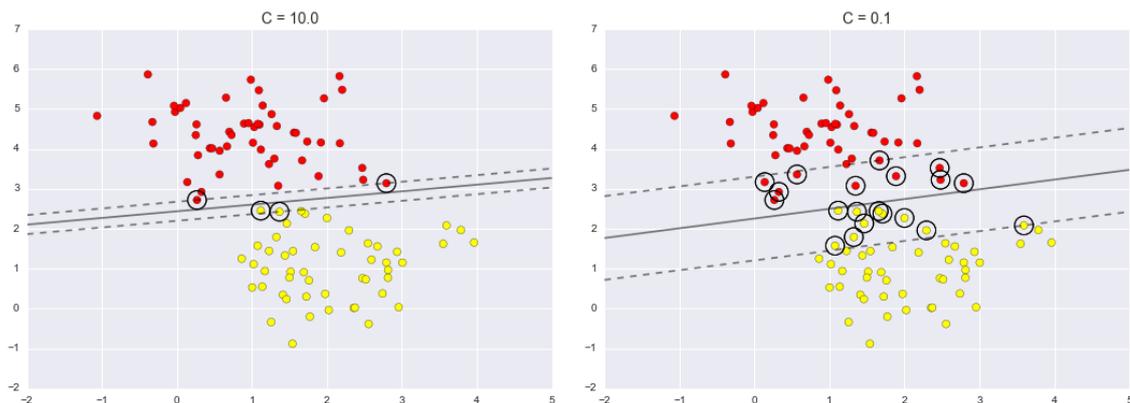


Figura 13 - Exemplo de diferentes valores de C em um *support vector machine* para classificação

Fonte: <https://jakevdp.github.io/PythonDataScienceHandbook/05.07-support-vector-machines.html>

A função $g_\sigma(\xi)$, por sua vez, é calculada através da fórmula apresentada em 69, na qual o valor de σ geralmente utilizado é igual a 1 ou 2, segundo Izenman (2008). Neste texto o valor adotado para este parâmetro deste ponto em diante será 1.

$$g_\sigma(\xi) = \sum_{i=1}^n \xi_i^\sigma. \quad (69)$$

A função primal para o caso de dados não linearmente separáveis depende dos valores de ξ e η além dos valores de β_0 , β e α , conforme representado na equação 70.

$$\begin{aligned} F_P(\beta_0, \beta, \xi, \alpha, \eta) &= \frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i \{y_i(\beta_0 + x_i^T \beta) - (1 - \xi_i)\} \\ &\quad - \sum_{i=1}^n \eta_i \xi_i, \end{aligned} \quad (70)$$

onde η é um vetor com n valores maiores ou iguais a zero.

Ao diferenciar a função primal acima em relação aos parâmetros β_0 , β e ξ , mantendo fixos α e η , obtêm-se os seguintes resultados de acordo com Izenman (2008):

$$\frac{\partial F_P}{\partial \beta_0} = - \sum_{i=1}^n \alpha_i y_i, \quad (71)$$

$$\frac{\partial F_P}{\partial \beta} = \beta - \sum_{i=1}^n \alpha_i y_i x_i, \quad (72)$$

$$\frac{\partial F_P}{\partial \xi_i} = C - \alpha_i - \eta_i. \quad (73)$$

Ao igualar a zero as equações 71 a 73, tem-se como solução as equações 74 a 76, que quando substituídas em 70 dão origem à função $F_D(\alpha)$ apresentada em 77, a qual é idêntica a $F_D(\alpha)$ apresentada no caso linearmente separável (equação 63).

$$\sum_{i=1}^n \alpha_i y_i = 0, \quad (74)$$

$$\beta^* = \sum_{i=1}^n \alpha_i y_i x_i, \quad (75)$$

$$\alpha_i = C - \eta_i, \quad (76)$$

$$F_D(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (x_i^T x_j). \quad (77)$$

Segundo Izenman (2008), as condições apresentadas nas equações 78 a 81 devem ser satisfeitas durante a busca dos valores α_i que maximizam a função $F_D(\alpha)$, além de que a seguinte restrição deve ser respeitada: $0 \leq \alpha_i \leq C$. Logo, percebe-se que o valor do parâmetro de regularização é uma limitante superior dos valores que α_i podem assumir.

$$y_i(\beta_0 + x_i^T \beta) - (1 - \xi_i) \geq 0, \quad (78)$$

$$\xi_i \geq 0, \alpha_i \geq 0 \text{ e } \eta_i \geq 0, \quad (79)$$

$$\alpha_i \{y_i(\beta_0 + x_i^T \beta) - (1 - \xi_i)\} = 0, \quad (80)$$

$$\xi_i(\alpha_i - C) = 0. \quad (81)$$

Encontrado o vetor $\hat{\alpha}$ que otimiza o problema em questão, os valores de β que definem o hiperplano de interesse podem ser obtidos substituindo $\hat{\alpha}$ na equação 75, enquanto o valor β_0 pode ser encontrado através das equações 80 e 81, após a obtenção de β .

3.4.3. Support vector machine não linear

A utilização de SVM lineares nem sempre traz bons resultados e isso é comum quando a técnica em questão é aplicada em situações nas quais um hiperplano não é suficiente para classificar corretamente um conjunto de dados. Problemas como esse são comuns e a solução é a construir um SVM não linear.

A técnica de SVM não linear tem como ideia principal transformar as variáveis de *input* em relação à sua escala original, levando-as para um espaço de Hilbert \mathcal{H} com $N_{\mathcal{H}}$ dimensões, no qual deve-se encontrar um hiperplano ótimo de separação. Desta forma, cada observação terá seu valor correspondente no espaço transformado, conforme representado na equação 82. A variável dependente y , por sua vez, continua inalterada, logo, os dados apresentarão a seguinte estrutura no cálculo do SVM não linear: $\{\Phi(x_i), y_i\}$.

$$\Phi(x_i) = (\phi_1(x_i), \dots, \phi_{N_{\mathcal{H}}}(x_i))^T \in \mathcal{H}, i = 1, 2, \dots, n. \quad (82)$$

De acordo com Izenman (2008), um ponto importante a ser observado é a dificuldade existente de se calcular o produto interno de dados transformados não linearmente, ou seja, calcular $\langle \Phi(x_i), \Phi(x_j) \rangle = \Phi(x_i^T) \Phi(x_j)$. Para contornar esta dificuldade utiliza-se o “truque de Kernel”, que ao invés de realizar o cálculo de $\langle \Phi(x_i), \Phi(x_j) \rangle$ no espaço \mathcal{H} , o faz através da utilização de uma função Kernel do tipo $K(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle$ no espaço original das variáveis estudadas, o que facilita os cálculos computacionais tendo em vista que a dimensionalidade do espaço original é muito menor do que a do espaço transformado \mathcal{H} . Além disso, a técnica de SVM tem como vantagem a necessidade de calcular o produto interno $\langle x_i, x_j \rangle = x_i^T x_j$ somente durante o cálculo da função $F_D(\alpha)$. Desta forma, ao utilizar o truque de Kernel a função $K(x_i, x_j)$ substituirá todos os produtos internos $\langle \Phi(x_i), \Phi(x_j) \rangle$ encontrados durante a aplicação do SVM.

Tendo em vista as facilidades de se utilizar Kernel para resolver problemas desta natureza, o próximo passo é determinar qual o formato que a função K deve assumir.

Existem uma série de condições que devem ser satisfeitas para que a função kernel seja válida, as quais não serão abordadas neste texto, mas podem ser lidas em Izenman (2008).

Dentre as várias funções K possíveis e que satisfazem todas as condições necessárias, existem duas que são amplamente utilizadas e estão apresentadas nas equações seguintes, onde $\sigma > 0$ é o parâmetro de escala (também conhecido como parâmetro *gamma*), $c \geq 0$ e d é um número inteiro.

$$\text{Kernel polinomial de grau } d = (\langle x, y \rangle + c)^d. \quad (83)$$

$$\text{Kernel de função de base radial gaussiana} = \exp\left\{-\frac{\|x - y\|^2}{2\sigma^2}\right\}. \quad (84)$$

A função sigmoide com parâmetros $a \geq 0$ e $b \geq 0$, apresentada na equação 85, não é necessariamente um kernel por não atender a todas as suposições dependendo dos valores adotados para os parâmetros a e b , mas de acordo com Izenman (2008), o uso desta função é comum em algumas situações.

$$\text{Kernel sigmoide} = \tanh(a\langle x, y \rangle + b). \quad (85)$$

Durante um estudo no qual é necessário utilizar SVM não linear é comum o questionamento de qual dentre as diversas funções kernel existentes utilizar para tratar os dados disponíveis. Na ausência de estudos anteriores realizados com as mesmas variáveis, Izenman (2008) sugere iniciar testes utilizando o kernel de função de base radial gaussiana (RBF) ou um kernel polinomial de grau 1 ou 2. Avaliando os resultados obtidos com estas funções pode-se concluir se há necessidade ou não de testar outros kernels no banco, visando a obtenção de melhores resultados.

Outro questionamento comum é qual valor do parâmetro de *gamma* utilizar quando estamos trabalhando com uma função RBF e qual valor do parâmetro de regularização C adotar durante um estudo. O valor de *gamma* é o inverso do raio de influência das amostras selecionadas pelo modelo como *support vectors*. Se o valor de *gamma* for muito pequeno, o modelo não é capaz de capturar a complexidade dos dados, pois a região de influência de qualquer *support vector* incluirá todo o *training set*. Por outro lado, quando o valor de *gamma* é muito grande, significa que o raio de influência é muito pequeno e por isso só incluirá o

próprio *support vector*, gerando assim *overfitting*. A Figura 14 ilustra a relação dos parâmetros C e γ em um problema de classificação. Percebe-se claramente, que valores grandes de γ limitam a influência do classificador às amostras mais próximas e que valores pequenos de C consideram poucos pontos como *support vectors*.

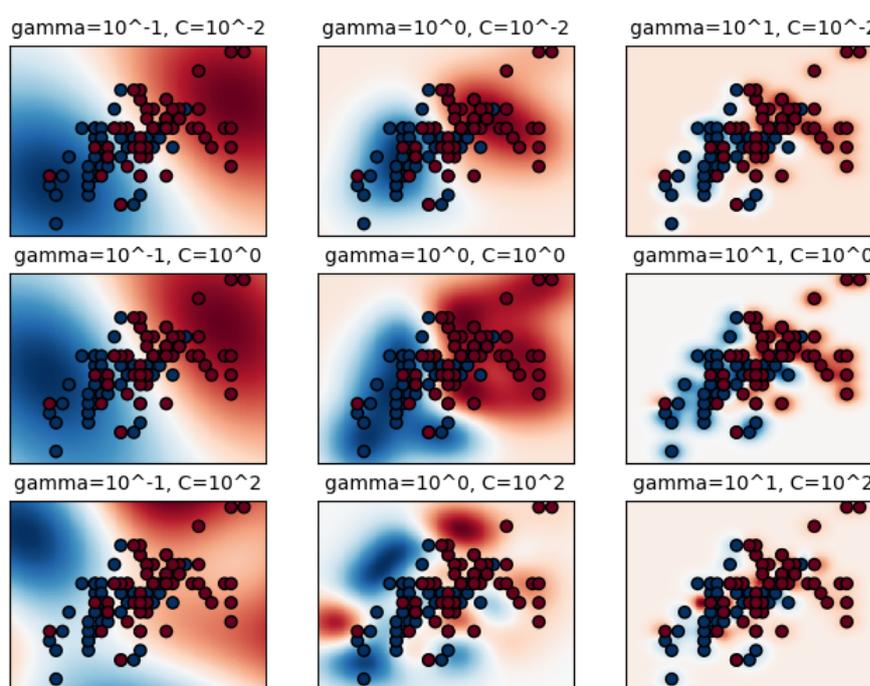


Figura 14 - Exemplo de valores dos parâmetros γ e C em um SVM de classificação

Fonte: https://scikit-learn.org/stable/auto_examples/svm/plot_rbf_parameters.html

Logo, para identificar os melhores valores de γ e C para um conjunto de dados é necessário testar diversos valores e avaliar as respostas obtidas para encontrar qual combinação fornece as melhores previsões. Estes testes devem ser realizados em um banco de teste diferente do utilizado para construir o SVM ou então a aplicação da técnica de validação cruzada faz-se necessária para garantir uma boa escolha dos parâmetros avaliados. A mesma resposta é válida para a definição dos melhores valores dos parâmetros a , b e c , necessários quando estamos trabalhando com funções sigmoide e polinomial.

Até o dado momento, todos os conceitos aqui apresentados trataram somente de SVM's aplicados para classificar variáveis binárias, mas é de conhecimento geral que em várias situações existem mais de duas categorias relacionadas à variável resposta estudada. Uma

estratégia comumente adotada para resolver este tipo de problema é dividir as diversas categorias em diferentes problemas binários, confrontando uma categoria contra todas as demais (problema conhecido como “todos contra um”) ou confrontando pares de categorias dois a dois (denominado “um contra um”). Entretanto, essas soluções envolvem problemas como desbalanceamento do número de amostras entre as categorias, quando falamos de um problema “todos contra um”, e um número reduzido de amostras para treinar o SVM, quando tratamos de um caso “um contra um”. Existem na literatura SVM’s construídos para tratar simultaneamente múltiplas categorias, sem que haja necessidade de dividir os dados em problemas binários. Estas metodologias não serão tratadas neste texto, mas detalhes podem ser lidas em Izenman (2008).

3.4.4. Support vector machine para regressão

A técnica de SVM foi inicialmente desenvolvida com o objetivo de prever categorias em um banco de dados. Entretanto, muitos dos problemas existentes têm como objetivo a predição de variáveis contínuas, o que levou à adaptação dos conceitos de *support vector machine* para tratar problemas de regressão.

No SVM para regressão, também conhecido como *support vector regression* (SVR), o hiperplano de classificação utilizado quando a variável dependente é categórica deve ser substituído por uma função de regressão linear, como a descrita em 86.

$$\mu(x) = \beta_0 + x^T \beta. \quad (86)$$

As margens adotadas no SVM para classificação, por sua vez, devem ser substituídas por um valor ϵ , o qual representa uma tolerância (positiva e negativa) em relação ao valor predito pela função de regressão aplicada. De acordo com Izenman (2008), geralmente adota-se um valor de ϵ próximo a 0.1. Estabelece-se assim um “corredor” $\mu(x) \pm \epsilon$, no qual os pontos amostrais podem ou não estar contidos.

Redefinido o conceito de hiperplano e margem para o caso de regressão, o próximo passo é construir uma função de perda que ignorará os erros de predição quando os pontos que

caírem dentro do corredor estabelecido e que calculará a perda caso os pontos caiam fora deste corredor. As equações 87 e 88 representam, respectivamente, dois tipos distintos de função de perda: a função ϵ de perda linear insensível (a qual será adotada deste ponto em diante no texto) e a função ϵ de perda quadrática insensível.

$$L_1^\epsilon(y, \mu(x)) = \max\{0, |y - \mu(x)| - \epsilon\}. \quad (87)$$

$$L_2^\epsilon(y, \mu(x)) = \max\{0, (y - \mu(x))^2 - \epsilon\}. \quad (88)$$

O próximo passo é redefinir a variável de folga ξ para o caso onde SVM é aplicado com a finalidade de regressão. Quando um ponto está dentro do corredor $\mu(x) \pm \epsilon$, a variável folga assume valor igual a zero. Em caso contrário, deve-se avaliar se a amostra caiu acima do corredor ou abaixo dele. Caso tenha caído acima, a variável folga é calculada de acordo com a equação 89 e é denotada por ξ'_i . Se a amostra caiu abaixo do corredor, o cálculo da variável folga deve ser feito considerando a equação 90 e a notação adotada é ξ_i .

$$\xi'_i = y_i - \mu(x_i) - \epsilon \geq 0. \quad (89)$$

$$\xi_i = \mu(x_i) - \epsilon - y_i \geq 0. \quad (90)$$

Para facilitar a compreensão dos conceitos acima apresentados, a Figura 15 ilustra a função $\mu(x)$, a tolerância ϵ e as variáveis folga ξ_i e ξ'_i .

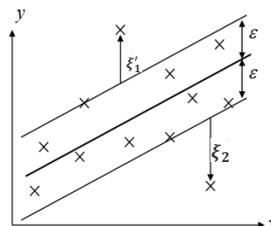


Figura 15 - Support Vector Machine para regressão

Fonte (com adaptações): Paisitkriangkrai (2012)

Assim como é feito no SVM de classificação, o objetivo do SVM para regressão é identificar quais são os parâmetros β_0 e β que minimizam os erros de predição, ou seja, minimizam $\frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^n (\xi_i + \xi'_i)$ garantindo que as seguintes suposições sejam satisfeitas:

$$y_i - (\beta_0 + x_i^T \beta) \leq \epsilon + \xi'_i, \quad (91)$$

$$(\beta_0 + x_i^T \beta) - y_i \leq \epsilon + \xi_i, \quad (92)$$

$$\xi'_i \geq 0 \text{ e } \xi_i \geq 0. \quad (93)$$

Neste caso, a função primal passa a ser a que se encontra equacionada em 94, onde a_i , b_i , c_i e d_i são os multiplicadores lagrangianos que devem ser todos maiores ou iguais a zero.

$$\begin{aligned} F_P(\beta_0, \beta, \xi, \xi') = & \frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^n (\xi_i + \xi'_i) - \sum_i a_i \{y_i - (\beta_0 + x_i^T \beta) - \epsilon - \xi'_i\} \\ & - \sum_i b_i \{(\beta_0 + x_i^T \beta) - y_i - \epsilon - \xi_i\} - \sum_i c_i \xi'_i - \sum_i d_i \xi_i. \end{aligned} \quad (94)$$

Derivando a função F_P em relação a cada um dos parâmetros que a compõe têm-se, de acordo com Izenman (2008):

$$\frac{\partial F_P}{\partial \beta_0} = \sum_i a_i - \sum_i b_i, \quad (95)$$

$$\frac{\partial F_P}{\partial \beta} = \beta + \sum_i a_i x_i - \sum_i b_i x_i, \quad (96)$$

$$\frac{\partial F_P}{\partial \xi_i} = C + b_i - d_i, \quad (97)$$

$$\frac{\partial F_P}{\partial \xi'_i} = C + a_i - c_i. \quad (98)$$

Ao igualar a zero as equações 95 a 98 são encontrados os resultados apresentados nas equações 99 a 102. Percebe-se que os lagrangianos c_i e d_i podem ser escritos em função de a_i , b_i e C , de modo que ao obter esses valores, obtêm-se também os dois primeiros citados.

$$\beta^* = \sum_i (b_i - a_i)x_i, \quad (99)$$

$$\sum_i (b_i - a_i) = 0, \quad (100)$$

$$C + b_i - d_i = 0, \quad (101)$$

$$C + a_i - c_i = 0. \quad (102)$$

Ao substituir as equações 99 e 100 na função primal obtêm-se a função F_D , apresentada em 103, que deve ser maximizada em relação aos valores de a_i e b_i escolhidos, sujeito a $0 \leq a_i \leq C$, $0 \leq b_i \leq C$ e $\sum_i (b_i - a_i) = 0$ e à satisfação das restrições apresentadas em 104 a 109.

$$F_D = -\frac{1}{2} \sum_{i,j=1}^n (b_i - a_i)(b_j - a_j) \langle x_i, x_j \rangle - \epsilon \sum_{i=1}^n (b_i - a_i) + \sum_{i=1}^n y_i (b_i - a_i), \quad (103)$$

$$a_i(\beta_0 + x_i^T \beta - y_i - \epsilon - \xi_i) = 0, \quad (104)$$

$$b_i(y_i - \beta_0 - x_i^T \beta - \epsilon - \xi'_i) = 0, \quad (105)$$

$$\xi_i \xi'_i = 0, \quad (106)$$

$$a_i b_i = 0, \quad (107)$$

$$(a_i - C) \xi_i = 0, \quad (108)$$

$$(b_i - C) \xi'_i = 0. \quad (109)$$

Após obter os valores de a_i e b_i que maximizam a função F_D , basta substituí-los na equação 99 para encontrar o valor de β^* . De acordo com Smola e Schölkopf (2003), quando $0 \leq a_i, b_i \leq C$ temos $\xi_i = 0$ e $\xi'_i = 0$ e assumindo essas premissas, as equações 104 e 105

passam a ser escritas como representado em 110 e 111, de modo que é possível encontrar o valor estimado de β_0^* . Assim, a função de regressão pode ser representada como equacionado em 112.

$$\beta_0^* = y_i - x_i^T \beta + \epsilon, \quad (110)$$

$$\beta_0^* = y_i - x_i^T \beta - \epsilon, \quad (111)$$

$$\mu^*(x) = \beta_0^* + \sum_{i=1}^n (b_i - a_i)(x^T x_i). \quad (112)$$

Os conceitos de SVM para regressão apresentados até o momento tratam de uma situação onde os dados não necessitam de uma transformação não-linear. Conforme explicado na seção 3.4.3, existem problemas nos quais é necessário efetuar uma transformação não-linear nas variáveis independentes visando identificar, em um espaço diferente do original, um modelo capaz de gerar previsões melhores que as obtidas no caso base. Para isso utiliza-se uma função kernel do tipo $K(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle$ que substitui o produto interno $\langle x_i, x_j \rangle$, o qual é necessário somente no cálculo da função $F_D(\alpha)$, de modo que a equação 103 se transforma na equação 113, seguindo o mesmo conjunto de restrições apresentado anteriormente.

$$F_D = -\frac{1}{2} \sum_{i,j=1}^n (b_i - a_i)(b_j - a_j)k(x_i, x_j) - \epsilon \sum_{i=1}^n (b_i - a_i) + \sum_{i=1}^n y_i(b_i - a_i). \quad (113)$$

Considerando a utilização da função kernel, β^* e $\mu^*(x)$ passam a ser definidos conforme representado em 114 e 115.

$$\beta^* = \sum_i (b_i - a_i)\Phi(x_i), \quad (114)$$

$$\mu^*(x) = \beta_0^* + \sum_{i=1}^n (b_i - a_i)k(x, x_i). \quad (115)$$

4. REDES NEURAIIS ARTIFICIAIS

“O cérebro é um computador (sistema de processamento de informação) altamente complexo, não-linear e paralelo” (Haykin, 2001). Sua capacidade de resolver problemas, aprender novas línguas, tomar decisões, interpretar textos e realizar outras tarefas, muitas vezes de forma simultânea e precisa, é o que vem inspirando os estudos de muitos cientistas ao longo dos anos, os quais almejam construir um “cérebro artificial” capaz de resolver problemas de diferentes naturezas.

As redes neurais artificiais (RNA), ou somente redes neurais (RN), são um ramo da ciência da computação que utiliza de conhecimentos de engenharia, estatística, matemática e neurociências para desenvolver uma “máquina” capaz de reproduzir a capacidade cerebral de resolver determinados problemas. Os primeiros estudos com este intuito datam de 1943, mas um lento avanço foi percebido durante muitas décadas devido à ausência de capacidade computacional que suportasse as ideias por trás das redes neurais. Com o avanço tecnológico das últimas décadas, estudos nesta área vem ganhando força e a aplicação de redes neurais vem se espalhando entre diversas áreas do conhecimento, como astronomia (Módolo, 2016), economia (Giacomel, 2016), computação (Silva, Almeida e Yamakami, 2012), medicina (Ambrósio, 2002), entre outras.

De acordo com Haykin (2001), o poder das redes neurais está associado a uma série de propriedades que outras técnicas não possuem, dentre as quais a autora destaca: a não-linearidade, que é a capacidade de incorporar neurônios não-lineares à rede, permitindo um tratamento adequado para problemas com *inputs* não-lineares; e a adaptabilidade, que corresponde à capacidade que as redes neurais possuem de adaptar os pesos atribuídos aos *inputs* (e camadas ocultas) sendo assim capaz de fornecer respostas adequadas caso mudanças no comportamento dos dados sejam percebidas.

O primeiro passo para elaborar algo com capacidade de aprendizado similar ao cérebro é compreender a estrutura física e o funcionamento de um cérebro real, o que será discutido na seção seguinte.

4.1. FUNCIONAMENTO DO CÉREBRO HUMANO E COMPREENSÃO DO PROCESSO DE APRENDIZAGEM

De acordo com Izenman (2008), a maior parte do cérebro é composta pelo córtex cerebral, que consiste em uma vasta rede de células interconectadas denominadas neurônios. Segundo o mesmo autor, o cérebro humano possui mais de 10 bilhões de neurônios de mais de 100 tipos diferentes.

Um neurônio é composto por três partes fundamentais, ilustradas na Figura 16:

- Dendritos, que recebem sinais (*inputs*) de outros neurônios através de processos eletroquímicos;
- Corpo celular, que contém o núcleo onde as informações recebidas são processadas. Caso o processamento exceda um determinado limite, o corpo celular gera um impulso, que será transmitido ao axônio;
- Axônio, que funciona como uma espécie de dispositivo de saída (*output*) que ao final se ramifica em terminais sinápticos que transmitirão informações ao dendrito ou corpo celular de outro neurônio, através das sinapses.

Representação simplificada de um neurônio

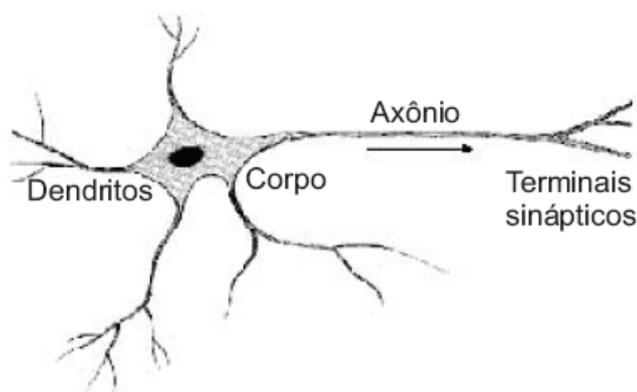


Figura 16 - Ilustração de um neurônio humano

Fonte: Ferneda (2006)

Mais importante do que compreender como é a estrutura de um neurônio real é entender como o processo de aprendizagem se dá sobre ele. O centro da aprendizagem cerebral está nas sinapses, que podem ser de dois tipos: inibitórias ou excitatórias. De modo simplificado, as primeiras evitam que o neurônio dispare informação para outros neurônios, enquanto as últimas contribuem para que o neurônio dispare. É importante lembrar que o disparo só ocorrerá quando as sinapses excederem uma determinada quantidade, ou seja, quando a contribuição das sinapses excitatórias superar as inibitórias e ultrapassar dado limite.

A capacidade de aprender está relacionada à força das conexões existentes entre os neurônios. É possível remover conexões ou adicioná-las para reforçar um dado comportamento. Logo, o neurônio dá diferentes pesos às sinapses recebidas, de modo que determinados *inputs* são mais relevantes que outros no processo de decisão de passar adiante ou não dada informação. São as experiências vividas pelo portador do cérebro que determinam quais sinapses devem ser reforçadas ou inibidas. A experiência vai sendo acumulada durante a vida de modo que o cérebro se torna capaz de desenvolver suas próprias regras para cada situação e assim tomar decisões futuras com base em suas experiências anteriores.

Após a compreender a estrutura de um neurônio real, o passo seguinte é representá-lo de forma artificial, pois o que forma uma rede neural é o conjunto de neurônios artificiais. Um neurônio artificial básico é formado por três partes: os *inputs* (X_i) e seus pesos (β_i), um contador (U) para contabilizar os *inputs* recebidos e uma função de ativação (φ) que determina qual será o *output* (Y) para um dado conjunto de *inputs*. A Figura 17 ilustra a estrutura descrita, sendo que detalhes sobre suas partes serão fornecidos nas seções seguintes.

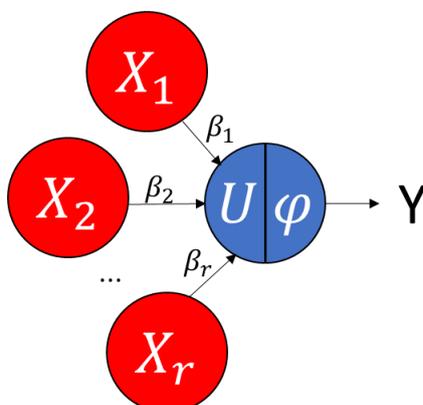


Figura 17 - Representação de um neurônio artificial

4.2. REDES NEURAIS: O COMEÇO DA HISTÓRIA

Em 1943, McCulloch e Pitts construíram o primeiro protótipo do que seria a atividade de um neurônio humano. O neurônio por eles construído tinha vários *inputs* que podiam assumir valor 0 ou 1, sendo 1 uma sinapse inibitória e zero uma sinapse excitatória. Se alguma das sinapses enviadas fosse inibitória, ou seja, enviasse o valor 1, o neurônio não transmitia nenhum sinal, logo, seu *output* era igual a 0. Se não houvesse nenhuma sinapse inibitória, os *inputs* eram somados e comparados com um valor de referência. Caso a soma dos *inputs* fosse maior ou igual este valor, o *output* era igual a 1 e o dado neurônio transmitia o sinal. Caso a soma fosse menor que o valor de referência, o neurônio não apresentava nenhuma atividade. Como explicado anteriormente, é necessário que o número de estímulos recebidos seja grande o suficiente para que a mensagem seja passada a diante.

Com o passar do tempo percebeu-se que o neurônio de McCulloch e Pitts não era uma boa aproximação do funcionamento real de um neurônio, pois as únicas alterações possíveis estavam relacionadas à mudança dos valores de *input* e do valor de referência. Os pesos que imitariam a força das conexões entre as sinapses, que é a base do aprendizado de redes neurais, não estavam representados.

Em 1949, Donald O. Hebb escreveu um livro intitulado “Organização do Comportamento”, no qual ele discutia como o sistema nervoso influenciava o comportamento e vice-versa. Neste livro, ele apresentou uma teoria conhecida como “regra de aprendizado de Hebb”, na qual ele considera que a força sináptica da conexão entre dois neurônios depende do histórico de disparos entre eles, ou seja, quanto mais esses neurônios interagem entre si, mais forte é a conexão que existe entre ambos. Posteriormente, sua teoria foi expandida, acrescentando a “regra de inibição” que diz que quando um neurônio interage frequentemente com outro, mas essa interação nunca gera um disparo, a chance futura de que um disparo seja efetuado por meio da interação entre esses neurônios cai. Lorente de Nó, no ano de 1944, apresentou sua teoria de que o cérebro continha circuitos fechados de neurônios. Hebb então propôs que a memória ficava armazenada no córtex cerebral em agrupamentos de neurônios interconectados e que estes agrupamentos diziam respeito a funções específicas, como sensações, pensamentos e percepções. As memórias, de acordo com Hebb, eram retidas no agrupamento correspondente à uma dada função e ainda que nenhum *input* fosse percebido no agrupamento, a memória continuava presente.

Após Hebb, o desenvolvimento de redes neurais deu um salto com os estudos de Rosenblatt em 1958, que elaborou o “Perceptron”, que é o neurônio de McCulloch e Pitts mas com uma importante adaptação: cada *input* vinha acompanhado de um peso de conexão, ou seja, alguns *inputs* podem influenciar mais que outros no *output*. Pesos positivos ficaram definidos como sinapses excitatórias e negativos como sinapses inibitórias, sendo que a força da conexão estava relacionada a magnitude do peso. A Figura 15, que ilustra um neurônio artificial, representa exatamente a ideia de um Perceptron, cuja formulação matemática é dada por:

$$u = \sum X_i \beta_i, \quad (116)$$

$$Y = \varphi(u), \quad (117)$$

onde X_i é o valor do i -ésimo *input*, β_i é o peso associado a X_i , U é o contador e φ é a função de ativação que determina o valor do *output* Y .

Outra forma de escrever matematicamente um Perceptron é adicionando na equação um elemento β_0 , denominado “bias”, juntamente com um *input* $X_0 = 1$. Incorporar o bias no modelo faz com que o contador aumente ou diminua (a depender do sinal de β_0), sendo que para facilitar os cálculos é interessante assumir um valor de bias que faça com que o valor de ativação se torne zero, o que se obtém ao igualar o valor de β_0 com o negativo do valor de ativação observado antes de sua incorporação. Logo, as equações 116 e 117 podem ser reescritas como representado em 118 e 119:

$$v = \beta_0 X_0 + \sum X_i \beta_i, \quad (118)$$

$$Y = \varphi(v). \quad (119)$$

Após os primeiros passos dados por McCulloch e Pitts, Hebb e Rosenblatt no estudo de redes neurais, muitos pesquisadores foram levados a aplicar as ideias por eles apresentadas e com isso novos estudos foram sendo elaborados. Em 1969, Minsky e Papert demonstraram matematicamente que o Perceptron com a estrutura apresentada por Rosenblatt, não era capaz

de resolver problemas matemáticos quando os *inputs* não eram linearmente separáveis, fazendo com que o problema fosse visto como “não computável por meio de um Perceptron”. De acordo com Izenman (2008), uma função $Y \in \{0,1\}$ é classificada como “perceptron computável” se para um dado valor de referência existir um hiperplano no qual seja possível dividir os *inputs* em dois espaços R_1 e R_0 , sendo que o primeiro corresponde aos *inputs* que retornam *output* igual a 1 e o segundo os *inputs* que retornam *output* igual a 0. Se for possível dividir os dois conjuntos de *inputs* sem erro, eles são ditos linearmente separáveis e foi na exceção a este comportamento que Minsky e Papert perceberam a falha do Perceptron, o que esfriou as pesquisas na área até os anos 80.

Com um aumento no desempenho computacional no meio da década de 80, encontrou-se uma solução para viabilizar cálculos mais complexos utilizando o Perceptron como pano de fundo: bastava considerar duas ou mais camadas de Perceptrons conjuntamente, gerando assim um Perceptron com Múltiplas Camadas (Multi-Layer Perceptron, em inglês, abreviado por MLP), solução que é empregada ainda hoje para resolver problemas mais complexos.

Desta forma, além dos nós de *input*, *output* e seus pesos equivalentes, incorpora-se uma ou mais camadas de Perceptrons, denominada “camada oculta” na qual pesos de entrada e saída são atribuídos.

A Figura 18 apresenta o desenho de uma rede com 3 *inputs* (em vermelho), 2 nós na camada oculta (em amarelo) e 2 nós de *output* (em azul). Na ilustração em questão, φ_1 e φ_2 representam, respectivamente, a função de ativação aplicada na camada oculta e na camada de *outputs*.

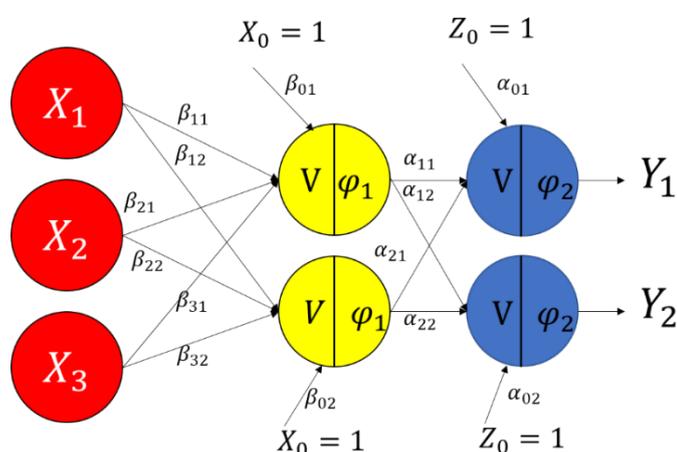


Figura 18 - Exemplo de rede com camada oculta

Para demonstrar como seria uma rede neural construída para prever variáveis geometalúrgicas, encontra-se na Figura 19 uma ilustração de rede com diversas variáveis de *input* (teores, litologias/mineralogias, granulometrias), 4 nós ocultos e diversos *outputs* (recuperações, teores no concentrado e no rejeito). Cada seta representada indica a necessidade de estimar um peso de conexão entre os nós da rede, sendo que os bias não foram ilustrados para facilitar a visualização.

No exemplo da Figura 19, somente uma camada de nós ocultos foi adicionada, mas é possível elaborar redes com quantas camadas e nós ocultos sejam necessários para uma boa predição. A implantação de uma ou mais camadas ocultas na rede neural abriu as portas para que novos trabalhos fossem iniciados nesta linha de pesquisa, fazendo com que o conhecimento em redes neurais avançasse de forma significativa. Todos os detalhes sobre a evolução do conhecimento de redes neurais até os tempos atuais podem ser lidos em Haykin (2001).

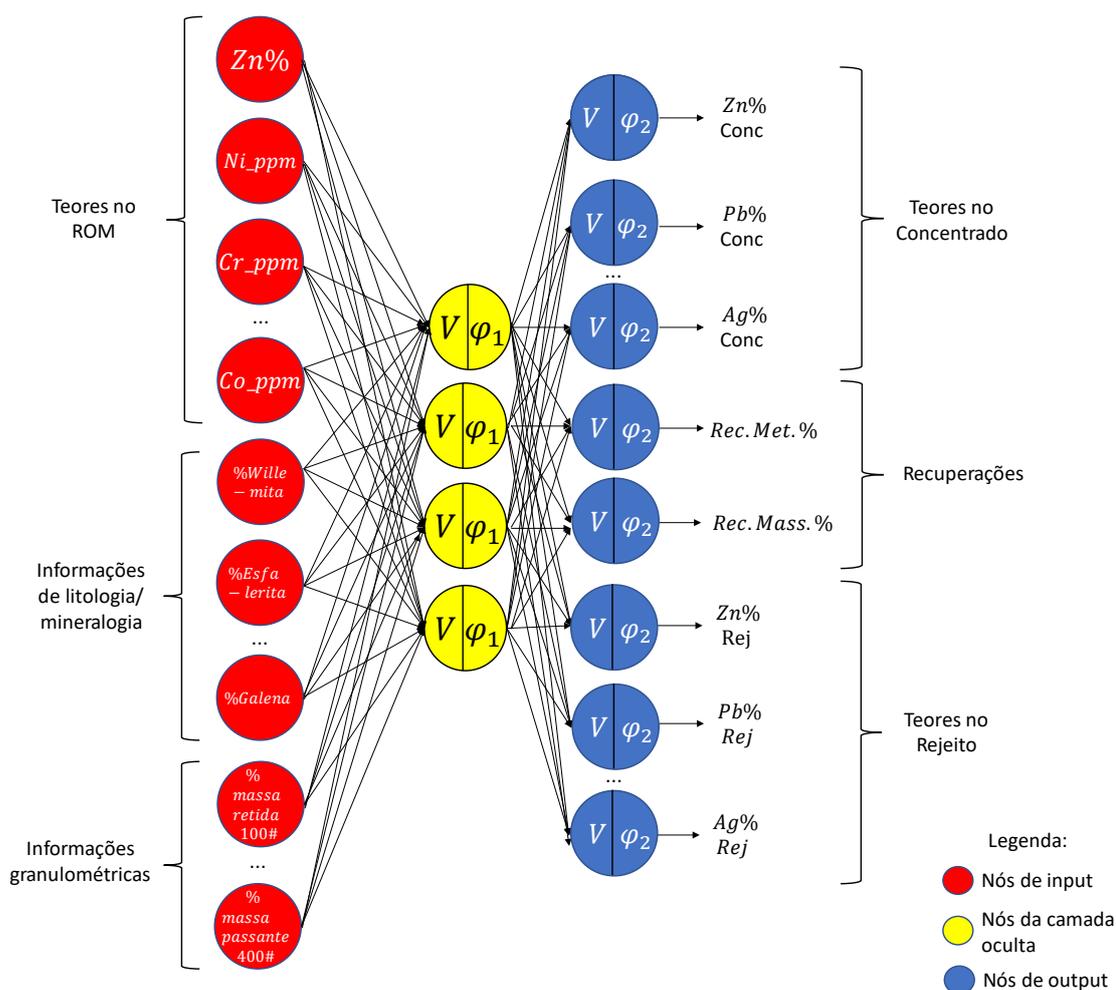


Figura 19 - Exemplo de RNA construída para prever variáveis geometalúrgicas

Para um bom funcionamento da rede neural, é importante que algumas precauções sejam tomadas, sendo que a primeira delas está relacionada a escolha da função de ativação que será aplicada em cada camada da rede. A seção seguinte apresentará as principais funções e suas características.

4.3. FUNÇÃO DE ATIVAÇÃO

Em uma rede neural, cada camada possui uma função de ativação associada, a qual analisa o contador e determina o valor de saída de determinado nó. A escolha da função de ativação pode ser decisiva entre ter uma rede robusta ou uma rede com péssima qualidade preditiva. Ela pode assumir diversos formatos e resultar em *outputs* contidos no intervalo de 0 a 1, de -1 a 1 ou até mesmo em valores do universo real.

Encontram-se equacionadas e explicadas a seguir as funções de ativação mais utilizadas, onde “ v ” é o valor do contador (considerando a existência de bias) no nó em estudo, ou seja, $v = \beta_0 X_0 + \sum X_i \beta_i$. A Figura 20 ilustra o gráfico das funções de ativação aqui representadas.

- A função de limiar, também conhecida como função de Heaviside, tem como característica retornar valores de saída iguais a zero ou um. Esta função representa o comportamento do neurônio elaborado por McCulloch e Pitts e sua propriedade de “tudo ou nada”. Sua equação está representada em 120.

$$\varphi(v) = \begin{cases} 1, & \text{se } v \geq 0 \\ 0, & \text{se } v < 0 \end{cases} \quad (120)$$

- A função sinal, cujo comportamento é similar ao da função de limiar exceto pelo fato de sua resposta apresentar os valores -1, 0 ou 1, possui sua equação representada em 121.

$$\varphi(v) = \begin{cases} 1, & \text{se } v > 0 \\ 0, & \text{se } v = 0 \\ -1, & \text{se } v < 0 \end{cases} \quad (121)$$

- A função de ativação linear, cujo range de valores está contido no universo real, tem como saída o valor do contador, sem que nenhum cálculo seja efetuado. Sua equação encontra-se apresentada a seguir:

$$\varphi(v) = v. \quad (122)$$

- A função linear por partes apresenta resposta contida no intervalo de 0 a 1, sendo que sua representação matemática está em 123.

$$\varphi(v) = \begin{cases} 1, & v \geq +\frac{1}{2} \\ v, & +\frac{1}{2} > v > -\frac{1}{2} \\ 0, & v \leq -\frac{1}{2} \end{cases} \quad (123)$$

- A função sigmoide, equacionada em 124, é amplamente utilizada na construção de redes neurais e seu gráfico apresenta curva em forma de “s”. Esta função é amplamente utilizada para representar probabilidades, uma vez que sua saída fornece valores contidos no intervalo de 0 a 1. O parâmetro “a” da equação define a inclinação da função em estudo, sendo que quando este valor se aproxima do infinito a função assume o formato de uma função de limiar.

$$\varphi(v) = \frac{1}{1 + \exp(-av)}. \quad (124)$$

- A função tangente hiperbólica também possui curva em formato de “s”. A grande diferença dessa para a função sigmoide é que seus valores de saída estão contidos no intervalo de -1 a 1. Sua equação encontra-se representada em 125.

$$\varphi(v) = \frac{(e^{av} - e^{-av})}{(e^{av} + e^{-av})}. \quad (125)$$

- A função Rectifier, também conhecida como ReLU (*Rectifier Linear Unit*), é amplamente utilizada em modelos de redes neurais. Esta função assume valor igual a v , caso v seja um número positivo e assume valor zero em caso contrário. Logo, a resposta pertence ao universo dos números reais. A equação 126 representa a função ReLU.

$$\varphi(v) = \max(0, v). \quad (126)$$

- A função Rectifier Softplus, equacionada em 127, tem seus valores de saída entre 0 e $+\infty$. Ao contrário da ReLU, que assume valor igual a zero quando v é um número negativo, a função Softplus apresenta um decaimento leve até zero quando valores negativos são observados.

$$\varphi(v) = \ln(1 + e^v). \quad (127)$$

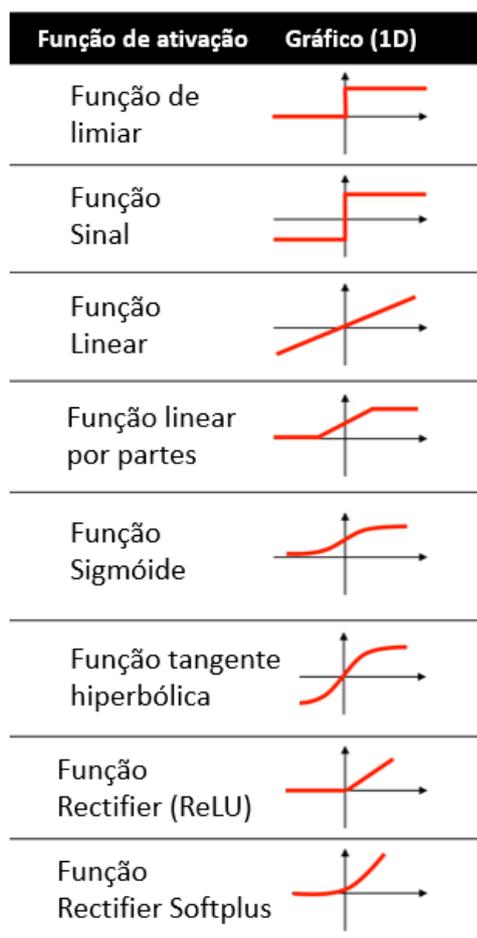


Figura 20 - Gráficos ilustrativos das funções de ativação

Fonte (com adaptações): <http://sebastianraschka.com>

4.4. ALGORITMO DE APRENDIZADO

Como dito anteriormente, uma das características das redes neurais é sua capacidade de aprender ao longo do tempo e analisar dados que inicialmente não foram utilizados para construir a rede, fornecendo previsões robustas para os dados em estudo. De acordo com Haykin (2001), define-se aprendizagem em redes neurais como: “um processo pelo qual os parâmetros livres de uma rede neural são adaptados através de um processo de estimulação pelo ambiente no qual a rede está inserida. O tipo de aprendizagem é determinado pela maneira pela qual a modificação dos parâmetros ocorre”.

Em redes neurais e em técnicas de aprendizado de máquina (*machine learning*) é possível classificar o tipo de aprendizado em dois: supervisionado e não-supervisionado. O

primeiro necessita dos dados de saída, os quais serão utilizados para comparar as previsões realizadas pela rede com os resultados reais, permitindo um ajuste nos pesos para melhorar os resultados obtidos. O segundo, por sua vez, aprende sem a necessidade de conhecer previamente os valores de saída, de modo que o processo de aprendizagem ocorre somente com base nos dados de *input*. Nesta tese, serão abordadas somente técnicas com aprendizado supervisionado, visto que os problemas a serem tratados por meio de redes neurais possuem os valores de saída conhecidos. Maiores detalhes sobre algoritmos de aprendizado não-supervisionado podem ser lidos em Haykin (2001).

4.4.1. Backpropagation

O algoritmo mais utilizado para conduzir um aprendizado supervisionado das redes neurais é conhecido como “backpropagation”, o qual consiste em calcular a primeira derivada da função de erro para encontrar os pesos que minimizem a diferença entre o valor real e o valor predito pela rede. A minimização se dá através da aplicação de um método de otimização, sendo o mais comumente utilizado o gradiente descendente, exemplificado a seguir. De forma simplificada, o processo funciona da seguinte forma:

- 1) Os pesos da rede são atribuídos de forma aleatória;
- 2) Utilizando os *inputs* combinados com os pesos aleatórios do passo anterior, os valores de saída da rede são calculados;
- 3) O valor real é comparado com o valor predito pela rede e a diferença entre eles é calculada, conforme apresentado na equação 128, onde *i* refere-se ao índice do vetor de *inputs*, ou seja, ao número da amostra no banco de dados e *k* refere ao índice do nó de *output*.

$$e_{i,k} = y_{i,k} - \hat{y}_{i,k}. \quad (128)$$

- 4) A soma dos quadrados dos erros de todos os *outputs* é calculada e esse valor é denominado “função de erro”, calculada acordo com a equação 129.

$$E_i = \frac{1}{2} \sum_{k=1}^K e_{i,k}^2. \quad (129)$$

- 5) A função do erro tem seu valor médio calculado considerando todos os n dados do banco analisado, gerando a “soma dos erros quadráticos” denotada por ESS e calculada conforme apresentado a seguir.

$$ESS = \frac{1}{n} \sum_{i=1}^n E_i. \quad (130)$$

- 6) O passo seguinte consiste em selecionar novos pesos que reduzam o valor de ESS, até que o mínimo da função seja atingido e o algoritmo convirja. Para isso, utiliza-se a técnica do gradiente descendente, que será explicada a seguir considerando uma rede com uma camada de nós oculta, além das camadas de *input* e *output*.

A notação e os conceitos apresentados a seguir foram retirados de Izenman (2008), mas pequenas alterações em relação à notação original do autor podem ser percebidas. Considere i como o índice do vetor de *input* com $i = 1, 2, \dots, n$, sendo n o número de dados do banco; M como o conjunto de r nós de *input*; T como o conjunto de t nós da camada oculta; K como o conjunto de s nós de *output*. Assim sendo, temos $k \in K$ como o índice dos nós de *output*; $j \in J$ como o índice dos nós da camada oculta; $m \in M$ como o índice dos nós de *input*; α como sendo o peso do nó da camada oculta; e z como sendo o valor do nó na camada oculta.

A soma ponderada dos nós da camada oculta que estão ligados ao k -ésimo nó de *output* é dada por (sendo $z_{i,0} = 1$):

$$v_{i,k} = \sum_{j=1}^J \alpha_{kj} z_{i,j} = \alpha_{k0} + z_i^T \alpha_k, \quad (131)$$

$$z_i = (z_{i,1}, \dots, z_{i,t})^T, \quad (132)$$

$$\alpha_k = (\alpha_{k1}, \dots, \alpha_{kt})^T. \quad (133)$$

Logo, o valor do *output* que corresponde à soma calculada pela equação 131 é dado por:

$$\hat{y}_{i,k} = g_k(v_{i,k}), \quad (134)$$

onde $g_k(\cdot)$ é a função de ativação relacionada ao *output*, a qual assume-se ser diferenciável.

Como dito anteriormente, a primeira etapa do algoritmo de backpropagation é a atribuição de valores aleatórios aos pesos da rede. Assumindo estes pesos iniciais, procura-se a direção que minimize a função do erro, de modo que os pesos sejam atualizados no passo seguinte, conforme a equação abaixo:

$$\alpha_{\tau+1} = \alpha_{\tau} + \Delta\alpha_{\tau}, \quad (135)$$

$$\alpha_{\tau} = (\alpha_{\tau,1}^T, \dots, \alpha_{\tau,s}^T)^T = (\alpha_{\tau,kj}), \quad (136)$$

onde τ neste caso significa o número da iteração em que o algoritmo se encontra e:

$$\Delta\alpha_{\tau} = -\eta \frac{\partial E_i}{\partial \alpha_{\tau}} = \left(-\eta \frac{\partial E_i}{\partial \alpha_{\tau,kj}} \right) = (\Delta\alpha_{\tau,kj}), \quad (137)$$

sendo η o parâmetro de aprendizado que determina o tamanho do passo que deverá ser dado em busca do mínimo da função. Observe que valores pequenos de η causam demora na convergência do algoritmo enquanto valores muito grandes podem sobrepassar o mínimo da função. Lembrando que o mesmo raciocínio utilizado até então é válido para $\alpha_{i,k0}$.

Em Izenman (2008) está a prova de que a derivada apresentada na equação 137 tem como resultado:

$$\alpha_{\tau+1,kj} = \alpha_{\tau,kj} + \eta \delta_{i,k} z_{i,j}, \quad (138)$$

onde:

$$\delta_{i,k} = e_{i,k} g'_k(v_{i,k}), \quad (139)$$

sendo esse valor o gradiente local da i -ésima observação no k -ésimo nó do *output*.

Resolvidos os pesos dados aos nós da camada oculta que está ligada aos nós do *output*, o próximo passo é determinar os pesos dos nós de *input* que estão ligados aos nós da camada oculta. A soma ponderada dos *inputs* e seus pesos, conectados ao j -ésimo nó oculto, é dada por (sendo $x_{i,0} = 1$):

$$u_{i,j} = \beta_{i,j0} + x_i^T \beta_{i,j}, \quad (140)$$

$$x_i = (x_{i,1}, \dots, x_{i,r})^T, \quad (141)$$

$$\beta_{i,j} = (\beta_{i,j1}, \dots, \beta_{i,jr})^T. \quad (142)$$

Desta forma, sabe-se que o valor do j -ésimo nó oculto será dado por:

$$z_{i,j} = f_j(u_{i,j}), \quad (143)$$

sendo $f_j(\cdot)$ a função de ativação (a qual assume-se ser diferenciável) relacionada aos nós de *input*.

Os pesos associados a cada nó de *input* são atualizados a cada iteração de acordo com a equação 145, da mesma forma como foi feito para os pesos dos nós da camada oculta, sendo que $\beta_\tau = (\beta_{\tau,1}^T, \dots, \beta_{\tau,t}^T)^T = (\beta_{\tau,jm})$.

$$\beta_{\tau+1} = \beta_{\tau} + \Delta\beta_{\tau}, \quad (144)$$

onde:

$$\Delta\beta_{\tau} = -\eta \frac{\partial E_i}{\partial \beta_{\tau}} = \left(-\eta \frac{\partial E_i}{\partial \beta_{\tau,jm}} \right) = (\Delta\beta_{\tau,jm}). \quad (145)$$

É importante lembrar que as fórmulas aqui apresentadas também são válidas para o valor do bias associado à camada de *input*.

Em Izenman (2008) está a prova de que a derivada apresentada na equação 145 resulta em:

$$\beta_{\tau+1,jm} = \beta_{\tau,jm} + \eta \delta_{i,j} x_{i,m}, \quad (146)$$

onde:

$$\delta_{i,j} = f'_j(u_{i,j}) \sum_{k=1}^K \delta_{i,k} \alpha_{\tau,kj}. \quad (147)$$

Em teoria, o cálculo das equações acima apresentadas deve ser repetido até que o algoritmo convirja e os pesos dos nós das camadas ocultas e dos nós de *input* sejam os que minimizam a função de erro. Mas segundo Haykin (2001), não é possível demonstrar que o algoritmo de backpropagation convergiu. Sendo assim, um critério de parada adotado leva em consideração a soma dos erros quadráticos, que deve ser atingir valores pré-determinados pelo pesquisador para que a parada ocorra. Outra opção é determinar um número específico de iterações que o algoritmo deve executar para considerar que encontrou resultados adequados. Pode-se ainda efetuar a parada do algoritmo quando observar-se que com o aumento do número de iterações a mudança do erro é ínfima, indicando que o mínimo já foi atingido.

4.4.2. Rprop

O algoritmo de aprendizado denominado Rprop (*Resilient Propagation*) foi desenvolvido em 1994 por Riedmiller e Braun com o objetivo de adaptar o algoritmo de backpropagation para torná-lo menos dependente da escolha do valor do parâmetro de aprendizado η e assim gerar resultados com maior rapidez e de melhor qualidade.

A primeira diferença entre o Rprop e o backpropagation consiste na introdução de um valor Δ_{kj} para cálculo dos pesos entre a camada oculta e a camada de *output* e de um valor Δ_{jm} para cálculo dos pesos entre a camada de *inputs* e a camada oculta, sendo que este valor é denominado valor individual de atualização. Neste texto, somente as fórmulas do Rprop para o cálculo dos pesos entre a camada oculta e a camada de *output* serão apresentadas, mas os resultados para os pesos entre os *inputs* e a camada oculta seguem o mesmo raciocínio. As fórmulas que serão apresentadas foram retiradas de Riedmiller e Braun (1993), entretanto as notações foram modificadas para estarem coerentes com as de Izenman (2008).

O cálculo do valor individual de atualização se dá através da seguinte equação:

$$\Delta_{\tau,kj} = \begin{cases} \eta^+ * \Delta_{\tau-1,kj}, & \text{se } \frac{\partial E_i}{\partial a_{\tau-1,kj}} * \frac{\partial E_i}{\partial a_{\tau,kj}} > 0 \\ \eta^- * \Delta_{\tau-1,kj}, & \text{se } \frac{\partial E_i}{\partial a_{\tau-1,kj}} * \frac{\partial E_i}{\partial a_{\tau,kj}} < 0 \\ \Delta_{\tau-1,kj}, & \text{em caso contrário} \end{cases} \quad (148)$$

onde $0 < \eta^- < 1 < \eta^+$, sendo que os valores de η^- e η^+ são determinados pelo usuário e serão discutidos mais a seguir.

Ao interpretar a equação 149 percebe-se a seguinte lógica: todas as vezes que a derivada parcial de a_{kj} muda seu sinal em relação ao passo anterior, tem-se um indicativo de que a última atualização de pesos foi muito grande e que o algoritmo pode ter pulado o mínimo local. Logo, o valor atualizado de $\Delta_{\tau,kj}$ é reduzido pelo fator η^- . Caso a derivada parcial mantenha seu sinal, o valor de $\Delta_{\tau,kj}$ é levemente aumentado ao ser multiplicado por η^+ , acelerando assim a convergência do algoritmo.

Atualizado o valor de $\Delta_{\tau,kj}$, o próximo passo é atualizar o próprio peso a_{kj} , de acordo com a seguinte equação:

$$\alpha_{\tau+1,kj} = \alpha_{\tau,kj} + \Delta\alpha_{\tau,kj}, \quad (149)$$

onde:

$$\Delta\alpha_{\tau,kj} = \begin{cases} -\Delta_{\tau,kj}, & \text{se } \frac{\partial E_i}{\partial a_{\tau,kj}} > 0 \\ +\Delta_{\tau,kj}, & \text{se } \frac{\partial E_i}{\partial a_{\tau,kj}} < 0 \\ 0, & \text{em caso contrário} \end{cases}. \quad (150)$$

Interpretando as equações 149 e 150, percebe-se que caso a derivada parcial da função de erro apresente um valor positivo, que o peso $\alpha_{\tau,kj}$ deverá ser subtraído do valor $-\Delta_{\tau,kj}$. Caso a derivada seja positiva, o valor $+\Delta_{\tau,kj}$ deverá ser somado a $\alpha_{\tau,kj}$, resultando no peso atualizado $\alpha_{\tau+1,kj}$.

Entretanto, os autores acrescentaram uma exceção à regra acima apresentada: caso a derivada parcial da função de erro tivesse seu sinal alterado entre os passos τ e $\tau - 1$, é necessário que o peso $\alpha_{\tau,kj}$ seja igual ao negativo deste mesmo peso no passo anterior, ou seja:

$$\Delta\alpha_{\tau,kj} = -\Delta\alpha_{\tau-1,kj}, \quad (151)$$

se:

$$\frac{\partial E_i}{\partial a_{\tau-1,kj}} * \frac{\partial E_i}{\partial a_{\tau,kj}} < 0. \quad (152)$$

Para que no próximo passo a derivada não tenha seu sinal invertido mais uma vez, é necessário considerar que $\frac{\partial E_i}{\partial a_{\tau-1,kj}}$ assuma o valor 0 na regra acima. Todos os valores apresentados até então deverão ser atualizados em cada iteração do algoritmo, até que a convergência seja atingida.

De acordo com Riedmiller e Braun (1993), em todos os testes realizados, o algoritmo Rprop mostrou resultados superiores ao backpropagation e demais algoritmos por eles avaliados, mostrando que sua aplicação faz com que o algoritmo convirja de forma mais rápida e obtenha resultados mais precisos e acurados.

Em relação aos parâmetros η^- e η^+ que devem ser determinados pelo usuário, os autores sugerem: um valor de $\eta^- = 0.5$, indicando que deve-se considerar a metade do valor de $\Delta_{\tau-1,kj}$ na iteração seguinte; e um valor de $\eta^+ = 1.2$, baseado em testes empíricos por eles realizados.

Assim, como o backpropagation, valores iniciais devem ser determinados para cada Δ_{kj} . De acordo com Riedmiller e Braun (1993), a escolha destes valores não é crítica para a obtenção de bons resultados de forma rápida, mas apesar disso, os autores sugerem iniciar o algoritmo com valores de $\Delta_{kj} = 0.1$.

4.4.3. Adam

O método de otimização estocástica denominado Adam (abreviação de *adaptive moment estimation*) foi elaborado por Kingma e Ba (2015). Assim como as demais técnicas apresentadas, seu foco é encontrar quais valores dos pesos da rede neural minimizam a função objetivo, que geralmente é a função de erro, conforme citado na seção 4.4.1.

Para acompanhar a notação dos autores, a função objetivo será denotada como $f(\theta)$, onde θ representa todos os pesos da rede em estudo. O gradiente desta função na iteração τ será denotado por g_τ , enquanto o valor estimado do primeiro momento do gradiente será denotado por m_τ e o segundo momento por v_τ . Os dois momentos citados são a base dos cálculos de otimização do algoritmo em questão, conforme será demonstrado a seguir.

O primeiro passo do Adam consiste em atribuir valores iniciais para os pesos da rede, o que é feito de forma aleatória. Em seguida, é necessário que os valores dos momentos também sejam inicializados. Os autores determinam que estes valores sejam igualados a zero para $\tau =$

0, logo, $m_0 = 0$ e $v_0 = 0$. O próximo passo é a definição do valor do parâmetro de aprendizado η , do valor das taxas de decaimento exponencial B_1 e B_2 (cujos valores devem estar no intervalo $[0,1)$) e do valor do parâmetro ϵ . De acordo com Kingma e Ba (2015), valores destes parâmetros que geraram bons resultados nos estudos de caso por eles realizados foram $\eta = 0.001$, $B_1 = 0.9$, $B_2 = 0.999$ e $\epsilon = 10^{-8}$, logo, estes valores podem ser utilizados como iniciais e posteriormente adaptados de acordo com o problema em estudo, caso haja necessidade.

Definidos os valores supracitados, o passo seguinte do algoritmo consiste em atualizar os valores dos momentos m_τ e v_τ , utilizando para isso as equações 153 e 154.

$$m_\tau \leftarrow B_1 m_{\tau-1} + (1 - B_1) g_\tau, \quad (153)$$

$$v_\tau \leftarrow B_2 v_{\tau-1} + (1 - B_2) g_\tau^2. \quad (154)$$

Uma vez que os momentos foram igualados a zero no primeiro passo do algoritmo ($\tau = 0$), sabe-se que as estimativas desses estão enviesadas em direção a zero necessitando que uma correção seja aplicada a estes valores para que eles não apresentem vieses. Esta correção é calculada através das equações 155 e 156, apresentadas a seguir.

$$\hat{m}_\tau \leftarrow \frac{m_\tau}{(1 - B_1^\tau)}, \quad (155)$$

$$\hat{v}_\tau \leftarrow \frac{v_\tau}{(1 - B_2^\tau)}. \quad (156)$$

Corrigidos os valores dos momentos, o passo seguinte consiste em atualizar o valor de θ , conforme representado em 157.

$$\theta_\tau \leftarrow \theta_{\tau-1} - \frac{\eta \hat{m}_\tau}{(\sqrt{\hat{v}_\tau} + \epsilon)}. \quad (157)$$

O processo é repetido até que a convergência do algoritmo seja alcançada, de modo que os valores de θ_τ encontrados sejam aqueles que minimizam a função objetivo. Maiores detalhes sobre o algoritmo e suas propriedades podem ser lidos em Kingma e Ba (2015).

4.4.4. Função de base radial aplicada a redes neurais

Outro método utilizado para treinar redes neurais artificiais é conhecido como função de base radial (Radial Basis Function, em inglês, abreviado por RBF), que de acordo com Bishop (1996) pode ser muito mais rápido que os demais utilizados para treinar Perceptrons de múltiplas camadas. A maior diferença entre esse e os métodos abordados nas seções 4.4.1 a 4.4.3 consiste no fato das redes com funções de base radial serem tratadas como um problema de ajuste de curva (aproximação) em um espaço de alta dimensionalidade, enquanto os algoritmos de backpropagation, Rprop e Adam tratam o problema aplicando uma técnica recursiva conhecida como aproximação estocástica (Haykin, 2001). Outra diferença existente é que o tipo de rede descrito nesta seção conta com somente uma camada oculta de nós, sendo que Perceptrons de múltiplas camadas podem contar com inúmeras camadas de nós ocultos.

O primeiro passo para compreender este método é considerá-lo como um interpolador exato e em seguida, extrapolar os conceitos apresentados para defini-lo como um interpolador suavizado. Os conceitos apresentados a seguir consistem em um perceptron com um único nó de *output*, mas todos os resultados podem ser extrapolados para considerar quantos nós sejam necessários na saída da rede. Os conceitos desta seção foram retirados de Bishop (1996) e as notações deste autor foram adaptadas para serem coerentes com as demais notações já apresentadas neste texto.

O objetivo de um interpolador exato é encontrar uma função $h(\cdot)$ que ao ser alimentada com o vetor de *inputs* x_i retorne exatamente o valor do *output* y_i , conforme equacionado em 158, sendo n o número de dados do banco.

$$h(x_i) = y_i, \quad i = 1, 2, \dots, n. \quad (158)$$

A técnica de função de base radial consiste em encontrar n funções de base $\phi(\cdot)$ que são não-lineares e assumem a seguinte forma:

$$\phi(\|x - x_i\|). \quad (159)$$

Interpretando a equação 159, vê-se que o valor da função está relacionado com a distância entre x e x_i , sendo que o cálculo da distância é geralmente o euclidiano e que os valores x_i são os centros das funções $\phi(\cdot)$ escolhidas. O *output*, que é o valor da função $h(\cdot)$, será a combinação linear de todas as n funções de base, conforme apresentado em 160.

$$h(x_i) = \sum_{i=1}^n w_i \phi(\|x - x_i\|). \quad (160)$$

O valor de cada peso w_i será obtido através da solução do sistema abaixo (equação 161), onde w é o vetor com todos os pesos, y é o vetor com os *outputs* e Φ é a matriz quadrada e não-singular que guarda todos os valores da função $\phi(\cdot)$.

$$w = \Phi^{-1}y. \quad (161)$$

Ao substituir os pesos encontrados através da solução da equação 161 na equação 160, obtêm-se um ajuste que passa exatamente em cada ponto amostral, definindo assim um interpolador exato.

Existem na literatura diversas formas que a função de base $\phi(\cdot)$ pode assumir para garantir que a matriz Φ seja não-singular, desde que os dados do vetor de *input* sejam todos diferentes entre si (Micchelli, 1986 apud Bishop, 1996). As funções mais comuns estão representadas a seguir.

- Função de base gaussiana, dada pela equação 162, é a mais utilizada de todas as funções de base. O parâmetro σ , que deve assumir valores positivos, controla a suavização da função de interpolação. É classificada como uma função de base localizada, pois $\phi \rightarrow 0$ quando $|x| \rightarrow \infty$.

$$\phi(x) = \exp\left(\frac{-x^2}{2\sigma^2}\right). \quad (162)$$

- Função Multiquádrica, equacionada em 163:

$$\phi(x) = (x^2 + \sigma^2)^{1/2}. \quad (163)$$

- Função Multiquádrica inversa, representada em 164:

$$\phi(x) = \frac{1}{(x^2 + \sigma^2)^{1/2}}. \quad (164)$$

- Função Thin-Plate-Spline, dada por:

$$\phi(x) = x^2 \ln(x). \quad (165)$$

- Função cúbica, equacionada em 166:

$$\phi(x) = x^3. \quad (166)$$

- Função linear, representada por:

$$\phi(x) = x. \quad (167)$$

Definida a rede de função de base radial com interpolador exato, o próximo passo é apresentar os resultados para o interpolador suavizado, que é uma adaptação do primeiro. Um interpolador exato é, por definição, aquele que não apresenta erros de predição, o que inicialmente pode parecer uma boa característica, mas que resulta em predições ruins, quando a função é aplicada em dados não utilizados para elaborá-la. Deseja-se, portanto, uma função suavizada que retorne predições mais acuradas para valores não pertencentes à base de dados inicial utilizada em sua elaboração.

A grande diferença entre a rede de função de base radial com interpolador exato e interpolador suavizado baseia-se no número “ l ” de funções de base que serão ajustadas aos dados, sendo l um número bem inferior a n . Outra diferença está no centro das funções de base utilizadas, que no caso exato assume o valor dos dados de *input* e que no caso suavizado se torna um parâmetro a ser determinado durante o treinamento da rede. Além disso, o parâmetro de suavização σ deixa de ser um valor único assumido para todas as funções de base e passa a ser um valor variável relacionado a cada função m ajustada e que também terá seu valor determinado durante o treinamento da rede. Por fim, acrescenta-se um bias w_0 na soma dos pesos e das funções, gerando a seguinte equação para o caso onde só existe um nó de *output* (lembrando que os resultados podem ser generalizados para quantos nós forem necessários):

$$\hat{y}(x) = \sum_{l=1}^L w_l \phi_l(x) + w_0. \quad (168)$$

A função de base gaussiana, apresentada em 162 para o interpolador exato, assume a seguinte forma para o interpolador suavizado, denominada função de base gaussiana esférica:

$$\phi_l(x) = \exp\left(-\frac{\|x - \mu_l\|^2}{2\sigma_l^2}\right), \quad (169)$$

onde μ_l é o vetor que determina o centro da função de base e σ_l é o parâmetro de suavização para cada função de base l .

É possível considerar que a função de base gaussiana esférica assuma uma matriz de covariância arbitrária (Σ_l) ao invés de valores σ_l , sendo assim equacionada conforme apresentado em 170.

$$\phi_l(x) = \exp\left\{-\frac{1}{2}(x - \mu_l)^T \Sigma_l^{-1}(x - \mu_l)\right\}. \quad (170)$$

Apresentada a forma matemática que as redes neurais de função de base radial assumem, o passo seguinte é apresentar como a rede é treinada neste caso. Conforme descrito por Bishop (1996), os pesos da primeira camada da rede, que relaciona os *inputs* com as funções de base, passam por um processo de aprendizado não-supervisionado, o qual determinará os parâmetros necessários das funções de base (os valores de μ_l e σ_l/Σ_l no caso da função de base gaussiana). Algumas maneiras de determinar os parâmetros supracitados encontram-se detalhadas a seguir, sendo que maiores detalhes podem ser obtidos através da leitura de Bishop (1996), onde também se encontra a explicação da utilização de modelos gaussianos mistos para determinação de μ_l , a qual não será abordada neste texto.

Uma das formas existentes de selecionar os centros das funções de base consiste em retirar uma amostra de tamanho l dos dados de *input* e fazer com que cada μ_l assuma exatamente o valor de cada ponto da amostra retirada. Utilizar a técnica de *K-means* para escolher os centros das funções de base, também é uma possibilidade e torna mais robusta a escolha de uma sub amostra do banco de dados. Existe ainda a opção de iniciar o algoritmo com todos os pontos sendo o centro de uma função de base (como no caso do interpolador exato) e ir reduzindo um a um o número de centros, iniciando o corte pelos que menos influenciam nos resultados, de modo que o número final de centros e funções de base seja suficiente para uma boa predição do *output*. Ambas as formas supracitadas dizem respeito somente ao ajuste de μ_l . Uma opção para a escolha dos valores de σ_l passa pela determinação de valores que sejam múltiplos da distância média do centro das funções de base.

Outra opção existente para ajustar os parâmetros das funções de base consiste em ajustar uma única função de base centrada em um ponto amostral, calcular os pesos da segunda camada da rede e verificar o erro obtido considerando este ajuste. Repetindo o mesmo procedimento para todos os dados amostrais, ficará retida somente a função de base cujo ajuste de pesos apresentou o menor erro. No passo seguinte, mantendo a função da etapa anterior, deve-se

adicionar mais uma função de base e avaliar o erro obtido, repetindo o procedimento para os $n-1$ dados restantes. Novamente, é retido o ponto amostral que forneceu o menor erro. O procedimento é repetido até que l pontos sejam selecionados.

Detalhadas algumas maneiras de encontrar os parâmetros das funções de base, o próximo passo consiste em compreender como são ajustados os pesos da segunda camada da rede. O objetivo durante o ajuste de pesos nesta etapa é de minimizar a função da soma de quadrados dos erros apresentada na equação 129 da seção 4.4.1. Os pesos são determinados de forma rápida através da solução do sistema de equações lineares apresentado em 171. Maiores detalhes podem ser obtidos em Bishop (1996).

$$\Phi^T \Phi W^T = \Phi^T y. \quad (171)$$

4.5. CONSIDERAÇÕES SOBRE OS DADOS UTILIZADOS

4.5.1. Divisão da base de dados

Estudos de redes neurais (e de outras técnicas de *machine learning*, como as utilizadas neste trabalho) necessitam de ao menos dois bancos de dados, da mesma forma como foi relatado na seção 3.1.5 para validação de modelos de regressão. O primeiro, denominado “banco de treinamento”, é utilizado para treinar a rede, permitindo modificações das funções de ativação, dos pesos iniciais, do número de variáveis imputadas, do número de nós na camada oculta e de outros fatores que modificam os pesos de conexão.

Finalizada a construção da rede para os dados de treinamento, o passo seguinte é testá-la no segundo banco de dados, denominado de “banco de teste”. Nesse banco, o comportamento preditivo da rede é avaliado considerando dados não utilizados em sua elaboração, pois o objetivo final do estudo é utilizá-la constantemente na predição de resultados de novas amostras.

Diversos autores sugerem diferentes partições do banco de dados total para gerar os bancos de treinamento e de teste. São valores amplamente praticados: 2/3 para dados de treinamento e 1/3 para dados de teste; 75% para dados de treinamento e 25% para dados de

teste; e 80% para dados de treinamento e 20% para dados de teste. Ainda que vários valores sejam sugeridos na literatura, a porcentagem aplicada deve ser determinada pelo pesquisador que conduz o estudo, devendo esse fazer diversos testes para determinar o número ótimo para o problema em questão.

Tendo em vista que nem sempre uma boa performance no banco de teste garante a ausência de super ajuste (*overfitting*) nos modelos, é comum separar parte dos dados do banco para avaliar a capacidade preditiva do modelo elaborado em dados não utilizados para treiná-lo ou testá-lo. Devido à reduzida quantidade de dados de algumas bases, nem sempre é possível separar amostras para pertencerem à este terceiro banco, mas quando o número amostral permite, esta é uma boa prática que deve ser adotada.

4.5.2. Alteração da escala das variáveis

As variáveis de um estudo estatístico podem ser categóricas ou contínuas e suas observações podem ser inteiras, fracionadas e até negativas. Devido a diferença de escala entre as variáveis, pode ocorrer que determinada variável influencie mais do que outras nos resultados, não pelo fato de ser a mais importante estatisticamente, mas sim pela sua escala observacional ser muito grande perante as demais.

Em redes neurais artificiais sugere-se sempre a mudança de escala das variáveis ou a estandardização das mesmas, visando mantê-las dentro de um intervalo conhecido. A normalização das variáveis é outro procedimento recomendado. Caso o banco de dados seja dividido em dados de treinamento e dados de teste é importante que o mesmo critério de transformação das variáveis seja aplicado em ambas as bases, visando manter a coerência entre os resultados.

4.6. CONSIDERAÇÕES SOBRE A REDE CONSTRUÍDA

4.6.1. Número de nós na camada oculta

Um fator que impacta fortemente no funcionamento e nos resultados obtidos com redes neurais é o número de nós da camada oculta e o número de camadas que deve ser considerado em cada rede. Sabe-se que quanto maior o número de nós e camadas, melhores serão os resultados obtidos no banco de treinamento, mas conseqüentemente o tempo para o ajuste dos pesos da rede será muito maior, além do risco de *overfitting* que pode comprometer a qualidade das predições em dados não utilizados na elaboração da rede.

Não existe um critério definido que diga o número ideal de nós e camadas ocultas, cabendo ao pesquisador realizar vários testes para determinar estes valores. A técnica de validação cruzada pode auxiliar na avaliação dos resultados obtidos, levando a conclusões sobre o número ótimo de nós e camadas ocultas. O princípio da parcimônia, que defende a simplicidade na elaboração de modelos estatísticos, deve ser aplicado também a redes neurais, de modo que redes com menos nós e camadas ocultas devem ser preferidas em relação a redes mais complexas em caso de comportamento preditivo similar entre elas.

5. IMPUTAÇÃO DE DADOS FALTANTES

A análise de dados multivariados é muitas vezes afetada pela ausência de informação em algumas amostras, gerando assim um banco de dados heterotópico que muitas vezes pode atrapalhar o andamento do estudo. Neste trabalho, foi necessário utilizar uma técnica de imputação de dados faltantes para que o estudo geometalúrgico das minas de ouro em questão pudesse avançar. Tendo em vista a importância desta técnica para o desenvolvimento do trabalho, este capítulo conta com a revisão bibliográfica de imputação de dados faltantes.

O primeiro passo para compreender a ausência de informação e como tratá-la corretamente é identificando o padrão de dados faltantes e o mecanismo de dados faltantes. Segundo Enders (2010), o padrão de dados faltantes refere-se à configuração de valores observados e valores faltantes em um banco de dados. Logo, o padrão de dados faltantes pode ser visto simplesmente como a descrição de onde os “buracos” de informação estão, o que não explica o motivo pelo qual o dado não foi coletado. Já o mecanismo de dados faltantes é aquele que descreve as possíveis relações entre as variáveis medidas e os dados faltantes. De acordo com Rubin (1976), são três os principais mecanismos de falta: dados faltantes aleatórios (*Missing at Random* - MAR), dados faltantes completamente aleatórios (*Missing Completely at Random* - MCAR) e dados faltantes não aleatórios (*Missing Not at Random* - MNAR).

Dados faltantes aleatórios são aqueles cuja falta da informação não está associada ao valor da variável na qual a falta de dados foi observada, mas sim ao valor de outra variável do banco de dados. Dados faltantes completamente aleatórios são aqueles cuja falta de dados não está relacionada nem ao valor da variável na qual a falta foi observada nem à qualquer outra variável do banco de dados. Dados faltantes não aleatórios são aqueles onde a falta de informação está associada ao valor da variável na qual a ausência de dados foi observada e também está associada a alguma outra variável do banco. Exemplos dos três mecanismos de falta relacionados ao contexto de mineração podem ser lidos em Silva (2018).

Confrontando o banco de dados geometalúrgico em estudo com os diferentes padrões de dados faltantes apresentados por Enders (2010) em seu livro, é possível concluir que a ausência da variável calcinação (que necessitará de imputação na seção

6.2) é definida como um padrão univariado, tendo em vista que todas as demais variáveis geometalúrgicas estão completamente disponíveis no banco de dados. Em relação ao mecanismo de dados faltantes, o estudo em questão pode ser classificado como MAR, uma vez que as amostras cuja calcinação não foi medida são pertencentes à primeira campanha de amostragem geometalúrgica, a qual, juntamente com a segunda campanha, focou em amostras com teor mais alto de ouro no ROM. Conseqüentemente, existe relação da variável com dados faltantes com outra variável do banco de dados, o que a torna um caso de MAR.

As técnicas de Máxima Verossimilhança e de Imputação Múltipla são vistas atualmente como o “estado da arte” das técnicas de imputação (Enders, 2010). Pelo fato de assumirem fracas suposições a respeito dos dados faltantes, tais técnicas produzem estimativas dos parâmetros com baixo viés e grande poder estatístico. Além disso, são técnicas amplamente recomendadas para imputação de casos MAR, o qual estamos lidando neste estudo. Logo, optou-se por realizar a imputação dos dados de calcinação através da técnica de Máxima Verossimilhança, que será detalhada a seguir seguindo as notações, fórmulas e derivadas apresentadas em Enders (2010).

O objetivo da estimativa por Máxima Verossimilhança é identificar os valores dos parâmetros populacionais que têm a maior probabilidade de produzir um conjunto particular informações. Logo, o primeiro passo da técnica consiste em especificar uma distribuição de probabilidade aos dados. Em seu texto, Enders (2010) assume a normalidade dos dados, logo, todas as fórmulas apresentadas pelo autor estão relacionadas ao universo Gaussiano.

No caso da distribuição normal univariada, a função de densidade de probabilidade é calculada através da fórmula 172, onde μ é a média populacional, σ^2 é a variância populacional e L_i é a probabilidade de se obter um valor y_i caso a distribuição dos dados populacionais seja normal com média μ e variância σ^2 .

$$L_i = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-0.5(y_i-\mu)^2}{\sigma^2}}. \quad (172)$$

Uma vez ser necessária uma medida que sumarie a probabilidade de todas as amostras do banco conjuntamente, deve-se calcular a probabilidade conjunta, a qual é dada pelo produto das probabilidades individuais, conforme equacionado em 173.

$$L = \prod_{i=1}^n \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{0.5(y_i-\mu)^2}{\sigma^2}} \right\}. \quad (173)$$

Tendo em vista que o resultado da equação 173 sempre fornece um número muito pequeno, expresso geralmente em notação científica, é comum adotar o logaritmo natural da função para colocar os valores em uma escala mais comum. Sabe-se que uma propriedade dos logaritmos é que o logaritmo de um produto é igual à soma dos logaritmos de cada valor individual que está sendo multiplicado. Logo, ao aplicar o logaritmo natural em 173 temos:

$$\log L = \sum_{i=1}^N \log \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{0.5(y_i-\mu)^2}{\sigma^2}} \right\}. \quad (174)$$

Desta forma, o passo seguinte é descobrir quais os valores μ e σ^2 maximizam o valor observado de $\log L$, produzindo assim o melhor ajuste aos dados em estudo. A busca pelos valores de μ e σ^2 pode ser feita de maneira iterativa ou através da utilização da primeira derivada da função $\log L$, objetivando encontrar o ponto em que esta função atinge seu valor máximo.

Ao derivar a equação 174 em relação a μ obtêm-se:

$$\frac{\partial \log L}{\partial \mu} = \frac{1}{\sigma^2} \left(-N\mu + \sum_{i=1}^N y_i \right). \quad (175)$$

Igualando a zero o lado direito da equação 175 e multiplicando ambos os lados por σ^2 temos que a estimativa de máxima verossimilhança da média é dada por:

$$\hat{\mu} = \sum_{i=1}^N \frac{y_i}{N}. \quad (176)$$

É imediato perceber, que a fórmula apresentada em 176 é a mesma utilizada para o cálculo da média.

Ao derivar a equação 174 em relação a σ^2 temos:

$$\frac{\partial \log L}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \sum_{i=1}^N \frac{(y_i - \mu)^2}{2\sigma^4}. \quad (177)$$

Igualando o lado direito a zero e isolando o termo σ^2 obtêm-se:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^N (y_i - \mu)^2}{N}. \quad (178)$$

Logo, a estimativa de máxima verossimilhança da variância é dada pela mesma fórmula utilizada para calcular a variância populacional.

Para extrapolar os conceitos supracitados para o caso de dados multivariados, devemos trabalhar com a distribuição normal multivariada para realizar as estimativas de máxima verossimilhança. Desta forma, a equação 172 que apresentava a função de densidade de probabilidade normal univariada deverá ser substituída pela equação representada em 179.

$$L_i = \frac{1}{(2\pi)^{k/2} |\Sigma|^{1/2}} e^{-0.5(y_i - \mu)^T \Sigma^{-1} (y_i - \mu)}, \quad (179)$$

onde μ é o vetor de médias, Σ é a matriz de covariâncias, k é o número de variáveis e y_i é o vetor com as k observações relacionadas à i -ésima amostra do banco.

Conforme descrito anteriormente, a escala dos valores é muito pequena, sendo recomendado trabalhar com os valores de L_i em escala logarítmica. Desta forma, ao aplicar o logaritmo natural na equação 179 temos a equação representada em 180.

$$\log L_i = -\frac{k}{2} \log(2\pi) - \frac{1}{2} \log|\Sigma| - \frac{1}{2} (y_i - \mu)^T \Sigma^{-1} (y_i - \mu). \quad (180)$$

Seguindo a mesma lógica do caso univariado, para encontrar o valor de $\log L$ que resume a informação para todo banco de dados no caso multivariado, basta somar os resultados da fórmula 180 para todas as observações, conforme apresentado em 181. Quanto maior for o valor obtido de $\log L$, melhor será o ajuste dos parâmetros escolhidos aos dados em estudo.

$$\log L = \sum_{i=1}^N \log L_i. \quad (181)$$

Tendo finalizada a apresentação da técnica de máxima verossimilhança, o próximo passo é utilizá-la na inserção de dados faltantes. Para isso, é necessário redefinir a equação 180 para que amostras incompletas sejam consideradas de maneira distinta no cálculo de máxima verossimilhança a ser realizado. Logo a equação passa a ser definida por:

$$\log L_i = -\frac{k_i}{2} \log(2\pi) - \frac{1}{2} \log|\Sigma_i| - \frac{1}{2} (y_i - \mu_i)^T \Sigma_i^{-1} (y_i - \mu_i), \quad (182)$$

onde k_i é o número de variáveis completas para a amostra i , y_i é o vetor com os valores observados para a amostra i , Σ_i é a matriz de covariâncias e μ_i é o vetor de médias, ambos considerando somente as variáveis completas para a amostra i .

Após redefinido o cálculo do logaritmo da máxima verossimilhança individual, o passo seguinte consiste em redefinir a medida que resume o valor de máxima verossimilhança para todas as amostras da base de dados, sejam essas completas ou com dados faltantes. A equação 183 mostra a fórmula utilizada nesta situação.

$$\log L = \log L_{\text{Dados Completos}} + \log L_{\text{Dados Incompletos}} \quad (183)$$

No caso em que mais de uma variável apresente incompletude, é necessário calcular o log L para os dados completos, seguido do log L considerando somente os dados completos para a primeira variável com observações faltantes e, finalmente, deve-se calcular o log L para os dados completos considerando a segunda variável com observações faltantes. Esta lógica de cálculo se estende para o caso de mais do que duas variáveis envolvidas.

Definido o log L para dados completos e incompletos, o passo seguinte consiste em apresentar o algoritmo EM - *Expectation Maximization* (Enders, 2010). O algoritmo consiste em duas partes: a primeira, chamada etapa E e a segunda, chamada etapa M ambas explicadas a seguir considerando um caso bivariado.

O algoritmo inicia considerando somente os dados completos e a partir deles são estimados o vetor de médias μ (calculado por meio das fórmulas 184 e 185) e a matriz de covariância Σ (calculada utilizando as fórmulas 186, 187 e 188). Deve-se calcular o valor do log L nesta primeira parte do algoritmo e armazená-lo para fins de comparação com os valores de log L que surgirão nas etapas seguintes.

$$\hat{\mu}_Y = \frac{1}{N} \sum Y, \quad (184)$$

$$\hat{\mu}_X = \frac{1}{N} \sum X, \quad (185)$$

$$\hat{\sigma}_Y^2 = \frac{1}{N} \left(\sum Y^2 - \frac{(\sum Y)^2}{N} \right), \quad (186)$$

$$\hat{\sigma}_X^2 = \frac{1}{N} \left(\sum X^2 - \frac{(\sum X)^2}{N} \right), \quad (187)$$

$$\hat{\sigma}_{X,Y} = \frac{1}{N} \left(\sum XY - \frac{\sum X \sum Y}{N} \right). \quad (188)$$

O passo E consiste em utilizar o vetor e a matriz supracitados para construir um modelo de regressão para prever a variável com dados faltantes (Y) a partir da informação disponível na variável completa (X). As equações utilizadas para o cálculo do intercepto e da inclinação da reta de regressão estão representadas em 189 e 190, enquanto a equação da variância residual da regressão de Y dado X está representada em 191 e a fórmula do valor de \hat{Y}_i predito através do modelo de regressão está representada em 192.

$$\hat{\beta}_1 = \frac{\hat{\sigma}_{X,Y}}{\hat{\sigma}_X^2}, \quad (189)$$

$$\hat{\beta}_0 = \hat{\mu}_Y - \hat{\beta}_1 \hat{\mu}_X, \quad (190)$$

$$\hat{\sigma}_{Y|X}^2 = \hat{\sigma}_Y^2 - \hat{\beta}_1^2 \hat{\sigma}_X^2, \quad (191)$$

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i. \quad (192)$$

Construído o modelo, os locais onde a variável dependente apresentou dados faltantes são completados pela estimativa gerada através do modelo de regressão em questão. Concluída esta etapa, inicia-se o passo M, no qual um novo vetor de médias μ e uma nova matriz de covariância Σ são calculados utilizando as equações 184 até 188, mas agora considerando os dados preenchidos pelo modelo de regressão no passo E. Além disso, o valor de $\sum Y^2$ na equação 186 deve ser substituído pelo valor de $\hat{Y}_i + \hat{\sigma}_{Y|X}^2$ obtido através do modelo de regressão calculado no passo E. Calcula-se o valor do log L e armazena-se o valor obtido. Utilizando o novo vetor de médias e a nova matriz de

covariância o modelo de regressão deve ser recalculado. Este processo deve se repetir até que não haja mudança significativa no valor do log L do passo atual em relação ao passo anterior. Logo, os valores finais de μ e Σ são os valores a serem considerados em cálculos futuros, por se tratarem de melhores estimativas para o conjunto de dados em questão. Conforme descrito por Enders (2010), o algoritmo EM não realiza a imputação de dados faltantes, mas é comum adotar o valor de \hat{Y}_i observado na última rodada do algoritmo como o valor imputado.

6. DESENVOLVIMENTO

Este capítulo contém uma breve descrição dos depósitos estudados e dos dados coletados para elaboração do modelo geometalúrgico para predição da recuperação mássica e metalúrgica dos produtos e dos rejeitos, apresenta as técnicas utilizadas e os resultados obtidos ao aplicar cada uma delas no conjunto de dados em questão.

6.1. MINA DE VAZANTE

6.1.1. Apresentação do depósito

A cidade de Vazante, onde fica localizada a mina de mesmo nome na qual o estudo será desenvolvido, está localizada à noroeste do estado de Minas Gerais, próximo às cidades de Paracatu e Patos de Minas. A cidade de Belo Horizonte, capital do estado, está distante 507 quilômetros da cidade de Vazante e Brasília, a capital do país, dista 348 quilômetros da mesma. O mapa apresentado na Figura 21 ilustra a localização geográfica do município de Vazante (ponto B).

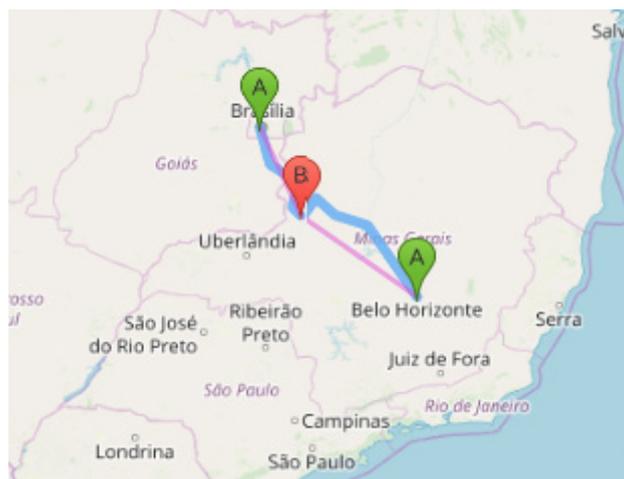


Figura 21 - Localização geográfica do município de Vazante (MG)

Atualmente, a Nexa Resources, antiga Votorantim Metais S/A, é principal produtora brasileira de zinco. De acordo com dados fornecidos pela própria empresa em seu *website* (<https://www.nexareport.com/2018/desempenho-economico-financeiro>, acessado em 11/09/2019), no ano de 2018 a produção de metal contido da mina de Vazante foi de 140.8 mil toneladas de zinco e 1.2 mil toneladas de chumbo. O preço médio da tonelada de zinco na London Metal Exchange foi de US\$ 2922 por tonelada no ano de 2018, enquanto o preço médio de chumbo no mesmo ano foi de US\$ 2242 por tonelada. Dentre as diversas aplicações do zinco se destacam a proteção do aço contra corrosão e oxidação (galvanização) e a composição de ligas zamac. Na forma de composto químico, é utilizado na produção de pilhas e baterias, na vulcanização de borrachas, na indústria cerâmica, têxtil, na indústria de cosméticos, entre outros.

O depósito de Vazante é definido como um depósito de zinco não sulfetado e, de acordo com Lemos (2011), os principais tipos de minério ali encontrados apresentam predominância de willemita (Zn_2SiO_4) que contém 58.5% de zinco contido e calamina ($2ZnO.SiO_2.H_2O$) que contém 54% do metal em questão. O principal mineral de chumbo encontrado no depósito é a galena (PbS). Por se tratar de um minério silicatado de zinco com chumbo, a rota de processo adotada em Vazante busca a concentração de ambos metais a partir de duas etapas de flotação. Na primeira, ocorre a flotação de chumbo, gerando o primeiro produto vendável: concentrado de chumbo. O rejeito desta etapa alimenta a flotação de zinco, na qual o concentrado de zinco é o produto vendável e o rejeito é enviado para a barragem. A Figura 22 apresenta o fluxograma simplificado da usina de Vazante.

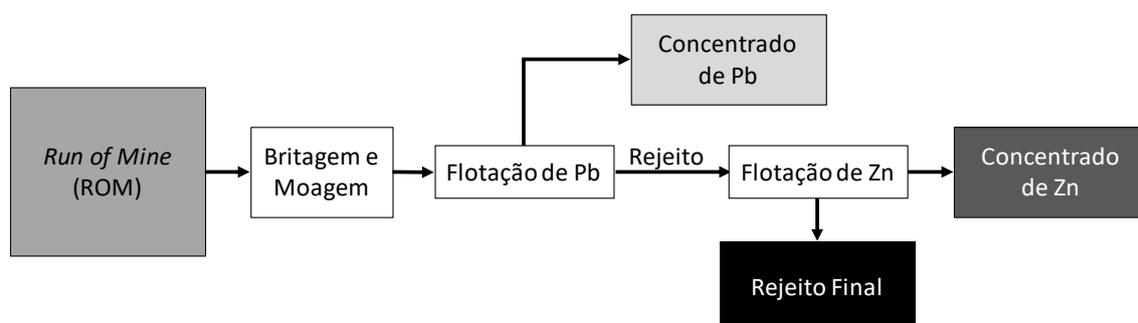


Figura 22 - Fluxograma simplificado da usina de Vazante

Os reagentes utilizados no processo de beneficiamento são: AGLP como dispersante, MICB como espumante, sulfato de sódio como ativador do zinco e como regulador de pH, amina como coletora da willemita (Bechir *et. al*, 2019) e o amil xantato como coletor da galena.

6.1.2. Apresentação dos dados e estatísticas descritivas

Os dados utilizados neste estudo fazem parte de dois estudos distintos realizados na mina de Vazante. O primeiro deles, realizado em 2017, consistiu na análise química de 28 elementos no ROM, seguida da flotação do material visando obter um concentrado de zinco, um concentrado de chumbo e um rejeito final. O segundo estudo, realizado em 2018, possuiu somente os teores de Zn e Pb no ROM e após as análises químicas as amostras seguiam para flotação visando obter os mesmos concentrados e rejeito do primeiro estudo realizado. Ambos estudos consideraram fixa a quantidade de reagentes utilizada nos testes de flotação de bancada, visando garantir a existência de um teste padrão.

As variáveis a serem preditas pelo modelo geometalúrgico da mina de Vazante podem ser calculadas através das fórmulas 193 a 198:

$$\text{Recuperação Mássica do Conc. de Zn} = \frac{\text{Massa do concentrado de Zn}}{\text{Massa do ROM}}, \quad (193)$$

$$\text{Recuperação Mássica do Conc. de Pb} = \frac{\text{Massa do concentrado de Pb}}{\text{Massa do ROM}}, \quad (194)$$

$$\text{Recuperação Mássica do Rejeito Final} = \frac{\text{Massa do Rejeito Final}}{\text{Massa do ROM}}, \quad (195)$$

$$\text{Rec. Met. de Zn no Conc. de Zn} = \frac{\text{Zn\% no Conc. de Zn}}{\text{Zn\% no ROM}} \times \text{Rec. Mássica do Conc. de Zn}, \quad (196)$$

$$\text{Rec. Met. de Zn no Conc. de Pb} = \frac{\text{Zn\% no Conc. de Pb}}{\text{Zn\% no ROM}} \times \text{Rec. Mássica do Conc. de Pb}, \quad (197)$$

$$\text{Rec. Met. de Zn no Rejeito Final} = \frac{\text{Zn\% no Rej. Final}}{\text{Zn\% no ROM}} \times \text{Rec. Mássica do Rej. Final}. \quad (198)$$

É importante ressaltar que as variáveis acima apresentadas não se encontram representadas em termos percentuais e sim com valores entre 0 e 1 para facilitar o tratamento estatístico de algumas técnicas que serão utilizadas nas seções seguintes. A interpretação dos números é imediata, bastando multiplicar por 100 os valores observados para colocá-los em termos percentuais.

Considerando ambos estudos, o banco de dados de amostras geometalúrgicas possui 140 dados, os quais serão utilizados para elaboração do modelo a ser proposto. Tendo em vista que é necessário um banco de dados isotópico para elaboração dos modelos preditivos, as informações das 28 variáveis analisadas no ROM na primeira parte do estudo foram reduzidas a utilização dos teores de Zn e Pb, uma vez que estes elementos eram os únicos disponíveis no banco de dados do estudo realizado em 2018. A Tabela 4 apresenta a estatística descritiva dos teores e a Figura 23 ilustra os histogramas de zinco e chumbo.

Tabela 4 - Estatística descritiva dos teores no ROM – Mina de Zinco

Variável	N	Média	Desvio Padrão	Mínimo	Q1	Mediana	Q3	Máximo
Zn	140	11.734	5.449	3.430	7.671	10.171	15.194	25.660
Pb	140	0.333	0.177	0.113	0.238	0.290	0.369	1.230

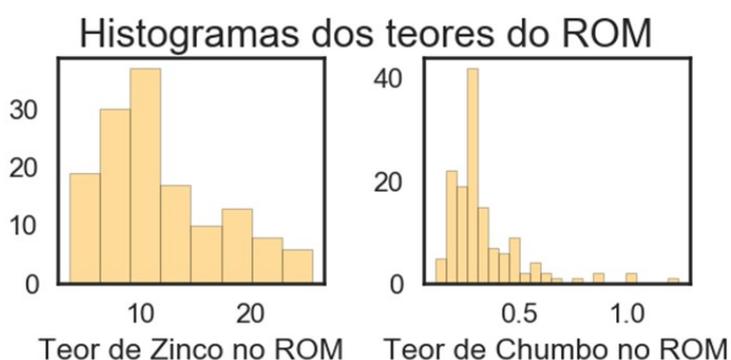


Figura 23 - Histogramas dos teores do ROM – Mina de Zinco

O teor médio de zinco dos dados coletados é igual a 11.734 e 50% dos dados apresentam teor entre 7.671 e 15.194. O teor mínimo observado é de 3.43% de zinco e o maior teor de zinco observado é igual a 25.66%. Para chumbo, o teor médio é de 0.333 e

50% dos dados coletados possuem teor entre 0.238 e 0.369. O menor teor de chumbo observado foi igual a 0.113% e o teor mais alto desta mesma variável foi de 1.23%. Analisando as estatísticas descritivas é possível concluir que os dados coletados são representativos, em termos de teor do ROM, em relação aos teores que se observa na mina de Vazante para os alvos atualmente lavrados, confirmando uma boa escolha de amostras para compor o banco de dados geometalúrgico.

Os 140 dados coletados são provenientes de três alvos distintos dentro da mina de Vazante: Extremo Norte (com 31 amostras), Lumiadeira (com 53 dados) e Sucuri (com 56 amostras). Não há informação geológica relacionada as amostras, logo, o alvo ao qual a amostra pertence é a única informação que pode agregar algum conhecimento adicional. Desta forma, optou-se por avaliar os teores de zinco e chumbo separados por alvo, conforme ilustrado nos box-plots da Figura 24.

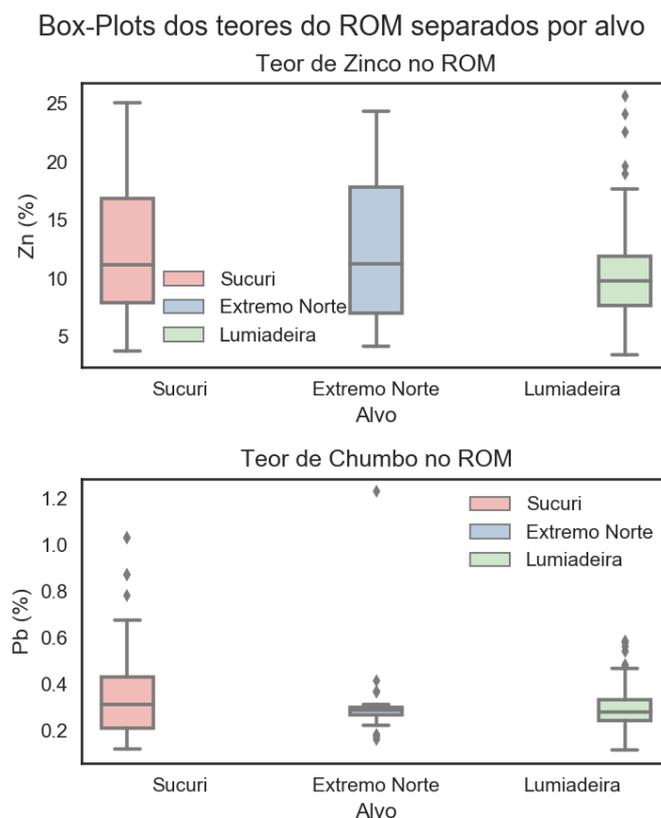


Figura 24 - Box-plots dos teores de zinco e chumbo separados por alvo

Analisando a figura, percebe-se que os diferentes alvos possuem teores médios similares, ainda que o alvo Lumiadeira apresente menor variabilidade do teor de zinco no ROM e que o alvo Extremo Norte apresente menor variabilidade do teor de chumbo.

Em relação as variáveis dependentes em estudo, as estatísticas descritivas encontram-se apresentadas na Tabela 5 e os histogramas estão ilustrados na Figura 25.

Tabela 5 - Estatística descritiva das variáveis de processo – Mina de Zinco

Variável	N	Média	Desvio Padrão	Mínimo	Q1	Mediana	Q3	Máximo
Rec. Mássica do Conc. de Zn	140	0.236	0.147	0.039	0.120	0.195	0.323	0.641
Rec. Mássica do Conc. de Pb	140	0.014	0.007	0.002	0.008	0.014	0.017	0.053
Rec. Mássica do Rej. Final	140	0.751	0.146	0.352	0.659	0.792	0.861	0.955
Rec. Met. de Zn no Conc. de Zn	140	0.722	0.158	0.358	0.613	0.728	0.850	0.961
Rec. Met. de Zn no Conc. de Pb	140	0.010	0.005	0.001	0.006	0.009	0.012	0.036
Rec. Met. de Zn no Rej. Final	140	0.268	0.158	0.029	0.137	0.265	0.377	0.629

Histogramas das variáveis de processo

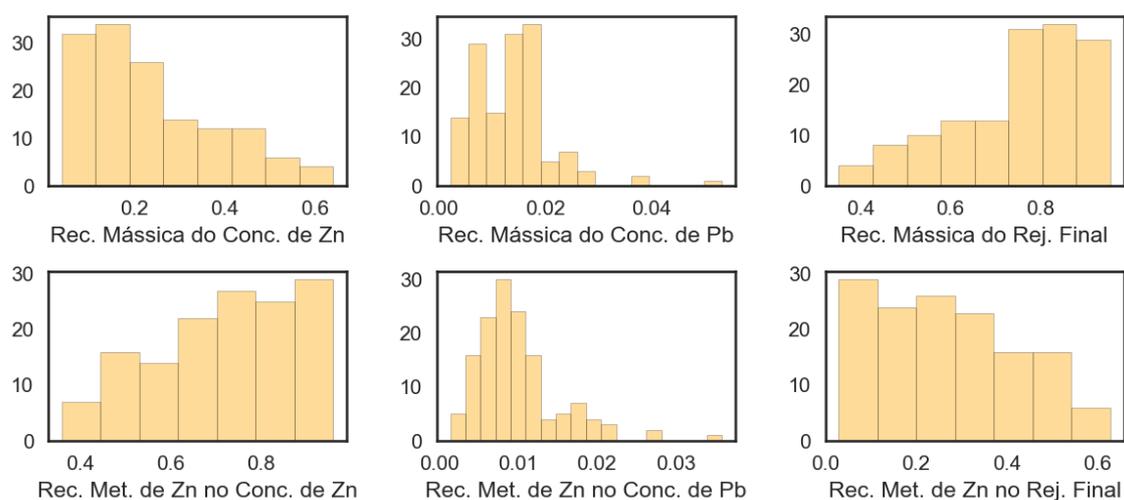


Figura 25 - Histogramas das variáveis de processo – Mina de Zinco

Analisando os resultados da Tabela 5 e da Figura 25, é possível perceber que na maioria das amostras o concentrado de chumbo possui a menor massa entre os produtos gerados no circuito de flotação, sendo responsável em média por 1.4% da massa total. Em relação à recuperação metalúrgica, vê-se que em média 1% da recuperação metalúrgica de zinco está no concentrado de chumbo.

Em média, 75.1% da massa gerada e 26.8% da recuperação metalúrgica de zinco é destinada ao rejeito final. 25% dos dados apresentaram recuperação metalúrgica de zinco no rejeito superior a 37.7%.

Em relação ao produto de interesse na mina de Vazante, o concentrado de zinco, vê-se uma recuperação mássica média de 23.6%, enquanto a recuperação metalúrgica média deste produto é de 72.2%. A recuperação metalúrgica mínima observada neste produto é de 35.8% e a máxima é 96.1%

Por mais que os teores de zinco e chumbo no ROM apresentem distribuições similares em todos os alvos estudados, é possível que a resposta metalúrgica por alvo não seja similar. Para avaliar a veracidade desta hipótese, o teste de Kruskal Wallis (Siegel e Castellan Jr, 2006) foi calculado para cada uma das seis variáveis dependentes em estudo, visando avaliar se os alvos pertencem a mesma população ou se existe indício da presença de diferentes domínios geometalúrgicos na amostra coletada. O fundamento teórico do teste de Kruskal Wallis encontra-se descrito a seguir.

Primeiramente, os dados são organizados em ordem crescente desconsiderando o grupo de origem (no caso em estudo, o alvo). Cada amostra recebe o número de sua posição no conjunto de dados, valor denominado posto, de modo que a amostra com menor valor observado receba o posto um e a amostra de maior valor receba o posto N. Em seguida, é calculada a média dos postos por grupo. Caso as amostras pertençam à mesma população, espera-se que os postos médios sejam quase os mesmos, enquanto que se as amostras vierem de populações com medianas diferentes, os postos médios devem diferir significativamente. O teste de Kruskal Wallis trabalha com as diferenças entre os postos médios de acordo com a estatística de teste apresentada na equação 199.

$$KW = \frac{12}{N(N+1)} \sum_{i=1}^k n_i (\bar{R}_i - \bar{R})^2, \quad (199)$$

onde k é o número de grupos estudados, N é o número de dados considerando todos os grupos conjuntamente, n_i é o número de amostras no i-ésimo grupo, \bar{R}_i é a média dos postos no i-ésimo grupo e \bar{R} é a média dos postos de todos os grupos conjuntamente.

O valor obtido através do uso da estatística de teste deve ser comparado com o da distribuição qui-quadrado com $k-1$ graus de liberdade. Uma vez que a hipótese nula do teste de Kruskal Wallis assume que os grupos (alvos) pertencem a mesma população, um valor-p maior ou igual a 0.05 indica a aceitação desta hipótese, enquanto um valor-p menor do que 0.05 aponta para a rejeição, considerando um nível de significância igual a 5%.

Os box-plots apresentados nas Figuras 26 e 27 ilustram as variáveis dependentes separadas por alvo e o valor-p do teste de Kruskal Wallis encontra-se apresentado em cada um deles. Observa-se a rejeição da hipótese nula para a Recuperação Mássica do Concentrado de Chumbo, para a Recuperação Metalúrgica de Zinco no Concentrado de Zinco e para a Recuperação Metalúrgica de Zinco no Rejeito Final, indicando, portanto, que existem diferentes domínios geometalúrgicos na amostra trabalhada.

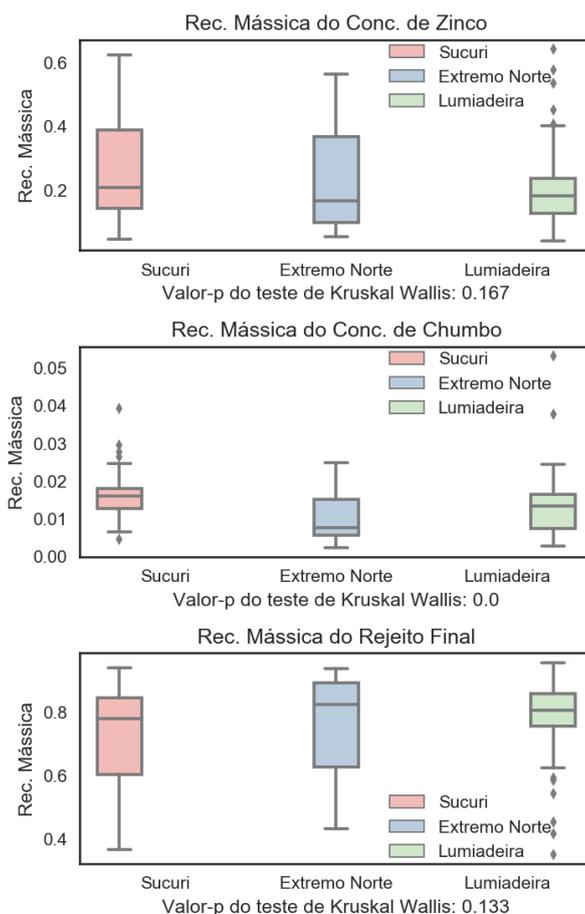


Figura 26 - Box-plots das variáveis de Recuperação Mássica por alvo – Mina de Zinco

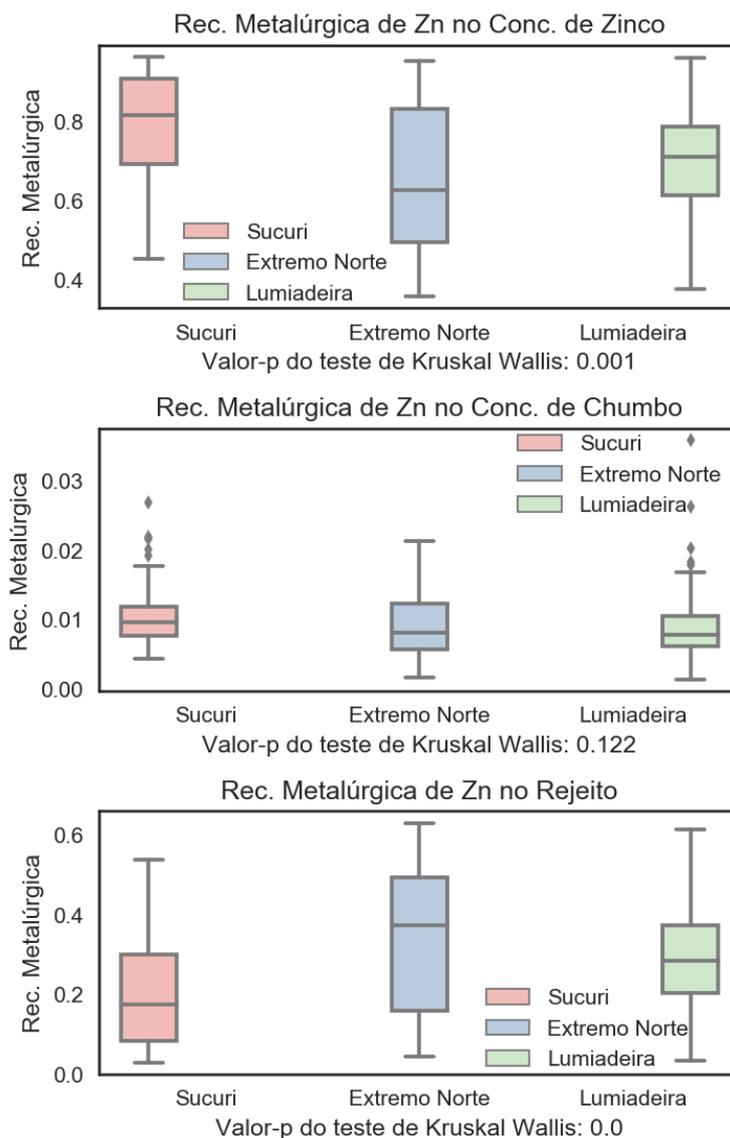


Figura 27 - Box-plots das variáveis de Recuperação Metalúrgica por alvo – Mina de Zinco

Note que as diferenças observadas entre os alvos estão associadas às características químicas ou mineralógicas específicas que não podem ser explicadas somente através dos teores de Zn e Pb no ROM. De acordo com estudos anteriormente realizados com o minério de Vazante, sabe-se, por exemplo, que o minério proveniente de Extremo Norte necessita de um maior tempo de cominuição para que as partículas estejam totalmente liberadas, logo, a recuperação metalúrgica de zinco no concentrado de zinco é negativamente afetada, fazendo com que tal alvo tenha respostas inferiores aos demais. Outras características mineralógicas certamente influenciam na resposta do minério no processo de beneficiamento, mas tal impacto é atualmente desconhecido,

sendo necessário conduzir, no futuro, estudos químicos e mineralógicos detalhados dos diferentes tipos de minério do depósito, visando esclarecer quais fatores impactam diretamente no comportamento metalúrgico de cada alvo.

Tendo em vista a impossibilidade de incorporar outras variáveis no modelo e a dificuldade de se construir um modelo para cada alvo (devido ao baixo número amostral em cada grupo), é necessário categorizar a informação da origem do minério para incluí-la como variável independente no conjunto de dados. Essa informação será utilizada para elaborar o modelo geometalúrgico, de modo que, indiretamente, as características mineralógicas de cada alvo sejam levadas em consideração. Para isso, duas variáveis categóricas (conhecidas como variáveis *dummy*, em estatística) foram construídas: a variável D1, que assume o valor 1 caso a amostra pertença ao alvo Extremo Norte e assume o valor 0 em caso contrário; e a variável D2, que assume o valor 1 caso a amostra pertença a Lumiadeira e assume o valor 0 em caso contrário. Caso a amostra pertença ao alvo Sucuri ambas variáveis, D1 e D2, assumirão o valor 0, de modo que cada alvo estará representado pelo seu próprio conjunto de códigos.

Desta forma, todo o estudo que será apresentado a seguir terá como variáveis independentes as variáveis categóricas D1 e D2 e os teores de Zn e Pb. As duas últimas variáveis foram transformadas para assumir valores entre 0 e 1, tendo em vista a diferença de grandeza entre elas (conforme ilustrado na Tabela 4). Assim, o menor valor amostral assumirá o valor 0 e o maior valor amostral assumirá o valor 1, sendo que os demais dados assumirão valores contidos neste intervalo. Esta transformação é importante uma vez que variáveis com maior escala observacional podem ter maior influência nos resultados devido à sua maior variabilidade. Logo, todos os modelos elaborados a seguir levarão em conta os dados de zinco e chumbo transformados.

O banco de dados da mina de zinco em questão possui 140 amostras, as quais foram divididas aleatoriamente da seguinte forma: 115 amostras foram utilizadas para treinar e testar a qualidade dos modelos ajustados, enquanto as 25 amostras restantes foram separadas das demais com o objetivo de avaliar a qualidade dos modelos construídos em dados não utilizados nas bases de treinamento e de teste, para comprovar a ausência de *overfitting* e a qualidade das previsões geradas.

6.1.3. Modelos construídos com dados de treinamento e teste

O primeiro passo na construção dos modelos de predição consistiu em elaborá-los utilizando os 115 dados de treinamento e teste disponíveis. Logo, todos os modelos elaborados nesta seção consideram este número amostral.

6.1.3.1. Análise de Regressão

A primeira técnica testada para fins preditivos no estudo em questão foi a análise de regressão, tendo em vista que esta é a técnica mais popularmente utilizada na criação de modelos geometalúrgicos. As variáveis independentes testadas no ajuste dos modelos são as variáveis D1 e D2 e os teores de zinco e chumbo transformados.

Para cada uma das variáveis dependentes, foi gerado um modelo de regressão que inicialmente tinha como preditoras as quatro variáveis independentes disponíveis. Em seguida, foi calculado o teste F para cada um dos modelos, visando identificar se ao menos uma das variáveis independentes era estatisticamente significativa para explicar a variável dependente em estudo. Assumindo um nível de significância de 5%, foi possível concluir que todos os modelos possuíam ao menos uma variável significativa, conforme pode ser visto no valor “Prob (F-statistic)” apresentado nas Figuras 28 a 33. Dada a rejeição da hipótese nula dos testes F para todas as variáveis dependentes, foi necessário calcular os testes t individuais, visando avaliar a significância individual, ao nível de 5%, de cada variável independente inserida no modelo. Quando alguma variável não era estatisticamente significativa, ela era removida e todos os coeficientes eram recalculados, até que o modelo final permanecesse somente com variáveis significativas estatisticamente (Valor-p < 0.05).

A Figura 28 ilustra os resultados obtidos para o modelo construído para prever a recuperação mássica do concentrado de zinco. As variáveis independentes utilizadas no modelo final foram o teor de zinco e as variáveis D1 e D2. A equação que relaciona as variáveis encontra-se apresentada em 200.

$$\text{Rec. Mass. Conc. Zn} = 0.0289 + 0.5907 * \text{Zn} - 0.0336 * \text{D1} - 0.0157 * \text{D2}. \quad (200)$$

Ao interpretar o modelo acima percebe-se que quando a amostra é proveniente do Extremo Norte (D1), o valor da recuperação mássica é reduzido em 0.0336, enquanto para as amostras oriundas de Lumiadeira (D2) esta redução é de 0.0157. O modelo reflete o fato de que o alvo Sucuri é o que apresenta os maiores valores de recuperação mássica de zinco quando um mesmo teor de zinco no ROM é considerado para todos os alvos. Além disso, é possível observar que um maior teor de zinco na alimentação leva a uma maior recuperação mássica do concentrado de zinco. O coeficiente de determinação do modelo em questão é igual a 0.978, indicando que 97.8% da variabilidade desta variável pode ser explicada através do modelo elaborado.

Analisando o histograma dos resíduos e o teste de Anderson Darling, apresentados na Figura 34, percebe-se a violação da suposição de normalidade dos resíduos deste modelo, uma vez que o valor da estatística de teste de Anderson Darling foi de 2.197 e o valor crítico com 5% de significância é de 0.762. O gráfico dos valores preditos *versus* os resíduos, ilustrado na Figura 35, mostra heterocedasticidade dos resíduos, violando mais uma suposição para a construção de um modelo válido.

OLS Regression Results						
Dep. Variable: Rec. Mássica Conc. Zn			R-squared: 0.978			
Model: OLS			Adj. R-squared: 0.977			
Method: Least Squares			F-statistic: 1615			
Date: Thu, 28 May 2020			Prob (F-statistic): 2.28E-91			
Time: 13:46:55			Log-Likelihood: 273.5			
No. Observations: 115			AIC: -539			
Df Residuals: 111			BIC: -528			
Df Model: 3						
Covariance Type: nonrobust						
	coef	std err	t	P> t	[0.025	0.975]
const	0.0289	0.005	5.918	0	0.019	0.039
Zn	0.5907	0.009	67.904	0	0.574	0.608
D1	-0.0336	0.006	-5.923	0	-0.045	-0.022
D2	-0.0157	0.005	-3.203	0.002	-0.025	-0.006
Omnibus: 10.551			Durbin-Watson: 2.067			
Prob(Omnibus): 0.005			Jarque-Bera (JB): 10.657			
Skew: -0.687			Prob(JB): 0.00485			
Kurtosis: 3.582			Cond. No. 5.28			

Figura 28 - Dados do modelo de regressão construído para Rec. Mássica do Conc. de Zn

A Figura 29 apresenta as características estatísticas do modelo construído para prever a recuperação mássica do concentrado de chumbo. A equação obtida encontra-se representada em 201.

$$\text{Rec. Mass. Conc Pb} = 0.0146 - 0.0054 * \text{Zn}. \quad (201)$$

O modelo construído indica que se a amostra pertence ao alvo Extremo Norte, sua recuperação mássica do concentrado de chumbo é menor do que a observada quando os outros dois alvos são alimentados. Analisando o coeficiente de determinação do modelo, é evidente perceber seu fraco poder preditivo, uma vez que somente 8.8% da variabilidade total da recuperação mássica do concentrado de chumbo pode ser explicada através do modelo elaborado.

Ao analisar as estatísticas dos resíduos, apresentadas nas Figuras 34 e 35, percebe-se resíduos homocedásticos mas não normalmente distribuídos, invalidando uma das suposições necessárias.

OLS Regression Results						
Dep. Variable: Rec. Mássica Conc. Pb			R-squared: 0.088			
Model: OLS			Adj. R-squared: 0.08			
Method: Least Squares			F-statistic: 10.95			
Date: Thu, 28 May 2020			Prob (F-statistic): 0.00125			
Time: 13:47:26			Log-Likelihood: 404.5			
No. Observations: 115			AIC: -805			
Df Residuals: 113			BIC: -799.5			
Df Model: 1						
Covariance Type: nonrobust						
	coef	std err	t	P> t	[0.025	0.975]
const	0.0146	0.001	19.14	0	0.013	0.016
D1	-0.0054	0.002	-3.31	0.001	-0.009	-0.002
Omnibus: 65.241			Durbin-Watson: 1.994			
Prob(Omnibus): 0			Jarque-Bera (JB): 314.433			
Skew: 1.914			Prob(JB): 5.27E-69			
Kurtosis: 10.139			Cond. No. 2.56			

Figura 29 - Dados do modelo de regressão construído para Rec. Mássica do Conc. de Pb

A Figura 30 apresenta os dados do modelo construído para prever a recuperação mássica do rejeito final. Os resultados mostram que 97.7% da variância total da variável em questão pode ser explicada através da equação 202.

$$Rec\ Mass.\ Rej.\ Final = 0.9538 - 0.5866 * Zn + 0.0399 * D1 + 0.0179 * D2. \quad (202)$$

Percebe-se, que quanto maior o teor de zinco na alimentação, menor será a recuperação mássica do rejeito final. Além disso, vê-se que quando a amostra é proveniente do alvo Extremo Norte sua recuperação mássica do rejeito é 3.99% maior que a recuperação observada no alvo Sucuri e que quando as amostras são de Lumiadeira o aumento em relação à Sucuri é de 1.79%.

A análise de resíduos apresentada nas Figuras 34 e 35 mostram resíduos heterocedásticos e que não seguem uma distribuição normal, logo, ambas suposições foram violadas.

OLS Regression Results						
Dep. Variable: Rec. Mássica Rej. Final			R-squared: 0.977			
Model: OLS			Adj. R-squared: 0.976			
Method: Least Squares			F-statistic: 1579			
Date: Thu, 28 May 2020			Prob (F-statistic): 7.81E-91			
Time: 13:47:50			Log-Likelihood: 272.89			
No. Observations: 115			AIC: -537.8			
Df Residuals: 111			BIC: -526.8			
Df Model: 3						
Covariance Type: nonrobust						
	coef	std err	t	P> t	[0.025	0.975]
const	0.9538	0.005	194.082	0	0.944	0.964
Zn	-0.5866	0.009	-67.072	0	-0.604	-0.569
D1	0.0399	0.006	6.985	0	0.029	0.051
D2	0.0179	0.005	3.653	0	0.008	0.028
Omnibus: 6.412			Durbin-Watson: 2.106			
Prob(Omnibus): 0.041			Jarque-Bera (JB): 6.005			
Skew: 0.545			Prob(JB): 0.0497			
Kurtosis: 3.253			Cond. No. 5.28			

Figura 30 - Dados do modelo de regressão construído para Rec. Mássica do Rejeito Final

A Figura 31 contém as informações estatísticas do modelo elaborado para prever a recuperação metalúrgica de zinco no concentrado de zinco. Analisando os dados, percebe-se que 75% da variabilidade da variável dependente encontra-se explicada pelo modelo elaborado, o qual está equacionado em 203.

$$\text{Rec Met Zn Conc Zn} = 0.5889 + 0.5125 * \text{Zn} - 0.1456 * \text{D1} - 0.0664 * \text{D2}. \quad (203)$$

De acordo com o modelo, quando a amostra é pertencente ao alvo Extremo Norte sua recuperação metalúrgica de zinco no concentrado de zinco cai 14.56% em relação à amostras de Sucuri, considerando um mesmo teor de zinco no ROM para ambos os alvos. Já para o alvo Lumiadeira, a queda em relação ao alvo Sucuri é de 6.64%, também considerando um mesmo valor de zinco na alimentação.

Os gráficos com as análises de resíduos mostram resíduos normais, porém heterocedásticos, o que viola uma das suposições estatísticas necessárias para a construção de um modelo válido.

OLS Regression Results						
Dep. Variable: Rec. Met. de Zn no Conc. de Zn			R-squared: 0.75			
Model: OLS			Adj. R-squared: 0.743			
Method: Least Squares			F-statistic: 110.9			
Date: Thu, 28 May 2020			Prob (F-statistic): 2.98E-33			
Time: 13:48:11			Log-Likelihood: 126.87			
No. Observations: 115			AIC: -245.7			
Df Residuals: 111			BIC: -234.8			
Df Model: 3						
Covariance Type: nonrobust						
	coef	std err	t	P> t	[0.025	0.975]
const	0.5889	0.017	33.661	0	0.554	0.624
Zn	0.5125	0.031	16.461	0	0.451	0.574
D1	-0.1456	0.02	-7.165	0	-0.186	-0.105
D2	-0.0664	0.017	-3.795	0	-0.101	-0.032
Omnibus: 6.622			Durbin-Watson: 2.064			
Prob(Omnibus): 0.036			Jarque-Bera (JB): 6.357			
Skew: -0.57			Prob(JB): 0.0416			
Kurtosis: 3.166			Cond. No. 5.28			

Figura 31 - Dados do modelo de regressão construído para Rec. Met. de Zn no Conc. de Zn

A Figura 32 apresenta as informações do modelo construído para explicar a recuperação metalúrgica de zinco no concentrado de chumbo. Percebe-se que o modelo possui capacidade preditiva muito ruim, visto que somente 2.5% da variabilidade total da variável dependente pode ser explicada através de sua utilização. Analisando o valor-p da estatística de teste F e do teste t individual, percebe-se valores superiores a 0.05, indicando que o teor de chumbo no ROM é não uma variável estatisticamente válida para explicar a recuperação metalúrgica de zinco no concentrado de chumbo. Como nenhuma outra variável apresentou valores significativos destas estatísticas, conclui-se que, com os dados existentes, não é possível elaborar um modelo válido para a previsão da recuperação metalúrgica de zinco no concentrado de chumbo através da técnica de análise de regressão. De toda forma, o modelo capaz de indicar a melhor relação entre as variáveis será apresentado na equação 204, ainda que este não seja estatisticamente significativo.

$$\text{Rec Met de Zn no Conc Pb} = 0.0088 + 0.0054 * \text{Pb}. \quad (204)$$

Analisando as Figuras 34 e 35, percebe-se a existência de resíduos homocedásticos porém não normalmente distribuídos, violando mais uma suposição para construção de um modelo estatisticamente válido.

OLS Regression Results						
Dep. Variable: Rec. Met. de Zn no Conc. de Pb		R-squared: 0.025				
Model: OLS		Adj. R-squared: 0.016				
Method: Least Squares		F-statistic: 2.863				
Date: Thu, 28 May 2020		Prob (F-statistic): 0.0934				
Time: 13:48:29		Log-Likelihood: 439.33				
No. Observations: 115		AIC: -874.7				
Df Residuals: 113		BIC: -869.2				
Df Model: 1						
Covariance Type: nonrobust						
	coef	std err	t	P> t	[0.025	0.975]
const	0.0088	0.001	10.904	0	0.007	0.01
Pb	0.0054	0.003	1.692	0.093	-0.001	0.012
Omnibus: 56.759		Durbin-Watson: 1.967				
Prob(Omnibus): 0		Jarque-Bera (JB): 194.847				
Skew: 1.776		Prob(JB): 4.89E-43				
Kurtosis: 8.296		Cond. No. 6.67				

Figura 32 - Dados do modelo de regressão construído para Rec. Met. de Zn no Conc. de Pb

Os dados do modelo de regressão elaborado para prever a recuperação metalúrgica de zinco no rejeito final podem ser vistos na Figura 33. Analisando os resultados percebe-se que 75.8% da variabilidade total da variável dependente pode ser explicada através do modelo construído, o qual encontra-se equacionado em 205.

$$\text{Rec Met Zn no Rej} = 0.4007 - 0.5122 * \text{Zn} + 0.147 * \text{D1} + 0.0667 * \text{D2}. \quad (205)$$

Percebe-se, que quando a amostra pertence ao alvo Extremo Norte sua recuperação metalúrgica de zinco no rejeito final aumenta em 14.7% em relação à uma amostra coletada no alvo Sucuri que contenha o mesmo teor de zinco no ROM. Para uma amostra coletada em Lumiadeira, este aumento em relação à Sucuri é de 6.67% para um mesmo teor de zinco na alimentação.

A análise de resíduos mostra resíduos normais, mas não homocedásticos.

OLS Regression Results						
Dep. Variable: Rec. Met. de Zn no Rej. Final		R-squared: 0.758				
Model: OLS		Adj. R-squared: 0.751				
Method: Least Squares		F-statistic: 115.7				
Date: Thu, 28 May 2020		Prob (F-statistic): 5.08E-34				
Time: 13:48:47		Log-Likelihood: 129.23				
No. Observations: 115		AIC: -250.5				
Df Residuals: 111		BIC: -239.5				
Df Model: 3						
Covariance Type: nonrobust						
	coef	std err	t	P> t	[0.025	0.975]
const	0.4007	0.017	23.376	0	0.367	0.435
Zn	-0.5122	0.031	-16.792	0	-0.573	-0.452
D1	0.147	0.02	7.382	0	0.108	0.186
D2	0.0667	0.017	3.895	0	0.033	0.101
Omnibus: 6.543		Durbin-Watson: 2.087				
Prob(Omnibus): 0.038		Jarque-Bera (JB): 6.19				
Skew: 0.557		Prob(JB): 0.0453				
Kurtosis: 3.221		Cond. No. 5.28				

Figura 33 - Dados do modelo de regressão construído para Rec. Met. de Zn no Rej. Final

Histograma dos resíduos - Valor Crítico AD (5%): 0.762

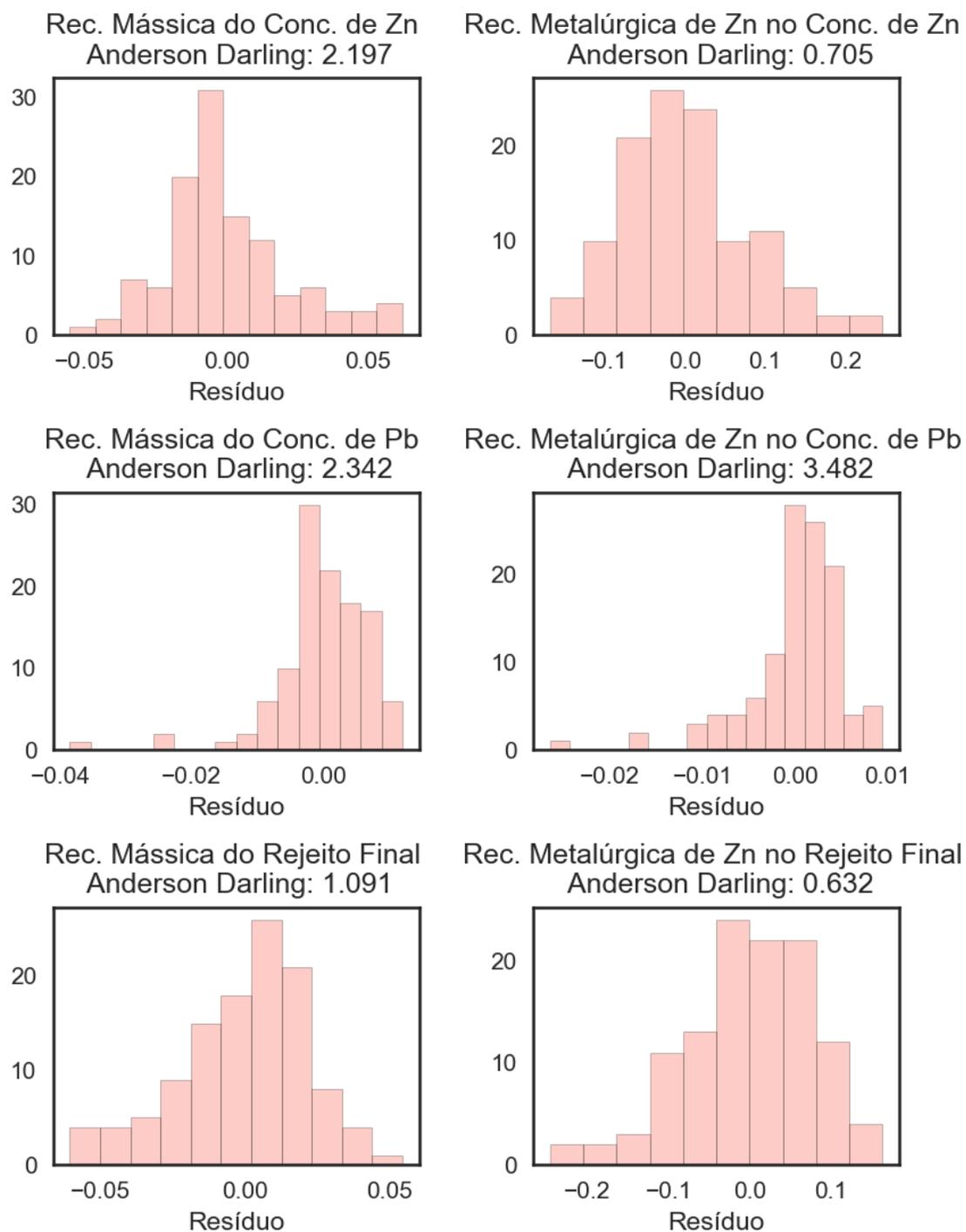


Figura 34 - Histograma dos resíduos – Mina de Zinco

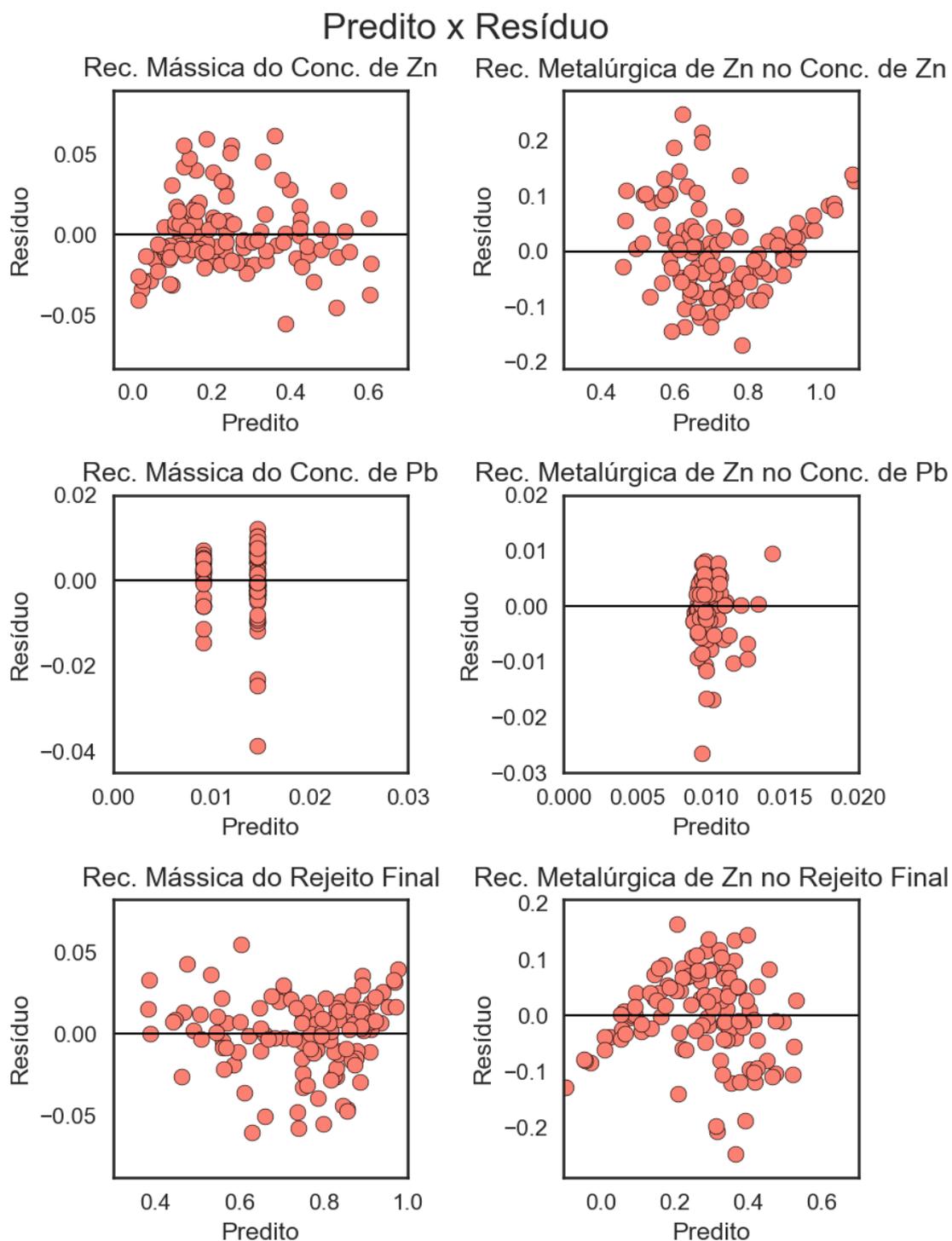


Figura 35 - Gráficos dos valores preditos x resíduos – Mina de Zinco

A Figura 36 ilustra os valores preditos versus os valores reais de recuperação mássica e metalúrgica, enquanto a Tabela 6 apresenta a correlação entre esses valores. Percebe-se uma distribuição dos pontos bem apertada ao redor da reta de 45 graus para a

recuperação mássica de zinco e recuperação mássica do rejeito final. Já para a recuperação mássica de chumbo e para recuperação metalúrgica de zinco no concentrado de chumbo, fica evidente a baixa qualidade preditiva dos modelos elaborados. A recuperação metalúrgica de zinco no concentrado de zinco apresenta um comportamento curvado para valores superiores a 0.8 desta variável, enquanto para a recuperação metalúrgica de zinco no rejeito final o mesmo comportamento curvado é percebido na cauda inferior, para valores menores que 0.2. Este comportamento é consequência da utilização de uma técnica de estimativa linear para predizer variáveis não aditivas, gerando resultados inesperados.

Tendo em vista que todos os modelos elaborados violaram pelo menos uma das suposições necessárias para a elaboração de um modelo de regressão estatisticamente válido (resíduos normalmente distribuídos e homocedásticos), optou-se por encerrar neste momento a análise dos dados da mina de Vazante utilizando a técnica de análise de regressão. Os testes t e F podem ter seus valores invalidados devido a violação de tais suposições, o que poderia gerar problemas na extrapolação dos modelos.

Tabela 6 - Correlações entre valores preditos x observados - Análise de regressão – Mina de Zinco

Variáveis	Correlação
Rec. Mássica no Conc. de Zn	0.989
Rec. Mássica no Conc. de Pb	0.345
Rec. Mássica no Rej. Final	0.988
Rec. Met. de Zn no Conc de Zn	0.866
Rec. Met. de Zn no Conc de Pb	0.157
Rec. Met. de Zn no Rej. Final	0.870

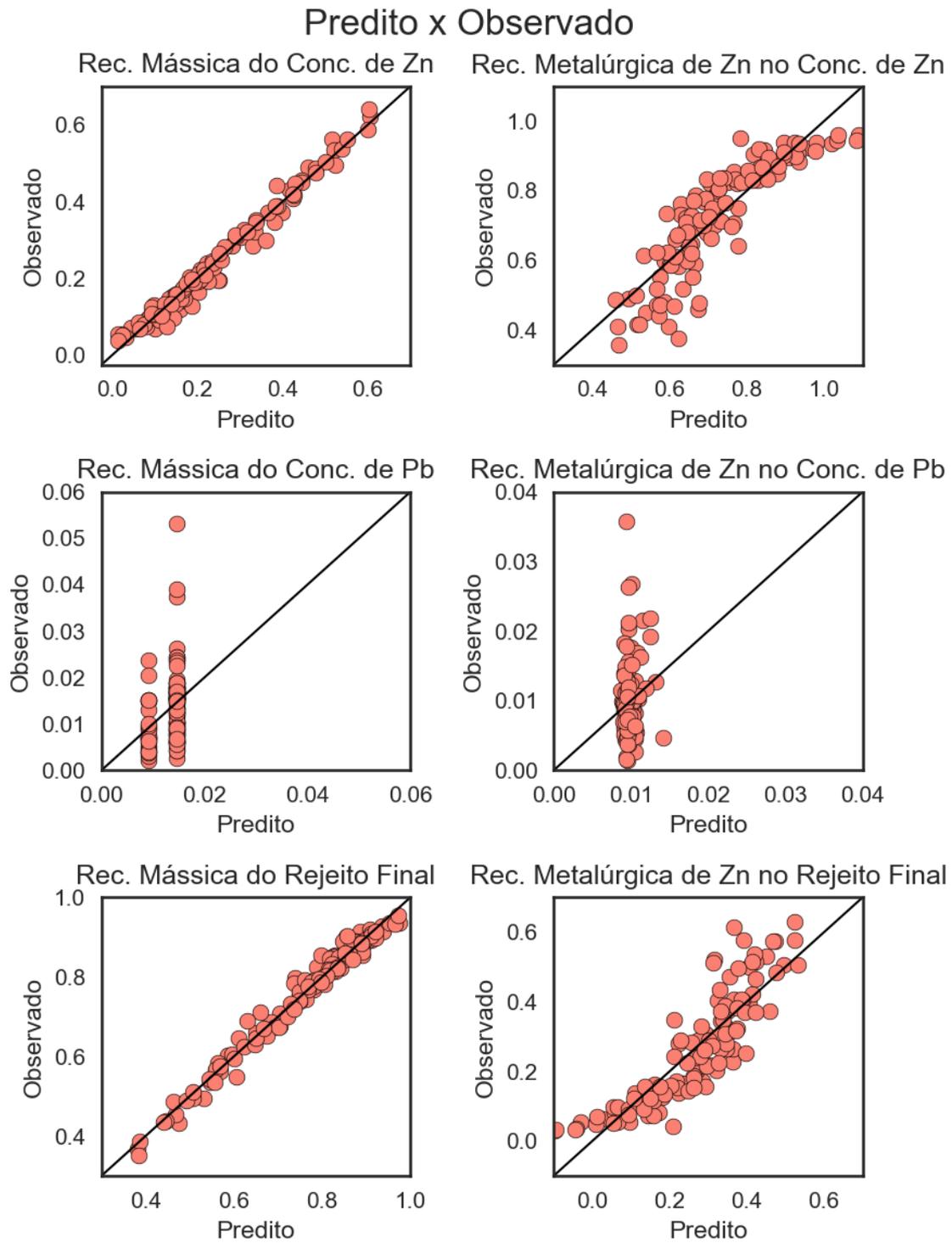


Figura 36 - Gráfico dos valores preditos x observados – Mina de Zinco

6.1.3.2. *Random forest*

Visando realizar a predição das variáveis dependentes em estudo, decidiu-se utilizar a técnica de *random forest* para regressão, considerando como variáveis independentes o teor de zinco e chumbo transformados e as variáveis D1 e D2. Tendo em vista o pequeno número de variáveis em estudo, decidiu-se adotar como quatro o número máximo de variáveis utilizadas na elaboração das árvores. Foram consideradas 1000 árvores para a elaboração do estudo e foi fixado o valor da semente em 8520 para que os resultados pudessem ser reproduzidos.

Para escolher o número mínimo de amostras por folha, foram testados valores entre 2 e 15. O valor do coeficiente de correlação obtido através do out-of-bag (OOB) foi calculado para cada teste de amostras por folha e escolheu-se como o valor final aquele que apresentou maior coeficiente de correlação. A Tabela 7 apresenta o número mínimo de amostras por folha considerado para cada uma das variáveis dependentes em estudo.

Tabela 7 - Número mínimo de amostras por folha - Mina de Zinco

Variáveis	Número mínimo de amostras por folha
Rec. Mássica no Conc. de Zn	2
Rec. Mássica no Conc. de Pb	6
Rec. Mássica no Rej. Final	2
Rec. Met. de Zn no Conc de Zn	3
Rec. Met. de Zn no Conc de Pb	6
Rec. Met. de Zn no Rej. Final	3

A Tabela 8 apresenta a importância das variáveis independentes nas árvores geradas através da técnica de *random forest*, sendo que quanto mais próximo de um o valor da importância, maior é a relevância da variável para predizer a respectiva variável dependente. Analisando os resultados, vê-se que o teor de zinco no ROM é responsável por mais de 80% da explicação das variáveis recuperação mássica de zinco, recuperação mássica do rejeito final, recuperação metalúrgica de zinco no concentrado de zinco e recuperação metalúrgica de zinco no rejeito final. O teor de chumbo na alimentação é a variável que mais influencia na predição da recuperação metalúrgica de zinco no

concentrado de chumbo. A recuperação mássica do concentrado de chumbo, por sua vez, recebe grande influência dos teores de zinco e chumbo e da variável D1.

Tabela 8 - Importância das variáveis independentes nas árvores elaboradas – Mina de Zinco

Variáveis	Zn	D1	D2	Pb
Rec. Mássica no Conc. de Zn	0.95	0.01	0.00	0.04
Rec. Mássica no Conc. de Pb	0.32	0.30	0.05	0.33
Rec. Mássica no Rej. Final	0.95	0.01	0.00	0.03
Rec. Met. de Zn no Conc de Zn	0.82	0.07	0.03	0.09
Rec. Met. de Zn no Conc de Pb	0.37	0.02	0.05	0.56
Rec. Met. de Zn no Rej. Final	0.81	0.07	0.03	0.09

A Figura 37 ilustra os valores preditos através da técnica de *random forest* versus os valores observados das variáveis dependentes em estudo. Os coeficientes de correlação entre os valores reais e as estimativas encontram-se apresentados na Tabela 9.

Percebe-se um comportamento mais uniforme dos dados ao redor da reta identidade quando as variáveis recuperação metalúrgica de zinco no concentrado de zinco e recuperação metalúrgica de zinco no rejeito final são analisadas, ao contrário do comportamento curvado que se observou quando a técnica de análise de regressão foi utilizada. Além disso, a boa correlação entre os valores preditos e os observados da recuperação mássica de zinco e da recuperação mássica do rejeito final foi mantida quando a técnica de *random forest* foi aplicada. As predições realizadas para a recuperação mássica do concentrado de chumbo e para a recuperação metalúrgica de zinco no concentrado de chumbo estão melhor distribuídas ao redor da reta identidade quando *random forest* foi utilizada, em comparação com os resultados obtidos através de análise de regressão.

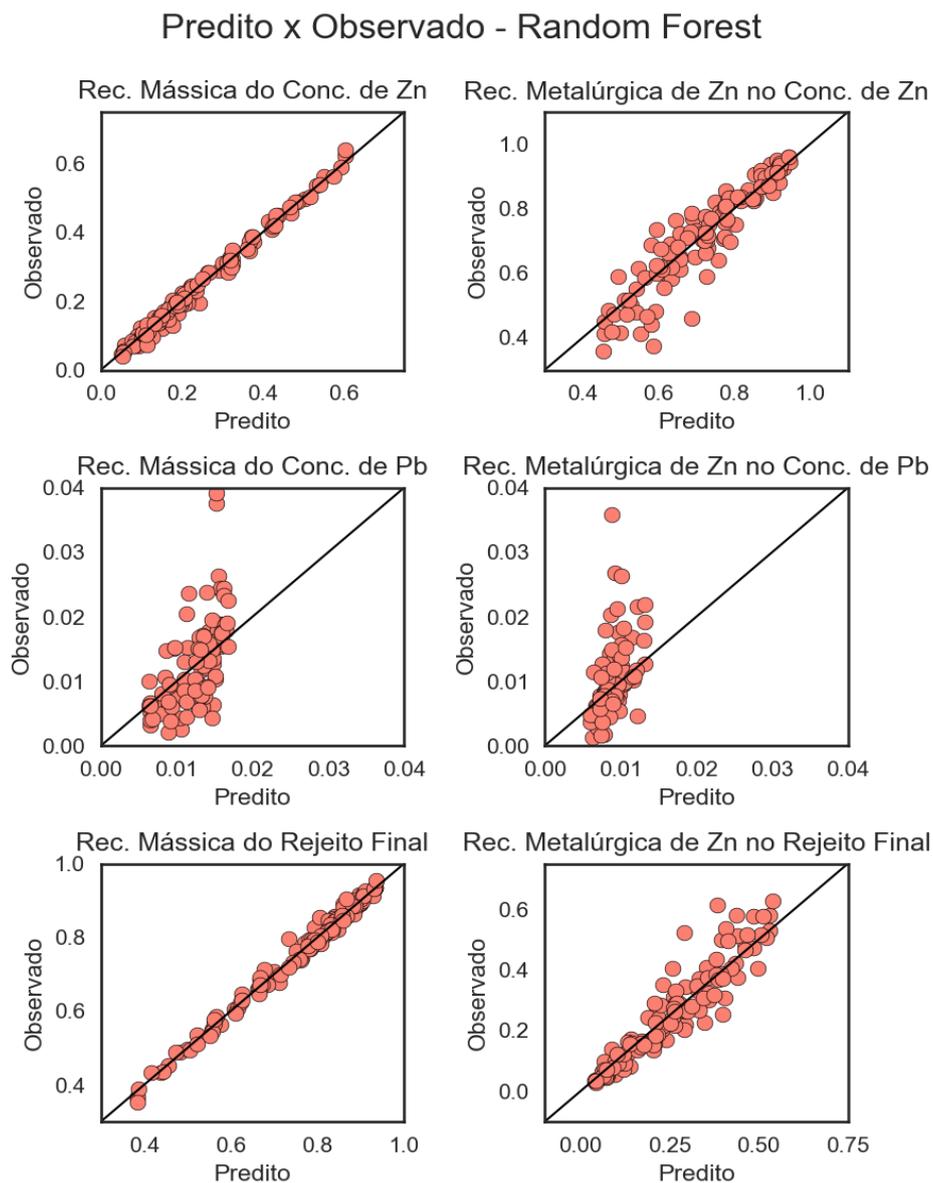


Figura 37 - Valores preditos x observados - Random Forest – Mina de Zinco

Tabela 9 - Correlações entre valores preditos x observados - Random Forest – Mina de Zinco

Variáveis	Correlação
Rec. Mássica no Conc. de Zn	0.995
Rec. Mássica no Conc. de Pb	0.587
Rec. Mássica no Rej. Final	0.994
Rec. Met. de Zn no Conc de Zn	0.932
Rec. Met. de Zn no Conc de Pb	0.516
Rec. Met. de Zn no Rej. Final	0.932

6.1.3.3. Support vector regression

A técnica de *support vector regression* (SVR) foi aplicada aos dados considerando as mesmas variáveis independentes utilizadas nas seções anteriores. O estudo de SVR foi dividido em duas partes: na primeira foi considerada a utilização de um *support vector regression* com função linear e na segunda foram testadas as três funções não-lineares existentes visando identificar qual delas apresentava as melhores respostas. Em ambas partes, foi adotado um valor de épsilon (ϵ) igual a 0.0001.

6.1.3.3.1. Support vector regression com função linear

O primeiro passo para a utilização de um SVR linear é a definição do valor do parâmetro C. Como já foi discutido na seção 3.4.2, a escolha deste parâmetro deve ser feita analisando diversos resultados e verificando a presença ou não de *overfitting* nos mesmos. Visando avaliar a performance dos valores de C em um banco de dados não utilizado na construção do SVR, o banco de dados foi dividido aleatoriamente em um banco de teste contendo 20% dos dados e um banco de treinamento, com os outros 80%. Considerando este banco, foram testados valores de C variando entre 0.001 e 100. Para cada variável dependente, foi escolhido o valor de C que retornou o menor erro absoluto médio (MAE) no banco de teste, garantindo assim que não exista *overfitting* e que os resultados sejam adequados. A Tabela 10 apresenta os valores do parâmetro C adotados neste estudo.

Tabela 10 - Valores de C escolhidos para construção do SVR linear – Mina de Zinco

Variáveis	C
Rec. Mássica no Conc. de Zn	35
Rec. Mássica no Conc. de Pb	96
Rec. Mássica no Rej. Final	15
Rec. Met. de Zn no Conc de Zn	52
Rec. Met. de Zn no Conc de Pb	0.2
Rec. Met. de Zn no Rej. Final	39

Considerando os valores de C acima, os SVR foram calculados considerando todos os 115 dados disponíveis. Os valores preditos através desta técnica foram plotados contra os valores reais e a correlação entre estes valores foi calculada, conforme ilustrado na Figura 38 e apresentado na Tabela 11, respectivamente.

Analisando o gráfico de valores preditos versus observados, vê-se na cauda superior da recuperação metalúrgica de zinco no concentrado de zinco e na cauda inferior da recuperação metalúrgica de zinco no rejeito final o mesmo comportamento curvado observado quando a análise de regressão foi utilizada. Tal comportamento é consequência da utilização de uma técnica de estimativa linear para prever variáveis não aditivas, gerando resultados insatisfatórios. De maneira geral, os resultados via SVR linear se assemelham muito aos obtidos através de análise de regressão.

Tabela 11 - Correlações entre valores preditos x observados - SVR Linear – Mina de Zinco

Variáveis	Correlação
Rec. Mássica no Conc. de Zn	0.989
Rec. Mássica no Conc. de Pb	0.350
Rec. Mássica no Rej. Final	0.988
Rec. Met. de Zn no Conc de Zn	0.861
Rec. Met. de Zn no Conc de Pb	0.197
Rec. Met. de Zn no Rej. Final	0.864

Predito x Observado - Support Vector Regression Linear

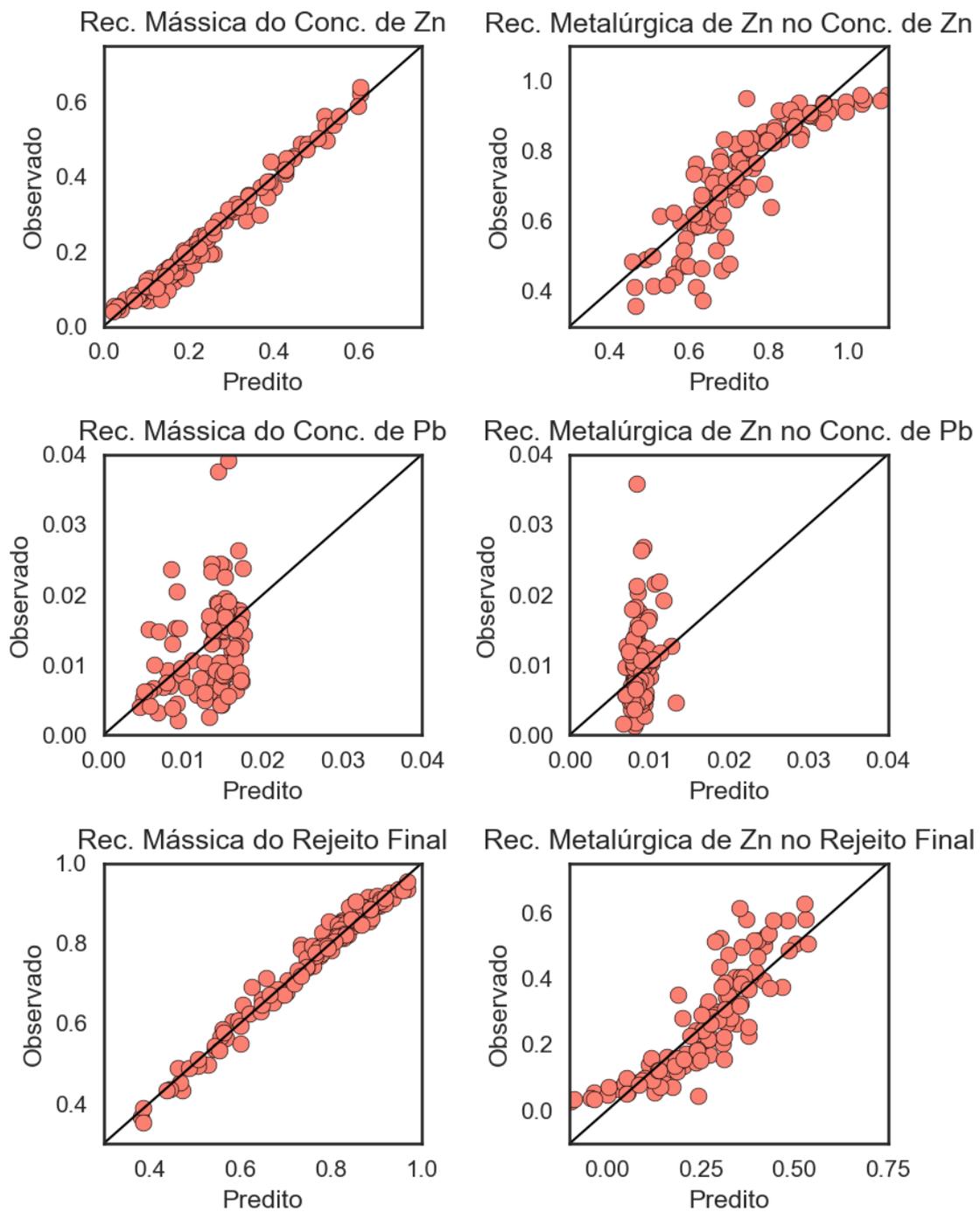


Figura 38 - Predito x Observado - *Support vector regression* linear – Mina de Zinco

6.1.3.3.2. Support vector regression não linear

O primeiro passo para ajustar um SVR não linear é definir, além do parâmetro C, qual tipo de kernel será utilizado e o valor do parâmetro de escala (σ), também conhecido como parâmetro *gamma*. Da mesma forma como foi feito no SVR linear, foi utilizado o banco de dados dividido entre banco de teste e de treinamento para determinar os melhores valores dos parâmetros em questão. Neste estudo, foram avaliados os valores obtidos utilizando um kernel *radial basis function* (RBF), uma vez que este apresentou melhores resultados que os demais. Foram avaliados os resultados obtidos com valor de *gamma* variando entre 0.001 e 100 e valores de C variando entre 0.001 e 100. A escolha da melhor combinação de parâmetros foi feita através da análise dos valores de MAE obtidos no banco de teste, sendo que os parâmetros que retornaram o menor valor de MAE para cada variável foram os selecionados.

A Tabela 12 apresenta os valores de C e *gamma* escolhidos para um kernel RBF, o qual apresentou os melhores resultados em todos os testes realizados.

Tabela 12 - Valores dos parâmetros C e *gamma* para o SVR não linear – Mina de Zinco

Variáveis	C	Gamma
Rec. Mássica no Conc. de Zn	74	0.08
Rec. Mássica no Conc. de Pb	13	2
Rec. Mássica no Rej. Final	96	0.07
Rec. Met. de Zn no Conc de Zn	76	0.05
Rec. Met. de Zn no Conc de Pb	19	0.006
Rec. Met. de Zn no Rej. Final	100	0.04

Utilizando os valores da tabela acima, o SVR não linear foi calculado para todos os 115 dados disponíveis. O gráfico dos valores preditos versus os valores reais encontra-se ilustrado na Figura 39 e as correlações entre esses valores está apresentada na Tabela 13.

Tabela 13 - Correlações entre valores preditos x observados - SVR não linear – Mina de Zinco

Variáveis	Correlação
Rec. Mássica no Conc. de Zn	0.992
Rec. Mássica no Conc. de Pb	0.567
Rec. Mássica no Rej. Final	0.991
Rec. Met. de Zn no Conc de Zn	0.894
Rec. Met. de Zn no Conc de Pb	0.320
Rec. Met. de Zn no Rej. Final	0.895

Predito x Observado - Support Vector Regression Não Linear

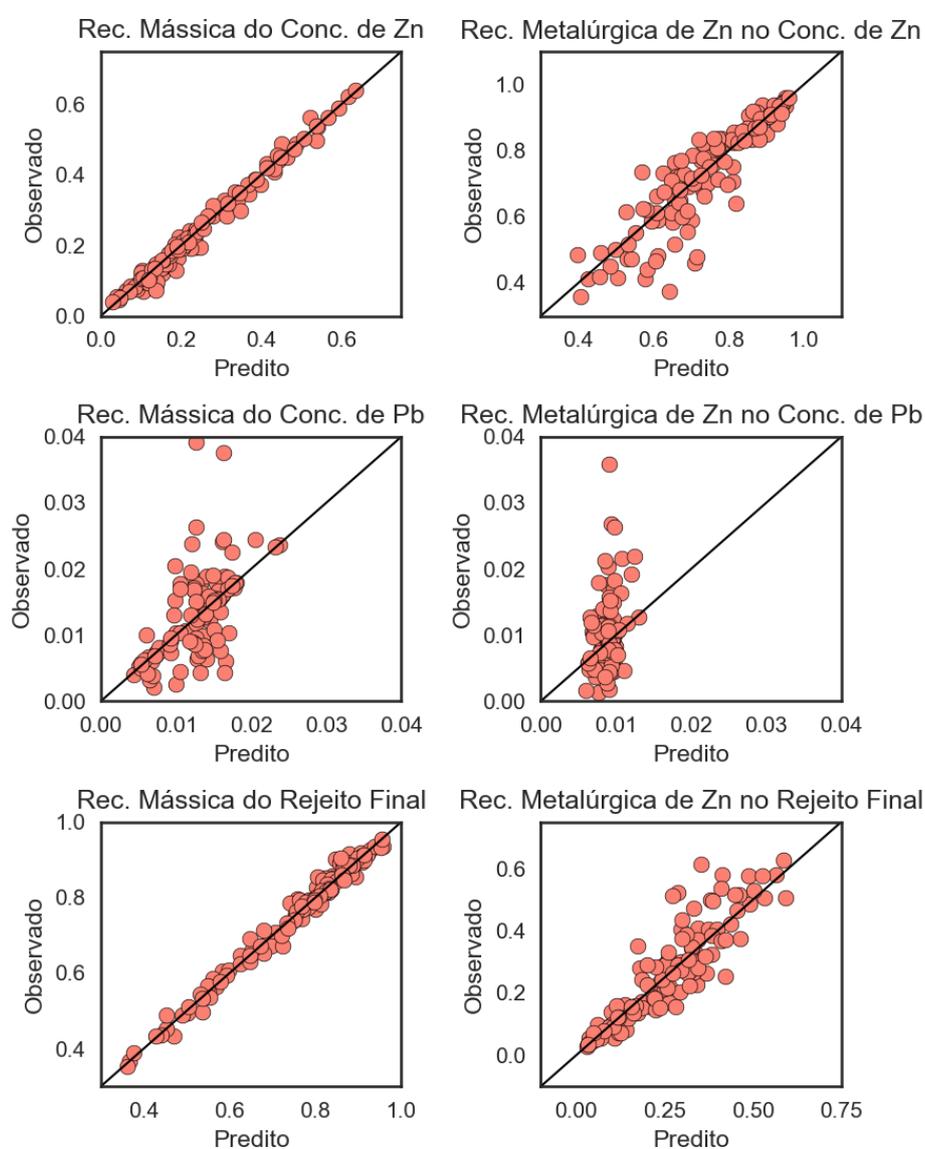


Figura 39 - Predito x Observado - Support vector regression não linear – Mina de Zinco

Analisando os resultados, percebe-se que as estimativas relacionadas à cauda superior da variável recuperação metalúrgica de zinco no concentrado de zinco e as previsões relacionadas à cauda inferior da recuperação metalúrgica de zinco no rejeito estão muito mais aderentes à reta de 45 graus, quando os mesmos valores são comparados com os preditos utilizando as outras técnicas até aqui apresentadas.

6.1.3.4. Redes Neurais Artificiais

A construção da rede neural utilizada para prever as variáveis geometalúrgicas da mina de Vazante começou pela definição de qual transformação aplicar aos dados de input, chegando à conclusão que a padronização dos dados para média zero e variância unitária era a opção que gerava os melhores resultados. Em seguida, seguiu-se com a determinação do número de camadas ocultas que a rede teria. Sabe-se, que devido ao número amostral não ser muito elevado, uma rede com mais de três camadas ocultas poderia causar *overfitting*, logo, limitou-se a testar redes com no máximo três camadas ocultas. Entre os diversos testes realizados, concluiu-se que utilizar três camadas ocultas gerava melhores resultados: menor MAE e maiores correlações em todas as variáveis de *output* no banco de teste. Logo, esta configuração de rede foi adotada para o modelo geometalúrgico a ser usado na mina de Vazante.

Em seguida, foram testadas diversas combinações de nós nas camadas ocultas, sendo que valores entre 2 e 7 foram avaliados em todas as camadas. Os resultados mostram uma melhor performance quando 5 nós são utilizados na primeira camada oculta, 4 nós utilizados na segunda e 5 nós utilizados na terceira camada.

Para determinar qual função de ativação deveria ser utilizada nas camadas ocultas e na camada de *output*, foram testadas as funções linear, sigmoide, tangente hiperbólica, relu e softplus, sendo que os melhores resultados foram obtidos quando a função linear foi utilizada na primeira e na segunda camada oculta, a softplus foi utilizada na terceira e a função de ativação linear foi utilizada na camada de *output*.

Para determinar o valor do parâmetro de aprendizado η para os dados em questão, foram testados valores entre 0.00001 e 0.5, sendo que os melhores resultados foram obtidos quando $\eta = 0.03$ foi utilizado.

Sendo assim, a rede neural utilizada para prever as seis variáveis dependentes em estudo tem a forma ilustrada na Figura 40.

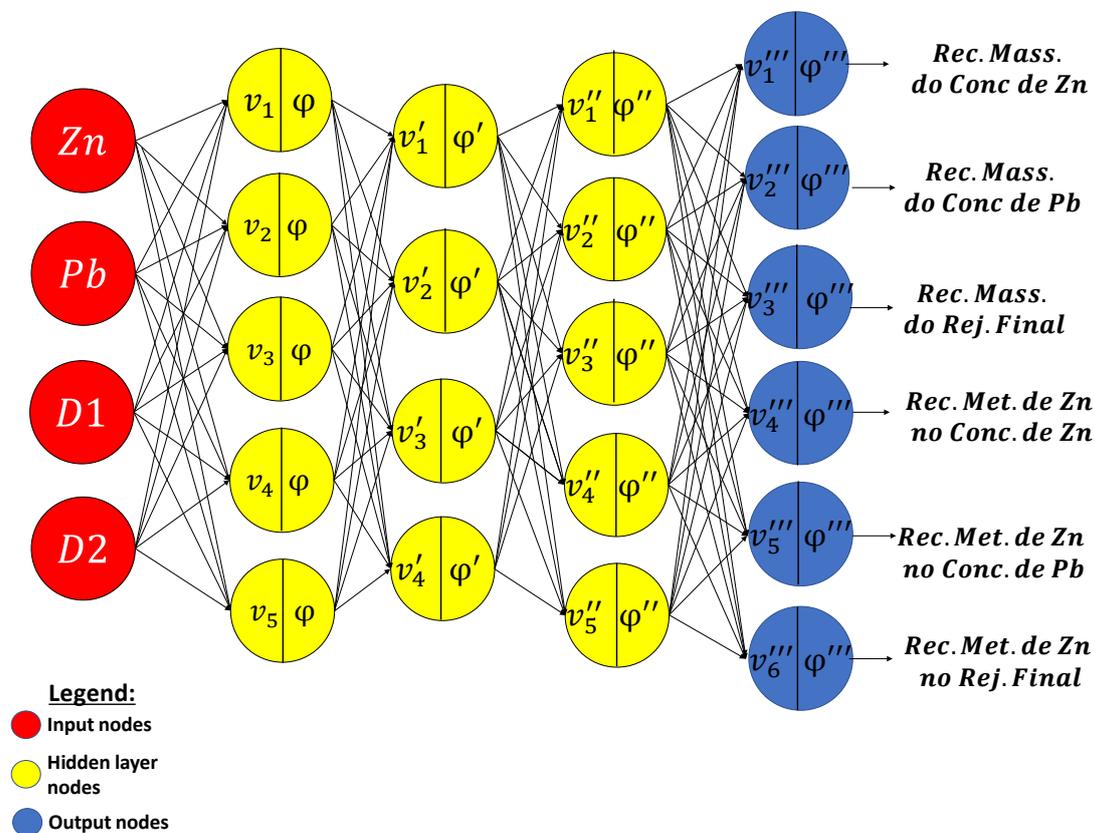


Figura 40 - Rede Neural utilizada para prever as variáveis geometalúrgicas da mina de Vazante

A Figura 41 ilustra os valores preditos pela rede em questão versus os valores reais das seis variáveis dependentes. Na Tabela 14, encontra-se a correlação entre estes valores.

Tabela 14 - Correlações entre valores preditos x observados – Redes Neurais – Mina de Zinco

Variáveis	Correlação
Rec. Mássica no Conc. de Zn	0.990
Rec. Mássica no Conc. de Pb	0.270
Rec. Mássica no Rej. Final	0.990
Rec. Met. de Zn no Conc de Zn	0.895
Rec. Met. de Zn no Conc de Pb	0.009
Rec. Met. de Zn no Rej. Final	0.898

Predito x Observado - Redes Neurais

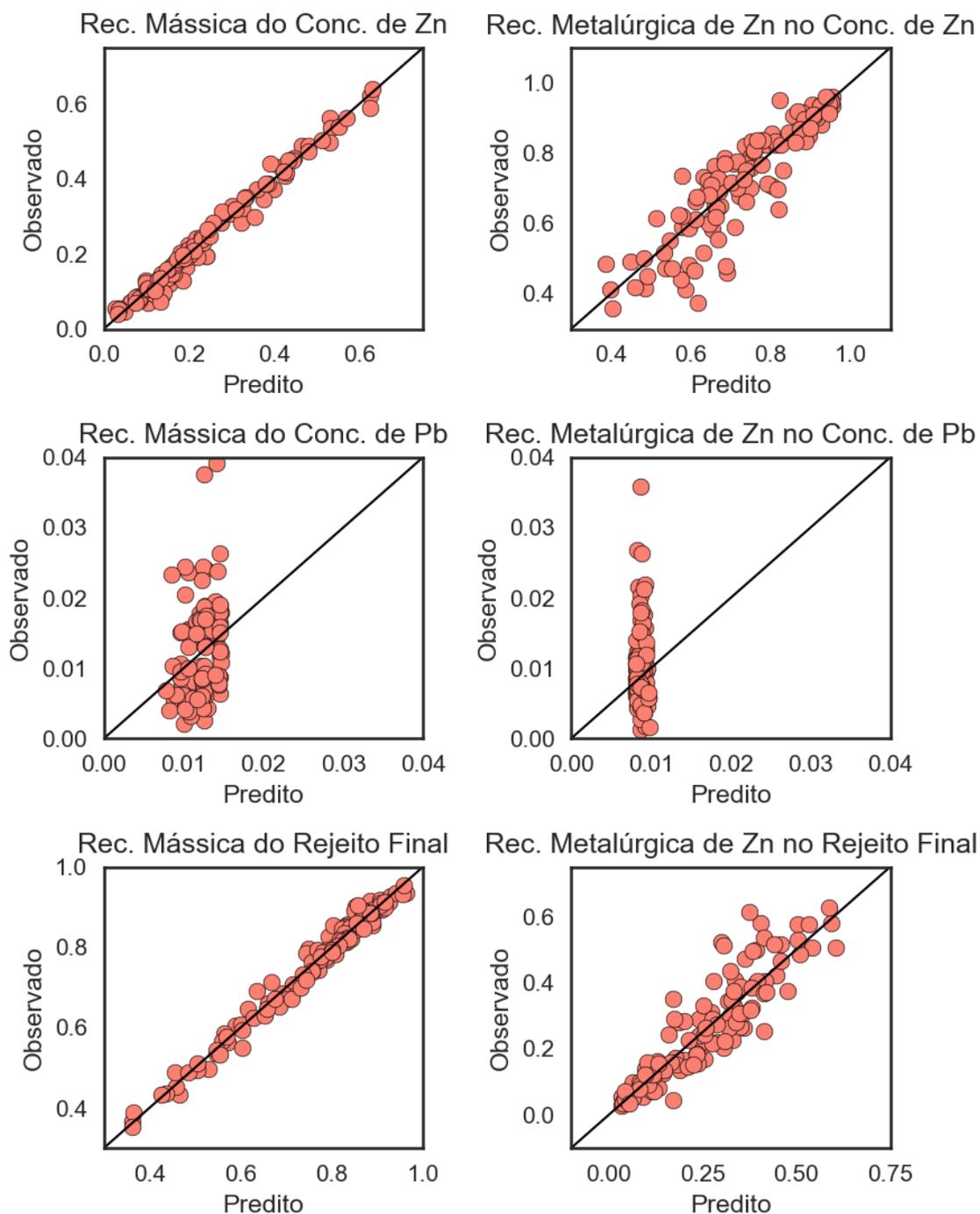


Figura 41 - Predito x Observado - Redes Neurais Artificiais – Mina de Zinco

Analisando os resultados vê-se que os dados de recuperação metalúrgica de zinco no concentrado de zinco e de recuperação metalúrgica de zinco no rejeito final se distribuem uniformemente em torno da reta de 45 graus, sem a geração de caudas como

as observadas quando as técnicas de análise de regressão e de SVR Linear foram utilizadas.

6.1.4. Validação cruzada k-fold

A validação cruzada é o método mais simples e mais amplamente utilizado para estimar o erro de predição (Hastie, Tibshirani e Friedman, 2009). A ideia da validação cruzada *k-fold* é dividir o banco de dados disponível 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. Suponha que $k = 5$, logo, a configuração da validação cruzada *5-fold* assume a forma apresentada na Figura 42. Na primeira rodada, as quatro primeiras partes são utilizadas como banco de treinamento para elaborar o modelo e a quinta parte é utilizada para avaliar os erros relacionados à utilização do modelo construído. Na segunda rodada, a quinta parte passa a ser considerada como banco de treinamento e a quarta como banco de teste. O algoritmo segue até que todas as partes tenham sido utilizadas como banco de treinamento e de teste em algum momento.

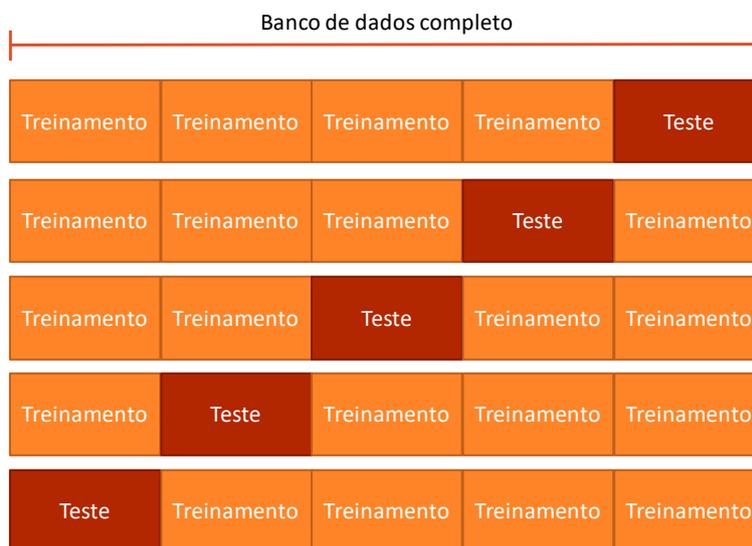


Figura 42 - Exemplo de funcionamento da técnica de validação cruzada com $k = 5$

A estimativa do erro de predição da validação cruzada é dada pela fórmula 206, apresentada em Hastie, Tibshirani e Friedman (2009):

$$CV(\hat{f}) = \frac{1}{N} \sum_{i=1}^N L(y_i, \hat{f}^{-k(i)}(x_i)), \quad (206)$$

onde: $\hat{f}^{-k(i)}$ é a função ajustada sem considerar o i -ésimo dado contido na k -ésima parte utilizada como banco de teste, L é a função de perda, que geralmente é o erro absoluto ou o erro quadrático entre o valor predito e o valor real, y_i é o valor real observado no banco de dados e N é o número de dados em estudo. Desta forma, o erro de predição pode ser resumido através do erro médio absoluto (MAE) e do erro quadrático médio (MSE) observados nos bancos de teste.

Além de monitorar o erro de predição, é possível calcular o erro padrão do erro de predição, o que possibilita a construção de um intervalo de confiança para este valor, levando a uma interpretação mais ampla desta estatística.

Visando determinar o erro de predição dos modelos ajustados utilizando todas as técnicas apresentadas até então, foi aplicada a técnica de validação cruzada *k-fold* assumindo $k = 5$. Optou-se por apresentar os resultados da raiz quadrada do erro quadrático médio (RMSE) ao invés de apresentar os valores de MSE devido aos pequenos valores relacionados à última métrica citada, a qual tornava a visualização dos resultados menos imediata por estar em notação científica. Além disso, foi avaliado o erro absoluto médio e optou-se por avaliar também a correlação entre os valores preditos e os valores reais do banco de dados. A Tabela 15 apresenta o valor médio das três métricas supracitadas acompanhado do erro padrão. As células coloridas de laranja representam o melhor resultado obtido para a variável em estudo.

Analisando os resultados é possível perceber que, com exceção das variáveis recuperação mássica no concentrado de chumbo e recuperação metalúrgica de zinco no concentrado de chumbo, a técnica de redes neurais quase sempre apresentou os melhores resultados de RMSE e correlação. Logo, as variáveis responsáveis pelo destino de 98.6% da massa e de 99% do metal, em média, apresentaram menor erro de predição quando a técnica de redes neurais artificiais foi utilizada, indicando que esta é a mais aconselhada

para ser utilizada nas previsões geometalúrgicas do minério destinado ao beneficiamento na mina de Vazante.

Tabela 15 - Avaliação dos resultados da validação cruzada 5-fold – Mina de Zinco

Banco de teste					
Métrica	Variáveis	RF	SVR Linear	SVR Não-Linear	Redes Neurais
MAE	Rec. Mássica no Conc. de Zn	0.0218 ± 0.0023	0.0175 ± 0.0015	0.0175 ± 0.002	0.0174 ± 0.0009
	Rec. Mássica no Conc. de Pb	0.005 ± 0.0005	0.0049 ± 0.0005	0.0062 ± 0.0006	0.0053 ± 0.0005
	Rec. Mássica no Rej. Final	0.0234 ± 0.002	0.0186 ± 0.0014	0.0184 ± 0.0021	0.0179 ± 0.0011
	Rec. Met. de Zn no Conc de Zn	0.0694 ± 0.0073	0.0648 ± 0.0018	0.054 ± 0.0026	0.0552 ± 0.0027
	Rec. Met. de Zn no Conc de Pb	0.0036 ± 0.0004	0.0037 ± 0.0003	0.0036 ± 0.0004	0.0039 ± 0.0004
	Rec. Met. de Zn no Rej. Final	0.0681 ± 0.0069	0.0633 ± 0.0018	0.0529 ± 0.003	0.0539 ± 0.003
RMSE	Rec. Mássica no Conc. de Zn	0.0279 ± 0.003	0.0237 ± 0.002	0.0261 ± 0.0044	0.0234 ± 0.0013
	Rec. Mássica no Conc. de Pb	0.0068 ± 0.001	0.0068 ± 0.001	0.0112 ± 0.0027	0.0072 ± 0.001
	Rec. Mássica no Rej. Final	0.0297 ± 0.0027	0.0242 ± 0.0015	0.0268 ± 0.0047	0.0234 ± 0.0014
	Rec. Met. de Zn no Conc de Zn	0.0922 ± 0.0105	0.0828 ± 0.0031	0.0763 ± 0.0033	0.0758 ± 0.0036
	Rec. Met. de Zn no Conc de Pb	0.0052 ± 0.0008	0.0052 ± 0.0007	0.0052 ± 0.0007	0.0054 ± 0.0006
	Rec. Met. de Zn no Rej. Final	0.0903 ± 0.0104	0.0815 ± 0.0028	0.0745 ± 0.0037	0.0744 ± 0.0037
Correlação	Rec. Mássica no Conc. de Zn	0.9302 ± 0.0028	0.9885 ± 0.0018	0.9831 ± 0.0067	0.988 ± 0.0017
	Rec. Mássica no Conc. de Pb	0.5205 ± 0.0297	0.3735 ± 0.0211	0.2445 ± 0.1278	0.2256 ± 0.0749
	Rec. Mássica no Rej. Final	0.9304 ± 0.0029	0.9874 ± 0.0019	0.9813 ± 0.0073	0.9878 ± 0.002
	Rec. Met. de Zn no Conc de Zn	0.8249 ± 0.0454	0.8594 ± 0.0123	0.8794 ± 0.0162	0.8851 ± 0.0135
	Rec. Met. de Zn no Conc de Pb	0.2192 ± 0.1154	0.1391 ± 0.0819	0.2601 ± 0.0585	-0.0758 ± 0.0376
	Rec. Met. de Zn no Rej. Final	0.831 ± 0.0425	0.8602 ± 0.0132	0.8845 ± 0.0168	0.8892 ± 0.0138

Ao interpretar os resultados obtidos via redes neurais artificiais vê-se que correlações superiores à 88% foram observadas no banco de teste quando as variáveis recuperação mássica no concentrado de zinco, recuperação mássica no rejeito final, recuperação metalúrgica de zinco no concentrado de zinco e recuperação metalúrgica de zinco no rejeito final foram avaliadas, indicando resultados muito promissores para tais variáveis. A correlação observada para a recuperação mássica no concentrado de chumbo foi baixa (0.2256) e a correlação da recuperação metalúrgica de zinco no concentrado de chumbo foi quase nula (-0.075), conforme já era esperado com base nos resultados observados considerando todas as amostras. Os resultados insatisfatórios observados para tais variáveis não são preocupantes uma vez que elas representam em média 1.4% da massa e 1% do destino do metal em estudo.

6.1.5. Análise do fechamento mássico e metalúrgico

Visando avaliar o comportamento dos modelos elaborados em relação ao fechamento mássico e metalúrgico obtido com cada um deles, foi avaliado o percentual de amostras que apresentaram erro de fechamento inferior a 1%, para mais ou para menos, quando os *5-folds* construídos para a validação cruzada foram utilizados. A Tabela 16 apresenta o percentual médio dos *5-folds* para cada uma das técnicas testadas. Analisando os resultados, percebe-se que a técnica de redes neurais artificiais se sobressai, indicando que 97% das amostras no banco de teste apresentam erros de fechamento das previsões inferiores a 1%, o que evita o pós-processamento das estimativas para fins de utilização dos resultados na prática.

Tabela 16 - Análise do fechamento mássico e metalúrgico – Mina de Zinco

Banco	Fechamento	RF	SVR Linear	SVR Não-Linear	Redes Neurais
Treinamento	Mássico	96%	100%	96%	100%
	Metalúrgico	96%	86%	98%	100%
Teste	Mássico	92%	100%	93%	97%
	Metalúrgico	94%	86%	95%	97%

6.1.6. Análise dos modelos em dados não utilizados nos bancos de treinamento e teste

Conforme explicado anteriormente, no início do estudo, os 140 dados disponíveis foram divididos de maneira que 115 fossem utilizados para treinar e testar os modelos elaborados e 25 amostras fossem separadas para uma validação final dos modelos, a fim de comprovar a eficiência destes na previsão de dados não utilizados para a calibração dos mesmos. Ao aplicar os modelos elaborados com os 115 dados iniciais nas 25 amostras restantes, os resultados apresentados na Tabela 17 foram observados.

Tabela 17 - Métricas observadas ao aplicar os modelos elaborados nas 25 amostras separadas para validação final

Métrica	Variáveis	RF	SVR Linear	SVR Não Linear	Redes Neurais
MAE	Rec. Mássica no Conc. de Zn	0.0212	0.0174	0.0163	0.0145
	Rec. Mássica no Conc. de Pb	0.0044	0.0046	0.0039	0.0050
	Rec. Mássica no Rej. Final	0.0225	0.0165	0.0164	0.0147
	Rec. Met. de Zn no Conc de Zn	0.0484	0.0562	0.0484	0.0455
	Rec. Met. de Zn no Conc de Pb	0.0035	0.0036	0.0034	0.0041
	Rec. Met. de Zn no Rej. Final	0.0480	0.0557	0.0497	0.0458
RMSE	Rec. Mássica no Conc. de Zn	0.0272	0.0212	0.0207	0.0172
	Rec. Mássica no Conc. de Pb	0.0064	0.0062	0.0056	0.0063
	Rec. Mássica no Rej. Final	0.0287	0.0199	0.0199	0.0173
	Rec. Met. de Zn no Conc de Zn	0.0614	0.0717	0.0673	0.0596
	Rec. Met. de Zn no Conc de Pb	0.0048	0.0051	0.0049	0.0052
	Rec. Met. de Zn no Rej. Final	0.0602	0.0716	0.0671	0.0585
Correlação	Rec. Mássica no Conc. de Zn	0.9806	0.9899	0.9903	0.9924
	Rec. Mássica no Conc. de Pb	0.3767	0.3863	0.6060	0.4238
	Rec. Mássica no Rej. Final	0.9771	0.9890	0.9887	0.9912
	Rec. Met. de Zn no Conc de Zn	0.9063	0.8729	0.8958	0.9154
	Rec. Met. de Zn no Conc de Pb	0.4451	0.3240	0.4357	0.0008
	Rec. Met. de Zn no Rej. Final	0.9089	0.8688	0.8913	0.9164

Conclui-se que a técnica de redes neurais artificiais apresentou os melhores valores de MAE, RMSE e correlação para as quatro variáveis mais importantes do estudo, confirmando mais uma vez a superioridade desta técnica em relação às demais e comprovando a ausência de *overfitting* no modelo.

6.1.7. Comparação dos resultados obtidos com o modelo de referência utilizado na mina de Vazante

Atualmente na mina de Vazante, utiliza-se de um modelo de regressão quadrático para prever a recuperação metalúrgica de zinco no concentrado de zinco, tendo como variável independente o teor de zinco no ROM. Este modelo foi concebido com uma base de dados diferente da adotada na elaboração dos modelos aqui apresentados. As demais variáveis geometalúrgicas não possuem um modelo próprio.

O modelo de regressão quadrática é uma ampliação do modelo de regressão múltipla, onde as interações entre as variáveis independentes e os quadrados dessas variáveis são incorporados no modelo visando uma predição mais acurada da variável

dependente em estudo. A fórmula 207 apresenta uma equação de regressão quadrática típica:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \dots + b_3x_1x_2 + \dots + b_kx_q^2, \quad (207)$$

onde b_0 é uma estimativa do valor real do intercepto da reta (β_0), b_i ($i = 1, \dots, k$) é uma estimativa do valor real da inclinação da reta (β_i) associada a cada variável x_j ($j = 1, \dots, q$), seu quadrado ou sua combinação com outra variável independente e \hat{y} é a estimativa do valor da variável dependente com base no modelo gerado.

As suposições e validações necessárias para um modelo de regressão quadrática são as mesmas feitas para um modelo de regressão múltipla, devendo ser observada com mais atenção a possibilidade de *overfitting* devido ao maior número de coeficientes ajustados em modelos quadráticos. Modelos desta natureza são frequentemente utilizados quando existem indícios de relação quadrática entre as variáveis estudadas ou quando a análise de resíduos dos modelos lineares aponta para uma relação não linear dos resíduos com a variável dependente.

Para tornar possível a comparação do modelo de redes neurais (que apresentou os melhores resultados globais) com o modelo utilizado atualmente na mina, optou-se por utilizar os mesmos 140 dados empregados na criação das redes neurais para ajustar um modelo quadrático utilizando o teor de zinco no ROM como variável independente, visto que uma diferença na base de dados poderia gerar resultados não comparáveis entre si.

Ao utilizar a mesma divisão da base de dados em dados de teste e dados de treinamento adotada na seção 6.1.4, obtêm-se os resultados de erro absoluto médio (MAE), raiz quadrada do erro quadrático médio (RMSE) e de correlação apresentados na Tabela 18.

Tabela 18 – Resultados da validação cruzada 5-fold para o modelo quadrático utilizado para prever a rec. metalúrgica de zinco no concentrado de zinco

Variável	MAE	RMSE	Correlação
Rec. Met. Zn no Conc. de Zinco	0.0676 ± 0.004	0.0877 ± 0.004	0.8356 ± 0.0188

Ao comparar tais resultados com os obtidos através da técnica de redes neurais, percebe-se a superioridade da última em relação a todas as métricas utilizadas, tendo em vista que a rede elaborada apresentou MAE de 0.0556, RMSE de 0.0728 e correlação de 0.8908. Além disso, a técnica de redes neurais é capaz de prever outras 5 variáveis geometalúrgicas que o modelo atualmente utilizado pela mina não prevê, garantindo um fechamento mássico e metalúrgico adequado das previsões. Logo, pode-se dizer que o modelo de redes neurais supera o modelo atualmente utilizado na operação.

6.1.8. Reconciliação

Com o objetivo de avaliar a eficácia do modelo elaborado através de redes neurais na previsão geometalúrgica das variáveis da usina de beneficiamento de zinco em estudo, decidiu-se aplicar o modelo construído em 24 pilhas monitoradas entre os meses de janeiro e abril de 2019. Todas as pilhas possuíam análise química de zinco e chumbo no ROM a cada 2 horas durante sua formação, juntamente com a massa de material enviado nas mesmas duas horas. Desta maneira, foi possível calcular os teores de Zn e Pb ponderados pela massa, de modo a obter os teores médios das pilhas. Para determinar a variável D1 em cada uma das pilhas analisadas, foi considerado o percentual de material proveniente de Extremo Norte em relação à massa total da pilha e para a variável D2 foi considerado o percentual de material proveniente de Lumiadeira em relação à massa total. Tendo em vista que algumas pilhas são formadas com material sem rastreabilidade (por exemplo: maticos, sobra de pilhas anteriores, minério do estoque, etc), optou-se por considerar apenas as pilhas cuja rastreabilidade do material fosse de, no mínimo, 80%. A Tabela 19 apresenta os teores de zinco e chumbo, as variáveis D1 e D2 e o percentual rastreável de cada uma das pilhas utilizadas no estudo de reconciliação.

Tabela 19 - Pilhas escolhidas para reconciliação

Pilha	% rastreável da pilha	Zn (%)	Pb (%)	D1	D2
4754	94.43	9.82	0.28	0.00	0.60
4755	89.13	9.86	0.28	0.00	0.63
4756	92.45	10.46	0.31	0.00	0.22
4757	89.60	9.92	0.33	0.00	0.10
4758	84.73	9.81	0.33	0.00	0.65
4770	94.95	10.70	0.28	0.20	0.80
4772	86.77	7.24	0.35	0.45	0.54
4773	94.33	11.18	0.33	0.49	0.51
4774	88.34	10.80	0.36	0.53	0.25
4775	90.87	11.02	0.38	0.17	0.83
4776	80.32	11.41	0.34	0.30	0.46
4777	94.89	10.97	0.32	0.22	0.36
4778	85.34	11.02	0.36	0.11	0.25
4784	83.52	11.12	0.38	0.14	0.26
4786	80.18	11.00	0.34	0.26	0.09
4787	97.85	10.68	0.37	0.15	0.48
4793	82.26	11.18	0.36	0.00	0.38
4797	83.70	11.34	0.39	0.37	0.00
4798	83.98	13.04	0.37	0.25	0.42
4799	90.32	12.58	0.37	0.40	0.19
4800	83.84	12.07	0.35	0.46	0.18
4801	93.07	11.79	0.35	0.46	0.21
4802	84.11	11.91	0.32	0.29	0.54
4805	96.08	12.23	0.26	0.27	0.64

Com as informações disponíveis na Tabela 19, foi possível calcular as estimativas geradas pela rede neural para as seis variáveis geometalúrgicas em estudo, conforme apresentadas na Tabela 20. Sabendo da inexistência de *overfitting* no modelo elaborado com 115 dados, optou-se por recalcular os pesos da rede considerando agora o banco completo, com 140 informações, utilizando a mesma estrutura da rede: 3 camadas ocultas com 5 nós ocultos na primeira camada e função de ativação linear, 4 nós ocultos na segunda camada e função de ativação linear, 5 nós ocultos na terceira camada e função de ativação softplus, função de ativação linear na camada de output e taxa de aprendizado igual a 0.03. Ao utilizar todos os dados, garante-se um modelo mais próximo da realidade, uma vez que o tamanho amostral se aproxima, ainda que de maneira sutil, do tamanho populacional.

Tabela 20 - Predições geradas pela rede neural para as seis variáveis geometalúrgicas estudadas

Pilha	Rec. Mass. Conc. Zn	Rec. Mass. Conc. Pb	Rec. Mass. Conc. Rejeito	Rec. Met. De Zn no Conc. Zn	Rec. Met. De Zn no Conc. Pb	Rec. Met. De Zn no Rejeito
4754	0.18	0.016	0.80	0.71	0.011	0.28
4755	0.18	0.016	0.80	0.71	0.011	0.28
4756	0.21	0.017	0.78	0.76	0.011	0.23
4757	0.19	0.017	0.79	0.75	0.011	0.24
4758	0.18	0.016	0.80	0.70	0.011	0.28
4770	0.19	0.016	0.79	0.69	0.011	0.30
4772	0.11	0.015	0.88	0.55	0.012	0.44
4773	0.20	0.015	0.78	0.69	0.011	0.30
4774	0.19	0.016	0.79	0.69	0.011	0.30
4775	0.20	0.016	0.78	0.70	0.011	0.29
4776	0.21	0.016	0.77	0.72	0.011	0.26
4777	0.21	0.016	0.78	0.73	0.011	0.26
4778	0.22	0.017	0.77	0.75	0.011	0.23
4784	0.22	0.016	0.77	0.75	0.011	0.24
4786	0.21	0.016	0.77	0.75	0.011	0.24
4787	0.20	0.016	0.78	0.72	0.011	0.27
4793	0.22	0.017	0.76	0.76	0.011	0.22
4797	0.22	0.016	0.76	0.75	0.011	0.24
4798	0.26	0.016	0.73	0.78	0.011	0.21
4799	0.25	0.016	0.74	0.76	0.011	0.23
4800	0.23	0.016	0.75	0.74	0.011	0.25
4801	0.22	0.016	0.76	0.73	0.011	0.26
4802	0.23	0.016	0.76	0.73	0.011	0.25
4805	0.23	0.016	0.75	0.74	0.011	0.25

Finalizada a estimativa das variáveis geometalúrgicas utilizando a rede construída, o passo seguinte foi rastrear as respostas metalúrgicas que as pilhas apresentaram ao passar pela usina de beneficiamento. O primeiro obstáculo encontrado durante esta etapa diz respeito à ausência de informação das massas. Não existe registro da massa alimentada na usina, da massa do concentrado de chumbo e do concentrado de zinco e nem do rejeito. Logo, as fórmulas tradicionalmente utilizadas para calcular a recuperação mássica e metalúrgica, as quais consideram a massa dos materiais envolvidos, não puderam ser utilizadas, dando lugar às equações de recuperação que consideram somente os teores da alimentação, concentrado e rejeito, conforme equacionado em 208 e 209.

$$Rec. Mássica = \frac{(Teor Alimentado - Teor Rejeito)}{(Teor Concentrado - Teor Rejeito)} \quad (208)$$

$$Rec. Met. = \frac{Teor Concentrado * (Teor Alimentado - Teor Rejeito)}{Teor Alimentado * (Teor Concentrado - Teor Rejeito)} \quad (209)$$

O segundo obstáculo diz respeito à ausência de informação do teor de zinco no concentrado de chumbo, o que impossibilitou o cálculo da recuperação mássica do concentrado de chumbo e da recuperação metalúrgica de zinco no concentrado de chumbo. As únicas informações disponíveis são: o teor alimentado na flotação de zinco, o teor do concentrado de zinco e o teor do rejeito da flotação de zinco. Desta maneira, é possível calcular, utilizando às equações 209 e 210, as recuperações mássica e metalúrgica para esta etapa do beneficiamento.

O modelo construído via redes neurais tem como *outputs* as recuperações mássica e metalúrgica em relação ao material alimentado na usina, ou seja, em relação ao ROM, pois só assim é possível fazer o balanço mássico e metalúrgico do minério (vide equações 194 a 199). Como os dados disponíveis na usina só permitem o cálculo das recuperações na etapa de flotação de zinco, foi necessário assumir duas premissas para transformá-las em resultados relativos ao ROM: a primeira premissa foi assumir que a recuperação mássica do concentrado de chumbo em relação ao ROM observada na usina foi igual ao valor médio observado nas 140 amostras utilizadas na elaboração das redes neurais, ou seja, 0.014; a segunda premissa, foi assumir que a recuperação metalúrgica de zinco no concentrado de chumbo em relação ao ROM observada na usina foi igual ao valor médio amostral, logo, 0.01. Desta maneira, os resultados da recuperação mássica do concentrado de zinco calculada na etapa de flotação de zinco devem ser multiplicados por 0.986 para se tornarem o valor aproximado da recuperação mássica de zinco em relação ao ROM. Para a recuperação metalúrgica de zinco no concentrado de zinco, deve-se multiplicar por 0.99 os resultados, levando-os assim para resultados em relação ao ROM. Os mesmos multiplicadores devem ser adotados para a recuperação mássica e metalúrgica do rejeito. A seguir encontra-se o exemplo do cálculo realizado para a pilha 4754, o qual facilitará a compreensão do problema e da solução adotada para esta questão:

A pilha 4754 apresentou 10.28% de zinco na alimentação da flotação de zinco, 39.66% de zinco no concentrado de zinco e 2.40% de zinco no rejeito da flotação. Tendo tais valores disponíveis é possível calcular, utilizando as equações 209 e 210, a recuperação mássica e metalúrgica de zinco no concentrado de zinco, as quais são respectivamente, 0.2116 e 0.8163. Como a recuperação mássica e metalúrgica de zinco no rejeito são o complementar de tais valores, temos 0.7884 e 0.1837, respectivamente. Uma vez que tais valores foram calculados considerando o teor na alimentação da flotação e não o teor do ROM, como é feito pela rede neural construída, é necessário colocar tais

valores de recuperação observados na usina na mesma base de comparação. Ao assumir que, em média, 1.4% da massa total do ROM e 1.0% do teor de zinco do ROM ficou retido na etapa de flotação de chumbo, é necessário reduzir os valores das recuperações calculadas na etapa de flotação de zinco considerando tais valores, conforme mostrado nas equações 210 a 213.

$$\begin{aligned} & \text{Rec. Mássica Conc. Zn em relação ao ROM} \\ & = 0.986 * \text{Rec. Mássica Conc. Zn na flotação de zinco,} \end{aligned} \quad (210)$$

$$\begin{aligned} & \text{Rec. Mássica do Rejeito em relação ao ROM} \\ & = 0.986 * \text{Rec. Mássica do Rejeito na flotação de zinco,} \end{aligned} \quad (211)$$

$$\begin{aligned} & \text{Rec. Met. de Zn no Conc. Zn em relação ao ROM} \\ & = 0.99 * \text{Rec. Met. de Zn no Conc. Zn na flotação de zinco,} \end{aligned} \quad (212)$$

$$\begin{aligned} & \text{Rec. Met. de Zn no Rejeito em relação ao ROM} \\ & = 0.99 * \text{Rec. Met. de Zn no Rejeito na flotação de zinco.} \end{aligned} \quad (213)$$

Para o exemplo da pilha 4754, a recuperação mássica no concentrado de zinco seria 0.2086 e a recuperação mássica do rejeito seria 0.7774, restando 0.014 para a recuperação mássica no concentrado de chumbo, todos valores em relação ao ROM. A recuperação metalúrgica de zinco no concentrado de zinco seria 0.8082 e a recuperação metalúrgica de zinco do rejeito seria 0.1818, restando 0.01 como a recuperação metalúrgica de zinco no concentrado de chumbo, valores também em relação ao ROM.

Calculadas as recuperações em relação ao ROM, o passo seguinte foi determinar os intervalos de tempo nos quais a usina processou o material proveniente das 24 pilhas em estudo. Foram considerados os resultados emitidos pela usina entre o horário que a pilha alimentou a usina mais 6 horas (tempo de residência do material) e o horário que a alimentação da pilha foi finalizada mais seis horas. Exemplo: se a pilha 4754 alimentou a planta no dia 01/01 às 12:00, foram considerados para fins de reconciliação os resultados emitidos a partir das 18:00. Se esta mesma pilha teve sua alimentação finalizada às 06:00 do dia 02/01, foram considerados os resultados emitidos até as 12:00 do dia 02/01. A Tabela 21 apresenta os resultados observados na usina para cada pilhas estudada.

Tabela 21 - Respostas da usina para cada uma das pilhas em estudo

Pilha	Início da alimentação	Fim da alimentação	Rec. Mássica Conc Zn	Rec. Mássica Conc Pb*	Rec. Mássica Rejeito	Rec. Met. de Zn no Conc Zn	Rec. Met. de Zn no Conc Pb*	Rec. Met. Rejeito
4754	05/01/2019 21:30	07/01/2019 00:31	0.220	0.014	0.766	0.812	0.010	0.178
4755	07/01/2019 00:31	08/01/2019 17:33	0.215	0.014	0.771	0.818	0.010	0.172
4756	08/01/2019 17:33	10/01/2019 06:35	0.231	0.014	0.755	0.830	0.010	0.160
4757	10/01/2019 08:35	12/01/2019 00:37	0.215	0.014	0.771	0.831	0.010	0.159
4758	12/01/2019 02:37	13/01/2019 21:39	0.205	0.014	0.781	0.818	0.010	0.172
4770	30/01/2019 17:58	01/02/2019 04:00	0.227	0.014	0.759	0.838	0.010	0.152
4772	01/02/2019 05:00	03/02/2019 04:02	0.200	0.014	0.786	0.822	0.010	0.168
4773	03/02/2019 04:02	04/02/2019 09:04	0.230	0.014	0.756	0.861	0.010	0.129
4774	04/02/2019 13:04	05/02/2019 17:05	0.229	0.014	0.757	0.857	0.010	0.133
4775	05/02/2019 18:05	06/02/2019 23:07	0.230	0.014	0.756	0.844	0.010	0.146
4776	07/02/2019 01:07	08/02/2019 04:08	0.241	0.014	0.745	0.863	0.010	0.127
4777	08/02/2019 06:08	09/02/2019 11:10	0.242	0.014	0.744	0.864	0.010	0.126
4778	09/02/2019 13:10	10/02/2019 17:11	0.240	0.014	0.746	0.859	0.010	0.131
4784	20/02/2019 10:22	22/02/2019 06:24	0.242	0.014	0.744	0.868	0.010	0.122
4786	24/02/2019 07:27	26/02/2019 04:29	0.247	0.014	0.739	0.869	0.010	0.121
4787	26/02/2019 06:29	28/02/2019 05:31	0.235	0.014	0.751	0.861	0.010	0.129
4793	09/03/2019 14:42	11/03/2019 05:44	0.242	0.014	0.744	0.860	0.010	0.130
4797	17/03/2019 03:51	19/03/2019 02:53	0.251	0.014	0.735	0.857	0.010	0.133
4798	19/03/2019 02:53	21/03/2019 01:55	0.298	0.014	0.688	0.882	0.010	0.108
4799	21/03/2019 01:55	22/03/2019 18:57	0.293	0.014	0.693	0.881	0.010	0.109
4800	23/03/2019 03:58	24/03/2019 12:59	0.308	0.014	0.678	0.891	0.010	0.099
4801	24/03/2019 23:00	26/03/2019 15:02	0.272	0.014	0.714	0.883	0.010	0.107
4802	26/03/2019 20:02	28/03/2019 15:04	0.271	0.014	0.715	0.883	0.010	0.107
4805	01/04/2019 08:08	03/04/2019 07:10	0.275	0.014	0.711	0.877	0.010	0.113

*Variáveis cujo valor foi fixado no valor médio amostral dos testes de bancada

Determinadas as respostas da usina, o passo seguinte consistiu em calcular o fator de escala entre os testes de flotação em bancada que são do tipo circuito aberto, os quais foram utilizados para construir o modelo, e a usina, que utiliza de um circuito de flotação fechado na flotação de zinco. Sabe-se que um circuito de flotação fechado permite a recirculação do material, aumentando assim a recuperação metalúrgica do concentrado de zinco. Desta maneira, é necessário estimar um fator de escala (*up-scaling*), para tornar os resultados gerados pelo modelo de redes neurais comparável com os resultados observados na usina. Assim, as 24 pilhas disponíveis foram divididas aleatoriamente em quatro partes: as três primeiras, com 6 amostras cada uma, foram utilizadas para calcular o fator de escala; e a quarta parte, também composta por 6 pilhas, foi utilizada para fins de avaliação da capacidade preditiva do modelo em escala industrial. A divisão das pilhas em bancos de teste, treinamento e dados futuros encontra-se representada na Tabela 22.

Tabela 22 - Divisão das pilhas em dados de treinamento, teste e dados futuros

Pilha	Parte 1	Parte 2	Parte 3	Parte 4
4754	-	-	-	Futuro
4758	-	-	-	Futuro
4805	-	-	-	Futuro
4773	-	-	-	Futuro
4799	-	-	-	Futuro
4778	-	-	-	Futuro
4787	Treinamento	Treinamento	Teste	-
4755	Treinamento	Treinamento	Teste	-
4756	Treinamento	Treinamento	Teste	-
4800	Treinamento	Treinamento	Teste	-
4784	Treinamento	Treinamento	Teste	-
4772	Treinamento	Treinamento	Teste	-
4775	Treinamento	Teste	Treinamento	-
4770	Treinamento	Teste	Treinamento	-
4797	Treinamento	Teste	Treinamento	-
4793	Treinamento	Teste	Treinamento	-
4757	Treinamento	Teste	Treinamento	-
4786	Treinamento	Teste	Treinamento	-
4801	Teste	Treinamento	Treinamento	-
4798	Teste	Treinamento	Treinamento	-
4776	Teste	Treinamento	Treinamento	-
4802	Teste	Treinamento	Treinamento	-
4777	Teste	Treinamento	Treinamento	-
4774	Teste	Treinamento	Treinamento	-

O grupo de pilhas pertencente à primeira partição dos dados teve seu valor da estimativa via redes neurais comparado com os resultados observados na usina. Desta forma, gerou-se informação da diferença relativa existente entre o predito e observado para cada pilha do primeiro banco de treinamento e a média da diferença relativa foi calculada para cada variável em estudo, conforme apresentado na Tabela 23. Tendo em vista a impossibilidade de confrontar as previsões das variáveis relacionadas à flotação de chumbo com o realizado pela usina, pelo fato de não serem monitoradas no cotidiano da planta, descarta-se a necessidade de estudar um fator para tais variáveis.

Tabela 23 – Diferença relativa entre predito pela rede neural versus realizado pela usina para as pilhas de treinamento do primeiro conjunto de dados

Pilha	Rec. Mass. Conc. Zn	Rec. Mass. Rejeito	Rec. Met. de Zn no Conc. Zn	Rec. Met. de Zn no Rejeito
4787	-14.3%	4.1%	-16.2%	106.8%
4755	-15.0%	3.8%	-13.3%	62.6%
4756	-11.0%	2.9%	-8.9%	45.4%
4800	-25.3%	11.1%	-16.8%	148.1%
4784	-10.5%	3.1%	-13.4%	94.4%
4772	-46.9%	11.8%	-32.9%	160.1%
4775	-12.1%	3.4%	-16.7%	95.2%
4770	-14.8%	4.2%	-17.3%	94.7%
4797	-12.9%	4.0%	-13.0%	82.8%
4793	-8.8%	2.5%	-11.2%	73.2%
4757	-9.5%	2.3%	-9.8%	50.9%
4786	-14.2%	4.4%	-14.3%	101.7%
Dif. relativa média	-16.3%	4.8%	-15.3%	93.0%

Analisando os resultados da Tabela 23, conclui-se que seria necessário multiplicar por 1.163 as previsões realizadas pela rede para a recuperação mássica no concentrado de zinco e multiplicar por 1.153 as previsões da recuperação metalúrgica de zinco no concentrado de zinco. Logo, tais valores são os fatores de escala calculados considerando a primeira partição realizada nos dados. Para evitar de utilizar mais um fator, o que comprometeria o fechamento mássico e metalúrgico das previsões, optou-se por calcular a recuperação mássica do rejeito e a recuperação metalúrgica de zinco no rejeito por diferença. Desta maneira, as seguintes fórmulas foram utilizadas para obter as previsões de tais variáveis no banco de teste da primeira partição:

$$Rec. Mass Rejeito = 1 - 1.163 * Rec. Mass Conc Zn - Rec. Mass Conc Pb, (214)$$

$$Rec. Met. de Zn no Rejeito$$

$$= 1 - 1.153 * Rec. Met. Zn no Conc Zn - Rec. Met. Zn no Conc Pb. (215)$$

Definidos os fatores de correção, o passo seguinte foi aplicá-los nas previsões realizadas pela rede neural para as pilhas do banco de teste da primeira partição dos dados.

Feito isso, os valores da diferença relativa entre predito e observado foram calculados, gerando a tabela a seguir:

Tabela 24 - Diferenças relativas entre valores preditos via redes neurais com fator de escala e respostas observadas na usina para as pilhas de teste do primeiro conjunto de dados

Pilhas	Rec. Mass. Conc. Zn	Rec. Mass. Rejeito	Rec. Met. de Zn no Conc. Zn	Rec. Met. de Zn no Rejeito
4801	-4.9%	1.6%	-4.5%	36.0%
4798	0.9%	-0.7%	1.7%	-14.3%
4776	3.2%	-1.3%	-3.3%	21.5%
4802	-3.1%	0.9%	-4.1%	32.5%
4777	-0.4%	-0.2%	-2.5%	16.7%
4774	-1.4%	0.2%	-7.4%	46.9%
Dif. relativa média	-0.96%	0.12%	-3.35%	23.22%

Percebe-se uma melhora significativa dos resultados com fator aplicado em relação às estimativas sem fator de escala: de -16.3% para -0.96% para recuperação mássica do concentrado de zinco, de 4.8% para 0.12% para a recuperação mássica do rejeito, de -15.3% para -3.35% para a recuperação metalúrgica de zinco no concentrado de zinco e de 93.0% para 23.22% para a recuperação metalúrgica de zinco no rejeito.

O passo seguinte foi realizar o mesmo procedimento para os dois conjuntos de dados seguintes: calcular o fator de escala no banco de treinamento e aplicá-lo ao banco de teste. Ao fazer isso, temos os seguintes fatores encontrados nos bancos de treinamento para as três partições de dados em estudo e também o fator médio entre as três partições, representado na Tabela 25.

Tabela 25 - Fatores de escala encontrados para as três partições de dados estudadas

Cálculo do fator de correção - Training Set	Rec. Mass. Conc. Zn	Rec. Mass. Rejeito	Rec. Met. de Zn no Conc. Zn	Rec. Met. de Zn no Rejeito
Parte 1	-16.3%	4.8%	-15.3%	93.0%
Parte 2	-17.7%	5.6%	-16.5%	111.0%
Parte 3	-13.4%	4.3%	-15.0%	101.1%
Média	-15.8%	4.9%	-15.6%	101.7%

Aplicando os fatores de escala em seus respectivos bancos de teste temos as diferenças relativas médias representadas na Tabela 26. A média das diferenças relativas médias representa o erro médio que será observado quando o modelo de redes neurais juntamente com o fator de escala médio, apresentado na Tabela 25, for aplicado em dados não utilizados no cálculo do fator de escala.

Tabela 26 - Fatores de escala aplicados em seus respectivos bancos de teste

Aplicação do fator de escala nas estimativas - Test Set	Rec. Mass. Conc. Zn	Rec. Mass. Rejeito	Rec. Met. de Zn no Conc. Zn	Rec. Met. de Zn no Rejeito
Parte 1	-0.96%	0.12%	-3.35%	23.22%
Parte 2	3.47%	-1.38%	0.55%	-4.09%
Parte 3	-9.80%	2.74%	-4.47%	25.12%
Média	-2.43%	0.49%	-2.42%	14.75%

Após o estudo dos resultados nas três partições de dados optou-se por adotar como o fator de escala final aquele que consiste na média dos fatores de escala obtidos para as três partições. Logo, os valores preditos pela rede para a recuperação mássica no concentrado de zinco devem ser multiplicados por 1.158 e os valores da recuperação metalúrgica de zinco no concentrado de zinco devem ser multiplicados por 1.156. As equações que serão utilizadas para ajustar os resultados das recuperações no rejeito encontram-se representadas em 216 e 217.

$$Rec. Mass Rejeito = 1 - 1.158 * Rec. Mass Conc Zn - Rec. Mass Conc Pb, (216)$$

$$Rec. Met. de Zn no Rejeito$$

$$= 1 - 1.156 * Rec. Met. Zn no Conc Zn - Rec. Met. Zn no Conc Pb. (217)$$

Aplicando tais fatores nas estimativas realizadas pela rede nos dados referentes à quarta partição, observa-se os resultados apresentados na Tabela 27.

Tabela 27 – Diferença relativa observada quando o fator de escala foi aplicado nas previsões da rede neural para a quarta partição dos dados

Pilha	Rec. Mass. Conc. Zn	Rec. Mass. Rejeito	Rec. Met. de Zn no Conc. Zn	Rec. Met. de Zn no Rejeito
4754	-3.9%	0.8%	1.0%	-5.2%
4758	2.5%	-1.0%	-0.4%	1.1%
4805	-2.2%	0.6%	-2.5%	18.4%
4773	0.6%	-0.4%	-7.9%	51.9%
4799	-3.1%	1.0%	0.0%	-0.8%
4778	3.6%	-1.5%	1.5%	-10.3%
Diferença relativa média	-0.4%	-0.1%	-1.4%	9.2%

As diferenças relativas médias da recuperação mássica do concentrado de zinco e da recuperação metalúrgica de zinco no concentrado de zinco foram muito próximas de zero (-0.4% e -0.1%, respectivamente), indicando um bom ajuste do modelo com fator de escala aos dados. A recuperação metalúrgica de zinco no concentrado de zinco também apresentou um ótimo resultado (-1.4%). Sabendo que a Mina de Vazante considera como adequadas previsões da recuperação metalúrgica de zinco no concentrado de zinco que tenham diferença de até ± 2 pontos percentuais em relação ao observado na usina, é possível confirmar a validade operacional do modelo geometalúrgico gerado via redes neurais na previsão da principal variável de controle da planta.

A recuperação metalúrgica de zinco no rejeito apresentou um erro mais alto, mas fortemente influenciado pela presença de um comportamento distinto da pilha 4773 na usina, o qual gerou uma diferença relativa de 51.9%. Se excluída tal pilha da análise a diferença passaria para 0.6%.

O erro médio absoluto (MAE) e a raiz quadrada do erro quadrático médio (RMSE) das previsões com o fator de escala aplicado encontram-se resumidos para cada variável na Tabela 28, juntamente com a correlação entre as estimativas e o observado na usina.

Tabela 28 - MAE, RMSE e correlação entre predito com fator de escala e observado na usina para a quarta partição dos dados

Pilha	Rec. Mass. Conc. Zn	Rec. Mass. Rejeito	Rec. Met. de Zn no Conc. Zn	Rec. Met. de Zn no Rejeito
MAE	0.006	0.007	0.019	0.019
RMSE	0.007	0.007	0.030	0.029
Correlação	0.978	0.977	0.581	0.583

Ao desconsiderar a pilha 4773, os resultados da Tabela 28 são atualizados para os que estão apresentados na Tabela 29. Observa-se valores de MAE e RMSE muito próximos de zero, indicando a qualidade do modelo elaborado combinado com o fator de escala. As correlações entre predito e observado também indicam uma excelente aderência entre os resultados.

Tabela 29 - MAE, RMSE e correlação entre predito com fator de escala e observado na usina para a quarta partição dos dados – desconsiderando a pilha 4773

Pilha	Rec. Mass. Conc. Zn	Rec. Mass. Rejeito	Rec. Met. de Zn no Conc. Zn	Rec. Met. de Zn no Rejeito
MAE	0.008	0.007	0.009	0.009
RMSE	0.008	0.008	0.012	0.012
Correlação	0.977	0.976	0.913	0.914

Com base nos resultados apresentados é possível concluir que a metodologia aqui descrita está validada e pode ser utilizada na rotina da mina de Vazante para prever o balanço mássico e metalúrgico da planta, garantindo previsão acurada das variáveis com maior impacto econômico e ambiental e garantindo o fechamento mássico e metalúrgico das previsões.

6.1.9. Discussão dos resultados e sugestões de trabalhos futuros

Os resultados obtidos na seção 6.1 mostram que o modelo elaborado via redes neurais artificiais apresentou melhores resultados que os obtidos quando as técnicas de análise de regressão, *random forest* e *support vector regression* foram utilizadas, superando, para a maioria das variáveis modeladas, os resultados das demais técnicas em termos de MAE, RMSE e correlação nos bancos de teste da validação cruzada *5-fold*. Além disso, a qualidade do fechamento mássico e metalúrgico obtido nas previsões realizadas através de redes neurais foram superiores às obtidas com as demais técnicas, gerando um erro inferior a $\pm 1\%$ em 97% das amostras do banco de teste.

Ao confrontar o modelo de redes neurais com o modelo atualmente utilizado pela mina de Vazante para prever a recuperação metalúrgica de zinco no concentrado de

zinco, observou-se resultados muito melhores quando a técnica proposta neste trabalho foi utilizada, sugerindo que esta deveria ser implementada na rotina da operação para predizer com maior precisão e acurácia as respostas da usina de beneficiamento.

Durante a etapa de reconciliação, observou-se, que ao aplicar um fator de mudança escala (laboratório para usina) nas previsões realizadas através da técnica de redes neurais, obtinha-se resultados muito próximos aos realizados pela usina de beneficiamento para as pilhas avaliadas, indicando erros inferiores a 2% para todas as variáveis monitoradas na reconciliação, após a remoção de uma pilha considerada outlier. Os pontos destacados acima reforçam a validade da metodologia proposta neste estudo na predição das variáveis geometalúrgicas da mina de zinco de Vazante.

O próximo passo deste estudo consiste na estimativa/simulação das variáveis de input da rede neural para que, em seguida, cada bloco do depósito possa alimentar a rede com seus teores de Zn e Pb no ROM e com o alvo ao qual o bloco pertence, gerando assim uma estimativa 3D das 6 variáveis geometalúrgicas em estudo, já corrigidas pelo fator de escala. Pode-se ainda considerar, em estudos futuros, a incorporação de novas variáveis no modelo, como o teor de zinco e chumbo nos concentrados, o teor de contaminantes que impactam negativamente no processo hidrometalúrgico ao qual o concentrado de zinco é posteriormente submetido, o teor de contaminantes nos rejeitos, dentre outras.

6.2. MINAS DE OURO

6.2.1. Apresentação dos depósitos

A localização dos depósitos de ouro em estudo, bem como informações sobre a geologia dos mesmos não puderam ser divulgadas devido ao sigilo solicitado pela empresa fornecedora dos dados. A única informação que pode ser divulgada é relativa ao estado brasileiro no qual as minas estão localizadas: Minas Gerais.

6.2.2. Apresentação dos dados e estatísticas descritivas

Os materiais retirados das duas minas de ouro em estudo são processados inicialmente na usina 1, a qual conta com circuitos de britagem e moagem, seguidos por uma etapa de separação gravimétrica e um circuito de flotação. Em seguida, o concentrado da flotação é enviado para usina 2, na qual o material é submetido a uma etapa de calcinação e seguida pelo circuito de lixiviação. A Figura 43 ilustra o fluxograma simplificado do material desde a alimentação até o rejeito gerado pela etapa de lixiviação.

Com o objetivo de replicar o comportamento observado na usina de beneficiamento, foram desenvolvidos testes em escala de bancada para que amostras com diferentes características químicas e litológicas fossem testadas, visando avaliar sua performance antes mesmo que os materiais alimentassem a planta industrial. No total, 37 amostras foram coletadas e submetidas aos testes de bancada, gerando informação suficiente para construir um modelo estatístico capaz de relacionar os teores do ROM com as recuperações mássica e metalúrgica dos produtos e dos rejeitos gerados na usina.

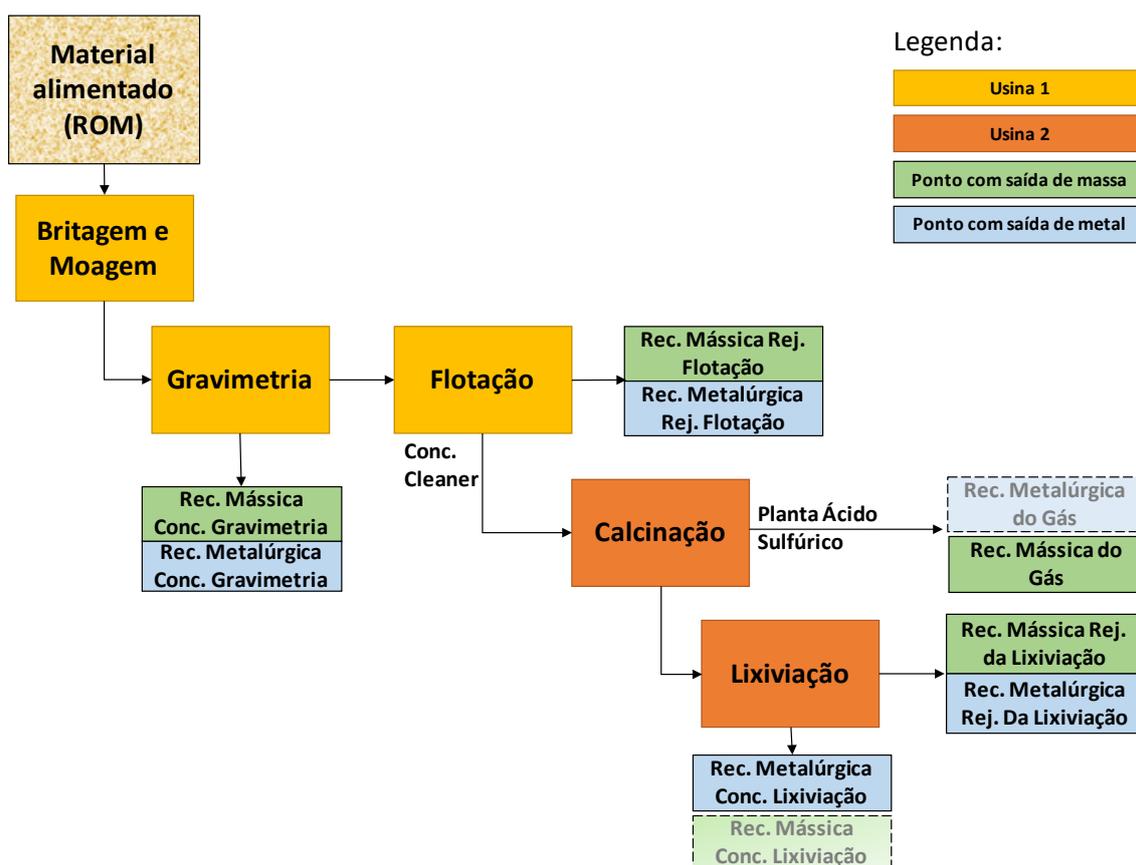


Figura 43 - Fluxograma simplificado das usinas 1 e 2

Para modelar o balanço mássico e metalúrgico do material que alimenta o circuito de concentração das minas em estudo, é necessário primeiramente mapear os pontos onde há concentração de material (visando a geração de produtos vendáveis) e os pontos onde o material é descartado (gerando um passivo ambiental), além de monitorar os teores em cada um destes pontos. Estas informações darão origem às variáveis dependentes do estudo. As siglas utilizadas para representar a recuperação mássica e a recuperação metalúrgica, deste ponto em diante, serão RMA e RME, respectivamente.

O primeiro ponto de concentração ocorre na etapa de concentração gravimétrica onde é possível quantificar o percentual da massa total alimentada no processo que é concentrado a partir deste processo de concentração físico. Após realizadas análises químicas do concentrado gravimétrico e do rejeito que será enviado para etapa de flotação, é possível calcular a recuperação metalúrgica desta etapa. As fórmulas utilizadas para calcular a recuperação mássica e metalúrgica desta etapa estão apresentadas nas equações 218 e 219.

$$RMAS \text{ Conc. Gravimétrico} = \frac{\text{Massa Conc. Gravimétrico}}{\text{Massa alimentada}}, \quad (218)$$

$$RMET \text{ Conc. Grav.} = \frac{\text{Massa Conc. Grav.} * \text{Teor Conc. Gravimétrico}}{\text{Massa alimentada} * \text{Teor Alimentação}}. \quad (219)$$

O segundo ponto de concentração ocorre na etapa de flotação. O rejeito da etapa de concentração gravimétrica alimenta o circuito de flotação, logo, uma importante informação é o percentual da massa inicial que alimentou a etapa de flotação e o percentual do metal inicial que alimentou a flotação, dados pelas equações 220 e 221.

$$\% \text{ Massa inicial que alimentou a flotação} = (1 - RMAS \text{ Conc. Grav.}), \quad (220)$$

$$\% \text{ Metal inicial que alimentou a flotação} = (1 - RMET \text{ Conc. Grav.}). \quad (221)$$

As recuperações mássica e metalúrgica do concentrado e do rejeito da flotação são calculadas considerando as fórmulas a seguir.

$$RMAS \text{ Conc. Flotação} = \frac{\text{Massa Conc. Flotação}}{\text{Massa Alimentada Flotação}}, \quad (222)$$

$$RMET \text{ Conc. Flotação} = \frac{\text{Massa Conc. Flotação} * \text{Teor Conc. Flotação}}{\text{Massa Alim. Flotação} * \text{Teor Alim. Flotação}}, \quad (223)$$

$$RMAS \text{ Rej. Flotação} = \frac{\text{Massa Rej. Flotação}}{\text{Massa Alimentada Flotação}}, \quad (224)$$

$$RMET \text{ Rej. Flotação} = \frac{\text{Massa Rej. Flotação} * \text{Teor Rej. Flotação}}{\text{Massa Alim. Flotação} * \text{Teor Alim. Flotação}}. \quad (225)$$

Obtido o resultado das fórmulas 222 e 224, o passo seguinte é multiplicá-los pelo percentual da massa inicial que alimentou a flotação (fórmula 220), para que seja possível compreender quanto por cento da massa inicial teve como destino o concentrado e o rejeito da flotação, respectivamente. O mesmo vale para a recuperação metalúrgica do

concentrado e do rejeito da flotação, valores que devem ser multiplicados pelo percentual do metal inicial que alimentou a flotação, tornando possível quantificar a porcentagem do metal inicial que foi destinado ao concentrado e ao rejeito da flotação. Logo, a recuperação mássica do concentrado e do rejeito da flotação em relação à massa inicial e a recuperação metalúrgica do concentrado e do rejeito da flotação em relação à quantidade inicial de metal são dadas pelas equações 226, 227, 228 e 229, respectivamente.

RMA_S Conc. Flotação Final

$$= RMA_{S \text{ Conc. Flot}} * \% \text{ Massa inicial alim. da flotação}, \quad (226)$$

RMA_S Rej. Flotação Final

$$= RMA_{S \text{ Rej. Flot}} * \% \text{ Massa inicial alim. da flotação}, \quad (227)$$

RME_T Conc. Flotação Final

$$= RME_{T \text{ Conc. Flot}} * \% \text{ Metal inicial alim. da flotação}, \quad (228)$$

RME_T Rej. Flotação Final

$$= RME_{T \text{ Rej. Flot}} * \% \text{ Metal inicial alim. da flotação}. \quad (229)$$

O rejeito da flotação é enviado para a barragem, logo, o monitoramento da massa e do percentual de ouro enviados para a barragem são duas importantes variáveis de controle. O concentrado da flotação alimenta a etapa de calcinação, logo, o percentual da massa inicial que alimentou a etapa de calcinação é idêntico ao valor obtido na equação 226, uma vez que não há perda de material entre as etapas. O mesmo vale para o percentual do metal inicial que alimentou a calcinação, sendo este valor igual ao resultado obtido na equação 228.

Na etapa de calcinação, a queima do concentrado da flotação faz com que parte da massa do material alimentado nesta etapa seja transformada em gás, o qual é enviado para fábrica de ácido sulfúrico da empresa. Desta maneira, para o fechamento do balanço mássico, é necessário considerar a massa perdida na etapa de calcinação. A fórmula 230 fornece a recuperação mássica após a calcinação, logo, a massa perdida na forma de gás é seu complementar, dado pela equação 231.

$$RMAS \text{ após Calcinação} = \frac{\text{Massa após etapa de calcinação}}{\text{Massa alimentada na calcinação}}, \quad (230)$$

$$RMAS \text{ Gás} = (1 - RMAS \text{ após Calcinação}). \quad (231)$$

O percentual de massa perdido na forma de gás em relação a massa inicial deve ser calculado utilizando a fórmula 232:

$$RMAS \text{ Gás Final} = RMAS \text{ Conc. Flotação Final} * RMAS \text{ Gás}. \quad (232)$$

Em relação ao teor de ouro na etapa de calcinação, é possível afirmar que todo o ouro que sai no concentrado da flotação e alimenta a calcinação é alimentado na etapa seguinte (lixiviação), não havendo perda do metal de interesse nesta etapa do processo, uma vez que a queima do concentrado faz com que somente o enxofre se transforme em gás, não afetando o ouro presente no material.

A última etapa de concentração é a lixiviação do material calcinado. O percentual da massa inicial que alimentou a lixiviação pode ser obtido através da fórmula 233:

$$\begin{aligned} \% \text{ Massa inicial que alimentou a lixiviação} \\ = RMAS \text{ após Calcinação} * RMAS \text{ Conc. Flotação Final}. \end{aligned} \quad (233)$$

O percentual do metal inicial que alimentou a lixiviação é igual ao percentual do metal inicial que alimentou a calcinação, logo, é dado pela equação 228, uma vez que não há perda de metal no gás, conforme explicado anteriormente.

Sabe-se que o concentrado produzido na etapa de lixiviação está diluído em uma solução de cianeto e que sua massa é muito reduzida. Por exemplo, para a amostra “12 BC”, dos 200g de material alimentado no teste de lixiviação, 6.58 gramas de ouro por tonelada estavam diluídos na solução de cianeto. Percebe-se então, que o percentual de massa do concentrado em relação a massa inicial é irrisório, sendo mais prudente adotar que 100% da massa alimentada na etapa de lixiviação é destinada ao rejeito, ao invés de

tentar quantificar uma massa tão reduzida que está sujeita a erros de medição na balança. Desta maneira, adotaremos neste estudo que não há recuperação mássica no concentrado da lixiviação, por se tratar de um valor muito pequeno. Logo, para calcular a porcentagem da massa inicial que foi destinada ao rejeito da lixiviação, basta utilizar a equação 234.

$$RMAS \text{ Rej. Lixiviação} = \% \text{ Massa inicial que alimentou a lixiviação.} \quad (234)$$

O cálculo da recuperação metalúrgica do rejeito da lixiviação pode ser feito utilizando a equação 235. Tendo em vista a massa reduzida do concentrado produzido na etapa de lixiviação, o cálculo da recuperação metalúrgica do concentrado da lixiviação deve ser feito através do complementar dos resultados observados para o rejeito da lixiviação, conforme ilustrado na fórmula 236.

$$RMET \text{ Rej. Lixiviação} = \frac{\text{Massa Rej. Lixiviação} * \text{Teor Rej. Lixiviação}}{\text{Massa Alim Lixiviação} * \text{Teor Alim Lixiviação}}, \quad (235)$$

$$RMET \text{ Conc. Lixiviação} = 1 - RMET \text{ Rej. Lixiviação.} \quad (236)$$

Para quantificar a recuperação metalúrgica do concentrado e do rejeito da lixiviação em relação ao metal inicialmente alimentado no processo, basta utilizar as equações 237 e 238, respectivamente:

$$\begin{aligned} RMET \text{ Conc. Lix. Final} \\ = RMET \text{ Conc. Lixiviação} * RMET \text{ Conc. Flotação Final,} \end{aligned} \quad (237)$$

$$\begin{aligned} RMET \text{ Rej. Lix. Final} \\ = RMET \text{ Rej. Lixiviação} * RMET \text{ Conc. Flotação Final.} \end{aligned} \quad (238)$$

Ao somar as informações de recuperação mássica do concentrado gravimétrico, do rejeito da flotação, do gás e do rejeito da lixiviação, os três últimos em relação à massa

inicial alimentada no processo, tem-se o valor um conforme representado na equação 239, uma vez que toda massa alimentada é destinada a um destes produtos ou rejeitos.

$$\begin{aligned}
 &RMAS\ Conc.\ Gravimétrico + RMAS\ Rej.\ Flotação\ Final + RMAS\ Gás\ Final \\
 &+ RMAS\ Rej.\ Lixiviação\ Final = 1.
 \end{aligned}
 \tag{239}$$

A soma igual a um também ocorre quando são somadas a recuperação metalúrgica do concentrado gravimétrico, do rejeito da flotação, do concentrado da lixiviação e do rejeito da lixiviação, os três últimos em relação à quantidade de metal inicialmente alimentada no processo. A equação 240 representa esta afirmação.

$$\begin{aligned}
 &RMET\ Conc.\ Gravimétrico + RMET\ Rej.\ Flotação\ Final \\
 &+ RMET\ Conc.\ Lixiviação\ Final + RMET\ Rej.\ Lixiviação\ Final \\
 &= 1.
 \end{aligned}
 \tag{240}$$

O fechamento da recuperação mássica e metalúrgica total representa que não houve perda de massa ou metal durante o processo de beneficiamento, tornando possível fazer o balanço mássico e metalúrgico da usina.

Definidas as variáveis em estudo, o passo seguinte é analisá-las do ponto de vista estatístico. A Tabela 30 mostra as estatísticas descritivas das amostras em estudo, apresentando informações dos teores de Au (g/t) e S (%) no ROM, da recuperação mássica do concentrado gravimétrico, do concentrado do rejeito da flotação final, do gás final e do rejeito da lixiviação final, além de informações da recuperação metalúrgica do concentrado gravimétrico, do rejeito da flotação final, do concentrado da lixiviação final e do rejeito da lixiviação final. As Figuras 44 e 45 apresentam os box-plots para as mesmas variáveis supracitadas. Note que as estatísticas se encontram divididas em duas partes, a primeira com os dados da Mina A e a segunda com as informações relativas à Mina B. O mesmo ocorre para os box-plots, sendo os cor de rosa relacionados à Mina A e os azuis à Mina B.

O teor médio de ouro no ROM na Mina A é bem superior ao observado na Mina B (11.42 g/t contra 2.38 g/t). Em relação ao teor de enxofre, esta superioridade se mantém, mas em menor grau, sendo o teor médio na Mina A de 8.55% e na Mina B de 7.53%.

Observando o destino do metal alimentado proveniente da Mina A, conclui-se que, em média, 60% do ouro é destinado ao concentrado gravimétrico, 7% é enviado para o rejeito da flotação, 2% é enviado para o rejeito da lixiviação e 30% é destinado ao concentrado da lixiviação. Para a Mina B, estes números são: 71% do ouro é destinado ao concentrado gravimétrico, 5% ao rejeito da flotação, 2% ao rejeito da lixiviação e 23% ao concentrado da lixiviação.

Ao observar a massa alimentada, proveniente da Mina A, conclui-se que, em média, 6% do material é destinado ao concentrado gravimétrico, 79% da massa é enviada para o rejeito da flotação, 11% é enviada para o rejeito da lixiviação e 3% é transformada em gás. Considerando a Mina B, conclui-se que, em média, 5% da massa alimentada é destinada ao concentrado gravimétrico, 87% da massa é enviada para o rejeito da flotação, 6% é enviada para o rejeito da lixiviação e 2% é transformada em gás.

Tabela 30 - Estatísticas descritivas – Minas de Ouro

Mina	Variável	N	Média	Desvio Padrão	Mínimo	Q1	Mediana	Q3	Máximo
Mina A	Au (g/t)	32	11.42	10.58	1.07	5.32	8.21	13.75	51.33
	S (%)	32	8.55	7.58	0.38	3.62	6.33	10.41	31.64
	RMA Conc. Gravimétrico	32	0.06	0.01	0.03	0.05	0.06	0.06	0.08
	RME Conc. Gravimétrico	32	0.60	0.14	0.26	0.51	0.59	0.71	0.87
	RMA Rej. Flotação Final	32	0.79	0.10	0.50	0.73	0.81	0.88	0.93
	RME Rej. Flotação Final	32	0.07	0.06	0.01	0.03	0.05	0.10	0.22
	RMA Gás Final	19	0.03	0.02	0.01	0.02	0.02	0.04	0.07
	RME Conc. Lixiviação Final	32	0.30	0.13	0.10	0.21	0.27	0.39	0.66
	RME Rej. Lixiviação Final	32	0.02	0.01	0.00	0.01	0.02	0.02	0.04
RMA Rej. Lixiviação Final	19	0.11	0.06	0.02	0.06	0.10	0.13	0.28	
Mina B	Au (g/t)	5	2.38	2.31	0.46	0.83	2.12	2.21	6.27
	S (%)	5	7.53	11.67	0.85	1.48	3.46	3.56	28.30
	RMA Conc. Gravimétrico	5	0.05	0.01	0.04	0.04	0.05	0.06	0.06
	RME Conc. Gravimétrico	5	0.71	0.10	0.57	0.67	0.70	0.78	0.82
	RMA Rej. Flotação Final	5	0.87	0.03	0.84	0.86	0.87	0.88	0.93
	RME Rej. Flotação Final	5	0.05	0.02	0.03	0.03	0.05	0.06	0.07
	RMA Gás Final	3	0.02	0.00	0.01	0.01	0.02	0.02	0.02
	RME Conc. Lixiviação Final	5	0.23	0.08	0.12	0.18	0.25	0.25	0.33
	RME Rej. Lixiviação Final	5	0.02	0.01	0.01	0.01	0.02	0.02	0.03
RMA Rej. Lixiviação Final	3	0.06	0.01	0.05	0.05	0.06	0.07	0.07	

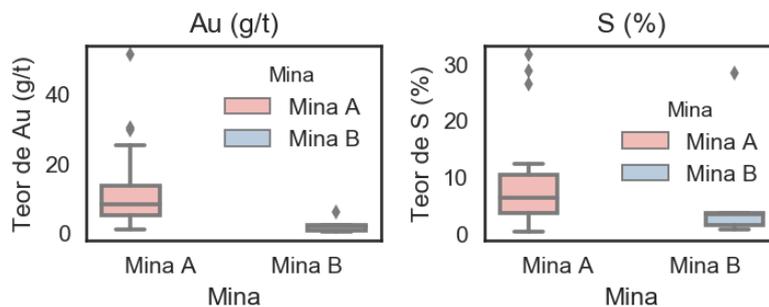


Figura 44 - Box-plots dos teores de Au e S separados por mina

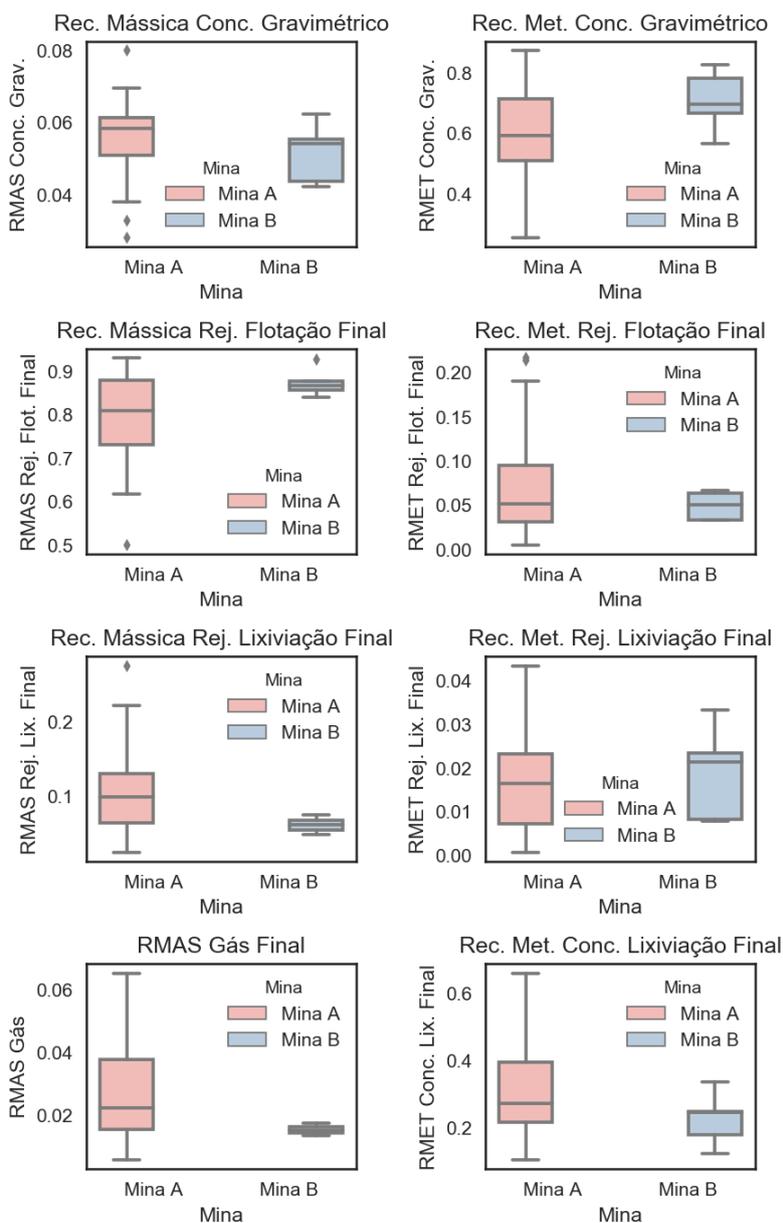


Figura 45 - Box-plots das variáveis metalúrgicas – Minas de Ouro

Ao avaliar as informações da Tabela 30 e das Figuras 44 e 45, é possível perceber que, tanto os teores de alimentação, quanto as respostas observadas no processo de beneficiamento, apresentam comportamentos distintos em relação à mina analisada. Com isso, decidiu-se criar uma variável categórica D1 que armazena a informação de qual mina a amostra pertence, de modo que esta variável possa ser utilizada como *input* no modelo gerado. A equação 241 demonstra como a variável D1 foi construída:

$$D1 = \begin{cases} 0, & \text{caso amostra pertença a Mina B} \\ 1, & \text{caso a amostra pertença a Mina A} \end{cases} \quad (241)$$

6.2.2.1. Pré-processamento dos dados

Para construir um modelo para fins de previsão, é necessário que o banco de dados seja isotópico, o que não é o caso do banco em questão, uma vez que somente 22 das 37 amostras existentes possuem informação da massa que sobrou após a etapa de calcinação. A ausência desta informação faz com que não seja possível calcular a recuperação mássica do gás final e a recuperação mássica do rejeito da lixiviação final. Tendo em vista que a exclusão dos dados que não possuem informação da massa após calcinação não é uma opção, devido ao reduzido número de amostras que restaria, optou-se por realizar a imputação dos dados faltantes. Escolheu-se a técnica de Máxima Verossimilhança (Enders, 2010) que é o estado da arte para imputação de dados faltantes aleatórios (MAR). Os dados faltantes deste estudo são classificados como MAR uma vez que as amostras cuja calcinação não foi medida são pertencentes à primeira campanha de amostragem geometalúrgica, a qual, juntamente com a segunda campanha, focou em amostras com teor mais alto de ouro no ROM. Consequentemente, existe relação da variável com dados faltantes com outra variável do banco de dados, o que a torna um caso de MAR.

Como Enders (2010) assume a normalidade dos dados realizar a imputação através da técnica de Máxima Verossimilhança, o primeiro passo foi verificar se a distribuição dos dados da variável calcinação era normal. Calculou-se, portanto, o teste de Anderson Darling, conforme ilustrado na Figura 46. Como o valor da estatística de Anderson Darling foi 0.151 e este valor é menor do que o valor crítico (0.696), aceita-se a hipótese de normalidade desta variável.

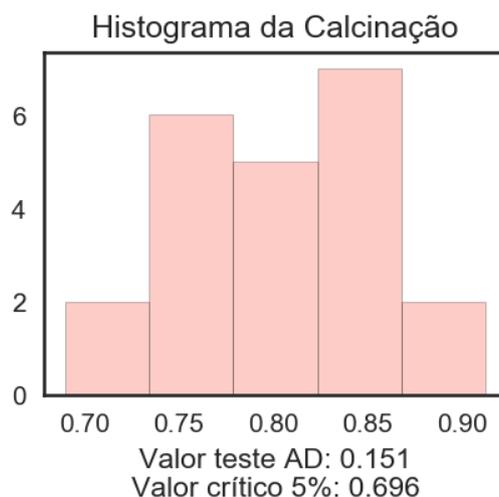


Figura 46 - Histograma da variável calcinação

O próximo passo consistiu em escolher qual variável seria utilizada para conduzir a imputação. Escolheu-se o teor de enxofre no concentrado da flotação por ser esta a variável com correlação mais forte com a variável a ser imputada (-0.608), conforme ilustrado na Figura 47.

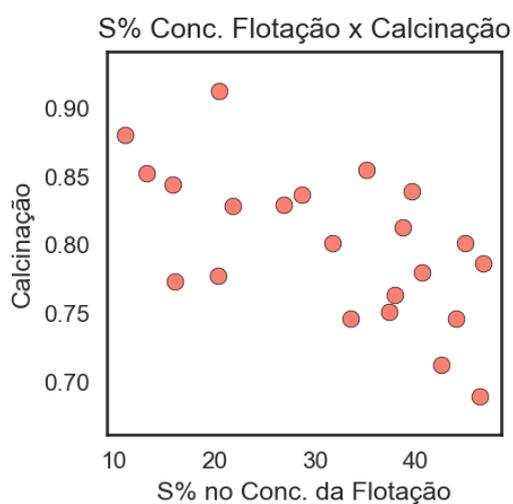


Figura 47 - Gráfico de dispersão entre teor de enxofre no concentrado da flotação e a massa que sobrou após etapa de calcinação

Realizada a imputação utilizando a técnica de Máxima Verossimilhança e o algoritmo *Expectation Maximization* (detalhados no capítulo 5), o passo seguinte foi

avaliar se o histograma da variável calcinação manteve sua forma original após a imputação dos dados. Analisando a Figura 48 percebe-se que o histograma reproduz as características estatísticas da variável calcinação.

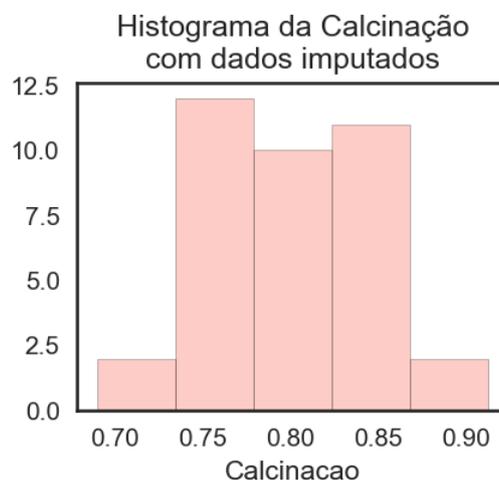


Figura 48 - Histograma da calcinação com dados imputados

Finalizada a imputação, o passo seguinte foi transformar os teores de Au e S para valores contidos no intervalo entre 0 e 1, tendo em vista as diferentes grandezas das variáveis independentes em estudo. Desta forma, evita-se que alguma variável tenha maior influência nos modelos estatísticos pelo simples fato de sua grandeza ser maior que a das demais variáveis. A transformação entre 0 e 1 foi escolhida por resultar em valores de Au e S contidos no mesmo intervalo observacional da variável D1.

6.2.3. Modelos construídos com todos os dados

Ao contrário do que foi realizado para o estudo de caso na mina de zinco, não foi possível dividir o banco de dados das minas de ouro em bancos de treinamento e teste e um banco final para validação dos modelos, devido à escassez de dados amostrais. Logo, as análises realizadas nesta seção consideram as 37 amostras disponíveis.

6.2.3.1. Análise de Regressão

A primeira técnica aplicada aos dados geometalúrgicos das minas de ouro em avaliação foi a análise de regressão multivariada. Foram construídos 8 modelos de regressão, um para cada variável dependente em estudo, os quais encontram-se apresentados a seguir. A Figura 59 apresenta os gráficos dos valores preditos versus os observados através dos modelos que serão apresentados e a Tabela 31 resume os coeficientes de regressão encontrados para os mesmos.

O primeiro modelo foi elaborado para prever a recuperação mássica do concentrado gravimétrico. A equação obtida encontra-se apresentada em 242 e os dados do modelo ajustado encontram-se na Figura 49. A única variável independente estatisticamente válida para construir um modelo para predição da recuperação mássica do concentrado gravimétrico foi o teor de S (%). Percebe-se que quanto maior o teor de enxofre no ROM, maior o valor da variável dependente. O coeficiente de determinação do modelo é igual a 0.351, indicando que 35.1% da variabilidade desta variável pode ser explicada através do modelo elaborado. A análise dos resíduos, apresentada nas Figuras 57 e 58 revelou que estes são normalmente distribuídos e homocedásticos.

$$RMAS\ Conc.\ Gravimétrico = 0.0489 + 0.0245 * S. \quad (242)$$

OLS Regression Results					
Dep. Variable: RMAS Conc. Grav.		R-squared: 0.351			
Model: OLS		Adj. R-squared: 0.333			
Method: Least Squares		F-statistic: 18.94			
Time: 16:48:01		Log-Likelihood: 124.12			
No. Observations: 37		AIC: -244.2			
Df Residuals: 35		BIC: -241			
Df Model: 1					
Covariance Type: nonrobust					
	coef	std err	t	P> t	[0.025 0.975]
const	0.0489	0.002	24.045	0	0.045 0.053
S	0.0245	0.006	4.352	0	0.013 0.036
Omnibus: 4.082		Durbin-Watson: 1.672			
Prob(Omnibus): 0.13		Jarque-Bera (JB): 3.209			
Skew: -0.719		Prob(JB): 0.201			
Kurtosis: 3.123		Cond. No. 4.22			

Figura 49 - Dados do modelo de regressão construído para RMAS Conc. Gravimétrico

O segundo modelo foi construído para prever a recuperação mássica do rejeito da flotação final e sua equação encontra-se representada em 243. Os teores de ouro e enxofre foram as variáveis independentes utilizadas e observa-se que quanto maiores os teores destas variáveis, menor é a recuperação mássica do rejeito da flotação. A Figura 50 apresenta os detalhes estatísticos do modelo em questão, cujo coeficiente de determinação é igual a 0.711. A análise dos resíduos indicou que estes são normalmente distribuídos e homocedásticos.

$$RMAS \text{ Rej. Flotação Final} = 0.8853 - 0.3099 * Au - 0.1067 * S. \quad (243)$$

OLS Regression Results							
Dep. Variable: RMAS Rej. Flotação		R-squared: 0.711					
Model: OLS		Adj. R-squared: 0.694					
Method: Least Squares		F-statistic: 41.86					
Time: 16:48:21		Log-Likelihood: 57.158					
No. Observations: 37		AIC: -108.3					
Df Residuals: 34		BIC: -103.5					
Df Model: 2							
Covariance Type: nonrobust							
	coef	std err	t	P> t	[0.025	0.975]	
const	0.8853	0.013	67.047	0	0.858	0.912	
Au	-0.3099	0.055	-5.686	0	-0.421	-0.199	
S	-0.1067	0.043	-2.479	0.018	-0.194	-0.019	
Omnibus: 10.649		Durbin-Watson: 2.148					
Prob(Omnibus): 0.005		Jarque-Bera (JB): 10.268					
Skew: -0.999		Prob(JB): 0.00589					
Kurtosis: 4.633		Cond. No. 7.43					

Figura 50 - Dados do modelo de regressão construído para RMAS Rej. Flotação Final

O modelo construído para prever a recuperação mássica do rejeito da lixiviação final encontra-se equacionado em 244. O teor de ouro foi a única variável independente estatisticamente significativa para prever esta variável dependente, sendo que quanto maior o teor de ouro, maior a recuperação mássica do rejeito da lixiviação.

$$RMAS \text{ Rej. Lixiviação Final} = 0.0635 + 0.2732 * Au. \quad (244)$$

A Figura 51 apresenta os dados do modelo ajustado. Analisando os resultados percebe-se que o coeficiente de determinação é igual a 0.608, indicando que 60.8% da variabilidade da variável dependente pode ser explicada através do modelo elaborado. A análise de resíduos mostrou que estes não são normalmente distribuídos apesar de serem homocedásticos, o que viola uma das suposições de um modelo estatisticamente válido.

OLS Regression Results						
Dep. Variable: RMAS Rej. Lixiviação		R-squared: 0.608				
Model: OLS		Adj. R-squared: 0.597				
Method: Least Squares		F-statistic: 54.33				
Time: 16:48:45		Log-Likelihood: 63.112				
No. Observations: 37		AIC: -122.2				
Df Residuals: 35		BIC: -119				
Df Model: 1						
Covariance Type: nonrobust						
	coef	std err	t	P> t	[0.025	0.975]
const	0.0635	0.01	6.186	0	0.043	0.084
Au	0.2732	0.037	7.371	0	0.198	0.348
Omnibus: 14.916		Durbin-Watson: 2.136				
Prob(Omnibus): 0.001		Jarque-Bera (JB): 16.971				
Skew: 1.268		Prob(JB): 0.000206				
Kurtosis: 5.139		Cond. No. 5.18				

Figura 51 - Dados do modelo de regressão construído para RMAS Rej. Lixiviação Final

O modelo elaborado para prever a recuperação mássica do gás encontra-se equacionado em 245 e os dados do modelo estão apresentados na Figura 52. Observa-se que quanto maiores os teores de ouro e enxofre no ROM maior é a recuperação mássica do gás. O coeficiente de determinação do modelo em questão é 0.772 e a análise de resíduos não indicou nenhuma violação das suposições de normalidade e homocedasticidade.

$$RMAS \text{ Gás Final} = 0.0104 + 0.0765 * Au + 0.0209 * S. \quad (245)$$

OLS Regression Results						
Dep. Variable: RMAS Gás		R-squared: 0.772				
Model: OLS		Adj. R-squared: 0.758				
Method: Least Squares		F-statistic: 57.48				
Time: 16:49:08		Log-Likelihood: 116.8				
No. Observations: 37		AIC: -227.6				
Df Residuals: 34		BIC: -222.8				
Df Model: 2						
Covariance Type: nonrobust						
	coef	std err	t	P> t	[0.025	0.975]
const	0.0104	0.003	3.951	0	0.005	0.016
Au	0.0765	0.011	7.033	0	0.054	0.099
S	0.0209	0.009	2.439	0.02	0.003	0.038
Omnibus: 5.044		Durbin-Watson: 2.239				
Prob(Omnibus): 0.08		Jarque-Bera (JB): 3.932				
Skew: 0.783		Prob(JB): 0.14				
Kurtosis: 3.309		Cond. No. 7.43				

Figura 52 - Dados do modelo de regressão construído para RMAS Gás Final

A equação 246 representa o modelo ajustado para prever a recuperação metalúrgica do concentrado gravimétrico. Segundo o modelo ajustado, quanto maior o teor de ouro no ROM menor é o valor da variável dependente em questão.

$$RMET\ Conc.\ Gravimétrico = 0.6938 - 0.3932 * Au. \quad (246)$$

Os dados do modelo ajustado encontram-se na Figura 53, na qual é possível ver um coeficiente de determinação igual a 0.314. A análise de resíduos, apresentada nas Figuras 57 e 58 não indicam violações das suposições de normalidade e homocedasticidade.

OLS Regression Results						
Dep. Variable: RMET Conc. Grav.		R-squared: 0.314				
Model: OLS		Adj. R-squared: 0.294				
Method: Least Squares		F-statistic: 16.03				
Time: 16:49:32		Log-Likelihood: 27.059				
No. Observations: 37		AIC: -50.12				
Df Residuals: 35		BIC: -46.9				
Df Model: 1						
Covariance Type: nonrobust						
	coef	std err	t	P> t	[0.025	0.975]
const	0.6938	0.027	25.498	0	0.639	0.749
Au	-0.3932	0.098	-4.003	0	-0.593	-0.194
Omnibus: 1.117		Durbin-Watson: 1.847				
Prob(Omnibus): 0.572		Jarque-Bera (JB): 0.943				
Skew: 0.137		Prob(JB): 0.624				
Kurtosis: 2.268		Cond. No. 5.18				

Figura 53 - Dados do modelo de regressão construído para RMET Conc. Gravimétrico

O modelo elaborado para prever a recuperação metalúrgica do rejeito da flotação final encontra-se equacionado em 247. Ao analisar os resultados apresentados na Figura 54 percebe-se que o valor-p relacionado ao teor de ouro é superior a 0.05, indicando que esta variável não é estatisticamente significativa para prever a variável em estudo. O valor-p relacionado às demais variáveis disponíveis (S e D1) também apresentou valores superiores a 0.05. Desta maneira é possível concluir que este modelo não é estatisticamente válido para prever a recuperação metalúrgica do rejeito da flotação e que nenhum outro pode ser encontrado para prever tal variável através da técnica de análise de regressão. A análise de resíduos mostra que estes não são normalmente distribuídos, fato que está relacionado ao valor do teste t para a variável ouro. Logo, o modelo apresentado é inválido para prever a variável em questão e não deve ser utilizado na prática.

$$RMET \text{ Rej. Flotação Final} = 0.0825 - 0.0633 * Au. \quad (247)$$

OLS Regression Results						
Dep. Variable: RMET Rej. Flotação		R-squared: 0.047				
Model: OLS		Adj. R-squared: 0.019				
Method: Least Squares		F-statistic: 1.711				
Time: 16:49:53		Log-Likelihood: 53.273				
No. Observations: 37		AIC: -102.5				
Df Residuals: 35		BIC: -99.32				
Df Model: 1						
Covariance Type: nonrobust						
	coef	std err	t	P> t	[0.025	0.975]
const	0.0825	0.013	6.154	0	0.055	0.11
Au	-0.0633	0.048	-1.308	0.199	-0.161	0.035
Omnibus: 9.467		Durbin-Watson: 1.861				
Prob(Omnibus): 0.009		Jarque-Bera (JB): 9.245				
Skew: 1.217		Prob(JB): 0.00983				
Kurtosis: 3.262		Cond. No. 5.18				

Figura 54 - Dados do modelo de regressão construído para RMET Rej. Flotação Final

A equação 248 representa o modelo encontrado para prever a recuperação metalúrgica do concentrado da lixiviação final. Quanto maior o teor de ouro no ROM, maior o valor da variável dependente em questão. A Figura 55 mostra os dados do modelo construído, o qual possui coeficiente de determinação igual a 0.48 e resíduos normalmente distribuídos e homocedásticos, de acordo com as Figuras 57 e 58.

$$RMET\ Conc.\ Lixiviação\ Final = 0.2093 + 0.4401 * Au. \quad (248)$$

OLS Regression Results						
Dep. Variable: RMET Conc. Lixiviação		R-squared: 0.48				
Model: OLS		Adj. R-squared: 0.465				
Method: Least Squares		F-statistic: 32.27				
Time: 16:50:12		Log-Likelihood: 35.84				
No. Observations: 37		AIC: -67.68				
Df Residuals: 35		BIC: -64.46				
Df Model: 1						
Covariance Type: nonrobust						
	coef	std err	t	P> t	[0.025	0.975]
const	0.2093	0.021	9.754	0	0.166	0.253
Au	0.4401	0.077	5.681	0	0.283	0.597
Omnibus: 1.585		Durbin-Watson: 1.995				
Prob(Omnibus): 0.453		Jarque-Bera (JB): 1.232				
Skew: -0.443		Prob(JB): 0.54				
Kurtosis: 2.878		Cond. No. 5.18				

Figura 55 - Dados do modelo de regressão construído para RMET Conc. Lixiviação Final

A Figura 56 apresenta os dados do modelo elaborado para estimar a recuperação metalúrgica do rejeito da lixiviação final, cujo coeficiente de determinação é de 0.099. Ao analisar o valor do teste t relacionado ao teor de enxofre no ROM, percebe-se que esta variável não é estatisticamente significativa para predizer a variável em questão. Apesar disso, os resíduos apresentaram distribuição normal e variância constante. No entanto, o modelo apresentado na equação 249 não deve ser utilizado na prática, uma vez que o coeficiente do teor de enxofre é estatisticamente igual a zero.

$$RMET \text{ Rej. Lixiviação Final} = 0.0139 + 0.0139 * S. \quad (249)$$

OLS Regression Results							
Dep. Variable: RMET Rej. Lixiviação				R-squared: 0.099			
Model: OLS				Adj. R-squared: 0.073			
Method: Least Squares				F-statistic: 3.831			
Time: 16:50:30				Log-Likelihood: 115.42			
No. Observations: 37				AIC: -226.8			
Df Residuals: 35				BIC: -223.6			
Df Model: 1							
Covariance Type: nonrobust							
	coef	std err	t	P> t	[0.025	0.975]	
const	0.0139	0.003	5.412	0	0.009	0.019	
S	0.0139	0.007	1.957	0.058	-0.001	0.028	
Omnibus: 0.785				Durbin-Watson: 1.963			
Prob(Omnibus): 0.676				Jarque-Bera (JB): 0.819			
Skew: 0.309				Prob(JB): 0.664			
Kurtosis: 2.613				Cond. No. 4.22			

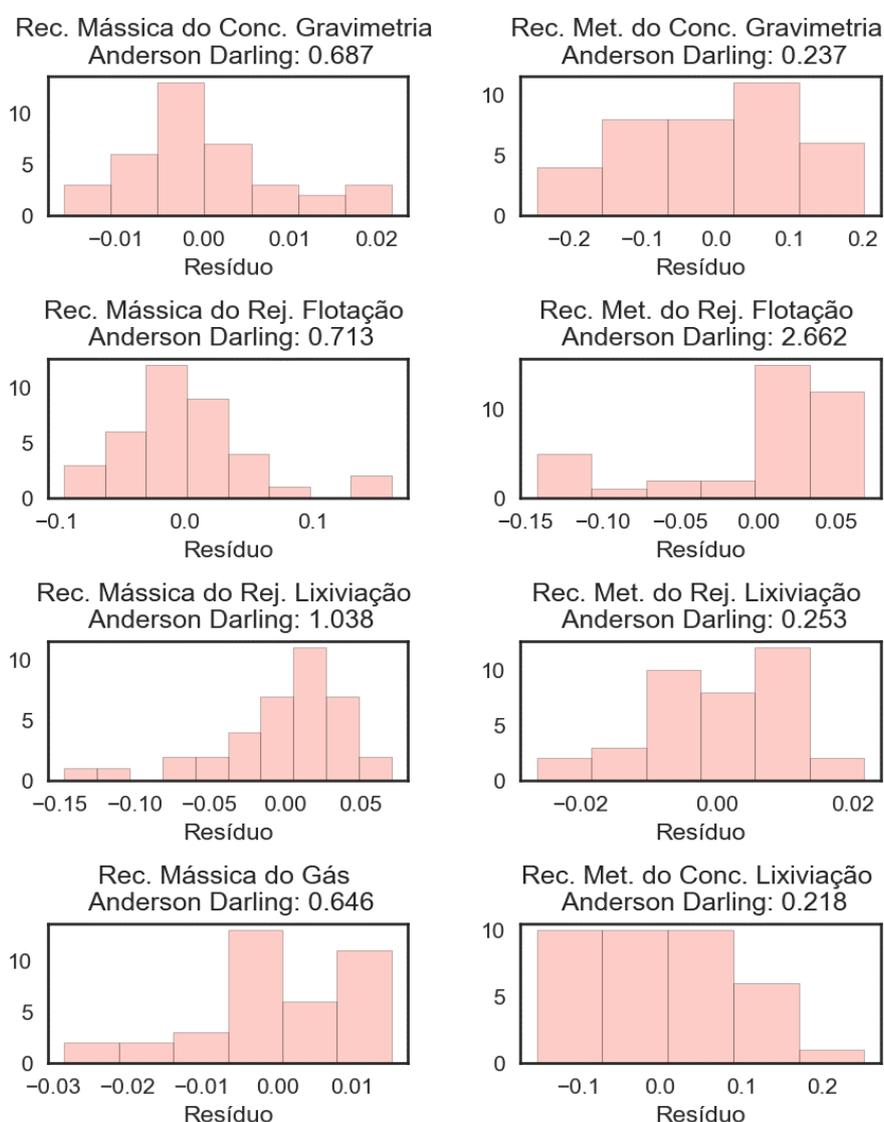
Figura 56 - Dados do modelo de regressão construído para RMET Rej. Lixiviação Final

Após a análise estatística dos modelos construídos percebe-se que 3 dos 8 modelos elaborados são inválidos para serem utilizados na prática. Desta forma, os resultados referentes a técnica de análise de regressão para ajuste dos modelos geometalúrgicos das minas de ouro em estudo não serão mais apresentados nem comparados com as demais técnicas sob avaliação.

Tabela 31 - Correlações obtidas via Análise de Regressão – Minas de Ouro

Variáveis	Correlação
RMAS Conc. Gravimétrico	0.593
RMAS Rej. Flotação Final	0.843
RMAS Rej. Lixiviação Final	0.780
RMAS Gás Final	0.878
RMET Conc. Gravimétrico	0.560
RMET Rej. Flotação Final	0.216
RMET Conc. Lixiviação Final	0.693
RMET Rej. Lixiviação Final	0.314

Histograma dos resíduos - Valor Crítico AD (5%): 0.722

**Figura 57 - Histograma dos resíduos – Minas de Ouro**

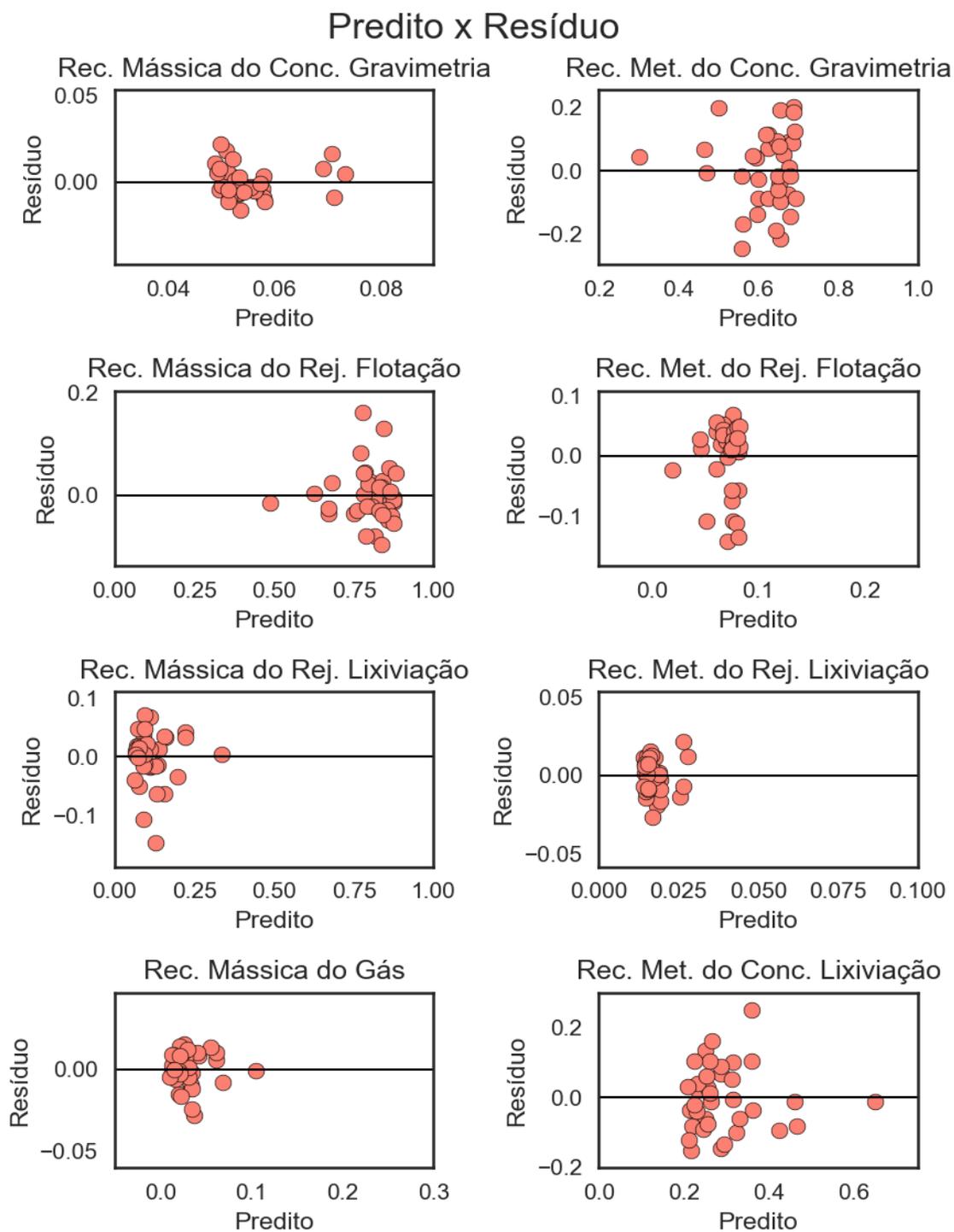


Figura 58 - Gráficos dos valores preditos x resíduos – Minas de Ouro

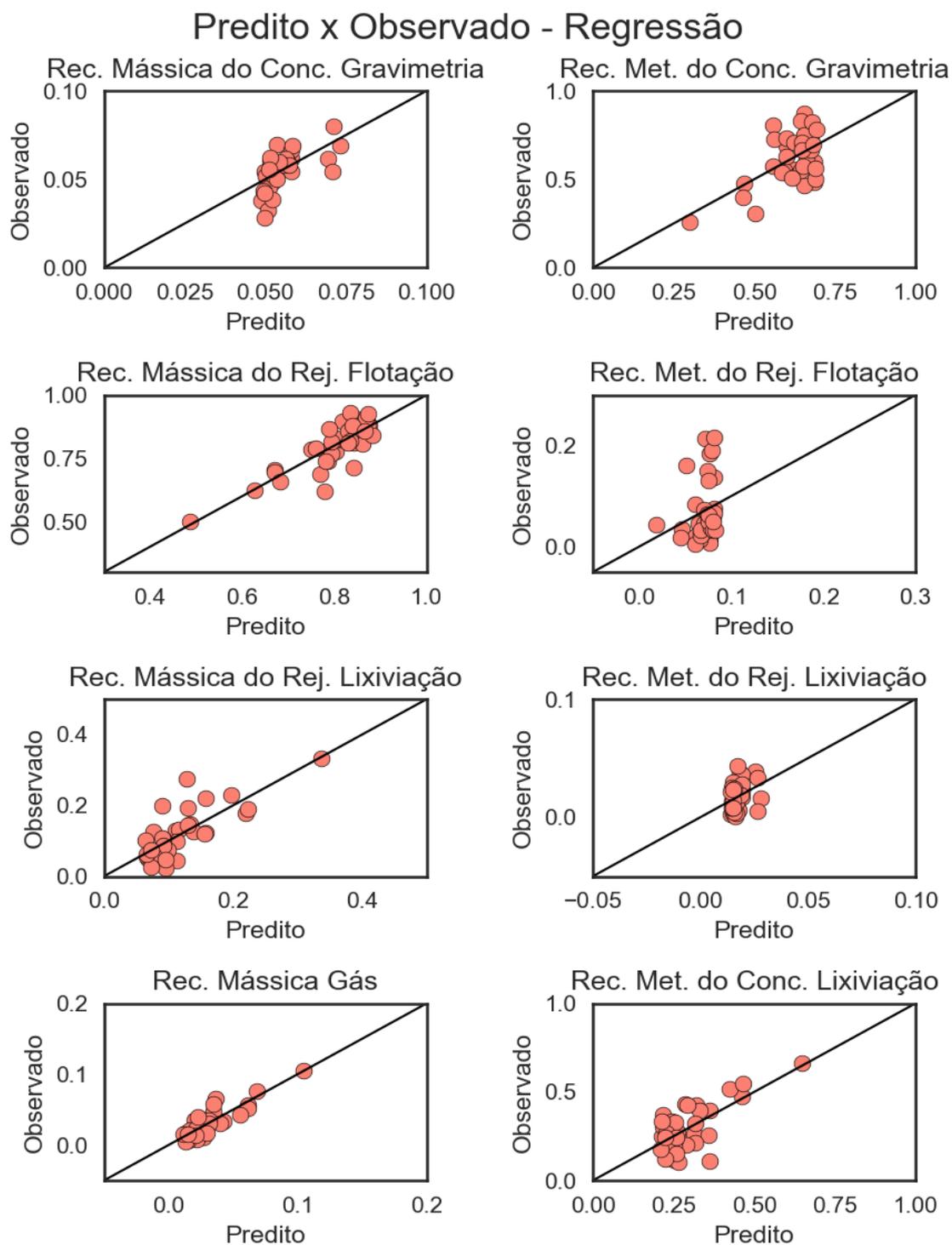


Figura 59 - Gráfico dos valores preditos x observados – Análise de regressão - Minas de Ouro

6.2.3.2. *Random forest*

Para aplicar a técnica de *random forest* na previsão geometalúrgica das variáveis relacionadas às minas de ouro, utilizou-se o mesmo número de árvores adotado no estudo realizado para a mina de zinco estudada neste trabalho, ou seja, 1000. Da mesma maneira, a técnica de *out-of-bag* foi utilizada para determinar o número mínimo de amostras por folha, sendo que valores entre 2 e 15 foram testados e escolheu-se aquele que retornou o maior coeficiente de correlação. A Tabela 32 apresenta o número mínimo de amostras por folha escolhido para cada variável.

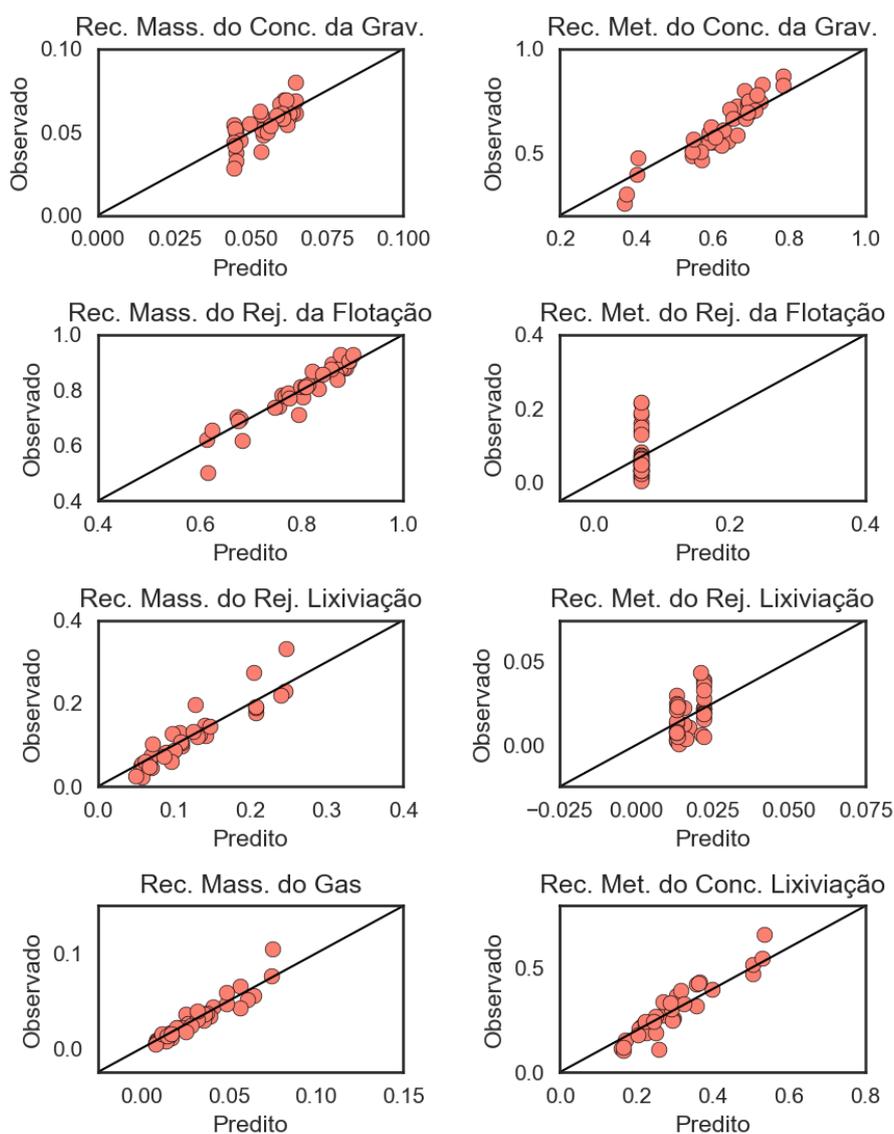
Tabela 32 - Número mínimo de amostras por folha – Minas de Ouro

Variáveis	Número mínimo de amostras por folha
RMAS Conc. Gravimétrico	5
RMAS Rej. Flotação Final	2
RMAS Rej. Lixiviação Final	2
RMAS Gás Final	2
RMET Conc. Gravimétrico	2
RMET Rej. Flotação Final	15
RMET Conc. Lixiviação Final	2
RMET Rej. Lixiviação Final	11

Os valores dos coeficientes de correlação obtidos quando a técnica de *random forest* foi utilizada encontram-se apresentados na Tabela 33 e os gráficos dos valores preditos versus os valores reais encontram-se na Figura 60. Observam-se valores altos de correlação para a maioria das variáveis em estudo, mas é necessário avaliar o erro de predição desta técnica através da validação cruzada *k-fold*, a fim de confirmar estes valores para dados não utilizados no ajuste dos modelos. Esta validação será apresentada nas seções seguintes.

Tabela 33 - Coeficiente de correlação - Random Forest – Minas de Ouro

Variáveis	Correlação
RMAS Conc. Gravimétrico	0.774
RMAS Rej. Flotação Final	0.944
RMAS Rej. Lixiviação Final	0.927
RMAS Gás Final	0.946
RMET Conc. Gravimétrico	0.929
RMET Rej. Flotação Final	0.226
RMET Conc. Lixiviação Final	0.933
RMET Rej. Lixiviação Final	0.488

Predito x Observado - Random Forest**Figura 60 - Valores preditos versus observados - Random Forest – Minas de Ouro**

6.2.3.3. Support vector regression

Da mesma forma como foi conduzida a elaboração dos modelos via *support vector regression* para a mina de zinco em estudo, conduziu-se o estudo para as minas de ouro em questão: primeiramente foi considerada a utilização de um kernel linear e em seguida foi utilizado um kernel RBF. Em ambas partes, foi adotado um valor de épsilon (ϵ) igual a 0.0001. Os resultados encontram-se detalhados a seguir.

6.2.3.3.1. Support vector regression linear

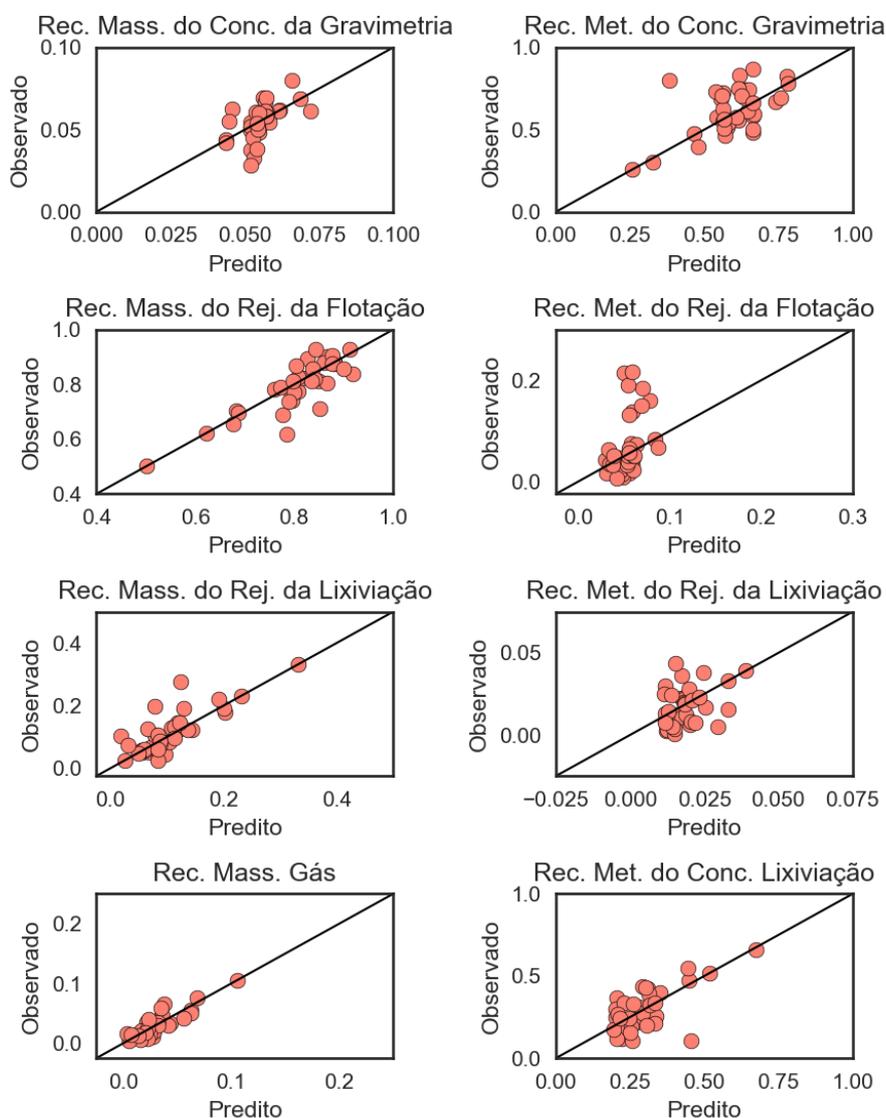
Para determinar o melhor valor do parâmetro C para o SVR linear, foi realizada uma busca em grid com valor de C variando entre 0.001 e 100. O banco de dados foi dividido em banco de teste e de treinamento e, para cada variável dependente, foi escolhido o valor de C que retornou o menor erro absoluto médio (MAE) no banco de teste, garantindo assim que não existisse *overfitting* e que os resultados fossem adequados. A Tabela 34 apresenta o valor escolhido para cada uma das 8 variáveis dependentes em estudo. Os coeficientes de correlação e os gráficos com os valores preditos versus valores reais encontram-se apresentados na Tabela 35 e na Figura 61, respectivamente.

Tabela 34 - Valores escolhidos do parâmetro C para o SVR linear – Minas de Ouro

Variáveis	C
RMAS Conc. Gravimétrico	64
RMAS Rej. Flotação Final	10
RMAS Rej. Lixiviação Final	87
RMAS Gás Final	0.05
RMET Conc. Gravimétrico	68
RMET Rej. Flotação Final	0.001
RMET Conc. Lixiviação Final	0.1
RMET Rej. Lixiviação Final	29

Tabela 35 - Coeficientes de correlação - SVR linear – Minas de Ouro

Variáveis	Correlação
RMAS Conc. Gravimétrico	0.552
RMAS Rej. Flotação Final	0.849
RMAS Rej. Lixiviação Final	0.809
RMAS Gás Final	0.881
RMET Conc. Gravimétrico	0.575
RMET Rej. Flotação Final	0.288
RMET Conc. Lixiviação Final	0.646
RMET Rej. Lixiviação Final	0.350

Predito x Observado - Support Vector Regression Linear**Figura 61 - Valores preditos versus observados - SVR linear – Minas de Ouro**

6.2.3.3.2. Support vector regression não linear

Para ajustar o SVR não linear aos dados, utilizou-se a mesma abordagem adotada para a mina de zinco aqui apresentada: dividiu-se o banco de dados em banco de teste e de treinamento e uma busca em grid foi realizada visando determinar os melhores valores dos parâmetros C e γ . Foram avaliados valores de γ entre 0.001 e 100 e valores de C entre 0.001 e 100. Os valores escolhidos foram aqueles que retornaram o menor valor de MAE no banco de teste, os quais encontram-se apresentados na Tabela 36. Diversos kernels foram testados e o que apresentou os melhores resultados foi o RBF, logo, este foi o escolhido.

Tabela 36 - Valores escolhidos dos parâmetros C e γ - SVR não linear – Minas de Ouro

Variáveis	C	Gamma
RMAS Conc. Gravimétrico	4	0.8
RMAS Rej. Flotação Final	99	1
RMAS Rej. Lixiviação Final	97	1
RMAS Gás Final	18	1
RMET Conc. Gravimétrico	82	3
RMET Rej. Flotação Final	38	4
RMET Conc. Lixiviação Final	48	2
RMET Rej. Lixiviação Final	95	0.04

A Tabela 37 apresenta a correlação entre os valores preditos e observados utilizando os parâmetros supracitados e a Figura 62 ilustra tais correlações.

Tabela 37 - Correlação entre valores preditos e observados - SVR não linear - Minas de Ouro

Variáveis	Correlação
RMA Conc. Gravimétrico	0.797
RMA Rej. Flotação Final	0.893
RMA Rej. Lixiviação Final	0.860
RMA Gás Final	0.929
RME Conc. Gravimétrico	0.846
RME Rej. Flotação Final	0.655
RME Conc. Lixiviação Final	0.859
RME Rej. Lixiviação Final	0.567

Predito x Observado - Support Vector Regression Não Linear

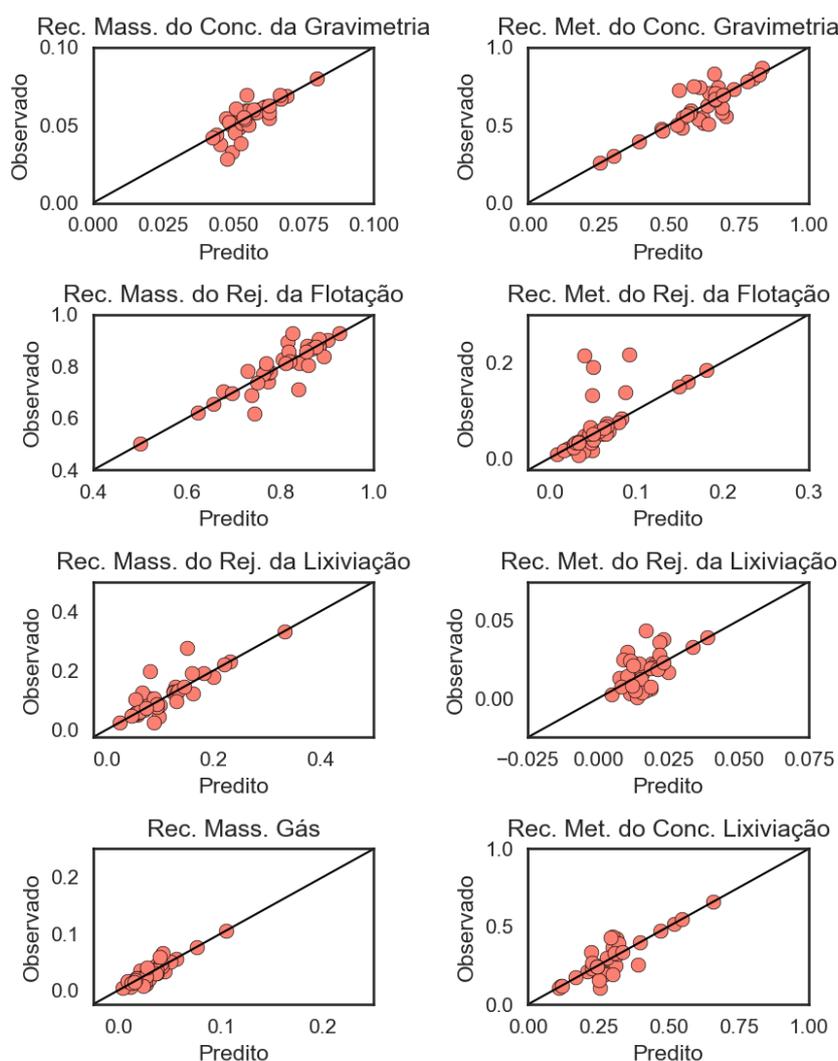


Figura 62 - Valores preditos versus observados - SVR não linear - Minas de Ouro

6.2.3.4. Redes Neurais Artificiais

Para iniciar a elaboração da rede neural utilizada para prever as variáveis geometalúrgicas das minas de ouro em estudo, foram testados os algoritmos de otimização do gradiente descendente e o Adam, sendo que o último apresentou os melhores resultados e foi o escolhido para prosseguir no estudo. Em seguida, foi necessário definir a taxa de aprendizado η que seria utilizada no algoritmo. Para o estudo em questão, foram testadas taxas entre 0.00001 e 0.5. Os melhores resultados de MAE e correlação foram obtidos quando a taxa de 0.00003 foi adotada. Outra definição necessária foi em relação ao número de camadas ocultas da rede e o número de nós ocultos em cada camada. Tendo em vista que o número de dados em estudo é reduzido, optou-se por avaliar a performance de redes com no máximo três camadas ocultas e com no máximo seis nós ocultos em cada camada, com o objetivo de evitar um super ajuste do modelo (*overfitting*). Avaliados os resultados de todas as combinações possíveis de camadas e nós ocultos, chegou-se à conclusão que uma rede com duas camadas ocultas, com 4 nós ocultos na primeira camada e 5 nós ocultos na segunda apresentava as melhores respostas. Por fim, foi necessário definir quais as funções de ativação devem ser colocadas em cada camada oculta da rede e na camada de *output*. Foram testadas as funções linear, sigmoide, relu, tangente hiperbólica e softplus. Observou-se que melhores resultados de MAE e correlação foram obtidos ao adotar a função linear na primeira e na segunda camada oculta, além de utilizá-la na camada de *output*.

A rede construída para prever as variáveis em estudo pode ser ilustrada conforme a estrutura apresentada na Figura 63.

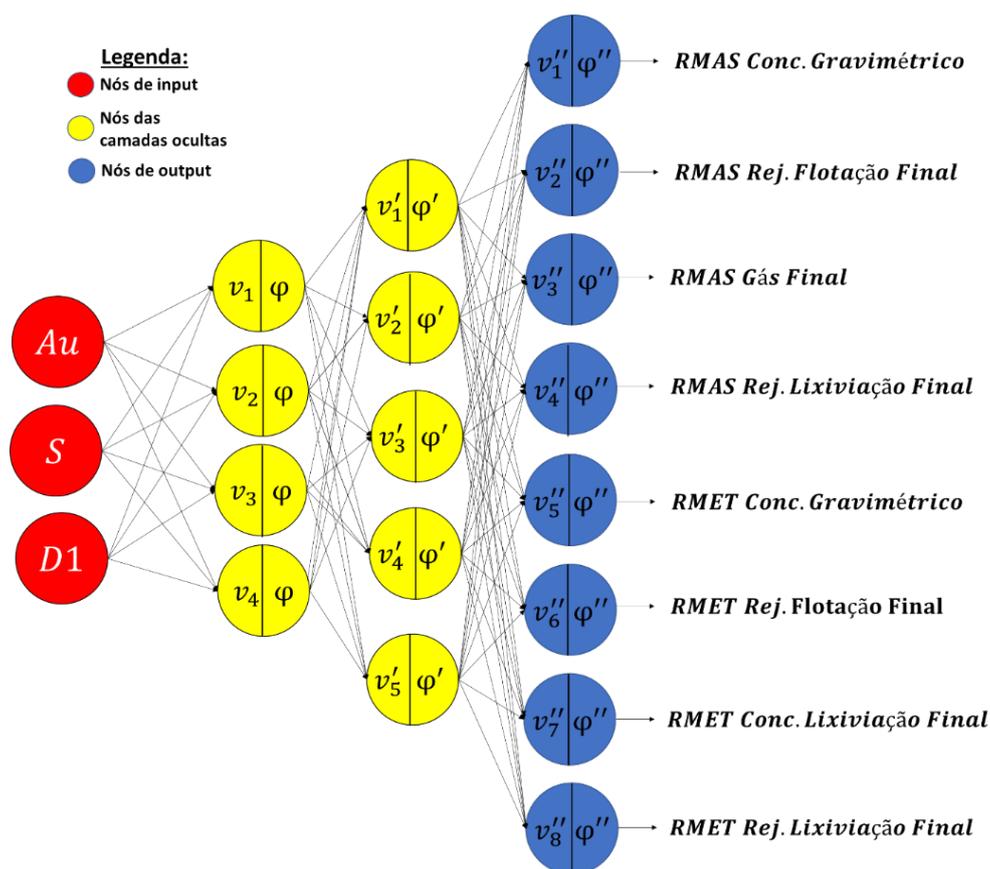


Figura 63 - Configuração da rede neural construída para as minas de ouro

A Tabela 38 apresenta os resultados obtidos quando a rede neural com os parâmetros anteriormente descritos foi ajustada para as 37 amostras do banco de dados e a Figura 64 ilustra tais correlações.

Tabela 38 – Correlações obtidas utilizando Redes Neurais – Minas de Ouro

Variáveis	Correlação
RMAS Conc. Gravimétrico	0.540
RMAS Rej. Flotação Final	0.849
RMAS Rej. Lixiviação Final	0.810
RMAS Gás Final	0.883
RMET Conc. Gravimétrico	0.589
RMET Rej. Flotação Final	0.121
RMET Conc. Lixiviação Final	0.678
RMET Rej. Lixiviação Final	0.307

Predito x Observado - Redes Neurais

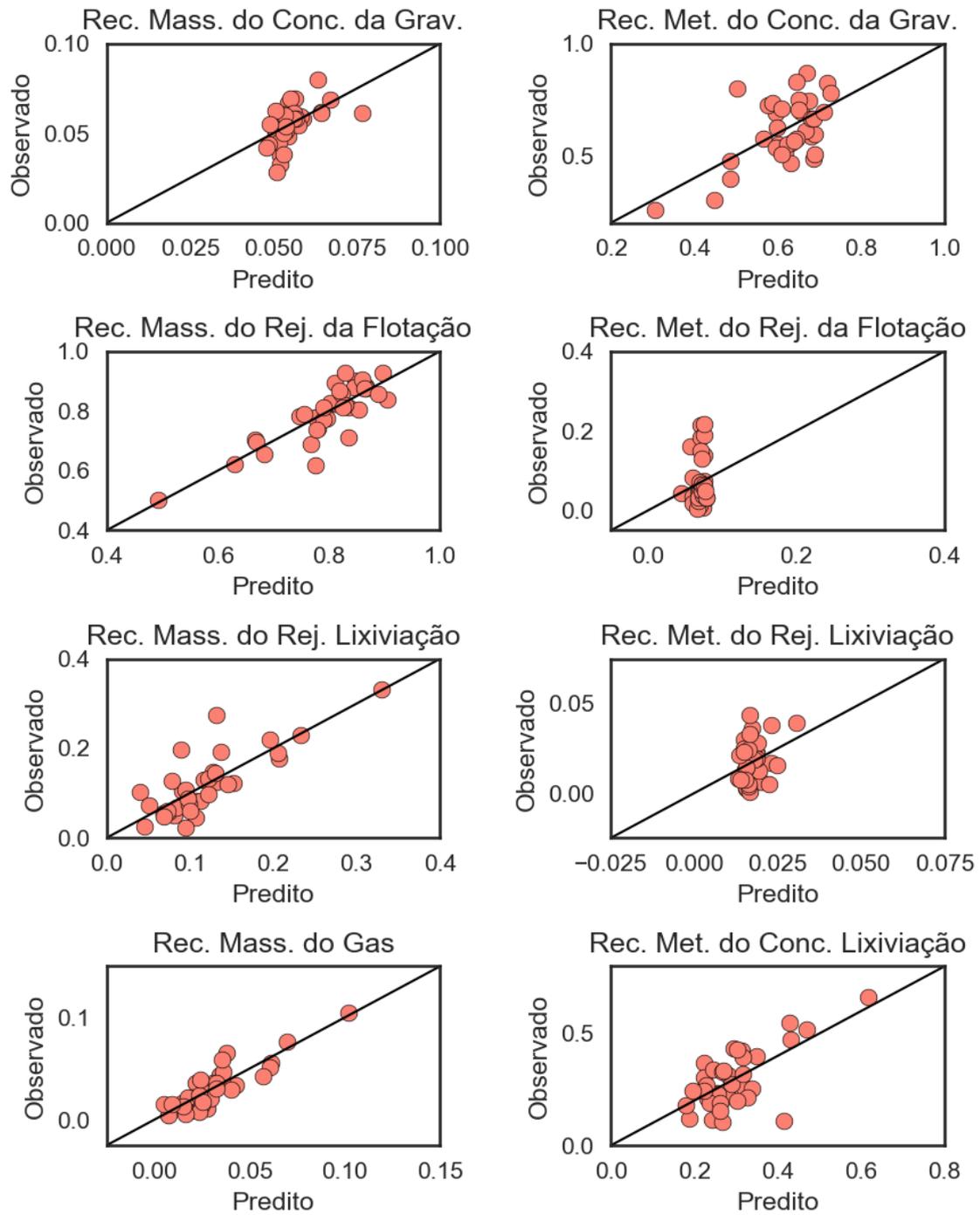


Figura 64 - Valores preditos versus observados - Redes Neurais - Minas de Ouro

6.2.4. Validação cruzada k-fold

Com o objetivo de avaliar o erro de predição associado a cada um dos modelos elaborados, optou-se por realizar a validação cruzada *5-fold*, da mesma maneira como foi feito para o estudo realizado na mina de zinco apresentada neste trabalho. Os resultados obtidos encontram-se na Tabela 39. Para as variáveis recuperação mássica do gás final e recuperação mássica do rejeito da lixiviação final, foram omitidos os resultados no banco de teste referentes às amostras que sofreram imputação da variável calcinação, a qual é utilizada para o cálculo de ambas variáveis.

Tabela 39 - Avaliação dos resultados da validação cruzada 5-fold – Minas de Ouro

Banco de teste					
Métrica	Variáveis	RF	SVR Linear	SVR Não Linear	Redes Neurais
MAE	RMA Conc. Gravimétrico	0.0068 ± 0.0002	0.0085 ± 0.0004	0.0073 ± 0.0022	0.0067 ± 0.001
	RMA Rej. Flotação Final	0.044 ± 0.0045	0.0397 ± 0.0039	0.0461 ± 0.005	0.0398 ± 0.0061
	RMA Rej. Lixiviação Final	0.0279 ± 0.0053	0.0332 ± 0.0056	0.034 ± 0.0038	0.0286 ± 0.007
	RMA Gás Final	0.0089 ± 0.0027	0.0117 ± 0.0022	0.0094 ± 0.0021	0.0118 ± 0.0034
	RMET Conc. Gravimétrico	0.1015 ± 0.0158	0.1239 ± 0.0194	0.1244 ± 0.0189	0.096 ± 0.0154
	RMET Rej. Flotação Final	0.0486 ± 0.0068	0.0452 ± 0.0117	0.0473 ± 0.0144	0.0501 ± 0.0061
	RMET Conc. Lixiviação Final	0.0807 ± 0.0171	0.0862 ± 0.008	0.0869 ± 0.0091	0.0793 ± 0.0095
	RMET Rej. Lixiviação Final	0.0094 ± 0.0009	0.0112 ± 0.0008	0.0102 ± 0.0012	0.0094 ± 0.0005
RMSE	RMA Conc. Gravimétrico	0.0084 ± 0.0004	0.0111 ± 0.0004	0.0107 ± 0.0037	0.0086 ± 0.0012
	RMA Rej. Flotação Final	0.0611 ± 0.0044	0.0554 ± 0.0056	0.0596 ± 0.0067	0.0503 ± 0.0073
	RMA Rej. Lixiviação Final	0.0367 ± 0.0089	0.0438 ± 0.0104	0.0489 ± 0.0066	0.0359 ± 0.0115
	RMA Gás Final	0.0098 ± 0.0025	0.0132 ± 0.0021	0.0107 ± 0.0024	0.0128 ± 0.0032
	RMET Conc. Gravimétrico	0.1168 ± 0.0143	0.1515 ± 0.0233	0.1715 ± 0.0296	0.1127 ± 0.0194
	RMET Rej. Flotação Final	0.0592 ± 0.0097	0.0601 ± 0.0165	0.0614 ± 0.0192	0.0602 ± 0.0092
	RMET Conc. Lixiviação Final	0.0964 ± 0.0197	0.1021 ± 0.0148	0.1146 ± 0.0134	0.095 ± 0.0165
	RMET Rej. Lixiviação Final	0.0112 ± 0.0007	0.0135 ± 0.001	0.0123 ± 0.0017	0.0111 ± 0.0006
Correlação	RMA Conc. Gravimétrico	0.6176 ± 0.0891	0.4788 ± 0.1866	0.6639 ± 0.1076	0.5528 ± 0.1183
	RMA Rej. Flotação Final	0.8128 ± 0.0509	0.7754 ± 0.0813	0.8102 ± 0.081	0.8039 ± 0.0707
	RMA Rej. Lixiviação Final	0.6089 ± 0.2258	0.5282 ± 0.1961	0.4923 ± 0.2066	0.7254 ± 0.0679
	RMA Gás Final	0.8866 ± 0.0393	0.6752 ± 0.1137	0.7834 ± 0.075	0.8073 ± 0.0545
	RMET Conc. Gravimétrico	0.4126 ± 0.1816	0.3117 ± 0.1417	0.1554 ± 0.2025	0.5372 ± 0.1072
	RMET Rej. Flotação Final	0 ± 0	0.1709 ± 0.1024	0.4619 ± 0.175	0.1176 ± 0.1471
	RMET Conc. Lixiviação Final	0.5276 ± 0.1667	0.6268 ± 0.1102	0.5216 ± 0.1849	0.6495 ± 0.1144
	RMET Rej. Lixiviação Final	0.4127 ± 0.1265	0.1494 ± 0.0838	0.1817 ± 0.2	0.1119 ± 0.208

Ao analisar o MAE, percebe-se que 4 das 8 variáveis estudadas apresentaram menor valor desta estatística quando a técnica de redes neurais foi utilizada e ao analisar o RMSE, vê-se que 5 das 8 variáveis apresentaram melhores resultados ao utilizar a

técnica de redes neurais. Em termos de correlação, observou-se que a recuperação mássica do rejeito da lixiviação final, a recuperação metalúrgica do concentrado gravimétrico e a recuperação metalúrgica do concentrado da lixiviação final apresentaram melhores resultados com a técnica de redes neurais, sendo que a recuperação mássica do concentrado gravimétrico e a recuperação metalúrgica do rejeito da flotação final apresentaram correlações mais altas quando o SVR não linear foi utilizado e a recuperação mássica do rejeito da flotação final, a recuperação mássica do gás final e a recuperação metalúrgica do rejeito da lixiviação final tiveram melhores correlações quando a técnica de *random forest* foi aplicada. De maneira geral, a técnica de redes neurais mostrou melhores resultados que as demais, mostrando-se mais promissora para modelar as variáveis geometalúrgicas em questão.

Analisando mais profundamente os resultados obtidos via redes neurais, percebe-se que as variáveis RMAJ Rej. Flotação Final, RMAJ Rej. Lixiviação Final e RMAJ Gás Final representam juntas, em média, 93% da massa considerando o material da Mina A e 95% considerando o material da Mina B. Logo, é possível concluir que as correlações superiores a 0.72 para essas variáveis indicam modelos muito bons para a modelagem do balanço mássico. Considerando que, em média, 90% do metal da Mina A e que 94% do metal da Mina B têm como destino o concentrado gravimétrico e o concentrado da lixiviação, é possível considerar que as variáveis com maior impacto metalúrgico tiveram modelos com bons resultados. Logo, as variáveis RMEJ Rej. Lixiviação e RMEJ Rej. Flotação, que representaram modelos ruins, são as que possuem menor saída de metal.

6.2.5. Análise do fechamento mássico e metalúrgico

Da mesma forma como o fechamento das estimativas de zinco foi avaliado, avaliou-se os resultados de fechamento mássico e metalúrgico obtidos para as predições realizadas para as minas de ouro em estudo, os quais encontram-se apresentados na Tabela 40.

Tabela 40 - Fechamento mássico e metalúrgico - Minas de Ouro

Banco	Fechamento	RF	SVR Linear	SVR Não-Linear	NN
Treinamento	Mássico	94%	88%	78%	97%
	Metalúrgico	24%	11%	48%	95%
Teste	Mássico	94%	86%	68%	97%
	Metalúrgico	11%	8%	41%	84%

Observa-se que a técnica de redes neurais apresentou resultados superiores às demais. Em média, 97% das amostras do banco de teste apresentaram erros de fechamento mássico inferiores a 1% e 84% das amostras apresentaram erros de fechamento metalúrgico inferiores a 1%, para mais ou para menos.

6.2.6. Reconciliação

A etapa de reconciliação dos resultados gerados pela rede neural com as respostas observadas na usina de beneficiamento de ouro não pode ser realizada, uma vez que a usina em questão não tem conhecimento do tempo de residência de cada etapa do processo. Desta forma, a avaliação da eficácia do modelo elaborado na previsão de respostas da usina não pode ser comprovada, ficando como um passo futuro deste estudo.

6.2.7. Discussão dos resultados e sugestão de trabalhos futuros

Os resultados apresentados na seção 6.2 mostraram, que ao utilizarmos nas minas de ouro a mesma metodologia proposta para a mina de Vazante, os resultados obtidos via redes neurais artificiais mostraram-se superiores aos das demais técnicas avaliadas, em termos de MAE, RMSE e correlação para a maioria das variáveis dependentes modeladas. Além disso, o fechamento mássico e metalúrgico foi melhor reproduzido quando a técnica de redes neurais foi aplicada, ainda que o fechamento obtido tenha sido inferior ao observado na mina de zinco. Acredita-se, que a dificuldade de um fechamento perfeito esteja relacionada ao baixo número amostral utilizado na elaboração da rede neural para as minas de ouro em estudo.

Devido às dificuldades operacionais da usina de beneficiamento de ouro, não foi possível validar a metodologia com base em resultados industriais, sendo a realização desta etapa uma sugestão para trabalhos futuros.

7. CONCLUSÕES

7.1. CONCLUSÕES INICIAIS

O estudo realizado comprovou a possibilidade de se prever simultaneamente as quantidades e qualidades dos produtos e rejeitos oriundos de plantas de beneficiamento, com base nas características físico-químicas do minério alimentado, atendendo assim à meta estabelecida na seção 1.4.

Os objetivos específicos traçados na seção 1.5 também foram atendidos, tendo em vista que as técnicas de análise de regressão, *random forest*, *support vector regression* e redes neurais artificiais foram estudadas e suas métricas (MAE, RMSE e correlação) avaliadas em busca dos melhores resultados na previsão das variáveis geometalúrgicas relacionadas às minas de zinco e ouro aqui apresentadas. Observou-se que a técnica de redes neurais gerava os melhores resultados em termos de fechamento mássico e metalúrgico e também em relação às demais métricas para a maioria das variáveis estudadas. Tendo em vista sua superioridade, tal técnica foi aplicada em pilhas de minério de zinco cujas variáveis de *input* da rede eram conhecidas e as previsões realizadas pela rede foram reconciliadas com as respostas observadas na usina de beneficiamento.

A reconciliação permitiu o cálculo de um fator de escala para tornar comparáveis os resultados dos testes de bancada (que utilizam de um circuito de flotação aberto) e da planta (que possui um circuito de flotação fechado), de modo que ao aplicar tal fator nos resultados gerados pelas redes neurais observou-se erros inferiores a 2% após a retirada de um outlier.

Conclui-se, portanto, que a metodologia proposta é adequada para prever a recuperação mássica e metalúrgica de produtos e rejeitos através das características físico-químicas do minério.

7.2. BENEFÍCIOS RELACIONADOS À UTILIZAÇÃO DA METODOLOGIA PROPOSTA

Ao utilizar a metodologia proposta, as mineradoras terão conhecimento das respostas metalúrgicas do minério antes mesmo que este alimente a planta, levando a um planejamento de lavra mais eficiente. O conhecimento da recuperação mássica e metalúrgica dos rejeitos permitirá, além de um conhecimento prévio dos impactos ambientais da mineração, enxergar oportunidades de ganho anteriormente desapercibidas. Alguns dos benefícios encontram-se citados a seguir:

- O teor de corte poderá ser determinado não considerando somente o teor do metal de interesse no ROM, mas sim visualizando a recuperação mássica e metalúrgica de um dado bloco, levando a um melhor aproveitamento econômico do recurso;
- Ao determinar uma região da mina com baixa recuperação metalúrgica no produto de interesse, é possível adaptar a usina (antes mesmo da alimentação do minério) para receber tal material considerando um fluxo de beneficiamento distinto do habitual, adicionando, por exemplo, uma etapa *cleaner* na flotação, mudando a dosagem e os reagentes utilizados, aumentando o tempo de moagem para atingir uma maior liberação do mineral-minério, entre muitas outras opções existentes com o intuito de elevar a recuperação metalúrgica para níveis desejáveis;
- Caso a barragem de rejeitos esteja com capacidade próxima ao limite máximo, deve-se optar pela lavra de regiões cuja massa de rejeitos gerada seja menor, o que será indicado pela recuperação mássica do rejeito predita através da metodologia aqui apresentada. Assim, seria possível a lavra de um maior número de blocos sem que o volume de rejeitos seja excedido, gerando ganho financeiro devido à lavra e beneficiamento de um maior volume de material, sem que a barragem seja comprometida. Essa solução poderia ser fator decisivo entre a manutenção da operação ou paralização temporária das atividades, até que uma nova barragem fosse licenciada e entrasse em operação;

- Outra possibilidade de ganho ao conhecer as características do rejeito está ligada ao teor do metal de interesse destinado à barragem final. Sabendo de antemão que determinado *stope* gerará uma grande quantidade do metal de interesse no rejeito é possível optar por estocá-lo em pilhas ao invés de enviá-lo para barragem, o que permitiria um novo processamento do material após adaptações no processo de beneficiamento (adição de mais reagentes, mudar o processo de deslamagem, adição de etapas *scavenger* na flotação, etc). Ainda que tais adaptações levem tempo para serem implementadas é melhor considerá-las como uma oportunidade de ganho financeiro do que enviar o material diretamente para a barragem, na qual dificilmente o metal será recuperado.

7.3. TRABALHOS FUTUROS

Para trabalhos futuros relacionados ao estudo realizado na mina de Vazante sugere-se estimar os valores das variáveis geometalúrgicas no modelo de blocos, permitindo assim uma visualização 3D das características metalúrgicas do minério. Para isso, será necessário estimar, bloco a bloco, os teores de Zn (%) e Pb (%) no ROM e aplicar a rede com fator de escala em cada um deles. Outra opção é simular os teores de zinco e chumbo em cada bloco, permitindo que as estimativas das variáveis metalúrgicas sejam aplicadas em todos os cenários simulados, o que levaria a um melhor conhecimento da variabilidade esperada das variáveis modeladas.

Em relação ao estudo relacionado nas minas de ouro, a sugestão de trabalhos futuros está relacionada à reconciliação dos modelos com as respostas observadas na usina. Além disso, sugere-se incorporar novos dados e/ou acrescentar outras variáveis independentes (de *input*) ao modelo, visando aumentar a qualidade preditiva do mesmo.

7.4. CONSIDERAÇÕES FINAIS

A metodologia proposta nesta tese é adequada para estimar variáveis geometalúrgicas, aditivas e não aditivas, relacionadas aos produtos e aos rejeitos gerados pelo beneficiamento mineral, possibilitando ainda um bom fechamento mássico e

metalúrgico das previsões sem que haja necessidade de realizar um pós-processamento dos resultados. Ao utilizar a metodologia aqui descrita, torna-se possível aumentar os lucros da operação e reduzir seus impactos ambientais, sem que dados adicionais aos já existentes nas mineradoras sejam coletados e sem que mais tempo seja gasto na elaboração de um modelo que considere fatores econômicos e ambientais simultaneamente.

8. REFERÊNCIAS BIBLIOGRÁFICAS

AMBRÓSIO, P. E. **Redes neurais artificiais no apoio ao diagnóstico diferencial de lesões intersticiais pulmonares**. 2002. Dissertação (Mestrado em Física Aplicada à Medicina e Biologia) - Programa de Pós-Graduação em Física Aplicada à Medicina e Biologia, Universidade de São Paulo, Ribeirão Preto, 2002.

ARMSTRONG, M. **Basic linear geostatistics**. Berlin: Springer-Verlag, 1998.

BECHIR, J. L. C., BAPTISTA, J. R., SOUZA, A. D., MARTINS, E. Determination of maximum zinc recovery of Vazante mine ore by flotation process. **REM, International Engineering Journal**, v. 72, n. 2, p. 315-320, Apr. 2019. DOI: 10.1590/0370-44672018720069.

BISHOP, C. M. **Neural networks for pattern recognition**. Oxford: Clarendon Press, 1995.

BOISVERT, J. B., ROSSI, M. E., EHRIG, K., DEUTSCH, C. V. Geometallurgical modeling at Olympic Dam mine, South Australia. **Mathematical Geosciences**, v. 45, p. 901-925, July 2013. DOI 10.1007/s11004-013-9462-5.

BULLED, D., MCINNES, C. Flotation plant design and production planning through geometallurgical modelling. *In*: CENTENARY OF FLOTATION SYMPOSIUM, 2015, Brisbane. **Proceedings...**, Brisbane, QLD: Australasian Institute of Mining and Metallurgy, 2015, p. 809-814.

BYE, A. R. Case studies demonstrating value from geometallurgy initiatives. *In*: THE FIRST AUSIMM INTERNATIONAL GEOMETALLURGY CONFERENCE, 2011,

Brisbane. **Proceedings...**, Brisbane, QLD: Australasian Institute of Mining and Metallurgy, 2011, p. 9-30.

CARRASCO, P., Chilès, J. P., Séguret, S. Additivity, metallurgical recovery and grade. *In: VIII INTERNATIONAL GEOSTATISTICS CONGRESS, GEOSTATS 2008, 2008, Santiago. Proceedings...*, Santiago: 2008, p. 237-246.

CORNAH, A. J. Assessments of the impact of non-additivity in the estimation of iron ore attributes. *In: IRON ORE CONFERENCE, 2013, Perth. Proceedings...*, Perth, WA: Australasian Institute of Mining and Metallurgy, 2013, p. 129-136.

COWARD, S., VANN, J., DUNHAM, S., STEWART, M. The primary-response framework for geometallurgical variables. *In: SEVENTH INTERNATIONAL MINING GEOLOGY CONFERENCE, 2009, Perth. Proceedings...*, Perth, WA: Australasian Institute of Mining and Metallurgy, 2009, p. 109-113.

DAVID, D. Geometallurgical guidelines for miners, geologists and process engineers – discovery to design. *In: THE SECOND AUSIMM INTERNATIONAL GEOMETALLURGY CONFERENCE, 2013, Brisbane. Proceedings...*, Brisbane, QLD: Australasian Institute of Mining and Metallurgy, 2013, p. 129-132.

DEUTSCH, C. V., JOURNEL, A. G. **GSLIB: geostatistical software library and user's guide**. 2nd edition, New York: Oxford University Press, 1998.

DEUTSCH, C. V. Geostatistical modelling of geometallurgical variables – problems and solutions. *In: THE SECOND AUSIMM INTERNATIONAL GEOMETALLURGY CONFERENCE, 2013, Brisbane. Proceedings...*, Brisbane, QLD: Australasian Institute of Mining and Metallurgy, 2013, p.7-15.

DEUTSCH, J. L., PALMER, K., DEUTSCH, C. V., SZVMANSKI, J. Spatial modeling of geometallurgical properties: techniques and a case study. **Natural Resources Research**, v. 25, n. 2, p. 161-181, July 2015. DOI: 10.1007/s11053-015-9276-x.

DEUTSCH, J. L. **Multivariate spatial modeling of metallurgical rock properties**. 2015. Thesis (Doctor of Philosophy, Mining Engineering) – Department of Civil and Environmental Engineering, University of Alberta, 2015.

DRUMOND, D. A. **Estimativa e classificação de variáveis geometalúrgicas a partir de técnicas de aprendizado de máquinas**. 2019. Tese (Doutorado em Engenharia de Minas, Metalurgia e de Materiais) – Programa de Pós-Graduação em Engenharia de Minas, Metalurgia e de Materiais, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2019.

DUNHAM, S., VANN, J. Geometallurgy, geostatistics and project value – Does your block model tell you what you need to know? *In: PROJECT EVALUATION CONFERENCE, 2007, Melbourne. Proceedings...*, Melbourne: Australasian Institute of Mining and Metallurgy, 2007, p. 189-196.

DUNHAM, S., VANN, J., COWARD, S. Beyond geometallurgy – gaining competitive advantage by exploiting the broad view of geometallurgy. *In: THE FIRST AUSIMM INTERNATIONAL GEOMETALLURGY CONFERENCE, 2011, Brisbane. Proceedings...*, Brisbane, QLD: Australasian Institute of Mining and Metallurgy, 2011, p. 115-123.

DUTTA, S., BANDOPADHYAY, S., GANGULI, R., MISRA, D. Machine learning algorithms and their application to ore reserve estimation of sparse and imprecise data. **Journal of Intelligent Learning Systems and Applications**, v. 2, p. 86–96, May 2010. DOI: 10.4236/jilsa.2010.22012.

ENDERS, C. K. **Applied missing data analysis**. New York: The Guilford Press, 2010.

FERNANDES, F. G. **Estudo do melhor método de extrapolação de regressão múltipla para construção do modelo geometalúrgico de uma mina de fosfato brasileira**. 2013. Dissertação (Mestrado em Engenharia Mineral) – Programa de Pós-Graduação em Engenharia Mineral, Universidade Federal de Ouro Preto, Ouro Preto, 2013.

FERNEDA, E. Redes neurais e sua aplicação em sistemas de recuperação de informação. **Ciência da Informação**, Brasília, v. 35, n. 1, p. 25-30, jan./abr. 2006. DOI: 10.1590/S0100-19652006000100003.

FRANCO-SEPÚLVEDA, G., DEL RIO-CUERVO, J. C., PACHÓN-HERNÁNDEZ M. A. State of the art about metaheuristics and artificial neural networks applied to open pit mining. **Resources Policy**, v. 60, p. 125-133, Dec. 2018. DOI: 10.1016/j.resourpol.2018.12.013.

GIACOMEL, F. S. **Um método algorítmico para operações na bolsa de valores baseado em ensembles de redes neurais para modelar e prever os movimentos dos mercados de ações**. 2016. Dissertação (Mestrado em Computação) – Programa de Pós-Graduação em Computação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2016.

HAYKIN, S. **Redes neurais: princípios e prática**. 2. ed. Porto Alegre: Bookman, 2001.

HAIR, J. F., ANDERSON, R. E., TATHM, R. L., BLACK, W. C., BABIN, B. J. **Análise multivariada de dados**. 6. ed. New Jersey: Prentice Hall, 2009.

ISAAKS, E. H., SRIVASTAVA, R. M. **Applied geostatistics**. New York: Oxford University Press, 1989.

IZENMAN, A. J. **Modern multivariate statistical techniques**. New York: Springer, 2008.

JACKSON, J., MCFARLANE, A. J., OLSON HOAL, K. Geometallurgy – back to the future: scoping and communicating geomet programs. *In: THE FIRST AUSIMM INTERNATIONAL GEOMETALLURGY CONFERENCE*, 2011, Brisbane. **Proceedings...**, Brisbane, QLD: Australasian Institute of Mining and Metallurgy, 2011, p. 125-131.

JENSEN, J. L. W. V. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. **Acta Mathematica**, v. 30, p. 175-193, 1905. DOI: 10.1007/BF02418571.

JOHNSON, R. A., WICHERN, D. W. **Applied multivariate statistical analysis**. 5th ed. Upper Saddle River: Prentice Hall, 2002.

JORC. 2012. Australasian code for reporting of exploration results, mineral resources and ore reserves (The JORC Code), Melbourne, Victoria: Australasian Institute of Mining and Metallurgy, 2012.

JORJANI, E., BAGHERIEH, A. H., MESROGHLI, S., CHEHREH CHELGANI, S., JORJANI, H. Prediction of yttrium, lanthanum, cerium and neodymium leaching recovery from apatite concentrate using artificial neural networks. *In: METPLANT*, 2008, Perth. **Proceedings...**, Perth, Western Australia: Australasian Institute of Mining and Metallurgy, 2008, p. 433–446.

KAHRAMAN, S. The prediction of penetration rate for percussive drills from indirect tests using artificial neural networks. **Journal of the South African Institute of Mining and Metallurgy**, v. 116, n. 8, p. 793–800, Aug. 2016. DOI: 10.17159/2411-9717/2016/v116n8a12.

KEENEY, L., WALTERS, S. G. A Methodology for geometallurgical mapping and orebody modelling. *In: THE FIRST AUSIMM INTERNATIONAL GEOMETALLURGY CONFERENCE*, 2011, Brisbane. **Proceedings...**, Brisbane, QLD: Australasian Institute of Mining and Metallurgy, 2011, p. 217-225.

KINGMA, D. P., BA, J. L. Adam: a method for stochastic optimization. *In: INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS*, 2015, San Diego. **Proceedings...**, San Diego, CA, 2015, p. 1-13.

KITTLER, P., LIEBEZEIT, V., EHRIG, K., MACMILLAN, E., LOWER, C. ‘It seemed like a good idea at the time...’ common mistakes in geometallurgy. *In: THE FIRST AUSIMM INTERNATIONAL GEOMETALLURGY CONFERENCE*, 2011, Brisbane. **Proceedings...**, Brisbane, QLD: Australasian Institute of Mining and Metallurgy, 2011, p. 133-138.

KNIGHT, R., OLSON HOAL, K., ABRAHAM, A. P. G. Three-dimensional geometallurgical data integration for predicting concentrate quality and tailings composition in a massive sulfide deposit. *In: THE FIRST AUSIMM INTERNATIONAL GEOMETALLURGY CONFERENCE*, 2011, Brisbane. **Proceedings...**, Brisbane, QLD: Australasian Institute of Mining and Metallurgy, 2011, p. 227-232.

LAMBERG, P. Particles – the bridge between geology and metallurgy. *In: CONFERENCE IN MINERAL ENGINEERING*, 2011, Luleå. **Proceedings...**, Luleå: 2011, p. 1-16, 2011.

LEMOS, M. G. **Caracterização geológica e tecnológica do minério de zinco do Extremo Norte da Mina de Vazante, Minas Gerais**. 2011. Dissertação (Mestrado em Geologia e Recursos Naturais) – Programa de Pós-Graduação em Geociências, Universidade Estadual de Campinas, Campinas, 2011.

LIEBEZEIT, V., SMITH, M., EHRIG, K., KITTLER, P., MACMILLAN, E., LOWER, C. Geometallurgy data management – a significant consideration. *In*: THE FIRST AUSIMM INTERNATIONAL GEOMETALLURGY CONFERENCE, 2011, Brisbane. **Proceedings...**, Brisbane, QLD: Australasian Institute of Mining and Metallurgy, 2011, p. 237-246.

LISHCHUK, V., LUND, C., GHORBANI, Y. Evaluation and comparison of different machine-learning methods to integrate sparse process data into a spatial model in geometallurgy. **Minerals Engineering**, v. 134, p. 156-165, Apr. 2019. DOI: 10.1016/j.mineng.2019.01.032.

LOPERA, P. A. **Geometallurgical mapping and mine modelling – comminution studies: La Colosa case study, AMIRA P843A**. 2014. Dissertation (Master of Science in Geology) - University of Tasmania, Tasmania, 2014.

LOUWRENS, E., NAPIER-MUNN, T., KEENEY, L. Geometallurgical characterisation of a tailings storage facility – a novel approach. *In*: TAILINGS AND MINE WASTE MANAGEMENT FOR THE 21ST CENTURY, 2015, Sydney. **Proceedings...**, Sydney, NSW: Australasian Institute of Mining and Metallurgy, 2015, p. 125-131.

MACMILLAN, E., EHRIG, K., LIEBEZEIT, V., P KITTLER, P., LOWER, C. Use of geometallurgy to predict tailings leach acid consumption at Olympic Dam. *In*: THE FIRST AUSIMM INTERNATIONAL GEOMETALLURGY CONFERENCE, 2011, Brisbane. **Proceedings...**, Brisbane, QLD: Australasian Institute of Mining and Metallurgy, 2011, p. 93-102.

MAGALHÃES, M. N., LIMA, A. C. P. Noções de probabilidade e estatística. 6. ed. São Paulo: Universidade de São Paulo, 2007.

MCCOY, J. T., AURET, L. Machine learning applications in minerals processing: a review. **Minerals Engineering**, v. 132, p. 95-109, Mar. 2019. DOI: 10.1016/j.mineng.2018.12.004.

MCCULLOUGH, G., MUTSAERS, B., JESSOP, J. Obstacles to embedding geometallurgy decision-making into operations and how to increase maturity of this critical capability. *In: THE SECOND AUSIMM INTERNATIONAL GEOMETALLURGY CONFERENCE*, 2013, Brisbane. **Proceedings...**, Brisbane, QLD: Australasian Institute of Mining and Metallurgy, 2013, p. 139-140.

MCQUISTION, F. W., BECHAUD, L. J. Metallurgical sampling and testing. *In: SURFACE MINING*, 1968. Edited by E.P. Pfeleider. **Proceedings...**, Littleton, CO: SME-AIME, 1968, p. 103-121.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada**. 1. ed. Belo Horizonte: Editora UFMG, 2005.

MÓDOLO, M. **Classificação automática de supernovas usando redes neurais artificiais**. 2016. Tese (Doutorado em Computação Aplicada) – Programa de Pós-Graduação em Computação Aplicada, Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2016.

MONTOYA, P. A., KEENEY, L., JAHODA, R., HUNT, J., BERRY, R., DREWS, U., CHAMBERLAIN, V., LEICHLITER, S. Techniques applicable to prefeasibility projects – La Colosa case study. *In: THE FIRST AUSIMM INTERNATIONAL*

GEOMETALLURGY CONFERENCE, 2011, Brisbane. **Proceedings...**, Brisbane, QLD: Australasian Institute of Mining and Metallurgy, 2011, p. 103-111.

MOTTA, E. G. M. **Definição de domínios mineralógicos em minério de ferro utilizando krigagem de indicadores**. 2014. Dissertação (Mestrado em Engenharia de Minas, Metalurgia e de Materiais) – Programa de Pós-Graduação em Engenharia de Minas, Metalurgia e de Materiais, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2014.

MWANGA, A., ROSENKRANZ, J., LAMBERG, P. Testing of ore comminution behavior in the geometallurgical context – a review. **Minerals**, v. 5, p. 276-297, May 2015. DOI: 10.3390/min5020276.

NAKHAEI, F., MOSAVI, M. R., SAM, A., VAGHEI, Y. Recovery and grade accurate prediction of pilot plant flotation column concentrate: neural network and statistical techniques. **International Journal of Mineral Processing**, v. 110–111, p. 140–154, Jul. 2012. DOI: 10.1016/j.minpro.2012.03.003.

NEWTON, M. J., GRAHAM, J. M. Spatial modelling and optimisation of geometallurgical indices. *In*: THE FIRST AUSIMM INTERNATIONAL GEOMETALLURGY CONFERENCE, 2011, Brisbane. **Proceedings...**, Brisbane, QLD: Australasian Institute of Mining and Metallurgy, 2011, p. 247-261.

PANDA, L., TRIPATHY, S. K. Performance prediction of gravity concentrator by using artificial neural network - a case study. **International Journal of Mining Science and Technology**, v. 24, p. 461-465, Jun 2014. DOI: 10.1016/j.ijmst.2014.05.007.

PRADES, C. F., DEUTSCH, C. V. **Comparison of machine learning techniques for predicting and learning from geometallurgical multivariate databases.** 2016. 20 p. Paper 308, CCG Annual Report 18.

PEARSON, K. 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. DOI: 10.1080/14786440109462720.

REIS, E. A. Notas de aulas, UFMG: Universidade Federal de Minas Gerais, 2009.

RIEDMILLER, M., BRAUN, H. A direct adaptative method for faster backpropagation learning: the RPROP algorithm. *In*: PROCEEDINGS OF IEEE INTERNATIONAL CONFERENCE ON NEURAL NETWORKS, 1993, San Francisco. **Proceedings...**, San Francisco, 1993, p. 586-591.

RICHMOND, A., SHAW, W. J. Geometallurgical modelling – quo vadis? *In*: SEVENTH INTERNATIONAL MINING GEOLOGY CONFERENCE, 2009, Perth. **Proceedings...**, Perth, WA: Australasian Institute of Mining and Metallurgy, 2009, p. 115-118.

RUBIN, D. B. Inference and missing data. **Biometrika**, v. 63, n. 3., p. 581-592, Dec. 1976.

SAMANTA, B., BANDOPADHYAY, S., GANGULI, R., DUTTA, S. A comparative study of the performance of single neural networks vs. adaboost algorithm based combination of multiple neural networks for mineral resource estimation. **Journal of the South African Institute of Mining and Metallurgy**, v. 105, p. 237–246, Apr. 2005.

SEPULVEDA, E., DOWD, P.A., XU, C., ADDO, E. Multivariate modelling of geometallurgical variables by projection pursuit. **Mathematical Geosciences**, v. 49, p. 121–143, Jan. 2017. DOI: 10.1007/s11004-016-9660-z

SILVA, R. M., ALMEIDA, T. A., YAMAKAMI, A. Análise de desempenho de redes neurais artificiais para classificação automática de web spam. **Revista Brasileira de Computação Aplicada**, Passo Fundo, v. 4, n. 2, p. 42-57, Out. 2012. DOI: 10.5335/rbca.2012.2195.

SILVA, C. Z. **Metodologias de inserção de dados sob mecanismo de falta MNAR para modelagem de teores em depósitos multivariados heterotópicos**. 2018. Tese (Doutorado em Engenharia de Minas, Metalurgia e de Materiais) – Programa de Pós-Graduação em Engenharia de Minas, Metalurgia e de Materiais, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2018.

SHAW, W. J., KHOSROSHAHI, S., WEEKS, A. Modelling geometallurgical variability – a case study in managing risk. *In*: THE SECOND AUSIMM INTERNATIONAL GEOMETALLURGY CONFERENCE, 2013, Brisbane. **Proceedings...**, Brisbane, QLD: Australasian Institute of Mining and Metallurgy, 2013, p. 247-252.

SRIVASTAVA, V., BANDOPADHYAY, S., AKDOGAN, G., GHOSH, T. Dynamic neural network based model for SAG mill power prediction. *In*: 38TH INTERNATIONAL SYMPOSIUM ON THE APPLICATION OF COMPUTERS AND OPERATIONS RESEARCH IN THE MINERAL INDUSTRY, 2017, Colorado. **Proceedings...**, Golden, Colorado: Colorado School Mines, p. 227–233.

TAWO, E. E., AL-ALAWI, S. M. A comparison between artificial neural network and a geostatistical technique in the estimation of regionalized variables. **Engineering Journal of the University of Qatar**, v. 12, p. 125–149, 1999.

TRIOLA, M. F. **Introdução à estatística**. 10. ed. Rio de Janeiro: LTC, 2008.

VANN, J., JACKSON, J., COWARD, S., DUNHAM, S. The geomet curve – a model for implementation of geometallurgy. *In: THE FIRST AUSIMM INTERNATIONAL GEOMETALLURGY CONFERENCE*, 2011, Brisbane. **Proceedings...**, Brisbane, QLD: Australasian Institute of Mining and Metallurgy, 2011, p. 35-43.

VIEIRA, M. C. A. **Metodologia para prever recuperação de zinco em planta de beneficiamento**. 2016. Dissertação (Mestrado em Engenharia de Minas, Metalurgia e de Materiais) – Programa de Pós-Graduação em Engenharia de Minas, Metalurgia e de Materiais, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2016.

WALTERS, S. G. An overview of new integrated geometallurgical research. *In: NINTH INTERNATIONAL CONGRESS FOR APPLIED MINERALOGY*, 2008, Brisbane. **Proceedings...**, Brisbane, QLD: Australasian Institute of Mining and Metallurgy, 2008, p.79-82.

WILLIAMS, S. R., M. B. A., RICHARDSON, J. M. **Geometallurgical mapping: a new approach that reduces technical risk**. SGS Minerals Services, 2004. 13 p. Technical paper 2004-01.

WILLIAMS, S. R. A Historical perspective of the application and success of geometallurgical methodologies. *In: THE SECOND AUSIMM INTERNATIONAL GEOMETALLURGY CONFERENCE*, 2013, Brisbane. **Proceedings...**, Brisbane, QLD: Australasian Institute of Mining and Metallurgy, 2013, p. 37-47.

WIRFIYATA, F., MCCAFFERY, K. Applied geo-metallurgical characterisation for life of mine throughput prediction at Batu Hijau. *In: INTERNATIONAL AUTOGENOUS GRINDING, SEMIAUTOGENOUS GRINDING AND HIGH PRESSURE GRINDING*

ROLL TECHNOLOGY, 2011, Westmount. **Proceedings...**, Westmount, QC: Canadian Institute of Mining, Metallurgy and Petroleum, 2011, p.1-22.

WU, X., ZHOU, Y. Reserve estimation using neural network techniques. **Computers and Geosciences**, v. 19, n. 4, p. 567–575, Apr. 1993. DOI: 10.1016/0098-3004(93)90082-G.