



UNIVERSIDADE FEDERAL DE CATALÃO (UFCAT)
INSTITUTO DE FÍSICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS E ENGENHARIA DE
MATERIAIS – PPGCEM

SILVÂNIA ALVES BRAGA DE CASTRO

AGRUPAMENTO DE DOMÍNIOS E MODELAGEM GEOMETALÚRGICA DA
MINA CHAPADÃO, OUVIDOR - GO

CATALÃO

2024



UNIVERSIDADE FEDERAL DE CATALÃO

INSTITUTO DE FÍSICA

Av. Dr. Lamartine Pinto de Avelar, número 1120, - Bairro Setor Universitário, Catalão/GO, CEP 75704-020
Telefone: - - <https://www.ufcat.edu.br>

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA)

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DE TESES E DISSERTAÇÕES DA UNIVERSIDADE FEDERAL DE CATALÃO (UFCAT)

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Catalão (UFCAT) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFCAT), sem ressarcimento dos direitos autorais, de acordo com a Lei 9.610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFCAT é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o(a) autor(a) e o(a) orientador(a) Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico (Dissertação/Tese)

TESE

2. Nome completo do(a) autor(a): SILVANIA ALVES BRAGA DE CASTRO

Nome completo do(a) orientador(a): ANDRE CARLOS SILVA

3. Título do trabalho

Título: AGRUPAMENTO DE DOMÍNIOS E MODELAGEM GEOMETALÚRGICA DA MINA CHAPADÃO, OUIDOR - GO

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento: SIM [] NÃO¹

[¹] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:

a) consulta ao(à) autor(a) e ao(à) orientador(a);

b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação.

O documento não será disponibilizado durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

Obs.: Este termo deverá ser assinado no SEI pelo orientador e pelo autor



Documento assinado eletronicamente por **ANDRE CARLOS SILVA, Orientador(a)**, em 29/11/2024, às 11:46, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **SILVANIA A B CASTRO, Usuário Externo**, em 29/11/2024, às 12:00, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufcat.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0134193** e o código CRC **B0619F2C**.

SILVÂNIA ALVES BRAGA DE CASTRO

AGRUPAMENTO DE DOMÍNIOS E MODELAGEM GEOMETALÚRGICA DA
MINA CHAPADÃO, OUVIDOR - GO

Tese apresentada ao Programa de Pós-Graduação em Ciências e Engenharia de Materiais, do Instituto de Física, da Universidade Federal de Catalão como requisito para obtenção do título de Doutora em Ciências e Engenharia de Materiais. Área de concentração: Metodologia da Ciência e Engenharia de Materiais. Linha de pesquisa: Caracterização e Desenvolvimento de Moléculas e Materiais

Orientador: Prof. Dr. André Carlos Silva
Coorientadora: Prof^a. Dr^a. Elenice Maria Schons
Silva

CATALÃO

2024

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UFCAT.

Castro, Silvânia Alves Braga de
AGRUPAMENTO DE DOMÍNIOS E MODELAGEM
GEOMETALÚRGICA DA MINA CHAPADÃO, OUVIDOR - GO /
Silvânia Alves Braga de Castro. - 2024.
207, f.

Orientador: Prof. André Carlos Silva; co-orientador Prof. Elenice Maria Schons Silva.

Tese (Doutorado) - Universidade Federal de Catalão, Instituto de Física, Catalão, Programa de Pós-Graduação em Ciências Exatas e Tecnológicas, Catalão, 2024.

Bibliografia. Apêndice.

Inclui siglas, mapas, símbolos, gráfico, tabelas, lista de figuras, lista de tabelas.

1. Geometalurgia. 2. Agrupamento de domínios. 3. Aprendizado de máquinas. 4. Depósitos de fosfatos. I. Silva, André Carlos, orient. II. Título.

CDU 669

ATA DE DEFESA DE TESE

Ata nº 1 da sessão de Defesa de Tese de Doutorado aluna **SILVANIA ALVES BRAGA DE CASTRO**, que confere o título de Doutor em **CIÊNCIA E ENGENHARIA DE MATERIAIS**, na área de concentração **Metodologia da Ciência e Engenharia de Materiais**.

Aos vinte e nove de novembro de dois mil e vinte e quatro, a partir das oito horas, em sala virtual, realizou-se a sessão pública de Defesa de Tese intitulada "AGRUPAMENTO DE DOMÍNIOS E MODELAGEM GEOMETALÚRGICA DA MINA CHAPADÃO, OUVIDOR - GO". Os trabalhos foram instalados pelo Orientador, **Professor Doutor ANDRE CARLOS SILVA (FENG/UFCAT)** com a participação dos demais membros da Banca Examinadora: **Professora Doutora ELENICE MARIA SCHONS SILVA (FENG/UFCAT)**, Coorientadora, **Professor Doutor LUCIANO NUNES CAPPONI (UFRGS)**, membro titular externo; **Professor Doutor MARCÉLIO PRADO FONTES (CEFET/MG)**, membro titular externo; **Professor Doutor MAURICIO GUIMARAES BERGERMAN (PMI/USP)**, membro titular externo e **Professor Doutor HERNAN FRANCISCO ANTICOI SUDZUKI (EMIT/UPC)**, membro titular externo. Após a arguição do candidato, a Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Tese sendo considerada a candidata aprovada pelos seus membros. Proclamado o resultado pelo **Professor Doutor ANDRE CARLOS SILVA**, Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora.



Documento assinado eletronicamente por **ANDRE CARLOS SILVA, Orientador(a)**, em 29/11/2024, às 14:46, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **ELENICE MARIA SCHONS SILVA, Professor(a) do Magistério Superior**, em 29/11/2024, às 14:46, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Hernán Francisco Anticoi Sudzuki, Usuário Externo**, em 02/12/2024, às 14:00, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Maurício Guimarães Bergerman, Usuário Externo**, em 02/12/2024, às 16:02, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Marcélio Prado Fontes, Usuário Externo**, em 02/12/2024, às 17:45, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **LUCIANO NUNES CAPPONI, Usuário Externo**, em 02/12/2024, às 20:04, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufcat.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0141019** e o código CRC **D7692117**.

*Dedico esta tese às minhas filhas Estela e Elena,
pelas brincadeiras não brincadas e pelo tempo retirado delas (...).*

AGRADECIMENTOS

Aos Professores Dr. André Carlos Silva e Dra. Elenice Maria Schons Silva, pela orientação, aprendizado e confiança passada nos momentos difíceis.

À CMOC Brasil, pelo apoio e incentivo à pesquisa, por fornecer os dados necessários para a realização deste trabalho.

À Universidade Federal de Catalão (UFCAT) e ao Centro Federal de Educação Tecnológica (CEFET-MG) pelo apoio. Agradeço a toda a equipe do LaMPPMin pela parceria e pelas amizades que cultivei.

Ao engenheiro Bruno Palhares Milanezi e a toda a equipe por colaborar com a pesquisa científica; e ao prof. Leandro Matioli pelo suporte.

Ao meu marido Gláucio pela paciência e apoio; aos meus pais Dolores e Hilton (em memória), aos meus irmãos Luciene, Leidiane, José e Sula, às minhas filhas Estela e Elena pelo amor e compreensão.

A Deus pela força concedida através da minha fé.

RESUMO

A modelagem de depósitos minerais tem sido aprimorada ao longo dos anos com a incorporação de informações geológicas, químicas e metalúrgicas. No entanto, a diversidade dessas informações torna essa atividade dispendiosa, já que os dados não são unificados nos procedimentos, no suporte amostral e nas quantidades. A não aditividade e a não linearidade de certos atributos impõem mais dificuldades. Recentemente, novas abordagens em geometurgia computacional têm mostrado progressos notáveis na análise de dados espaciais para contornar os desafios que restringiriam a criação desses modelos. Este estudo é impulsionado por essas novas metodologias e emprega técnicas de aprendizado de máquina para viabilizar o modelo da Mina Chapadão, enfrentando dificuldades, como a complexidade geológica dos contatos transicionais e a sobreposição de estruturas, além da falta de uniformidade no banco de dados, isto é, as informações metalúrgicas são significativamente menores em comparação com as outras. Essa situação não é exclusiva de Chapadão, mas sim comum à maioria dos depósitos minerais, que frequentemente limitam a modelagem dos minérios aos atributos primários da rocha. A generalização dos modelos metalúrgicos é opção válida quando as relações entre as características geológicas e as respostas de processamento são bem definidas. Sob essa perspectiva, tipologias de minério foram definidas, cujas características geológicas e químicas são compatíveis com os resultados metalúrgicos. O algoritmo fuzzy foi testado para gerenciar a incerteza do fenômeno estudado e para classificar as amostras de sondagem em domínios tipológicos modelados por krigagem dos indicadores. Três técnicas de aprendizado de máquinas foram analisadas para predição dos atributos de processo. Os resultados satisfatórios dos algoritmos de predição possibilitaram a simulação geoestatística das variáveis primárias para cada domínio do modelo geometúrgico que foram reconciliadas com os dados de produção. Foi possível observar que o modelo geometúrgico obtido pode ser utilizado para prever o desempenho do minério na planta de processamento mineral.

Palavras-chave: Geometurgia, Agrupamento de domínios, Aprendizado de máquinas.

ABSTRACT

Mineral deposit modeling has been improved over the years with the incorporation of geological, chemical, and metallurgical information. However, the diversity of this information makes this activity expensive since the data are not unified in procedures, sample support, or quantities. The non-additivity and non-linearity of certain attributes impose further difficulties. Recently, new approaches in computational geometallurgy have shown remarkable progress in the analysis of spatial data to overcome the challenges that would restrict the creation of these models. This study is driven by these new methodologies and employs machine learning techniques to enable the Chapadão Mine model, facing difficulties such as the geological complexity of transitional contacts and overlapping structures, in addition to the lack of uniformity in the database, that is, the metallurgical information is significantly smaller compared to the others. This situation is not exclusive to Chapadão, but is common to most mineral deposits, which often limit mineral modeling to the primary attributes of the rock. The generalization of metallurgical models is a valid option when the relationships between geological characteristics and processing responses are well defined. From this perspective, ore typologies were defined, whose geological and chemical characteristics are compatible with the metallurgical results. The fuzzy algorithm was tested to manage the uncertainty of the studied phenomenon and to classify the drill samples into typological domains modeled by kriging of the indicators. Three machine learning techniques were analyzed to predict the process attributes. The satisfactory results of the predictive models allowed the geostatistical simulation of the primary variables for each domain of the geometallurgical model that were reconciled with the production data. It was possible to observe that the geometallurgical model obtained can be used to predict the performance of the ore in the plant.

Keywords: Geometallurgy, Domain clustering, Machine learning.

LISTA DE FIGURAS

Figura 1. Exemplo das etapas de um programa geometalúrgico. Adaptado de P. Lamberg, 2011.....	23
Figura 2. Desigualdade de Jensen (adaptado de Coward <i>et al.</i> , 2009).	25
Figura 3. Procedimento gráfico para transformação da função de distribuição dos dados originais x em valores de uma distribuição normal x' . Modificado de Yamamoto, 2020.....	38
Figura 4. Dendrograma. Fonte (Shwartz; David, 2014).....	43
Figura 5. Exemplo de diferentes configurações de agrupamento para um mesmo conjunto de dado. Adaptado de (Shwartz; David, 2014).	43
Figura 6. Formas típicas de funções de pertinência: 1- triangular; 2- trapezoidal; 3- gaussiana; 4- em forma de sino; 5-singleton. Fonte: Kovacic; Bogdan, 2006.....	47
Figura 7. Implicação de Mamdani aplicada à regra fuzzy FR1.....	50
Figura 8. Implicação produto aplicada à regra fuzzy FR1	51
Figura 9. Procedimento de inferência individual baseado em regras.	52
Figura 10. Adaptado de Vendramin et al. (2010).	55
Figura 11. Conjunto de treinamento representado pelas classes vermelhas e azuis. Fonte: Shwartz e David (2014).	62
Figura 12. Em a) o efeito underfitting, cujo modelo não é suficiente para ajustar as amostras de treinamento (pontos). Em b) o modelo bem ajustado. Em c) o modelo polinomial aproxima a função verdadeira quase perfeitamente, aprendendo o ruído dos dados de treinamento (overfitting). Fonte: https://scikit-learn.org/stable/auto_examples/model_selection/plot_underfitting_overfitting.html	68
Figura 13. Exemplo de validação cruzada. Caixas cinzas correspondem ao grupo de teste, enquanto as brancas correspondem ao grupo de treino. No eixo x, são representados os diferentes grupos K. Fonte: (adaptado de Müller et al., 2016). ...	68
Figura 14. Localização geográfica do Complexo Catalão I.	69
Figura 15. Produtos potenciais do complexo Catalão I. Em vermelho direito, minério da mina Chapadão; em azul, o limite da cava. Adaptado de CMOC 2023.....	71
Figura 16. À esquerda, localização dos complexos alcalinos pertencentes à PIAP (modificado de Gibson et al., 1995); e à direita, condução do calor da pluma do manto subjacente (modificado de Barbosa, 2005).....	72
Figura 17. Classificação informal das rochas da série bebedourítica proposta por Brod et al. (2004).	74
Figura 18. Classificação das rochas da série foscorítica segundo Yegorov (1993).....	75
Figura 19. Classificação química de rochas da série carbonatítica segundo Woolley & Kempe (1989).	76
Figura 20. Mapa geológico da mina. Modificado de CMOC (2023).....	78
Figura 21. Em I, o perfil intempérico proposto por Oliveira e Imbernon (1998); em II, adaptação de Ribeiro (2008) para as classes de minérios.	79
Figura 22. Fluxograma do processo de concentração mineral de fosfatos.....	82
Figura 23. Informações contidas no banco de dados.	85
Figura 24. Análise de agrupamento pelos métodos k-means e hierárquico.	87
Figura 25. Informações geológicas da mina Chapadão e redefinição dos domínios.	87
Figura 26. Análise estatística dos teores utilizada como entrada <i>crisp</i> no modelo fuzzy.	89
Figura 27. Função de pertinência e termos linguísticos dos conjuntos fuzzy de entrada.	89

Figura 28. Arquitetura básica do algoritmo fuzzy.	90
Figura 29. Componentes do modelo de predição fuzzy.	91
Figura 30. Histogramas dos indicadores das tipologias da mina.	92
Figura 31. Variograma do FO nos eixos de maior continuidade e perpendicular.	92
Figura 32. Variograma do FO nos eixos no eixo vertical.	93
Figura 33. Quantitativo dos ensaios de flotabilidade por domínio.	94
Figura 34. <i>Box-plots</i> das variáveis respostas por domínio tipológico.	95
Figura 35. Análise de Variância para as cinco tipologias de minério e as variáveis respostas.	96
Figura 36. Divisão do banco de dados para os modelos de predição.	96
Figura 37. Número de realizações versus o desvio padrão médio das simulações.	98
Figura 38. Distribuição dos furos de sondagem na cota 800.	99
Figura 39. <i>Box plot</i> dos teores para o perfil intempérico.	101
Figura 40. Matriz de correlação dos teores por horizonte intempérico.	102
Figura 41. Matriz de correlação dos teores por horizonte intempérico.	103
Figura 42. Gráficos <i>box plot</i> dos teores dos litotipos.	104
Figura 43. Gráficos <i>box plot</i> dos teores dos litotipos.	105
Figura 44. Gráficos <i>box plot</i> dos teores dos litotipos.	106
Figura 45. Resultados dos agrupamentos. Em a) fuzzy; b) k-means e c) hierárquico. .	107
Figura 46. Verificação visual em seções horizontais: amostras sondagem, agrupamento fuzzy e k-means, nessa ordem.	108
Figura 47. Análise estatística dos teores nas amostras precisas e nos domínios fuzzy. .	109
Figura 48. Blocos com pertinência fuzzy em uma ou mais classes, cota 750.	110
Figura 49. Histograma dos dados desagrupados à esquerda; histograma dos indicadores krigados à direita.	111
Figura 50. Histograma dos dados desagrupados à esquerda; histograma dos indicadores krigados à direita.	112
Figura 51. Histograma das maiores probabilidades escolhidas para categorizar os blocos.	113
Figura 52. Ocorrências proporcionais das tipologias no banco de dados e no modelo de blocos.	114
Figura 53. Verificação visual do modelo de blocos e as amostras na seção horizontal Z850.	114
Figura 54. Verificação visual do modelo de blocos e as amostras na seção horizontal Z750.	115
Figura 55. Verificação visual do modelo de blocos e as amostras na seção vertical 7.992.500E.	115
Figura 56. Verificação visual do modelo de blocos e as amostras na seção vertical 7.993.000.	115
Figura 57. À esquerda a variância explicada e, à direita as cargas fatoriais das sete primeiras PC's.	116
Figura 58. Em a) variáveis selecionadas na PCA, em b) similaridade entre as variáveis.	116
Figura 59. Matriz de correlação entre as variáveis químicas e de processo.	117
Figura 60. Gráficos do coeficiente de correlação preditos contra reais para DIST_MAS.	119
Figura 61. Gráficos do coeficiente de correlação preditos contra reais para MMAG. .	120
Figura 62. Gráficos do coeficiente de correlação preditos versus reais para P2O5_C. .	121
Figura 63. Gráficos do coeficiente de correlação preditos versus reais para DIST_MAS.	124

Figura 64. Gráficos do coeficiente de correlação preditos versus reais para MMAG..	125
Figura 65. Gráficos do coeficiente de correlação preditos versus reais para P ₂ O ₅ _C. .	126
Figura 66. Gráficos do coeficiente de correlação preditos versus reais para DIST_MAS.	129
Figura 67. Gráficos do coeficiente de correlação preditos versus reais para MMAG. .	130
Figura 68. Gráficos do coeficiente de correlação preditos versus reais para P ₂ O ₅ _C. .	131
Figura 69. Histogramas dos teores P ₂ O ₅ , Fe ₂ O ₃ e MgO na parte superior e dos teores normalizados para o FO no oxidado (na parte inferior).....	134
Figura 70. Histogramas dos teores P ₂ O ₅ , Fe ₂ O ₃ e MgO na parte superior e dos teores normalizados para o FO no micáceo (na parte inferior da figura).....	135
Figura 71. Histogramas dos teores P ₂ O ₅ , Fe ₂ O ₃ e MgO na parte superior e dos teores normalizados para o FLPI no oxidado (na parte inferior da figura).	136
Figura 72. Histogramas dos teores P ₂ O ₅ , Fe ₂ O ₃ e MgO na parte superior e dos teores normalizados para o FLPI no micáceo (na parte inferior da figura).....	137
Figura 73. Histogramas dos teores P ₂ O ₅ , Fe ₂ O ₃ e MgO na parte superior e dos teores normalizados para o CBMS no oxidado (na parte inferior da figura).	138
Figura 74. Histogramas dos teores P ₂ O ₅ , Fe ₂ O ₃ e MgO na parte superior e dos teores normalizados para o CBMS no micáceo (na parte inferior da figura).....	139
Figura 75. Histogramas dos teores P ₂ O ₅ , Fe ₂ O ₃ e MgO na parte superior e dos teores normalizados para o CBMG no oxidado (na parte inferior da figura).....	140
Figura 76. Histogramas dos teores na parte superior e normalizados para o CBMG no micáceo (na parte inferior da figura).	141
Figura 77. Variogramas dos dados na vertical (à esquerda), direção 0°(centro) e 90° (à direita) FO oxidado.	144
Figura 78. Variogramas dos dados na vertical (à esquerda), direção 0°(centro) e 90° (à direita) FLPI oxidado.	145
Figura 79. Variogramas dos dados na vertical (à esquerda), direção 0°(centro) e 90° (à direita) CBMS oxidado.	146
Figura 80. Variogramas dos dados na vertical (à esquerda), direção 0°(centro) e 90° (à direita) CBMG oxidado.....	147
Figura 81. Variogramas dos dados na vertical (à esquerda), direção 0°(centro) e 90° (à direita) FO micáceo.....	148
Figura 82. Variogramas dos dados na vertical (à esquerda), direção 0°(centro) e 90° (à direita) FLPI micáceo.....	149
Figura 83. Variogramas dos dados na vertical (à esquerda), direção 0°(centro) e 90° (à direita) CBMS micáceo.....	150
Figura 84. Variogramas dos dados na vertical (à esquerda), direção 0°(centro) e 90° (à direita) CBMG micáceo.	151
Figura 85. Validação da reprodução da distribuição dos teores de P ₂ O ₅ , MgO e Fe ₂ O ₃ . Em preto estão representadas as distribuições das simulações e em vermelho o a distribuição desagrupada para o FO oxidado e micáceo.	152
Figura 86. Validação da reprodução da distribuição dos teores de P ₂ O ₅ , MgO e Fe ₂ O ₃ . Em preto estão representadas as distribuições das simulações e em vermelho o a distribuição desagrupada para o FLPI oxidado e micáceo.....	153
Figura 87. Validação da reprodução da distribuição dos teores de P ₂ O ₅ , MgO e Fe ₂ O ₃ . Em preto estão representadas as distribuições das simulações e em vermelho o a distribuição desagrupada para o CBMS oxidado e micáceo.	154
Figura 88. Validação da reprodução da distribuição dos teores de P ₂ O ₅ , MgO e Fe ₂ O ₃ . Em preto estão representadas as distribuições das simulações e em vermelho o a distribuição desagrupada para o CBMG oxidado e micáceo.	155

Figura 89. Validação dos variogramas da esquerda para direita maior e menor continuidade. Em verde os 35 variogramas gerados pela simulação, em vermelho o variograma dos dados originais para o FO oxidado.	156
Figura 90. Validação dos variogramas da esquerda para direita maior e menor continuidade. Em verde os 35 variogramas gerados pela simulação, em vermelho o variograma dos dados originais para o FO micáceo.	157
Figura 91. Validação dos variogramas da esquerda para direita maior e menor continuidade. Em verde os 35 variogramas gerados pela simulação, em vermelho o variograma dos dados originais para o FLPI oxidado.	158
Figura 92. Validação dos variogramas da esquerda para direita maior e menor continuidade. Em verde os 35 variogramas gerados pela simulação, em vermelho o variograma dos dados originais para o FLPI micáceo.	159
Figura 93. Validação dos variogramas da esquerda para direita maior e menor continuidade. Em verde os 35 variogramas gerados pela simulação, em vermelho o variograma dos dados originais para o CBMS oxidado.....	160
Figura 94. Validação dos variogramas da esquerda para direita maior e menor continuidade. Em verde os 35 variogramas gerados pela simulação, em vermelho o variograma dos dados originais para o CBMS micáceo.	161
Figura 95. Validação dos variogramas da esquerda para direita maior e menor continuidade. Em verde os 35 variogramas gerados pela simulação, em vermelho o variograma dos dados originais para o CBMG oxidado.	162
Figura 96. Validação dos variogramas da esquerda para direita maior e menor continuidade. Em verde os 35 variogramas gerados pela simulação, em vermelho o variograma dos dados originais para o CBMG micáceo.....	163
Figura 97. Gráfico dos teores simulados (9 cenários em azul) e os dados originais desagrupados (verde).	164
Figura 98. Gráfico das médias das variáveis de processo por domínio tipológico.....	165
Figura 99. Gráfico do coeficiente de variação médio das variáveis de processo por domínio tipológico.	165
Figura 100. Gráficos das médias e desvio padrão do modelo de blocos e do banco de dados.	166
Figura 101. Mapa do coeficiente de variação da recuperação mássica na cota 810 m.	166
Figura 102. Mapa do coeficiente de variação da recuperação mássica na cota 840 m, à esquerda distribuição da amostragem.	166
Figura 103. Relação DIST_MAS e P ₂ O ₅ _C entre usina e modelo de blocos.	168
Figura 104. Recuperação mássica entre usina e modelo de blocos após fator de mudança de escala. À direita os desvios entre modelo geometalúrgico, amostras e dados usina.	168
Figura 105. P ₂ O ₅ _C entre usina e modelo de blocos. À direita os desvios entre modelo, amostras e dados usina.	169
Figura 106. Proporções das tipologias na programação das pilhas.	169
Figura 107. Resultados dos modelos preditivos aplicados aos teores do volume de alimentação proporcional a participação das tipologias na composição das pilhas.	170

LISTA DE TABELAS

Tabela 1. Variáveis de processo obtidas nos ensaios de bancada.....	83
Tabela 2. Parâmetros testados nos algoritmos.....	97
Tabela 3. Simulações fuzzy e aderência dos domínios com as amostras.....	107
Tabela 4. Verificação dos agrupamentos pelo CR.....	109
Tabela 5. Incerteza geológica dada pelo grau de pertinência das amostras.	110
Tabela 6. Teste de normalidade dos resíduos para os modelos gerados.	118
Tabela 7. Equações da regressão linear para os domínios testados.	118
Tabela 8. Índice de importância da variável obtidos no RF para o horizonte oxidado.	122
Tabela 9. Índice de importância da variável obtidos no RF para o horizonte micáceo.	122
Tabela 10. Número mínimo de amostras por folha, horizonte oxidado.....	123
Tabela 11. Número mínimo de amostras por folha, horizonte micáceo	123
Tabela 12. Resultados dos modelos SVM não linear para diferentes parâmetros C, zona oxidado. À direita as correlações e à esquerda o RMSE.....	127
Tabela 13. Resultados dos modelos SVM não linear para diferentes parâmetros C na zona micáceo. direita as correlações e à esquerda o RMSE.....	127
Tabela 14. Comparação dos coeficientes de correlação entre os modelos SVM linear e não linear.	128
Tabela 15. Avaliação dos resultados da validação cruzada 5-fold – oxidado.....	132
Tabela 16. Avaliação dos resultados da validação cruzada 5-fold – micáceo	132
Tabela 17. Comparação entre os modelos preditivos utilizando todas as variáveis primárias e as três variáveis selecionadas para oxidado e micáceo.....	133
Tabela 18. Sumário estatístico dos histogramas dos teores do banco de dados.	134
Tabela 19. Relação madograma pela raiz do variograma das variáveis normalizadas.	141
Tabela 20. Cálculo do fator de mudança de escala.	168

LISTA DE SIGLAS

CB – Dolomita carbonatito

CBMG – Foscorito com carbonatito

CBMS – Apatita carbonatitos

COG – Centro de gravidade

CR - Índice de Rand corrigido

DIST_MAS – recuperação mássica (ensaios de bancada)

ERM - minimização do risco empírico

FL – Flogopititos

FO – Foscorito

KI- Krigagem dos indicadores

MMAG –Porcentagem da massa do material magnético do separador de baixo campo

NL - Nelsonito

NPK - abreviação para fórmula de fertilizantes Nitrogênio (N), Fósforo (P) e Potássio (K)

P₂O₅_C – Teor de P₂O₅ no concentrado

PCA - Análise de Componentes Principais

PI – Picrito com xenólitos

PIAP – Província ígnea do Alto Paranaíba

RF- *Random Forest* ou Floresta aleatória

RL – Regressão Linear

SGS - Simulação sequencial gaussiana

SQG soma dos quadrados entre os grupos

SQR soma dos quadrados das amostras dentro do grupo

SQT soma dos quadrados das variações das populações

SVM - *Support Vector Machine*, ou Máquina de Vetores de Suporte

LISTAS DE SÍMBOLOS

b - Constante que determina o deslocamento do hiperplano no SVM

$C(h)$ - Covariância

$F(X)$ - Função de distribuição acumulada da variável aleatória X

h_n - Direções de continuidade espacial

$I(x)$ - Variável indicatriz categórica

L - Pesos da matriz de descorrelação

$L_s^{hinge}(w)$ - Função de Lagrange

$p^*_{SK}(u, k)$ - Probabilidade estimada de k categoria

Sph - Modelo esférico de variograma

T - Operador de transposição de matriz

T -norm - Lógica da t-norma Lukasiewicz (operador AND)

U - Universo de discurso

Var - Variância

w - Vetor do hiperplano SVM

σ - Desvio padrão

ρ - Coeficiente de correlação de Pearson

β - Coeficientes de regressão

$\gamma(h)$ - Variograma

$\mu_{f(x)}$ - Função de pertinência

μ - Grau de pertinência

$\hat{\mu}$ - Média

ξ_n - Variável de folga para SVM suave

SUMÁRIO

1 INTRODUÇÃO	17
1.1 Contextualização.....	17
1.2 Problema e hipótese	20
1.3 Objetivos	21
1.4 Contribuição da tese	21
2 GEOMETALURGIA.....	21
2.1 Banco de dados	22
2.2 Variáveis geometalúrgicas	24
2.3 Modelagem espacial e geoestatística.....	26
2.3.1 Definição de domínios	26
2.3.2 Krigagem dos indicadores.....	28
2.3.3 Simulação sequencial gaussiana	30
2.4 Incerteza dos modelos geometalúrgicos.....	33
2.5 Avanços da geometalurgia	35
3 ANÁLISE MULTIVARIADA E APRENDIZADO DE MÁQUINA	37
3.1 Análise exploratória e modelos de distribuição.....	37
3.2 Coeficiente de correlação.....	39
3.3 Análise de componentes principais - PCA.....	40
3.4 Técnicas de Agrupamento	41
3.4.1 Agrupamento hierárquico	42
3.4.2 Agrupamento por partição	43
3.4.3 Algoritmo difuso	45
3.4.3.1 Conjunto fuzzy, operadores e variáveis linguísticas.....	46
3.4.3.2 Regras e implicação fuzzy	48
3.4.3.3 Desfuzzificação	52
3.4.4 Métricas de verificação dos agrupamentos.....	53
3.5 Análise de regressão.....	55

3.5.1 Regressão simples	56
3.5.2 Regressão múltipla.....	57
3.5.3 Suposições dos modelos e verificação da regressão	58
3.6 Floresta aleatória	59
3.7 Máquinas de vetores de suporte (SVM)	61
3.7.1 SVM de margem rígida (hard SVM)	61
3.7.2 SVM de margem suave	64
3.7.3 Função Kernel.....	66
3.8 Métricas de verificação dos modelos preditivos	67
4. ESTUDO DE CASO	69
4.1 A Mina Chapadão.....	69
4.2 Geologia	71
4.2.1 Protólitos	71
4.2.2 Perfil de alteração intempérica	79
4.3 Processamento mineral.....	81
4.3.1 Ensaios de flotabilidade	82
5. METODOLOGIA.....	84
6. RESULTADOS.....	100
6.1 Apresentação dos dados e estatísticas descritivas.....	100
6.2 Agrupamento de domínios tipológicos.....	106
6.3 Modelagem dos indicadores.....	111
6.4 Seleção de variáveis	115
6.5 Modelos preditivos	117
6.6 Simulação das variáveis primárias	133
6.7 Reconciliação	167
7. CONCLUSÕES	170
7.1 Aplicação do algoritmo difuso para definição dos domínios com amostras imprecisas.	170
7. 2 Utilização de variáveis primárias para modelagem de atributos metalúrgicos por técnicas computacionais	171

7.3 Cumprimento das metas e recomendações.....	171
8. REFERÊNCIAS	173
APÊNDICE.....	183
APÊNDICE A – Histograma dos indicadores das tipologias do banco de dados.	183
APÊNDICE B – Variogramas verticais dos indicadores das tipologias do banco de dados.....	184
APÊNDICE C - Variogramas azimute 0° dos indicadores das tipologias.....	185
APÊNDICE D - Variogramas azimute 90° dos indicadores das tipologias.....	186
APÊNDICE E - Mapa variograma à esquerda para os teores e à direita para os indicadores do dos principais domínios nas zonas mineralizadas.	187
APÊNDICE F - Gráfico de resíduos da regressão linear.	188
APÊNDICE G –Variogramas verticais dos cenários simulados e dos dados para os teores P2O5 do Fe2O3 e do MgO.....	200
APÊNDICE H – Mapas de coeficiente de variação da recuperação mássica	203

1 INTRODUÇÃO

1.1 Contextualização

Estudos apontam que o Brasil é um dos maiores produtores e exportadores de alimentos. O país é responsável, atualmente, por cerca de 8% do consumo global de fertilizantes, ocupando a quarta posição, atrás apenas da China, Índia e dos Estados Unidos. No entanto, mais de 87% dos fertilizantes utilizados no Brasil são importados, evidenciando um elevado nível de dependência externa (Ministério da Agricultura, 2022). Dos elementos químicos presentes nos fertilizantes, conforme quantidade e proporção, o nitrogênio, o fósforo e o potássio são os que suprem as maiores deficiências do solo, por isso a fórmula básica dos fertilizantes NPK é usada.

O fósforo é o elemento presente na *commoditie* desta pesquisa, encontrado na natureza como fosfatos de rocha em depósitos de origens sedimentares, ígneas e biogénéticas. Os minérios de fosfatos sedimentar estão localizados nos Estados Unidos, no Marrocos e no Oriente Médio. Já os minérios de fosfatos ígneos estão presentes na África do Sul, Rússia, Finlândia e Brasil. No Brasil, os maiores depósitos minerais de rocha fosfática estão nos estados de Minas Gerais, Goiás e São Paulo, e mais de 80% dessa produção provém dos municípios de Tapira e Araxá, em Minas Gerais, e de Catalão e Ouidor, em Goiás. A exploração do fosfato ocorre nas rochas carbonatíticas com teores de P_2O_5 em torno de 10% (Saab; Paula, 2008).

Ademais, pesquisas recentes realizadas pelo Serviço Geológico Brasileiro (SGB-CPRM), relacionadas aos recursos minerais para fertilizantes no Brasil, indicam que apesar de a indústria nacional de fertilizantes fosfatados apresentar dados crescentes de produção ao longo dos anos, o Brasil tem hoje uma situação desconfortável em relação ao fosfato face à limitada fonte natural de recursos, pois não existe descoberta de novas reservas. Dessa forma, a tendência é a produção em minério de menor teor com alto custo e o esgotamento de parte de suas reservas até 2050. O maior tempo de vida para as minas é de 36 anos. Assim, vale destacar que, se descoberta novas reservas os empreendimentos mineiros em fosfato levam de 6 a 10 anos para o início de produção (Ministério da Agricultura, 2022, n.p).

Grasso (2015) cita vinte complexos alcalino-carbonatíticos com ocorrências de fosfatos no Brasil, todavia, apenas alguns desses complexos podem ser considerados depósitos econômicos de fosfato, e somente sete possuem minas em operação. É o caso das minas de Araxá, Tapira, Salitre em Minas Gerais; Catalão I e II em Goiás, Jacupiranga

e Juquiá em São Paulo; e Angico dos Dias na Bahia. Portanto, faz-se urgente a tomada de medidas para mitigar a situação do Brasil no cenário de rochas fosfáticas. A SBG e o Centro de Tecnologia Mineral realizaram vários estudos e apontaram a necessidade de (Mapa, 2022):

- I) Investimentos em pesquisas e descoberta de novas reservas para a manutenção da produção brasileira a longo prazo. Cerca de 60% e 21% do território nacional estão mapeados nas escalas 1:250.000 e 1:100.000, e menos de 5% têm cartografia geológica compatível com a escala de 1:50.000;
- II) Incentivo ao diálogo entre instituições governamentais e o setor privado para o desenvolvimento de novos projetos e novas cadeias emergentes em fertilizantes para viabilizar o ingresso da mineração no mercado de capitais, em especial na fase da pesquisa mineral, com o estímulo a *júnior companies*;
- III) Formatar e ampliar a rede de pesquisa sobre beneficiamento de minérios de fosfato. Bem como estimular pesquisas e estudos de melhorias na cadeia produtiva, da geologia à concentração. Isso envolveria pesquisadores de instituições técnico-científicas brasileiras e as empresas atuantes na produção de rochas fosfáticas, buscando aumento da vida útil das minas. Um exemplo desta medida, é o projeto de pesquisa fruto da parceria entre Universidade Federal de Catalão e CMOC Brasil.

Visando a cumprir tais medidas, o Conselho Nacional de Fertilizantes e Nutrição de Plantas (CONFERT), sob a liderança do vice-presidente Geraldo Alckmin, ratificou, em novembro de 2023, as diretrizes, os objetivos e as iniciativas do novo Plano Nacional de Fertilizantes (PNF). As ações imediatas e de médio prazo incluem a reativação ou expansão de fábricas de fertilizantes essenciais para o Brasil, principalmente os nitrogenados e fosfatados. Além disso, o plano sugere um extenso mapeamento geológico para avaliar o potencial de fosfato e potássio no país, prevendo pesquisas regionais em até 60 áreas nas próximas décadas, especialmente nos estados de Goiás, Tocantins, Bahia e Mato Grosso. O objetivo central é alcançar, até 2050, uma produção nacional que satisfaça de 45% a 50% da demanda interna. (Ministério do Desenvolvimento, Indústria, Comércio e Serviços-MDIC, 2023).

Sabe-se que a preocupação com a demanda de fertilizantes pelo setor agrícola no país é uma questão urgente, que exige ações efetivas para promover a descoberta e viabilidade de novas jazidas, assim como aumentar a vida útil das minas em operação. Nesse sentido, este estudo se alinha com tais ações necessárias.

A Mina Chapadão está situada no estado de Goiás e faz parte do Complexo Catalão I. O depósito mineral é resultante da concentração supergênica de apatita durante o intemperismo das rochas e contém abundante apatita residual, pirocloro, monazita, minerais de titânio (perovskita, ilmenita e anatase) e vermiculita. A lavra da jazida iniciou-se em uma porção superficial do manto de intemperismo na zona oxidada e mais homogênea. Porém, com o aprofundamento da cava, a lavra está se aproximando da zona rochosa, o que acarreta a diminuição dos teores dos minérios e as significativas mudanças em suas características físicas e químicas com implicações no seu processamento. Isso resulta em altos custos de produção e baixa recuperação mássica e metalúrgica (Braga, 2016). Portanto, reconhecer as características mineralógicas é importante para o desenvolvimento do processo de beneficiamento, como a relação minério-ganga, características do mineral de minério, granulometria de liberação entre outras.

Um modelo de recursos e reservas bem-feito envolve a criação de uma matriz que inclui mineralogia, teor e parâmetros metalúrgicos (Williams; Richardson, 2004). Esses modelos são denominados de modelos geometalúrgicos e visam o melhor aproveitamento dos minérios.

Diante do exposto, este trabalho pretende avaliar técnicas que contribuem para a elaboração do modelo geometalúrgico da Mina Chapadão, integrando as informações geológicas e os ensaios de flotabilidade. É relevante ressaltar que, na maioria dos depósitos de fosfatos residuais, os modelos geológicos limitam-se à definição das zonas de intemperismo. Sendo assim, a compreensão dos protolitos e das características da apatita é geralmente alcançada por meio dos mapas geológicos das frentes de lavra. Esse conhecimento é essencial para prever as respostas dos minérios, tornando a modelagem das litologias um processo indispensável. Um desafio na delimitação espacial das litologias em depósitos ígneos como o de Chapadão está nas características estruturais desses complexos (*stockworks*). Essas características dificultam as descrições precisas dos furos de sondagem. Assim, esse estudo propõe, além da predição das respostas de processamento mineral no modelo de blocos, uma abordagem para medir a incerteza geológica na base de dados das descrições das amostras.

1.2 Problema e hipótese

A elaboração dos modelos geometalúrgicos, exige um banco de dados conciso no suporte amostral e homogêneo no que se refere às quantidades de amostras para modelagem de qualquer atributo. O problema é que o banco de dados em estudo não atende a tais critérios e carece de mais ensaios de flotabilidade para a estimativa direta dessas variáveis. Logo, uma alternativa para suprir essa dificuldade é estudar uma função de transferência que expresse as relações entre as variáveis geológicas e químicas exaustivas com as respostas de processamento mineral. Para isso, é necessária a delimitação de domínios homogêneos que caracterizem o depósito e podem ser usados para a inferência das variáveis preditoras.

Dessa forma, este é o segundo problema deste trabalho, pois mais de 53% das amostras dos furos de sondagem que compõem o banco de dados têm descrições litológicas ambíguas, visto que as rochas ocorrem em corpos ou veios estreitos, heterogêneos, sobrepostos e orientados aleatoriamente. Ademais, o tamanho restrito dos testemunhos de sondagem dificulta a identificação das litologias predominantes e pode induzir a uma descrição equivocada. Isso demanda que o geólogo de recursos ou geoestatístico tenha um trabalho adicional de verificar as amostras para além da região de contato dos domínios litológicos, tornando-se uma tarefa laboriosa.

Portanto, a hipótese da pesquisa é estudar uma ferramenta que valide as informações dos furos de sondagem e, ao mesmo tempo, gerar uma medida para quantificar a incerteza geológica das descrições. Alguns trabalhos sobre incerteza geológica são relatados na literatura, nas etapas de ajuste dos contatos das unidades modeladas; mas nada foi encontrado sobre a incerteza geológica na base de dados em depósitos ígneos de fosfatos. Assim, se um algoritmo conseguir atingir tais objetivos, as simulações preditivas dos atributos de processo, utilizando ferramentas do aprendizado de máquina, podem ser avaliadas, aumentando a chance de sucesso do modelo geometalúrgico.

Para testar essa hipótese investigativa, foi necessário encontrar uma técnica capaz de mensurar o grau de pertencimento ou não pertencimento das litologias a um determinado domínio geológico, o que é um desafio para a maioria das técnicas de agrupamento existentes que não consideram a imprecisão dos *clusters*.

Acredita-se que, com o desenvolvimento dos algoritmos, a subjetividade no agrupamento da amostra e a dependência do conhecimento especialista para correção de descrições equivocadas sejam reduzidas, uma vez que a replicabilidade e a padronização

dos conceitos são mantidas. Bem como a redução do tempo, significativo em trabalhos de modelagem geológica e geometalúrgica.

1.3 Objetivos

Para atingir a meta proposta, são considerados os seguintes objetivos:

- Estudar um algoritmo de agrupamento capaz de quantificar a incerteza geológica.
- Realizar as estimativas espaciais dos domínios tipológicos por krigagem dos indicadores.
- Investigar técnicas de estatística multivariada e aprendizado de máquina para determinar qual ferramenta apresenta os melhores resultados na predição dos atributos geometalúrgicos.
- Estimar as variáveis preditoras do comportamento do minério utilizando ferramentas geoestatísticas.
- Reconciliar os resultados obtidos no modelo de blocos com os dados da usina.

1.4 Contribuição da tese

Esta tese pretende contribuir com a confecção do modelo geometalúrgico da Mina de Fosfato Chapadão por meio da interface multidisciplinar entre geociência, estatística e ciências da computação sem que informações adicionais tenham que ser coletadas, permitindo o aproveitamento máximo dos recursos materiais e do banco de dados que a empresa tem disponível. A intenção é que as ferramentas utilizadas facilitem o gerenciamento da variabilidade geológica e o entendimento da relação entre geologia e processamento mineral.

2 GEOMETALURGIA

A geometalurgia foi inicialmente difundida através da caracterização tecnológica. De acordo com Hoal (2008), Mcquiston e Bechaud já haviam discutido esse tema em 1968, focando na amostragem de depósitos minerais durante a etapa de pesquisa. Williams (2013) cita os primeiros trabalhos de caracterização tecnológica, ferramenta importante da geometalurgia, na década de 70; mas foi a partir dos anos 2000 que a geometalurgia passou a ser reconhecida como ciência até então não era identificada com este nome.

Segundo Kanh (1988), alguns trabalhos de caracterização tecnológica no Brasil foram realizados nas décadas de 70 e 80 por Paulo Abib Engenharia em jazidas de fosfatos, porém poucos estudos eram divulgados nessas épocas. Um aumento das pesquisas em geometurgia ocorreu nos anos 2000, com destaque em jazidas de minérios de ferro (Hsieh, 2005; Debrincat; Loo; Hutchens, 2004; Goldrinng 2003; Vieira *et al.*, 2003; Ottoni, 2002; Ishikawa *et al.*, 1983 *apud* Chemale *et al.*, 2013).

De acordo com Deutsch (2015), a geometurgia se consolidou devido à diminuição dos teores nas maiores reservas em lavra associada a um crescente aumento do conhecimento geológico e mineral. Estudos apontam que as características mineralógicas impactam na recuperação mássica e metalúrgica dos minérios (Braga; Costa, 2016). Portanto, a geometurgia tem um caráter multidisciplinar e exige um entendimento das informações geológicas, mineralógicas e metalúrgicas para construir um modelo espacial para a gestão da produção (Lamberg, 2011).

Uma boa avaliação de recursos via modelos tridimensionais com informações geológicas e metalúrgicas reduz significativamente as incertezas da planta. Dessa forma, durante o planejamento de lavra, fatores que poderiam afetar negativamente a concentração do minério podem ser previamente identificados e compreendidos para a otimização das atividades. Dunham e Vann (2007) sugerem que é possível tomar medidas durante a etapa de planejamento para administrar as consequências de um material com resposta metalúrgica inferior à esperada, por exemplo, adotando a blendagem do minério em casos nos quais há conhecimento de que este não atingiria a qualidade do concentrado desejada.

Os modelos geometúrgicos são necessários e podem ser inicialmente avaliados a partir de amostras de sondagem de exploração. Para Williams e Richardson (2004), a confecção desses modelos segue um fluxo que começa com a seleção de um número de amostras informadas pela geologia para ensaios que determinam parâmetros metalúrgicos dos minérios. A interpolação desses parâmetros por técnicas reconhecidas de geoestatística é utilizada para apoiar o modelo de processos.

2.1 Banco de dados

A estrutura de dados e a modelagem de dados em geometurgia dependem de qual abordagem geometúrgica é usada em uma mina. O processo de criação, manutenção e utilização de um modelo geometúrgico é chamado de programa geometúrgico (Figura 1). Dentro de um programa geometúrgico é definido o uso de

dados geológicos, a estratégia de amostragem, os métodos de teste e os resultados da simulação (Liebezeit *et al.*, 2011).

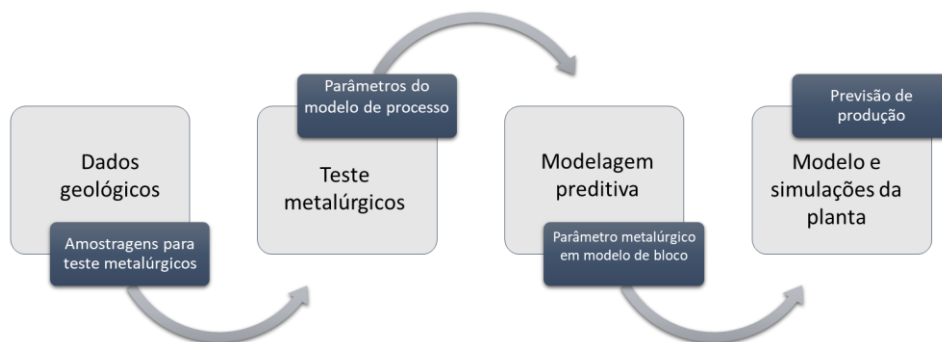


Figura 1. Exemplo das etapas de um programa geometalúrgico. Adaptado de P. Lamberg, 2011.

Entende-se que o levantamento do banco de dados e a seleção das amostras é o passo inicial e, por isso, o mais crítico do programa geometalúrgico (Williams; Richardson, 2004). Ademais, um desafio que um programa geometalúrgico precisa resolver a princípio é definir quantas amostras são necessárias para diferentes ensaios e quais são as técnicas envolvidas no desenvolvimento de um sistema geometalúrgico confiável. Se o número de amostras for muito pequeno, o modelo pode ser impreciso ou até mesmo defeituoso. Por certo, um número maior de amostras proporciona maior precisão de previsão, mas custos e tempo gastos em amostragem adicional e seus ensaios devem ser justificados (Lyshchuk, 2019).

Constata-se que algumas propriedades primárias, como conteúdo de metal (teores) e características geológicas qualitativas (litologia, alteração e estilo de mineralização) estão abundantemente disponíveis na maioria dos projetos de mineração. Os teores são propriedades fundamentais, pois medem metal contido, e possuem atributos valiosos: são quantitativos, contínuos e aditivos. Assim, as feições geológicas são qualitativas e sujeitas à interpretação dos geólogos (que podem ser reinterpretadas e variam com o tempo). Variáveis menos abundantes são as de caracterização mineralógica, uma vez que métodos de aquisição mais caros são necessários para obter informações da mineralogia, por exemplo, difração de raios-X (XRS) e microscopia eletrônica de varredura (Sepulveda *et al.*, 2018).

Destaca-se que o critério de amostragem adensada e os procedimentos confiáveis para obtenção dos teores de metal não são utilizados para incorporar as informações metalúrgicas ao modelo (Lozano; Bennet, 2003). Os ensaios e testes de processo são

relativamente caros, usam grandes volumes de amostras e requerem instalações de laboratório comerciais (Lozano; Bennet, 2003; Coward *et al.*, 2009).

As diferenças significativas entre o grande número de amostras registradas em bancos de dados geológicos e as relativamente poucas amostras dos bancos de dados de testes metalúrgicos dificultam a inferência das respostas metalúrgicas no modelo de recursos usando métodos geoestatísticos (Hunt *et al.*, 2013). Na atual abordagem geometalúrgica, a informação geológica continua sendo usada, inicialmente para classificação de domínios e depois avaliada criticamente em relação aos resultados dos testes geometalúrgicos a fim de suprir a carência dos ensaios de processo.

Outro ponto importante em relação ao banco de dados geometalúrgico é que muitas variáveis apresentam característica de não aditividade e não linearidade, o que pode exigir métodos de aumento da escala de laboratório (tamanho da amostra) para escalas de produção (Dunham; Vann, 2007) impondo ainda mais dificuldades para a modelagem geometalúrgica.

2.2 Variáveis geometalúrgicas

A relevância geometalúrgica para o planejamento de lavra e operação da planta de beneficiamento é baseada em reconhecer variáveis que contribuem para a valorização do recurso. Esses atributos compreendem a caracterização do material, dureza e moabilidade, liberação mineral, recuperação mássica e metalúrgica, consumo de reagentes, além dos atributos tradicionais, como o teor (Chibaya, 2013).

Uma discussão relacionada ao tratamento desses dados é a não aditividade e não linearidade de algumas variáveis. Segundo Carrasco *et al.* (2008), a aditividade é a propriedade que permite que a média de algumas variáveis seja calculada por meio de uma simples média linear, ou seja, a quantidade média é igual a média das quantidades, como exemplo dos teores médios de uma jazida. Nesse mesmo trabalho, os autores demonstram que a recuperação mássica é aditiva, mas a recuperação metalúrgica e o teor de concentrado não são. A não aditividade implícita das variáveis de resposta pode ser simplesmente ilustrada por meio da desigualdade de Jensen. Desse modo, quando a relação entre duas variáveis é não linear, uma média linear simples irá subestimar ou superestimar o valor verdadeiro.

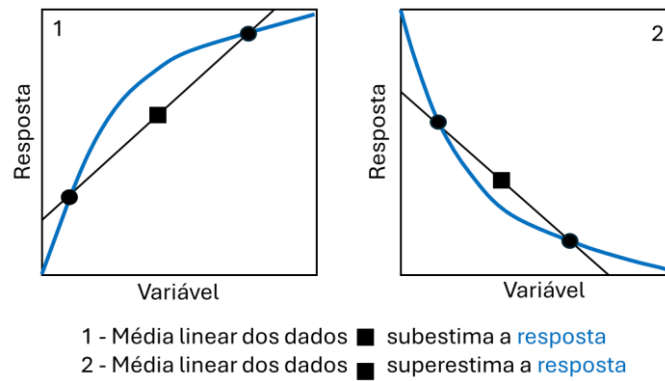


Figura 2. Desigualdade de Jensen (adaptado de Coward *et al.*, 2009).

As variáveis independentes e aditivas permitem o uso de métodos de estimativa lineares, como a krigagem ordinária e a krigagem simples (Matheron, 1963), pois podem ser calculadas linearmente e aumentadas em escala. Já para os atributos que apresentam características de não aditividade, como a maioria das variáveis de processo, a estimativa linear aplicada a esse tipo de variável pode resultar em médias tendenciosas (Dunham; Vann, 2007).

Por essa razão, outras ferramentas como geoestatística não linear e a simulação geoestatística vêm sendo empregadas para construção de modelos geometalúrgicos. Richmond e Shaw (2009) recomendam a utilização de krigagem dos indicadores para estimar variáveis geometalúrgicas não aditivas, estimando uma probabilidade ao invés da variável em questão. Coward e Dowd (2015) e Coward *et al.* (2009, 2013) sugerem que, quando possível, as propriedades primárias da rocha (atributos medidos diretamente nas amostras de rocha) sejam usadas para prever propriedades de resposta de processamento, como recuperação e desempenho de cominuição. Assim, usar variáveis primárias aditivas para prever variáveis de resposta não aditivas minimiza os vieses relacionados à não aditividade (Carrasco *et al.*, 2008; Newton e Graham, 2001) e resolve o problema do número reduzido de variáveis metalúrgicas recorrentes na maioria dos bancos de dados mineiros. Ademais, Deutsch (2013) sugere utilizar simulação para prever as propriedades metalúrgicas, visto que suposições de linearidade da variável não são assumidas neste caso e a simulação ainda tem a vantagem de avaliar a incerteza dos atributos, o que é importante para informar a variabilidade à planta de beneficiamento. Esse mesmo autor apresenta alternativas, como a criação de um modelo numérico ou função de transferência capaz de correlacionar as propriedades da rocha (variáveis primárias) com as variáveis secundárias. Boisvert *et al.* (2013) utilizaram a simulação sequencial gaussiana combinada com componentes principais, análise de regressão e

outras técnicas de geoestatística multivariada para lidar com os atributos metalúrgicos na mina australiana de Olympic Dam. Outros artigos e pesquisas utilizando as alternativas citadas acima para variáveis geometalúrgica podem ser vistos em Newton e Graham, (2011), Hunt *et al.*, (2013); Deutsch, (2015); Coward, (2015); Sepúlveda, (2018); Fernandes (2013); Motta, (2014); Braga, (2015); Vieira (2016); Braga Jr., (2017); Batelochi, (2018); Drumond (2019); Niquini, (2020).

Existe uma série de procedimentos e técnicas na literatura para estimativa de variáveis metalúrgicas, sejam aditivas ou não aditivas. Cabe aos geoestatísticos ou geomodeladores determinarem qual resposta se deseja obter ao elaborar o modelo em busca de informações coesas com o banco de dados e a realidade observada em cada depósito estudado.

2.3 Modelagem espacial e geoestatística

Para construir modelos tridimensionais, a estrutura clássica da geoestatística precisa ser adotada. Um pré-requisito para modelagem geoestatística é o estabelecimento de relações estatísticas representativas. As relações estatísticas univariadas são definidas nos histogramas e modelos de distribuição; nas estatísticas espaciais bivariadas na forma de funções de covariância ou variogramas; e nas estatísticas multivariadas, incluindo matrizes de correlação (Deutsch, 2015).

Essas relações estatísticas são calculadas em um domínio espacial estacionário para cada variável aleatória. As variáveis aleatórias podem ser de natureza contínua ou categórica. Uma variável contínua, como concentrações de metal ou parâmetros metalúrgicos, é medida em uma escala quantitativa. Por outro lado, uma variável categórica refere-se a condições de estado geralmente não ordenadas, como tipos de rocha ou solo, conforme Goovaerts (1997). Uma sequência simplificada para elaboração dos modelos 3D pode ser dada pela definição dos domínios espaciais, seguida da interpolação dos atributos categóricos ou contínuos, utilizando ferramentas geoestatísticas.

2.3.1 Definição de domínios

A elaboração dos modelos de recursos ocorre após as etapas de amostragem e ensaios de laboratórios e necessita de dados suficientes para permitir uma inferência viável das estatísticas necessárias (Larrondo; Deutsch, 2004). Segundo Larrondo e Deutsch (2004), as técnicas geoestatísticas mais comuns são baseadas em fortes suposições de estacionaridade de segunda ordem, ou seja, a média, a variância e a

covariância dos atributos permanecem constantes em todo o domínio e não dependem da localização dos pontos, mas apenas da distância entre eles. Portanto, a definição dos domínios é uma das primeiras tarefas na modelagem espacial (Braga; Costa, 2016).

A modelagem dos domínios tem sido restrita aos grupos geológicos com base na litologia e teor. Os domínios metalúrgicos são modelados com menos frequência, dada a quantidade inferior de informação. Vale ressaltar que modelos 3D baseados em unidades geológicas, podem ou não ser estatisticamente homogêneos para os parâmetros de processo (Dominy *et al.*, 2018). Entende-se que o agrupamento geometalúrgico é semelhante ao agrupamento geológico, mas se concentra nas características geometalúrgicas do minério para fornecer uma base para a otimização integrada da mineração ao processamento. Em ambos os casos é necessário separar as amostras e definir os limites entre os domínios.

Um aspecto a ser pontuado em relação à delimitação dos domínios é o fato de ser comum na literatura encontrar referência aos limites entre as unidades geológicas ou corpos de minérios como “*hard*” ou “*soft*” (Larrondo; Deutsch, 2004), mas geralmente, sobreposições de estruturas geológicas ocorrem durante e após o processo de formação dos depósitos, agregando complexidade às mineralizações e conseqüentemente ao processo de concentração.

Da perspectiva de sistemas complexos, os processos responsáveis pela formação de depósitos minerais são multiescalares (escala regional a microscópica) e multivariados (uma interação complexa de processos dinâmicos crustais e superficiais) (Frimmel; Nwaila, 2020), ou seja, uma descrição exata e uma amostragem completa de um sistema geológico são impossíveis em todas as escalas de mapeamento. A exemplo da Mina Chapadão, na qual são observadas estruturas transitórias resultantes do processo geológico de formação da jazida associado à alteração intempérica.

Como consequência dessa variabilidade dos eventos geológicos comuns nos depósitos minerais, uma busca por técnicas computacionais de estatística multivariada e aprendizado de máquina estão sendo empregadas para auxiliar nos agrupamentos e na determinação da imprecisão dos domínios de estimativa.

Vale ressaltar que a precisão e qualidade das estimativas do modelo de recursos depende da solidez do conhecimento geológico (variáveis categóricas) e da integridade dos dados de amostragem (variáveis contínuas) (Chibaya, 2013).

2.3.2 Krigagem dos indicadores

As variáveis categóricas que expressam as informações geológicas são frequentemente modeladas por métodos tradicionais que ainda são os métodos mais empregados na mineração. No entanto, esses métodos convencionais, conhecidos como modelagem explícita, apresentam algumas limitações quanto ao tempo de construção do modelo, já que a interpretação é feita de forma manual, e influenciada pela gênese dos depósitos minerais. Dessa forma, quanto mais complexo é um depósito, mais laboriosa é a modelagem para o intérprete ou modelador e, muitas vezes, essas interpretações são estabelecidas de maneira subjetiva, contando com a experiência do especialista.

As metodologias usadas para modelar domínios além da interpretação manual de seções e construção de sólidos são: estimativa por vizinho mais próximo e métodos geoestatísticos avançados como a krigagem da indicatriz, simulação sequencial dos indicadores, geoestatística multiponto e algoritmos que utilizam funções de base radial (RBFs).

De acordo com Newton e Graham (2011), nos dados avaliados, a krigagem da indicatriz apresentou melhores resultados quando comparada ao vizinho mais próximo. Já a interpretação de seções mostrou-se uma tarefa trabalhosa e demorada.

Em trabalhos anteriores de modelagem de domínios na Mina Chapadão, a krigagem dos indicadores apresentou bons resultados, sendo utilizada neste estudo devido a sua consistência e ampla utilização. A krigagem dos indicadores (KI) se baseia na construção de uma função de distribuição de probabilidades acumuladas por meio de uma transformação não linear (Journel, 1983). A interpolação não linear inclui métodos para estimar a expectativa condicional, e ainda a distribuição condicional de um valor em um local, ao invés de simplesmente prever o próprio valor.

A estimativa tem início com a transformação da variável original em variável indicadora que revela se a variável transformada pertence a uma dada classe de valores. Se a variável pertence àquela classe de valores, a sua correspondente variável indicadora recebe o valor um (1), e em caso contrário, o valor zero (0). Para a interpolação de dados categóricos em modelagem de depósitos minerais, a KI irá produzir a probabilidade de um local estar associado a um determinado tipo de rocha ou domínio e, assim, delineará os contatos geológicos do depósito (Sinclair; Blackwell, 2004).

Soares (2006) define a variável indicatriz categórica $I(x)$ como uma variável que pode assumir dois valores possíveis no espaço x , em um sistema binário composto por dois corpos, X e X_c , dentro de uma determinada área A , ou seja:

$$I(x) = \begin{cases} 1 & \text{se } x \in X \\ 0 & \text{se } x \in Xc, \end{cases} \quad \text{sendo } A = X \cup Xc \quad (1)$$

Assim, a variável categórica $I(x)$ é transformada por meio de codificação binária (0 e 1) e tem probabilidade de pertencer aos corpos X ou Xc , o que garante que os eventos sejam mutuamente exclusivos. Os variogramas para a variável categórica $I(x)$ devem ser construídos e, conhecendo os variogramas, inicia-se o processo de interpolação das probabilidades. O variograma é uma função que caracteriza o grau de continuidade espacial de um fenômeno, em que no eixo horizontal do gráfico plotam-se distâncias, enquanto no eixo vertical as meias-médias das diferenças quadráticas dos valores são separadas por um vetor h . Em se tratando de indicadores, a equação do variograma pode ser descrita (Goovaerts, 1997):

$$\gamma(h;sk) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} (i(u_{\alpha};s_k) - i(u_{\alpha}+h;s_k))^2 \quad (2)$$

As aplicações de variações em torno da krigagem dos indicadores são apresentadas por Goovaerts (1997), como a krigagem simples dos indicadores, krigagem ordinária dos indicadores, krigagem dos indicadores da mediana e co-krigagem dos indicadores. Entre as mais citadas estão a krigagem simples (KS) dos indicadores e a krigagem ordinária (KO). Na KS, o conceito de média conhecida e constante em todo o domínio que realiza uma estimativa por krigagem simples se mantém como definido na equação 3:

$$p_{SK}^*(u, k) = p_k + \sum_{\alpha=1}^n l_{\alpha}^{SK}(u, k) \cdot [i(u_{\alpha}, k) - p_k] \quad (3)$$

em que $p_{SK}^*(u, k)$ é a estimativa da probabilidade da categoria k prevalecer no local; u , $l_{\alpha}^{SK}(u, k)$ é o peso de krigagem simples para o indicador; $i(u_{\alpha}, k)$, p_k é a probabilidade média global desagrupada para cada categoria; e n é o número de dados usados para a estimativa (Goovaerts, 1997).

Na krigagem ordinária, os conceitos sobre as flutuações locais são considerados sobre a média dos indicadores, limitando a hipótese de estacionariedade a uma vizinhança local (Goovaerts, 1997):

$$p_{OK}^*(u, k) = \sum_{\alpha=1}^n l_{\alpha}^{OK}(u, k) \cdot i(u_{\alpha}, k) \quad (4)$$

em que $p^{*OK}(u, k)$ é a estimativa da KO da probabilidade da categoria k prevalecer no local; u , $l_{\alpha}^{OK}(u, k)$ é o peso de krigagem ordinária para o indicador $i(u_{\alpha}, k)$; e n é o número de dados utilizados na estimativa (Goovaerts, 1997).

Tanto a krigagem simples quanto a krigagem ordinária dos indicadores são bons interpoladores e irão fornecer um valor resultante entre o intervalo 0 e 1 para cada local ou ponto estimado, e as probabilidades devem somar 1 (Glacken; Blackney, 1998), mas essas condições nem sempre são satisfeitas se as K categorias são estimadas de maneira independente. Esse problema é conhecido como desvio de relação de ordem. Por isso, para contornar essa situação Goovaerts (1997) propõe corrigir as estimativas já realizadas: substituindo o valor fora do intervalo $[0,1]$ pelo valor de fronteira mais próximo, ou dividindo as estimadas pela soma das probabilidades. Outra possibilidade é utilizar um mesmo variograma e uma mesma estratégia de busca para a estimativa de todos os indicadores das K classes. A utilização de um mesmo variograma e estratégia de busca faz com que os mesmos pesos de krigagem sejam utilizados nas estimativas das K classes. Portanto, o resultado é que a soma das probabilidades estimadas fica sendo 1. Porém, para a estimativa de muitas classes com características espaciais distintas, essa alternativa pode não ser viável.

Uma vez estimadas as probabilidades e verificados os desvios de relação de ordem, deve-se retornar com o tipo (X ou X_c) correspondente ao ponto não amostrado, dado pelo valor mais provável (Teng; Koike, 2007, *apud* Yamamoto; Landim, 2013). É importante mencionar que essa tarefa nem sempre é simples quando não há uma categoria cuja probabilidade é bastante superior às demais. Por exemplo, em casos de modelagem com 5, ou mais categorias, as probabilidades divididas entre as cinco categorias podem levar a uma situação em que a categoria definida pelo maior valor apresenta uma probabilidade menor que a soma das estimativas das outras categorias, assim, o bloco tem maior chance de não ser a categoria de maior valor; cabendo uma discussão mais ampla sobre assunto.

2.3.3 Simulação sequencial gaussiana

Os modelos numéricos, relacionadas à quantificação das variáveis contínuas de um depósito mineral, podem ser amplamente classificados como determinísticos, onde há um único modelo do depósito; ou estocásticos, onde há múltiplas realizações do depósito. De uma perspectiva estatística, eles correspondem a modelos estimados e simulados do depósito, e os métodos mais comumente utilizados são a krigagem ordinária e a simulação

estocástica. O propósito do modelo e os objetivos do estudo determinarão se o modelo construído será determinístico com uma única melhor estimativa ou estocástico com múltiplas realizações do atributo (Deutsch, 2015).

Por isso, se o objetivo do estudo é inferir uma única melhor estimativa e minimizar os erros locais de um conjunto de propriedades de rochas em um domínio para estimativa e classificação de recursos, uma abordagem determinística como a krigagem ordinária pode ser usada, pois fornece um método robusto para a estimativa espacial de propriedades de rochas em um domínio geológico (Rossi; Deutsch, 2014). Esse método apresenta os melhores modelos lineares não enviesados, pois o estimador tenta aproximar a média residual de zero ($m_r = 0$) e é considerado o melhor modelo, uma vez que a variância dos erros (σ^2_R) é a menor possível. Nota-se, então, que a minimização da variância dos erros acaba por suavizar os resultados e, portanto, caso o objetivo seja modelar corretamente a variabilidade do fenômeno, o método mais indicado é a simulação estocástica (Olea, 2012).

Geralmente, em modelagem geometalúrgica, o propósito é informar a taxa de variação na alimentação da planta e, para quantificar essa incerteza, a simulação estocástica tem um papel importante, porque, além de informar a variabilidade dos minérios, as suposições de linearidade das variáveis não são assumidas (Deutsch, 2015). Ou seja, é uma técnica capaz de lidar com as dificuldades relacionadas às características das variáveis metalúrgicas.

Existem diversas técnicas de simulações divididas em dois tipos: as sequenciais e as espectrais. O método de simulação sequencial gaussiana (SGS), apresentado por Isaaks e Srivastava (1989), é o mais utilizado na modelagem de recursos minerais por sua simplicidade, flexibilidade e razoável eficiência (Deutsch, 2002; Soares, 2001). O procedimento da simulação sequencial gaussiana (SGS) foi descrito por Deutsch, em 1992. Todo o processo deste tipo de simulação é desenvolvido em ambiente gaussiano e admite a hipótese de multigaussianidade (formalismo multigaussiano) para a variável que se pretende simular (Soares, 2006).

Uma das maneiras de validar a multinormalidade do modelo gaussiano é checando se o valor 0.564 é reproduzido pela relação entre o madograma e raiz quadrada do variograma (Matheron, 1982):

$$\frac{\gamma_1(h)}{\sqrt{\gamma_2(h)}} = 0.546, \quad (5)$$

sendo:

$\gamma_1(h)$ o mandograma normalizado da variável analisada.

$\gamma_2(h)$ o variograma desta variável.

O algoritmo da simulação sequencial inicia considerando N variáveis aleatórias $\{Z(u_j), j = 1, N\}$. O objetivo da simulação sequencial é a geração das várias realizações conjuntas dessas N variáveis aleatórias: $\{z^l(u_j), j = 1, N\}, l = 1, L$ condicionadas ao conjunto de dados $\{z(u_\alpha), \alpha = 1, n\}$. Segundo Goovaerts (1997), a simulação pode ser condicional quando passa exatamente pelos pontos amostrais ou condicionantes. A simulação condicional conjunta de z em somente dois pontos u_1 e u_2 , na qual se obtém um conjunto de pares de realizações geradas por amostragem da função de distribuição acumulada condicional bivariada (Goovaerts, 1997) é dada pela função:

$$F(u_1, u_2; z_1, z_2 | (n)) = P_{rob}\{Z(u_1) \leq z_1, Z(u_2) \leq z_2 | (n)\} \quad (6)$$

ou seja, o valor $z^l(u_1)$ simulado com base na função de distribuição acumulada condicional $F(u_1; z_1 | (n))$ que pode ser atualizada pelo valor previamente simulado $z^l(u_1)$ dos n pontos de dados. A equação 6 pode ser generalizada para N variáveis:

$$F_N(u_1, \dots, u_N; z_1, \dots, z_N | (n)) = P_{rob}\{Z(u_i) \leq z_i, i = 1, N | (n)\} \quad (7)$$

e pode ser aproximada como produto de N funções de distribuição acumulada condicional, que são determinadas sequencialmente.

$$\begin{aligned} F(u_1, \dots, u_N; z_1, \dots, z_N | (n)) &= F(u_N; z_N | (n + N - 1)) \\ &\quad \cdot F(u_{N-1}; z_{N-1} | (n + N - 2)) \dots \\ &\quad \cdot F(u_2; z_2 | (n + 1)) \cdot F(u_1; z_1 | (n)) \end{aligned} \quad (8)$$

A fundamentação teórica mostra que cada novo ponto simulado é usado para atualizar a função de distribuição acumulada condicional. Como os valores simulados são utilizados para simular um novo nó, é importante que o processo de simulação sequencial utilize de um caminho aleatório sem repetição (*random walk*), pois o caminho aleatório é o processo estocástico com o menor efeito no modelo final conforme o aumento das simulações realizadas.

Pode-se resumir a realização da SGS conforme os seguintes passos (Goovaerts, 1997):

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

$$Z^l(u_j) = \varphi^{-1} \left(y^l(u_j) \right), \quad j = 1, N \quad (9)$$

Segundo Koltermann e Gorelick (1996), o método de simulação sequencial gaussiana (SGS) é o mais poderoso dos algoritmos gaussianos de geração de campos aleatórios. Sabe-se que existem algumas limitações da abordagem de simulação sequencial gaussiana, principalmente em relação aos requisitos para gaussianidade multivariada e entropia máxima, levando a valores extremos desconectados e desestruturados (Deutsch, 2015). Em algumas aplicações, as variáveis de interesse podem ser simuladas independentemente se não existir correlações entre elas. Caso contrário é necessário utilizar um método estocástico de variabilidade espacial conjunta por um modelo prévio de covariâncias e covariâncias cruzadas, os quais requerem uma solução numérica intensiva, exigindo muito processamento computacional dos sistemas de krigagem. Uma alternativa é a descorrelação das variáveis que são transformadas em fatores independentes.

Para verificar se existe correlação entre as múltiplas variáveis em um determinado banco de dados gráficos do tipo q-q plot, e matriz de correlação são métodos utilizados para avaliar a correlação entre variáveis (Isaaks; Srivastava, 1989).

2.4 Incerteza dos modelos geometalúrgicos

A cadeia de valor completa de um projeto de mineração está sujeita a incertezas que complicam o processo de tomada de decisão. Qualitativamente, sabe-se o que se quer

dizer com “incerteza” e suas conotações que giram em torno do reconhecimento de que o valor único que se relata é, em algum sentido, apenas um palpite razoável ou útil sobre o que o valor desconhecido pode ser (Isaaks; Srivastava, 1989). A incerteza está presente em modelos de recursos, nos processos da lavra e no tratamento mineral, sendo estes considerados fatores técnicos, mas, outras fontes de incertezas estão presentes nos fatores financeiros e ambientais. Tais incertezas na geologia e no processamento mineral podem ser classificadas como incertezas geometalúrgicas (Sepulveda *et al.*, 2018).

Observa-se que a maioria dos desafios relacionados à incerteza dos projetos de mineração podem ser atribuídos à tomada de decisão com base na interpretação e estimativa das informações de geociências. Segundo Berry *et al.* (2006), a entrega de projetos de mineração é insatisfatória quando ocorrem problemas como: produção menor do que o esperado, desafios geotécnicos, desempenho metalúrgico abaixo do esperado e custos de capital ou operacionais não orçados. Esses autores realizaram uma pesquisa em 105 projetos de mineração sem sucesso e o estudo ilustrou que a principal causa das falhas técnicas são os planos e o sequenciamento de lavra com uma contribuição de 32%; o risco metalúrgico (erros de projeto da planta e viés nos testes metalúrgicos) é a segunda maior causa de erros, responsável por 27%; e o terceiro contribuinte para as falhas dos projetos foram as estimativas de recursos e reservas, em torno de 17%. Berry *et al.* (2006), destacou que embora o planejamento de lavra tenha tido a maior contribuição na pesquisa realizada, alguns elementos da estimativa geológica podem causar desafios ao planejamento e à lavra, o que implica em sobreposição de erros em todas as subclasses determinadas no estudo.

Um primeiro passo na avaliação da incerteza na estimativa dos modelos geológicos é considerar os fatores que influenciam o erro. Um fator óbvio é o número de amostras e a distribuição espacial das amostras. As amostras próximas ajudam a tornar a estimativa mais confiável. Por isso, se as amostras adicionais estiverem extremamente próximas da amostra existente, a confiança não aumentará tanto quanto aumentaria se as amostras adicionais estivessem mais uniformemente distribuídas em torno do ponto que se tenta estimar (Isaaks; Srivastava, 1989).

O segundo fator que complica a questão da incerteza é a natureza do fenômeno em estudo. Sabe-se que cada unidade geológica é um volume irregular com características distintas, marcadas pelos limites ou contatos entre unidades. Estes contatos são determinados por descontinuidades complexas geradas por erosões, falhas e diferentes processos deposicionais (Amarante, 2018). Ao lidar com uma variável extremamente

suave e bem-comportada, as estimativas serão mais confiáveis do que ao lidar com depósitos geológicos complexos e com variável muito errática (Isaaks; Srivastava, 1989). Assim, a incerteza associada à delimitação dos domínios impacta diretamente no volume a ser estimado na avaliação de recursos, e a diferença entre a estimativa e o valor recuperado tem impacto significativo na lucratividade do empreendimento mineiro (Amarante, 2018).

Para compensar a incerteza relacionada à natureza dos fenômenos geológicos e geoestatísticos, recorre-se a diversas técnicas. Em alguns casos, busca-se a definição de domínios com base na mudança do teor médio local, que geralmente é gradacional em vez de abrupto, estimando, assim, variáveis neste tipo de domínio por uma análise de vizinhança móvel. Em outros casos, usa-se krigagem de indicadores probabilísticos para definir limites de zonas de transição (Zhang *et al.*, 2023). Técnicas estocásticas baseadas em simulação de indicadores, simulação gaussiana truncada e simulação multipontos (MPS), também, permitem acessar a incerteza relacionada ao modelo geológico, porém demandam grande esforço computacional. Outros métodos matemáticos com foco na geometria incluem interpolação de funções de distância assinaladas de base radial (RBF; Cowan *et al.*, 2003), e diferentes técnicas de agrupamento.

Já para quantificar a incerteza dos atributos contínuos de um modelo de recursos, a simulação geoestatística é uma ferramenta poderosa para superar os problemas de viés das abordagens convencionais, como a krigagem, sendo capaz de mensurar o risco das variáveis primárias (Deutsch, 2015). Quanto às variáveis de processo, se forem obtidas por modelos preditivos das variáveis primárias, a estrutura de resposta primária permite a quantificação de incertezas geometalúrgicas (Coward *et al.*, 2009).

Portanto, muitas são as fontes de incertezas no modelo de recursos, sendo importante a utilização de técnicas que possam mensurar a confiabilidade através de uma medida de variabilidade. Isso deve ser feito em todas as etapas, desde a descrição geológica e quantificação dos teores nas amostras de sondagem, passando aos métodos de delimitação de domínios e estimativa dos atributos. Essa abordagem é muito extensa constituindo por si um tema de pesquisa. Nesta tese, em determinadas etapas, será possível tratar e discutir algumas medidas de incerteza.

2.5 Avanços da geometalurgia

Estudos de caso com foco em geometalurgia, particularmente aqueles que destacam os vínculos com a caracterização geológica, estão aumentando na literatura.

Atualmente, a lavra de minérios exige melhor conhecimento dos depósitos, diferente do que era aceito no passado. Isso se deve, em grande parte, ao fato de haver a priorização da lavra ao longo dos anos em zonas mais homogêneas e mais ricas, restando níveis mais complexos de mineralização.

Como visto no item 2.0, os primeiros trabalhos de geometurgia surgiram na década de 70, mas, de acordo com Hunt *et al.* (2019), um aumento significativo na pesquisa geometúrgica ocorreu após 2005. E, semelhante ao grande projeto de modelagem de minérios e estudos geometúrgicos GeM, financiados pelo Australian Research Council, outros projetos ganharam espaço e outras instituições de pesquisas foram envolvidas.

em 2005, a pesquisa focada em geometurgia também estava ganhando espaço em outras instituições, como: Cambourne School of Mines (Reino Unido), COREM (Canadá), University of Cape Town (África do Sul) e Colorado School of Mines (EUA). A primeira conferência dedicada inteiramente à geometurgia ocorreu em 2011 e foi seguida por muitas outras (Australian Institute of Mining and Metallurgy (AusIMM); Gecamin; Institute of Materials, Minerals and Mining (IOM3); South African Institute of Mining and Metallurgy (SAIMM) (Hunt *et al.*, 2019, p.1475).

Ainda segundo Hunt *et al.* (2019), essas conferências não só levantaram a discussão em torno do tema geometurgia como levou a um rápido aumento na disponibilidade de artigos, citando:

Isso inclui modelos usados para dados de geometurgia que variam de domínio simples (Burger *et al.*, 2006) a árvores de decisão (Lishchuk *et al.*, 2019). Escobar e Jara (2012) descrevem desafios no desenvolvimento de um modelo geometúrgico para lixiviação de pilha de óxidos. Artigos que ilustram como a geometurgia pode melhorar o valor presente líquido (VPL) de um depósito em Dunham *et al.* (2011), Bye (2011), Wolff *et al.* (2012), King e MacDonald (2016) e Lotter *et al.* (2018). (Hunt *et al.*, 2019, p.1475).

Essa tendência pela busca por modelos geometúrgicos também é uma realidade nos depósitos de fosfatos no Brasil, e pode ser observada nos trabalhos de Cordeiro (2008), que traz uma abordagem sobre caracterização mineralógica e metalúrgica aplicada aos controles de lavra em Catalão I, descrevendo detalhadamente a mineralogia do depósito; Uliana (2010), que realiza uma análise da composição mineralógica por DRX e MEV/EDS e caracteriza o minério residual de Salitre - MG; Fernandes (2013) utiliza de métodos de extrapolação de regressão múltipla para construção do modelo

geometalúrgico de Tapira - MG; Braga (2015) utiliza a krigagem dos indicadores para modelagem de domínios geometalúrgicos na mina F4, Araxá – MG; Grasso (2015) emprega análises de principais componentes e regressão linear na predição metalúrgica no horizonte residual de Araxá; Braga Jr. (2017) prevê recuperação mássica no modelo de recursos, usando simulação conjunta colocada na mina de nióbio em Araxá - MG; Avelar (2018) avalia o impacto das substituições e dos parâmetros cristalinos das apatitas de Catalão I no processo de flotação, via métodos analíticos de caracterização; Pereira (2019) viabiliza predição metalúrgica inserindo dados faltantes para estimativa dos atributos por regressão de Esperança Condicional Alternada (Alternating Conditional Expectation - ACE); Teixeira (2019) realiza testes de flotação e caracterização tecnológica nos materiais cimentados na mina de fosfato em Catalão-GO; e Ramos (2021) caracteriza os minérios de nióbio e os materiais de alimentação na usina de beneficiamento.

3 ANÁLISE MULTIVARIADA E APRENDIZADO DE MÁQUINA

3.1 Análise exploratória e modelos de distribuição

Como parte inicial da modelagem de recursos, ou de qualquer análise mais detalhada de um depósito mineral, o pré-processamento dos dados e a análise exploratória consiste na preparação, organização e estruturação do banco de dados, de forma que se tenha uma visão global da variação dos valores, bem como a descrição dos dados por meio de tabelas e gráficos. Essas tabelas e gráficos vão expressar, de forma simples, características, como medidas de tendência central, medidas de dispersão, forma e modelo da distribuição de probabilidade, correlação simples, relações entre os grupos de variáveis e uma variedade de enunciados probabilísticos relacionados a alguns desses temas (Sinclair; Blackwell, 2002). Os modelos de distribuição de probabilidades têm um papel importante, pois algumas das ferramentas de estimativa funcionam melhor se a distribuição dos valores dos dados estiver próxima de uma distribuição normal, visto que a distribuição gaussiana é uma das muitas distribuições para as quais existe uma descrição matemática concisa (Isaaks; Srivastava, 1989). Os modelos de distribuição de probabilidade normal (gaussiana) dos dados são considerados como o modelo teórico mais comum, e muitos fenômenos na natureza seguem distribuições gaussianas.

Nota-se que qualquer distribuição pode sofrer uma transformação gaussiana $X' = G^{-1}(F(X))$, em que G^{-1} corresponde à inversa da função de distribuição acumulada

gaussiana; e $F(X)$ corresponde à função de distribuição acumulada da variável aleatória X (Figura 3).

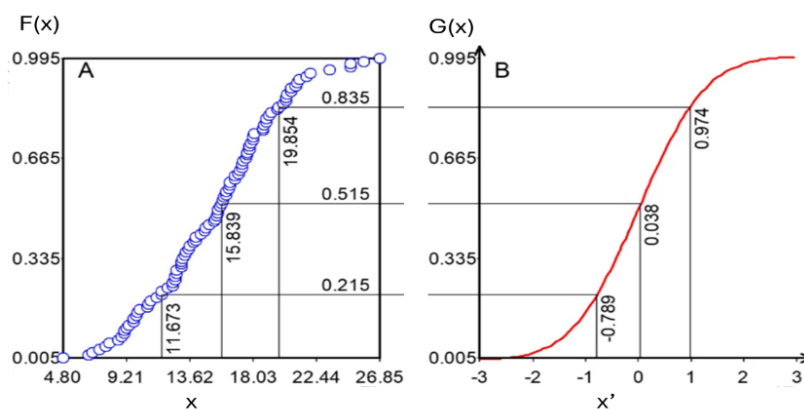


Figura 3. Procedimento gráfico para transformação da função de distribuição dos dados originais x em valores de uma distribuição normal x' . Modificado de Yamamoto, 2020.

As características importantes de um modelo de distribuição são o seu valor central, suas medidas de dispersão e simetria (Goovaerts, 1997). As estatísticas do primeiro grupo (média, mediana e moda) dão informações sobre onde está o centro da distribuição. A localização de outras partes da distribuição pode ser dada por valores que separam os dados ordenados em quatro (quartis), dez (decis) ou em cem (percentis) partes iguais. Muitos trabalhos encontrados na literatura adotam as medidas de localização para expressar as informações dos dados: o quantil inferior Q_1 (valor situado de tal modo na série de dados que 25% das observações são menores do que ele); a mediana (divide os dados ao meio); e o quantil superior Q_3 (valor situado de tal modo na série de dados que 75% das observações são menores do que ele). Assim, as estatísticas do segundo grupo incluem a variância, o desvio padrão e o coeficiente de variação. Eles são usados para descrever variabilidade dos valores dos dados (Isaaks; Srivastava, 1989).

A forma de distribuição de probabilidade é descrita pelo coeficiente de assimetria que fornece informações sobre a simetria, terceiro grupo para análise das características dos dados. O coeficiente de variação também fornece informações sobre o comprimento da cauda para certos tipos de distribuições. Na simulação dos teores, o coeficiente de variação é uma medida importante para a previsão das oscilações metalúrgicas e pode ser usado como uma medida de incerteza.

Tomadas em conjunto, todas essas características fornecem um resumo valioso da informação contida nos gráficos de probabilidade ou histogramas dos dados (Isaaks; Srivastava, 1989).

3.2 Coeficiente de correlação

Nos trabalhos que envolvem o tratamento de mais de uma variável, deseja-se, muitas vezes, conhecer o grau de relação existente entre elas (Triola, 2008). Uma forma de avaliar esta relação é por intermédio do cálculo da covariância que se constitui no produto das dispersões das variáveis em relação a uma média probabilidade comum. A divisão da covariância pelo produto dos desvios padrão de cada variável fornece o coeficiente de correlação linear (r) de Pearson. O coeficiente de correlação de Pearson é uma medida do grau da relação linear entre duas variáveis quantitativas. Este coeficiente varia entre os valores -1 e 1. O valor 0 (zero) sinaliza que não há relação linear, o valor 1 indica uma relação linear perfeita e o valor -1 também indica uma relação linear perfeita, mas inversa, ou seja, quando uma das variáveis aumenta a outra diminui. Dessa forma, quanto mais próximo estiver de 1 ou -1, mais forte é a associação linear entre as duas variáveis.

Por conveniência, rotulam-se duas medidas de X e Y . Associadas a essas duas variáveis estão duas quantidades que desempenham um papel importante na estatística: a covariância e a correlação de Pearson (Welcox, 2003). A covariância entre X e Y é:

$$\sigma_{xy} = E[(X - \mu_x)(Y - \mu_y)] \quad (10)$$

Em palavras, se para alguma população de indivíduos subtraímos a média de X de todos os valores possíveis para X e fazemos o mesmo para Y , então a covariância entre X e Y é definida como a média dos produtos dessas diferenças. Observe que a covariância de X consigo mesmo é apenas sua variância, e o mesmo vale para Y . Ou seja, a ideia de covariância generaliza a noção de variância para duas variáveis. A correlação de Pearson é a covariância dividida pelo produto dos desvios padrão e é normalmente rotulada como ρ (Welcox, 2003). Ou seja,

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad (11)$$

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. Isso é muito útil para construir modelos geometalúrgicos e minimizar os problemas relacionados às variáveis não aditivas e não lineares. Logo, o coeficiente de correlação é o primeiro indicativo de quais variáveis podem ser utilizadas para construir um modelo de regressão explicativo (Niquini, 2020).

3.3 Análise de componentes principais - PCA

A Análise de Componentes Principais (PCA) é uma ferramenta poderosa utilizada em estatística multivariada para a redução do número de variáveis, pois a redução da dimensão dos dados facilita as análises em todas as etapas e fornece uma visão estatisticamente privilegiada do conjunto de dados. Na mineração, muitos trabalhos atuais estão sendo feitos usando PCA associados a outras técnicas de *machine learning* com as mais diversas abordagens (Castro; Silva 2024). Na análise multivariada, quando muitas variáveis, em média 10 ou mais, estão disponíveis, os resultados geralmente são pouco alterados se só um subconjunto das variáveis for usado; as demais variáveis são, de certo modo, redundantes e podem ser descartadas (Jolliffe, 1972). Alguns estudos sobre agrupamento de domínios mostraram que um número reduzido de variáveis é mais efetivo no agrupamento (Fogliatto; Anzanello, 2011).

Segundo Jolliffe (1972), a ideia principal do PCA é reduzir a dimensionalidade do problema conservando o máximo da variabilidade. Isso pode ser feito por decomposição em autovalores (valores próprios) de uma matriz de covariância, geralmente depois de centralizar (e normalizar ou usar pontuações-Z) a matriz de dados para cada atributo. Abaixo há uma descrição sobre redução de dimensionalidade segundo Lyra *et al.*, (2010).

A realização da PCA consiste em fatorar a matriz de dados X , de modo que $X=TL^T+E$, em que L é a matriz dos pesos; T a matriz dos escores; e E a matriz dos resíduos. O símbolo T (T sobrescrito) é o operador de transposição de matriz. A primeira componente principal é $PC1= t_1l_1^T$, que é a melhor aproximação de posto 1 para X e corresponde à direção de maior variância no espaço multivariado. $E_1=X - t_1l_1^T$ é o resíduo de X , descontado PC1. A segunda componente principal é $PC2=t_2l_2^T$, que é a melhor aproximação de posto 1 para E_1 e corresponde à direção de maior variância no espaço multivariado não modelada por PC1, ou seja, ortogonal a ela. $E_2 =E_1 - t_2l_2^T$ é o resíduo deixado por PC1 e PC2. As componentes subsequentes modelam sempre a direção de maior variância no espaço multidimensional não modelado pelas PCs anteriores e são sempre ortogonais a todas elas. É possível realizar uma truncagem na sequência das

componentes principais mantendo apenas um número pequeno de PCs que já respondem por uma parcela significativa da informação total contida na estrutura de dados.

Na prática, para fazer uma análise de componentes principais, calcula-se inicialmente a matriz de covariância, C , para dados centrados na média, ou a matriz de correlação, R , para dados auto escalonados.

$$C(\text{ou } R) = \frac{E_0^T E_0}{n-1} \quad (12)$$

em que E_0 é a matriz pré-processada; e n é o número de linhas da matriz. Em seguida, calculam-se os autovalores e autovetores normalizados de C ou R (Lyra *et al.*, 2010).

$$CL = \Lambda L \quad (13)$$

e Λ é a matriz diagonal dos autovalores. Cada autovetor l é um vetor de pesos de uma componente principal. Cada autovalor λ fornece a quantidade de variância explicada pela respectiva componente, de modo que PC1 tem o maior autovalor, PC2 o segundo maior, e assim por diante, ou seja, são representados em ordem decrescente da maior participação da variância para a menor participação.

3.4 Técnicas de Agrupamento

O agrupamento é uma das técnicas mais amplamente utilizadas para a análise exploratória de dados. Todos os dias, uma grande quantidade de informações é gerada e um dos meios vitais para lidar com estes dados é classificá-los ou agrupá-los num conjunto de categorias ou *clusters* (Xu; Wunsch, 2005). Essa divisão é feita de modo que os grupos sejam diferentes uns dos outros, mas os elementos pertencentes ao mesmo grupo sejam parecidos entre si. Sendo assim, os métodos de agrupamentos podem ser utilizados com diversas finalidades e são comuns em análise de dados estatísticos nas ciências sociais, na engenharia, ciência da computação, nas ciências biológicas e nas geociências. (Izbicki; Santos, 2020).

Em geociências, a estimativa dos modelos de recursos requer a definição de domínios estacionários que estão associados às características geológicas. No entanto, apenas a partir de um conjunto de dados que reproduzam as características do fenômeno pode-se acessar essa informação, sendo necessário um criterioso processo de análise. A rigor, cada domínio estacionário apresenta sua própria função de distribuição de probabilidades e continuidade espacial, que são distintas de outros domínios, por essa

razão, a análise de agrupamento, tem sido empregada para reconhecer padrões em dados multivariados e, assim, auxiliar na definição de domínios geoestatísticos (Moreira, 2020).

Uma grande quantidade de desenvolvimentos e aplicações de técnicas de agrupamento são criadas para atender problemas específicos de uma área especializada que geralmente fazem suposições a favor da sua aplicação (Xu; Wunsch, 2005; Berkhin, 2006), contudo, as técnicas de agrupamento mais utilizadas são k-means e hierárquico.

3.4.1 Agrupamento hierárquico

Os algoritmos baseados em grafos ou hierárquicos procedem em uma sequência de rodadas. Duas abordagens podem ser derivadas do clustering hierárquico: aglomerativo e divisivo. No aglomerativo, eles começam com o clustering trivial, que tem cada ponto dos dados como um cluster de ponto de partida. Então, repetidamente, esses algoritmos mesclam os *clusters* “mais próximos” do agrupamento anterior com base em uma medida de dissimilaridade. Consequentemente, o número de *clusters* diminui a cada rodada. Se continuassem, tais algoritmos acabariam resultando no agrupamento trivial no qual todos os pontos de domínio compartilham um grande agrupamento (Shwartz; David, 2014). Na abordagem divisiva, o processo inicia-se com apenas um agrupamento contendo todos os dados e segue dividindo recursivamente, segundo alguma métrica ou medida de dissimilaridade, até que alcance algum critério de parada, frequentemente o número de *clusters* desejados (Metz; Monard, 2014).

Diferente dos algoritmos de agrupamento como K-means ou K-medoids, os métodos de agrupamento hierárquico (aglomerativos) não requerem a escolha do número de agrupamentos a serem pesquisados a priori, mas exigem que o usuário especifique uma medida de dissimilaridade entre grupos (disjuntos) de observações (Romary *et al.*, 2012). As medidas de dissimilaridade quantificam as diferenças entre os pares de pontos com base nos valores avaliados para um conjunto de variáveis, ou seja, utilizam de uma medida de distância para agrupar os indivíduos. Existem várias medidas de dissimilaridades, por exemplo: distância mínima, distância máxima, média grupal e critério de Ward's (Tan *et al.*, 2006).

Nota-se que, sem empregar uma regra de parada, o resultado de tal algoritmo pode ser descrito por um dendrograma de agrupamento tendo os conjuntos *singleton* em suas folhas e o domínio completo como sua raiz (Shwartz; David, 2014). Por exemplo, na Figura 4, se as entradas são os elementos $X = \{a, b, c, d, e\} \subset \mathbb{R}^2$, tendo como métrica a

distância euclidiana representada pelos pontos à esquerda, o dendrograma resultante é o representado à direita:

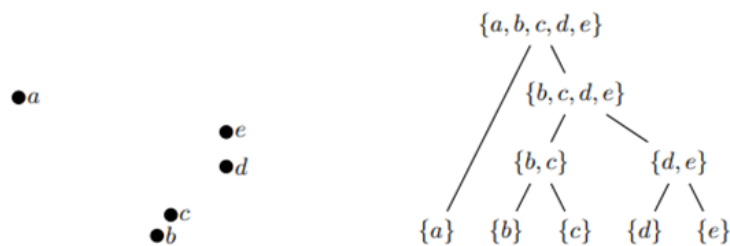


Figura 4. Dendrograma. Fonte (Shwartz; David, 2014).

3.4.2 Agrupamento por partição

O agrupamento por partição divide os dados em vários subconjuntos e, diferente do hierárquico, os objetos são divididos em grupos no mesmo nível, ou seja, sem a sobreposição de *clusters* ou não-aninhada. Trata-se de uma técnica inserida no contexto de aprendizado de máquina não supervisionada, o que significa que não precisamos de dados rotulados para os algoritmos de clustering.

Shwartz e David (2014) trazem o exemplo de um conjunto de pontos R^2 para ser agrupado (Figura 5). Suponha que temos que dividir o conjunto de pontos em dois grupos, então as alternativas são:

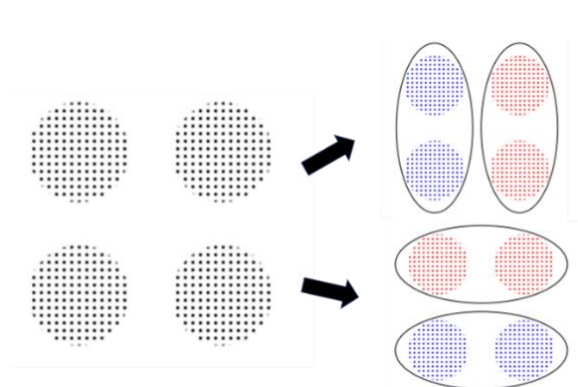


Figura 5. Exemplo de diferentes configurações de agrupamento para um mesmo conjunto de dado. Adaptado de (Shwartz; David, 2014).

Nesse exemplo simples (Figura 5), pode-se concluir que existem diferentes soluções de agrupamento concebíveis para um mesmo conjunto de dados. Como resultado, há uma grande variedade de agrupamentos e algoritmos que, em alguns dados de entrada, produzirão agrupamentos similares e para outros podem produzir agrupamentos muito diferentes. Cavalcanti Jr. (2006) cita alguns algoritmos de partição:

- Baseado em erro quadrado: K-means (MacQueen, 1967); K-medoids (Kauffman *et al.*, 1989).
- Probabilísticos: EM (maximização da expectativa, McLachlan *et al.* 1997); Modelos gaussianos mistos (descrito por VanderPlas, 2016).

O k-means é o método de agrupamento por partição baseado em protótipo mais utilizado devido à sua facilidade e simplicidade de implementação. Esse método realiza a partição dos dados em U grupos nos quais o valor de U pode ser descrito conforme equação retirada de Yasojima (2020):

Dado um conjunto de observações $\{S_1, S_2, \dots, S_N\}$, o algoritmo busca particionar as N observações em U grupos $u = \{u_1, u_2, \dots, u_n\}$ com o objetivo de minimizar a soma dos quadrados inter-*clusters* (variância). Formalmente, o objetivo é encontrar:

$$\arg_u \min \sum_{i=1}^U \sum_{p \in u_i} \|p - \mu_i\|^2 = \arg_u \min \sum_{i=1}^U |u_i| \text{Var}_{u_i} \quad (14)$$

onde μ_i é a média dos pontos em u_i .

Inicialmente o algoritmo escolhe aleatoriamente k pontos que serão os centros dos *clusters* iniciais (centroides). Em seguida, determina para cada instância do conjunto de dados o cluster ao qual pertence. Isso é feito calculando a distância entre o exemplo e o centro de cada cluster. As distâncias ou semelhanças entre pares podem ser usadas para calcular medidas de relações inter e intra-cluster. O exemplo pertence ao cluster do qual está mais próximo do centro. Um novo centroide é calculado. O centroide calculado para cada cluster torna-se o novo centro. Essa interação de cálculo da distância média dos pontos ao centroide ocorre em um *loop* até que nenhum ponto mude de centroide. Uma limitação do método é que o agrupamento converge para um mínimo local. Assim, para encontrar os melhores agrupamentos é necessário executar o algoritmo várias vezes, com diferentes centros iniciais e escolher o melhor resultado (Sanches, 2003).

Vale ressaltar que no k-means, um *cluster* é representado por seu centróide, que é uma média (geralmente média ponderada) de pontos dentro de um *cluster*. Portanto, ele funciona convenientemente apenas com atributos numéricos e pode ser afetado negativamente por um único *outlier*. Por outro lado, os centróides têm a vantagem de clareza geométrica e significado estatístico (Berkhin, 2006).

3.4.3 Algoritmo difuso

Um dos maiores desafios em agrupamento de litologias é que, geralmente, as variáveis dentro de um grupo são estimadas com uma suposição de limites rígidos, e as estruturas geológicas são, na maioria dos casos, transicionais e com mais de um evento geológico associado, levando a uma incerteza geológica. Diversas técnicas de agrupamento utilizadas para auxiliar o trabalho de classificação de domínios não consideram a imprecisão do agrupamento.

Nos últimos anos, a síntese entre algoritmos de agrupamento e a teoria dos conjuntos difusos levou ao desenvolvimento de vários algoritmos, chamados de agrupamento difuso, cujo objetivo é modelar padrões difusos (isto é, ambíguos) de forma eficiente. Zadeh (1965) introduziu a teoria de objetos matemáticos, que chamou de conjuntos nebulosos, ou seja, conjuntos em que os limites não são precisos.

A maioria das pesquisas de geologia e na área de geociências se concentram em sistemas naturais de complexidade organizada. A lógica fuzzy é uma ferramenta matemática inicialmente desenvolvida por cientistas de sistemas para lidar com problemas de complexidade organizada. A abordagem para a modelagem de sistemas complexos é baseada no senso comum, na intuição e na linguagem natural, em que formulações matemáticas precisas de componentes químicos e físicos de um sistema são substituídas por regras linguísticas baseadas na compreensão humana especializada do sistema natural (Demicco, 2004).

Embora a lógica difusa ou lógica dos conjuntos nebulosos seja uma teoria com base matemática, ela carrega uma percepção errônea, refletindo a conotação de seu rótulo, de que a lógica difusa é imprecisa e mal fundamentada (Zimmermann, 1996). Na verdade, a lógica fuzzy pode ser vista como uma tentativa de lidar precisamente com a imprecisão, assim como a teoria da probabilidade pode ser vista como uma tentativa de lidar precisamente com a incerteza (Demicco, 2004).

Os métodos de agrupamento rígidos não são capazes de quantificar incerteza nos agrupamentos. Nesses casos, o algoritmo difuso é desejável, já que atribui um grau de associação do agrupamento a todas as amostras. Esse grau de pertinência pode ser usado como uma medida de probabilidade e, portanto, pode fornecer uma maneira simples de quantificar a incerteza dos grupos (Sepulveda *et al.*, 2018). Uma ênfase em conceitos e definições básicas do sistema fuzzy é apresentada a seguir.

3.4.3.1 Conjunto fuzzy, operadores e variáveis linguísticas

Ao contrário do conceito clássico de conjunto, ou conjunto nítido, o limite de um conjunto fuzzy não é preciso. Ou seja, a mudança de não-pertencimento para pertencimento a um conjunto difuso pode ser gradual em vez de abrupta. Essa mudança gradual é expressa por uma função de pertinência, que caracteriza completa e exclusivamente um determinado conjunto fuzzy ou um conjunto nebuloso (Zimmermann, 1996). Seja A um conjunto de todos os inteiros maiores que 10.

$$\text{Escrevemos } A = \{x: x \in \mathbb{N}, x > 10\} \quad (15)$$

Seja B um conjunto de todos os inteiros muito maiores do que 10, matematicamente, essa declaração pode ser escrita como:

$$B = \{x: x \in \mathbb{N}, x \gg 10\} \quad (16)$$

A principal diferença entre esses dois conjuntos é que a relação (15) define completamente o conjunto A , enquanto a relação (16) não é suficiente para uma definição completa do conjunto B . A razão é a imprecisão do termo muito maior. É claro que 11, 12, 1178 e 2.075 são elementos do conjunto A . A maioria das pessoas concorda que 11.234 e 2.310 sem dúvida pertencem ao conjunto B , mas é duvidoso que 15 ou 50 sejam elementos de B (Kovacic; Bogdan, 2006).

O problema é como determinar o menor inteiro, que é muito maior que 10. Esse problema pode ser resolvido se usarmos uma forma alternativa de descrever um conjunto. De acordo com a teoria tradicional dos conjuntos, um conjunto pode ser definido por sua função característica. Em outras palavras, em vez de declarar individualmente cada elemento de um conjunto, definimos uma função, que pode assumir valores 1 ou 0, dependendo da pertinência total ou nenhuma associação de um elemento específico, respectivamente.

Seja S um conjunto de domínio X . Uma função característica do conjunto S atinge o valor:

$$\mu_S(X) = 1 \text{ se } X \in S, \text{ e} \quad (17)$$

$$\mu_S(X) = 0 \text{ se } X \notin S, \quad (18)$$

$$\mu: X \rightarrow \{0,1\} \quad (19)$$

O conjunto S com sua função característica é chamado de conjunto nítido. Definida como está, a função característica não pode descrever o conjunto B , ou seja, não pode lidar com a imprecisão na determinação do menor inteiro que pertenceria ao conjunto B . No entanto, ampliar a noção de função característica oferece uma maneira elegante de definir o conjunto B (teoria tradicional dos conjuntos – um conjunto pode ser definido por sua função característica) (Kovacic; Bogdan, 2006).

Em vez de determinar o menor inteiro pertencente ao conjunto B , pode-se dizer que todos os inteiros maiores que 10 pertencem ao conjunto B , mas com um grau de pertinência diferente. A função característica, obtendo valores parciais ou graduados do intervalo $[0, 1]$, agora se torna uma função de pertinência.

A função de pertinência converte o grau de nebulosidade no intervalo normalizado $[0, 1]$, em que os valores de limite 0 e 1 se assemelham aos graus de pertinência dos membros do conjunto nítido. As funções dos membros podem assumir diferentes formas. No entanto, a triangular, a trapezoidal e a gaussiana, mostradas na Figura 6, são as mais usadas (Kovacic; Bogdan, 2006).

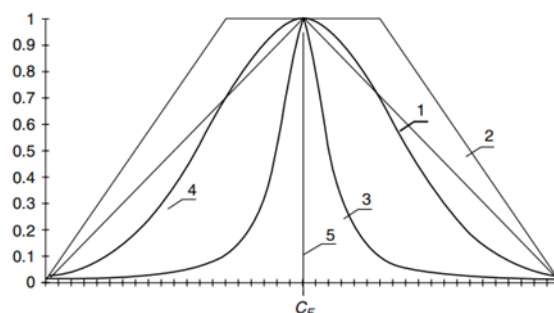


Figura 6. Formas típicas de funções de pertinência: 1- triangular; 2- trapezoidal; 3- gaussiana; 4- em forma de sino; 5-singleton. Fonte: Kovacic; Bogdan, 2006.

Na teoria dos conjuntos fuzzy, a gama de valores quantitativos possíveis considerados para os membros do conjunto fuzzy é chamada de universo de discurso (Zadeh, 1978). O universo do discurso pode ser contínuo ou discreto. A medida de imprecisão de cada elemento do conjunto fuzzy é determinada usando uma função de pertinência espalhada por uma parte ou por todo o universo do discurso.

Na lógica fuzzy, os conjuntos nebulosos são na verdade os valores das palavras, e a ideia central é realizar operações com as palavras. Para tanto, é preciso considerar uma variável linguística (ou fuzzy) como uma entidade utilizada para representar de modo impreciso – e, portanto, linguístico – um conceito ou uma variável de um dado problema.

Uma variável linguística, diferentemente de uma variável numérica, admite apenas valores definidos na linguagem fuzzy que está utilizando-se dela (Marro *et al.*, 2000). Portanto, a base de regras fuzzy IF - THEN contém as variáveis linguísticas que geralmente são os nomes de entradas e saídas, ao invés de seus valores concretos (números).

3.4.3.2 Regras e implicação fuzzy

A maneira mais comum de serem armazenadas informações em uma base de conhecimento fuzzy é a representação através de regras fuzzy. Uma base de regra fuzzy normalmente é formada de duas partes principais:

IF <antecedente> *THEN* <consequente>

O antecedente é composto por um conjunto de condições que, quando satisfeitas (mesmo parcialmente), determinam o processamento do consequente da regra por um mecanismo de inferência fuzzy. O processo acima descrito denomina-se disparo de uma regra. Por sua vez, o consequente é composto de um conjunto de ações ou diagnósticos que são gerados com o disparo da regra. Os consequentes das regras disparadas são processados em conjunto para gerar uma resposta determinística para cada variável de saída do sistema (Kovacic; Bogdan, 2006).

Portanto, a semântica é que vai definir, para o mecanismo de inferência (implicação), como são processados os antecedentes, quais são os indicadores de disparo das regras e quais os operadores utilizados sobre os conjuntos fuzzy existentes para executar o processamento de conhecimento, ou seja, o modelo de processamento definido para o sistema de conhecimento depende basicamente da forma de armazenamento das informações escolhida (Almeida; Evsukoff, 2003). Os conjuntos fuzzy obedecem a certas propriedades e podem ser operados de diversas maneiras. As operações entre conjuntos são extremamente importantes para os sistemas que se utilizam dessa lógica, pois os cálculos proposicionais, por exemplo, são baseados nessas operações (Marro *et al.*, 2021).

Existem muitas sugestões para determinar a função de pertinência de um conjunto fuzzy, que é o resultado da união, da interseção e do complemento de outros conjuntos fuzzy (Zimmermann, 1996). Em geral, os operadores em conjuntos difusos usam normas triangulares, uma classe de funções binárias, que podem ser divididas em normas T (operadores AND) e normas S (operadores OR). As T-norms executam uma operação de interseção em conjuntos fuzzy e têm uma importância particular no controle da lógica fuzzy. T-norma é geralmente denotada como T (a, b). S-normas representam uma

operação de união denotada como $S(a, b)$. Quase todas as normas T usadas em aplicações de controle fuzzy podem ser derivadas de quatro normas T básicas listadas abaixo:

$$T(\mu_B, \mu_C) = \min(\mu_B, \mu_C) \quad (20)$$

$$T(\mu_B, \mu_C) = \mu_B \cdot \mu_C \quad (21)$$

$$T(\mu_B, \mu_C) = \max(0, \mu_B + \mu_C - 1) \quad (22)$$

$$T(\mu_B, \mu_C) = \begin{cases} \mu_B, & \text{if } \mu_C = 1 \\ \mu_C, & \text{if } \mu_B = 1 \\ 0, & \text{if } \mu_B, \mu_C > 1 \end{cases} \quad (23)$$

Das diferentes maneiras de definir uma implicação difusa, duas delas são preferidas em aplicações de controle: implicação de produto (também chamada de Larsen, equação 21) e a implicação mínimos ou Mamdani (equação 20). Estas são retratadas com mais detalhes no exemplo de Kovacic e Bogdan (2006):

$$\mu_{FR^i} = \mu_{R_{pq}} \cdot \mu_{P_m} \quad (24)$$

$$\mu_{FR^i} = \min(\mu_{R_{pq}}, \mu_{P_m}) \quad (25)$$

O índice FR^i nas Equações (24, 25) pode ser substituído pelo índice $R_{pq} \rightarrow P_m$ designando de forma mais expressiva uma regra fuzzy selecionada, FR^i : **SE** R_{pq} **ENTÃO** P_m . Dados os conjuntos difusos discretos T_1 e F_1 com funções de pertinência triangular determinadas como:

$$T_1(x) = \{(0,25; 2); (0,5; 3); (0,75; 4); (1; 5); (0,75; 6); (0,5; 7); (0,25; 8)\} \text{ e} \\ F_1(y) = \{(0,33; 10); (0,67; 11); (1; 12); (0,67; 13); (0,33; 14)\} \quad (26)$$

Considere uma regra difusa na forma FR^1 : **SE** R_{11} **ENTÃO** P_1 , e a relação R_{11} ter uma forma bidimensional, R_{11} : x é T_1 **E** y é F_1 . O operador mínimo é usado como uma norma T .

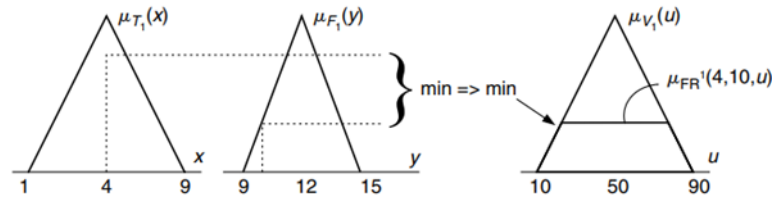


Figura 7. Implicação de Mamdani aplicada à regra fuzzy FR1.

Então, para os valores $x = 4$ e $y = 10$, a relação R_{11} tem o seguinte grau de filiação:

$$\mu_{R_{11}}(4, 10) = \min [\mu_{T_1}(4), \mu_{F_1}(10)] = \min [0,75; 0,33] = \mathbf{0,33} \quad (27)$$

O procedimento que atribui grau de pertinência $\mu_{T_1}(x)$ ao valor numérico 4 é chamado de fuzzificação. Os graus de pertinência $\mu_{T_1}(4) = 0,75$ e $\mu_{F_1}(10) = 0,33$ representam equivalentes difusos dos números 4 e 10, respectivamente.

Agora, considere a proposição P_1 na regra FR^1 : ter uma forma, $P_1: u \in V_1$, em que $u \in U$ é uma variável linguística e V_1 é um valor linguístico associado a um conjunto fuzzy,

$$V_1(x) = \{(\mu_{V_1}(u), u) : u \in U\} \quad (28)$$

$$V_1(x) = (0,25; 20); (0,5; 30); (0,75; 40); (1; 50); (0,75; 60); (0,5; 70); (0,25; 80)\} \quad (29)$$

Então, a interpretação da regra fuzzy FR^1 é dada pela função de pertinência μ_{FR^1} obtida após a aplicação de uma implicação de produto:

$$\mu_{FR^1}(4, 10, u) = \mu_{R_{11}}(4, 10) \cdot \mu_{V_1}(u) = \mathbf{0,33} \cdot \mu_{V_1}(u) \quad (30)$$

A implicação do produto difuso pode ser apresentada graficamente, como mostrado na Figura 8.

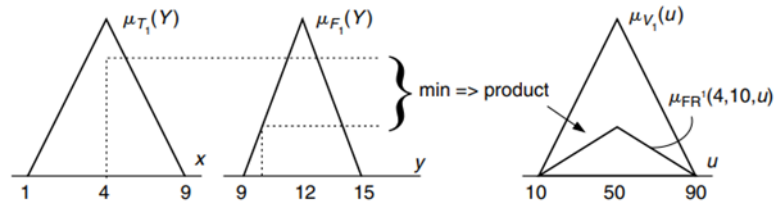


Figura 8. Implicação produto aplicada à regra fuzzy FR1

A diferença entre os resultados de duas implicações nebulosas é óbvia. Com a implicação do produto, a função de pertinência $\mu_{FR^1}(\cdot)$ é formada pela escala $\mu_{V_1}(u)$ e retém uma forma triangular após a implicação, enquanto a implicação min “corta” a função de pertinência original $\mu_{V_1}(u)$, que resulta em sua forma trapezoidal. A diferença entre os resultados das implicações sugere que o tipo de implicação utilizada no projeto do controlador fuzzy terá influência na estrutura do algoritmo de controle fuzzy.

Uma implicação fuzzy produz um conjunto fuzzy de saída resultante para cada regra fuzzy ativada, mas não define como esse conjunto fuzzy realmente contribui para o valor nítido de saída de um controlador fuzzy. Ou seja, o valor crisp de qualquer variável de entrada geralmente pertence a mais de um conjunto fuzzy de entrada que, por sua vez, ativa mais de uma regra fuzzy e, portanto, mais de um conjunto fuzzy de saída contribui para a saída.

A composição de todas as regras fuzzy ativadas gera um processo chamado de agregação, que conclui a inferência baseada em regras individuais com um conjunto fuzzy de saída, podendo ser usado posteriormente para o cálculo do valor de saída *crisp*. Uma apresentação gráfica desse procedimento é mostrada na Figura 9. Considerando os conjuntos fuzzy (x, y) e os respectivos graus de pertinência u , produzidos pela aplicação de duas regras fuzzy FR^1 e FR^2 , a agregação desses conjuntos resulta no conjunto de saída fuzzy da Figura 9.

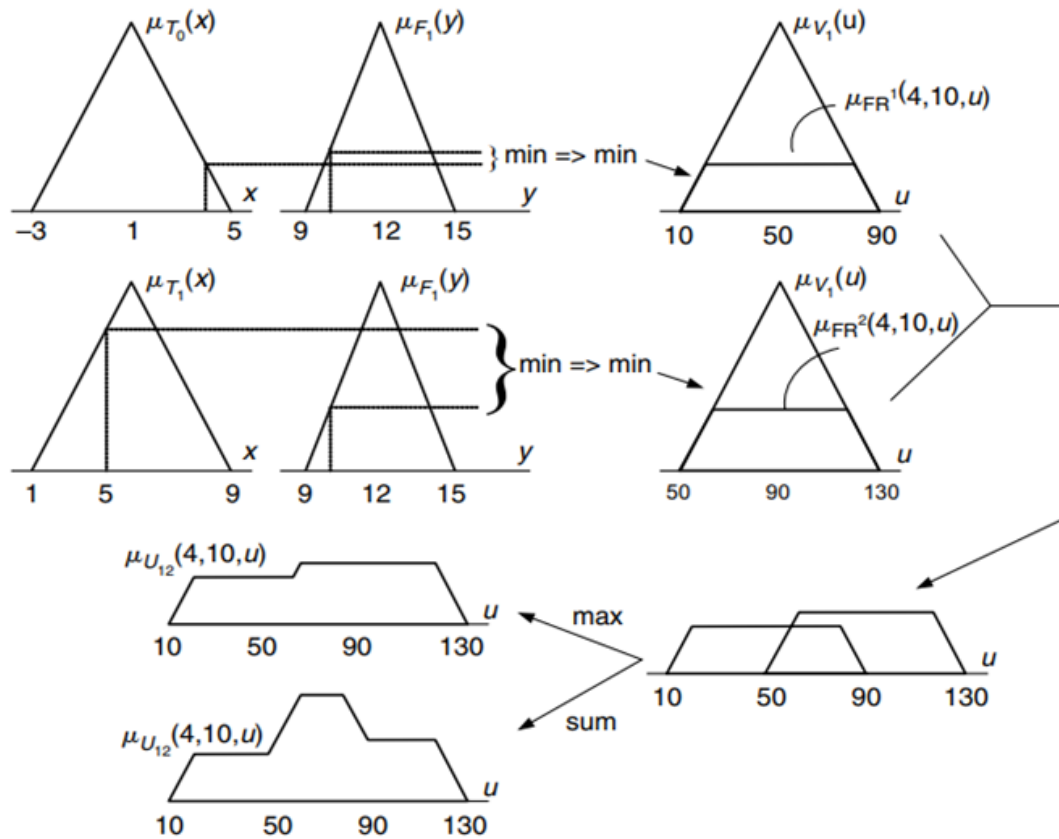


Figura 9. Procedimento de inferência individual baseado em regras.

No entanto, a saída do controlador fuzzy por si só não é suficiente e implicará um valor crisp. O procedimento que extrai o valor de saída nítido de um conjunto de saída fuzzy é chamado de defuzzificação (Kovacic; Bogdan, 2006).

3.4.3.3 Defuzzificação

Para se obter uma saída numérica, é necessário defuzzyficar a saída obtida na etapa anterior. O método de defuzzificação mais comum é a técnica do centroide, que obtém o ponto onde uma linha vertical divide ao meio um conjunto agregado. A fórmula matemática que obtém esse ponto é expressa da seguinte forma (equação 31):

$$u_{FC}(x_k, y_k) = \frac{\sum_i u_i \sum_{j=1}^r \mu_{FR^j}(x_k, y_k, u_i)}{\sum_i \sum_{j=1}^r \mu_{FR^j}(x_k, y_k, u_i)} \quad (31)$$

r é o número de regras nebulosas ativadas pelas entradas nítidas x_k e y_k . Agora, suponha que duas regras fuzzy fossem aplicadas a um conjunto fuzzy, cujo resultado da agregação fosse usado, e o conjunto de saída fuzzy atinge este formato:

$$U_{12}(u) = \left\{ \begin{array}{l} (0.25,20);(0.25, 30);(0.25,40);(0.25,50), (0.5,60);(0.58,70); \\ (0.58,80);(0.33,90);(0.33, 100);(0.33, 110);(0.25,120) \end{array} \right\} \quad (32)$$

Considerando o conjunto fuzzy obtido (4, 10, u), o resultado numérico (valor crisp) obtido com a aplicação da técnica do centroide (COG) é dado por:

$$COG(4,10) = \frac{[0.25 \cdot 20 + 0.25 \cdot 30 + 0.25 \cdot 40 + 0.25 \cdot 50 + (0.5) \cdot 60 + (0.58) \cdot 70 + \dots + 0.25 \cdot 120]}{[0.25 + 0.25 + 0.25 + 0.25 + (0.5) + (0.58) + \dots + 0.25]} = \frac{281}{3.9} = 72.0513 \quad (33)$$

3.4.4 Métricas de verificação dos agrupamentos

Diferentes algoritmos de agrupamento tendem a gerar diferentes respostas na base de dados. Avaliar a qualidade dos resultados é uma das considerações mais importantes a se fazer quando se deseja implementar algoritmos de aprendizado de máquina (Moreira, 2020).

Existem algumas métricas na literatura que possibilitam a avaliação do método escolhido, permitindo verificar se os resultados obtidos possuem algum significado prático ou que justifique a escolha de um determinado algoritmo em detrimento de outros.

Segundo Pacífico (2016), existem três tipos principais de critérios de validação:

- I- Os critérios externos que comparam duas partições formadas de uma mesma base de dados; geralmente, a partição gerada pelo algoritmo de agrupamento é comparada a uma partição fornecida *a priori*, por exemplo: alguns grupos pré-rotulados por um especialista ou bases geradas sinteticamente com distribuições conhecidas. Alguns dos critérios externos mais conhecidos na literatura são o Rand Index, Jaccard e o Rand Index Ajustado (Vendramin *et al.*, 2010);
- II- Os critérios internos que examinam a validade dos grupos formados pela verificação da adequabilidade intrínseca dos resultados obtidos, no que diz respeito ao conjunto de dados em análise, sendo desnecessário qualquer conhecimento prévio sobre as relações desses dados. São exemplos a Soma dos Erros Quadráticos dos Agrupamentos (*Within-Cluster Sum of Squares*, ou WCSS), o Erro de Quantização (*Quantization Error*), a Distância Intra-Cluster (*Intra-Cluster Distance*, ou D_{max}), a Distância Inter-Cluster (*Inter-Cluster Distance*, ou D_{min});

- III- Os critérios relativos que fazem uma análise com ênfase na comparação entre diferentes estruturas de agrupamentos obtidas para os dados, avaliando o mérito relativo das estruturas obtidas em relação à norma comparativa adotada. As estruturas consideradas mais representativas das características internas dos dados, de acordo com o critério adotado, são selecionadas. Critérios relativos encontrados na literatura são Largura de Silhueta, Coeficiente de partição de Dunn, Critério Davies-Bouldin etc.

O índice de Rand corrigido (CR) é interessante pois ele não é sensível à distribuição dos *clusters* e ao número de classes (Cavalcanti, 2006). O índice de Rand foi utilizado nessa tese para avaliação do agrupamento, uma vez que os padrões utilizados no sistema de inferência fuzzy foram baseados nas estatísticas de um subconjunto do banco de dados (47% das amostras com litologias precisas). O CR mede a qualidade da partição gerada pelo algoritmo através do grau concordância entre duas partições (U e V), em que a partição U é o resultado do agrupamento, e a partição V é a informação prévia ou a classe rotulada (Macário; Carvalho, 2012). Dadas as observações, o CR é então definido como:

$$CR = \frac{a - \frac{(a+c)(a+b)}{M}}{\frac{(a+c)+(a+b)}{2} - \frac{(a+c)(a+b)}{M}} \quad (34)$$

onde $M = a + b + c + d$.

As relações observadas pelo CR em sua avaliação, para cada par de padrões, são as seguintes:

- a: Número de pares de objetos de dados pertencentes à mesma classe V e ao mesmo *cluster* U;
- b: Número de pares de objetos de dados pertencentes à mesma classe em V e a diferentes *clusters* em U.
- c: Número de pares de objetos de dados pertencentes a diferentes classes em V e ao mesmo *cluster* em U.
- d: Número de pares de objetos de dados pertencentes a diferentes classes em V e a diferentes *clusters* em U.

Os termos a e d são medidas de classificações consistentes (acordos), enquanto os termos b e c são medidas de classificações inconsistentes (desacordos). Esse índice é corrigido para o acaso, sob a suposição de que o número de grupos (classes/*clusters*,

Figura 10) em ambas as partições U e V a serem comparadas é o mesmo (Vendramin *et al.*, 2010).

O CR tem seus valores limitados ao intervalo $[-1, 1]$, sendo o valor 1 indicativo de uma perfeita concordância entre as duas partições em comparação, enquanto o valor 0 (ou negativos) é um indicativo de partições formadas ao acaso (Arabie *et al.*, 1996).

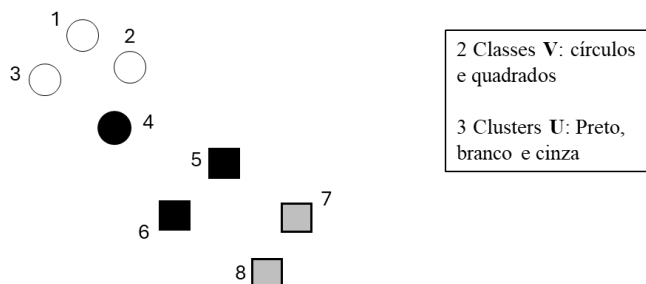


Figura 10. Adaptado de Vendramin *et al.* (2010).

Outra maneira de validar o agrupamento é comparar diferentes técnicas e algoritmos. Utilizaremos os algoritmos k-means e hierárquico, que são técnicas consolidadas para verificar a qualidade do agrupamento, como também a inspeção visual dos grupos.

Inspirado nos trabalhos de Moreira (2020), outra medida de verificação da qualidade dos agrupamentos pode ser obtida por meio da medida de continuidade espacial dos indicadores dos domínios gerados, comparando-se aos variogramas das amostras. Segundo esse autor, o padrão de distribuição das amostras de uma dada categoria pode ser caracterizado pela análise da continuidade espacial dos indicadores definidos por aquela categoria.

3.5 Análise de regressão

A regressão é uma ferramenta estatística comum para modelar a relação entre algumas variáveis “explicativas” e algum resultado real avaliado (Shwartz; David, 2014). É a técnica de dependência mais amplamente utilizada em tomada de decisões de 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. Isso explica a grande quantidade de trabalhos utilizando modelos de regressão (Hair *et al.*, 2009).

Segundo Deutsch (2015), muitas variáveis metalúrgicas apresentam comportamento não linear, e pesquisas substanciais na área de compreensão do

comportamento das variáveis de processamento mineral foram publicadas recentes. Esse autor ainda afirma que nem todas as variáveis metalúrgicas são não lineares e, mais importante, uma relação não linear entre variáveis não implica que as próprias variáveis tenham uma média não linear (por exemplo, a relação entre a área e o raio de um círculo é quadrática, embora tanto a área como o raio sejam lineares).

3.5.1 Regressão simples

Se uma relação linear é válida para sumarizar a dependência observada entre duas variáveis quantitativas, então a equação que descreve tal relação é dada por (Montgomery *et al.*, 2012):

$$Y = \beta_0 + \beta X \quad (35)$$

A relação entre X e Y é determinística, ou seja, espera-se que todos os pontos caiam exatamente em cima da reta de regressão. No entanto, existe uma diferença denominada erro (ε), que quantifica a falha do modelo em ajustar-se aos dados exatamente. Adicionando o erro à equação acima, temos:

$$Y = \beta_0 + \beta X + \varepsilon \quad (36)$$

β e β_0 são os parâmetros do modelo (coeficientes obtidos pelos mínimos quadrados); X é a variável regressora (independente); e Y é denominada de variável resposta ou dependente obtida através de X .

O melhor ajuste do modelo é encontrado para o conjunto de dados, de maneira a minimizar a diferença entre a soma dos quadrados do valor estimado e dos dados observados (Montgomery *et al.*, 2012).

$$\beta_0 = Y - \beta \bar{X} \quad (37)$$

$$\beta = \frac{\sum(x_i - \bar{X})(y_i - \bar{Y})}{\sum(x_i - \bar{x})^2} \quad (38)$$

3.5.2 Regressão múltipla

Segundo Faria (2011) uma regressão é dita múltipla quando se tem uma única variável dependente, porém duas ou mais variáveis independentes. A Equação 39 ilustra um modelo multivariado apresentando k variáveis regressoras:

$$y = \beta_0 + \beta_1 x_{1(i)} + \beta_2 x_{2(i)} + \dots + \beta_k x_{k(i)} + \varepsilon_i \quad \forall i = 1, \dots, n \quad (39)$$

- y é a variável dependente;
- β_0 e β_j , $j = 0, 1, 2, \dots, k$ são os coeficientes de regressão;
- x_j , $j = 1, 2, 3, \dots, k$ são as variáveis independentes;
- ε é um componente do erro aleatório.

Os parâmetros β_j representam a variação esperada em y por unidade de variação em x_j , quando todas as restantes variáveis x_i , com $i \neq j$ permanecem constantes. A equação apresentada em (39) refere-se a um elemento genérico i (Faria, 2011). A relação entre as n observações é dada pelo seguinte conjunto de equações:

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{1(1)} + \beta_2 x_{2(1)} + \dots + \beta_k x_{k(1)} + \varepsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_{1(2)} + \beta_2 x_{2(2)} + \dots + \beta_k x_{k(2)} + \varepsilon_2 \\ &\dots \\ y_n &= \beta_0 + \beta_1 x_{1(n)} + \beta_2 x_{2(n)} + \dots + \beta_k x_{k(n)} + \varepsilon_n \end{aligned} \quad (40)$$

De forma a simplificar a escrita, pode-se recorrer à notação matricial para representar as equações anteriores. Consideram-se as seguintes matrizes representativas de cada uma das variáveis em questão:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n \times 1}, X = \begin{bmatrix} 1 & x_{1(1)} & x_{2(1)} & \dots & x_{k(1)} \\ 1 & x_{1(2)} & x_{2(2)} & \dots & x_{k(2)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{1(n)} & x_{2(n)} & \dots & x_{k(n)} \end{bmatrix}_{n \times (k+1)}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}_{(k+1) \times 1}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n \times 1}$$

(41)

Assim sendo, tem-se a seguinte equação matricial:

$$Y = X\beta + \varepsilon \quad (42)$$

Também na regressão múltipla as estimativas dos mínimos quadrados são obtidas pela escolha dos estimadores que minimizam a soma dos quadrados dos desvios entre os valores observados e os valores ajustados. As demonstrações das operações algébricas podem ser vistas em Faria (2011), sendo a seguir apresentado o melhor estimador para os coeficientes da regressão com duas ou mais variáveis independentes.

$$b = (X^T X)^{-1} X^T Y \quad (43)$$

3.5.3 Suposições dos modelos e verificação da regressão

Os modelos de regressão possuem suposições e premissas que devem ser satisfeitas, sob o risco do modelo falhar em fornecer bons resultados. São algumas dessas premissas:

- 1- Independência entre os dados e linearidade das variáveis;
- 2- Normalidade dos resíduos;
- 3- Homocedasticidade dos resíduos.

Se as variáveis independentes são correlacionadas entre si, um problema chamado multicolinearidade levará à mudança drástica do modelo se uma variável independente for removida; portanto, um modelo como esse é impreciso. A matriz de correlação permite avaliar a existência de dependência linear entre par de variáveis. Os elementos da diagonal dessa matriz são chamados de fatores de inflação de variância (VIF) e representam o incremento da variância devido à presença de multicolinearidade, devendo seu valor ser menor do que dez. Na hora de selecionar as variáveis independentes, recomenda-se escolher as que têm uma correlação alta com a variável-alvo e que não tenham correlação com outras variáveis independentes, ou que tenham uma correlação muito fraca (Montgomery *et al.*, 2012).

Para avaliar a normalidade dos resíduos, é comum a utilização do gráfico de probabilidade normal associado a um teste de normalidade. Para verificar se os resíduos apresentam uma distribuição normal, podem-se usar gráficos (histograma, *Q-Qplot*) ou testes estatísticos (Shapiro-Wilk e Kolmogorov-Smirnov) para os resíduos. A violação da suposição de normalidade não invalida o modelo para bancos de dados com muitas amostras (Hair *et al.*, 2009).

Por fim, a homocedasticidade é o termo para designar variância constante dos erros/resíduos para observações distintas. Quando não há variação constante, temos o

problema de heterocedasticidade, ou seja, a variância dos erros é diferente para cada valor condicional de X. Esse problema é observado de 2 formas: análise gráfica ou testes estatísticos (Hair *et al.*, 2009).

Para verificar se a regressão é significativa, deve-se inicialmente observar o resultado da ANOVA. A Análise de Variância (ANOVA) se baseia no pressuposto de normalidade e de homocedasticidade (homogeneidade da variância) entre os grupos (Fan; Hancock, 2012). Assumir a normalidade é dizer que dois ou mais grupos têm médias iguais, ou seja, seguem distribuição gaussiana. Embora o objetivo seja testar a hipótese de médias iguais, o processo se dá pela decomposição da variância total existente entre as observações. O cálculo da ANOVA está ilustrado pela Equação 44 (Triola, 2008):

$$\sum_{i=1}^n (y_i - \bar{Y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{Y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (44)$$

A Equação 44 pode ser expressa da seguinte forma:

$$SQT = SQG + SQR \quad (45)$$

Em que SQT representa a soma dos quadrados das variações das populações; SQG simboliza a soma dos quadrados entre os grupos; e SQR é a soma dos quadrados das amostras dentro do grupo. No modelo de regressão, o parâmetro SQR pode ser entendido como o resíduo, e o SQG reporta o quanto da variável dependente pode ser explicada pelo modelo de regressão.

Se o p-valor > 0,050, não temos evidência para dizer que o modelo de regressão linear é importante para explicar a variável desfecho, ou seja, as variáveis independentes não exercem influência na variável dependente. Por outro lado, se o p-valor ≤ 0,050, pode-se dizer que ao menos uma das variáveis do modelo é importante para explicar a variável desfecho, ou seja, pelo menos uma variável independente exerce influência na variável dependente (Capp *et al.*, 2020).

3.6 Floresta aleatória

Floresta aleatória ou *Random Forest* é um método do aprendizado de máquina amplamente utilizado para tarefas de classificação e regressão. Foi proposto por Breiman (2001) a partir da observação de outras pesquisas, como: *bagging* de Breiman (1996), em que, para desenvolver cada árvore, uma seleção aleatória (sem substituição) era feita a

partir dos exemplos no conjunto de treinamento; seleção de divisão aleatória (Dietterich, 1998), onde, em cada nó, a divisão é selecionada aleatoriamente entre as K melhores divisões; subespaço aleatório (Ho, 1998) e o trabalho de reconhecimento de caracteres escritos por Amit e Geman (1997) (Breiman, 2001).

O *bagging* é o primeiro passo para elaboração de uma *random forest*. A partir de um banco de dados \mathcal{L} com n observações de variáveis independentes (x) e uma variável dependente (y), o método *bagging* inicia retirando uma amostra de tamanho n com reposição do conjunto de dados \mathcal{L} . Esse passo é repetido B vezes, com objetivo de criar B conjuntos de dados retirados de uma mesma distribuição, mas com elementos individuais distintos entre uma amostra e outra (Izenman, 2008).

A técnica de *bootstrap* é aplicada no banco de dados, e o *bagging* gera B árvores de decisão, distintas entre si com diferentes observações e conjuntos únicos de variáveis utilizadas:

$$\mathcal{L}^{*b} = (x_i^{*b}, y_i^{*b}), i=1, 2, \dots, n \text{ e } b=1, 2, \dots, B. \quad (46)$$

Árvores de decisão distintas entre si são então formadas em relação aos cortes que serão aplicados em cada nó pai e ao valor predito da variável y . A previsão será dada pela média agregada de todas as árvores:

$$\hat{\mu}_{bag}(x) = \frac{1}{B} \sum_{b=1}^B \hat{\mu}^{*b}(x) \quad (47)$$

É possível avaliar o erro de predição do modelo *bagging*, aplicando as árvores geradas em um banco distinto. O algoritmo busca as n_i árvores que não utilizaram a amostra x_i em sua formação (sendo n_i um valor menor ou igual a B). Posteriormente, y_i é calculado utilizando cada uma das árvores. A média das estimativas de todas as n_i árvores utilizadas para a predição de x_i é então utilizada no valor predito final:

$$\hat{\mu}_{bag}(x_i) = \frac{1}{n_i} \sum_{b \in N_i} \hat{\mu}^{*b}(x_i), \quad (48)$$

onde N_i é o conjunto de árvores que não utilizou a amostra x_i na etapa de treinamento. O erro é então calculado por meio do erro quadrático médio entre os valores reais e os valores estimados:

$$Erro_{bag} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\mu}_{bag}(x_i))^2 \quad (49)$$

A técnica de *random forest* cria uma combinação (*ensemble*) de árvores de decisão, na maioria dos casos treinados com o método de *bagging*, porém o RF utiliza de mais uma componente aleatória: em cada nó da árvore, são escolhidas aleatoriamente 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). Ou seja, o algoritmo de floresta aleatória adiciona aleatoriedade extra ao modelo quando está criando as árvores. Em vez de buscar a melhor característica para a partição dos nós, ele explora a melhor característica em um subconjunto aleatório das variáveis. Essa abordagem permite uma maior diversidade nas decisões, contribuindo para a robustez do modelo. Como o *random forest* considera, em cada nó da árvore m , variáveis ao acaso, é importante avaliar quais são as variáveis que fornecem os melhores resultados quando utilizadas para elaboração da árvore.

Em relação ao tamanho da árvore, Hastdie *et al.* (2009) citam que, ao introduzir mais árvores ao modelo, a variância da floresta reduz e normalmente ela se estabiliza com 200 a 500 árvores.

3.7 Máquinas de vetores de suporte (SVM)

As máquinas de vetores suporte (*Support Vector Machine*, SVM) são técnicas de aprendizado de máquina supervisionado e foram desenvolvidas por Vapnik em 1999, fazendo parte do grupo de técnicas que buscam maximizar a capacidade de generalização. Foram inicialmente construídas para a solução de problemas de classificação de padrões. Posteriormente, foram estendidas para a modelagem de dados de processos por regressão (Schölkopf; Smola, 2004). A formulação teórica é apresentada a seguir e adotou a mesma notação apresentada em Shwartz e David (2014).

3.7.1 SVM de margem rígida (hard SVM)

Seja $S = (x_1, y_1), \dots, (x_m, y_m)$ um conjunto de treinamento de exemplos, onde cada $x_i \in \mathbb{R}^d$ e $y_i \in \{\pm 1\}$. Diz-se que esse conjunto de treinamento é linearmente separável, se existir um meio-espaço, (w, b) , tal que $y_i = \text{sign}(\langle w, x_i \rangle + b)$ para todo i . Alternativamente, essa condição pode ser reescrita como:

$$\forall i \in [m], \quad y_i (\langle w, x_i \rangle + b) > 0 \quad (50)$$

Todos os meios-espacos (w, b) que satisfazem essa condiçao sao hipoteses para minimizaçao do risco empirico ERM (o erro entre 0-1 deve ser zero, que e o erro mınimo possıvel). Para qualquer amostra de treinamento separavel, ha muitos meio-espacos ERM, devendo o aprendizado escolher o melhor. Entao, consideram-se duas classes representadas pelos pontos azuis e vermelhos, como um conjunto de treinamento descrito na figura a seguir (Figura 11):

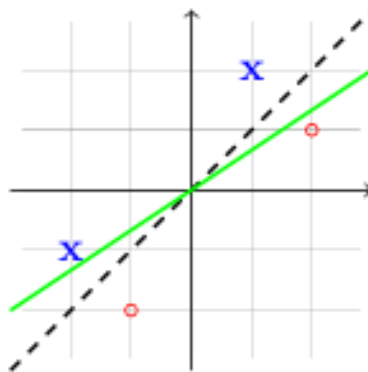


Figura 11. Conjunto de treinamento representado pelas classes vermelhas e azuis. Fonte: Shwartz e David (2014).

As duas classes sao separadas por hiperplanos (pretos tracejados e verdes solidos); intuitivamente, o hiperplano preto e preferıvel sobre o verde para separaçao dos conjuntos, e uma maneira de formalizar essa intuiçao e usar o conceito de margem.

A margem de um hiperplano em relaçao a um conjunto de treinamento e definida como a distancia mınima entre um ponto no conjunto de treinamento e o hiperplano. Se um hiperplano tiver uma margem grande, ele ainda separa o conjunto de treinamento, mesmo se for perturbado levemente cada instancia. Note a seguir que o erro verdadeiro de um meio-espaco pode ser limitado em termos da margem que ele tem sobre a amostra de treinamento (quanto maior a margem, menor o erro), independentemente da dimensao euclidiana na qual esse meio-espaco reside. A margem rıgida e a regra de aprendizado na qual se retorna um hiperplano ERM que separa o conjunto de treinamento com a maior margem possıvel (Shwartz; David, 2014). Para definir formalmente o SVM rıgido, primeiro se expressa a distancia entre um ponto x e um hiperplano usando os parametros que definem o meio-espaco (equaçoes 51-56). Seja a distancia entre um ponto x e o hiperplano definido por (w, b) , onde:

$$\|w\| = 1 \text{ e } |\langle w, x \rangle + b| \quad (51)$$

A distancia entre um ponto x e o hiperplano e definida como:

$$\min\{\|x - v\|: \langle w, v \rangle + b = 0\}, \quad (52)$$

tomando $v = x - (\langle w, x \rangle + b)w$, temos que:

$$\begin{aligned} \langle w, v \rangle + b &= \langle w, x \rangle - (\langle w, x \rangle + b)\|w\|^2 + b = 0 \quad \text{e,} \\ \|x - v\| &= |\langle w, x \rangle + b|\|w\| = |\langle w, x \rangle + b| \end{aligned} \quad (53)$$

Portanto, a distância é no máximo $|\langle w, x \rangle + b|$. Se qualquer outro ponto u no hiperplano, de modo que $\langle w, u \rangle + b = 0$. Tem-se:

$$\begin{aligned} \|x - u\|^2 &= \|x - v + v - u\|^2 = \|x - v\|^2 + \|v - u\|^2 + 2\langle x - v, v - u \rangle \\ &\geq \|x - v\|^2 + 2\langle x - v, v - u \rangle = \|x - v\|^2 + 2(\langle w, x \rangle + b)\langle w, v - u \rangle \\ &= \|x - v\|^2 \end{aligned} \quad (54)$$

Em que a última igualdade é porque $\langle w, v \rangle = \langle w, u \rangle = -b$. Então, a distância entre x e u é pelo menos a distância entre x e v . Com base nessa afirmação, o ponto mais próximo no conjunto de treinamento do hiperplano de separação é $\min_{i \in [m]} |\langle w, x_i \rangle + b|$. A regra do SVM rígido é dada pela equação:

$$\frac{\operatorname{argmax}}{(w,b):\|w\|=1} \min_{i \in [m]} |\langle w, x_i \rangle + b| \quad \text{s. t.} \quad \forall_i, y_i (\langle w, x_i \rangle + b) > 0 \quad (55)$$

Sempre que houver uma solução para o problema anterior (no caso separável), escreve-se um problema equivalente da seguinte forma:

$$\frac{\operatorname{argmax}}{(w,b):\|w\|=1} \min_{i \in [m]} y_i (\langle w, x_i \rangle + b) \quad (56)$$

Observe abaixo outra formulação equivalente da regra do SVM rígido como um problema de otimização quadrática. Exemplo:

Entrada: $(x_i, y_i), \dots, (x_m, y_m)$

Resolvendo:

$$\begin{aligned} (w_0, b_0) &= \frac{\operatorname{argmax}}{(w,b)} \|w\|^2 \quad \text{s. t.} \quad \forall_i, y_i (\langle w, x_i \rangle + b) \geq 1 \\ \text{Saída: } \hat{w} &= \frac{w_0}{\|w_0\|}, \quad \hat{b} = \frac{b_0}{\|w_0\|} \end{aligned} \quad (57)$$

Nas etapas seguintes, veremos que a saída do SVM rígido é de fato o hiperplano de separação com a maior margem. Intuitivamente, o SVM rígido busca por w de norma mínima entre todos os vetores que separam os dados e para os quais $|\langle w, x_i \rangle + b| \geq 1$ para todo i . Em outras palavras, considere que a margem seja 1, mas agora as unidades

nas quais medimos a margem escalam com a norma de w . Portanto, encontrar o maior meio-espço de margem se resume a encontrar w cuja norma é mínima. Formalmente, a saída do SVM rígido é uma solução dada na Equação (56).

Agora considere (w^*, b^*) uma solução da Equação (56) e defina a margem obtida por (w^*, b^*) como sendo $\gamma^* = \min_{i \in [m]} y_i (\langle w^*, x_i \rangle + b^*)$. Para todo i , tem-se que:

$$y_i (\langle w^*, x_i \rangle + b^*) \geq \gamma^* \quad (58)$$

ou equivalente:

$$y_i \left(\left\langle \frac{w^*}{\gamma^*}, x_i \right\rangle + \frac{b^*}{\gamma^*} \right) \geq 1 \quad (59)$$

Assim, o par $\left(\frac{w^*}{\gamma^*}, \frac{b^*}{\gamma^*} \right)$ satisfaz as condições da otimização quadrática do problema dado na Equação (57). Portanto, $\|w_0\| \leq \left\| \frac{w^*}{\gamma^*} \right\| = \frac{1}{\gamma^*}$. Isso segue para todo i ,

$$y_i (\langle \hat{w}, x_i \rangle + \hat{b}) = \frac{1}{\|w_0\|} y_i (\langle w_0, x_i \rangle + b_0) \geq \frac{1}{\|w_0\|} \geq \gamma^* \quad (60)$$

Como $\|\hat{w}\|=1$ obtemos que (\hat{w}, \hat{b}) . É uma solução ótima da equação 56.

3.7.2 SVM de margem suave

A formulação SVM rígido assume que o conjunto de treinamento é linearmente separável, o que é uma suposição bastante forte. O SVM suave não linearmente separável pode ser visto como um relaxamento da regra SVM rígido, que pode ser aplicada mesmo se o conjunto de treinamento não for linearmente separável. O problema de otimização na Equação (57) impõe as restrições rígidas $y_i (\langle w, x_i \rangle + b) \geq 1$ para todos os i . Um relaxamento permite que a restrição seja violada para alguns dos exemplos no conjunto de treinamento. Isso pode ser modelado pela introdução de variáveis de folga não negativas, ξ_1, \dots, ξ_m , e pela substituição de cada restrição $y_i (\langle w, x_i \rangle + b) \geq 1$ pela restrição $y_i (\langle w, x_i \rangle + b) \geq 1 - \xi_i$. Isto é, ξ_i mede o quanto a restrição $y_i (\langle w, x_i \rangle + b) \geq 1$ está sendo violada. O SVM suave minimiza conjuntamente a norma de w (correspondente à margem) e a média de ξ_i (correspondente às violações das restrições).

A troca entre os dois termos é controlado por um parâmetro. Isso leva ao problema de otimização do SVM suave:

Entrada: $(x_i, y_i), \dots, (x_m, y_m)$

Parâmetros: $\lambda > 0$

Resolvendo:

$$\min_{w,b,\xi} \left(\lambda \|w\|^2 + \frac{1}{m} \sum_{i=1}^m \xi_i \right)$$

$$\forall_i, \quad y_i(\langle w, x_i \rangle + b) \geq 1 - \xi_i \text{ and } \xi_i \geq 0 \quad (61)$$

Saída: w, b

Reescrevendo a Equação (61) como um problema de minimização de perdas regularizado. Lembrando a definição da função de perda:

$$\rho^{hinge}((w, b), (x, y)) = \max\{0, 1 - y(\langle w, x \rangle + b)\}$$

Dado (w, b) e um conjunto de treinamento S , a função de perda média em S é denotada por $L_s^{hinge}((w, b))$. Agora, considerando o problema de minimização de perda regularizada:

$$\min_{w,b} (\lambda \|w\|^2 + L_s^{hinge}((w, b))). \quad (62)$$

A equação (61) e a equação (62) são equivalentes. Corrigindo w, b e considerando a minimização sobre ξ na equação (58), a melhor atribuição para ξ_i seria 0 se $y_i(\langle w, x_i \rangle + b \geq 1)$ e seria $1 - y_i(\langle w, x_i \rangle + b)$ caso contrário. Em outras palavras $\xi_i \rho^{hinge}((w, b), (x_i, y_i))$, para todo i , a afirmação segue. Assim, note que o *Soft-SVM* cai no paradigma de minimização de perdas regularizada. Um algoritmo *Soft-SVM*, ou seja, uma solução para a Equação (62), tem um viés em direção a separadores de baixa norma. A função objetivo que se pretende minimizar na Equação (62) penaliza não apenas erros de treinamento, mas também normas grandes.

Muitas vezes é mais conveniente considerar o *Soft-SVM* para aprender um meio-espaço homogêneo, onde o termo de viés b é definido como zero, o que produz o seguinte problema de otimização:

$$\min_w (\lambda \|w\|^2 + L_s^{hinge}(w)), \quad (63)$$

Em que:

$$L_s^{hinge}(w) = \frac{1}{m} \sum_{i=1}^m \max\{0, 1 - y\langle w, x_i \rangle\} \quad (64)$$

Conclui-se que o SVM é um algoritmo para aprender meio-espaço com um certo tipo de conhecimento prévio, ou seja, preferência por margem grande. O SVM rígido busca o meio-espaço que separa os dados perfeitamente com a maior margem, enquanto SVM suave ou não linear não assume a separabilidade dos dados e permite que as restrições sejam violadas até certo ponto (Shwartz; David, 2014).

3.7.3 Função Kernel

Uma função kernel calcula um produto interno em um espaço de alta dimensão (chamado espaço de características) sem a necessidade de calcular explicitamente a transformação dos dados para esse espaço. Esse conceito é crucial para métodos como a Máquina de Vetores de Suporte (SVM), em que o kernel permite realizar a classificação em espaços de alta dimensão, mesmo que os dados originais estejam em um espaço de baixa dimensão. É uma ferramenta usada para aumentar a flexibilidade de modelos lineares, permitindo a modelagem de relações não lineares entre as características dos dados cuja formulação segue abaixo:

Seja X um espaço de entrada. Um kernel $k: X * X \rightarrow \mathbb{R}$ é uma função para todo $x, x' \in X$, satisfaz:

$$k(x, x') = \langle \phi(x), \phi(x') \rangle \quad (65)$$

onde $\phi: X \rightarrow \mathcal{H}$ é uma função de mapeamento que leva o espaço de entrada X a um espaço de características \mathcal{H} , e $\langle \cdot, \cdot \rangle$, denotando o produto interno \mathcal{H} . Algumas das funções kernel mais comuns incluem:

- i. Kernel linear: $k(x, x') = \langle x, x' \rangle$
- ii. Kernel polinomial: $k(x, x') = (\langle x, x' \rangle + c)^d$
- iii. Kernel RBF (função de base radial): $K(x, x') = \exp\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right)$ (66)

As funções ii e iii permitem que algoritmos como SVM encontrem hiperplanos de separação em espaços de alta dimensão, viabilizando a separação de classes que não são

linearmente separáveis no espaço original. As vantagens do SVM é que são eficazes quando o número de dimensões é maior que o número de amostras; e capacidade de se adaptar a diversos tipos de dados e problemas ao utilizar diferentes funções de kernel.

3.8 Métricas de verificação dos modelos preditivos

Qualquer estimativa ou predição é uma inferência baseada no valor real, e não uma medida; portanto, é necessário verificar se o modelo apresenta resultados consistentes. Utilizar conjunto de teste e treino para testar os modelos analíticos não é suficiente, e efeitos indesejados podem ser mascarados, como *overfitting* e *underfitting*. É indicado utilizar mais uma métrica para avaliação. *Overfitting* é considerado um grande problema dos modelos de AM, devido à facilidade que ocorre em casos em que não se tome cuidado ao gerar um modelo. O *overfitting* ocorre quando o modelo memoriza rigorosamente os padrões de treino e falha ao prever dados de teste. Isso significa que um modelo que apresenta alta precisão em um conjunto de dados de treinamento pode não ter o mesmo desempenho com novas instâncias. Por exemplo, um classificador que alcança 100% de precisão nos dados de treinamento, mas somente 50% em dados de teste, está sofrendo de *overfitting* nos dados de treinamento (Carvalho, 2014). As causas do *overfitting* podem ser: conjunto de dados de treino que não representa adequadamente o banco de dados completo; base de dados muito pequena; e base de dados com muito ruídos. Já o *underfitting* acontece quando há uma simplificação excessiva dos dados pelo modelo que não consegue capturar as relações entre as características e o alvo. Ou seja, diferente do *overfitting*, que trabalha muito bem com o conjunto de treinamento, mas não com o banco de teste, o *underfitting* apresenta problemas no próprio conjunto de treino. Isso ocorre quando o modelo gerado por determinado algoritmo não consegue trabalhar bem com as instâncias do conjunto de dados utilizados (Figura 12).

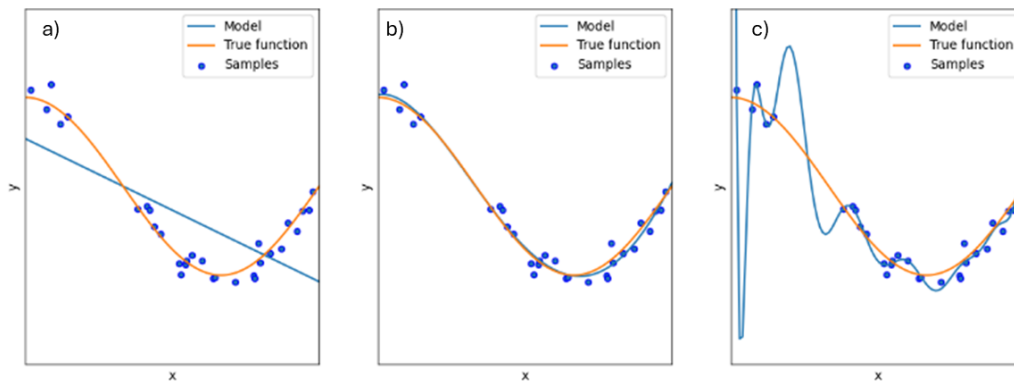


Figura 12. Em a) o efeito underfitting, cujo modelo não é suficiente para ajustar as amostras de treinamento (pontos). Em b) o modelo bem ajustado. Em c) o modelo polinomial aproxima a função verdadeira quase perfeitamente, aprendendo o ruído dos dados de treinamento (overfitting). Fonte: https://scikit-learn.org/stable/auto_examples/model_selection/plot_underfitting_overfitting.html

A validação cruzada é amplamente utilizada para evitar tanto o *overfitting* quanto o *underfitting*. A técnica de validação cruzada é utilizada para verificar se o conjunto de dados de treinamento está representativo o suficiente em relação à base de dados, na qual se deseja prever variáveis. Um tipo comum de validação cruzada é a *k-fold*, uma técnica de reamostragem na qual os dados são divididos em k partes iguais e utiliza cada uma das partes como conjunto de teste em uma das rodadas e como conjunto de treinamento nas demais (Müller *et al.*, 2016). Suponha que $k = 5$, então 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 13).



Figura 13. Exemplo de validação cruzada. Caixas cinzas correspondem ao grupo de teste, enquanto as brancas correspondem ao grupo de treino. No eixo x, são representados os diferentes grupos K . Fonte: (adaptado de Müller *et al.*, 2016).

A cada iteração realizada, a estimativa de erro do modelo, que consiste em quantas classificações incorretas foram feitas no conjunto de dados, é calculada. A estimativa total do erro do modelo é a soma das estimativas obtidas em cada iteração dividida por 5. Dessa forma, o erro de predição pode ser resumido por meio do erro médio absoluto (MAE) e do erro quadrático médio (MSE) observados nos bancos de teste.

4. ESTUDO DE CASO

4.1 A Mina Chapadão

A mina de fosfato em estudo, denominada Mina Chapadão, está localizada no município de Ouvidor–GO, próxima à cidade de Catalão–GO, no sudeste do estado, distante 263 km da capital Goiânia, pela BR352, e 300 km de Brasília (Figura 14). O complexo industrial Catalão I, do qual a mina faz parte, é acessado pela rodovia GO-210.

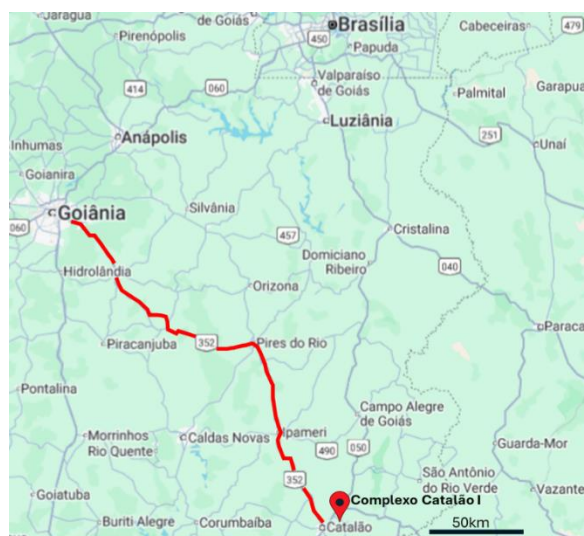


Figura 14. Localização geográfica do Complexo Catalão I.

A produção mineral no município de Catalão teve início a partir da década de 1970. Em 1976, a Mineração Catalão começou a exploração de nióbio na mina em Ouvidor (GO). No ano seguinte, a empresa realizou a produção da primeira liga de ferronióbio. Em 1984, já sob o controle da Copebrás, se inicia a produção de fertilizantes e fosfatados em Catalão, com uma ampliação em 2003, quando foi inaugurado o Complexo Mineraloquímico de Catalão, realizando a produção de fertilizantes e fosfatados. Ainda em 2003 a Copebrás também deu início à produção do Copefós – fosfato bicálcico

usado na composição de ração animal. Essas trocas de controle acionário se repetiram ao longo dos anos, e, em 2014, a Mineração Catalão e a Copebrás passam a integrar o capital ativo do grupo Anglo American, que foi posteriormente vendido em 2016 ao grupo chinês CMOC Internacional, o qual atua na mineração e no processamento de nióbio e fosfatos, minerais estratégicos para o desenvolvimento do mercado global e para a agricultura brasileira, ainda dependente de importação.

A mina Chapadão apresenta reserva lavrável em torno de 191.33 milhões de toneladas com teor médio de 12,5% de P_2O_5 . A vida útil da mina, considerando as informações sobre os recursos e as reservas de minério da companhia, é de 35 anos (CMOC, 2022). O método de lavra é a céu aberto, com uma cava de aproximadamente 2 km de raio E-W (Figura 15), que faz divisa a sudeste com a lavra de fosfatos da Mosaico Fertilizantes. A movimentação anual de material (*run of mine*) é da ordem de 6 milhões de toneladas e a produção é de 1,4 milhão t/ano (CMOC, 2022). O desmonte do minério ocorre quase que predominantemente de maneira mecânica. Perfuração por circulação reversa é utilizada para desmonte nas regiões de cimentação e para auxiliar a amostragem das frentes de lavra no curto prazo. A amostragem também é realizada por sondagem rotativa diamantada para atender à produção de curto, médio e longo prazo, seguido de análise química dos teores.

Toda lavra e produção de minério ocorrem no horizonte intemperizado. Embora exista alguns estudos e testes em rochas, dado o potencial de apatita controlado pelas ocorrências de foscórito, até o momento, esse depósito primário não está sendo explorado.

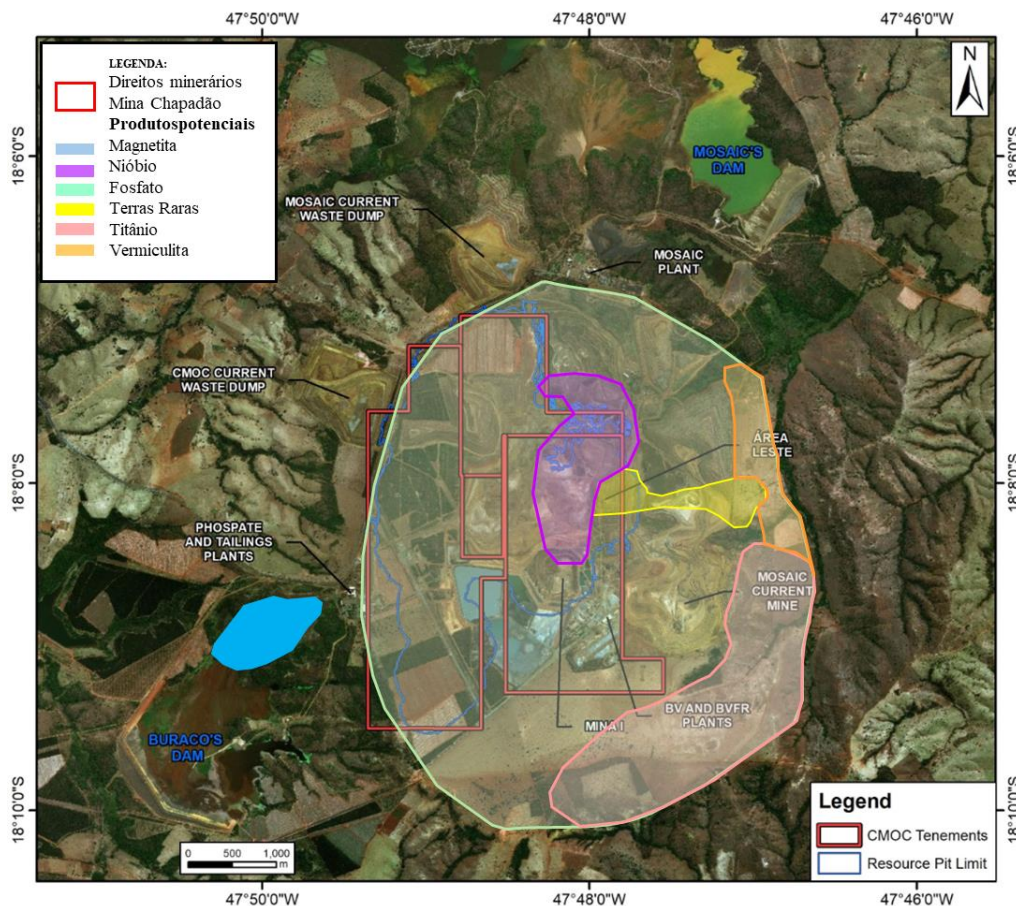


Figura 15. Produtos potenciais do complexo Catalão I. Em vermelho direito, minerário da mina Chapadão; em azul, o limite da cava. Adaptado de CMOC 2023.

4.2 Geologia

4.2.1 Protólitos

A Província Ígnea do Alto Paranaíba (PIAP), a qual o depósito Chapadão faz parte, é constituída por um conjunto de rochas ígneas kamafugíticas, kimberlíticas e carbonatíticas, que ocorrem sob a forma de diques, *pipes*, *plugs*, diatremas, derrames de lavas, depósitos piroclásticos e grandes complexos plutônicos (Gibson *et al.*, 1995). As intrusões deformaram as rochas metassedimentares da Faixa Brasília, quase sempre com geração de estruturas dômicas com aproximadamente 6 km de diâmetro. Segundo Gibson *et al.* (1995), os complexos alcalinos fazem parte de um contexto geológico maior associado a um intenso magmatismo ultrapotássico que atingiu a região durante o Cretáceo Superior. A Província Ígnea do Alto Paranaíba apresenta como limites na parte sudeste a bacia do Paraná e na parte nordeste o Cráton São Francisco (Figura 16). Estima-se que a quantidade de magma potássico nessa região a torna uma das maiores províncias alcalinas intracontinentais do mundo.

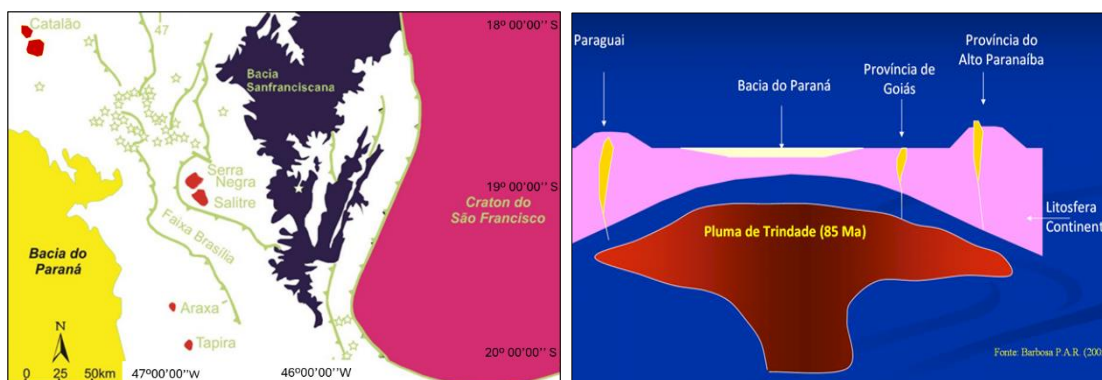


Figura 16. À esquerda, localização dos complexos alcalinos pertencentes à PIAP (modificado de Gibson et al., 1995); e à direita, condução do calor da pluma do manto subjacente (modificado de Barbosa, 2005).

A intensa atividade magmática que gerou a província tem, como hipótese mais aceita, o resultado do impacto da pluma mantélica de Trindade (Figura 16) sob a litosfera continental do Brasil Central, acerca de 90Ma, ocasionando a fusão parcial do manto litosférico metassomatizado sobrejacente e a produção volumosa de magmas ultrapotássicos (Gibson *et al.*, 1995; Brod *et al.*, 2004; Palmieri, 2011). Contrário a esses autores, Bizzi *et al.* (1995) defendiam que o *Hot Spot* Tristão da Cunha, que deu origem aos derrames basálticos da Bacia do Paraná, é também responsável pelo magmatismo alcalino da PIAP.

Segundo Ribeiro (2008), a classificação das rochas do Complexo Catalão I é difícil a considerar a classificação de rochas ígneas propostas por Le Maitre (1989), pois vários minerais, como apatita, perovskita, flogopita, magnetita e granada, que são os minerais essenciais nos complexos da PIAP, não são considerados nos modelos de classificações oficiais. Além disso, ocorreram transformações matassomáticas intensas nos minerais primários. Brod (1999), Cordeiro *et al.* (2010), Grasso (2010), Ribeiro (2008) e Palmieri (2011) descrevem três séries petrogenéticas para o Complexo Alcalino Catalão I: Bebedourítica, Foscorítica e Carbonatítica, relacionadas entre si por uma complexa combinação de cristalização fracionada, imiscibilidade de líquidos e metassomatismo.

As rochas da Série Bebedourítica são produzidas por cristalização fracionada de um magma silicático ultrapotássico e incluem bebedouritos, dunitos e piroxenitos. Essas rochas representam os termos finais da diferenciação magmática, com olivinas, flogopitas, perovskitas, magnetitas e apatitas. Segundo Brod *et al.* (2004), essa série está

frequentemente intrudida por *stockworks* de carbonatito e foscorito, o que ocasiona um metassomatismo potássico de intensidade variável e transforma as rochas ultramáficas da série bebedourítica em flogopititos. Os flogopititos derivados do metassomatismo apresentam-se potencialmente mineralizados em apatita e normalmente na cor castanho-avermelhado. Por outro lado, os flogopititos magmáticos, com raras exceções, não demonstram a mesma potencialidade. Aparentemente, alguns flogopititos magmáticos, mineralizados em apatita, o são às custas da injeção de veios de apatitito e foscorito, que ocorrem em todo complexo. Os flogopititos de origem magmática apresentam coloração preta ou cinza escuro (Ribeiro, 2008).

Os piroxenitos e os dunitos são raros e estão presentes na borda do domo, em núcleo centimétrico preservados no interior dos flogopititos (Brod, 1999; Brod *et al.*, 2004; Ribeiro, 2008). As rochas desse grupo possuem composição ultramáfica, predominando a serpentina, derivada da olivina e/ou piroxênio.

Segundo Tröger (1928), o bebedourito é um piroxenito rico em perovskita, com aproximadamente 54% de diopsídio, 21% de flogopita, 14% de perovskita, 10% de acessórios de apatita, feldspato potássico e olivina. Embora a composição típica descrita por Tröger contenha mais de 50% de diopsídio, diversos estudos sobre os bebedouritos na PIAP (Brod, 1999; Ribeiro, 2008; Grasso, 2010) mostram uma mineralogia mais variada, em que nenhum mineral atinge mais de 50% modal. Por exemplo, nos bebedouritos de Salitre e Tapira, a apatita é uma fase essencial, e não acessória (Grasso, 2015). Por essa razão, Brod *et al.* (2004) sugeriram um diagrama para classificação de bebedouritos adaptados para essas rochas da PIAP, conforme a Figura 17.

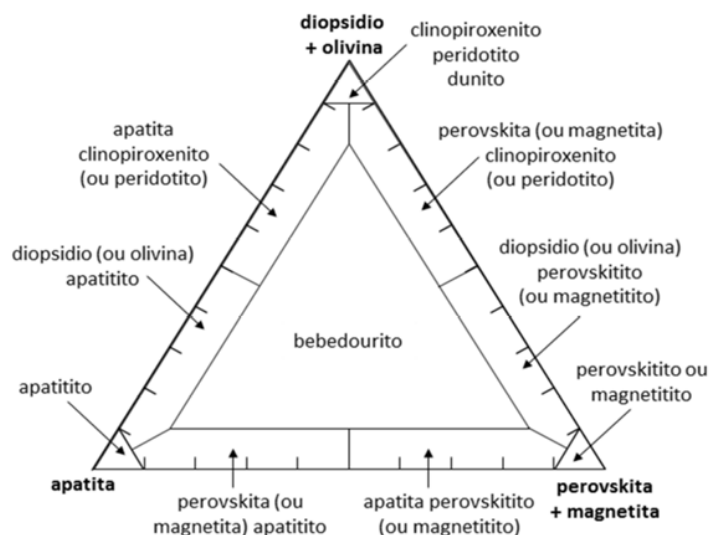


Figura 17. Classificação informal das rochas da série bebedourítica proposta por Brod et al. (2004).

A Série Foscorítica foi definida por Yegorov (1993) por variações modais de apatita, magnetita e olivina (Figura 18) e é constituída por rochas cristalizadas, a partir de resfriamento de magmas fosfáticos, representadas por: foscoritos, apatititos, nelsonitos e magnetititos. Os foscoritos observados no complexo são basicamente rochas de granulação média a grossa, compostos essencialmente por apatita, magnetita e um silicato magnésiano (flogopita primária ou pseudomorfos da olivina flogopitizadas). Segundo Brod (2004), nelsonitos podem ser considerados um equivalente do foscorito, porém mais pobre em silicatos, podendo conter ou não o mineral pirocloro (mineral de minério do nióbio). Carbonatos (calcita e/ou dolomita) são comuns, como material intersticial e como bolsões centimétricos. Já o Apatitito é uma rocha composta essencialmente de apatita, com pequenas quantidades de magnetita, flogopita e carbonato.

Essa série está presente por todo o depósito mais consistentemente entre a zona carbonatítica no centro e a zona de predominância ultramáfica nas bordas do domo. Conseqüentemente ocorre um aumento de apatita e magnetita da borda para o centro do complexo. Os foscoritos ocorrem em veios, diques, pequenos bolsões e *plugs* cortando diversos tipos de rochas, desde as silicáticas até as carbonatíticas (Ribeiro, 2008). A Série Foscorítica é importante tanto do ponto de vista petrológico quanto do ponto de vista econômico, pois essas rochas são responsáveis em grande parte pelas duas minas em exploração em Catalão: nióbio e fosfatos (Ribeiro, 2008).



Figura 18. Classificação das rochas da série foscorítica segundo Yegorov (1993).

Brod (2004) define a Série Carbonatítica por rochas compostas por mais de 50% em volume de carbonatos, dolomita, calcita, ou dolomita/ankerita, com grande variedade de acessórios. Definir o tipo de carbonato presente em descrição petrográfica nem sempre é possível, sendo necessário uma análise química para classificação como visto em Woolley & Kempe (1989) (Figura 19).

De maneira simplificada, os carbonatitos do complexo podem variar de Ca-carbonatitos a Mg-carbonatitos, sendo os grandes volumes presentes no depósito de dolomita carbonatitos. Os Ca-carbonatitos ocorrem com maior frequência como glóbulos aglutinados de calcita, formando bolsões carbonatíticos que alimentam veios e diques espalhados por todo o complexo (Ribeiro, 2008).

A ocorrência da apatita difere nos carbonatitos. Nos calcita carbonatitos, a apatita forma, predominantemente, cristais prismáticos curtos, com terminações arredondadas, disseminados na rocha. Nos carbonatitos dolomíticos, a apatita ocorre preferencialmente em bolsões monominerálicos, às vezes de contornos ameboides, sugerindo tratar-se de um produto de imiscibilidade de líquido fosfático a partir do magma carbonatítico.

Rochas de origem carbo/hidrotermal também estão presentes no complexo, ocorrendo como brechas e veios, cuja principal característica é a grande quantidade de magnetita, que, adicionada à dolomita presente, eleva o teor de MgO consideravelmente para valores superiores a 40%. O termo hidrotermalismo tem o mesmo sentido de metassomatismo, que é uma alteração na composição química da rocha por adição ou remoção de material pela ação dos fluidos carbo-hidrotermais (Ribeiro, 2008).

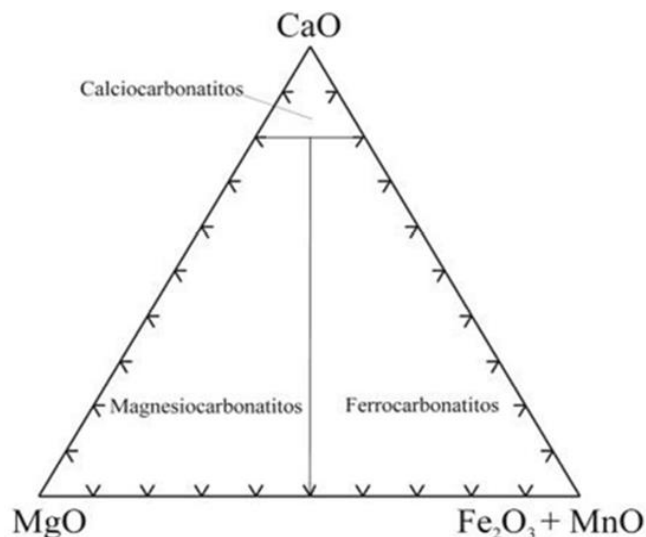


Figura 19. Classificação química de rochas da série carbonatítica segundo Woolley & Kempe (1989).

No complexo em geral, ocorrem também brechas de conduto, as quais são produzidas por incorporação de fragmentos (xenólitos) de rochas preexistentes durante eventos de magmatismo explosivo. Sua composição varia amplamente em função da carga de xenólitos, podendo ser a matriz constituída predominantemente por carbonatito ou flogopita picrito.

A intrusão magmática provocou deformação dômica e fenitização dos xistos e quartzitos encaixantes do Grupo Araxá. O Complexo Catalão I pode ser dividido, segundo a disposição espacial, em três grandes unidades (Ribeiro, 2008):

- A parte central onde predominam os carbonatitos e foscoritos sobre as rochas silicáticas;
- A parte intermediária onde foscoritos e carbonatitos se equivalem em proporção às rochas silicáticas;
- A região da borda onde predominam as rochas de filiação silicática sobre carbonatitos e foscoríticos.

Os trabalhos de mapeamento da mina e a descrição dos testemunhos de sondagem sintetizaram as principais feições geológicas locais, sendo duas sequências de cristalização fracionada, dominadas pela cristalização e acumulação de calcita, magnetita, apatita e flogopita (Figura 20), as quais, para efeitos de referenciamento, são denominadas de sequência norte e sequência sul. A principal diferença mineralógica entre as duas sequências é a maior abundância de calcita acumulada na sequência sul, o que resulta, do ponto de vista petrográfico, na maior abundância de calcita carbonatitos intercalados com

os foscoritos, enquanto, na sequência norte, nelsonito é o tipo petrográfico mais abundante, que ocorre intercalado com os foscoritos. Intrusões magmáticas posteriores de dolomita carbonatitos recortam as sequências cumuláticas na forma de plugs, diques e veios, sendo mais abundantes na região sul e nordeste (CMOC, 2023).

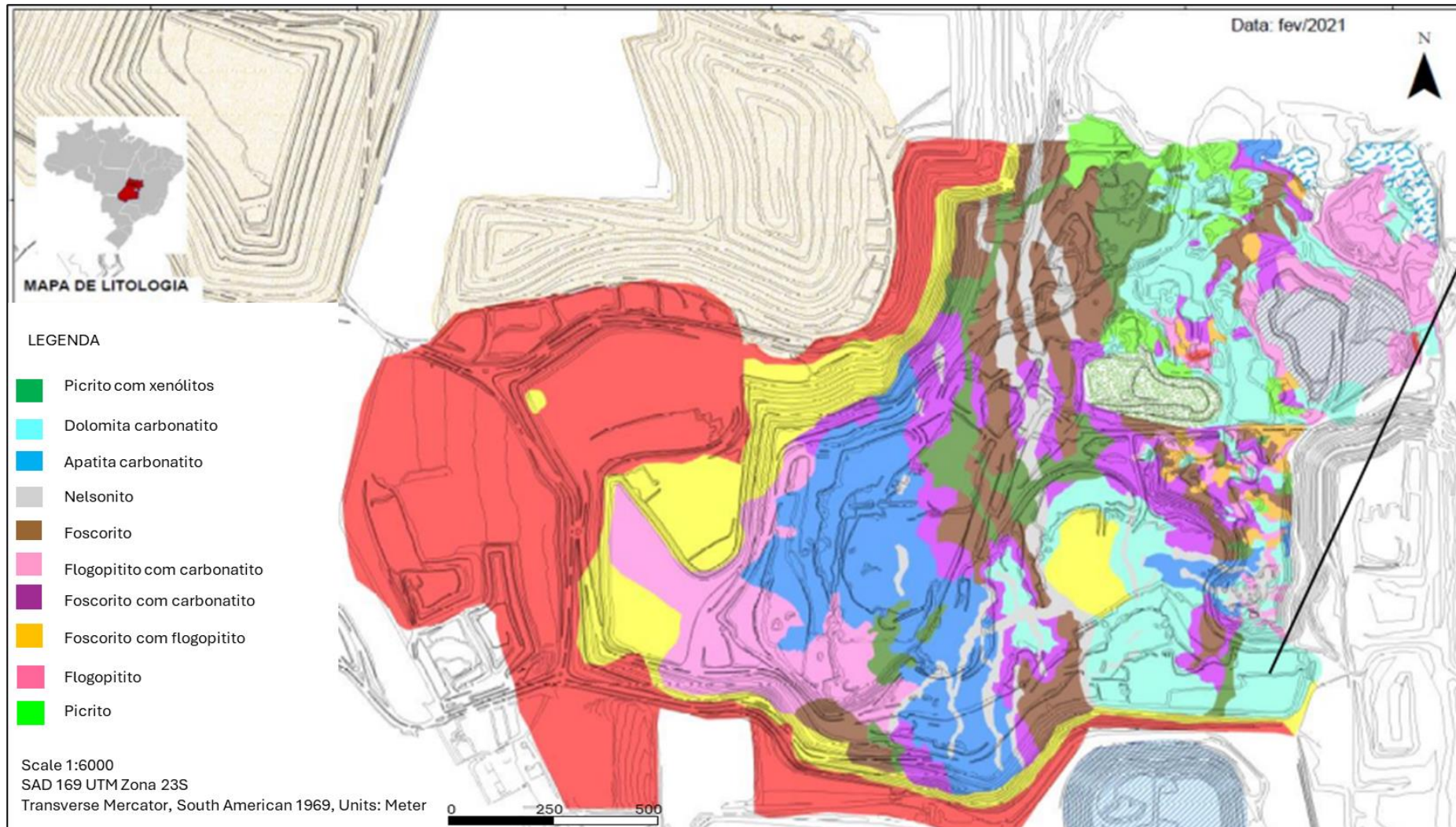


Figura 20. Mapa geológico da mina. Modificado de CMOC (2023).

4.2.2 Perfil de alteração intempérica

As reservas de fosfatos no Brasil estão relacionadas à alteração das rochas ígneas, onde um espesso manto de intemperismo com aproximadamente 150 m de profundidade se desenvolveu. Os protolitos têm originalmente teores elevados de nióbio (Nb) e fósforo (P) e são facilmente alteráveis, pois seus constituintes principais são carbonatos.

O perfil intempérico de Catalão I foi descrito por Oliveria e Imbernon (1998), com base em características petrológicas e químicas, e foi dividido em quatro horizontes: rocha alterada, saprólito isalterito, saprólito aloterito e solo. A rocha alterada é o horizonte com preservação parcial dos carbonatos. O saprólito isalterito é caracterizado pela lixiviação de carbonatos e por uma maior concentração residual de apatita e pela presença de goethita, quartzo e pequenas quantidades de fosfatos aluminosos. Esse horizonte ainda é caracterizado por baixos conteúdos de MgO e elevados teores de P_2O_5 e CaO. A apatita quase não apresenta indícios de alteração, mostrando texturas similares às observadas nas apatitas da rocha fresca. Já nas porções mais superiores desse horizonte, limítrofe ao aloterito, a superfície das apatitas pode apresentar feições de corrosão, que indicam dissolução incipiente. No saprólito aloterito, ocorre a dissolução completa da apatita e a formação de goethita, gorceixita e quartzo. Esse horizonte é marcado por conteúdos elevados de Fe_2O_3 e TiO_2 . Todo o P_2O_5 ocorre sob a forma de fosfatos aluminosos secundários. A cobertura superficial é constituída por caulinita, goethita, quartzo, fosfatos aluminosos e anatásio, ou seja, teores elevados de SiO_2 e Al_2O_3 e Fe_2O_3 (Oliveira; Imbernon, 1998).

A classificação dos horizontes intempéricos adotado na Mina Chapadão segue conceitos apresentados por Oliveira e Imbernon (1998) e foram adaptados pela equipe de geologia da CMOC. Esse modelo esquemático admite da base para o topo o seguinte perfil vertical: rocha sã ou fresca, rocha alterada, isalterito, aloterito e solo (Figura 21).

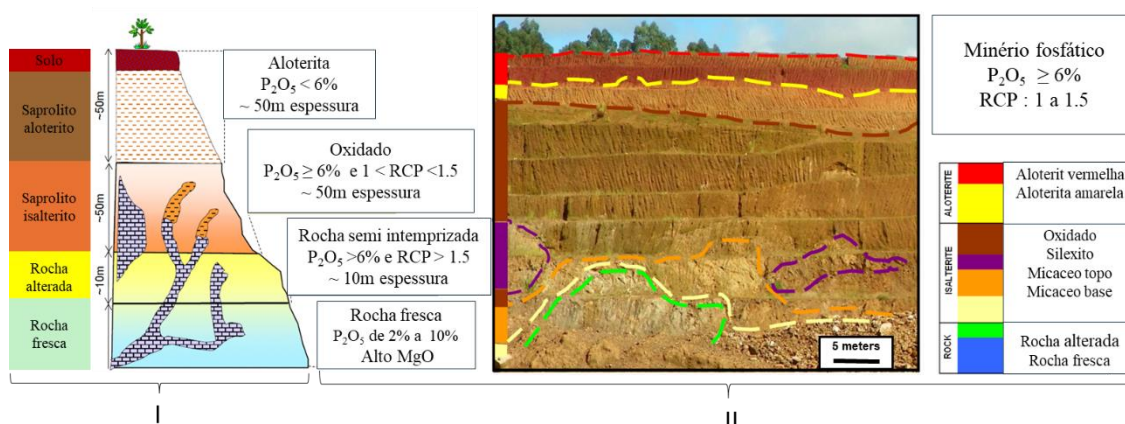


Figura 21. Em I, o perfil intempérico proposto por Oliveira e Imbernon (1998); em II, adaptação de Ribeiro (2008) para as classes de minérios.

A zona de Rocha Alterada tem apatita em potencial, porém ainda não existe uma rota de processo economicamente viável. Outras mineralizações de titânio, terras raras e vermiculitas (Figura 15) também não foram, até o momento, viabilizadas economicamente. Nesse horizonte, as feições e as características geométricas e estruturais da rocha fresca são preservadas, e se diferencia pela sua coloração verde-amarelado ou verde-acinzentado, possivelmente decorrente da presença de serpentina e clorita. Apresenta um pequeno enriquecimento em Fe_2O_3 e TiO_2 e empobrecimento em K_2O comparativamente à rocha fresca. Os minerais constituintes principais são apatita, flogopita, magnetita, ilmenita, perovskita, calcita, dolomita e quantidades subordinadas de vermiculita. A presença de carbonatos preservados ocorre pela baixa taxa de intensidade de intemperismo, que ocorre nesse horizonte. É caracterizado quimicamente por conteúdos de $\text{P}_2\text{O}_5 > 6\%$, e relação $\text{CaO}/\text{P}_2\text{O}_5 \geq 1,5$ (RCP). Os teores de P_2O_5 e MgO e a RCP são critérios para classificação de minérios na mina. A relação $\text{CaO}/\text{P}_2\text{O}_5$ é amplamente utilizada por indicar a presença de carbonatos ($\text{RCP} > 1,5$). O minério micáceo situa-se na parte inferior do horizonte saprolito isalterito. É formado por apatita, flogopita, magnetita, quartzo, ilmenita e vermiculita (em quantidades que vão aumentando da base para o topo). Ainda podem ser reconhecidas as feições da rocha fresca, porém nesse nível não ocorre mais a presença de carbonatos, evidenciada por uma coloração mais amarelada. O minério micáceo possui teores de $\text{P}_2\text{O}_5 \geq 6\%$, e relação $\text{CaO}/\text{P}_2\text{O}_5 < 1,5$. Acima do minério micáceo está o minério oxidado que é formado por apatita, fosfatos secundários aluminosos, magnetita, ilmenita, anatásio, óxidos/hidróxidos de ferro e argilominerais. Caracterizado geoquimicamente por conteúdos de $\text{P}_2\text{O}_5 \geq 6\%$, $\text{MgO} < 3\%$ e RCP entre 1% a 1,5%, possui coloração variando do ocre, do róseo ao marrom escuro e uma natureza friável, geralmente arenosa a areno-siltosa. Há bolsões micáceos e algumas massas e vênulas de silicificação. Localmente, há domínios ricos em magnetita ($> 50\%$) e apatita de granulação fina, bem como porções silicosas (silexitos e brechas silicosas) e domínios apatito-magnéticos maciços (cimentados) contendo até 17% de P_2O_5 . O horizonte oxidado apresenta melhor desempenho na usina, com alto teor de P_2O_5 e baixo MgO , além de pouca lama. A aloterita é o horizonte decorrente da evolução do saprolito isalterítico e estéril da mina. Sua espessura se aproxima de 25 m. É um nível mais homogêneo, composto de material argilo-arenoso solto, amarelo-avermelhado, em que as estruturas originais não podem mais ser identificadas.

O processo de laterização age concentrado a apatita, ao mesmo tempo que gera modificações químicas e morfológicas em toda a paragênese mineral. Essas alterações

impactam o processo de beneficiamento mineral, ou seja, apesar das características químicas verificadas serem próximas do normal (P_2O_5 e CaO/P_2O_5) ao longo do perfil isalterito, estudos evidenciaram composições, morfologias e texturas que prenunciam um comportamento diferente da apatita original durante os processos industriais de concentração (Toledo *et al.*, 2001). Segundo Toledo *et al.* (2001), são as propriedades da apatita que determinam quão bem ou quão mal a apatita pode ser recuperada industrialmente. As apatitas primárias de origem ígneas são limpas, com formas granulares, ovoides e arredondadas, variando de vítreas a translúcidas, hábito prismático, com poucas inclusões e estrias de dissolução. Já apatitas secundárias (hidrotermal) contêm, em sua superfície, inclusões e revestimentos, principalmente de óxidos e hidróxidos de ferro, bem como de óxidos de magnésio. Ocorrem em agregados semiesféricos com aspecto “rugoso” ou apresentam um aspecto “nebuloso”. Em alguns casos, formam estrias.

Outro exemplo relacionado às modificações morfológicas e químicas dos minerais são os problemas associados à presença de filossilicatos no beneficiamento. Segundo Azzone (2010), em algumas usinas, foram relatadas perdas da ordem de 20% na recuperação de P_2O_5 , quando identificado aumento dos filossilicatos. A presença de flogopita, mais comum na base da zona de minérios, acarreta uma maior resistência desse mineral à cominuição, aumentando o tempo de moagem e diminuindo a eficiência na deslamagem quando comparada aos argilominerais de topo, a exemplo da vermiculita. Portanto, o mapeamento e a modelagem dos protolitos é fundamental para entender as modificações esperadas nos horizontes de alteração.

4.3 Processamento mineral

O beneficiamento dos minérios fosfáticos de ocorrência na PIAP são geralmente semelhantes e possuem as seguintes etapas: britagem, empilhamento e homogeneização, moagem primária e separação magnética de baixo campo, moagem secundária e classificação, deslamagem, concentração por flotação e espessamento, separação magnética de alto campo (Luz; Lins, 2005). O processo de concentração da planta de fosfato de Chapadão está representado no fluxograma na Figura 22.

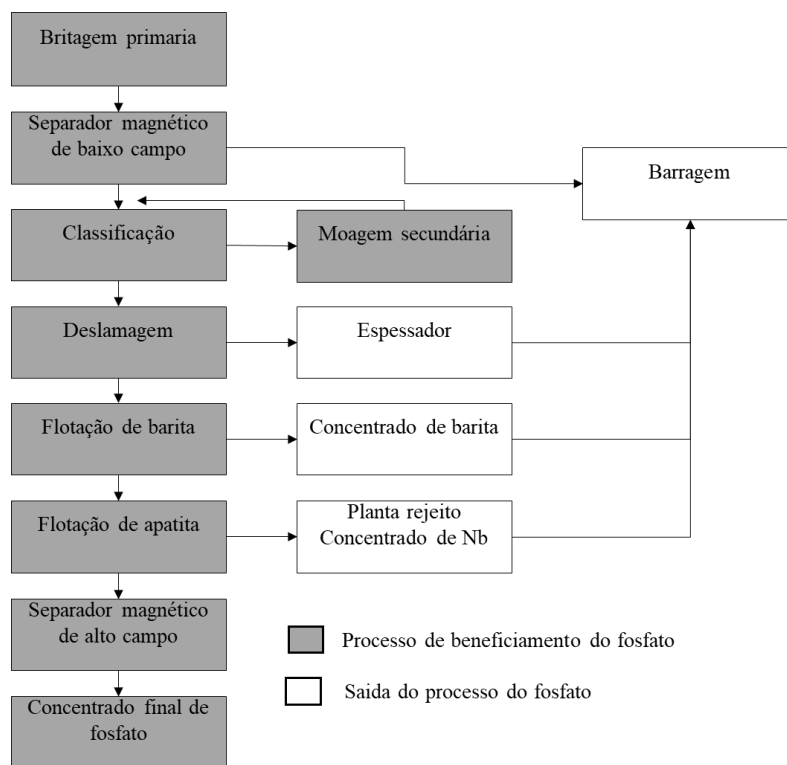


Figura 22. Fluxograma do processo de concentração mineral de fosfatos.

4.3.1 Ensaios de flotabilidade

O teste de flotabilidade reflete a sequência do processo de beneficiamento da planta em escala de bancada. Inicia-se com a moagem das amostras de furo de sondagem em moinho de barras para classificação da granulometria em 65 mesh (0,021 mm). Depois de moída e classificada na malha de corte, a amostra é deslamada em hidrociclone de 40mm, onde se separa a lama, ou seja, partículas de tamanho inferior a 30 μm . A amostra deslamada é então submetida ao processo de separação magnética de baixa intensidade, a uma intensidade de campo de 1.000 Gauss, onde se separa a magnetita.

Após as etapas de separação física, a amostra é submetida ao processo de flotação de barita, onde adiciona-se coletor de barita e espumante. O concentrado de barita é então quantificado e seu teor analisado quimicamente. O afundado da flotação de barita é destinado ao processo de flotação de apatita. A polpa destinada ao processo de flotação de apatita é adicionada em uma proveta de volume e peso conhecido. Através da densidade específica do sólido, é possível calcular a densidade da polpa e o percentual de sólidos da amostra. O objetivo é que a flotação de apatita na etapa *rougher* ocorra nas condições de 47% a 50% de sólidos. Com o conhecimento da massa disponível na cuba, calculam-se os volumes de reagentes a serem adicionados durante o teste. O próximo passo é a flotação propriamente dita, onde

adiciona-se a polpa à cuba e, em seguida, o depressor. Faz-se o condicionamento, sob agitação durante dois minutos, e adiciona-se à polpa 50% da dosagem estimada de coletor de apatita, condicionando a polpa sob agitação durante meio minuto. Após o condicionamento, adiciona-se o tensoativo. Em seguida, ajusta-se o ar e inicia-se a coleta de espuma/flotado durante dois minutos: o flotado é separado para ser flotado subsequentemente na etapa cleaner, porém, antes disso, ao afundado da etapa *rougher*, adiciona-se o restante de 50% da dosagem estimada de coletor e condiciona-se este durante meio minuto. Depois, ajusta-se o ar e inicia-se a coleta de espuma de concentrado *scavenger* durante um minuto e meio. O concentrado *scavenger* e o rejeito *scavenger*, também denominado rejeito final, são separados para quantificação em massa e análise química dos teores através da fluorescência de raios X. O concentrado *rougher* separado é disposto em uma cuba menor para realização da etapa de flotação *cleaner*, nesta etapa não há adição de reagentes apenas coleta de espuma, esse produto flotado é o concentrado final da flotação de apatita. O concentrado flotado na etapa de flotação de apatita passa pelo último estágio de concentração denominado separação magnética de alta intensidade, a um campo magnético de cerca de 10.000 Gauss com objetivo de retirar a fração magnética de óxidos de ferro ainda presente no concentrado de apatita e atender às especificações em teor de óxido de ferro (Teixeira, 2019).

Ao final do processo de flotação os produtos são separados, e a massa é quantificada e analisada quimicamente para se obter os teores dos elementos no concentrado através da técnica de fluorescência de raios X. A recuperação mássica do minério testado em bancada é denotada por DIST_MAS e representa o percentual da massa de alimentação no teste de bancada que se tornou produto. As variáveis obtidas nos testes podem ser vistas na Tabela 1. Os resultados das amostras de furo de sondagem ensaiadas são armazenados em planilhas para compor as informações do longo prazo.

Tabela 1. Variáveis de processo obtidas nos ensaios de bancada.

SIGLAS	DESCRIÇÃO
P2O5_C	% Teor de P2O5_C estimado para o concentrado de fosfato
CAO_C	% Teor de CAO_C estimado para o concentrado de fosfato
FE2O3_C	% Teor de FE2O3_C estimado para o concentrado de fosfato
MGO_C	% Teor de MGO_C estimado para o concentrado de fosfato
AL2O3_C	% Teor de AL2O3_C estimado para o concentrado de fosfato
SIO2_C	% Teor de SIO2_C estimado para o concentrado de fosfato
NB2O5_RF	% Teor de Nb2O5 no rejeito final de fosfato
LAMA	% de massa de lama
MMAG	% de massa do material magnético do separador de baixo campo
DIST_MASS	Recuperação mássica (%)

5. METODOLOGIA

Existem na literatura diferentes técnicas para modelar atributos geometalúrgicos. Para esse trabalho, foi utilizada uma função de transferência representada pelo modelo *Random Forest* em que as variáveis respostas, recuperação mássica, teor de P_2O_5 no concentrado e massa de magnético são obtidas por simulação sequencial das variáveis primárias. Um banco de dados de sondagem de longo prazo foi fornecido pela CMOC com dados até a data de 03/2022, contendo informações de teores, geologia e ensaios de flotabilidade. Os dados referentes aos teores e atributos metalúrgicos tiveram as grandezas numéricas ajustadas por uma constante a fim de preservar a confidencialidade da informação. O trabalho foi dividido em cinco etapas.

Etapa 1 – Análise exploratória

A análise exploratória do banco de dados é importante para verificar questões relacionadas ao suporte amostral, sobreposição de dados, verificação de coordenadas, valores anômalos e descrição geológica. Após a conferência, o banco de dados totalizou em um mapa geológico das frentes de lavra na escala 1:6.000 e 44.803 amostras de furos de sondagem com amostras regularizadas no suporte amostral de 5 metros. Nessas amostras, 18.583 são isotópicas para informações geológicas e de teores, e apresentam descrição de 32 litotipos com análise química de até nove óxidos. Os ensaios de bancada com resultados para todas as variáveis de processo estão presentes em 1.548 amostras. Três pontos relacionados ao banco de dados merecem destaques:

- O primeiro é sobre as descrições litológicas dúbias (mais de uma litologia reconhecida por amostras), resultando em 32 combinações litológicas presentes em 53% do banco. A causa das incertezas nas descrições tem várias origens, como mudanças de procedimentos, adequação de conceitos, suporte amostral restrito dos furos de sondagem e a gênese do depósito. Essa última é a mais significativa, pois as rochas ocorrem preferencialmente na forma de diques finos que se recortam em diferentes direções. Ou seja, unidades expressivas de uma única litologia não ocorrem nesses tipos de depósitos, que apresentam, além da variabilidade lateral dos protólitos, variações verticais intensas pela ação intempérica. Tudo isso dificulta o direcionamento das amostras para compor os grupos tipológicos.

- O segundo é a redução significativa das informações de processo ao considerar a isotopia dos atributos medidos (Figura 23). Ou seja, os testes de bancada não são comuns a todas as litologias, e as quantidades de testes não foram igualmente divididas. Além disso, a distribuição espacial das amostras não é uniforme, formando amostragem preferencial nas regiões de lavra. Esse é um problema comum em mineração, e pode ser amenizado quando as equipes de longo e curto prazo, da geologia ao processo mineral, definem juntas as amostras para os testes de bancada. Ainda sobre os atributos de processo, a equipe da CMOC informou vieses nas variáveis recuperação metalúrgica e Nb_2O_5 no rejeito final, e por isso não foram consideradas neste trabalho¹.

- O terceiro ponto está relacionado ao tratamento dos dados metalúrgicos e a não aditividade e não linearidade de algumas variáveis. Essas características restringem o uso de estimadores lineares, e dificulta comparações dos resultados de bancada com os dados de produção.

2113 furos de sondagem - 44.803 amostras		
Geologia	Litologia	28 053
	Perfil intemperico	34 920
Isotopia dados de geologia		24 737

2113 furos de sondagem - 44.803		
Químico	NB2O5	42 305
	P2O5	41 580
	SIO2	34 498
	FE2O3	34 485
	BAO	34 259
	CAO	34 398
	MGO	34 321
	TiO2	25 357
	AL2O3	33 278
Istotópico teores		23 753
Isotopia dos dados de geologia e químico		18 584

2113 furos de sondagem - 44.803		
Processo	P2O5_C	8 017
	CAO_C	8 015
	FE2O3_C	8 016
	MGO_C	7 239
	DIST_MAS	7 932
	LAMA	8 006
	MMAG	8 007
	AL2O3_C	7 290
	SIO2_C	7 944
NB2O5_RF		2 326
Isotopia dados geologia, químico e processo total		1 548
Isotopia dos dados geologia, químico e parcial processo (P2O5_C, MMAG e DIST_M)		4473*

* Ao final das análises multivariadas e seleção de variáveis, esses foram os atributos processo selecionados para modelagem geometalúrgica

Figura 23. Informações contidas no banco de dados.

Com o banco de dados verificado empregou-se técnicas de estatística e análise multivariada no Minitab® como: matriz de correlação, estatística descritiva (histogramas, *box plot*), análise de principais componentes e análise de agrupamento (hierárquico e *k-means*). O objetivo foi compreender as relações entre as variáveis, e selecionar as variáveis primárias mais importantes.

Etapa 2 – Agrupamento de domínios

Em trabalhos de modelagem 3D se faz necessário a definição de domínios estacionários e para isso, é necessário separar unidades geológicas. Em operações da lavra, os trabalhos de mapeamento geológico, aliado ao planejamento de curto prazo são utilizados para definir grupos geológicos e direcionar as atividades da mina. Portanto, mapas geológicos fornecem, por vezes, informações preciosas acerca do modo de ocorrência do minério, sua distribuição geométrica, variação mineralógica (textura, granulometria, associações minerais), relação de contato entre estéril e minério, presença de descontinuidades, entre outras. Por essa razão, as informações do mapeamento geológico são necessárias para suprir as necessidades das frentes de lavra, e devem compor a base de conhecimento dos modelos de blocos.

Na Mina Chapadão, o mapeamento geológico é realizado na escala 1:6000. As tipologias mapeadas são: aloterita vermelha e amarela, flogopititos, flogopititos com veios de carbonatito, foscorito, foscorito com veios de carbonatito, foscoritos com flogopititos, picrito, picrito com xenólitos, dolomita carbonatito, apatita carbonatito e nelsonito. Essas tipologias podem ser organizadas em domínios predominantes, seja pela semelhança geológica ou pelo fato de que nem todas as tipologias são mineralizadas e consideradas na composição das pilhas formadas por *blends* de minérios.

Uma análise multivariada preliminar utilizando os algoritmos de agrupamento *k-means* e hierárquico foi realizada a fim de direcionar o número ideal de domínios fundamentados em análises puramente estatísticas. O método Elbow, que envolve uma análise visual gráfica sobre o número de grupos com a variância média entre eles, foi usado para indicar a melhor configuração de agrupamentos. Essa análise considerou no primeiro momento somente as variáveis químicas, já que essas são analisadas em todas as amostras, independente de minério ou estéril. Num segundo momento, foram consideradas todas as variáveis de processo e suas análises químicas. Na Figura 24, é possível verificar que oito grupos (no universo geológico) a quatro grupos (no universo das unidades mineralizadas em apatita) são suficientes para explicar a variabilidade do fenômeno. De fato, a equipe de curto prazo reconhece em termos de importância para a mineralização de fosfato quatro unidades geológicas: foscorito, foscorito com carbonatito, apatita carbonatito e flogopititos.

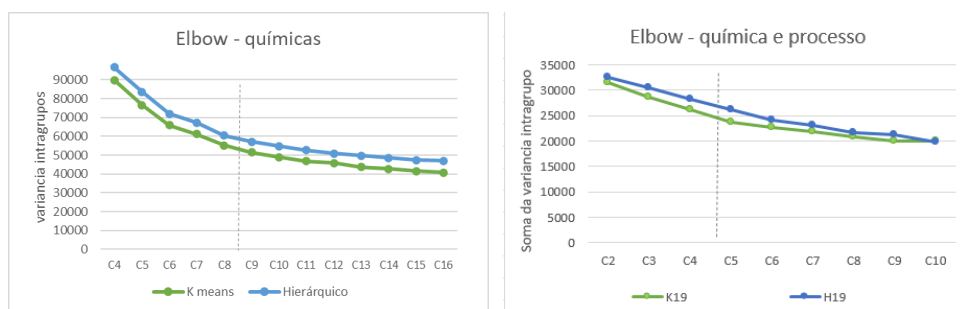


Figura 24. Análise de agrupamento pelos métodos k-means e hierárquico.

Com base no conhecimento especializado da equipe da mina, no mapa geológico e nas análises estatísticas, estabeleceu-se que oito domínios tipológicos seriam modelados (Figura 25).

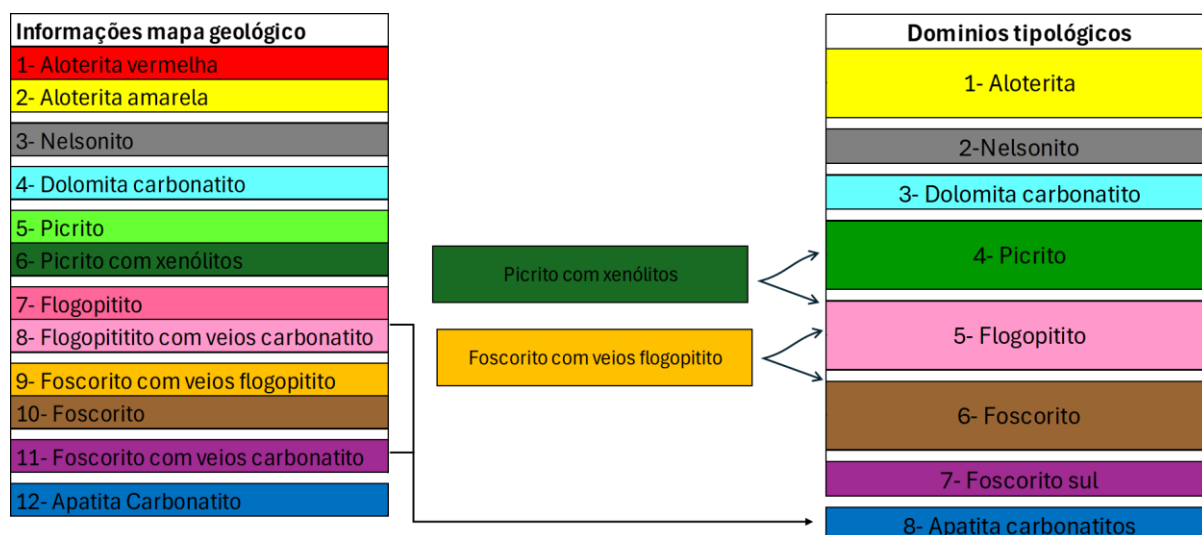


Figura 25. Informações geológicas da mina Chapadão e redefinição dos domínios.

Porém, como já foi dito, as litologias com forte sobreposição de veios são mais difíceis de serem identificadas nas descrições de sondagens devido ao suporte amostral limitado, diferente do que acontece no mapeamento geológico, onde a visão espacial permite uma maior assertividade. Por essa razão a teoria dos conjuntos fuzzy foi testada na tentativa de apresentar uma abordagem viável para lidar com a incerteza do fenômeno geológico e com ambiguidade das descrições das amostras. A implementação do algoritmo foi realizada em Python[®].

É sabido que minerais e rocha são caracterizados por uma estrutura cristalina e uma composição química definida (Teixeira et. al., 2009) portanto, uma combinação entre as análises estatísticas dos teores e o conhecimento geológico foi realizada para montar o algoritmo de agrupamento fuzzy. Uma adaptação do fenômeno estudado para um sistema de

modelagem fuzzy foi realizada cuja classificação das unidades pode ser representada como uma função $f: R(n+1) \rightarrow R$, com $y = f(x)$, onde a saída do sistema y é o domínio a qual a amostra pertence em função da entrada de x , com dimensão $n+1$, em que x_1, \dots, x_n são as expressões químicas extraídas das variáveis significativas para cada tipologia. Assim, o processo de agrupamento pode ser conhecido por uma função, mas não se tem uma expressão matemática explícita que relacione as variáveis de entrada com as variáveis de saída. É nessa etapa que o conhecimento especializado é utilizado, para dosar as relações entre os parâmetros de entrada (teores) com as informações geológicas.

Na linguagem fuzzy estas relações são escritas via termos linguísticos, como no exemplo: “As rochas **dolomita carbonatito** do depósito possuem alto teor de **MgO**”. O problema é como determinar o menor inteiro considerado alto (3, 5, 10, ...). Ao invés de determinar o menor inteiro pertencente ao conjunto, associa-se diferentes grau de pertinência para os teores daquele conjunto, obtendo valores parciais ou graduados no intervalo $[0, 1]$, que agora se torna uma função de pertinência.

Em regiões com maior taxa de transição ou sobreposição de veios a variabilidade dos teores é maior e a média local oscila muito. Por essa razão um subconjunto das amostras representas por 47% da base de dados original foi gerado contendo somente as amostras com uma única descrição tipológica, consideradas mais uniformes ou precisas. Os teores foram analisados sobre critérios estatísticos a fim de criar uma identidade química para cada domínio predominante e para formar a base de regras fuzzy.

Por exemplo, na Figura 26 é possível observar que teores elevados de P_2O_5 (acima de 10%) são características dos foscoritos e caso contrário (teores muito baixos) são típicos da dolomita carbonatito e a aloterita. A média, mediana e os quartis foram considerados para definir os conjuntos de números reais a, b, c, d (Figura 27) que são os parâmetros da função de pertinência fuzzy e para contornar *outliers*. Diferentes valores para os conjuntos de números reais foram testados resultando em vinte simulações até obter os melhores resultados dos limiares da função de pertinência. O processo de fuzzificação, transformou os valores das variáveis de entrada para a linguagem usual, ou seja, os teores foram classificados qualitativamente em termos linguísticos (baixo, médio e alto), utilizando o conhecimento especialista e, quantitativamente, atribuindo-se um grau de compatibilidade aos termos linguísticos através da função de pertinência trapezoidal (Figura 27).

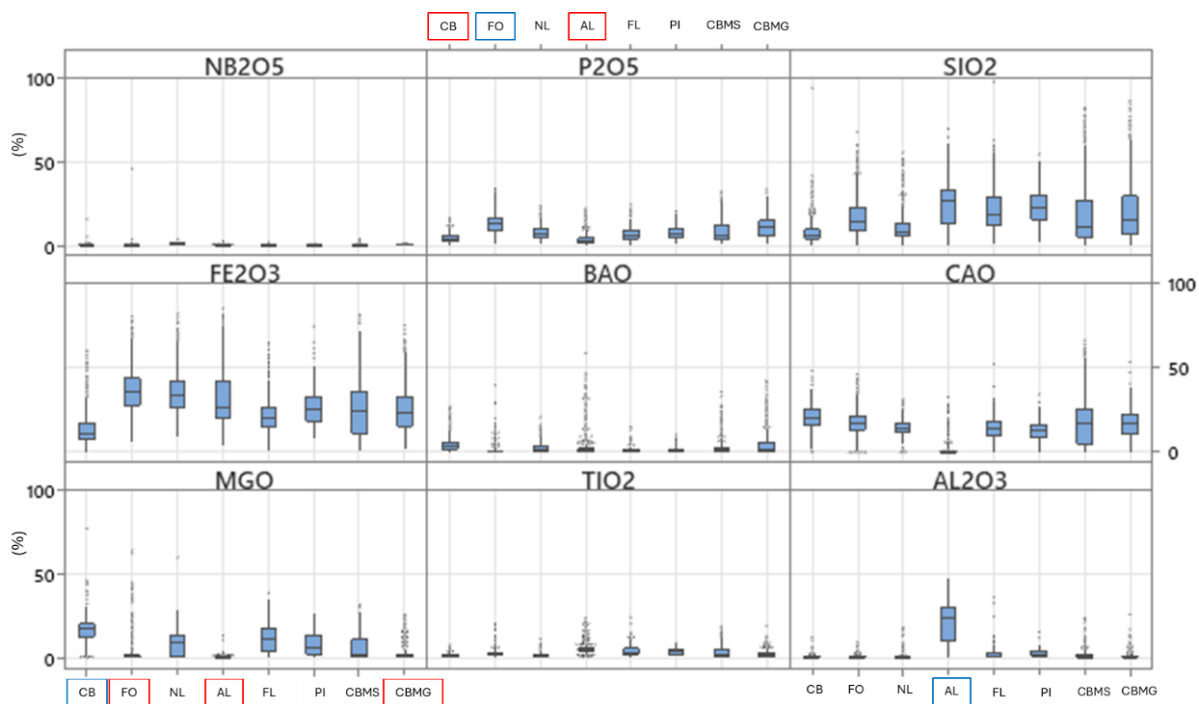
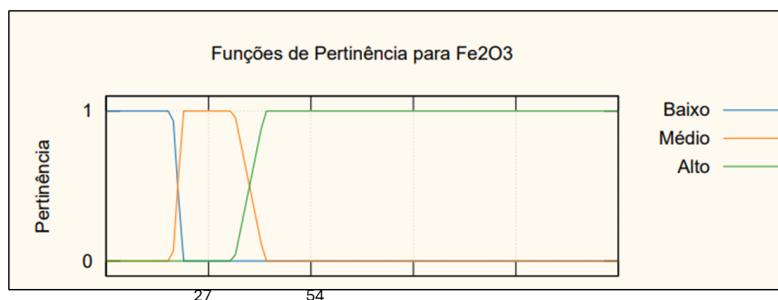


Figura 26. Análise estatística dos teores utilizada como entrada *crisp* no modelo fuzzy.



Teor	Baixo				Médio				Alto			
	a	b	c	d	a	b	c	d	a	b	c	d
NB2O5	-10	-9	0.3	0.5	0.3	0.5	0.7	0.8	0.7	0.8	110	120
P2O5	-10	-9	4.1	6.8	4.1	6.8	10.8	14.9	10.8	14.9	110	120
SiO2	-10	-9	12.2	14.9	12.2	14.9	27.1	40.7	27.1	40.7	110	120
Fe2O3	-10	-9	17.63	20.34	17.63	20.34	33.90	42.04	33.90	42.04	110	120
BAO	-10	-9	1.4	2.7	1.4	2.7	6.8	9.5	6.8	9.5	110	120
CAO	-10	-9	1.1	1.4	1.1	1.4	25.8	27.1	25.8	27.1	110	120
MGO	-10	-9	2.7	5.4	2.7	5.4	19.0	27.1	19.0	27.1	110	120
TiO2	-10	-9	1.8	2.7	1.8	2.7	5.4	8.1	5.4	8.1	110	120
AL2O3	-10	-9	0.4	1.1	0.4	1.1	9.5	12.2	9.5	12.2	110	120
RCP	-10	-9	1.8	2.0	1.8	2.0	5.4	6.8	5.4	6.8	110	120

Figura 27. Função de pertinência e termos linguísticos dos conjuntos fuzzy de entrada.

O núcleo do controlador fuzzy é constituído pela base de regras e por um método de inferência. Todas as proposições da base de regras do modelo de predição utilizado são da forma *IF-THEN*, exemplo:

“SE Fe₂O₃ é baixo, SE MgO é alto, SE Al₂O₃ é baixo, SE TiO₂ é baixo, ENTÃO CB”.

A base de regras foi montada para os oito domínios tipológicos: dolomita carbonatita (CB), foscoritos (FO), flogopititos (FL), picrito brechado (PI), apatita carbonatita (CBMS), foscorito com carbonatito (CBMG) aloterita (AL) e nelsonito (NL) (Figura 28).

```

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

# =====
# Arquitetura básica
# =====

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

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

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

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

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

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

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

# =====
# Base de regras
# =====

regras = [
    Regra([('FE2O3', Niveis.BAIXO), ('AL2O3', Niveis.BAIXO), ('MGO', Niveis.ALTO), ('TIO2',
        ('CB', Niveis.ALTO)
    ),
    (.....)
]

```

Figura 28. Arquitetura básica do algoritmo fuzzy.

Feito isso, o método de inferência fuzzy Mamdani uniu cada uma das proposições da base de regras, por meio de operações lógicas (*AND*), para simular o processo de dedução humana e produzir a variável fuzzy de saída. O passo seguinte foi a defuzzificação do conjunto fuzzy de saída [0,1], em um número real y , nesse caso o valor de saída y foi uma estimativa da classe pertencente. Os resultados das análises deste trabalho são baseados no método de defuzzificação do centroide (COG), considerando intervalos percentuais variando de 0% a 100%. O modelo fuzzy esquemático proposto para predição das tipologias é mostrado na Figura 29.

Ao final das simulações seguiu-se os três passos: primeiro, ordenou-se do maior para o menor as porcentagens das amostras com possibilidade de pertencer a mais de uma tipologia (medida de incerteza); segundo a saída fuzzy foi verificada com a descrição geológica dos furos de sondagem; terceiro, agrupou-se os domínios conforme os passos 1 e 2.

Assumiu-se neste trabalho que seriam utilizadas todas as informações confiáveis e disponíveis, no que diz respeito tanto à geologia quanto ao processo. As informações

primordiais para a modelagem espacial das tipologias de minério são as amostras de saída do agrupamento fuzzy validadas com as descrições de sondagem que constituem um total de 10.744 amostras. As demais amostras classificadas pelo algoritmo com informação divergente das descrições litológicas não foram consideradas na interpolação.

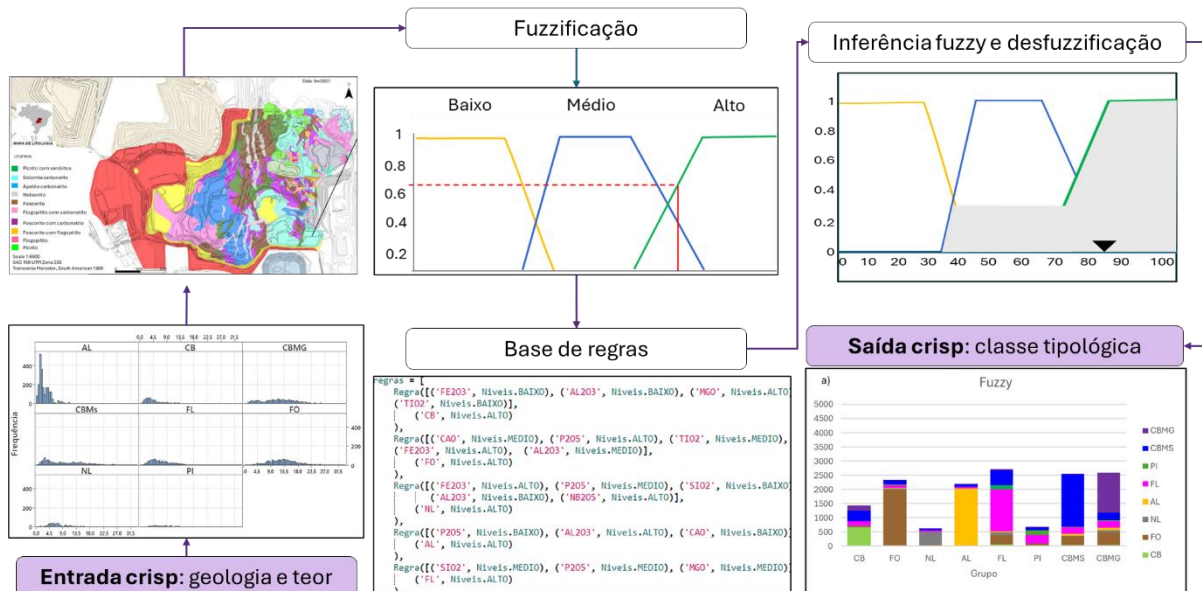


Figura 29. Componentes do modelo de predição fuzzy.

Etapa 3 – Modelagem espacial 3D

Os domínios tipológicos agrupados pelo algoritmo fuzzy foram codificados em variável indicadora (Figura 30). A variável indicadora aponta se a variável que sofreu a transformação binária pertence a uma dada classe de valores. Exemplo, para a modelagem espacial dos dolomita carbonatito (CB), todas as amostras classificadas no algoritmo fuzzy como CB recebeu código binário 1 e as classificadas nas demais litologias recebeu código zero. Com o banco de dados categorizado, os histogramas (Apêndice A) e os variogramas foram gerados, primeiro na direção vertical e depois no plano de varredura XY (Apêndice B a D). Os azimutes 0° e 90° representaram, nessa ordem, as direções de maior e menor continuidade espacial, e são direções aproximadas das estruturas geológicas locais e regionais com *trends* ora NNW ora NNE. Mapas de variograma foram gerados para os quatro domínios principais predominantes na mina FO, CBMS, FL e CBMG em que é possível observar direções de menor variabilidade NS para o FO, CBMS e CBMG. O mapa de variograma é uma ferramenta auxiliar, útil na definição das direções de anisotropia (Apêndice E).

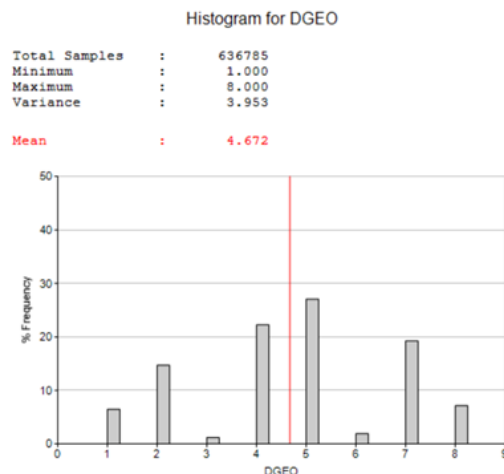


Figura 30. Histogramas dos indicadores das tipologias da mina.

Um mesmo modelo de variograma foi adotado para todos os indicadores, i.e., mesma estrutura e mesma estratégia de busca. Essa é uma solução adotada para atenuar problemas de relação de ordem durante a krigagem, que gera valores fora do intervalo $[0,1]$ e não garante que a somatória dos indicadores seja 1 (Goovaerts, 1997). Para modelar um variograma comum para todos os indicadores, foi utilizado como base o variograma do indicador 2 (foscorito), devido a sua importância tanto do ponto de vista petrológico quanto do ponto de vista econômico e pela semelhança das estruturas com os variogramas do foscorito sul (CBMG) e apatita carbonatita (CBMS). As Figuras 31 e 32 mostram o modelo teórico ajustado ao variograma experimental do FO.

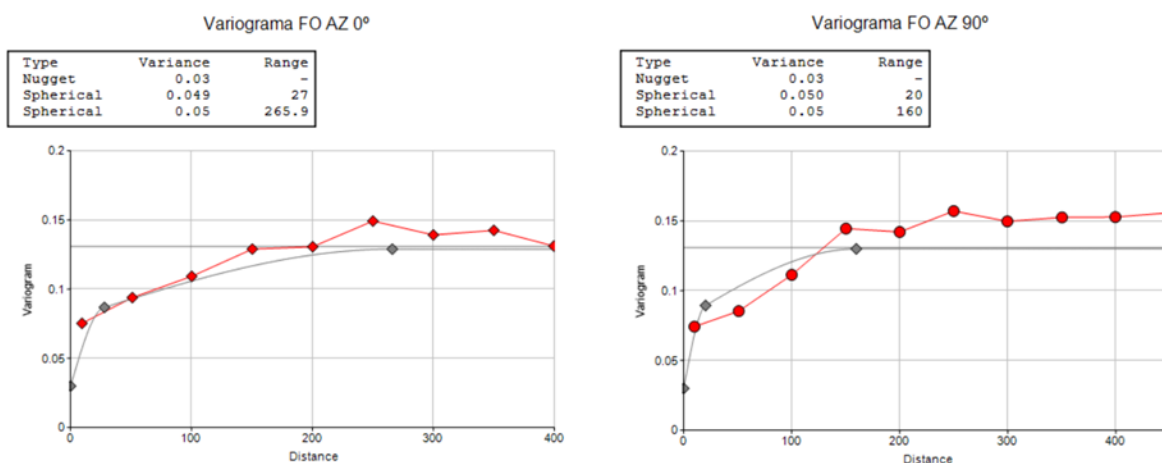


Figura 31. Variograma do FO nos eixos de maior continuidade e perpendicular.

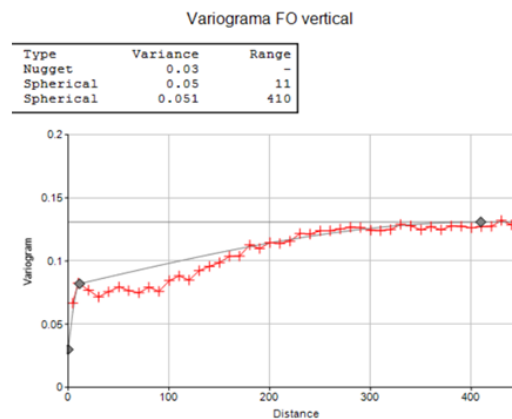


Figura 32. Variograma do FO nos eixos no eixo vertical.

O modelo de blocos utilizado para a estimativa possui as mesmas dimensões e origem praticados pelo modelo de recurso em uso na mina com blocos de 25x25x5 metros e foi estimado no Datamine®. O método de interpolação utilizado para a estimativa dos indicadores foi a krigagem ordinária dos indicadores. Alguns parâmetros de estimativa (número mínimo e máximo de amostras, número octantes etc.) foram testados a fim de obter o melhor modelo estimado. Uma estimativa por vizinho mais próximo limitada a um raio de busca equivalente a 60% do tamanho do bloco foi realizada nos modelos krigados para garantir que caso houvesse uma amostra descrita dentro do bloco essa litologia prevalecesse sobre a estimativa.

Os modelos geológicos interpolados seguiram o processo de validação praticado na literatura, tais como: verificação visual seção por seção entre modelo de bloco e furos; análise e correção de desvios de ordem; verificação da soma dos indicadores; verificação das proporções das tipologias.

Etapa 4 – Seleção de variáveis e modelos de predição

Com os domínios tipológicos modelados, uma análise do banco de dados de processo nesses domínios revelou que somente cinco dos oito grupos, tinham um número razoável de testes de bancada. Isso ocorre porque apesar da apatita estar presente nos outros grupos, eles não apresentam potencialidade (não são minérios) e isso explica a baixa quantidade de ensaios nessas litologias (Figura 33), pois, os testes são realizados para subsidiar a lavra de minérios determinada pelos teores de P_2O_5 e relação CaO/P_2O_5 (RCP). Portanto, para a predição das informações de processo nas etapas seguintes somente os domínios FO, FL, PI, CBMS e CBMG foram considerados.

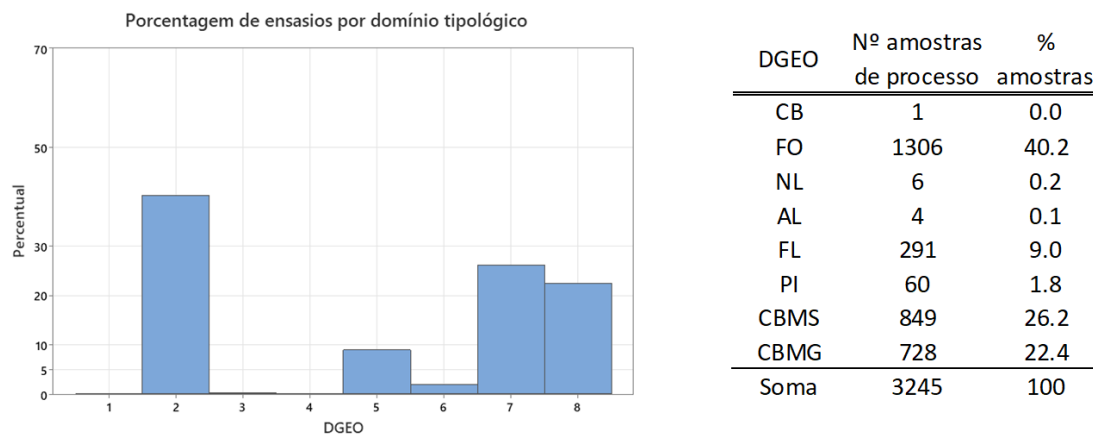


Figura 33. Quantitativo dos ensaios de flotabilidade por domínio.

Como já foi discutido as amostras de processo não são suficientes para a interpolação direta das variáveis metalúrgicas, e para contornar esse problema, uma opção foi testar uma função de transferência que adequasse as relações entre as variáveis primárias, aditivas e exaustivas com as variáveis de respostas. O banco de dados dos teores é aditivo e exaustivo, contendo mais de 40 mil amostras, contudo, não é isotópico para todos os elementos analisados; já os testes metalúrgicos correspondem 15% das informações dos teores, número considerável para a maioria dos depósitos minerais. Porém, ao considerar a isotopia de todos os atributos, esse valor cai para 4%. Por esta razão, a seleção das variáveis é fundamental tanto para preservar o maior número de amostras possíveis, quanto para otimizar as etapas futuras de simulação de múltiplas variáveis.

A análise de componentes principais (PCA) é uma técnica de análise multivariada muito utilizada na seleção de atributos e redução da dimensão dos dados. Os problemas de PCA envolvem a realização de uma transformação linear de dados para obter autovalores (λ) e autovetores (α), cujo autovalores λ são ordenados segundo a porcentagem de covariância. O algoritmo PCA foi aplicado no banco de dados e foram analisadas três abordagens para indicar as variáveis que seriam selecionadas em cada PC, nomeadamente métodos B2 e B4 propostos por Jolliffe (1972), e um índice de importância de variáveis (IIV) apresentado por Cervo (2015). Os dados foram normalizados e o critério utilizado para descarte das variáveis foi o recomendado por Jolliffe (1972), ou seja, o número de variáveis descartadas deve ser igual ao número de componentes em que a variância explicada (autovalor) é inferior a 0,7.

A PCA permitiu selecionar as variáveis de maior significado estatístico otimizando os modelos de predição. Os modelos preditivos são amplamente utilizados para conectar as

propriedades da rocha com as variáveis de processo. Regressão linear (RL), *Random forest* (RF) e support-vector machine (SVM) são ferramentas do aprendizado de máquina com resultados promissores (Deutsch *et al.*, 2013; Lishchuk *et al.*, 2019) e foram testados para prever as informações metalúrgicas da mina Chapadão para os domínios tipológicos mineralizados.

A tipologia PI (picrito com xenólitos), tem um número pequeno de amostras (Figura 32) e por essa razão foi agrupada com o flogopitito (FL) formando um único domínio geometalúrgico. Essas litologias apresentam características químicas muito semelhantes como foi apontado nas análises estatísticas para montagem do sistema de inferência fuzzy, embora o PI tenha teor de MgO inferior. São representadas por um mesmo magma parental silicático e baseado nesse fato optou-se por reorganizar os domínios FL e PI em um só grupo. Eles também apresentam características estatísticas semelhantes quanto às respostas de recuperação mássica e teor de P_2O_5 no concentrado (Figura 34).

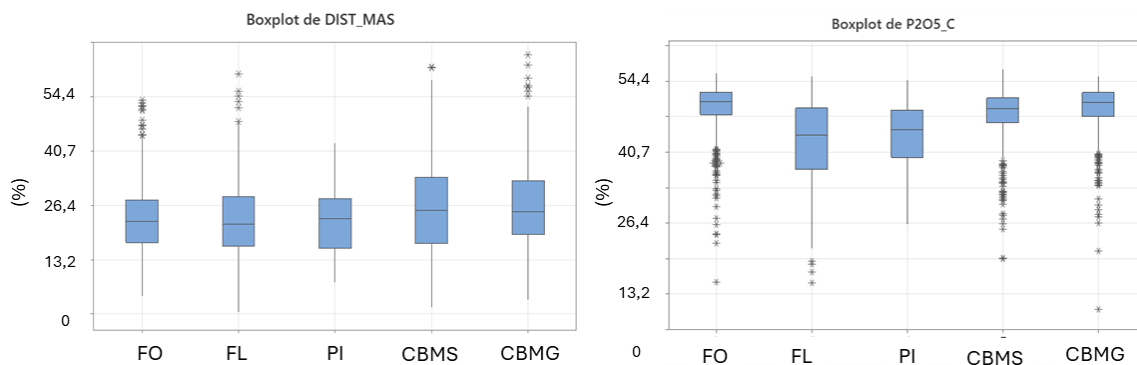


Figura 34. *Box-plots* das variáveis respostas por domínio tipológico.

Para reiterar a existência de diferenças significativas entre os domínios tipológicos que seguiriam para as simulações nos algoritmos de predição, adotou-se a técnica Análise de Variância (ANOVA), como mostra a Figura 35 (número 2 corresponde ao domínio FO, 7 CBMS, 8 CBMG e 56 FLPI).

Na ANOVA dizemos haver significância estatística quando o *p-value* é menor que o nível de significância estabelecido, nesse estudo o valor P ficou abaixo de 0,005. Quando a hipótese nula de médias iguais é verdadeira, o valor F dado pela relação: variação entre médias da amostra sobre a variação dentro das amostras deve ter valores semelhantes, e o teste estatístico deve ter resultado próximo de um. Nos casos em que é relativamente maior, como mostra a Figura 35, a hipótese de médias iguais é então descartada evidenciando que existem diferenças significativas entre pelo menos um par de médias dos domínios tipológicos

modelados, quando estes foram analisados do ponto de vista das variáveis DIST_MAS, MMAG e P₂O₅_C.

ANOVA com um fator: DIST_MAS versus DGEO						ANOVA com um fator: MMAG versus DGEO						ANOVA com um fator: P2O5_C versus DGEO					
Método						Método						Método					
Hipótese nula			Todas as médias são iguais			Hipótese nula			Todas as médias são iguais			Hipótese nula			Todas as médias são iguais		
Hipótese alternativa			Nem todas as médias são iguais			Hipótese alternativa			Nem todas as médias são iguais			Hipótese alternativa			Nem todas as médias são iguais		
Nível de significância			α = 0.05			Nível de significância			α = 0.05			Nível de significância			α = 0.05		
Linhas não usadas			4301			Linhas não usadas			4265			Linhas não usadas			4262		
<i>Assumiu-se igualdade de variâncias para a análise</i>						<i>Assumiu-se igualdade de variâncias para a análise</i>						<i>Assumiu-se igualdade de variâncias para a análise</i>					
Informações dos Fatores						Informações dos Fatores						Informações dos Fatores					
Fator		Níveis		Valores		Fator		Níveis		Valores		Fator		Níveis		Valores	
DGEO		4		2, 7, 8, 56		DGEO		4		2, 7, 8, 56		DGEO		4		2, 7, 8, 56	
Análise de Variância						Análise de Variância						Análise de Variância					
Fonte	GL	SQ (Aj.)	QM (Aj.)	Valor F	Valor-P	Fonte	GL	SQ (Aj.)	QM (Aj.)	Valor F	Valor-P	Fonte	GL	SQ (Aj.)	QM (Aj.)	Valor F	Valor-P
DGEO	3	4753	1584.25	30.79	0.000	DGEO	3	121798	40599.3	341.81	0.000	DGEO	3	6597	2198.84	167.52	0.000
Erro	3234	166411	51.46			Erro	3270	388398	118.8			Erro	3273	42962	13.13		
Total	3237	171164				Total	3273	510196				Total	3276	49558			

Figura 35. Análise de Variância para as cinco tipologias de minério e as variáveis respostas.

O banco de dados usado para a predição geometalúrgicas foi então dividido por tipologias e pelos horizontes de minério oxidado e micáceo (Figura 36), uma vez que os testes de processo estão condicionados a essas zonas e para evitar extrapolações para além das regiões amostradas. A base de dados resultou em 3.234 amostras. Os três algoritmos de predição foram avaliados tendo como base duas premissas: que fossem utilizadas as variáveis selecionadas na análise multivariada PCA; e que as variáveis utilizadas estivessem presentes em todas as amostras do banco de dados de processo, sendo elas P₂O₅, Fe₂O₃ e MgO.

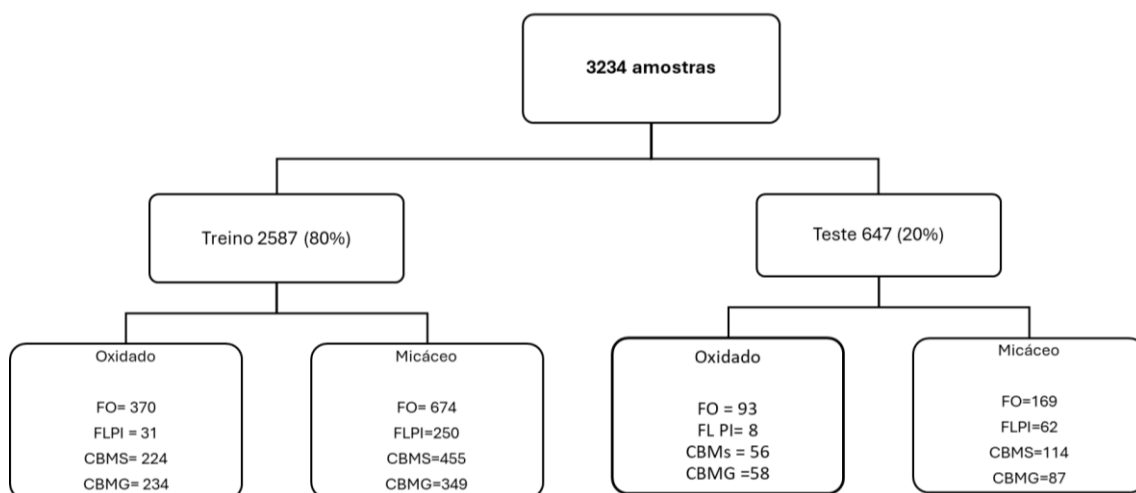


Figura 36. Divisão do banco de dados para os modelos de predição.

A análise de regressão foi a primeira ferramenta testada por ser uma técnica muito empregada e bem intuitiva. Os testes foram realizados no programa Minitab®. Para cada uma das variáveis dependentes, foi gerado um modelo de regressão com as variáveis independentes.

Os algoritmos RF e SVM foram extraídos da biblioteca *scikit-learning* e adaptados para o estudo de caso em linguagem Python®. No RF foram consideradas 1000 árvores para a elaboração do estudo e foi testado diferentes valores para o número mínimo de amostras por folha (Tabela 2). O SVM, em sua forma básica, é um classificador binário linear, que identifica um único limite entre duas classes e assume que os dados multidimensionais são linearmente separáveis no espaço de entrada. Existem casos em que não é possível separar bem as instâncias de forma linear e, neste caso, mapeia-se os dados para um espaço dimensional maior (na qual os dados são linearmente separáveis por estarem mais dispersos) e cria-se um classificador linear nesse espaço. Essas duas formas do SVM foram testadas para os dados em análise. O kernel RBF foi selecionado para o SVM não linear devido à sua popularidade em problemas de aprendizado e ao seu baixo número de parâmetros (apenas dois) que afetam o resultado do aprendizado, já que outras funções do kernel (polinomial) requerem mais parâmetros além do γ e do C , incluindo o grau do polinômio, o que pode aumentar significativamente o desempenho e a complexidade do classificador.

Tabela 2. Parâmetros testados nos algoritmos.

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

Os resultados para as três técnicas testadas foram verificados por validação cruzada através do algoritmo *K-fold*. A validação cruzada é um método simples e muito utilizado para verificar os modelos de predição. A ideia da validação cruzada é dividir o banco de dados em k partes iguais e utilizar cada uma das partes como conjunto de teste em uma das rodadas e como conjunto de treinamento nas demais. Para cada subdivisão, dispõe-se de dados diferentes para treino e teste. Ou seja, serão obtidos resultados diferentes de acordo com cada iteração k do modelo. Ao final das k iterações, calcula-se a acurácia sobre os erros encontrados e obtém-se uma medida mais confiável sobre a capacidade do modelo de representar o processo gerador dos dados.

Etapa 5 – Simulação sequencial gaussiana

Na última etapa deste trabalho, a simulação sequencial gaussiana das variáveis primárias preditoras das respostas metalúrgicas foi realizada, testando trinta e cinco cenários em cada um dos domínios tipológicos em um grid de pontos de 5x5x5 metros no Ar2GeMS. Destaca-se que a mina tem quase 2 km de extensão e o modelo de blocos possui 2.704.708 blocos. Para tornar esse modelo viável para as rotinas de curto prazo foi verificado um número satisfatório de cenários sem que o modelo de blocos aumentasse excessivamente em tamanho.

Costa (1997) propôs uma metodologia de análise do número de cenários a partir dos desvios padrão das simulações. Os gráficos da Figura 37 mostram que, a partir de 25 cenários, já existe uma estabilização da variância para os teores de P_2O_5 , Fe_2O_3 e MgO . Portanto, trinta e cinco cenários foram simulados para informar o acesso ao espaço de incerteza do fenômeno.

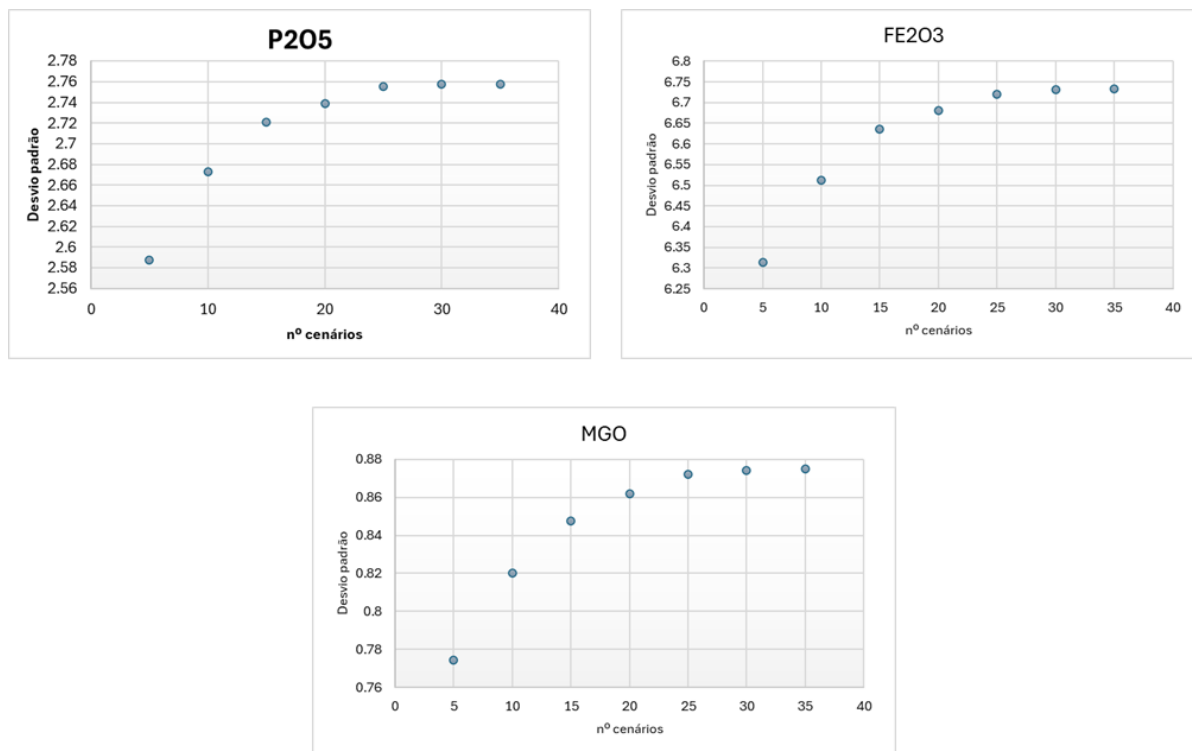


Figura 37. Número de realizações versus o desvio padrão médio das simulações.

As variáveis preditoras foram interpolados por simulação sequencial gaussiana, pois a simulação fornece uma medida de variabilidade importante para previsão da planta de concentração. Diferente do banco de dados utilizado na etapa 4, com isotopia para teores, tipologias e ensaios de flotabilidade, os dados para SGS são homogêneos para teores e

tipologias. As amostras foram separadas conforme as superfícies dos domínios krigados e as superfícies de intemperismo totalizando 7.539 amostras.

A malha de amostragem da sondagem é 75x75 m nas regiões da borda e 50x50 m no centro da mina, com amostragem preferencial nos locais de maior teor de P_2O_5 (Figura 38).

Os dados foram transformados para o espaço gaussiano, os histogramas foram gerados e a variabilidade espacial das amostras definidas a partir dos variogramas. Os variogramas foram construídos primeiro no eixo vertical para analisar o efeito pepita responsável por medir a descontinuidade a curta distância, e depois em diferentes direções na horizontal. Os variogramas modelados têm direções de maior e menor continuidade no azimute 0° e 90° seguindo as mesmas estruturas geológicas do modelo de indicadores das tipologias.

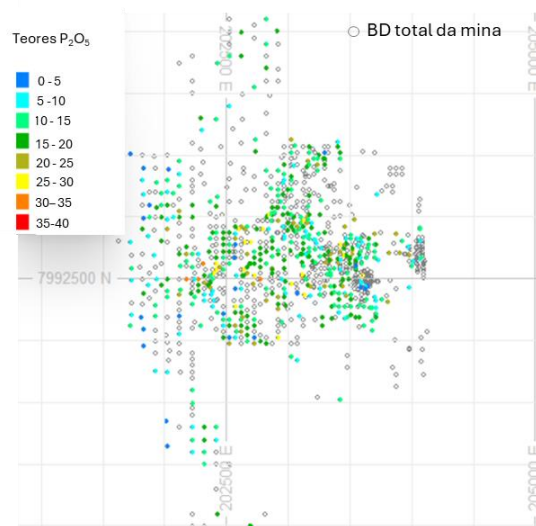


Figura 38. Distribuição dos furos de sondagem na cota 800.

As simulações dos teores foram verificadas pela reprodução dos histogramas e variogramas com os dados originais. Os modelos de predição foram aplicados nos pontos simulados, os valores médios das simulações (*e-type*) e o coeficiente de variação calculados. Os pontos foram posteriormente reescalados para o grid do modelo de blocos 25x25x10 m resultando no modelo geometalúrgico.

Para atestar a validade do modelo, avaliou-se a reconciliação com dados de produção. A reconciliação foi realizada a partir das programações das pilhas nos meses de janeiro a dezembro de 2023, exceto a pilha do mês de junho que contou com material de estoque. O modelo geometalúrgico foi separado segundo os sólidos de volume das pilhas e as massas conferidas.

6. RESULTADOS

6.1 Apresentação dos dados e estatísticas descritivas

As análises estatísticas (Figura 39) mostram para o perfil intempérico que os teores de P_2O_5 , Fe_2O_3 e MgO são excelentes marcadores dos horizontes de minérios, como descrito na literatura. O alto teor de alumínio caracterizado pela presença de argilo minerais e fosfatos aluminosos está presente no nível aloterita (A6). Nos níveis A3, A4 e A5 (micaceo de base, topo e oxidado) é possível notar o aumento dos teores de P_2O_5 e SiO dado a maior concentração de apatita residual e silicificação; e uma diminuição do MgO quando comparado aos níveis A1 e A2 cujas médias de MgO se mantem acima de 10% devido a maior presença de flogopita e filossilicatos. As matrizes de correlação (Figuras 40 e 41) indicam correlações significativas entre as variáveis P_2O_5 e CaO no horizonte A3, A4 e A5 (horizontes de minério, cujo teores de cálcio estão relacionados a presença da apatita) e entre o Fe_2O_3 e Al_2O_3 no nível A6 (horizonte de estéril).

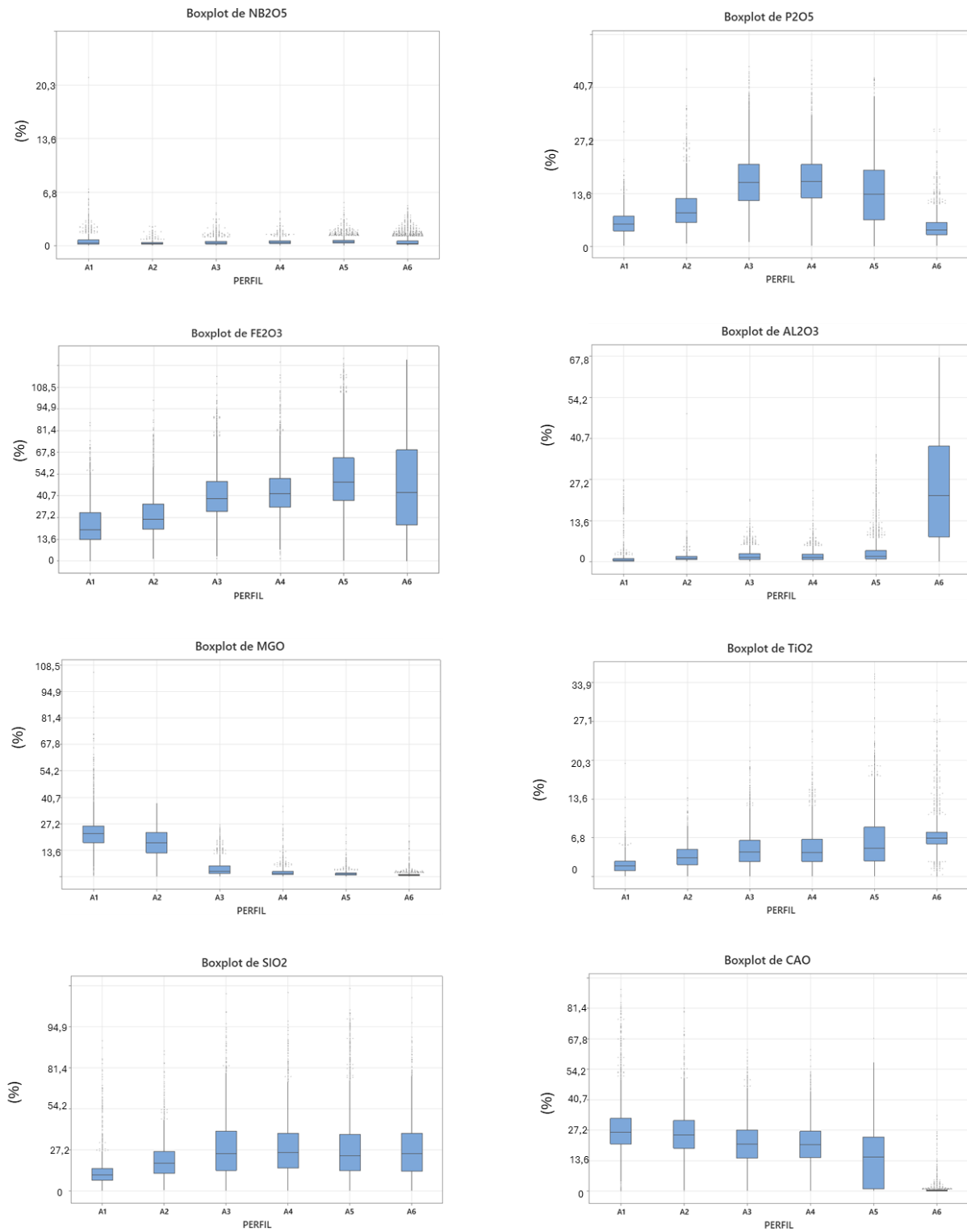


Figura 39. *Box plot* dos teores para o perfil intempérico.

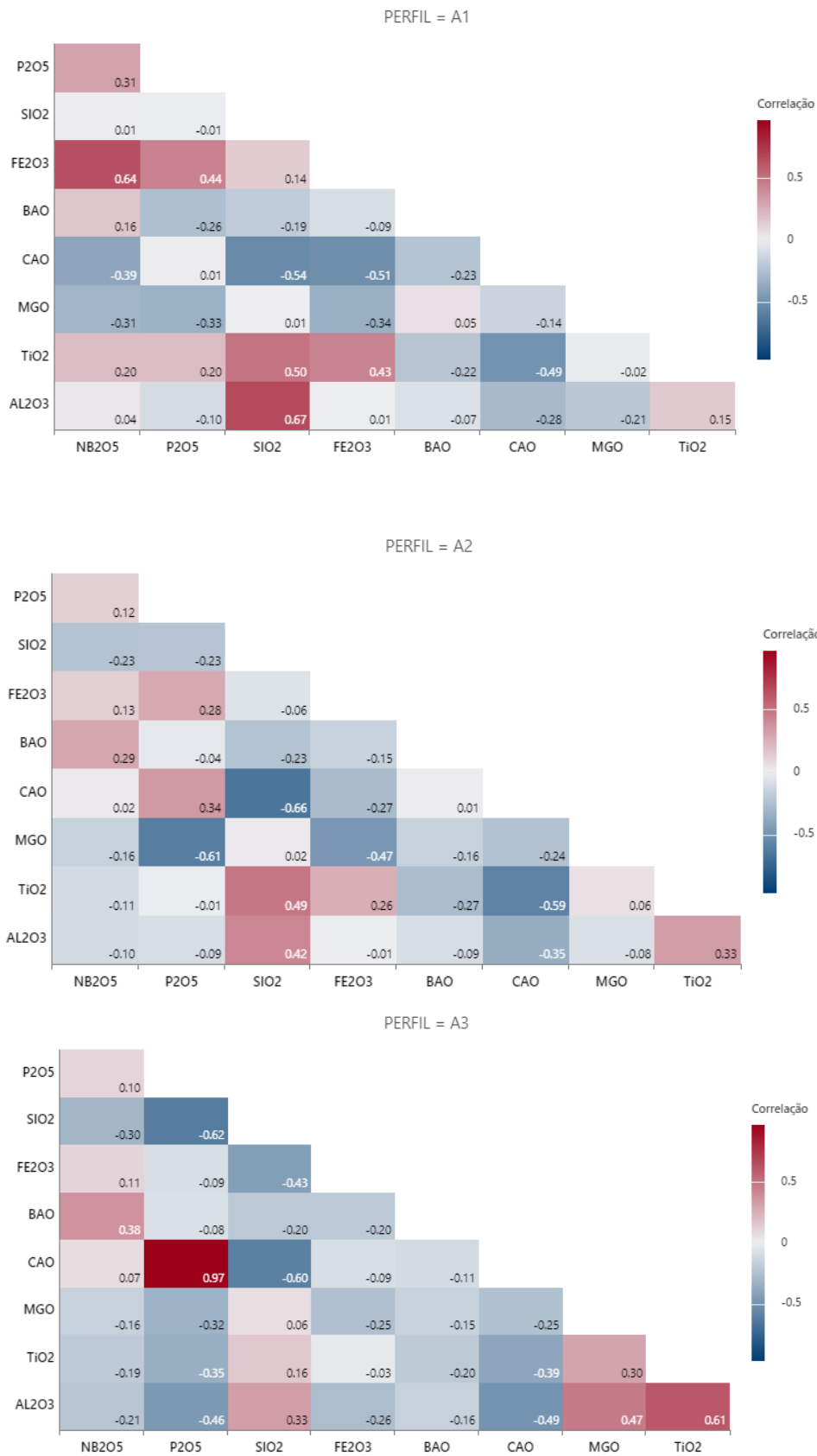


Figura 40. Matriz de correlação dos teores por horizonte intempérico.

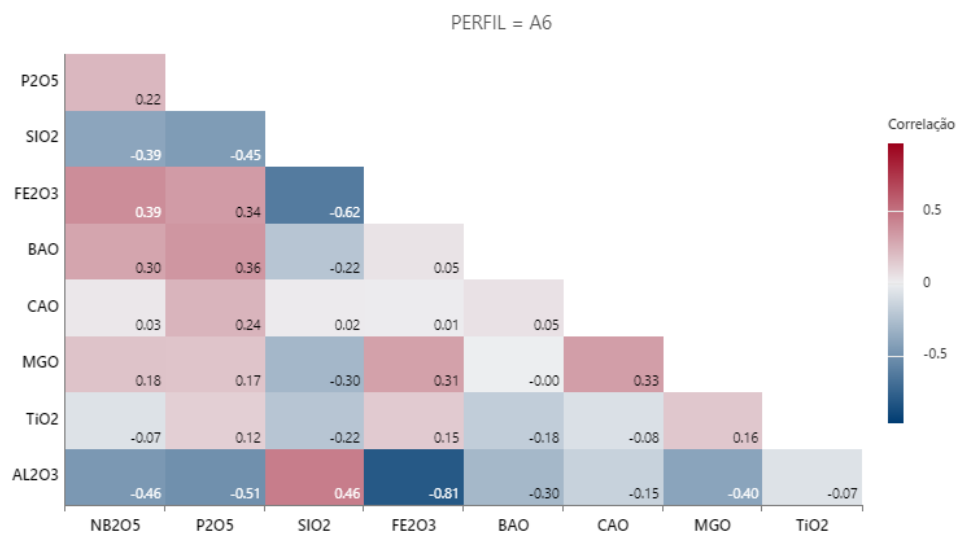
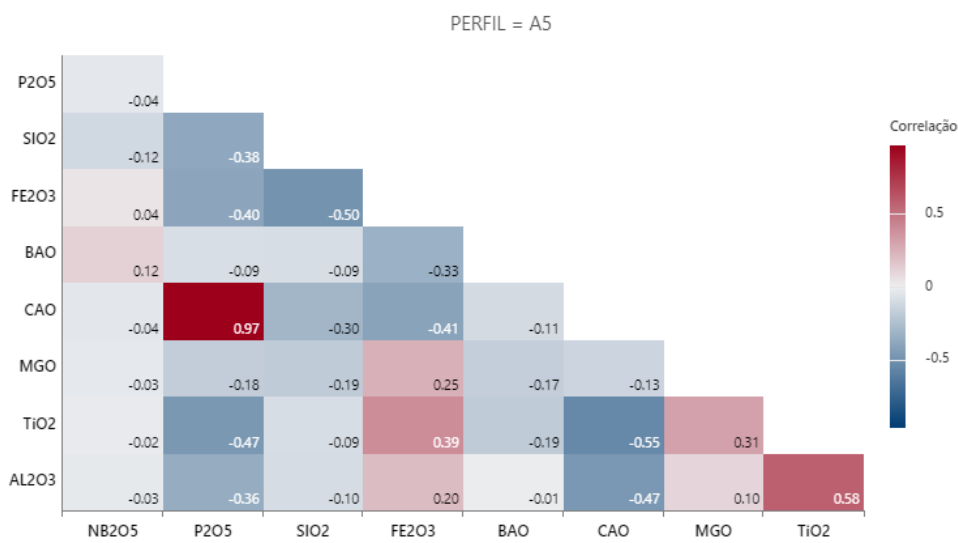
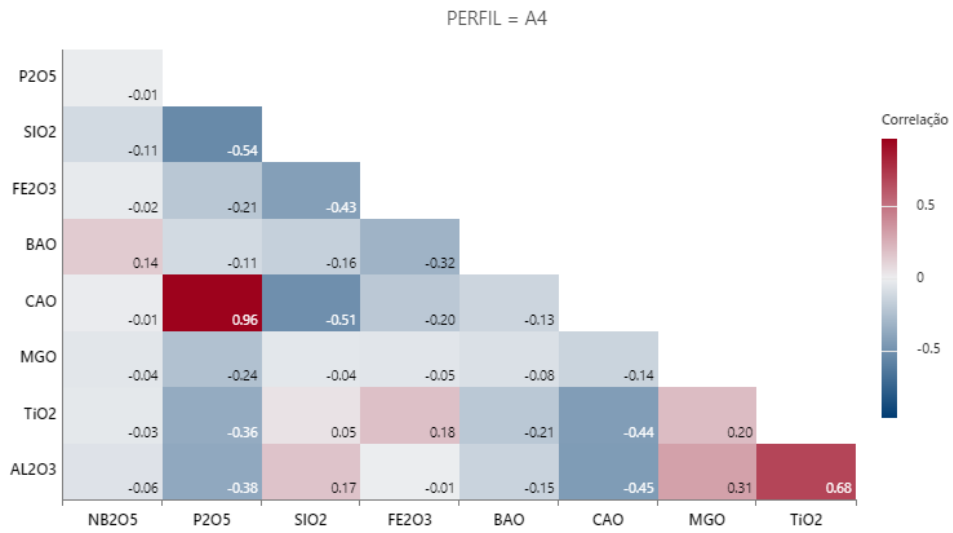


Figura 41. Matriz de correlação dos teores por horizonte intempérico.

Em relação as análises realizadas para os protolitos (Figura 42 a 44) percebe-se que apatita carbonatitos (APCBCA e APCBMG) e as combinações litológicas (litotipos) com os foscoritos possuem teores médios similares, principalmente para os óxidos P_2O_5 , Fe_2O_3 e MgO . Já os teores de MgO acima de 5% estão presentes nos flogopititos, bebedouritos e suas combinações. A litologia dolomita carbonatito apresenta um aumento significativo de MgO próximo a 20% sendo um excelente marcador para essa rocha. Os nelsonitos apresentam teores elevados de Nb_2O_5 e baixos teores de SiO_2 . A sílica baixa também é observada nas amostras que contêm carbonatitos. É possível observar que alguns litotipos apresentam semelhanças, podendo constituir grupos. Outros, porém, como picrito, brechas e as variações dos carbonatitos, possuem características químicas que não são tão facilmente diferenciadas.

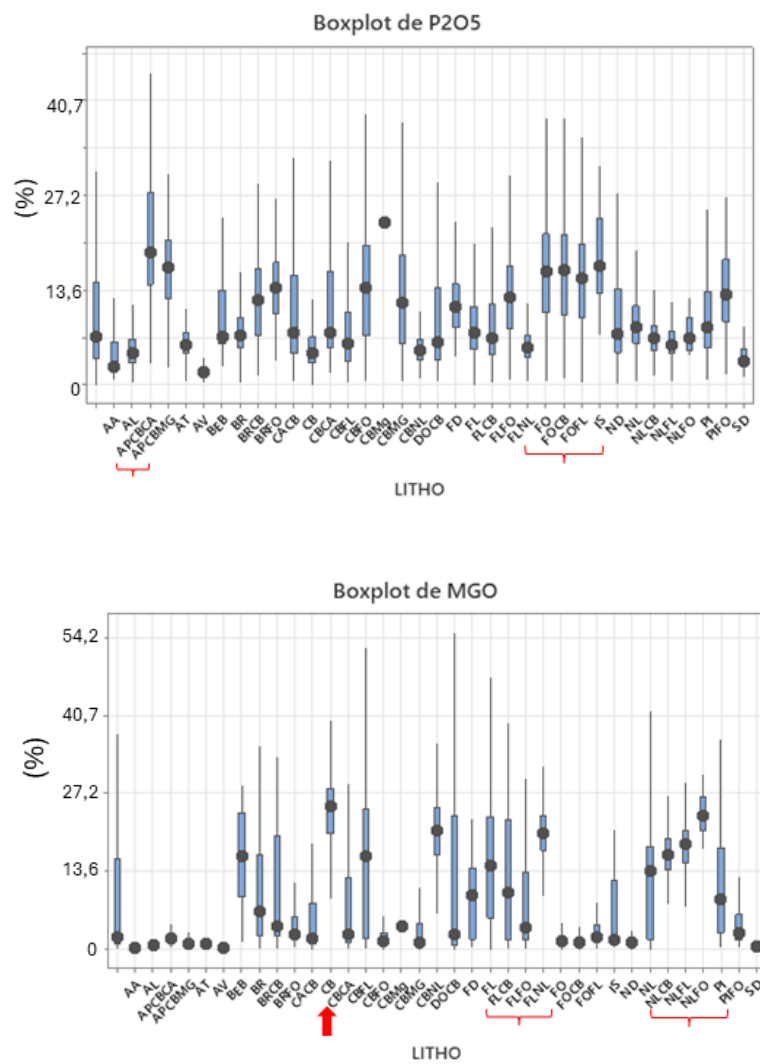


Figura 42. Gráficos *box plot* dos teores dos litotipos.

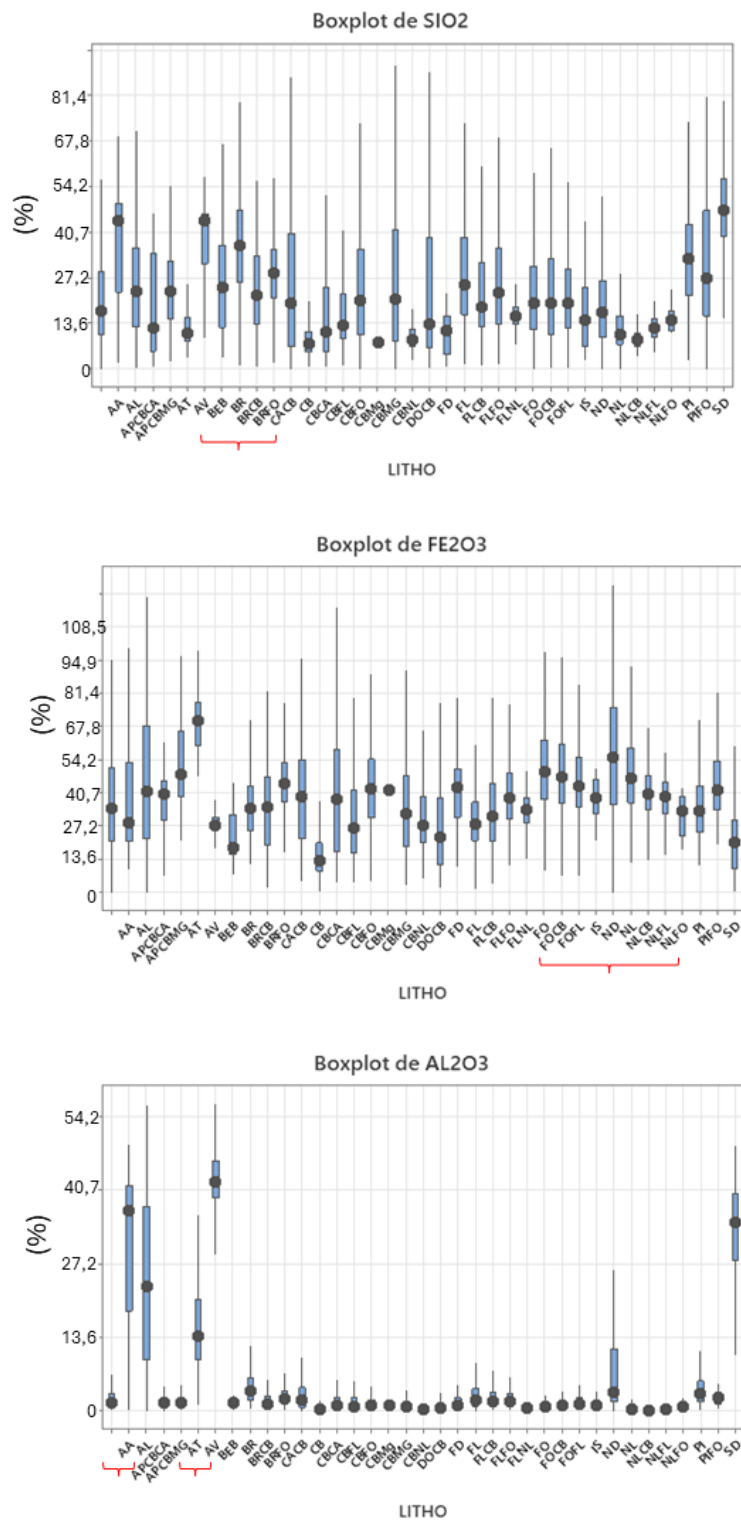


Figura 43. Gráficos *box plot* dos teores dos litotipos.

Tabela 3. Simulações fuzzy e aderência dos domínios com as amostras.

Tipologias- amostras precisas	Classe fuzzy aderente com a descrição geológica (%)				
	1	2	3	4	5*
CB	91.6	97.6	98.5	98.2	77.1
FO	71.8	64.0	61.7	64.5	70.01
NL	71.7	84.7	83.6	86.1	73.8
AL	87.6	91.5	91.5	91.7	91.6
FL	63.8	63.0	65.4	65.9	66.4
PI	28.9	53.3	53.4	46.0	61.0
CBMS	61.4	97.6	63.9	50.0	67.9
CBMG	64.0	65.2	76.9	51.6	84.2
Quantidade de amostras classificadas	15.733	9.532	10.634	14.618	15.101

A Figura 45 traz uma comparação do método fuzzy, com o k-means e hierárquico. É possível observar que o algoritmo fuzzy mostrou-se mais eficiente para classificar e, conseqüentemente, separar os domínios com menos misturas de tipologias quando comparado aos demais. Nos algoritmos tradicionais a tipologia predominante em cada *cluster* não foi facilmente identificada. Em relação ao comportamento e o padrão das distribuições estatísticas das variáveis é possível observar que elas foram reproduzidas pelos domínios fuzzy quando comparadas às amostras consideradas precisas (Figura 47).

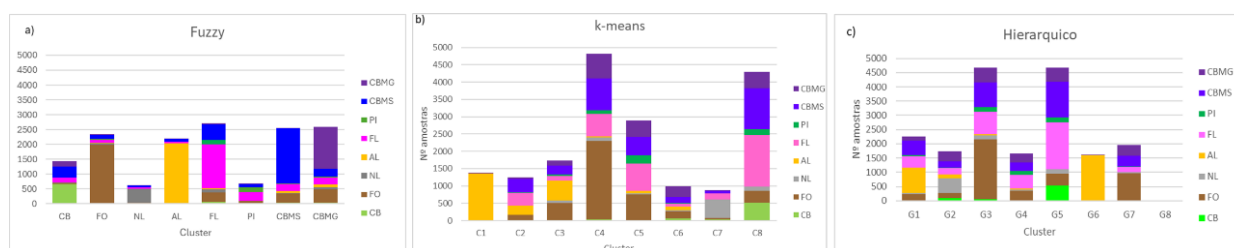


Figura 45. Resultados dos agrupamentos. Em a) fuzzy; b) k-means e c) hierárquico.

A inspeção visual seção a seção variando as cotas entre os dados de sondagem (amostras precisas) e os agrupamentos fuzzy e k-means mostrou uma similaridade maior entre as distribuições dos domínios obtidos pela lógica fuzzy com o banco de dados geológico.

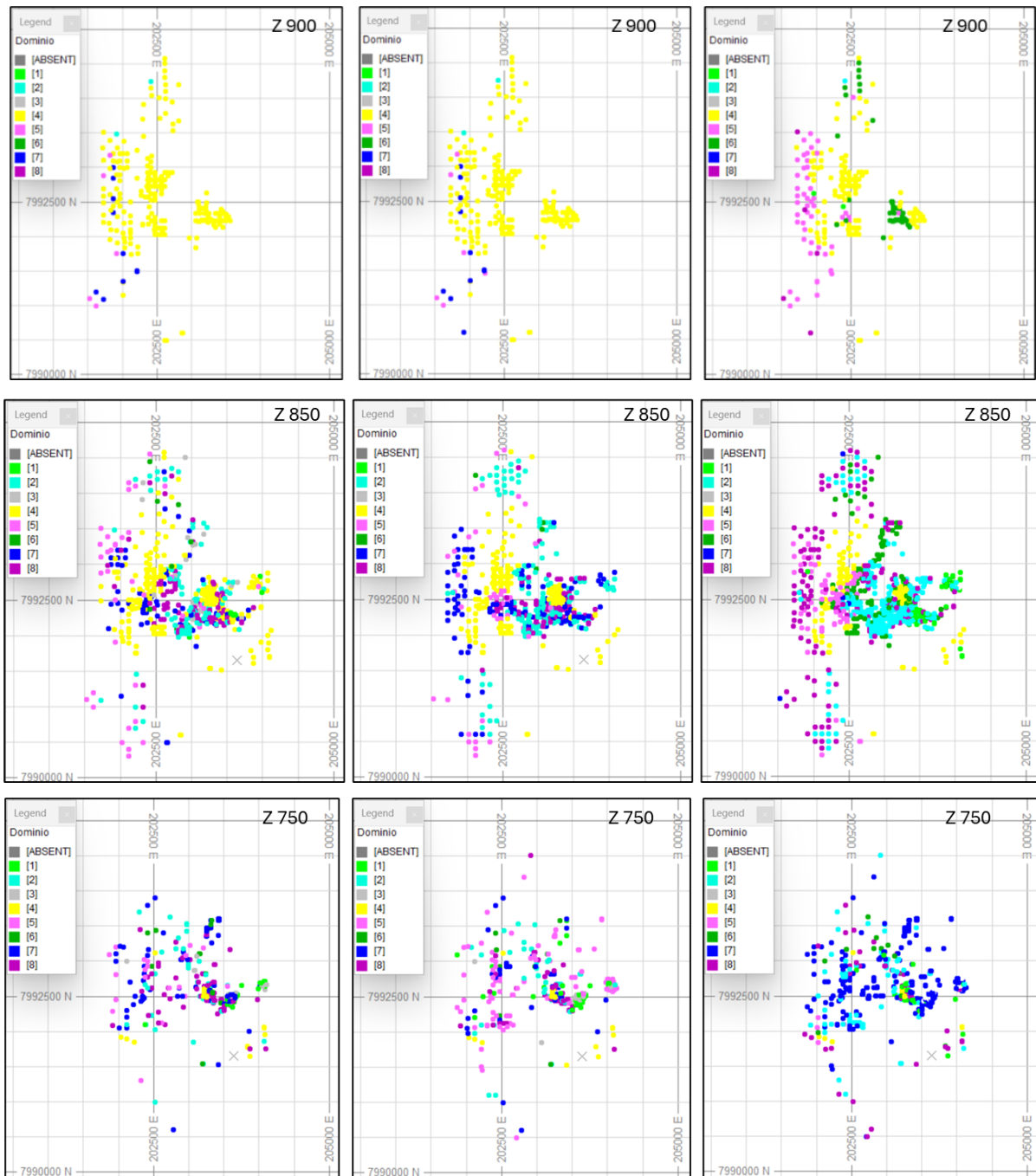


Figura 46. Verificação visual em seções horizontais: amostras sondagem, agrupamento fuzzy e k-means, nessa ordem.

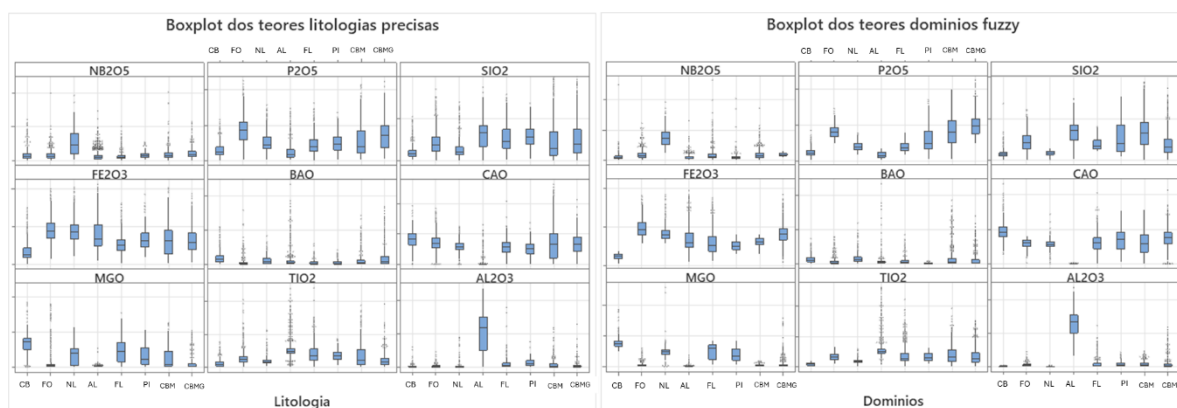


Figura 47. Análise estatística dos teores nas amostras precisas e nos domínios fuzzy.

O índice de Rand (CR) foi calculado para os resultados das três técnicas de agrupamento. O CR mede a qualidade da partição gerada pelo algoritmo através do grau concordância com a informação prévia ou a classe rotulada, isto é, com o subconjunto do banco de dados (47% das amostras com litologias precisas). O índice Rand ajustado varia entre 0 e 1, cujo valor próximo a 0,0 indica uma rotulagem aleatória independentemente do número de agrupamentos e amostras; e exatamente 1,0 quando os agrupamentos são idênticos. A Tabela 4 mostra que os resultados do agrupamento fuzzy apresentou um CR superior aos demais.

Tabela 4. Verificação dos agrupamentos pelo CR.

Rand Index Ajustado			
Hierárquico	k-means	Fuzzy (todas as classificações)	Fuzzy (classes validada com descrição)
0.25	0.25	0.49	0.86

Outra maneira de avaliar a qualidade do agrupamento bem como a coesão geográfica dos grupos é medindo a continuidade espacial de seus indicadores. Se os variogramas dos grupos gerados se apresentarem contínuos e com estruturação aceitável a pequenas distâncias isso indica que há uma certa homogeneidade nas unidades predominantes. Nos Apêndices B a D é possível ver que os variogramas dos indicadores atenderam tais requisitos.

Como informação adicional à classificação dos domínios, o algoritmo difuso atribuiu um grau de associação das tipologias a todas as amostras. Esse grau de pertinência pode ser usado como uma medida de probabilidade e, portanto, pode fornecer uma maneira simples de quantificar a incerteza do agrupamento. Em uma representação nítida do conceito “foscrito”, atribuir-se-ia 1 às combinações de variáveis que se encaixam perfeitamente na definição de FO

e zero para as demais. Na representação dos conjuntos nebulosos, a consideração da natureza transitória dos limites é expressa nas funções de pertinências, podendo uma determinada amostra ser simultaneamente membro de mais de um conjunto difuso, por exemplo: foscorito (FO) e foscorito com carbonatito (CBMG). A Tabela 5 apresenta o número de amostras que receberam grau de pertinência para uma ou mais classes. É possível notar que 54% das amostras receberam uma única classe (valor próximo ao observado no banco de dados de sondagem, onde aproximadamente 47% das amostras têm uma única descrição) e que 87% têm chances de pertencer no máximo a duas classes. Essa informação auxilia os debates sobre onde exatamente estão as regiões de maior variabilidade redirecionando os trabalhos de geologia. Observe na Figura 48 que as maiores incertezas litológicas dos blocos estão na região central; e no mapa geológico (Figura 20) essa região é marcada por uma quantidade maior de veios.

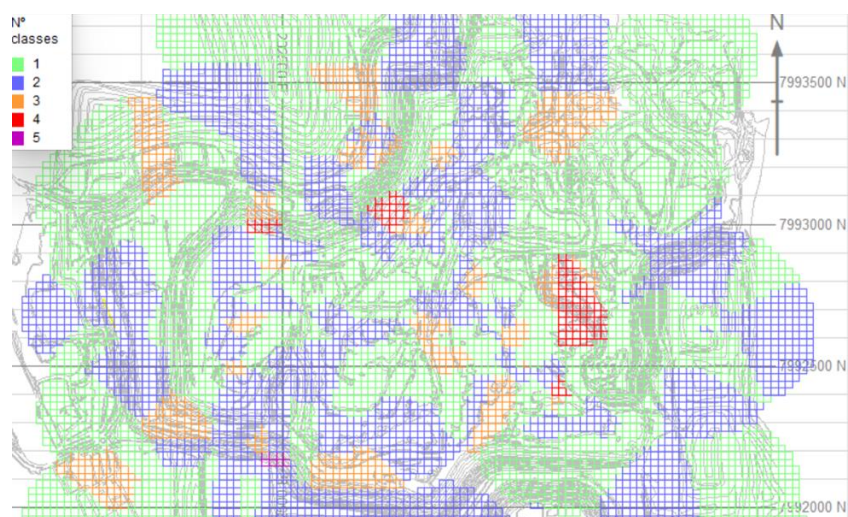


Figura 48. Blocos com pertinência fuzzy em uma ou mais classes, cota 750.

Tabela 5. Incerteza geológica dada pelo grau de pertinência das amostras.

Pertinência fuzzy em uma ou mais classes				Acertos fuzzy	
Quantidade de classes	Total amostras classificadas	%	% acumulada	Nº amostras corretas	%
1	8.150	54.0	54.0	5.386	66.1
2	4.929	32.6	86.6	3.596	73.0
3	1.663	11.0	97.6	1.436	86.3
4	313	2.1	99.7	286	91.4
5	46	0.3	100.0	40	87.0
6	0	0.0	-	-	-
7	0	0.0	-	-	-
8	0	0.0	-	-	-
Soma	15.101	100	-	10.744	71.1

6.3 Modelagem dos indicadores

Os resultados da krigagem dos indicadores mostraram-se consistentes com os dados originais desagrupados (Figuras 49 a 50); as médias da estimativa mantiveram-se próximas às médias dos dados originais e a variância diminuiu, como esperado, devido à suavização causada pela estimativa.

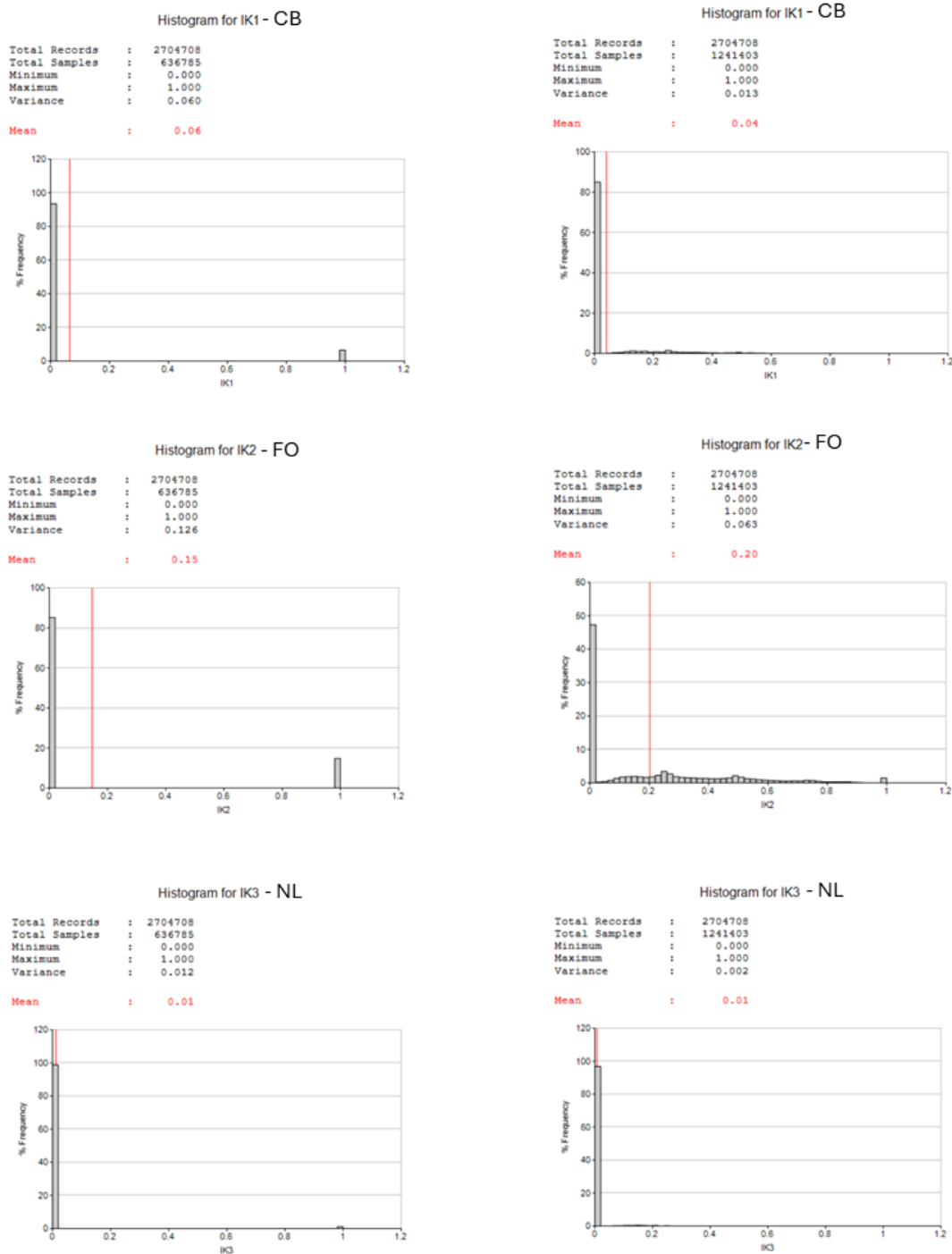


Figura 49. Histograma dos dados desagrupados à esquerda; histograma dos indicadores krigados à direita.

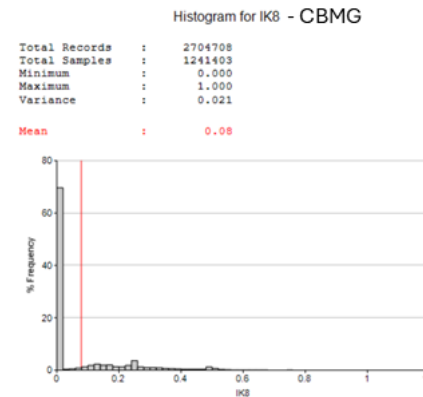
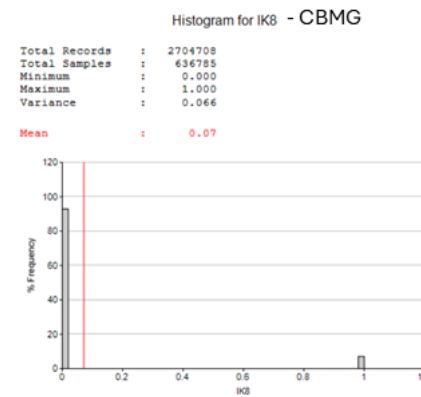
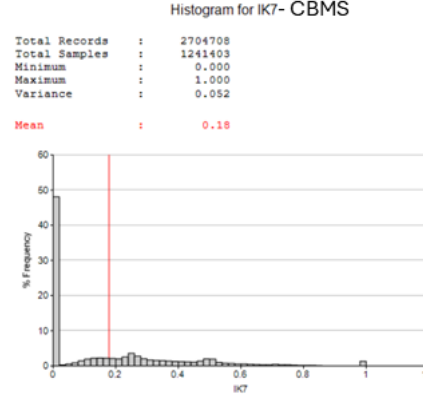
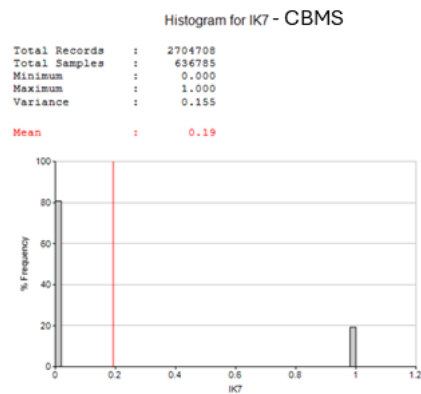
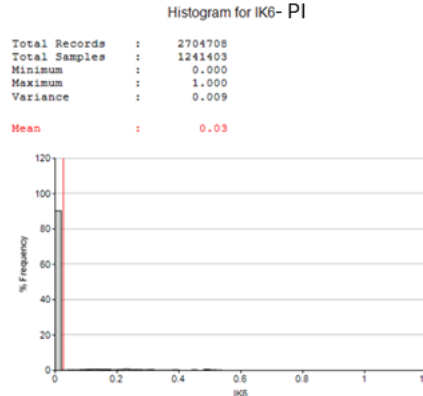
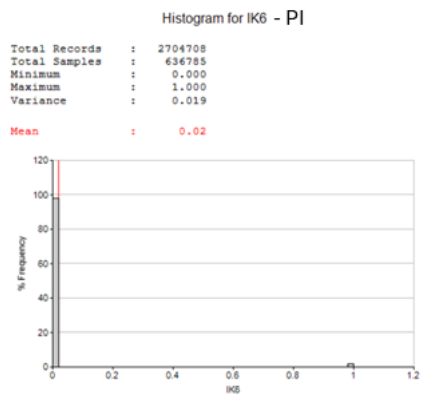
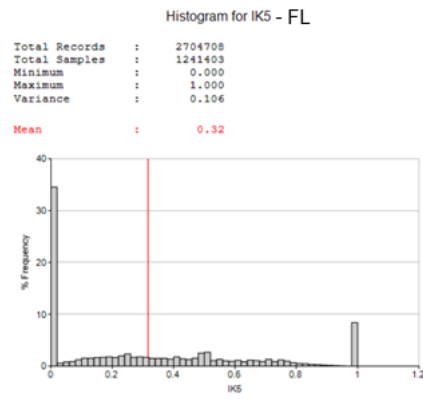
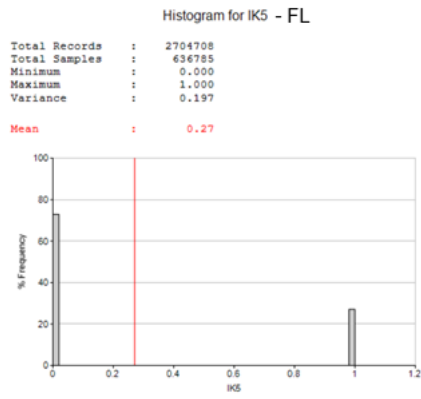


Figura 50. Histograma dos dados desagrupados à esquerda; histograma dos indicadores krigados à direita.

Após a estimativa, classificou-se os blocos com a tipologia de maior probabilidade de ocorrência. A Figura 51 mostra que mais de 90% das categorias foram definidas com probabilidades acima de 39.2% e que 75% das categorias foram definidas com probabilidades próximas a 50%. Esse resultado revela que a possibilidade de um bloco ter mais chance de não pertencer àquela categoria do que pertencer, no caso de uma divisão equilibrada da probabilidade entre as categorias, não foi impactante.

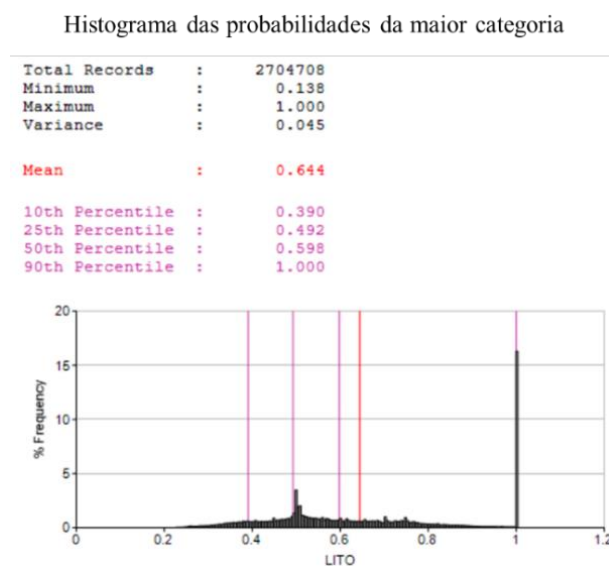


Figura 51. Histograma das maiores probabilidades escolhidas para categorizar os blocos.

Dependendo do volume de busca e alcance do variograma, os dados que se encontram nas bordas podem sofrer com a extrapolação, principalmente nos casos em que o modelo de blocos ultrapassa a região de amostragem como acontece na Mina Chapadão. Na Figura 52 é possível notar maiores diferenças entre as amostras desagrupadas e o modelo krigado para a litologia flogopitito, mas, ainda assim os resultados se mantiveram com desvios menos que 8%. Essa litologia tem ocorrência na borda do depósito, como visto no item 4.2.1. Vale ressaltar que o variograma do foscrito, utilizado como estrutura comum para a estimativa, tem um alcance bem menor quando comparado ao flogopitito, e não implicou no efeito de extrapolação do FL.

Em continuidade à análise dos dados interpolados, foi realizada a validação visual entre os blocos krigados e os indicadores do banco de dados em seções horizontais e verticais (Figuras 53 a 56). Essa análise qualitativa é frequentemente utilizada para investigar a existência de possível viés nas estimativas e é altamente recomendada. Não foi observada nenhuma discrepância visual entre os dados amostrais e os blocos estimados. A distribuição das amostras foi reproduzida pelos valores estimados, no entanto, nota-se uma quantidade de blocos

abaixo das amostras de sondagem, efeito da boa continuidade espacial medida nos variogramas verticais.

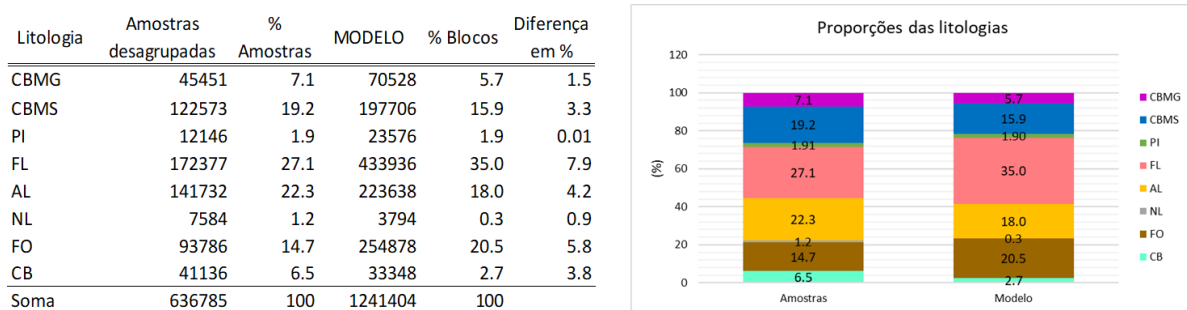


Figura 52. Ocorrências proporcionais das tipologias no banco de dados e no modelo de blocos.

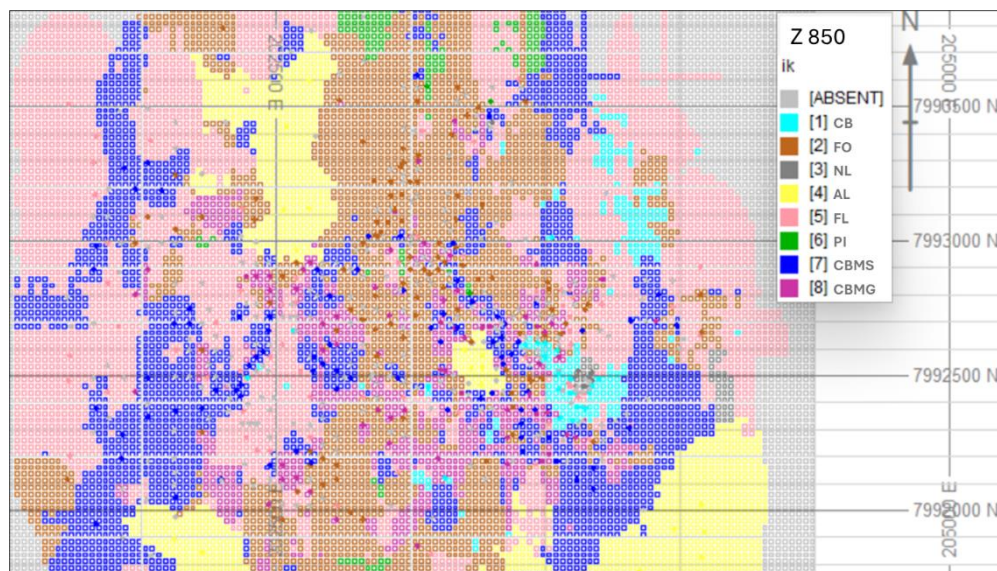


Figura 53. Verificação visual do modelo de blocos e as amostras na seção horizontal Z850.

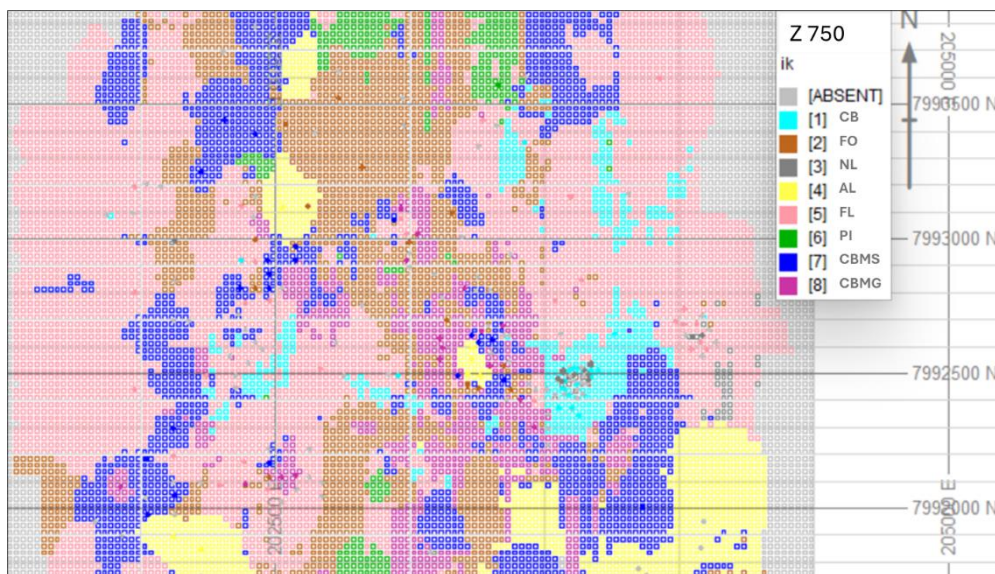


Figura 54. Verificação visual do modelo de blocos e as amostras na seção horizontal Z750.

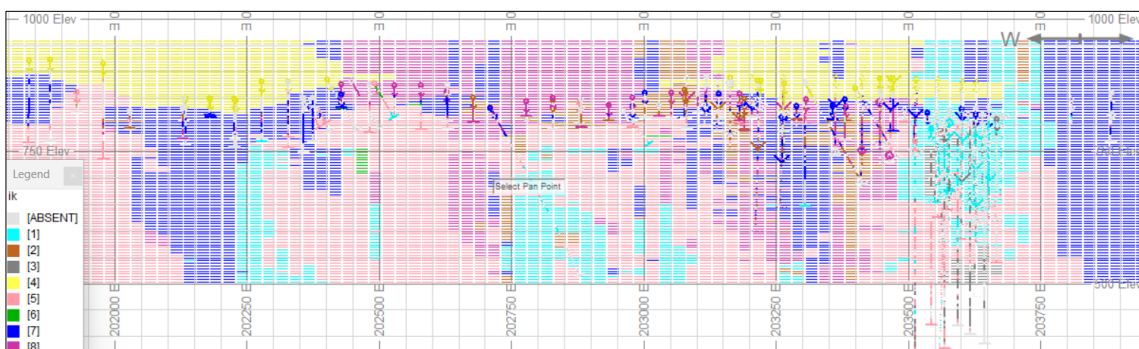


Figura 55. Verificação visual do modelo de blocos e as amostras na seção vertical 7.992.500E.

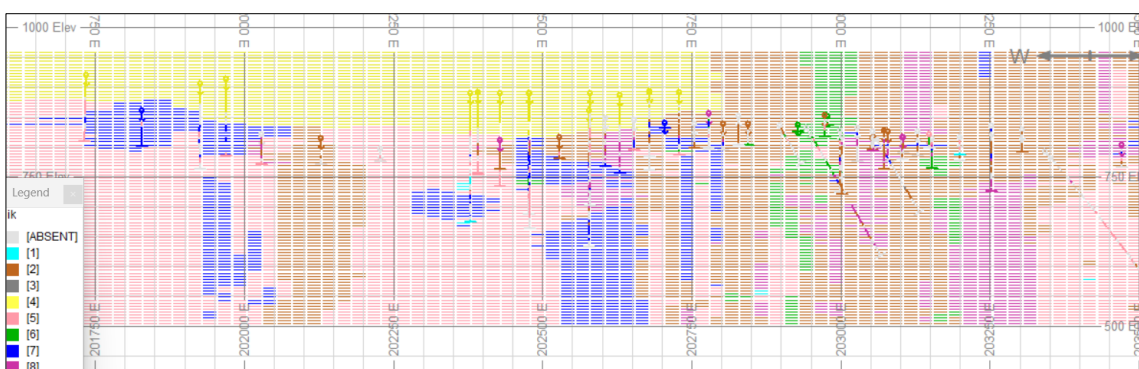


Figura 56. Verificação visual do modelo de blocos e as amostras na seção vertical 7.993.000

6.4 Seleção de variáveis

A seleção de variáveis se mostrou fundamental nesse trabalho, otimizando as etapas de simulação. A análise de componentes principais apontou que um número reduzido de sete componentes principais é responsável por explicar mais de 80% da variabilidade do fenômeno como mostra a Figura 57.

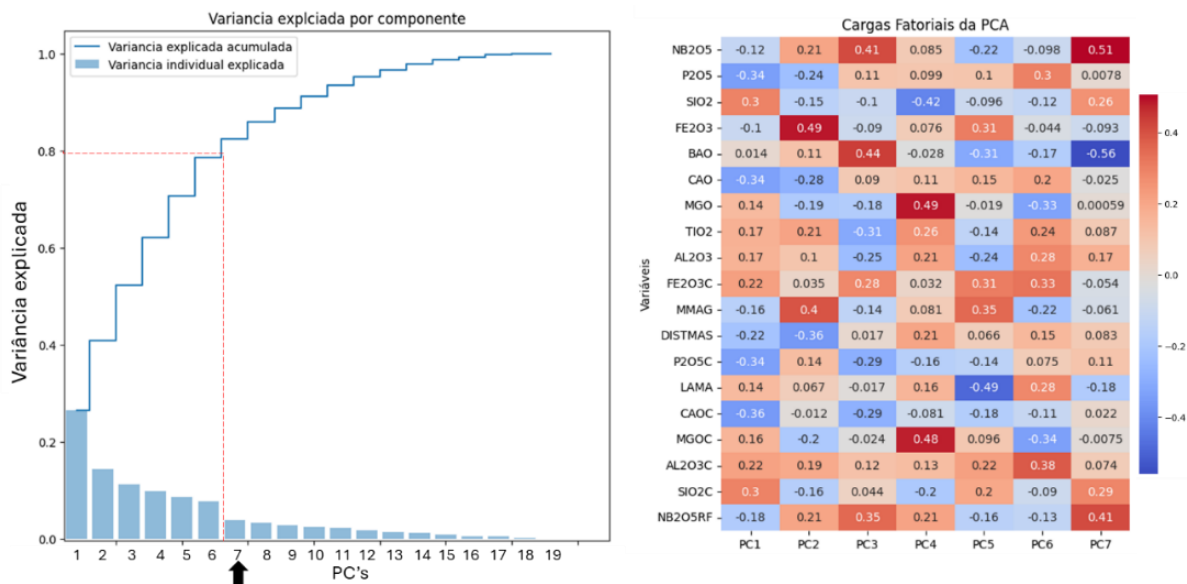


Figura 57. À esquerda a variância explicada e, à direita as cargas fatoriais das sete primeiras PC's.

A Figura 58 mostra que a variável Al_2O_3C foi a variável selecionada em todos os métodos testados e que Nb_2O_5 , Fe_2O_3 , SiO_2 e BaO aparecem em dois das três análises. Vale ressaltar que em todas as metodologias, quatro das sete variáveis selecionadas são químicas e, as três variáveis de processo (em negrito na Figura 58a) apresentam valores elevados de similaridade com as outras variáveis químicas (Figura 58b), sendo consideradas redundantes. O Fe_2O_3 apresenta alta correlação com $MMAG$ consequentemente, os valores altos de um induzem os valores altos do outro e por isso não precisam coexistir simultaneamente. O mesmo acontece com P_2O_5C e CaO_C ; MgO e MgO_C , P_2O_5 e $DIST_MAS$ (Figura 58 e 59).

Assim, conclui-se que as variáveis químicas exaustivas são significativas para explicar a variabilidade dos dados e podem ser utilizadas para prever comportamento metalúrgico dos atributos $DIST_MAS$, $P_2O_5_C$ e $MMAG$, que são variáveis de destaque no controle da produção na planta de beneficiamento.

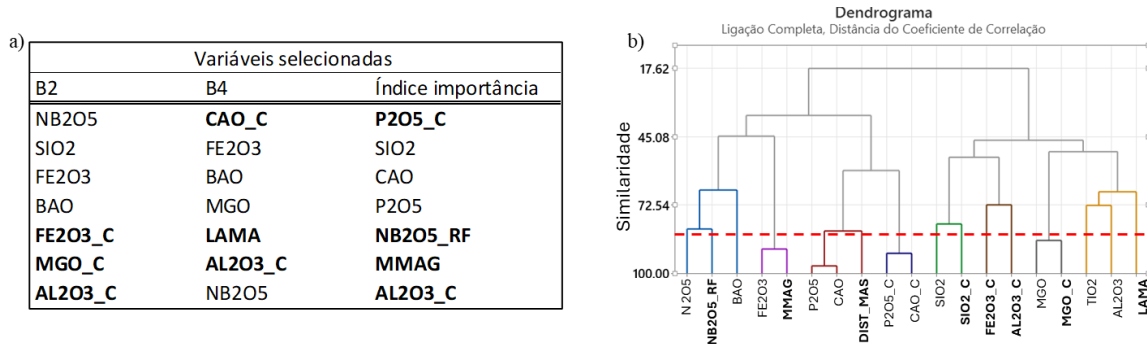


Figura 58. Em a) variáveis selecionadas na PCA, em b) similaridade entre as variáveis.

	Nb ₂ O ₅	P ₂ O ₅	SiO ₂	Fe ₂ O ₃	BaO	CaO	MgO	TiO ₂	Al ₂ O ₃	P ₂ O ₅ _C	CaO_C	Fe ₂ O ₃ _C	MgO_C	DIST_MA	LAMA	MMAG	Al ₂ O ₃ _C	SiO ₂ _C	Nb ₂ O ₅ _RF	
Nb ₂ O ₅	1.00																			
P ₂ O ₅	0.10	1.00																		
SiO ₂	-0.28	-0.61	1.00																	
Fe ₂ O ₃	0.16	-0.14	-0.47	1.00																
BaO	0.33	-0.14	-0.14	-0.10	1.00															
CaO	0.03	0.94	-0.56	-0.17	-0.16	1.00														
MgO	-0.21	-0.24	0.01	-0.22	-0.16	-0.13	1.00													
TiO ₂	-0.17	-0.36	0.02	0.17	-0.21	-0.43	0.23	1.00												
Al ₂ O ₃	-0.15	-0.30	0.08	-0.04	-0.13	-0.33	0.18	0.46	1.00											
P ₂ O ₅ _C	0.09	0.42	-0.34	0.30	-0.22	0.35	-0.36	-0.05	-0.07	1.00										
CaO_C	0.05	0.45	-0.35	0.14	-0.19	0.47	-0.14	-0.19	-0.16	0.84	1.00									
Fe ₂ O ₃ _C	0.00	-0.17	0.12	0.06	-0.01	-0.19	-0.07	0.05	0.07	-0.58	-0.65	1.00								
MgO_C	-0.11	-0.19	0.02	-0.19	-0.08	-0.10	0.74	0.09	0.10	-0.52	-0.28	0.06	1.00							
DIST_MAS	-0.04	0.67	-0.36	-0.27	-0.16	0.66	0.10	-0.22	-0.16	0.17	0.28	-0.18	0.09	1.00						
LAMA	0.06	-0.21	0.05	-0.13	0.14	-0.27	0.11	0.38	0.35	-0.14	-0.18	0.09	0.02	-0.19	1.00					
MMAG	0.08	-0.02	-0.46	0.81	-0.10	0.00	-0.10	0.04	-0.11	0.35	0.25	-0.19	-0.12	-0.19	-0.33	1.00				
Al ₂ O ₃ _C	-0.03	-0.24	0.07	0.15	-0.01	-0.28	-0.04	0.34	0.28	-0.38	-0.60	0.45	0.04	-0.27	0.13	0.02	1.00			
SiO ₂ _C	-0.25	-0.40	0.60	-0.29	-0.12	-0.36	0.13	0.00	0.04	-0.58	-0.57	0.29	0.18	-0.19	-0.02	-0.26	0.20	1.00		
Nb ₂ O ₅ _RF	0.64	0.22	-0.46	0.20	0.33	0.19	-0.11	-0.13	-0.13	0.15	0.14	-0.12	-0.08	0.06	-0.04	0.25	-0.05	-0.33	1.00	

Figura 59. Matriz de correlação entre as variáveis químicas e de processo.

6.5 Modelos preditivos

Regressão linear

As Figuras 60 a 62 apresentam a relação entre os valores calculados e os valores reais de recuperação mássica, massa de magnético e teor de P₂O₅ no concentrado. É possível notar para DIST_MAS resultados razoáveis de correlação entre 0,60 e 0,70 nas tipologias CBMS e CBMG, nas demais tipologias a nuvem de pontos é dispersa e as correlações são baixas. Para o MMAG todos os domínios apresentaram valores acima de 0,7, exceto CBMS. Já para o P₂O₅_C uma boa correlação (0,88 e 0,68) foi identificada somente para a tipologia FLPI na zona oxidada e micáceo.

Todos os pressupostos para validade do teste de regressão foram verificados e seguem no Apêndice F. É possível observar que os resíduos são heterocedásticos (lembrando um cone inclinado) ou não normalmente distribuídos para o P₂O₅_C na maioria dos domínios. Nos outros atributos, os resíduos exibem um comportamento homocedásticos. Com relação à normalidade dos resíduos, o teste de Anderson Darling foi realizado. Se o valor p for menor que o nível de significância (geralmente 0,05), rejeite a hipótese nula de que os dados têm de distribuição normal. A Tabela 6 mostra que os dados não seguem a distribuição normal para a maioria dos modelos, exceto nos modelos que se encontram marcados em negrito. Nota-se para a variável P₂O₅_C uma maior violação dos pressupostos de distribuição normal e variância dos resíduos. Diferente da recuperação mássica e massa de magnético, o teor de P₂O₅_C no concentrado não é uma variável aditiva, e espera-se que resultados inesperados sejam obtidos com a utilização de uma técnica de estimativa linear para prever variáveis não aditivas.

Vale ressaltar que, apesar do R-quadrado baixo obtido nos testes de regressão, as predictoras estatisticamente significativas podem ser utilizadas para tirar conclusões importantes sobre como as alterações nos valores das variáveis independentes estão associadas a alterações no valor da variável resposta. Ao interpretar o modelo (Tabela 7) percebe-se que o aumento do P_2O_5 e MgO e a diminuição do Fe_2O_3 são responsáveis pelo aumento da recuperação em mássica em todos os domínios, e que o P_2O_5 é a variável mais expressiva. Enquanto o teor de $P_2O_5_C$ no concentrado é reduzido em presença do MgO exceto para FO e CBMG no oxidado. Já a massa de magnético é influenciada pelo aumento de todas as variáveis com destaque para o Fe_2O_3 e MgO.

Tabela 6. Teste de normalidade dos resíduos para os modelos gerados.

Teste de normalidade dos resíduos Anderson-Darling (AD) $\alpha=0.005$			
Variável	DG	Oxidado p-valor	Micáceo p-valor
DIST_MASS	2	<0.005	<0.005
	56*	0.158	<0.006
	7	0.069	0.059
	8	0.58	0.017
P2O5_C	2	<0.005	<0.005
	56*	0.454	<0.010
	7	<0.005	<0.005
	8	<0.005	<0.005
MMAG	2	<0.005	<0.005
	56*	0.477	0.702
	7	<0.015	<0.015
	8	0.064	<0.016

Tabela 7. Equações da regressão linear para os domínios testados.

Equação de Regressão FO Oxidado DIST_MAS = 5.98 + 0.901 P2O5 - 0.0615 FE2O3 + 2.047 MGO	Equação de Regressão FO Oxidado P2O5_C = 3 1.15 + 0.1947 P2O5 + 0.0597 FE2O3 + 1.422 MGO	Equação de Regressão FO Oxidado MMAG = -10.36 + 0.170 P2O5 + 0.8357 FE2O3 + 6.726 MGO
Equação de Regressão FO micáceo DIST_MAS = 9.31 + 0.8885 P2O5 - 0.0965 FE2O3 + 0.857 MGO	Equação de Regressão FO micáceo P2O5_C = 2 8.370 + 0.3194 P2O5 + 0.1210 FE2O3 - 0.4217 MGO	Equação de Regressão FO micáceo MMAG = -12.18 + 0.035 P2O5 + 1.1293 FE2O3 + 0.580 MGO
Equação de Regressão CBMS oxidado DIST_MAS = 5.34 + 1.1592 P2O5 - 0.186 FE2O3 + 1.496 MGO	Equação de Regressão CBMS oxidado P2O5_C = 26.28 + 0.2875 P2O5 + 0.1631 FE2O3 + 1.232 MGO	Equação de Regressão CBMS oxidado MMAG = -5.82 + 0.2250 P2O5 + 0.7973 FE2O3 + 0.316 MGO
Equação de Regressão CBMS micáceo DIST_MAS = 4.17 + 1.0499 P2O5 - 0.0419 FE2O3 + 0.952 MGO	Equação de Regressão CBMS micáceo P2O5_C = 24.869 + 0.2985 P2O5 + 0.2405 FE2O3 - 0.007 MGO	Equação de Regressão CBMS micáceo MMAG = -2.69 + 0.0980 P2O5 + 0.683 FE2O3 + 2.688 MGO
Equação de Regressão CBMG oxidado DIST_MAS = -0.83 + 1.299 P2O5 - 0.0400 FE2O3 + 1.434 MGO	Equação de Regressão CBMG oxidado P2O5_C = 28.963 + 0.2066 P2O5 + 0.1111 FE2O3 + 1.214 MGO	Equação de Regressão CBMG oxidado MMAG = -8.99 + 0.165 P2O5 + 0.9210 FE2O3 + 2.045 MGO
Equação de Regressão CBMG micáceo DIST_MAS = 5.34 + 1.0890 P2O5 - 0.0871 FE2O3 + 0.985 MGO	Equação de Regressão CBMG micáceo P2O5_C = 20.06 + 0.5105 P2O5 + 0.2420 FE2O3 - 0.229 MGO	Equação de Regressão CBMG micáceo MMAG = -12.27 + 0.2628 P2O5 + 1.0228 FE2O3 + 0.717 MGO
Equação de Regressão FLPI oxidado DIST_MAS = 8.9 + 1.210 P2O5 - 0.138 FE2O3 - 0.52 MGO	Equação de Regressão FLPI oxidado P2O5_C = 12.39 + 2.724 P2O5 - 0.019 FE2O3 - 0.052 MGO	Equação de Regressão FLPI oxidado MMAG = -32.5 + 1.215 P2O5 + 1.212 FE2O3 + 1.34 MGO
Equação de Regressão FLPI micáceo DIST_MAS = 8.38 + 0.914 P2O5 - 0.0833 FE2O3 + 0.563 MGO	Equação de Regressão FLPI micáceo P2O5_C = 30.02 + 0.374 P2O5 + 0.0792 FE2O3 - 0.6374 MGO	Equação de Regressão FLPI micáceo MMAG = -21.73 + 0.593 P2O5 + 1.1759 FE2O3 + 0.718 MGO

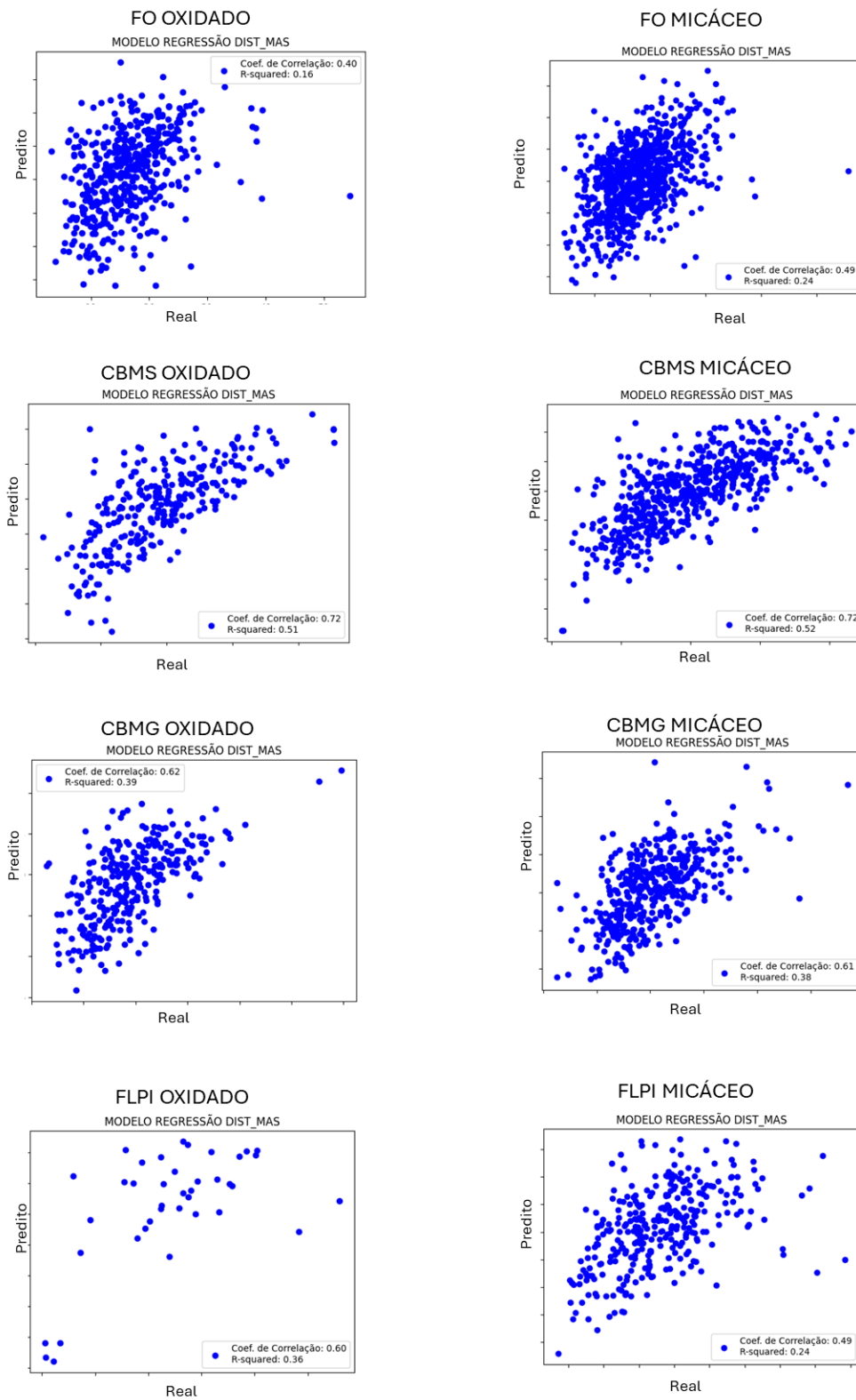


Figura 60. Gráficos do coeficiente de correlação preditos contra reais para DIST_MAS.

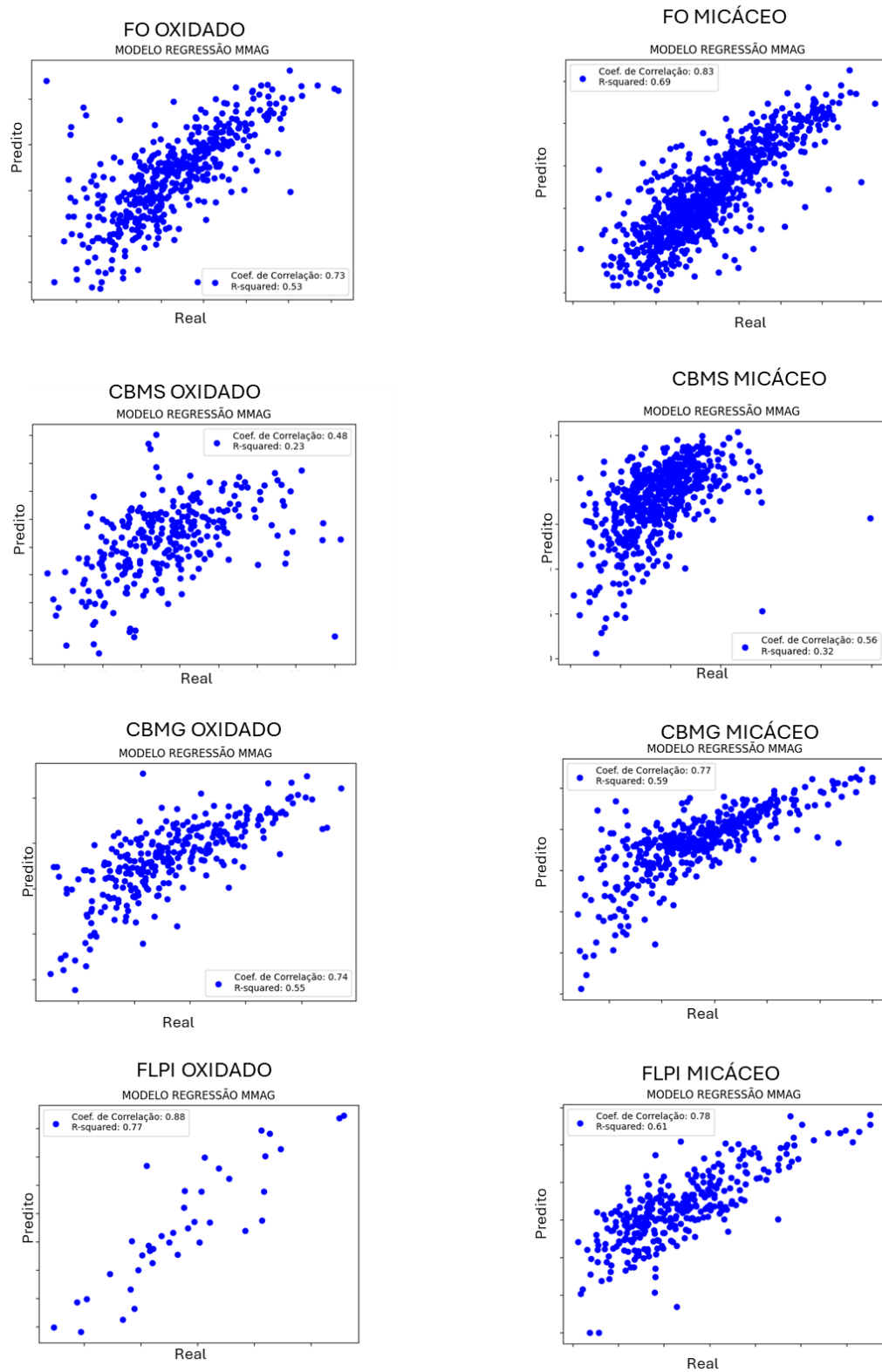


Figura 61. Gráficos do coeficiente de correlação preditos contras reais para MMAG.

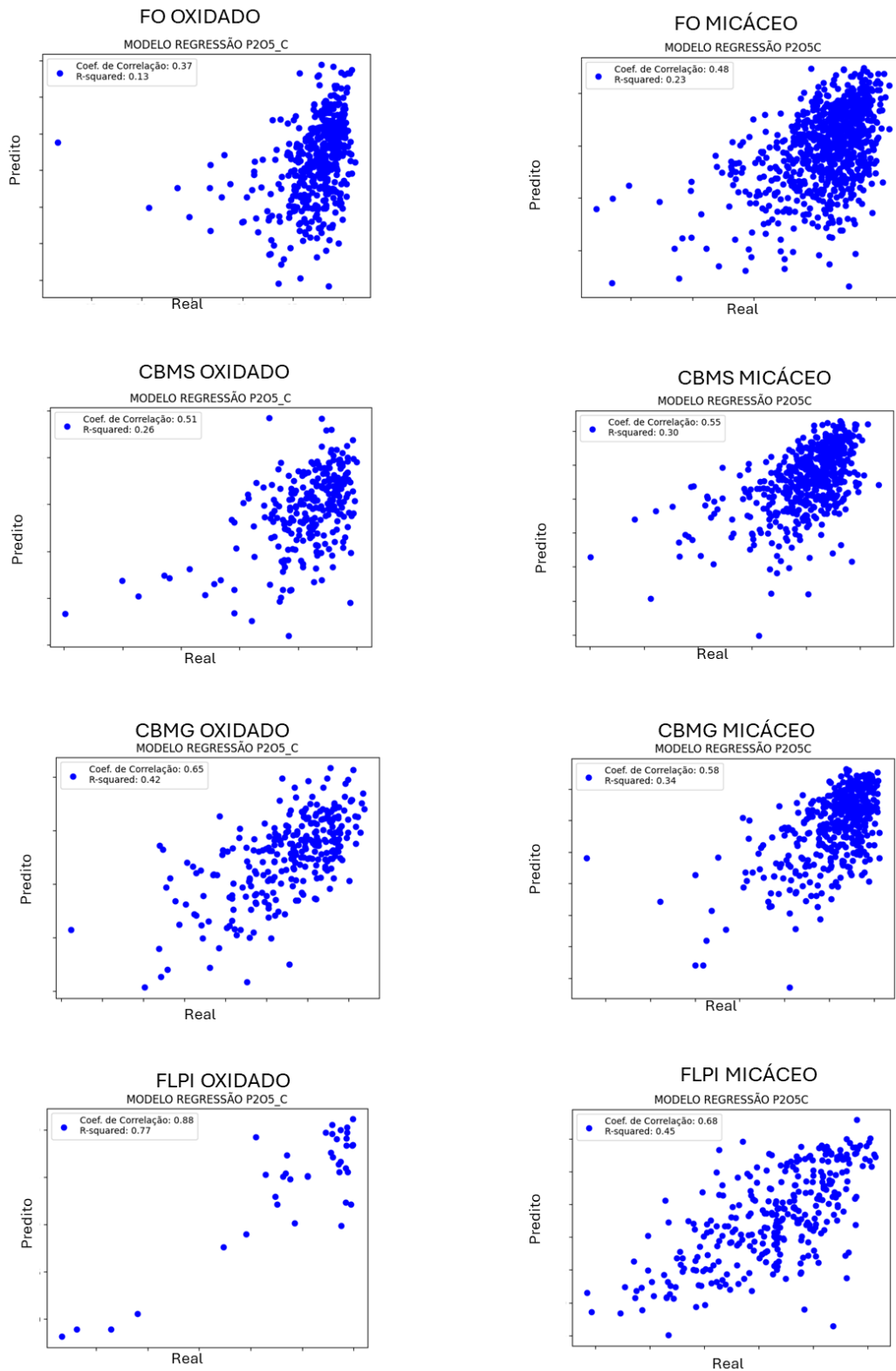


Figura 62. Gráficos do coeficiente de correlação preditos versus reais para P205_C

Random Forest

O RF foi a segunda técnica testada para os mesmos domínios e tipos de minérios. Porém, as árvores foram elaboradas, primeiro, utilizando todas as variáveis químicas a fim de reafirmar se as variáveis selecionadas na análise de PCA se confirmavam através do índice de importância como as variáveis mais significativas. As Tabelas 8 e 9 apresentam os índices de importância (variando de 0 a 1) obtidos através da técnica de *random forest* que corroboram com os resultados das análises multivariadas, nas quais os atributos P_2O_5 , Fe_2O_3 e MgO se destacam entre as demais. O BaO também é uma variável de destaque.

Tabela 8. Índice de importância da variável obtidos no RF para o horizonte oxidado.

OXIDADO	DISTA_MASS				P2O5_C				MMAG			
	FO	FLPI	CBMS	CBMG	FO	FLPI	CBMS	CBMG	FO	FLPI	CBMS	CBMG
NB2O5	0.11	0.13	0.09	0.09	0.06	0.03	0.06	0.03	0.05	0.03	0.12	0.05
P2O5	0.16	0.09	0.52	0.10	0.09	0.02	0.05	0.04	0.03	0.02	0.03	0.02
SiO2	0.10	0.07	0.04	0.09	0.09	0.03	0.09	0.09	0.05	0.01	0.10	0.05
FE2O3	0.11	0.16	0.04	0.10	0.13	0.16	0.08	0.06	0.64	0.74	0.35	0.59
BAO	0.09	0.15	0.04	0.22	0.29	0.06	0.38	0.42	0.03	0.03	0.06	0.08
CAO	0.15	0.12	0.14	0.10	0.05	0.05	0.08	0.02	0.03	0.04	0.06	0.02
MGO	0.09	0.14	0.04	0.13	0.14	0.56	0.10	0.21	0.11	0.05	0.09	0.11
TiO2	0.11	0.08	0.05	0.08	0.07	0.03	0.10	0.06	0.03	0.02	0.06	0.03
AL2O3	0.08	0.07	0.04	0.07	0.06	0.05	0.07	0.06	0.04	0.07	0.14	0.05

Tabela 9. Índice de importância da variável obtidos no RF para o horizonte micáceo.

MICACEO	DISTA_MASS				P2O5_C				MMAG			
	FO	FLPI	CBMS	CBMG	FO	FLPI	CBMS	CBMG	FO	FLPI	CBMS	CBMG
NB2O5	0.07	0.06	0.05	0.08	0.04	0.03	0.04	0.04	0.04	0.02	0.10	0.08
P2O5	0.09	0.30	0.16	0.04	0.04	0.08	0.08	0.08	0.02	0.01	0.03	0.09
SiO2	0.07	0.05	0.05	0.08	0.11	0.09	0.08	0.08	0.02	0.04	0.04	0.07
FE2O3	0.19	0.05	0.14	0.10	0.11	0.07	0.06	0.06	0.70	0.77	0.40	0.09
BAO	0.12	0.08	0.05	0.36	0.11	0.49	0.53	0.53	0.04	0.03	0.08	0.09
CAO	0.11	0.27	0.34	0.04	0.05	0.03	0.03	0.03	0.02	0.02	0.04	0.23
MGO	0.16	0.05	0.09	0.18	0.43	0.08	0.09	0.09	0.05	0.01	0.08	0.17
TiO2	0.12	0.06	0.06	0.06	0.07	0.08	0.03	0.03	0.05	0.02	0.09	0.08
AL2O3	0.07	0.07	0.06	0.07	0.06	0.04	0.05	0.05	0.04	0.08	0.13	0.10

Para escolher o número mínimo de amostras por folha, foram testados valores entre 2 e 10, obtendo-se melhores resultados com duas ou três amostras por folhas (Tabelas 10 e 11).

Tabela 10. Número mínimo de amostras por folha, horizonte oxidado.

BANCO DE TREINO/TESTE - COEFICIENTE TESTES - OXIDADO										
Variável resposta	DG	NÚMERO MÍNIMO DE AMOSTRAS POR FOLHA - RF (3 VARIÁVEIS)								
		2	3	4	5	6	7	8	9	10
DIST_MASS	FO	0.82	0.79	0.75	0.71	0.68	0.65	0.64	0.62	0.61
	FLPI	0.82	0.68	0.57	0.47	0.42	0.38	0.36	0.3	0.3
	CBMS	0.89	0.85	0.83	0.81	0.8	0.79	0.78	0.78	0.77
	CBMG	0.85	0.85	0.81	0.79	0.78	0.76	0.75	0.74	0.74
P2O5_C	FO	0.79	0.74	0.71	0.68	0.66	0.65	0.63	0.62	0.61
	FLPI	0.95	0.76	0.69	0.65	0.63	0.61	0.58	0.55	0.52
	CBMS	0.83	0.77	0.73	0.7	0.67	0.65	0.63	0.62	0.61
	CBMG	0.87	0.83	0.81	0.78	0.77	0.75	0.74	0.73	0.72
MMAG	FO	0.88	0.88	0.86	0.85	0.83	0.82	0.82	0.81	0.8
	FLPI	0.94	0.89	0.87	0.86	0.85	0.84	0.82	0.8	0.78
	CBMS	0.81	0.72	0.67	0.63	0.59	0.57	0.55	0.53	0.51
	CBMG	0.9	0.86	0.83	0.82	0.8	0.79	0.78	0.77	0.77

Tabela 11. Número mínimo de amostras por folha, horizonte micáceo

BANCO DE TREINO/TESTE - COEFICIENTE TESTES - MICACEO										
Variável resposta	DG	NÚMERO MÍNIMO DE AMOSTRAS POR FOLHA - RF (3 VARIÁVEIS)								
		2	3	4	5	6	7	8	9	10
DIST_MASS	FO	0.84	0.78	0.74	0.71	0.68	0.66	0.64	0.62	0.61
	FLPI	0.84	0.68	0.63	0.58	0.54	0.5	0.48	0.46	0.45
	CBMS	0.88	0.83	0.8	0.79	0.77	0.76	0.75	0.74	0.73
	CBMG	0.86	0.82	0.8	0.78	0.76	0.74	0.73	0.71	0.7
P2O5_C	FO	0.82	0.76	0.72	0.68	0.66	0.63	0.61	0.6	0.59
	FLPI	0.88	0.8	0.76	0.72	0.7	0.68	0.66	0.64	0.64
	CBMS	0.84	0.8	0.76	0.73	0.71	0.69	0.68	0.66	0.65
	CBMG	0.86	0.86	0.81	0.79	0.76	0.74	0.73	0.71	0.7
MMAG	FO	0.94	0.9	0.88	0.87	0.87	0.86	0.85	0.85	0.84
	FLPI	0.92	0.92	0.91	0.9	0.9	0.89	0.87	0.86	0.86
	CBMS	0.85	0.83	0.8	0.77	0.75	0.73	0.72	0.7	0.69
	CBMG	0.93	0.92	0.91	0.9	0.89	0.89	0.88	0.87	0.87

Os modelos gerados pelo RF apresentaram comportamento mais uniforme dos dados ao redor da reta identidade (Figuras 63 a 66) para todas as variáveis respostas em todos os domínios avaliados quando comparados a regressão linear. Os coeficientes de correlação entre os valores preditos e os valores observados também foram maiores.

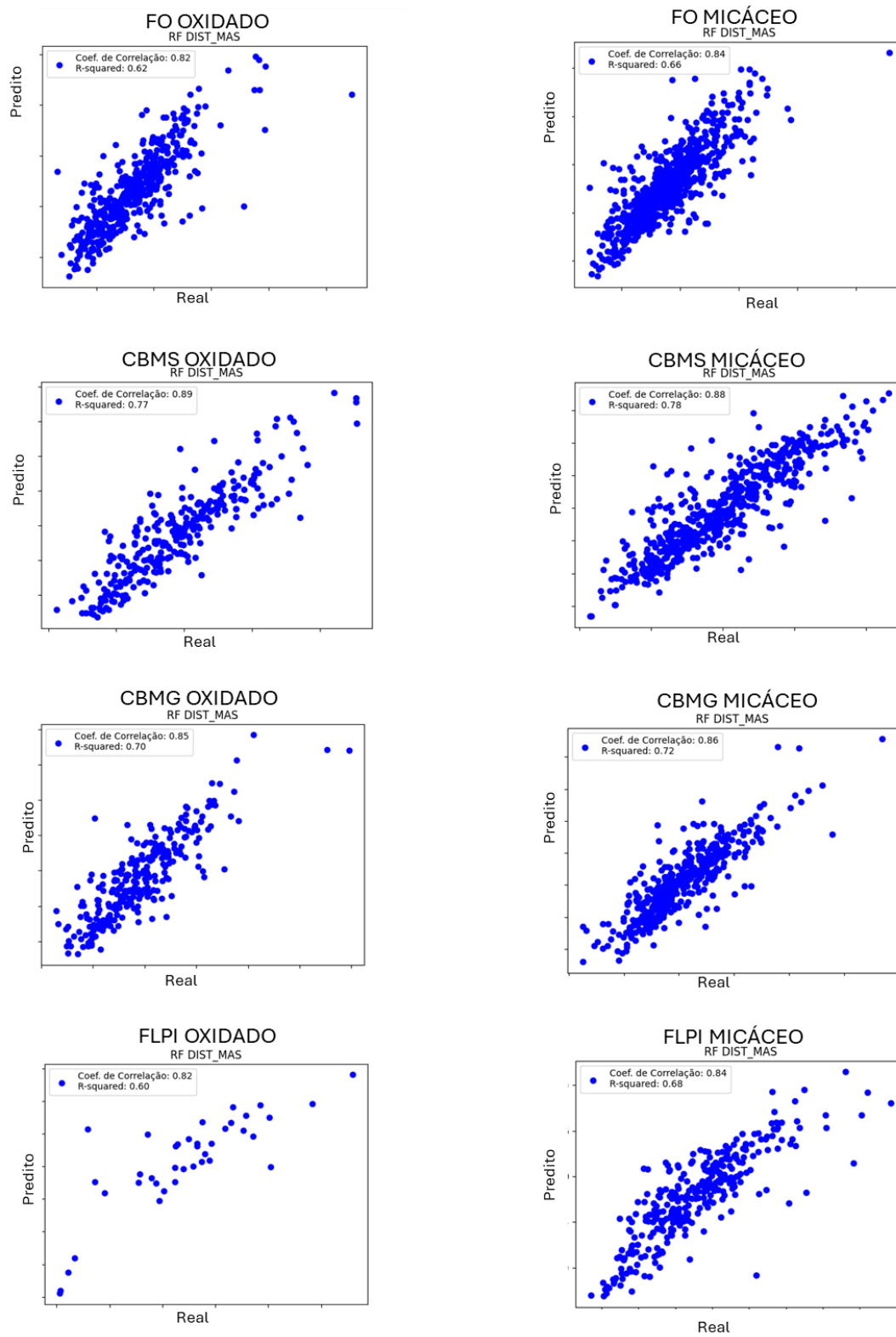


Figura 63. Gráficos do coeficiente de correlação preditos versus reais para DIST_MAS.

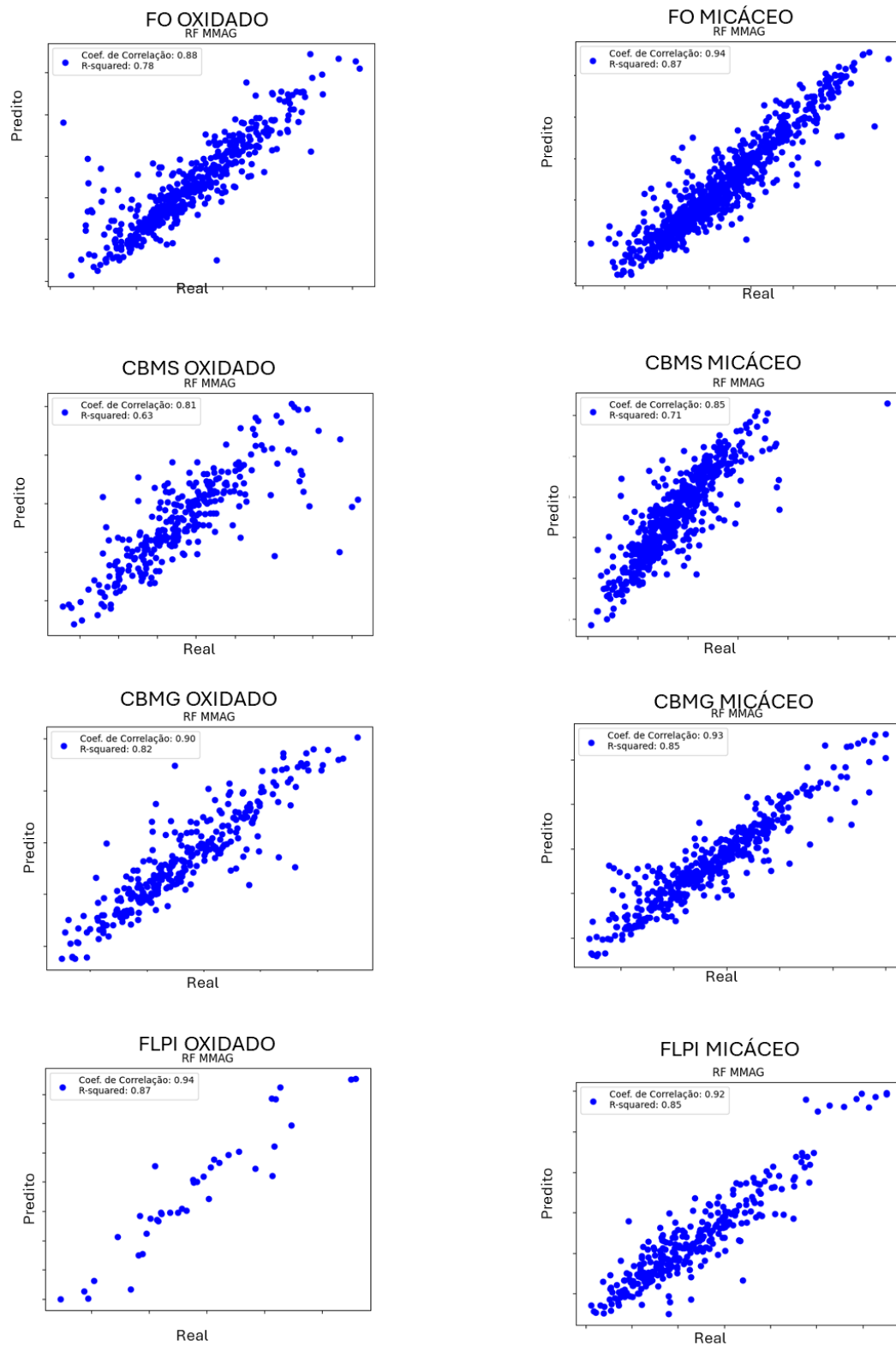


Figura 64. Gráficos do coeficiente de correlação preditos versus reais para MMAG

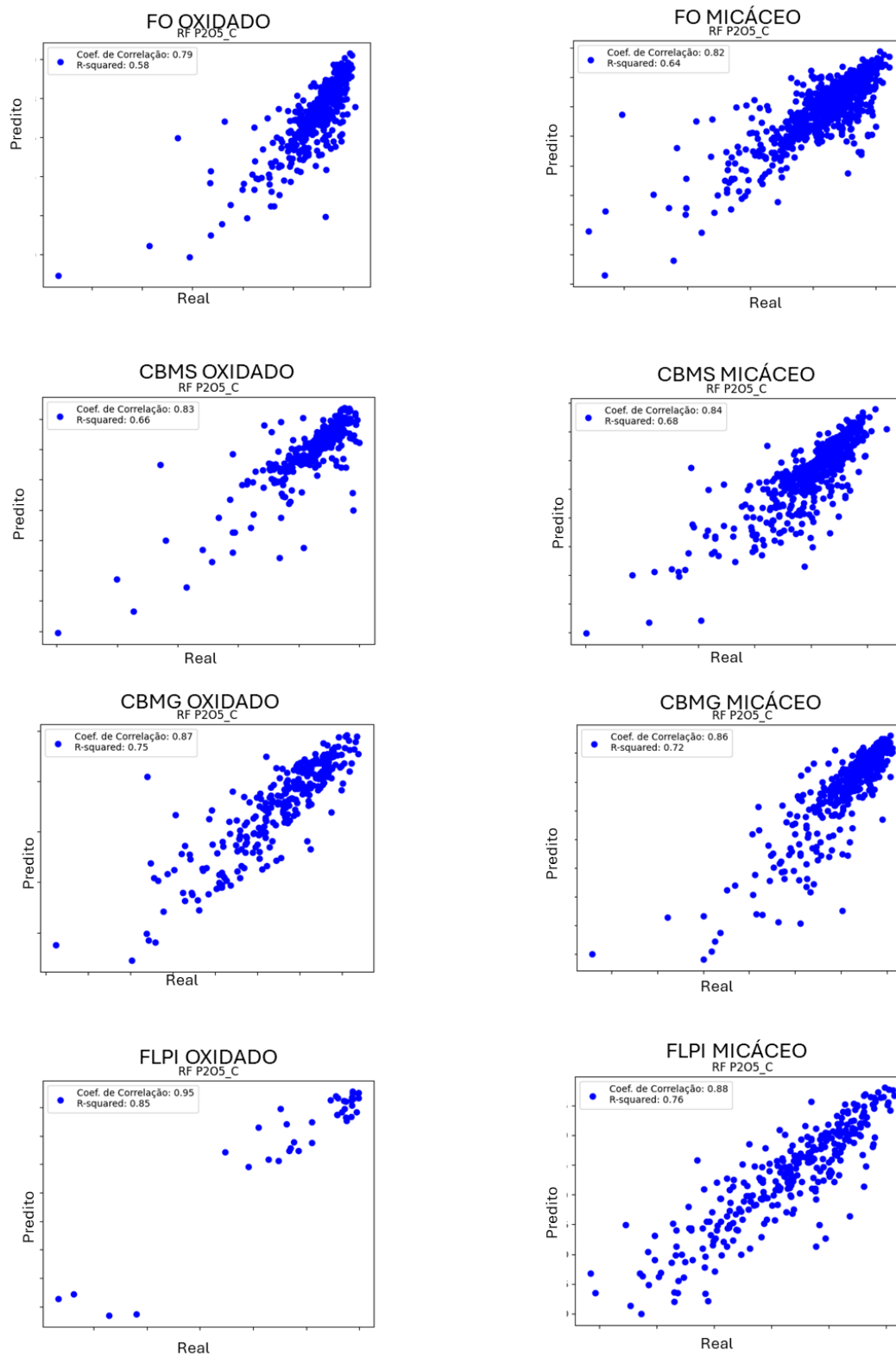


Figura 65. Gráficos do coeficiente de correlação preditos versus reais para $P_2O_5_C$.

SVM

Ao aplicar o SVM, alguns parâmetros de ajuste do modelo se tornam necessários. Esses parâmetros são inseridos diretamente no algoritmo e afetam diretamente os resultados,

influenciando a velocidade de treinamento, a propensão ao *overfitting*, a tolerância a erros de classificação, entre outros aspectos. O primeiro parâmetro é o valor C e está presente nas duas formas da SVM sendo responsável por comedir a tolerância nos erros de classificação, ou seja, é uma penalização imposta ao modelo durante o treinamento para cada amostra classificada erroneamente. Um valor muito alto de C faz com que o algoritmo treine o modelo visando a separação completa entre classes podendo causar *overfitting*. Por outro lado, baixos valores de C flexibilizam a etapa de treinamento e permitem fronteiras de decisão com erros, podendo levar a um *underfitting*. A fim de avaliar a performance do parâmetro C foram testados alguns valores aleatórios e os resultados seguem nas Tabelas 12 e 13. Os parâmetros escolhidos (em negrito) foram os que retornaram a menor raiz do erro quadrático médio (RMSE) no banco de teste, diminuído a possibilidade de *overfitting*, com os melhores coeficientes dos valores preditos e os valores observados.

Tabela 12. Resultados dos modelos SVM não linear para diferentes parâmetros C, zona oxidado. À direita as correlações e à esquerda o RMSE.

Variável resposta	DG	COEFICIENTES REAL vs PREDITOS DOS MODELOS						RMSE DO BANCO DE TESTE					
		Modelo SVM, kernel RBF, VARIANDO C_ OXIDADO						Modelo SVM, kernel RBF, VARIANDO C_ OXIDADO					
		0.01	1	10	35	50	100	0.01	1	10	35	50	100
DIST_MASS	FO	0.37	0.41	0.44	0.44	0.44	0.44	5.88	5.45	5.57	5.91	5.99	6.08
	FLPI	0.31	0.35	0.59	0.67	0.67	0.66	7.31	6.93	6.46	6.81	6.87	6.96
	CBMS	0.70	0.71	0.71	0.71	0.71	0.72	7.56	5.30	5.19	5.23	5.25	5.29
	CBMG	0.62	0.63	0.68	0.69	0.69	0.76	9.05	8.30	7.76	7.10	6.62	6.43
P2O5_C	FO	0.44	0.39	0.45	0.45	0.44	0.44	2.66	2.53	2.57	2.56	2.56	2.56
	FLPI	0.72	0.79	0.93	0.94	0.94	0.95	10.19	8.39	3.84	3.51	3.45	3.50
	CBMS	0.43	0.44	0.58	0.56	0.57	0.61	3.18	2.82	2.83	3.10	3.16	3.04
	CBMG	0.63	0.67	0.68	0.68	0.68	0.70	2.13	1.80	1.71	1.73	1.71	1.73
MMAG	FO	0.66	0.72	0.75	0.76	0.76	0.76	12.33	9.81	10.04	10.35	10.43	10.60
	FLPI	0.74	0.78	0.85	0.85	0.85	0.88	9.87	8.70	6.35	6.31	6.21	6.13
	CBMS	0.39	0.54	0.49	0.51	0.51	0.58	8.15	7.53	7.93	7.98	7.95	7.89
	CBMG	0.67	0.71	0.79	0.77	0.77	0.81	9.74	7.45	7.34	7.40	7.40	7.51

Tabela 13. Resultados dos modelos SVM não linear para diferentes parâmetros C na zona micáceo. À direita as correlações e à esquerda o RMSE.

Variável resposta	DG	COEFICIENTES DOS MODELOS						RMSE DO BANCO DE TESTE					
		Modelo SVM, kernel RBF, VARIANDO C_ MICACEO						Modelo SVM, kernel RBF, VARIANDO C_ MICACEO					
		0.01	1	10	35	50	100	0.01	1	10	35	50	100
DIST_MASS	FO	0.40	0.46	0.52	0.51	0.51	0.53	5.51	4.78	4.65	4.69	4.72	4.81
	FLPI	0.45	0.51	0.57	0.59	0.59	0.6	6.25	5.62	5.85	6.24	6.40	6.61
	CBMS	0.63	0.72	0.67	0.69	0.69	0.74	7.83	5.99	6.14	6.22	6.18	6.09
	CBMG	0.43	0.52	0.63	0.63	0.64	0.64	6.73	5.70	5.58	5.43	5.46	5.55
P2O5_C	FO	0.39	0.50	0.50	0.49	0.50	0.55	3.09	2.85	2.87	2.91	2.91	2.92
	FLPI	0.50	0.56	0.74	0.58	0.58	0.77	5.84	3.89	3.57	3.61	3.67	3.83
	CBMS	0.51	0.56	0.59	0.59	0.59	0.61	3.09	2.75	2.81	2.85	2.85	2.87
	CBMG	0.57	0.63	0.64	0.67	0.64	0.67	4.15	3.37	3.12	3.03	3.04	3.09
MMAG	FO	0.73	0.79	0.85	0.85	0.82	0.85	11.21	6.82	6.56	6.57	6.56	6.54
	FLPI	0.77	0.81	0.84	0.84	0.87	0.85	13.13	9.09	6.60	6.56	6.61	6.89
	CBMS	0.57	0.60	0.61	0.63	0.63	0.63	6.61	5.71	5.66	5.73	5.77	5.76
	CBMG	0.74	0.81	0.85	0.84	0.85	0.84	12.45	8.38	6.93	6.40	6.44	6.51

O outro parâmetro é gamma (parâmetro de escala) introduzido somente no SVM não linear e foi avaliado com os valores 0,0001, 0,001, 0,1, 1, 10 e 20. Observou-se uma pequena variação nos resultados de correlação e erro absoluto para os valores entre 0,001 e 1, e resultados inferiores para valores de 10 ou mais. Portanto, o valor de 0,001 foi selecionado para os modelos de predição. Uma comparação entre os modelos SVM linear e RBF é apresentada na Tabela 14, onde se observa que os modelos SVM não lineares apresentaram resultados superiores. Os gráficos que comparam os valores preditos com os valores reais estão disponíveis nas Figuras 66 a 68.

Tabela 14. Comparação dos coeficientes de correlação entre os modelos SVM linear e não linear.

Variável	DGEO	SVM - OXIDADO		SVM - MICACEO	
		LINEAR	RBF	LINEAR	RBF
DIST_MASS	FO	0.37	0.41	0.49	0.52
	FLPI	0.57	0.67	0.49	0.51
	CBMS	0.72	0.71	0.71	0.72
	CBMG	0.63	0.76	0.61	0.63
P2O5_C	FO	0.35	0.39	0.49	0.5
	FLPI	0.8	0.95	0.67	0.74
	CBMS	0.43	0.58	0.55	0.56
	CBMG	0.62	0.7	0.58	0.67
MMAG	FO	0.72	0.72	0.83	0.85
	FLPI	0.87	0.88	0.83	0.84
	CBMS	0.45	0.54	0.57	0.61
	CBMG	0.76	0.79	0.81	0.84

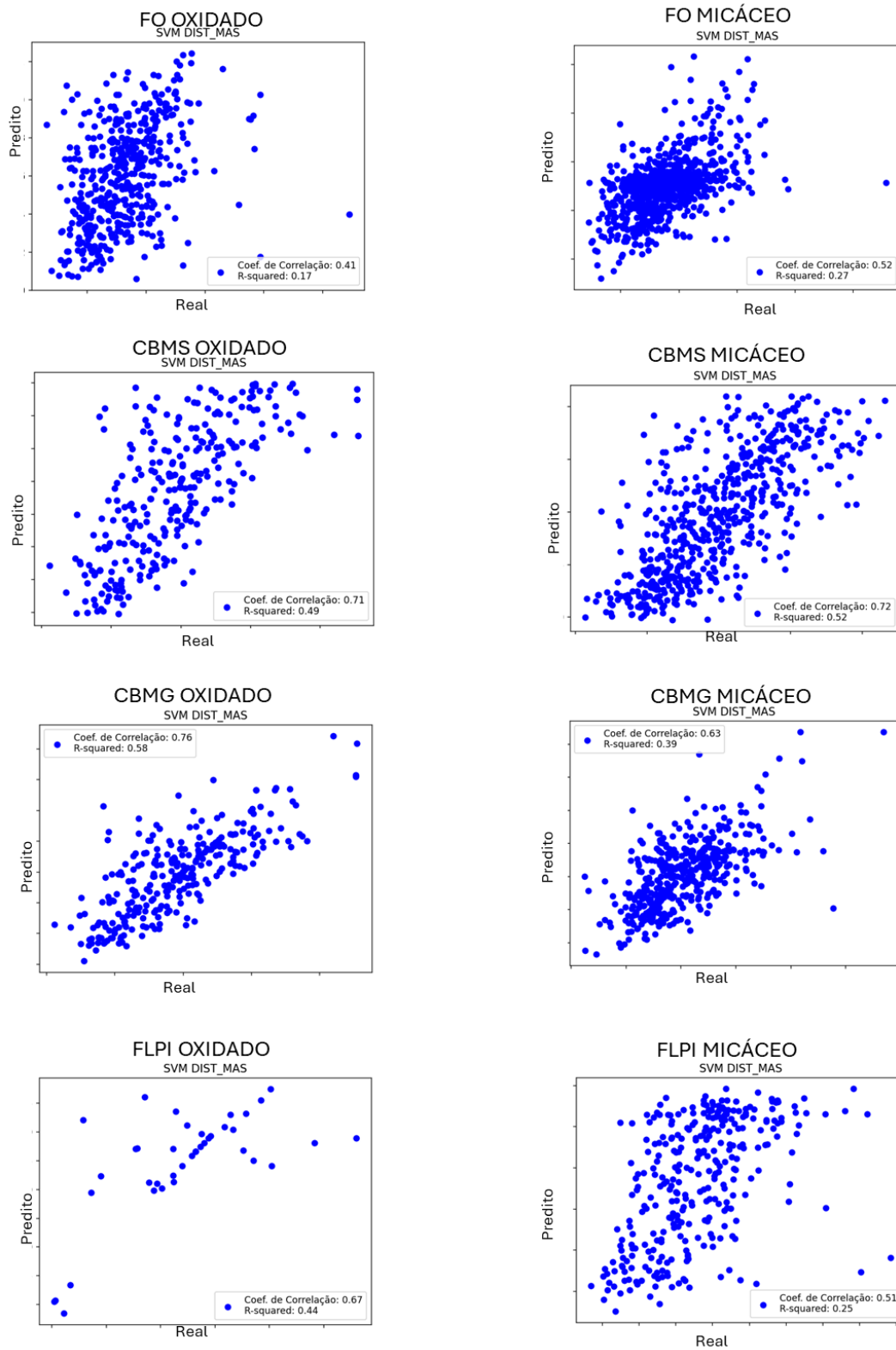


Figura 66. Gráficos do coeficiente de correlação preditos versus reais para DIST_MAS.

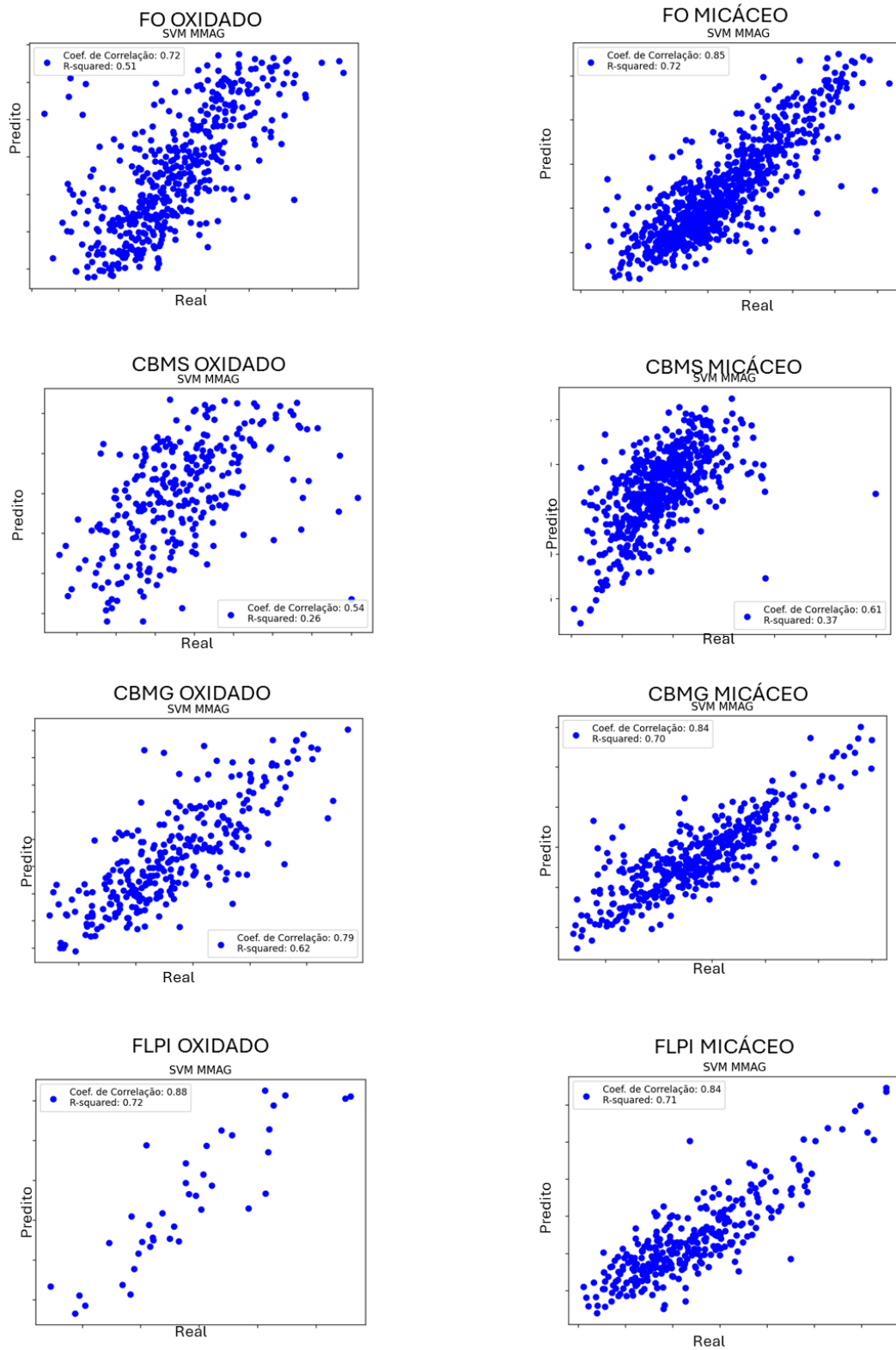


Figura 67. Gráficos do coeficiente de correlação preditos versus reais para MMAG.

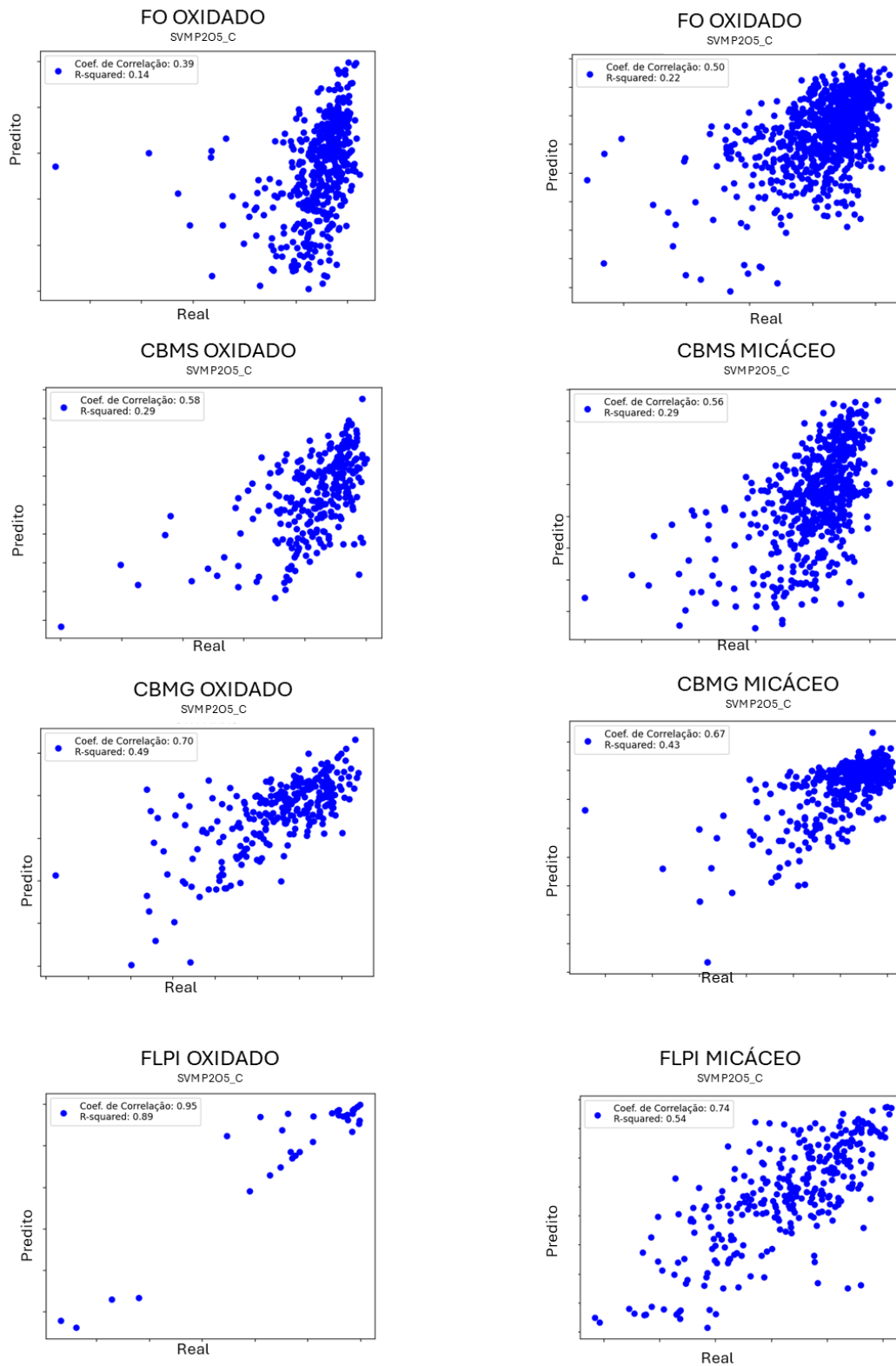


Figura 68. Gráficos do coeficiente de correlação preditos versus reais para $P_2O_5_C$.

Verificação das técnicas de predição

Os modelos de predição apresentaram, de maneira geral, índices satisfatórios para previsão das respostas metalúrgicas. O *Random forest* gerou os maiores coeficientes de correlação, porém, os melhores resultados para as métricas: erro médio absoluto (MAE) e erro quadrático médio (MSE) foram obtidos no modelo SVM e RL. Os valores de MAE e MSE (Tabelas 15 e 16) entre os três modelos se mantiveram próximos. Contudo, o coeficiente de correlação indicou uma diferença mais significativa do modelo RF em comparação com RL e SVM, tornando o RF o método mais indicado para a predição das variáveis do processo. Esses resultados são uma importante contribuição das ciências da computação para previsibilidade da usina, pois permitiram a simulação dos teores e aplicação da função de transferências nos dados simulados para compor o modelo geometalúrgico final.

Tabela 15. Avaliação dos resultados da validação cruzada 5-fold – oxidado.

Variável resposta		OXIDADOS								
		RL			RF			SVM		
		MAE	MSE	CORRELAÇÃO	MAE	MSE	CORRELAÇÃO	MAE	MSE	CORRELAÇÃO
DIST_MAS	FO	4.42	5.99	0.4	4.69	6.39	0.82	4.41	6.00	0.41
	FLPI	3.98	5.40	0.6	4.86	6.14	0.82	3.52	5.02	0.67
	CBMS	4.53	5.96	0.72	5.16	6.55	0.89	4.5	5.95	0.71
	CBMG	4.71	6.21	0.62	4.83	6.56	0.85	4.66	6.25	0.76
P2O5_C	FO	1.71	2.75	0.37	1.81	2.84	0.79	1.66	2.82	0.39
	FLPI	3.01	4.04	0.88	3	4.12	0.95	2.06	2.89	0.95
	CBMS	2	2.91	0.51	2.08	3.03	0.83	1.93	2.93	0.58
	CBMG	1.4	1.84	0.65	1.48	1.97	0.87	1.48	1.96	0.7
MMAG	FO	5.77	7.89	0.73	6.25	8.45	0.88	5.56	7.81	0.72
	FLPI	4.81	6.37	0.88	6.54	8.37	0.94	5.11	6.91	0.88
	CBMS	4.59	5.96	0.48	4.58	5.91	0.81	4.58	5.90	0.54
	CBMG	5.69	7.16	0.74	5.15	6.62	0.9	5.35	6.85	0.79

Tabela 16. Avaliação dos resultados da validação cruzada 5-fold – micáceo

Variável resposta		MICACEO								
		RL			RF			SVM		
		MAE	MSE	CORRELAÇÃO	MAE	MSE	CORRELAÇÃO	MAE	MSE	CORRELAÇÃO
DIST_MAS	FO	3.91	5.20	0.49	4.06	5.40	0.84	3.92	5.21	0.52
	FLPI	4.66	6.22	0.49	4.86	6.37	0.84	4.69	6.11	0.51
	CBMS	4.53	5.80	0.72	4.75	6.15	0.88	4.45	5.77	0.72
	CBMG	4.69	6.12	0.61	5.08	6.66	0.86	4.71	6.26	0.63
P2O5_C	FO	2.02	2.75	0.48	2.07	2.85	0.82	1.97	2.80	0.5
	FLPI	3.27	4.15	0.68	3.33	4.25	0.88	3.25	4.25	0.74
	CBMS	2.00	2.87	0.55	2.19	3.10	0.84	1.97	2.92	0.56
	CBMG	2.29	3.28	0.58	2.38	3.36	0.86	2.29	3.37	0.67
MMAG	FO	5.25	7.16	0.83	5.31	7.05	0.94	5.02	6.89	0.85
	FLPI	6.25	7.85	0.78	6.22	7.81	0.92	5.87	7.54	0.84
	CBMS	4.37	6.06	0.56	4.54	6.30	0.85	4.19	5.91	0.61
	CBMG	5.25	7.09	0.77	4.93	6.83	0.93	4.58	6.43	0.84

Por fim, para avaliar a implicação de operar com um número limitado de variáveis, os modelos preditivos calibrados com os parâmetros selecionados foram testados usando todas as variáveis químicas e comparados aos resultados alcançados usando apenas P_2O_5 , Fe_2O_3 e MgO .

Observou-se que não houve um impacto significativo nos resultados (Tabela 17), particularmente no modelo RF.

Tabela 17. Comparação entre os modelos preditivos utilizando todas as variáveis primárias e as três variáveis selecionadas para oxidado e micáceo.

COEFICIENTES DE CORRELAÇÃO PREDITO vs OBSERVADO													
Variável resposta	DG-MC	OXIDADO						MICÁCEO					
		Todas variáveis			P2O5, FE2O3, MGO			Todas variáveis			P2O5, FE2O3, MGO		
		RL	RF	SVM	RL	RF	SVM	RL	RF	SVM	RL	RF	SVM
DIST_MASS	FO	0.44	0.86	0.54	0.4	0.82	0.41	0.5	0.88	0.59	0.49	0.84	0.52
	FLPI	0.75	0.91	0.87	0.6	0.82	0.67	0.6	0.89	0.62	0.49	0.84	0.51
	CBMS	0.73	0.92	0.81	0.72	0.89	0.71	0.73	0.91	0.77	0.72	0.88	0.72
	CBMG	0.64	0.87	0.79	0.62	0.85	0.76	0.64	0.89	0.69	0.61	0.86	0.63
P2O5_C	FO	0.57	0.87	0.71	0.37	0.79	0.39	0.6	0.87	0.7	0.48	0.82	0.5
	FLPI	0.91	0.97	0.97	0.88	0.95	0.95	0.71	0.91	0.85	0.68	0.88	0.74
	CBMS	0.65	0.89	0.79	0.51	0.83	0.58	0.71	0.92	0.86	0.55	0.84	0.56
	CBMG	0.71	0.9	0.85	0.65	0.87	0.7	0.74	0.91	0.83	0.58	0.86	0.67
MMAG	FO	0.77	0.9	0.82	0.73	0.88	0.72	0.85	0.95	0.87	0.83	0.94	0.85
	FLPI	0.91	0.96	0.97	0.88	0.94	0.88	0.84	0.95	0.86	0.78	0.92	0.84
	CBMS	0.64	0.87	0.78	0.48	0.81	0.54	0.61	0.88	0.69	0.56	0.85	0.61
	CBMG	0.79	0.92	0.88	0.74	0.9	0.79	0.81	0.95	0.86	0.77	0.93	0.84

6.6 Simulação das variáveis primárias

Análise exploratória dos dados

As Figuras 69 a 76 apresentam os histogramas dos dados originais com média dos teores e dos dados transformados para o espaço gaussiano sendo essa uma condição para a simulação sequencial gaussiana. A normalização da distribuição amostral tem como objetivo obter o conhecimento da distribuição condicional para a estimativa e amenizar dados discrepantes.

Em todos os domínios nota-se nos histogramas que, a separação nos horizontes oxidado e micáceo impactaram nos teores médios de MgO (Tabela 18). Isso corrobora com a descrição do perfil intempérico cujos valores de MgO são indicadores da evolução do intemperismo (redução do MgO da base para o topo do perfil). Observam-se valores elevados deste óxido no domínio FLPI, especificamente no nível micáceo; também se identifica mais de uma população (Figura 72), efeito da junção do domínio flogopitito e o domínio picrito com xenólitos. Em relação ao P₂O₅, não é observado grandes variações nos diferentes grupos, exceto no FLPI, que mostra valores inferiores. O Fe₂O₃ apresenta semelhanças nos domínios FO e CBMG, com redução no CBMS. No FLPI, os valores de Fe₂O₃ diminuem significativamente no horizonte micáceo, o que também está relacionado ao protólito e ao perfil de alteração (Tabela 18). Essas observações apoiam a divisão dos domínios segundo os horizontes de intemperismo. Em geral,

os histogramas apresentam uma distribuição assimétrica positiva e tem características estatísticas distintas entre os domínios analisados.

Tabela 18. Sumário estatístico dos histogramas dos teores do banco de dados.

DOMÍNIO		P2O5		Fe2O3		MgO	
		Média	Variância	Média	Variância	Média	Variância
OXI	FO	16.4	5.8	55.1	85.6	1.4	0.3
	FLPI	8.3	7.4	59.0	116.9	3.4	1.3
	CBMS	15.6	38.8	34.5	20.4	0.9	0.2
	CBMG	20.2	18.4	43.8	84.7	1.0	0.4
MICACEO	FO	16.0	5.2	51.4	79.2	2.8	3.8
	FLPI	8.4	6.1	28.9	76.7	18.6	30.8
	CBMS	17.9	34.8	33.2	20.5	1.9	1.0
	CBMG	19.9	17.0	41.1	96.1	2.6	7.7

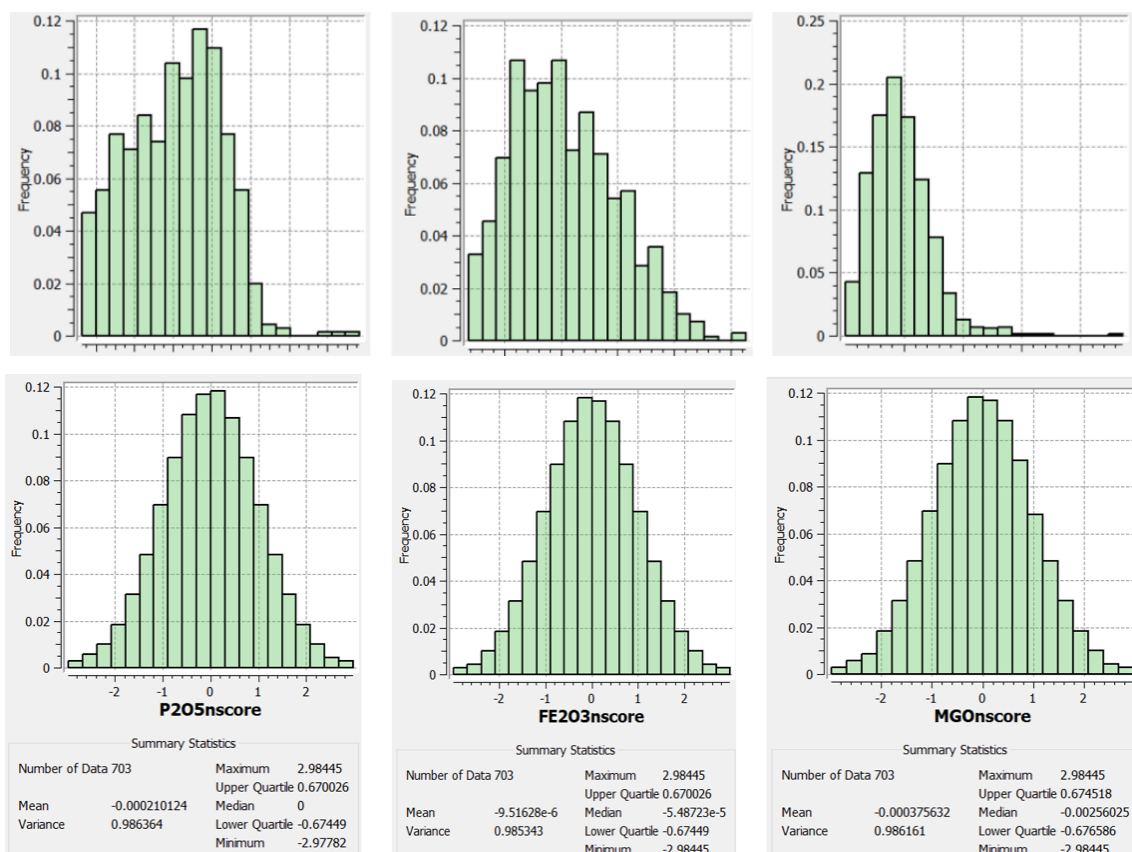


Figura 69. Histogramas dos teores P_2O_5 , Fe_2O_3 e MgO na parte superior e dos teores normalizados para o FO no oxidado (na parte inferior).

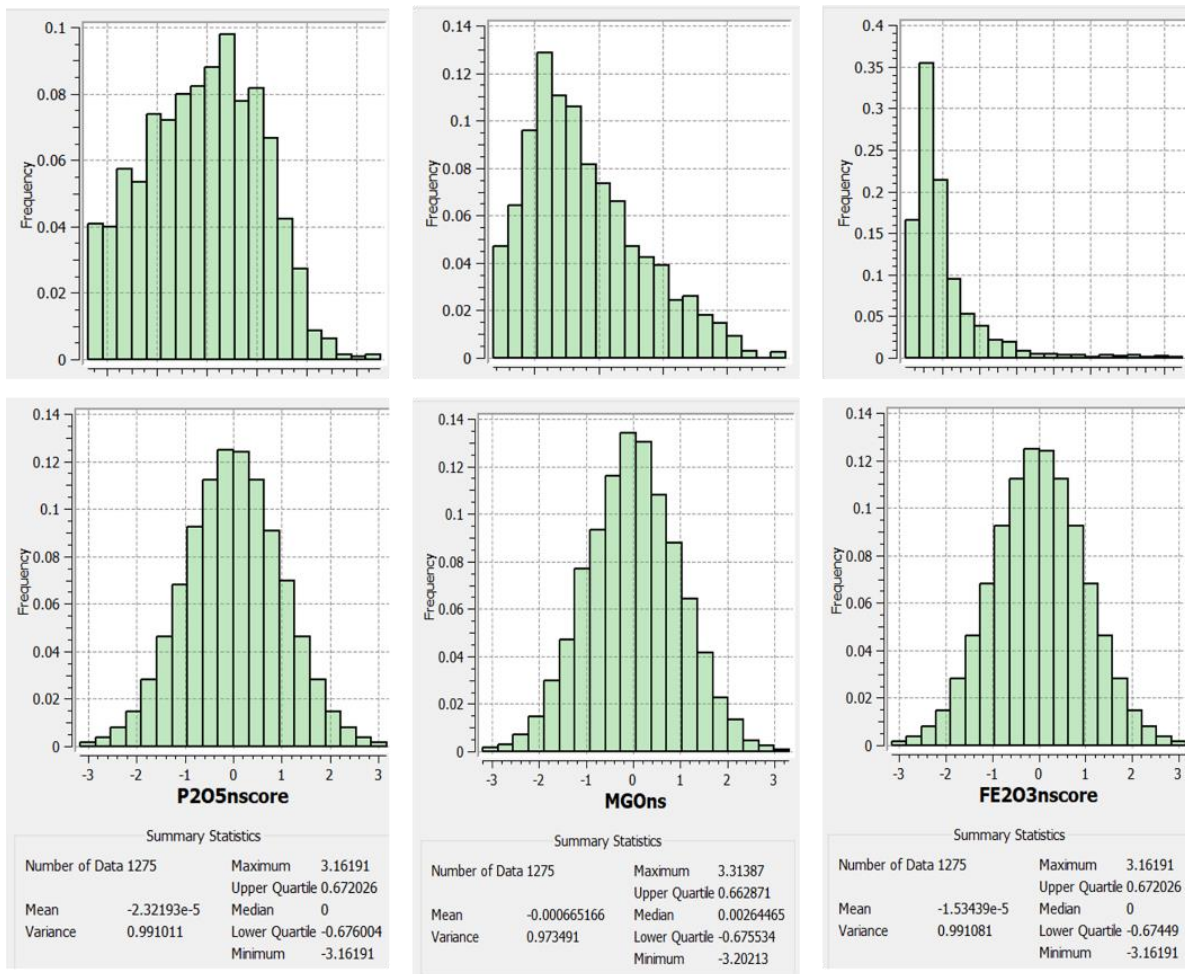


Figura 70. Histogramas dos teores P₂O₅, Fe₂O₃ e MgO na parte superior e dos teores normalizados para o FO no micáceo (na parte inferior da figura).

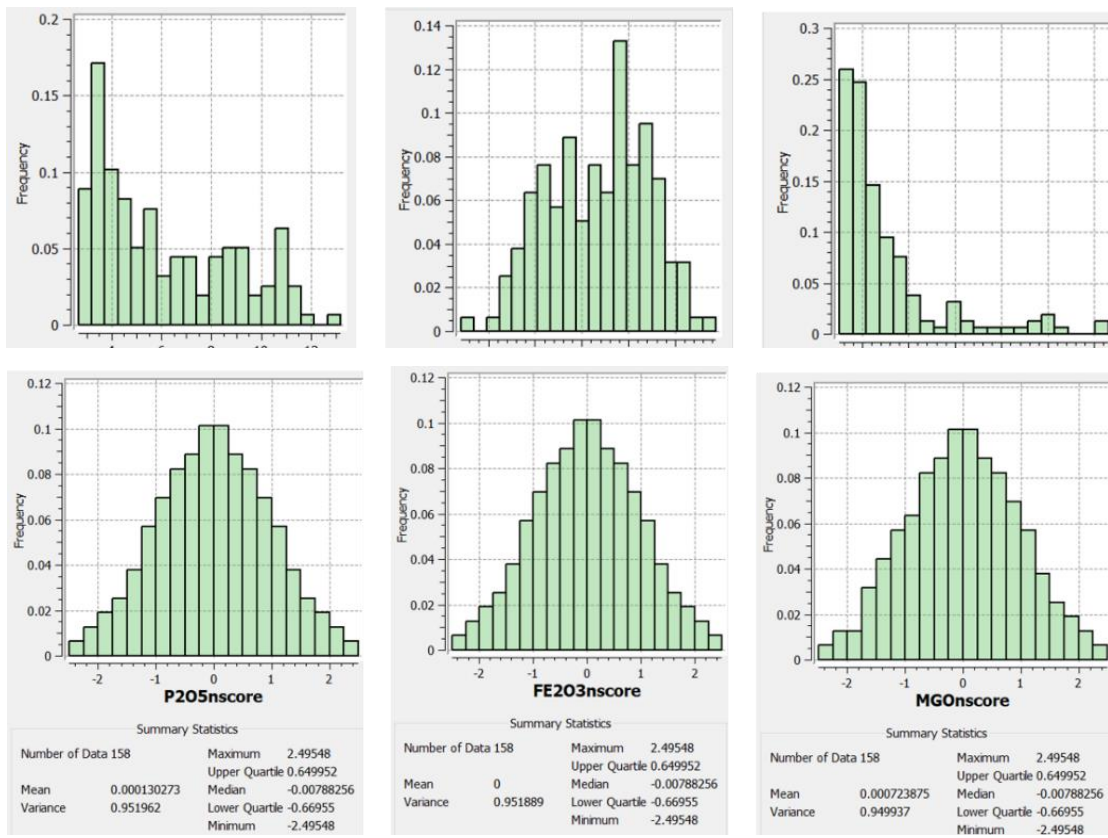


Figura 71. Histogramas dos teores P₂O₅, Fe₂O₃ e MgO na parte superior e dos teores normalizados para o FLPI no oxidado (na parte inferior da figura).

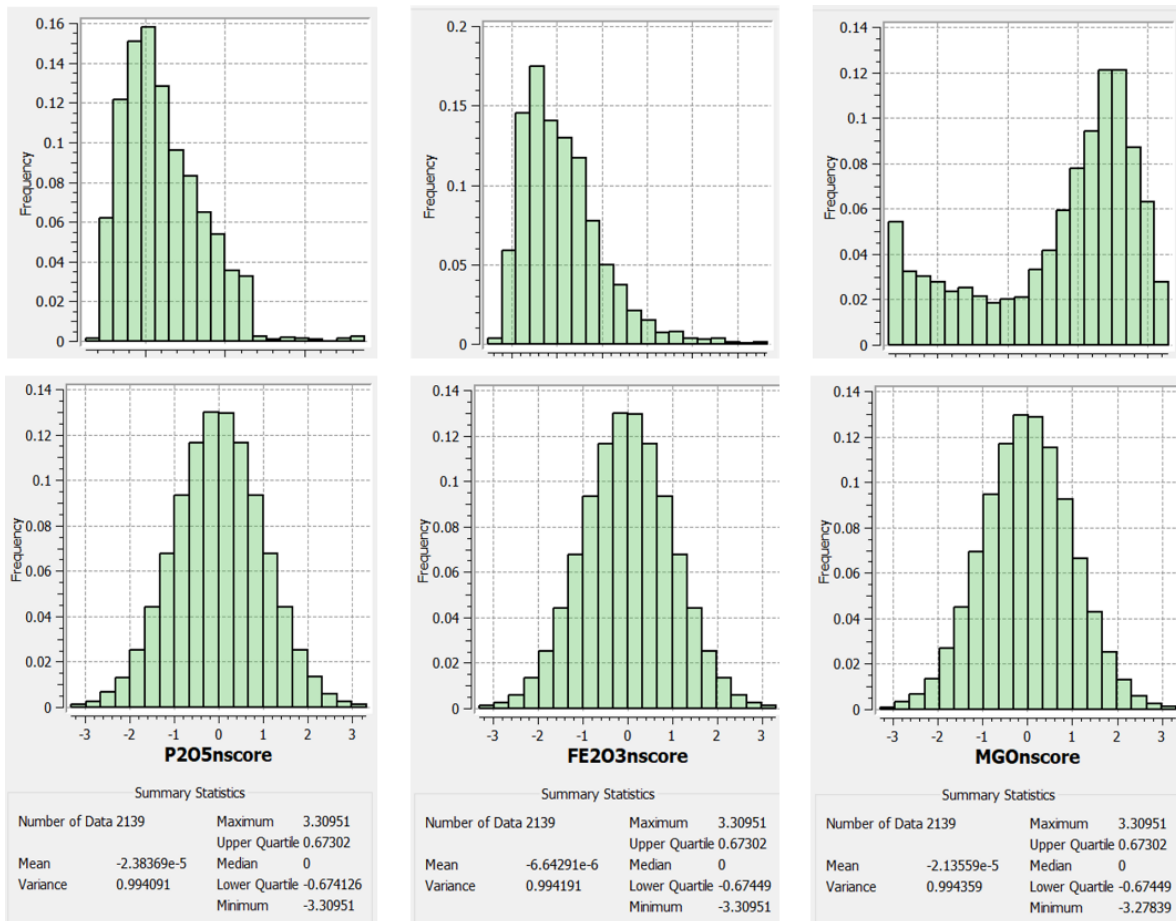


Figura 72. Histogramas dos teores P₂O₅, Fe₂O₃ e MgO na parte superior e dos teores normalizados para o FLPI no micáceo (na parte inferior da figura).

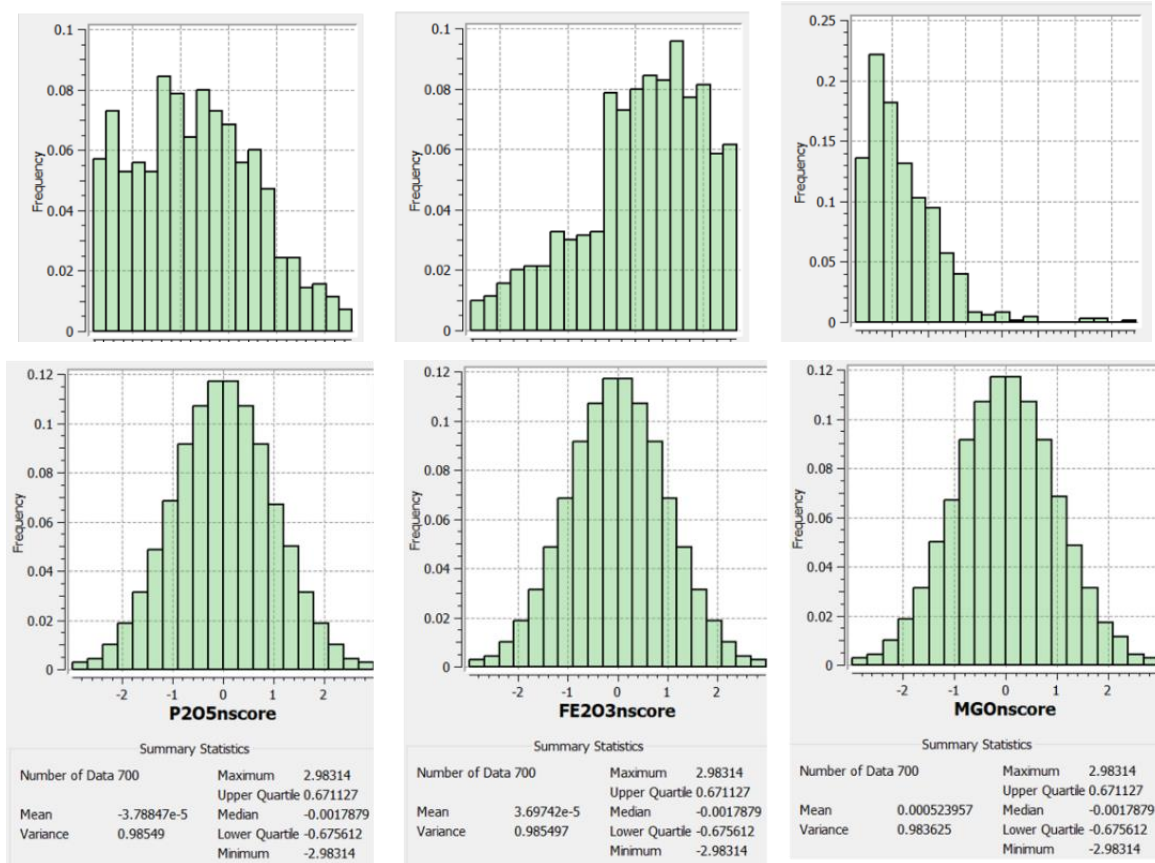


Figura 73. Histogramas dos teores P₂O₅, Fe₂O₃ e MgO na parte superior e dos teores normalizados para o CBMS no oxidado (na parte inferior da figura).

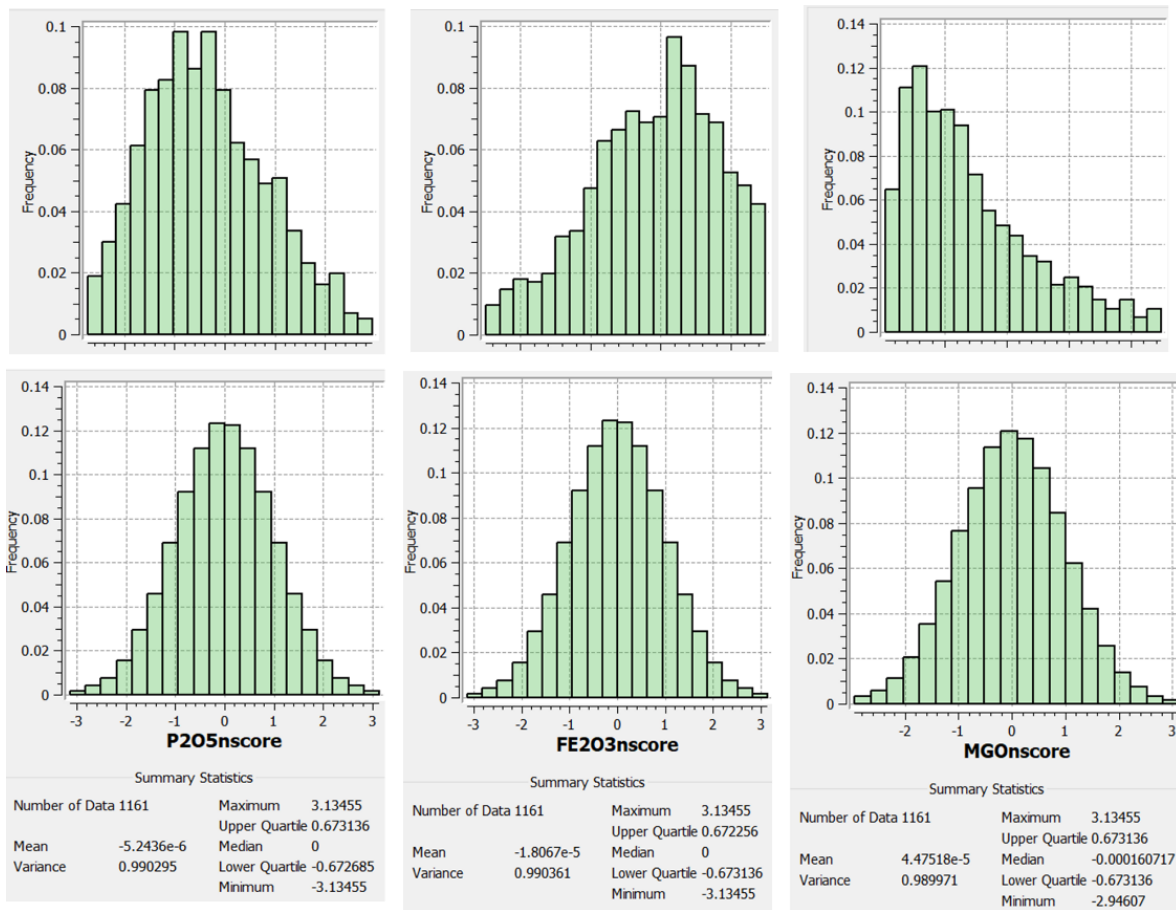


Figura 74. Histogramas dos teores P₂O₅, Fe₂O₃ e MgO na parte superior e dos teores normalizados para o CBMS no micáceo (na parte inferior da figura).

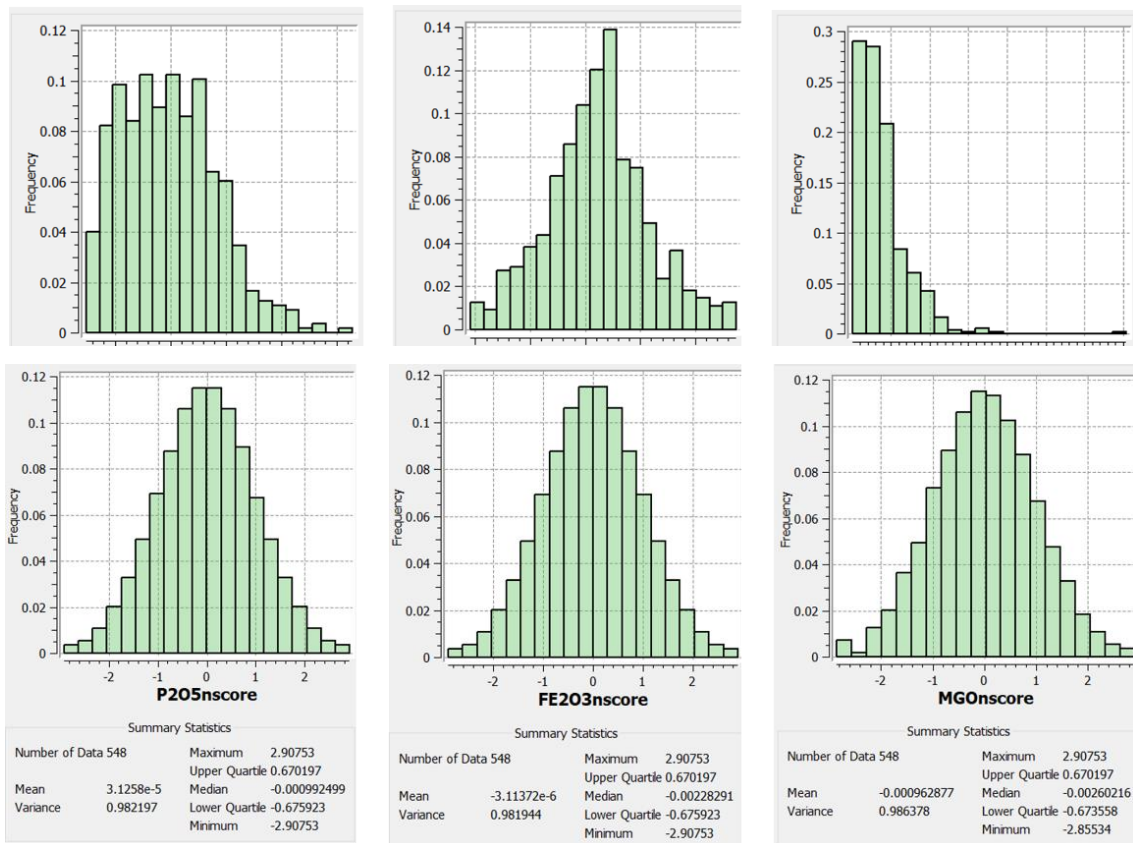


Figura 75. Histogramas dos teores P₂O₅, Fe₂O₃ e MgO na parte superior e dos teores normalizados para o CBMG no oxidado (na parte inferior da figura).

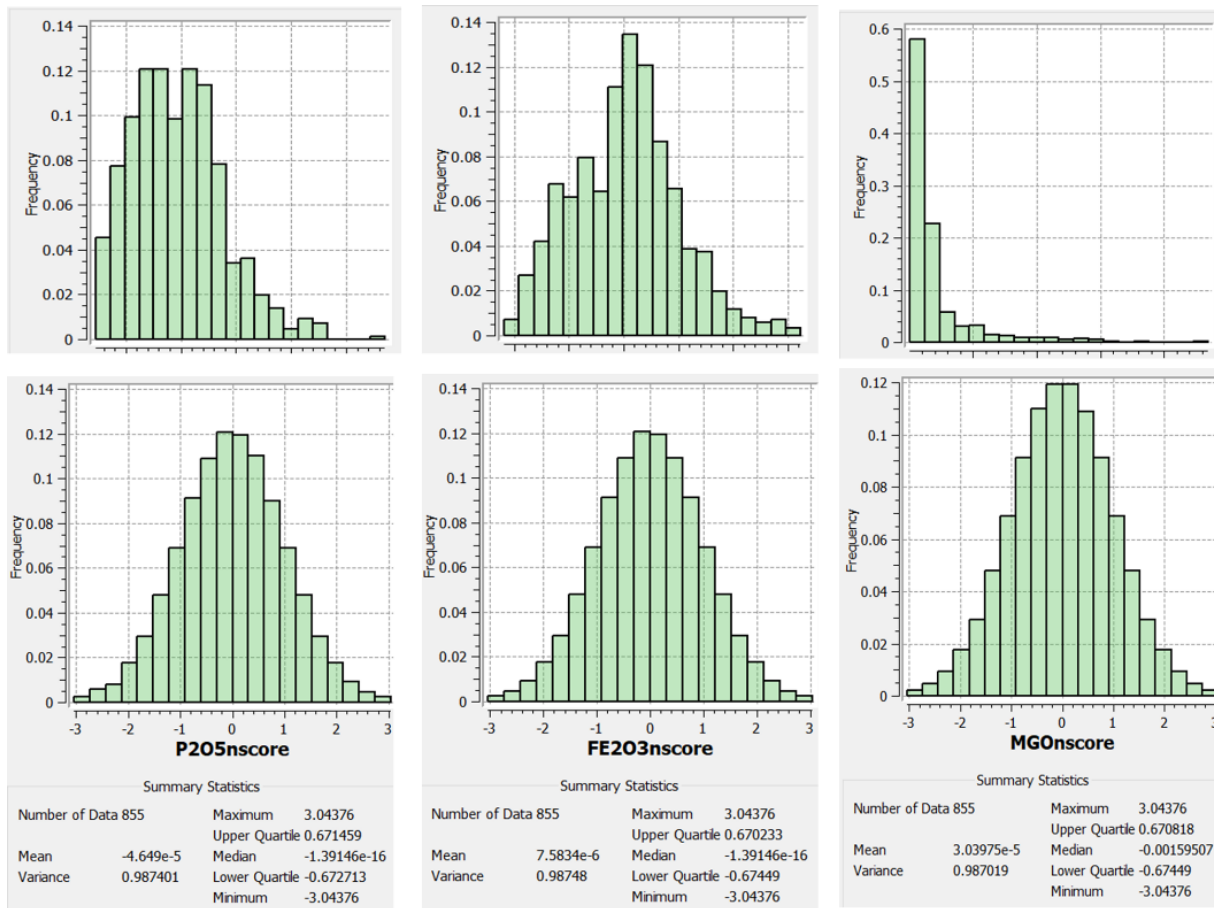


Figura 76. Histogramas dos teores na parte superior e normalizados para o CBMG no micáceo (na parte inferior da figura).

É importante destacar que além da normalidade da distribuição univariada, é necessário checar a normalidade de distribuições multivariadas para a utilização do modelo multigaussiano. Se a relação madograma e a raiz quadrada do variograma for próxima de 0,564 (Matheron, 1982) o formalismo multigaussiano pode ser assumido como mostra a Tabela 19.

Tabela 19. Relação madograma pela raiz do variograma das variáveis normalizadas.

P205			Fe203			MGO		
Número de pares	Distância (m)	Madograma/Raiz do variograma	Número de Pares	Distância média(m)	Madograma/Raiz do variograma	Número de pares	Distância (m)	Madograma/Raiz do variograma
21250	75	0.540	21250	75	0.549	21250	75	0.551
40946	150	0.548	40946	150	0.555	40946	150	0.556
58121	225	0.554	58121	225	0.560	58121	225	0.554
73379	300	0.554	73379	300	0.557	73379	300	0.562
81540	375	0.558	81540	375	0.560	81540	375	0.560
96273	450	0.556	96273	450	0.562	96273	450	0.565
101600	525	0.557	101600	525	0.559	101600	525	0.564
98944	600	0.558	98944	600	0.559	98944	600	0.563
93756	675	0.561	93756	675	0.561	93756	675	0.557
95509	750	0.558	95509	750	0.562	95509	750	0.559
95876	825	0.558	95876	825	0.565	95876	825	0.563
94508	900	0.557	94508	900	0.565	94508	900	0.567
84857	975	0.555	84857	975	0.564	84857	975	0.566
75103	1050	0.559	75103	1050	0.561	75103	1050	0.559
70788	1125	0.556	70788	1125	0.565	70788	1125	0.564

Modelagem dos variogramas

De modo geral, bons modelos de continuidade espacial foram obtidos. Os variogramas verticais do P_2O_5 no domínio FO oxidado, Fe_2O_3 no FLPI micáceo e MgO no CBMG micáceo não apresentaram boa estruturação e, por essa razão, os variogramas omnidirecionais foram preferidos. Essa dificuldade de obter bons variogramas pode estar relacionada ao número de amostras não expressivo, dada a subdivisão em domínios litológicos e em horizonte de minério.

As Figuras 77 a 84 mostram os variogramas experimentais dos dados. Os modelos adotados são esféricos e as direções de continuidade principal, intermediária e vertical são identificadas como h_1 , h_2 e h_3 para os teores de Fe_2O_3 , P_2O_5 e MgO representadas nas equações 67 a 78 para a camada oxidado e nas equações 79 a 90 para o micáceo.

$$\gamma_{P_2O_5_FO}(h) = 0,10 + \left[0,45.Sph_1 \left[\frac{h_1}{70}, \frac{h_2}{70}, \frac{h_3}{17} \right] + 0,45.Sph_2 \left[\frac{h_1}{170}, \frac{h_2}{170}, \frac{h_3}{46,5} \right] \right] \quad (67)$$

$$\gamma_{Fe_2O_3_FO}(h) = 0,10 + \left[0,40.Sph_1 \left[\frac{h_1}{75}, \frac{h_2}{75}, \frac{h_3}{30} \right] + 0,50.Sph_2 \left[\frac{h_1}{505}, \frac{h_2}{480}, \frac{h_3}{60} \right] \right] \quad (68)$$

$$\gamma_{MgO_FO}(h) = 0,05 + \left[0,45.Sph_1 \left[\frac{h_1}{78}, \frac{h_2}{60}, \frac{h_3}{20} \right] + 0,50.Sph_2 \left[\frac{h_1}{350}, \frac{h_2}{150}, \frac{h_3}{58} \right] \right] \quad (69)$$

$$\gamma_{P_2O_5_FLPI}(h) = 0,10 + \left[0,50.Sph_1 \left[\frac{h_1}{200}, \frac{h_2}{150}, \frac{h_3}{35} \right] + 0,40.Sph_2 \left[\frac{h_1}{400}, \frac{h_2}{300}, \frac{h_3}{65} \right] \right] \quad (70)$$

$$\gamma_{Fe_2O_3_FLPI}(h) = 0,05 + \left[0,45.Sph_1 \left[\frac{h_1}{400}, \frac{h_2}{146}, \frac{h_3}{30} \right] + 0,50.Sph_2 \left[\frac{h_1}{750}, \frac{h_2}{270}, \frac{h_3}{55} \right] \right] \quad (71)$$

$$\gamma_{MgO_FLPI}(h) = 0,05 + \left[0,5.Sph_1 \left[\frac{h_1}{370}, \frac{h_2}{105}, \frac{h_3}{22} \right] + 0,45.Sph_2 \left[\frac{h_1}{400}, \frac{h_2}{360}, \frac{h_3}{50} \right] \right] \quad (72)$$

$$\gamma_{P_2O_5_CBMS}(h) = 0,05 + \left[0,3.Sph_1 \left[\frac{h_1}{225}, \frac{h_2}{225}, \frac{h_3}{45} \right] + 0,65.Sph_2 \left[\frac{h_1}{490}, \frac{h_2}{490}, \frac{h_3}{100} \right] \right] \quad (73)$$

$$\gamma_{Fe_2O_3_CBMS}(h) = 0,1 + \left[0,50.Sph_1 \left[\frac{h_1}{120}, \frac{h_2}{90}, \frac{h_3}{22} \right] + 0,4.Sph_2 \left[\frac{h_1}{460}, \frac{h_2}{225}, \frac{h_3}{45} \right] \right] \quad (74)$$

$$\gamma_{MgO_CBMS}(h) = 0,05 + \left[0,50.Sph_1 \left[\frac{h_1}{220}, \frac{h_2}{180}, \frac{h_3}{41} \right] + 0,45.Sph_2 \left[\frac{h_1}{980}, \frac{h_2}{585}, \frac{h_3}{85} \right] \right] \quad (75)$$

$$\gamma_{P_2O_5_CBMG}(h) = 0,10 + \left[0,45.Sph_1 \left[\frac{h_1}{78}, \frac{h_2}{78}, \frac{h_3}{30} \right] + 0,45.Sph_2 \left[\frac{h_1}{304}, \frac{h_2}{304}, \frac{h_3}{60} \right] \right] \quad (76)$$

$$\gamma_{Fe_2O_3_CBMG}(h) = 0,05 + \left[0,45.Sph_1 \left[\frac{h_1}{90}, \frac{h_2}{90}, \frac{h_3}{2} \right] + 0,50.Sph_2 \left[\frac{h_1}{500}, \frac{h_2}{340}, \frac{h_3}{45} \right] \right] \quad (77)$$

$$\gamma_{MgO_CBMG}(h) = 0,05 + \left[0,50.Sph_1 \left[\frac{h_1}{315}, \frac{h_2}{160}, \frac{h_3}{55} \right] + 0,45.Sph_2 \left[\frac{h_1}{933}, \frac{h_2}{510}, \frac{h_3}{80} \right] \right] \quad (78)$$

$$\gamma_{P_2O_5_FO}(h) = 0,1 + \left[0,40.Sph_1 \left[\frac{h_1}{90}, \frac{h_2}{90}, \frac{h_3}{15} \right] + 0,5.Sph_2 \left[\frac{h_1}{260}, \frac{h_2}{200}, \frac{h_3}{65} \right] \right] \quad (79)$$

$$\gamma_{Fe_2O_3_FO}(h) = 0,10 + \left[0,40.Sph_1 \left[\frac{h_1}{202}, \frac{h_2}{170}, \frac{h_3}{40} \right] + 0,50.Sph_2 \left[\frac{h_1}{1125}, \frac{h_2}{1000}, \frac{h_3}{120} \right] \right] \quad (80)$$

$$\gamma_{MgO_FO}(h) = 0,05 + \left[0,45.Sph_1 \left[\frac{h_1}{90}, \frac{h_2}{90}, \frac{h_3}{25} \right] + 0,50.Sph_2 \left[\frac{h_1}{350}, \frac{h_2}{270}, \frac{h_3}{55} \right] \right] \quad (81)$$

$$\gamma_{P2O5_FLPI}(h) = 0,10 + \left[0,45.Sph_1 \left[\frac{h_1}{145}, \frac{h_2}{130}, \frac{h_3}{20} \right] + 0,45.Sph_2 \left[\frac{h_1}{415}, \frac{h_2}{247}, \frac{h_3}{60} \right] \right] \quad (82)$$

$$\gamma_{FE2O3_FLPI}(h) = 0,05 + \left[0,45.Sph_1 \left[\frac{h_1}{180}, \frac{h_2}{150}, \frac{h_3}{35} \right] + 0,50.Sph_2 \left[\frac{h_1}{440}, \frac{h_2}{350}, \frac{h_3}{100} \right] \right] \quad (83)$$

$$\gamma_{MGO_FLPI}(h) = 0,075 + \left[0,46.Sph_1 \left[\frac{h_1}{135}, \frac{h_2}{112}, \frac{h_3}{30} \right] + 0,465.Sph_2 \left[\frac{h_1}{450}, \frac{h_2}{315}, \frac{h_3}{100} \right] \right] \quad (84)$$

$$\gamma_{P2O5_CBMS}(h) = 0,05 + \left[0,45.Sph_1 \left[\frac{h_1}{415}, \frac{h_2}{270}, \frac{h_3}{55} \right] + 0,5.Sph_2 \left[\frac{h_1}{810}, \frac{h_2}{750}, \frac{h_3}{98,5} \right] \right] \quad (85)$$

$$\gamma_{FE2O3_CBMS}(h) = 0,10 + \left[0,5.Sph_1 \left[\frac{h_1}{50}, \frac{h_2}{50}, \frac{h_3}{16,95} \right] + 0,40.Sph_2 \left[\frac{h_1}{240}, \frac{h_2}{240}, \frac{h_3}{50} \right] \right] \quad (86)$$

$$\gamma_{MGO_CBMS}(h) = 0,1 + \left[0,50.Sph_1 \left[\frac{h_1}{101}, \frac{h_2}{60}, \frac{h_3}{25} \right] + 0,4.Sph_2 \left[\frac{h_1}{420}, \frac{h_2}{240}, \frac{h_3}{75} \right] \right] \quad (87)$$

$$\gamma_{P2O5_CBMG}(h) = 0,05 + \left[0,45.Sph_1 \left[\frac{h_1}{80}, \frac{h_2}{80}, \frac{h_3}{17} \right] + 0,5.Sph_2 \left[\frac{h_1}{250}, \frac{h_2}{250}, \frac{h_3}{60} \right] \right] \quad (88)$$

$$\gamma_{Fe2O3_CBMG}(h) = 0,1 + \left[0,40.Sph_1 \left[\frac{h_1}{150}, \frac{h_2}{150}, \frac{h_3}{35} \right] + 0,5.Sph_2 \left[\frac{h_1}{500}, \frac{h_2}{450}, \frac{h_3}{58} \right] \right] \quad (89)$$

$$\gamma_{MGO_CBMG}(h) = 0,05 + \left[0,50.Sph_1 \left[\frac{h_1}{120}, \frac{h_2}{80}, \frac{h_3}{28} \right] + 0,45.Sph_2 \left[\frac{h_1}{280}, \frac{h_2}{250}, \frac{h_3}{73} \right] \right] \quad (90)$$

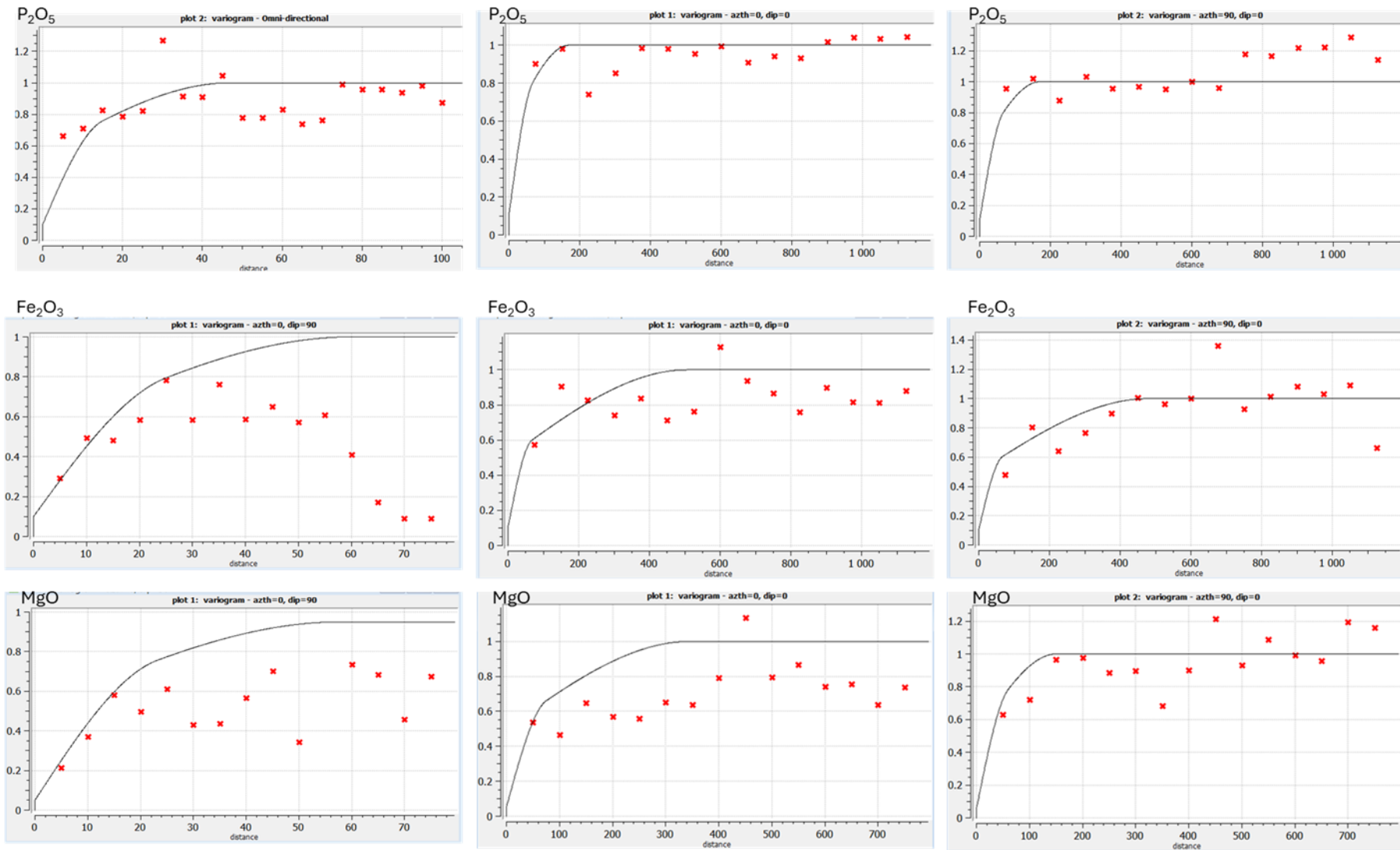


Figura 77. Variogramas dos dados na vertical (à esquerda), direção 0°(centro) e 90° (à direita) FO oxidado.

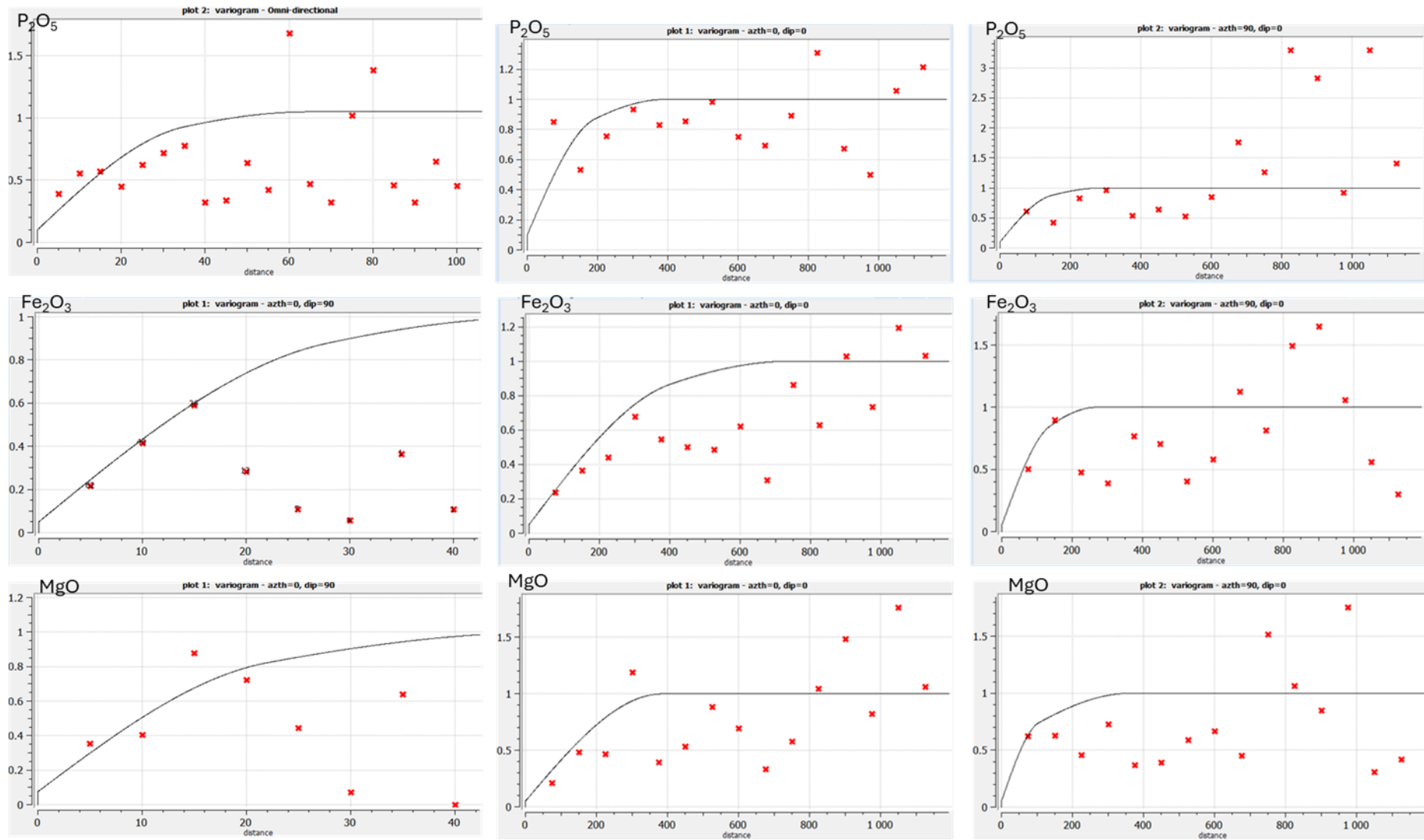


Figura 78. Variogramas dos dados na vertical (à esquerda), direção 0° (centro) e 90° (à direita) FLPI oxidado.

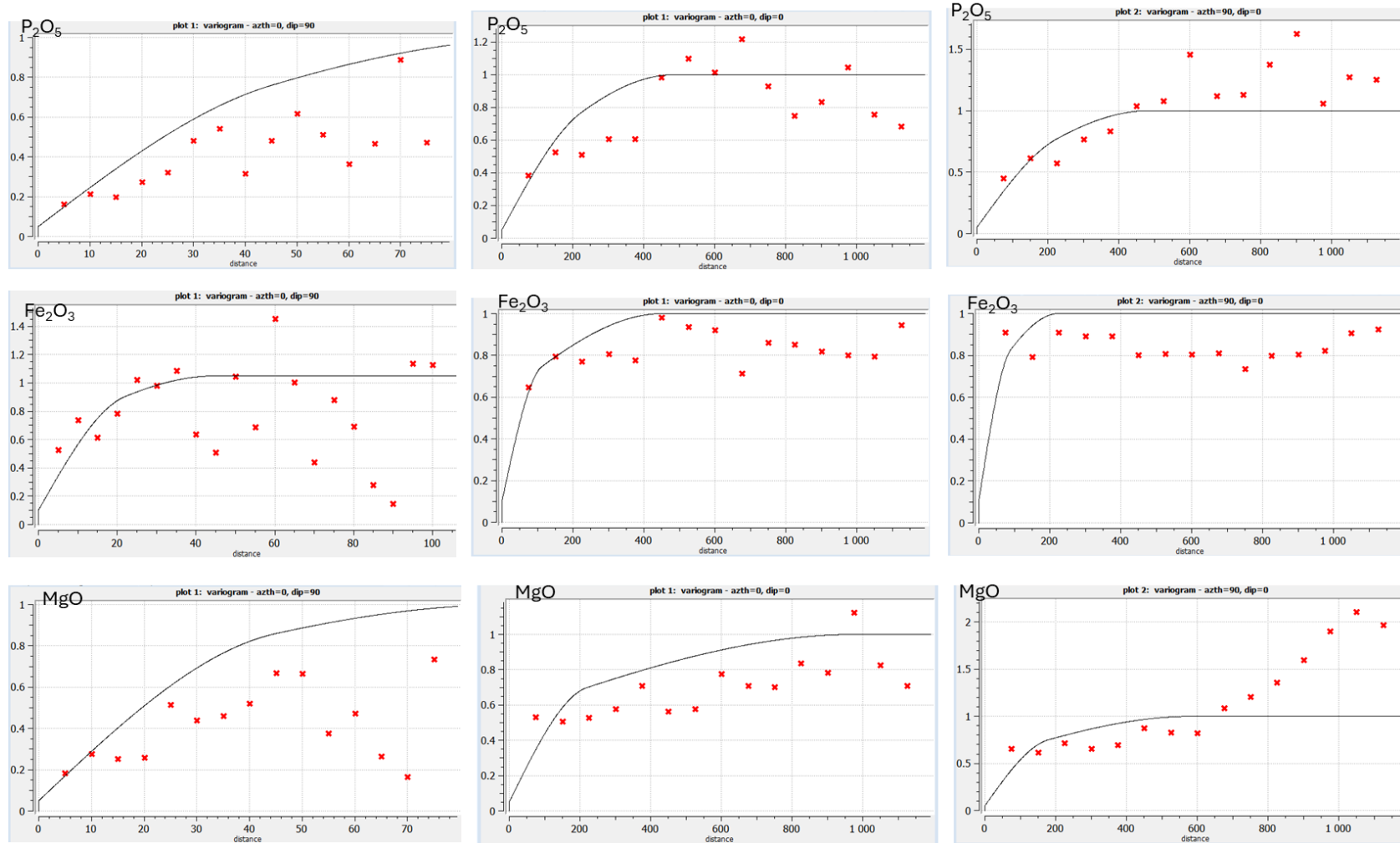


Figura 79. Variogramas dos dados na vertical (à esquerda), direção 0° (centro) e 90° (à direita) CBMS oxidado.

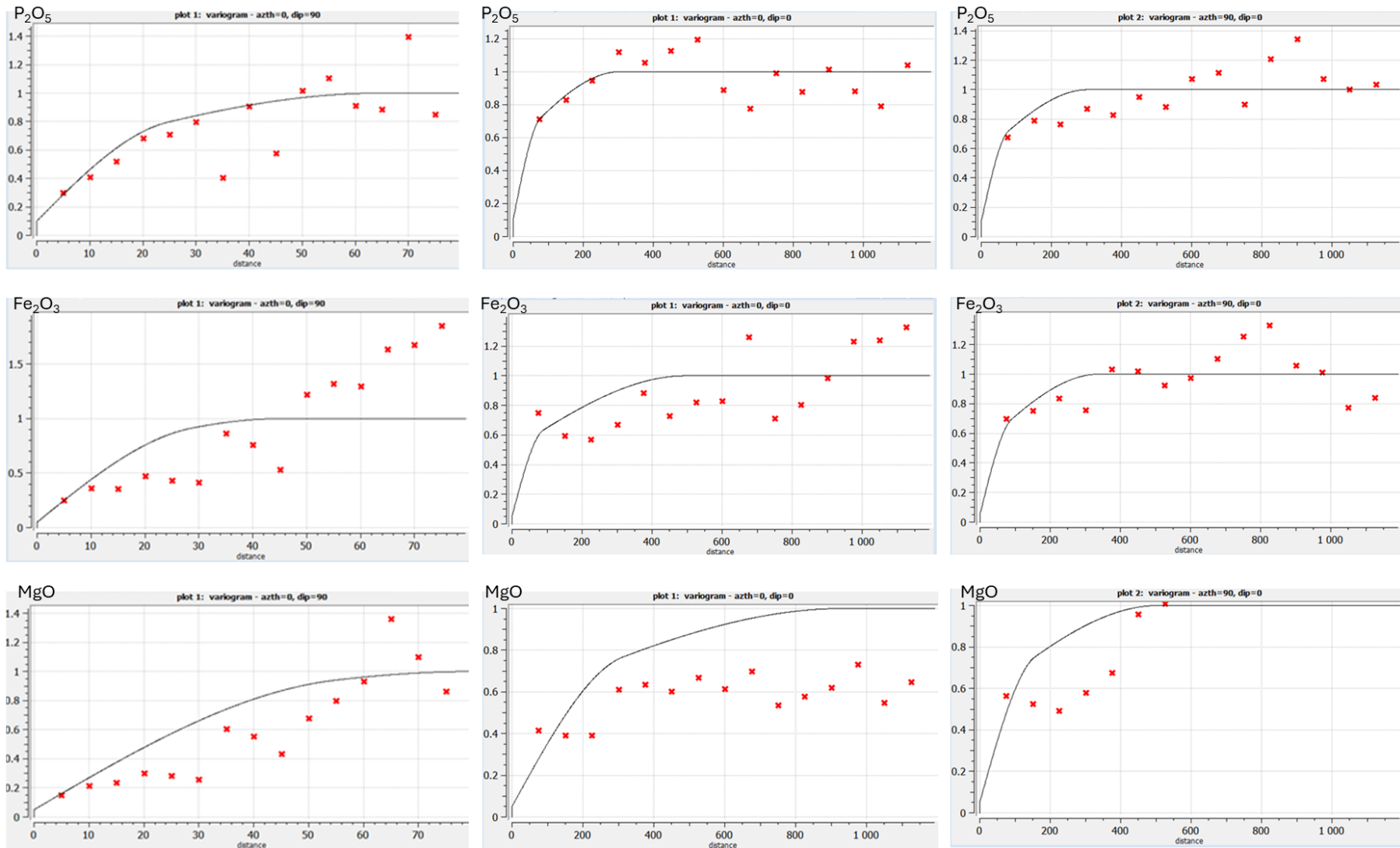


Figura 80. Variogramas dos dados na vertical (à esquerda), direção 0° (centro) e 90° (à direita) CBMG oxidado.

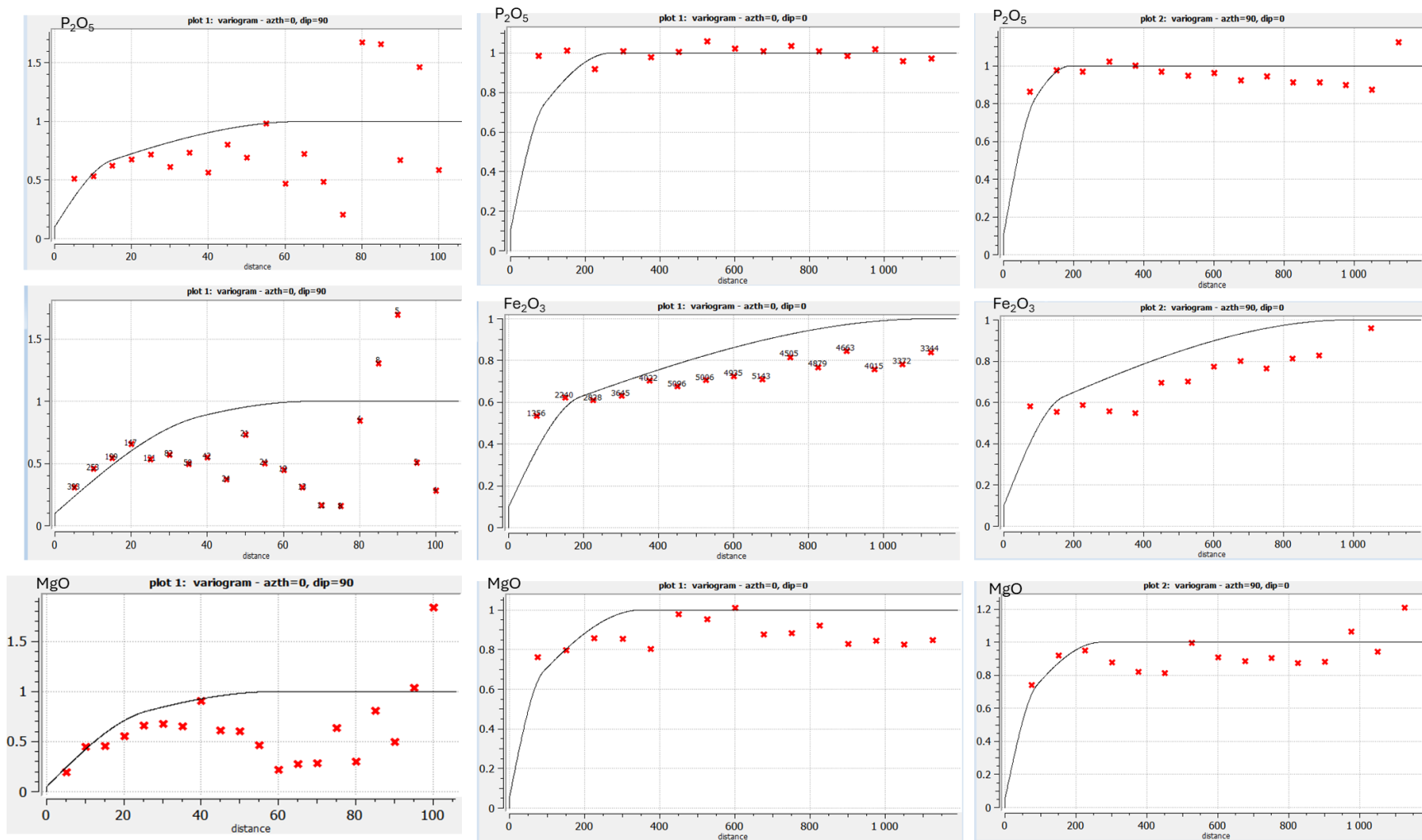


Figura 81. Variogramas dos dados na vertical (à esquerda), direção 0°(centro) e 90° (à direita) FO micáceo.

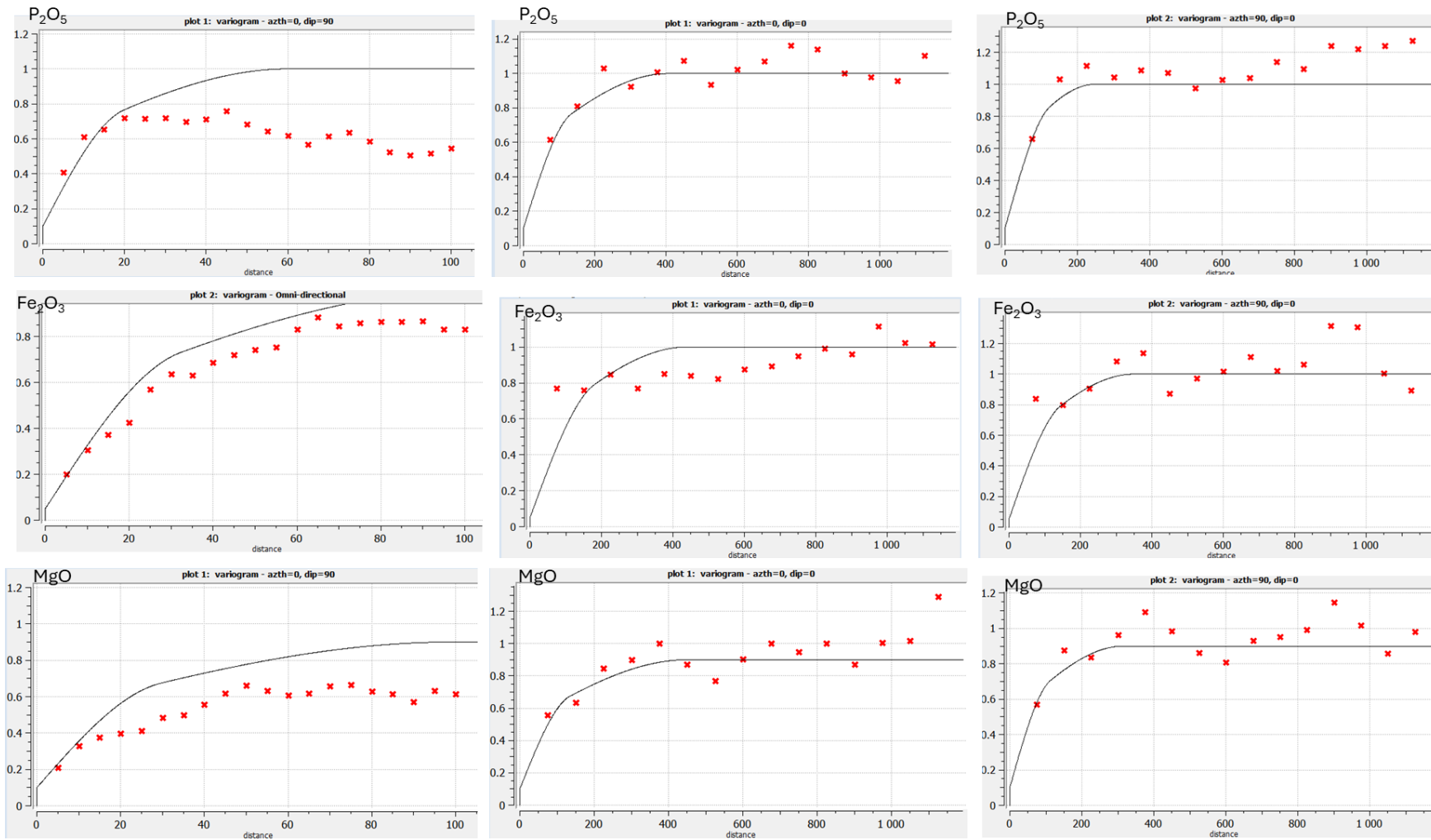


Figura 82. Variogramas dos dados na vertical (à esquerda), direção 0° (centro) e 90° (à direita) FLPI micáceo.

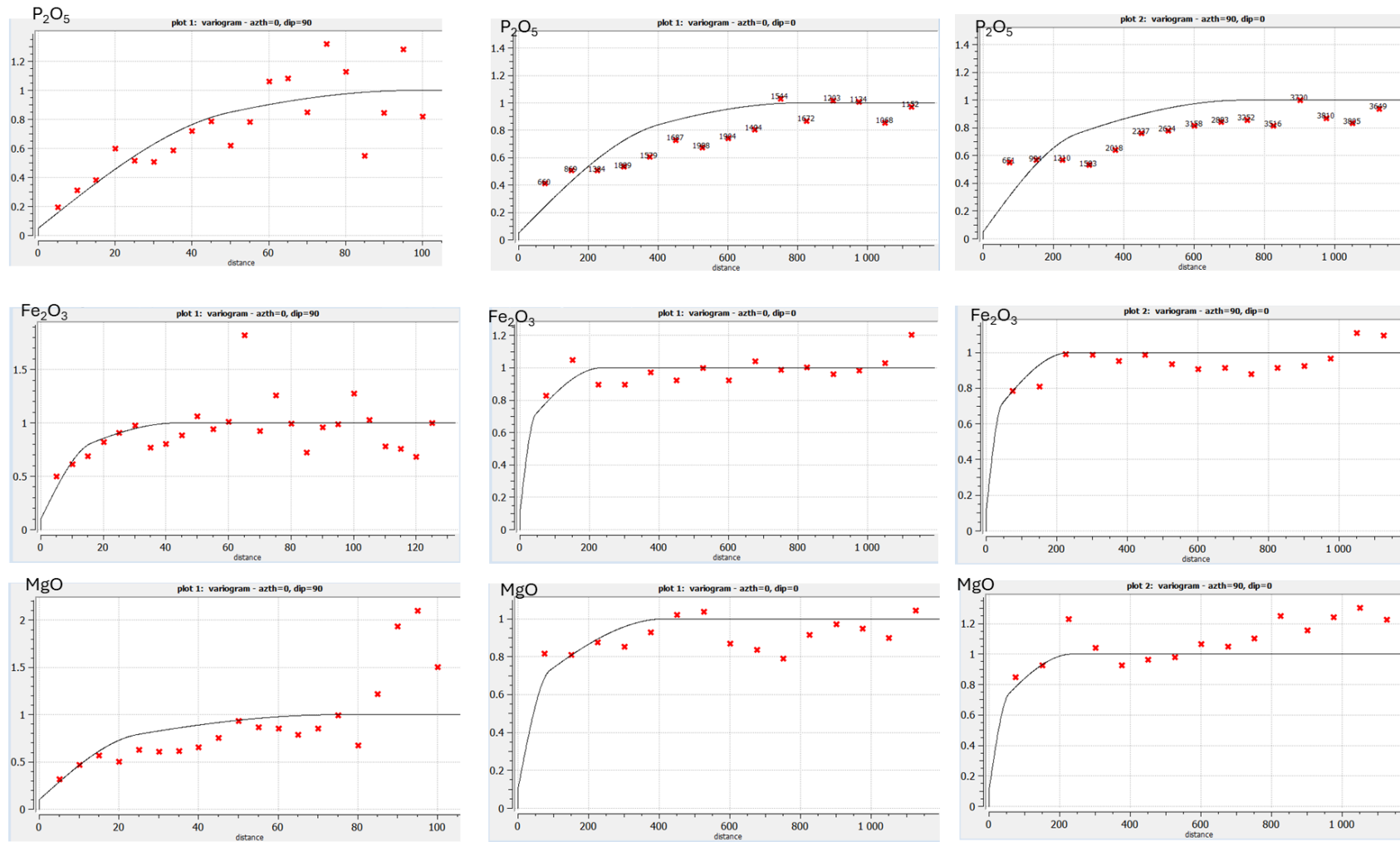


Figura 83. Variogramas dos dados na vertical (à esquerda), direção 0°(centro) e 90° (à direita) CBMS micáceo.

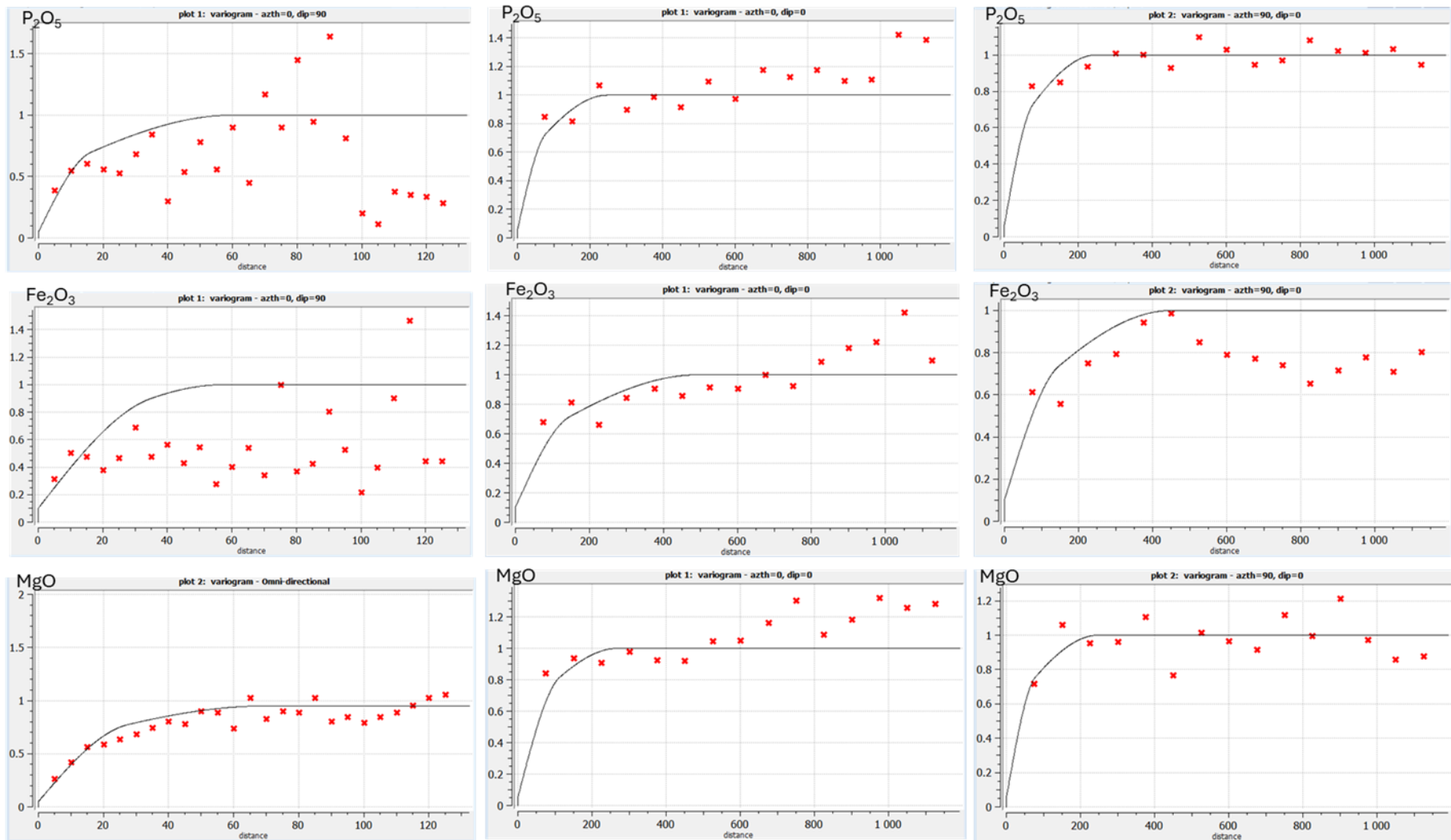


Figura 84. Variogramas dos dados na vertical (à esquerda), direção 0° (centro) e 90° (à direita) CBMG micáceo.

Validação das simulações e predição nos pontos simulados

A validação das simulações é uma fase importante para atestar a qualidade da informação gerada. Essa verificação é realizada por reprodução dos histogramas e do modelo de continuidade espacial, ou seja, estacionaridade de primeira e segunda ordem. As flutuações ergódicas também merece destaque pois são utilizadas para se referir às variações nas estatísticas das realizações em relação aos dados e identificam o espaço de incertezas. As Figuras 85 a 88 apresentam a validação das estatísticas das simulações com os dados amostrais desagrupados através dos histogramas acumulados. Nas Figuras 89 a 96 é possível notar que os variogramas experimentais obtidos para cada realização honraram a continuidade espacial dos dados amostrais.

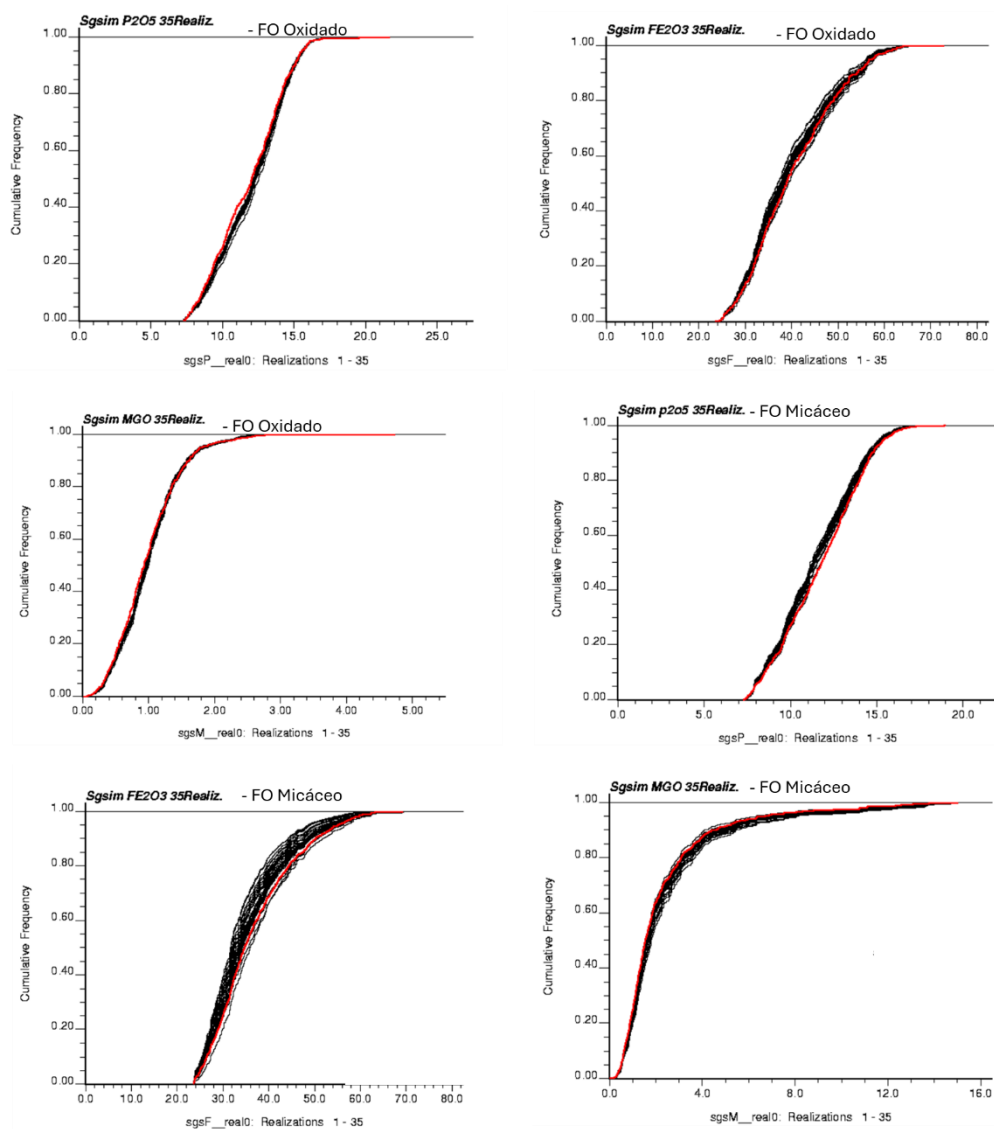


Figura 85. Validação da reprodução da distribuição dos teores de P_2O_5 , MgO e Fe_2O_3 . Em preto estão representadas as distribuições das simulações e em vermelho o a distribuição desagrupada para o FO oxidado e micáceo.

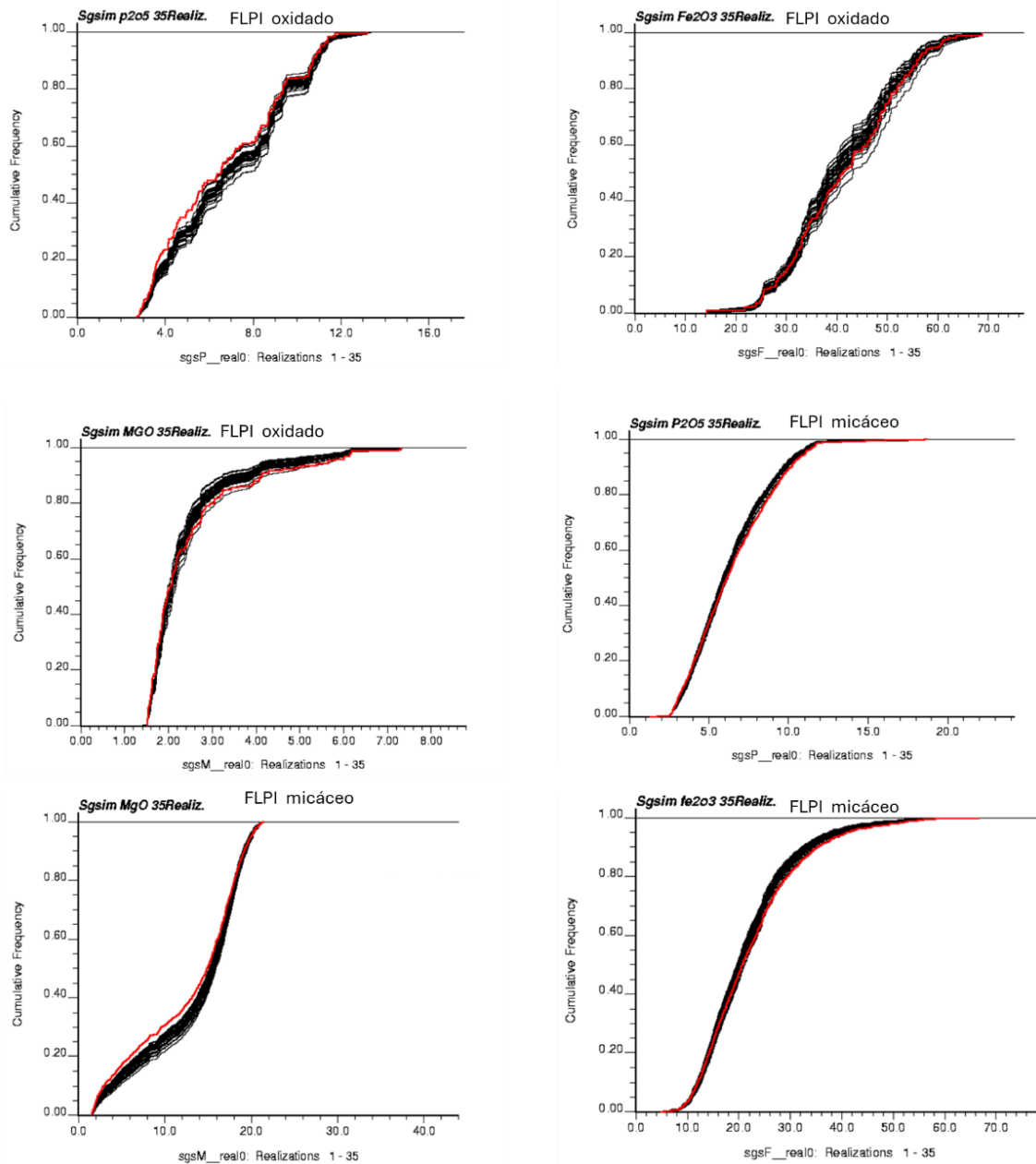


Figura 86. Validação da reprodução da distribuição dos teores de P_2O_5 , MgO e Fe_2O_3 . Em preto estão representadas as distribuições das simulações e em vermelho o a distribuição desagrupada para o FLPI oxidado e micáceo.

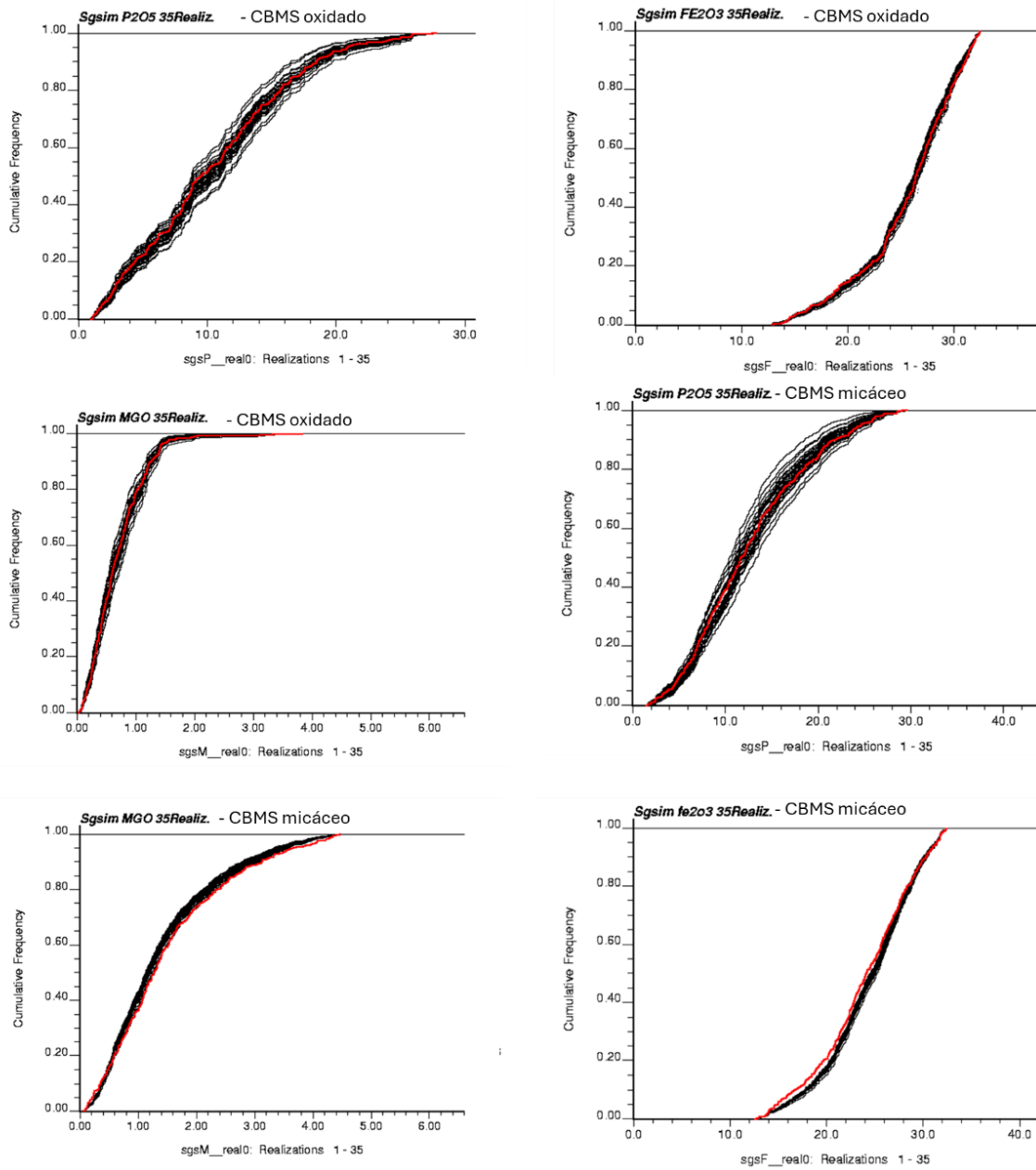


Figura 87. Validação da reprodução da distribuição dos teores de P_2O_5 , MgO e Fe_2O_3 . Em preto estão representadas as distribuições das simulações e em vermelho o a distribuição desagrupada para o CBMS oxidado e micáceo.

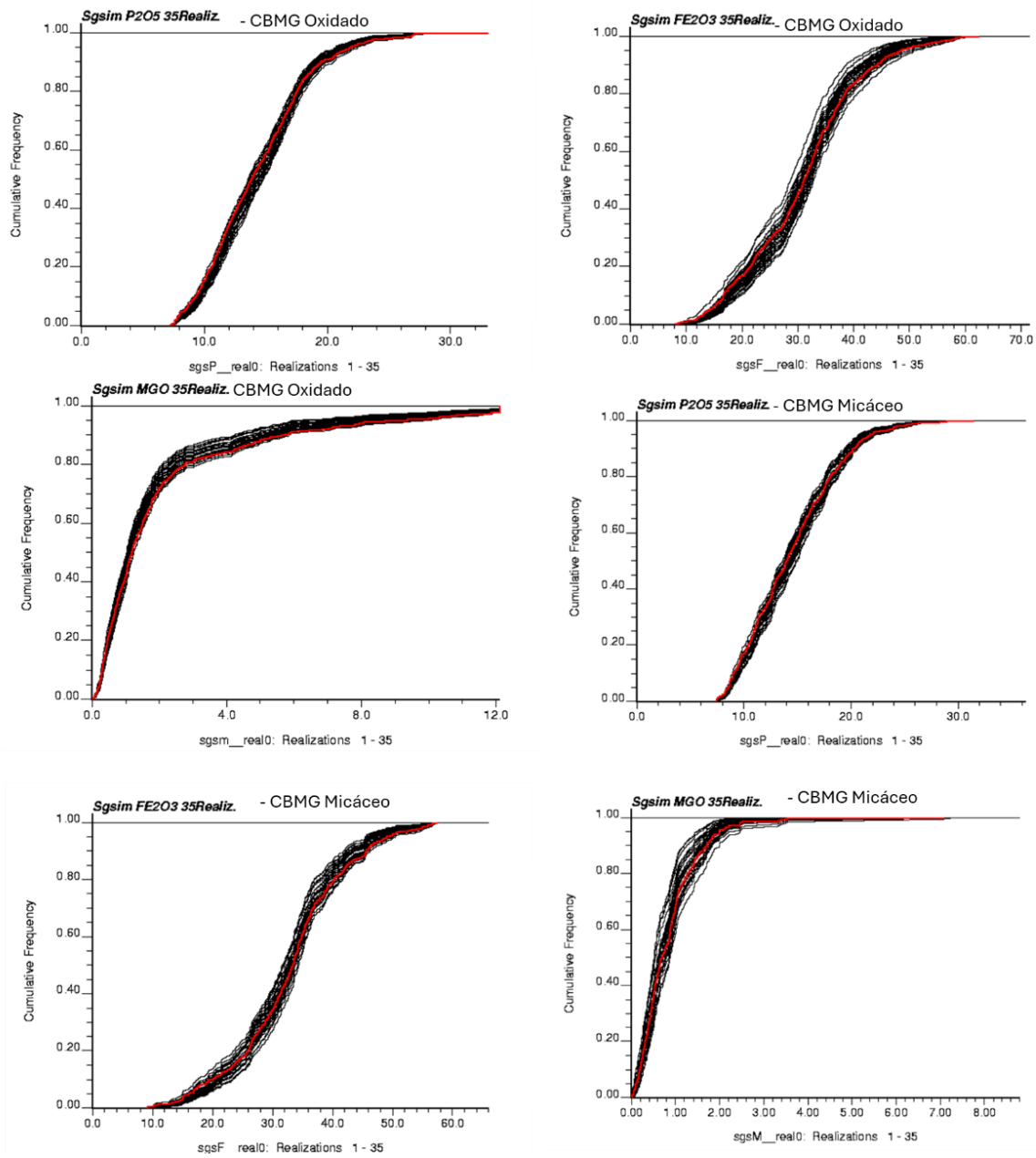


Figura 88. Validação da reprodução da distribuição dos teores de P_2O_5 , MgO e Fe_2O_3 . Em preto estão representadas as distribuições das simulações e em vermelho o a distribuição desagrupada para o CBMG oxidado e micáceo.

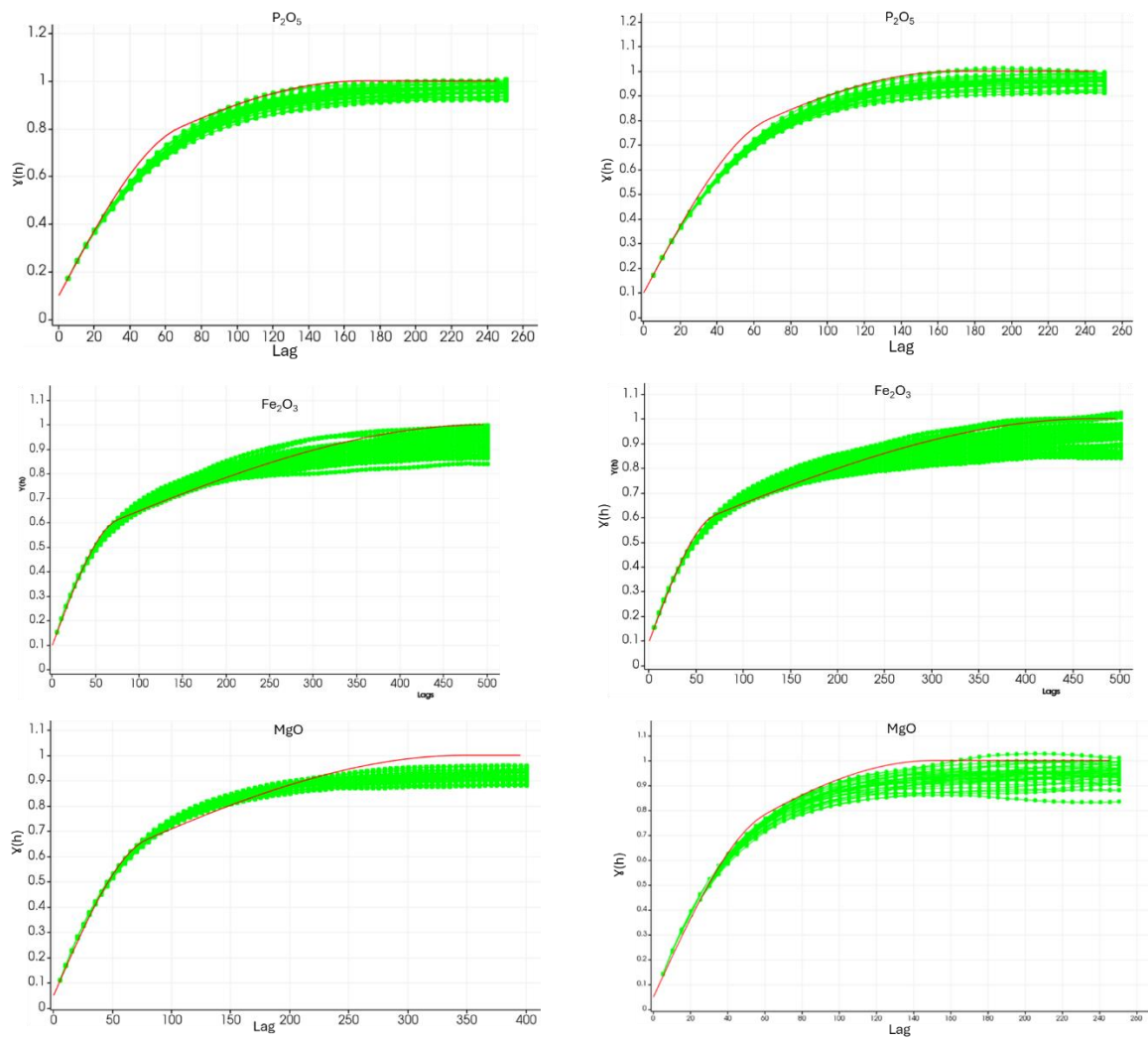


Figura 89. Validação dos variogramas da esquerda para direita maior e menor continuidade. Em verde os 35 variogramas gerados pela simulação, em vermelho o variograma dos dados originais para o FO oxidado.

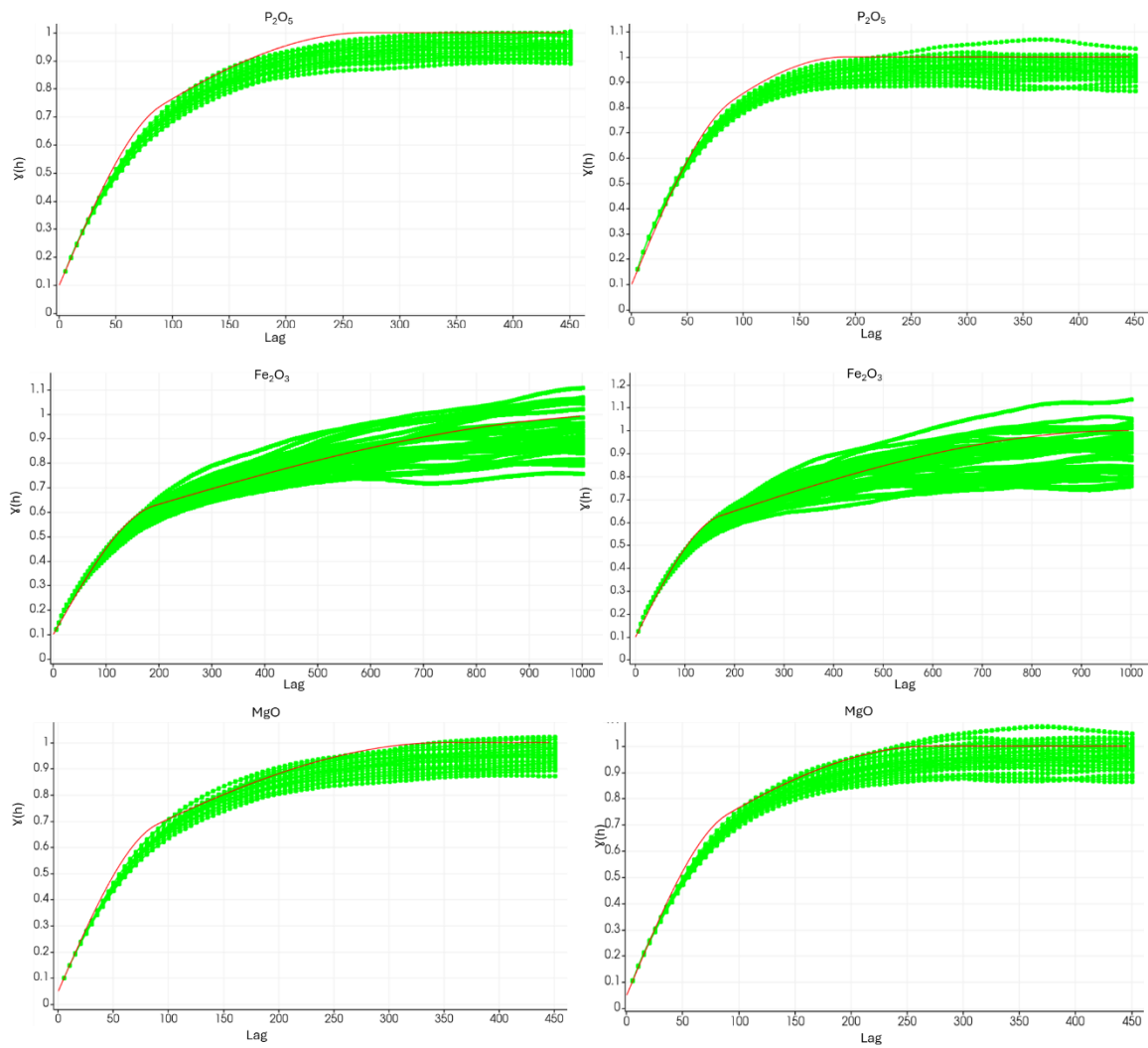


Figura 90. Validação dos variogramas da esquerda para direita maior e menor continuidade. Em verde os 35 variogramas gerados pela simulação, em vermelho o variograma dos dados originais para o FO micáceo.

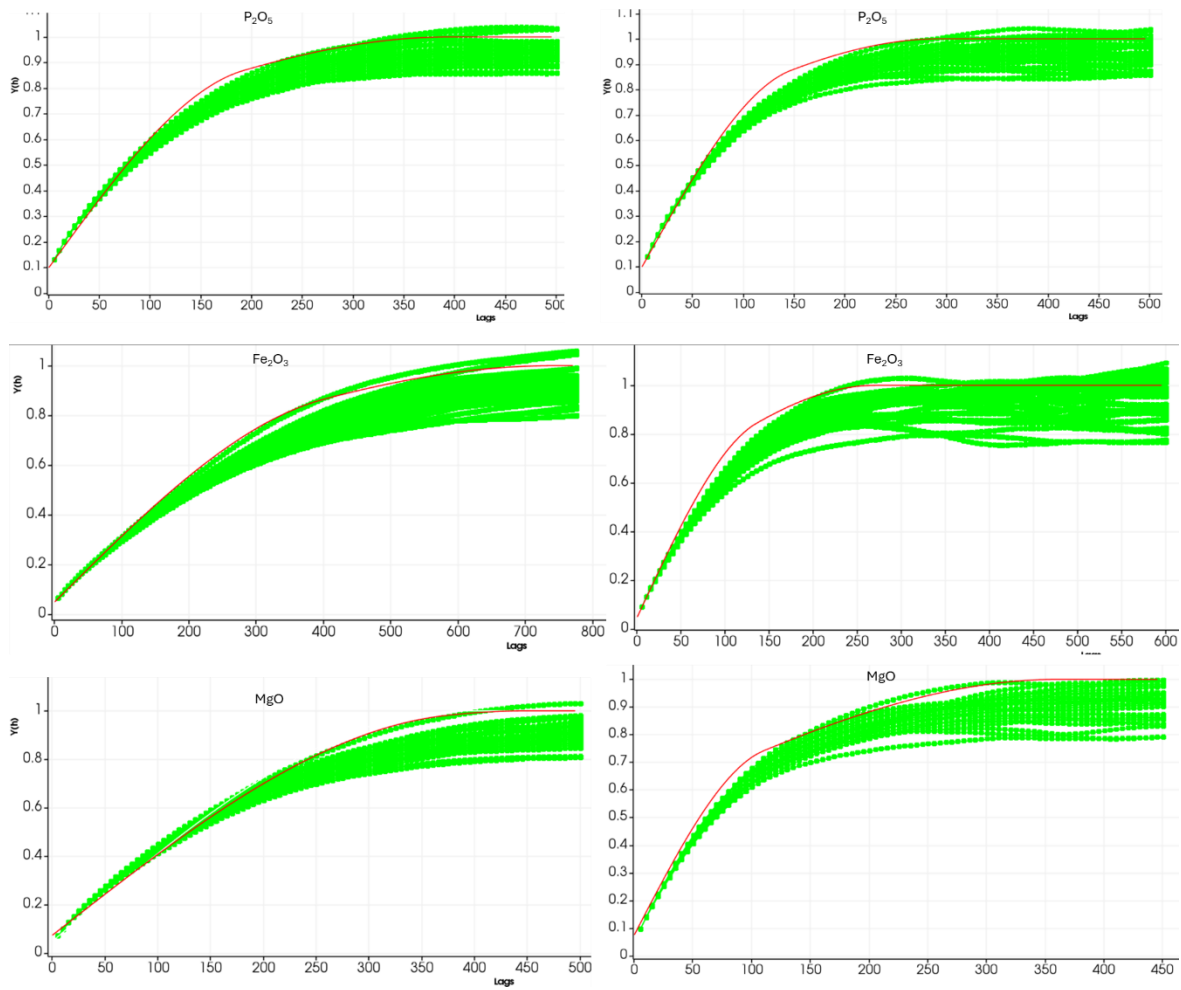


Figura 91. Validação dos variogramas da esquerda para direita maior e menor continuidade. Em verde os 35 variogramas gerados pela simulação, em vermelho o variograma dos dados originais para o FLPI oxidado.

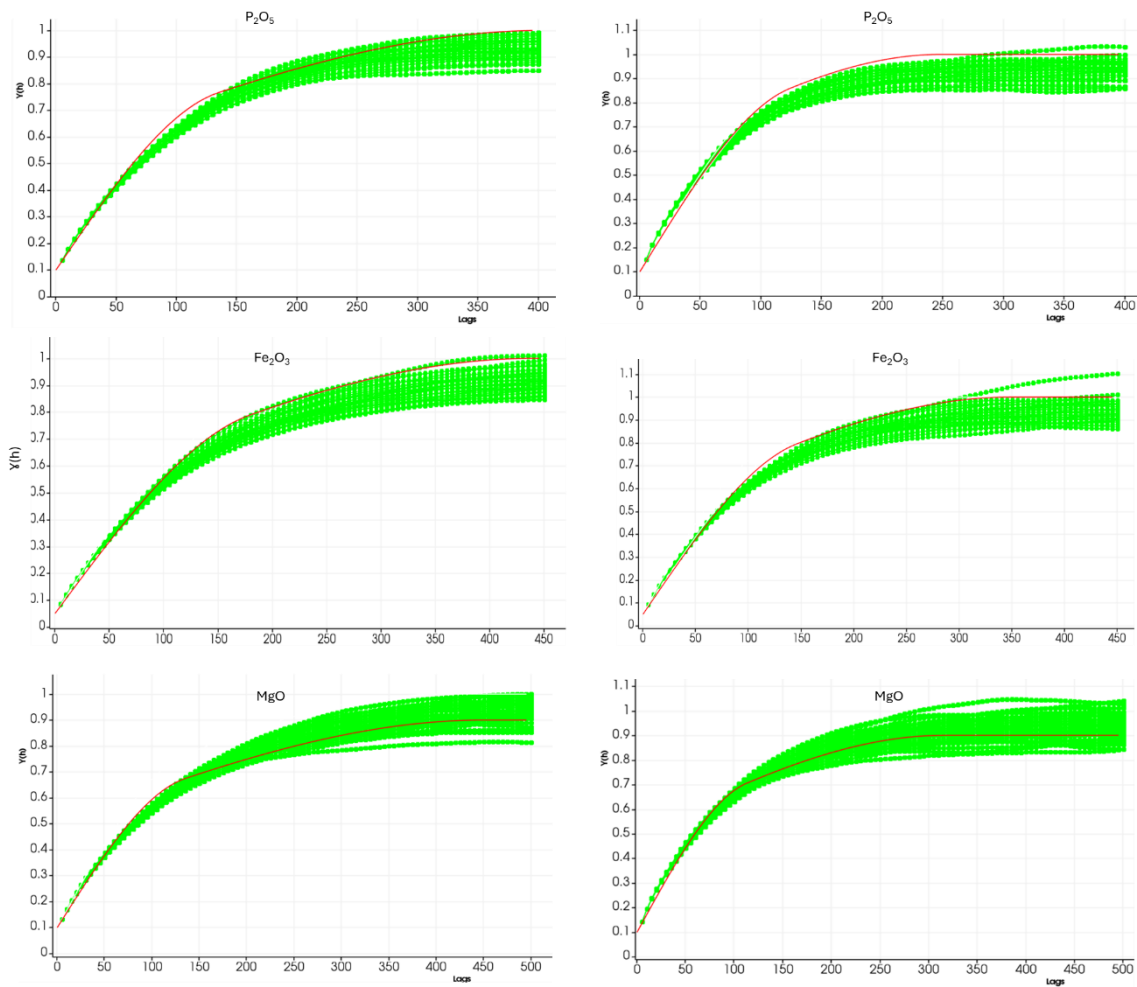


Figura 92. Validação dos variogramas da esquerda para direita maior e menor continuidade. Em verde os 35 variogramas gerados pela simulação, em vermelho o variograma dos dados originais para o FLPI micáceo.

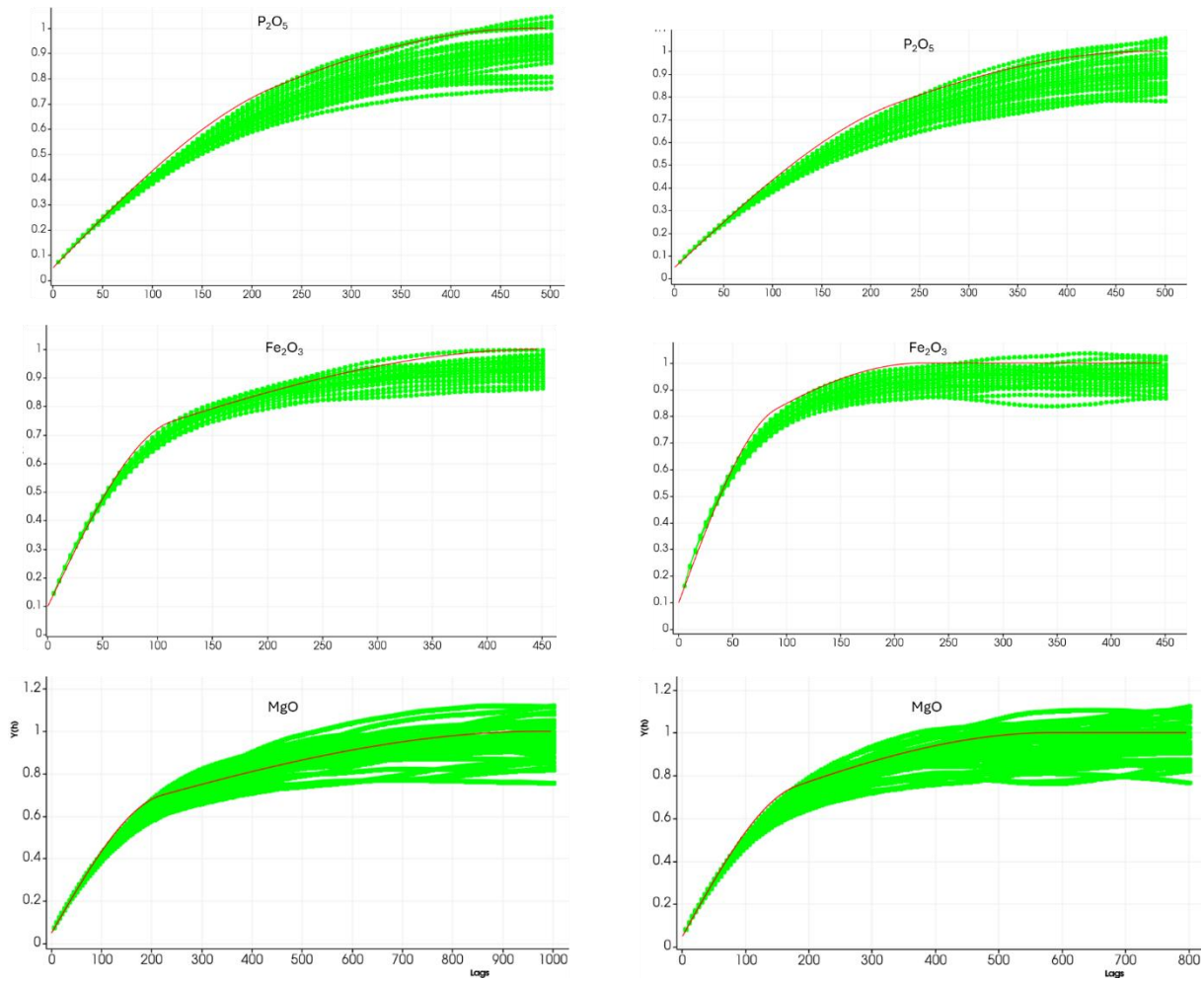


Figura 93. Validação dos variogramas da esquerda para direita maior e menor continuidade. Em verde os 35 variogramas gerados pela simulação, em vermelho o variograma dos dados originais para o CBMS oxidado.

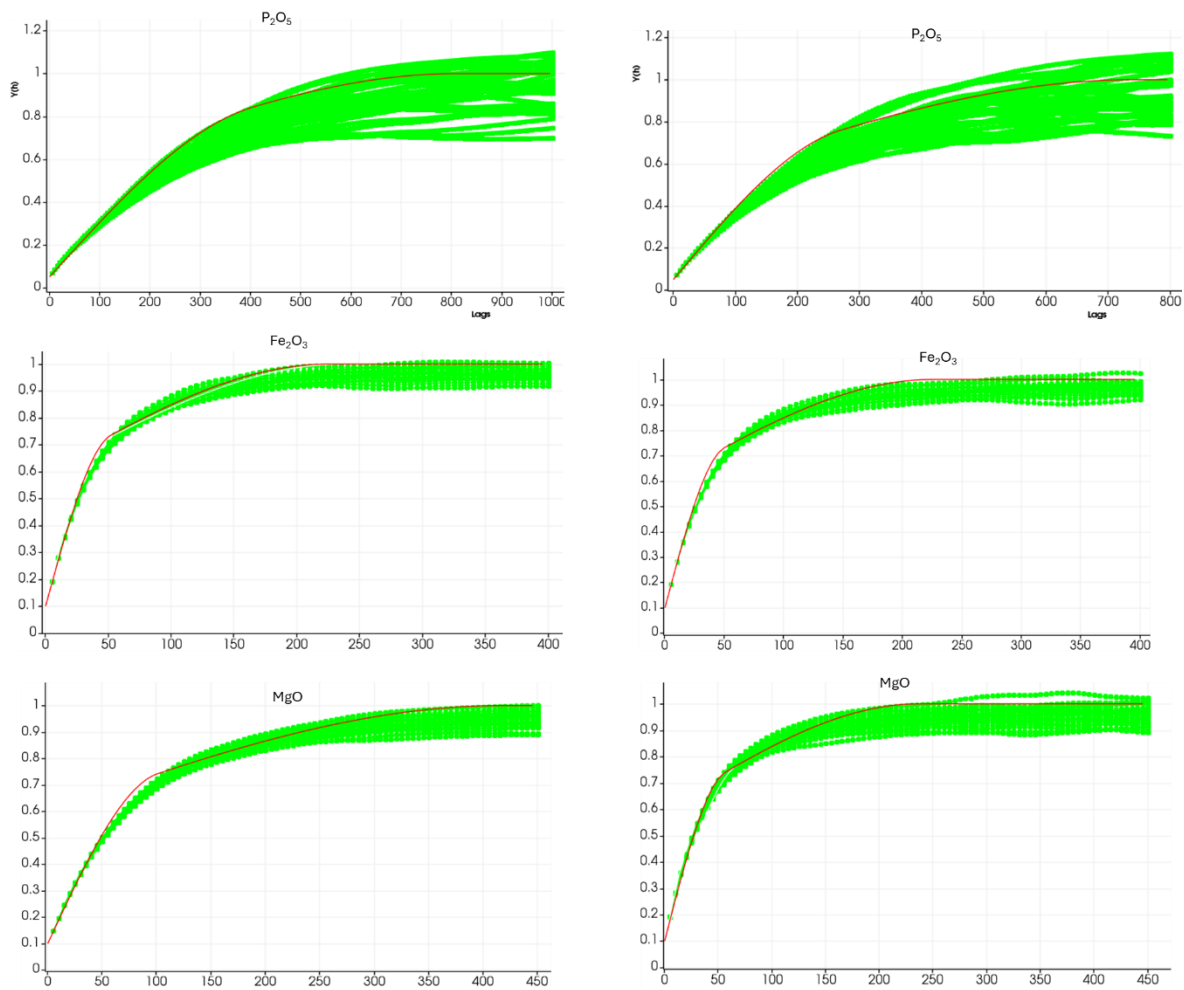


Figura 94. Validação dos variogramas da esquerda para direita maior e menor continuidade. Em verde os 35 variogramas gerados pela simulação, em vermelho o variograma dos dados originais para o CBMS micáceo.

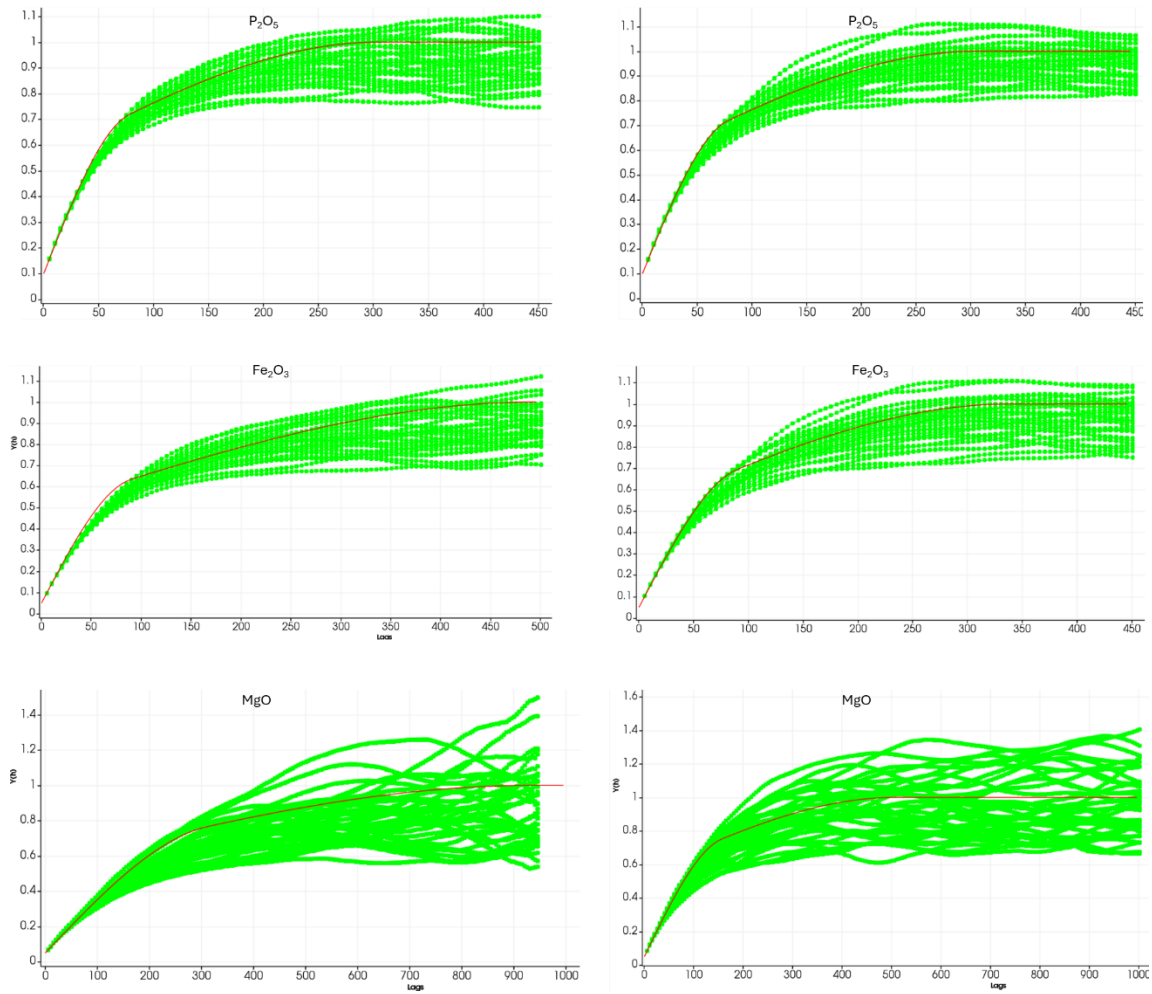


Figura 95. Validação dos variogramas da esquerda para direita maior e menor continuidade. Em verde os 35 variogramas gerados pela simulação, em vermelho o variograma dos dados originais para o CBMG oxidado.

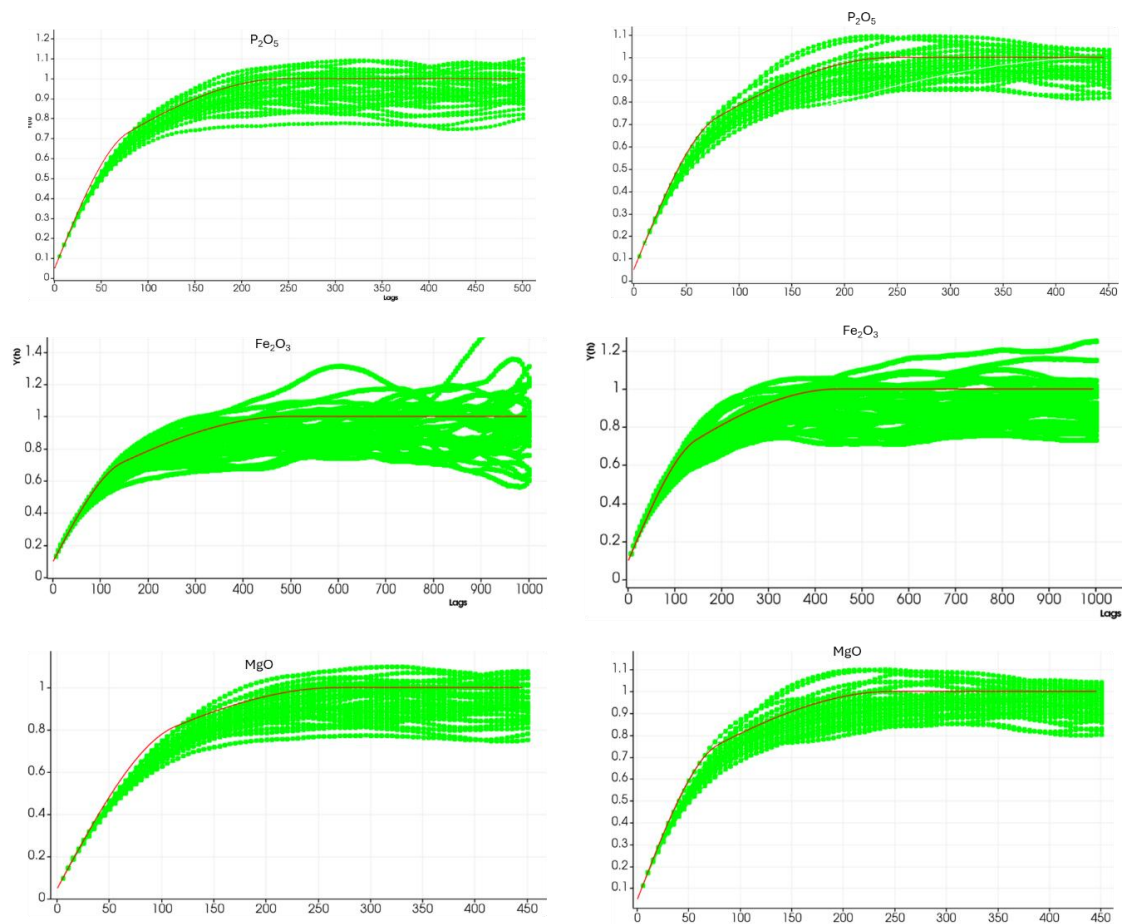


Figura 96. Validação dos variogramas da esquerda para direita maior e menor continuidade. Em verde os 35 variogramas gerados pela simulação, em vermelho o variograma dos dados originais para o CBMG micáceo.

A Figura 97 ilustra os resultados de nove cenários simulados escolhidos ao acaso com os dados originais.

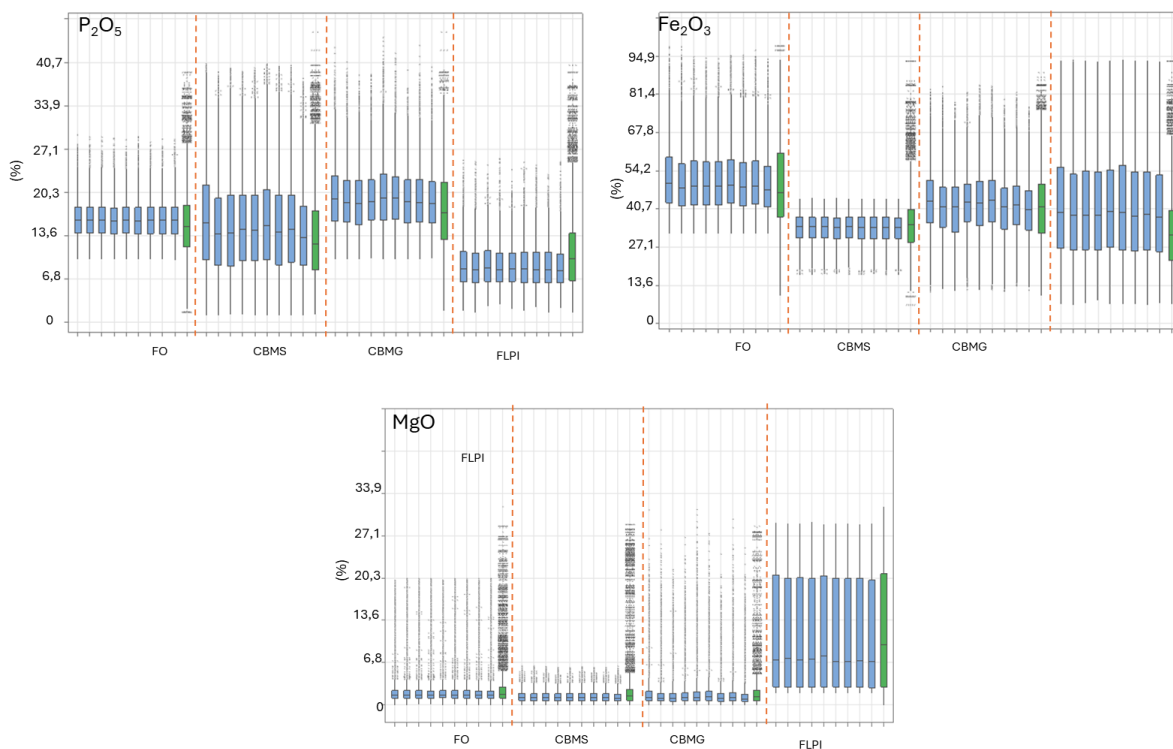


Figura 97. Gráfico dos teores simulados (9 cenários em azul) e os dados originais desagrupados (verde).

Observa-se nas estatísticas que os teores médios simulados estão aderente aos dados. Após a verificação das estimativas foi aplicado o modelo preditivo *Random Forest* nos pontos simulados (5x5x5 m), obtendo 35 cenários com informações metalúrgicas de DIST_MAS, MMAG e $P_2O_5_C$. O e-type e coeficiente de variação (CV) para os atributos metalúrgicos de previsibilidade na usina foram calculados. Uma mudança de suporte foi efetuada transferindo os pontos simulados (5x5x5 m) para blocos de dimensões 25x25x10 m. Esse procedimento suaviza as variâncias das estimativas, mas preserva as médias, por essa razão o coeficiente de variação médio de cada bloco foi calculado ainda nos pontos 5x5x5 m. As Figuras 98 e 99 mostram que a maior variabilidade nos atributos de processo está na massa de material magnético (MMAG). Em relação a recuperação mássica e teor de $P_2O_5_C$ o foscrito tem a menor variação em relação aos outros domínios tipológicos CBMG e CBMS que são tipologias com proporções importantes na composição das pilhas conforme será visto no item reconciliação, e ocorrem mais ao sul da mina. Os horizontes de minério têm importância tanto em relação às médias como a variabilidade dos atributos em uma mesma litologia; por exemplo a recuperação mássica é ligeiramente mais alta para as litologias do perfil micáceo que oxidado, isso pode ocorrer devido a característica mais límpida da apatita sem incrustações (primárias). Já a massa de magnetico é mais elevada no horizonte de oxidado que é a zona de maior presença

de óxidos de ferro. Uma análise das variáveis previstas no modelo simulado e as amostras de processo é apresentada nas Figura 100 (teores médios ajustados para confidencialidade da informação). É possível observar que o valor médio das amostras originais estão aderente aos valores preditos, ficando sempre dentro do intervalo de dispersão dos dados de um desvio padrão. A Figura 101 traz o coeficiente de variação da recuperação mássica obtida a partir dos teores simulados para a cota 810. Nota-se um aumento do CV nas regiões de borda do modelo, o que pode estar associado a malha de amostragem mais espaçada (Figura 102).

A próxima etapa consiste na reconciliação dos modelos com os dados da usina.

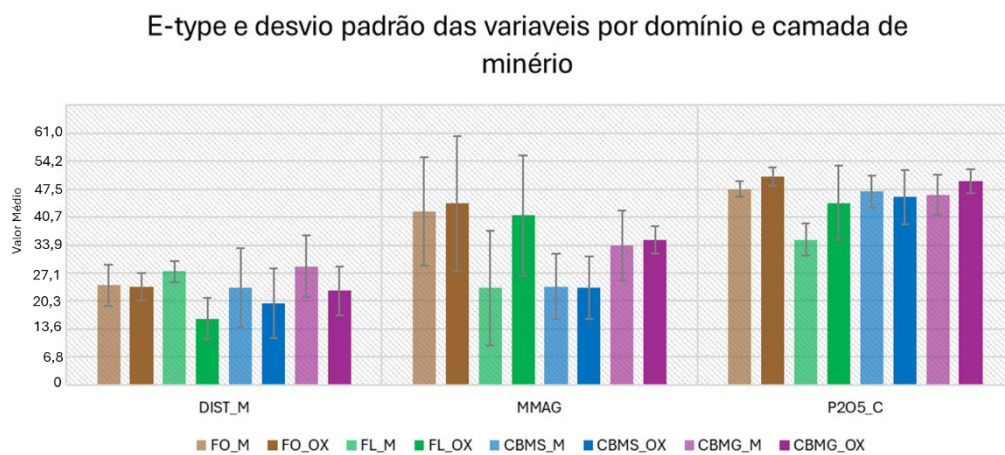


Figura 98. Gráfico das médias das variáveis de processo por domínio tipológico.

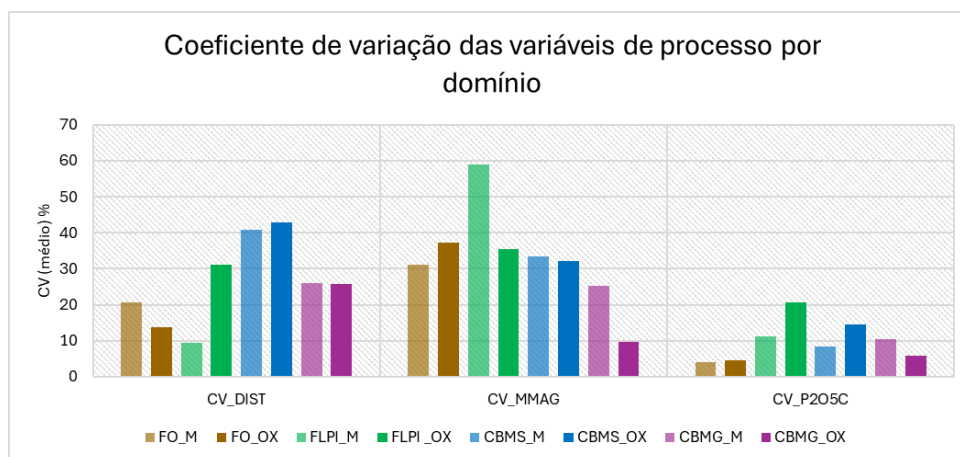


Figura 99. Gráfico do coeficiente de variação médio das variáveis de processo por domínio tipológico.

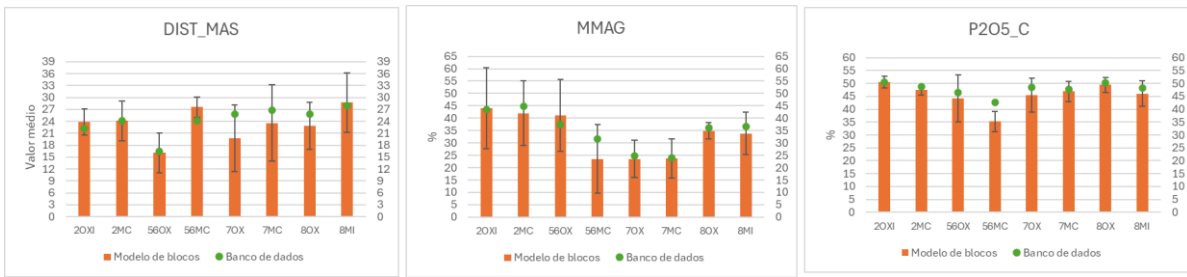


Figura 100. Gráficos das médias e desvio padrão do modelo de blocos e do banco de dados.

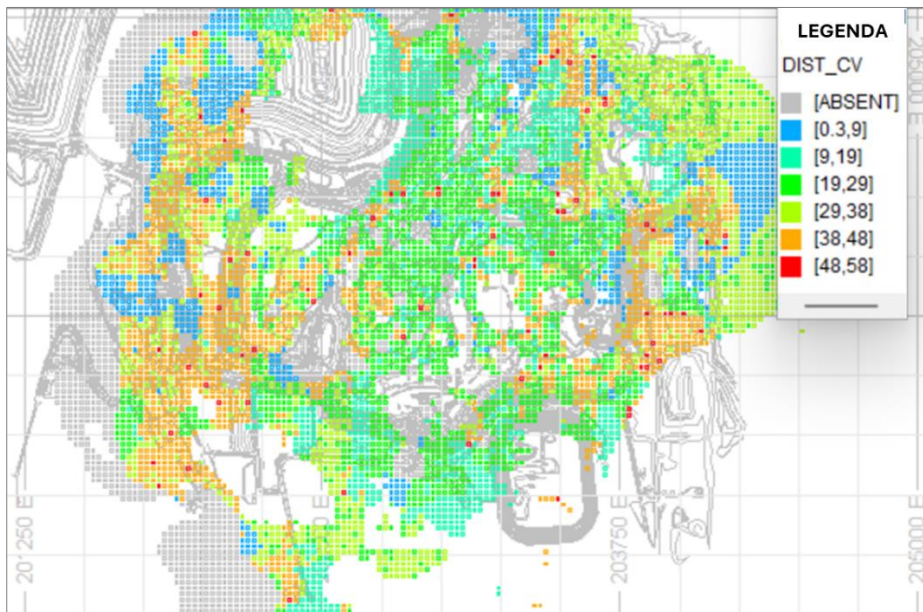


Figura 101. Mapa do coeficiente de variação da recuperação mássica na cota 810 m.

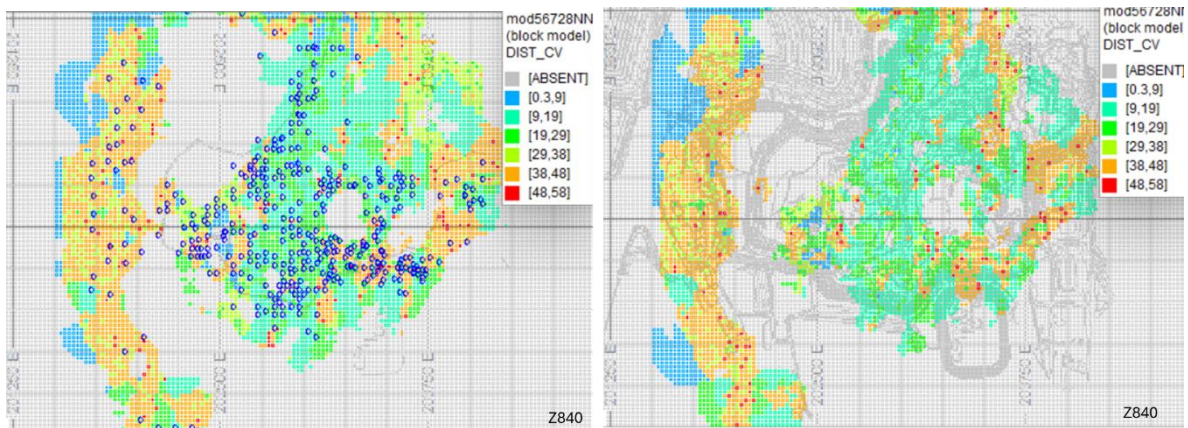


Figura 102. Mapa do coeficiente de variação da recuperação mássica na cota 840 m, à esquerda distribuição da amostragem.

6.7 Reconciliação

A fim de verificar a eficácia do modelo geometalúrgico, foi realizada a reconciliação utilizando as programações das pilhas nos meses de janeiro a dezembro de 2023, exceto as pilhas do mês de junho pois teve contribuição de material de estoque não sendo possível rastrear o local de origem do material. A reconciliação foi realizada para os atributos informados nas planilhas de indicadores da CMOC, que apresenta dados de recuperação, massas e teores calculados diariamente. As amostras para análises químicas dos teores de alimentação são realizadas na alimentação do magnético (AMG) a cada 24 horas, ou seja 2 a 3 amostras por pilha. As superfícies dos volumes lavrados mês a mês foram utilizadas para separar os blocos do modelo geometalúrgico e as médias de recuperação mássica e teor de $P_2O_5_C$ foram calculadas.

As Figuras 103 a 105 trazem uma comparação entre o realizado pela usina e o predito pelo modelo geometalúrgico. É possível notar para a recuperação mássica que os valores obtidos na usina estão a uma constante acima dos valores do modelo (Figura 103). Diferenças entre os valores observados e preditos de alguns atributos metalúrgicos são discutidos na literatura (Deutsch, 2015; Niquini, 2022; Campos2022) associados às características como não aditividade e não linearidade de algumas variáveis; relação volume alimentação e volume de bloco, e composições ou *blends* de diferentes tipos de minérios. Além desses, outros pontos merecem atenção na reconciliação como a taxa de produção, o tempo de residência do material da planta de processamento, eficiência operacional etc. O fato é que os ensaios metalúrgicos deveriam prever composições tipológicas de minérios e os modelos preditivos aplicados nos volumes de alimentação da usina, porém, esse é um problema recorrente, uma vez que o minério a ser processado depende da programação de mina que é influenciadas por diversos fatores, tornando essa abordagem impraticável.

Segundo Wambeke *et al.* (2018), as estimativas de blocos geometalúrgicos são imprecisas e para contornar esses problema necessitam de ajustes que devem ser feitos com base em observações de usinas. Assim, a fim de definir um ajuste que corrobore com as estimativas das variáveis no modelo de blocos, duas análises foram realizadas: primeiro buscou-se informação com a equipe da mineradora sobre o histórico da reconciliação dos ensaios de bancada com os resultados da usina, sendo constatado que, os testes de bancada normalmente matem-se abaixo da recuperação mássica da usina e são ajustados. A partir dessa informação verificou-se a relação entre a recuperação mássica da usina e a recuperação do modelo geometalúrgico a fim de encontrar um fator de mudança de escala que representasse tal

relação (Tabela 20). Esse fator foi calculado com base nos valores médios observados e estimados; e nos erros entre eles durante todo o ano de 2023.

A reconciliação permitiu observar que, ao aplicar um fator de mudança de escala (laboratório para usina) nas previsões obtidas através da técnica de *random forest*, os resultados se aproximaram dos valores de produção da recuperação mássica da usina. Os desvios entre os atributos preditos em relação à produção foram inferiores a |5|% (Figura 104).

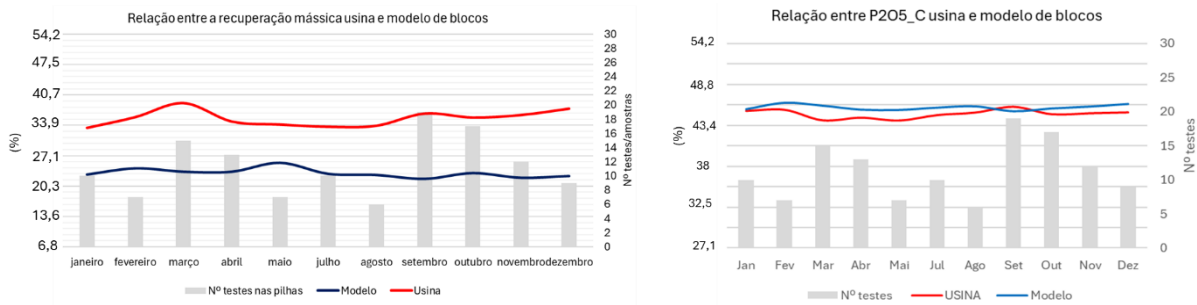


Figura 103. Relação DIST_MAS e P₂O₅_C entre usina e modelo de blocos.

Tabela 20. Cálculo do fator de mudança de escala.

Mês	MODELO vs USINA		
	Erro absoluto	Erro relativo	FATOR
janeiro	7.69	0.37	1.37
fevereiro	8.42	0.38	1.38
março	11.28	0.49	1.49
abril	8.27	0.39	1.39
maio	6.28	0.29	1.29
julho	7.76	0.37	1.37
agosto	8.12	0.39	1.39
setembro	10.75	0.50	1.50
outubro	9.15	0.42	1.42
novembro	10.33	0.48	1.48
dezembro	11.08	0.50	1.50
MÉDIA ANUAL			1.42

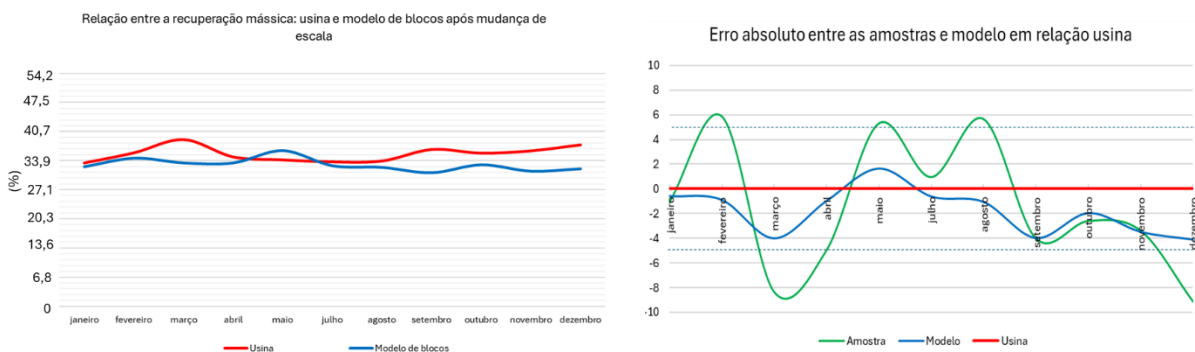


Figura 104. Recuperação mássica entre usina e modelo de blocos após fator de mudança de escala. À direita os desvios entre modelo geometalúrgico, amostras e dados usina.

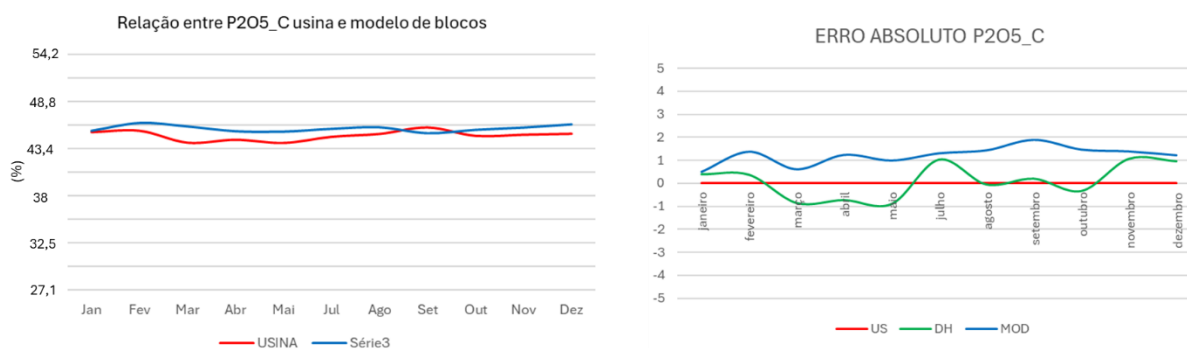


Figura 105. P₂O₅_C entre usina e modelo de blocos. À direita os desvios entre modelo, amostras e dados usina.

Como o volume de alimentação é composto por uma composição (*blends*) de tipologias de minérios (Figura 106), e os modelos preditivos foram obtidos individualmente para cada tipologia, uma outra avaliação foi realizada a fim de verificar a reprodução das relações entre previsto e realizado no volume processado. Inicialmente verificou-se as proporções das tipologias para composição das pilhas, que corresponde aproximadamente 30% de FO, 30% CBMS, 30% CBMG e 10% até 15% de FLPI tanto no nível oxidado como micáceo. O modelo preditivo do FO foi então aplicado nos teores de alimentação (resultados diários) e uma recuperação mássica obtida. O processo foi repetido utilizando os modelos preditivos das demais tipologias e uma recuperação mássica final proporcional a participação dos tipos de minério foi calculada (Figura 107). As proporções das misturas de minérios poderiam resultar em diferenças nas estimativas da recuperação mássica nos modelos preditos, mas esse fato não foi observado.

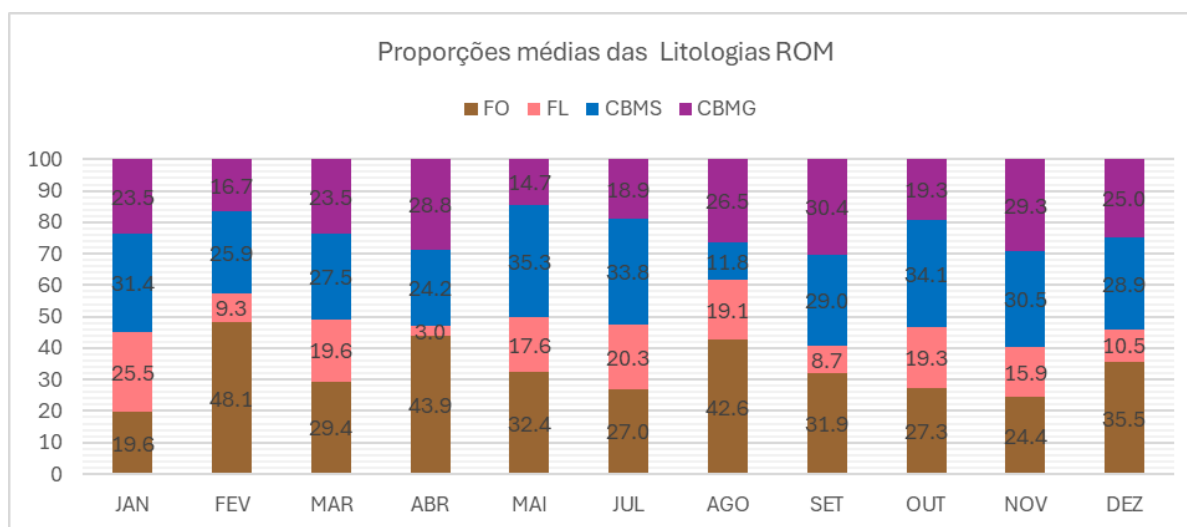


Figura 106. Proporções das tipologias na programação das pilhas.

Recuperação mássica média do ano				Recuperação mássica média do ano			
	Erro Abs	Erro relativo	Fator		Erro Abs	Erro relativo	Fator
FO oxidado	5.69	0.25	1.25	FO micáceo	8.07	0.37	1.37
CBMS oxidado	9.26	0.43	1.43	FLPI micáceo	8.19	0.37	1.37
CBMG oxidado	9.96	0.47	1.47	CBMG micáceo	8.43	0.39	1.39
FLPI oxidado	12.13	0.61	1.61	CBMS micáceo	7.77	0.35	1.35
Média proporcional oxidado			1.41	Média proporcional oxidado			1.37
				Média final			1.39

Figura 107. Resultados dos modelos preditivos aplicados aos teores do volume de alimentação proporcional a participação das tipologias na composição das pilhas.

7. CONCLUSÕES

7.1 Aplicação do algoritmo difuso para definição dos domínios com amostras imprecisas.

A escolha da teoria dos conjuntos fuzzy foi utilizada na tentativa de apresentar uma abordagem viável para lidar com a incerteza do fenômeno geológico e com as amostras com descrições ambíguas. Algumas amostras do banco de dados receberam pertinência zero para os oito domínios pré-definidos. Possivelmente por apresentar variações pontuais (problemas multiclases com estruturas geológicas mais intensas), já que apresentaram baixa compatibilidade com as expressões químicas utilizadas no conjunto fuzzy de entrada. Ou até mesmo por viés analítico em campanhas mais antigas, necessitando revisitar os furos de sondagem para averiguação. Nesses casos, fontes de informação adicionais de caracterização mineralógica, poderiam ser úteis para ajuste do modelo tornando a predição mais assertiva. Ainda assim, o número de amostras classificadas foi suficiente para permitir a definição dos domínios com mais eficiência que os métodos de agrupamento tradicionais.

Uma outra vantagem da metodologia fuzzy é o grau de pertencimento das amostras, que pode direcionar futuros trabalhos de mapeamento geológico, campanhas de amostragem, e auxiliar na escolha de amostras para caracterização e realização de testes em pequena escala.

Pode-se concluir que o modelo fuzzy cumpriu os objetivos iniciais ao maximizar a separação entre os grupos utilizando informações dos teores combinadas ao conhecimento especializado e descrição geológica dos furos de sondagem. Também reduziu o tempo de análise pelo geomodelador. Outros benefícios atingidos utilizando a inteligência computacional foi a replicabilidade e padronização dos conceitos para a definição dos grupos e automação das etapas permitindo atualização ágil do modelo quando houver inserção de novos dados.

7.2 Utilização de variáveis primárias para modelagem de atributos metalúrgicos por técnicas computacionais

Nas principais etapas da construção do modelo geometalúrgico nota-se que a integração entre o conhecimento geológico e as ciências da computação foi fundamental para a viabilidade do modelo. A seleção dos atributos por PCA otimizou todas as análises subsequentes relacionadas aos modelos preditivos, não só por eliminar recursos redundantes, garantido maior acurácia e agilidade dos algoritmos e procedimentos geoestatísticos, mas também por evidenciar que as variáveis aditivas eram suficientes para captar a variabilidade do fenômeno neutralizando questões relacionadas não aditividade das variáveis de processo. Os modelos preditivos apresentaram bons índices com os dados de reconciliação tornando possível não só prever o valor do atributo metalúrgico mais informar através das simulações das variáveis preditivas a variabilidade esperada nas frentes de lavra. Reconhecer a variabilidade ajuda a melhorar o desempenho da planta a tipos e ganhos de produtividade no planejamento de lavra e beneficiamento. Nesse sentido, as técnicas de computação utilizadas se mostraram fortes aliadas para suprir as dificuldades relacionadas às características e limitações dos bancos de dados.

Em última análise, é possível concluir que obter funções de transferência, que associem as características das propriedades primárias do depósito com o desempenho da planta é plausível quando o banco de dados metalúrgico não consegue suprir as premissas para uma estimativa consistente. Assim como a reconciliação, é fundamental ao permitir aumentar a confiabilidade nos modelos e detectar e corrigir vieses que poderiam invalidar as estimativas.

7.3 Cumprimento das metas e recomendações

Esse estudo teve como objetivo a determinação de uma metodologia capaz de suprir as deficiências do banco de dados e contornar as dificuldades relacionadas às características das variáveis metalúrgicas.

Na análise exploratória dos dados observou-se que a quantidade de informação uniformemente distribuída no banco de dados não seria suficiente para uma estimativa direta de alguns atributos metalúrgicos e que seria necessário utilizar funções de transferências a partir das variáveis primárias da rocha. Para entender a mineralogia do depósito e o comportamento esperado na planta de concentração era necessário modelar os protólitos e definir domínios de estimativa. Um grande problema encontrado no banco de dados foi a falta de consistência da descrição litológica o que inviabilizaria a construção do modelo geometalúrgico. A aplicação

da teoria do conjunto fuzzy associado às análises geológicas e geoestatísticas permitiram encontrar relações importantes que auxiliaram na definição dos domínios. A modelagem das tipologias foi fundamental para o cumprimento dos objetivos deste estudo pois possibilitou analisar as relações entre as características primárias da rocha com as respostas de processo.

Diferentes modelos de predição foram testados apresentando resultados plausíveis com a base de dados e as métricas de validação. A meta de definição de uma metodologia a ser aplicada para construção do modelo geometalúrgico estava, portanto, atingida.

Por fim, a reconciliação entre o modelo gerado e os dados de produção foi verificada a fim de confirmar se a metodologia produzia resultados satisfatórios. A reconciliação permitiu encontrar um fator de correção e diminuir as diferenças entre o predito e observado em função do volume bloco e volume de alimentação, tornando as estimativas válidas.

Conclui-se que a metodologia possibilitou a construção do modelo geometalúrgico cujo objetivo final é orientar o planejamento de lavra e otimizar recursos minerais, custos e receitas.

Sugere-se para trabalhos futuros a verificação das descrições litológicas, a iniciar pelas amostras que não tiveram uma classificação pelo algoritmo fuzzy validada pela informação geológica, que representa mais de 30% do banco de dados, valor que impacta de forma significativa o recurso disponível. Bem como verificar as amostras que apresentaram probabilidades para mais de uma categoria. A incerteza quantificada no agrupamento é uma importante fonte de informação para direcionar os estudos de geologia e validação do banco de dados. Associado a isso, sugere-se a verificação dos critérios e adequação do algoritmo sempre que novos dados forem inseridos.

Por fim recomenda-se adensar a amostragem dos testes de bancada nas regiões de maior variabilidade da recuperação mássica predita, ou seja, as regiões dos apatita carbonatito e flogopitito, regiões de borda do complexo. Uma amostragem mais efetiva, além de permitir uma estimativa mais precisa das variáveis modeladas, também pode possibilitar a inferência dos valores de recuperação metalúrgica não utilizada nesse estudo de caso.

8. REFERÊNCIAS

ALMEIDA, P. E. M.; EVSUKOFF, A. G. Sistemas *Fuzzy*. In: REZENDE, S. O (Org.). *Sistemas Inteligentes: fundamentos e aplicações*. São Paulo: Manole, 2003, p. 169-202.

AMARANTE, F. A. N. *Avaliação da incerteza associada ao modelo geológico através de métodos geoestatísticos estocásticos*. Dissertação (Mestrado), Universidade Federal do Rio Grande do Sul, Programa de Pós-Graduação em Engenharia de Minas, Metalúrgica e de Materiais, Porto Alegre, 2018.

ANNUAL REPORT CMOC. Relatório interno. 2022. Disponível em: <https://en.cmoc.com/html/InvestorMedia/Performance>. Acesso em: 04 jan. 2024.

ARAÚJO, C. D. P. *Uso de informação secundária imprecisa como distribuição de probabilidade local em conjuntos de dados completamente heterotópicos*. Tese (Doutorado), Universidade Federal do Rio Grande do Sul, Programa de Pós-Graduação em Engenharia de Minas, Metalúrgica e de Materiais, 2019.

AVELAR, A. N. *Caracterização dos minerais do grupo da apatita e carbonatos no minério sílico-carbonatado de Catalão, GO, e sua relevância no processo de flotação*. Tese (Doutorado), Universidade Federal de Minas Gerais, Programa de Pós-Graduação em Engenharia Metalúrgica, Materiais e de Minas, 2018.

BATELOCHI, M. *Simulação geoestatística e regressão multivariada aplicadas na modelagem geometalúrgica de um depósito de zinco*. Dissertação (Mestrado), Universidade Federal do Rio Grande do Sul, Programa de Pós-Graduação em Engenharia de Minas, Metalúrgica e de Materiais. 2018.

BERKHIN, P. A survey of clustering data mining techniques. In: KOGAN, J.; NICHOLAS, C.; TEBoulLE, M (Ed.). *Grouping multidimensional data: recent advances in clustering*. Heidelberg: Springer, 2006, p. 25-71.

BERRY, M.; MCCARTHY, P. Practical Consequences of Geological Uncertainty. In: *6th International Mining Geology Conference*, Darwin, NT, Aug., 2006.

BIZZI, L. A.; DE WIT, M. J.; SMITH, C. B.; MCDONALD, I.; ARMSTRONG, R. A. Heterogeneous enriched mantle materials and Dupal-type magmatism along the SW margin of the São Francisco Craton, Brazil. *Journal of Geodynamics*, 20(4), 469-491. 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, 2013.

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

BRAGA JÚNIOR, J. M. *Utilização de simulação conjunta colocada com variável supersecundária para construção de modelo geometalúrgico de nióbio Araxá-MG*. Dissertação

(Mestrado), Universidade Federal do Rio Grande do Sul, Programa de Pós-Graduação em Engenharia de Minas, Metalúrgica e de Materiais, 2017.

BREIMAN, L. Random forests. In: SHAPIRE, R. E. (Org.). *Machine learning*. Amsterdam: Kluwer Academic Publishers, v. 45, 2001, p. 5-32.

BROD, J.A. *Petrology and geochemistry of the Tapira alkaline complex, Minas Gerais State, Brazil*. PhD Thesis, University of Durham, UK, 1999.

BROD, J.A.; RIBEIRO, C.C.; GASPAR, J.C.; JUNQUEIRA-BROD, T.C.; BARBOSA, E.S.R.; RIFFEL, B.F.; SILVA, J.F.; CHABAN, N.; FERRARRI, A.J.D. Geologia e Mineralizações dos Complexos Alcalino Carbonatíticos da Província Ígnea do Alto Paranaíba. In: *42º Congresso Brasileiro de Geologia*, Araxá, Minas Gerais, Excursão 1: p. 1-29, 2004.

CAMPOS, P. H. A., COIMBRA LEITE COSTA, J. F., KOPPE, V. C., & ARCARI BASSANI, M. A. Geometallurgy-oriented mine scheduling considering volume support and non-additivity. *Mining Technology*, 131(1), 1-11, 2022.

CAPP, E.; NIENOV, O. H. (Orgs.). *Bioestatística quantitativa aplicada*. Universidade Federal do Rio Grande do Sul. Porto Alegre: UFRGS, 2020.

CARACTERIZAÇÃO DOS TIPOS GEOLÓGICOS DE MINÉRIOS_FINAL. *Relatórios internos da CMOB Brasil*. 2023.

CARRASCO, P.; CHILÈS, J. P.; SÉGURET, S. Additivity, metallurgical recovery and grade. In: *8th international Geostatistics Congress*, Geostats, Santiago, 2008.

CARVALHO, H. M. *Aprendizado de máquina voltado para mineração de dados: árvores de decisão*. Trabalho de Conclusão de Curso (Graduação), Universidade de Brasília, 2014.

CAVALCANTI JR., N. L. *Clusterização baseada em algoritmos fuzzy*. Dissertação (Mestrado), Universidade Federal de Pernambuco, Ciência da Computação, 2006.

CASTRO, S. A. B. D.; SILVA, A. C. Evaluation of PCA with variable selection for cluster typological domains. *REM-International Engineering Journal*, 77(2), 2024.

CHEMALE, F.; TAKEHARA, L. *Minério de ferro, geologia e geometalurgia*. São Paulo: Editora Blucher, 2013.

CHIBAYA, A. *Geometallurgical analysis – implications of operating flexibility: a case for Geometallurgy for Orapa A/K1 deposit*. Doctoral dissertation, University of the Witwatersrand, Johannesburg, 2013.

CORDEIRO, P. F. O.; BROD, J. A.; DANTAS, E. L.; BARBOSA, E. S. R. Mineral chemistry, isotope geochemistry and petrogenesis of niobium-rich rocks from the Catalão I carbonatite-phoscorite complex, Central Brazil. *Lithos*, v. 118, p. 223-237, 2010.

COWAN, J.; BEATSON, R.; ROSS, H. J.; FRIGHT, W. R.; MCLENNAN, T. J.; EVANS, T. R.; CARR, J. C.; LANE, R. G.; BRIGHT, D. V.; GILLMAN, A.J.; OSHUST, P.A.; TITLEY,

M. Practical implicit geological modelling. *In: 5th international mining geology conference, the Australian institute of mining and metallurgy*, v. 8, p. 89–99, 2003.

COWARD, S.; VANN, J.; DUNHAM, S.; STEWART, M. The primary-response framework for geometallurgical variables. *In: Seventh International Mining Geology Conference*, p. 109–113, 2009.

DEMICO, R. V.; KLIR, G. J. *Fuzzy Logic in Geology*. Academic Press. An imprint of Elsevier Science, 2004.

DEUTSCH, C. V. *Geostatistical Reservoir Modeling*. New York: Oxford University Press, 376 p., 2002.

DEUTSCH, C. V. Geostatistical modelling of geometallurgical variables – problems and solutions. *In: The Second AUSIMM International Geometallurgy conference*, Brisbane. Australasian Institute of Mining and Metallurgy, 2013.

DEUTSCH, J. L. *Multivariate spatial modeling of metallurgical rock properties*. Thesis (Doctor of Philosophy, Mining Engineering) – Department of Civil and Environmental Engineering, University of Alberta, 2015.

DOMINY, S. C.; O'CONNOR, L.; PARBHAKAR-FOX, A.; GLASS, H. J.; PUREVGEREL, S. "Geometallurgy—A route to more resilient mine operations." *Minerals*, 8.12: 560, 2018. <https://doi.org/10.3390/min8120560>.

DRUMOND, D. A. *Estimativa e classificação de variáveis geometalúrgicas a partir de técnicas de aprendizado de máquinas*. Tese (Doutorado), Universidade Federal do Rio Grande do Sul, Programa de Pós-Graduação em Engenharia de Minas, Metalúrgica e de Materiais, 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*, Melbourne, Victoria, p. 19–20, 2007.

FAN, W.; HANCOCK, G. R. Robust Means Modeling: An Alternative for Hypothesis Testing of Independent Means Under Variance Heterogeneity and Nonnormality. *Journal of Educational and Behavioral Statistics*, v. 37, 21p., 2012.

FARIA, B. F. P. *Teste F na regressão linear múltipla para dados temporais em correlação serial*. Tese (Doutorado em Estatística, Matemática e Computação). Universidade Aberta, Lisboa, 2011.

FERNANDES, F. G. *Estudo do melhor método de extrapolação de regressão múltipla para construção do modelo geometalúrgico de uma mina de fosfato brasileira*. Dissertação (Mestrado em Engenharia Mineral), Programa de Pós-Graduação em Engenharia Mineral, Universidade Federal de Ouro Preto, Ouro Preto, 2013.

FOGLIATTO, F. S.; ANZANELLO, M. J. Selecting the best clustering variables for grouping mass-customized products involving workers' learning. *International Journal of Production Economics*, 9p., 2011.

FOUEDJIO, F.; HILL, E. J.; LAUKAMP, C. Geostatistical clustering as an aid for ore body domaining: case study at the Rocklea Dome channel iron ore deposit, Western Australia. *Applied Earth Science, Taylor & Francis*, v. 127, n. 1, p. 15–29, 2017. <https://doi.org/10.1080/03717453.2017.1415114>

FRIMMEL, H. E.; NWAAILA, G. T. Geologic evidence of syngenetic gold in the Witwatersrand Goldfields, South Africa. In: SILLITOE, T.; GOLDFARB, R.; ROBERT, F.; SIMMONS, S. (Eds.). *Geology of the major gold deposits and provinces of the world*, (Special Publication 23, p. 645–668). Society of Economic Geologists. 2020. <https://doi.org/10.5382/SP.23.31>.

GIBSON, S.A.; THOMPSON, R.N.; LEONARDOS, O.H.; DICKIN, A.P.; MITCHELL, J.G. The Late Cretaceous impact of the Trindade mantle plume – evidence from large-volume, mafic, potassic magmatism in SE Brazil. *Journal of Petrology*, 36, p. 189-229, 1995.

GLACKEN, I.; BLACKNEY, P. A Practitioners Implementation of indicator kriging. In: *Symposium on Beyond Ordinary Kriging*, 14p., 1998.

GOOVAERTS, P. *Geostatistics for Natural Resources Evaluation*. New York: Oxford University Press, 483p., 1997.

GRASSO, C. B. *Petrologia do complexo alcalino-carbonatítico de Serra Negra*. Dissertação (Mestrado), Universidade de Brasília, Brasil, 2010.

GRASSO, C. B. *Geologia e geometurgia dos minérios de fosfato na base do manto de intemperismo da mina do Barreiro, Araxá*. Tese (Doutorado), Universidade de Brasília, 2016.

HAIR, J. F.; ANDERSON, R. E.; TATHM, R. L.; BLACK, W. C.; BABIN, B. J. *Análise multivariada de dados*. 6ª ed. New Jersey: Prentice Hall, 2009.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. Random forests. In: HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The elements of statistical learning: Data mining, inference and prediction*. Berlim: Springerp, p. 587-604, 2009.

HOAL, K. O. Getting the Geo into Geomet. *Society of Economic geologist*, n. 73, p. 52, 2008.

HUNT, J.; KOJOVIC, T.; BERRY, R. Estimating comminution indices from ore mineralogy, chemistry and drill core logging. In: *GeoMet 2013: The Second AusIMM International Geometallurgy Conference*, The Australasian Institute of Mining and Metallurgy (AusIMM), p. 173–176, 2013.

HUNT, J.; BERRY, R.; BECKER, M.; BAUMGARTNER, R. A special issue dedicated to geometallurgy: Preface. *Economic Geology*, 114(8), p. 1473-1479, 2019.

ISAAKS, E. H.; SRIVASTAVA, R. M. *Applied Geostatistics*. New York: Oxford University Press, p. 561, 1989.

IZBICKI, R., SANTOS, T. M. *Aprendizado de máquina: uma abordagem estatística*. Carlos: UICLAP, 272p., 2020.

IZENMAN, A. J. Committee machines. In: IZENMAN, A. J. *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*. Berlin: Springer, p. 505-550, 2008.

JOLLIFFE, I.T. Discarding Variables in a Principal Component Analysis. I: Artificial Data. *Journal of the Royal Statistical Society*, n. 2, p. 160-173, 1972.

JOURNEL, A. G. Nonparametric estimation of spatial distributions. *Journal of the International Association for Mathematical Geology*, Springer, v. 15, n. 3, 1983.

JOURNEL, A. G. *Fundamentals of geostatistics in five lessons*. New Jersey: John Wiley & Sons, v. 8, 1989.

KAHN, H. Caracterização mineralógica e tecnológica da jazida de fosfato do maciço alcalino de Anitápolis, SC. Tese de Doutorado. Universidade de São Paulo. 1988.

KOLTERMANN, C. E.; GORELICK, S. M. Heterogeneity in sedimentary deposits: A review of structure-imitating, process-imitating, and descriptive approaches. *Water Resources Research*, v. 32, 1996.

KOVACIC, Z. BOGDAN, S. *Fuzzy Controller Design Theory and Application*. London: Taylor & Francis Group, 416p., 2006.

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

LARRONDO, P.F.; DEUTSCH, C.V. Methodology for Geostatistical Model of Gradational Geological Boundaries: Local Non-stationary LMC. *Cent. Comput. Geostat.*, v. 6, p. 1–17. 2004.

LE MAITRE, R. W. *Igneous Rocks: Classification and Glossary of Terms: Recommendations of the International Union of Geological Sciences Subcommittee on the Systematics of Igneous Rocks*. Cambridge: Cambridge University Press, 2002.

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, Brisbane, QLD*: Australasian Institute of Mining and Metallurgy, 2011.

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*, 134, p. 156-165, 2019.

LYRA, W. D. S.; SILVA, E. C. D.; ARAÚJO, M. C. U. D.; FRAGOSO, W. D.; VERAS, G. Classificação periódica: um exemplo didático para ensinar análise de componentes principais. *Química Nova*, 33, p. 1594-1597, 2010.

LOUREIRO, F. E. L.; MONTE, M. B. de M.; NASCIMENTO, M. Agrominerais – Fosfato. In: LUZ, A. B.; LINS, F. A. F. (Orgs.). *Rochas & Minerais Industriais: usos e especificações*. Rio de Janeiro: CETEM – MCT, 2005.

LOZANO, C.; BENNET, C. *Geometallurgical modeling applied to production forecasting, plant design and optimisation*. Universidad de Chile – Departamento Ingeniería de Minas Taller, PROCEMIN, 11p., 2003.

MACÁRIO, V.; CARVALHO, F. D. A. de. An adaptive semi-supervised fuzzy clustering algorithm based on objective function optimization. *In: 2012 IEEE International Conference on Fuzzy Systems*, p. 1-8, 2012.

MARQUES, D. M. *Métodos estocásticos aplicados a definição de estratégias de amostragem e homogeneização*. Tese (Doutorado), Universidade Federal do Rio Grande do Sul, Programa de Pós-Graduação em Engenharia de Minas, Metalúrgica e de Materiais, 2014.

MARRO, A. A.; SOUZA, A. M. C.; CAVALCANTE, E. R. S.; BEZERRA, G. S.; NUNES, R. O. *Lógica Fuzzy: Conceitos e aplicações*. Universidade Federal do Rio Grande do Norte, 2010. Disponível em: https://www.researchgate.net/publication/228595876_Logica_Fuzzy_Conceitos_e_aplicacoes. Acesso em: 24 abr. 2021.

MARTIN, R.; BOISVERT, J. A random-path spatial clustering algorithm. *CCG Annual Report*, University of Alberta, v. 19, 2017.

MATHERON, G. Principles of Geostatistics. *Economic Geology*, Littleton, v. 58, n. 8, p. 1246-1266, 1963.

MATHERON, G. *La d'Estructuration des Hautes Teneurs et le Krigeage des Indicatrices*. Internal report N-761, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau, 33p. 1982. Disponível em: https://cg.ensmp.fr/bibliotheque/cgi-bin/public/bibli_index. Acesso em: 22 jan. 24.

MATHERON, G.; BEUCHER, H.; DE FOUQUET; C., GALLI, A.; GUERILLOT, D.; AND RAVENNE, C. Conditional simulation of the geometry of fluvio-deltaic reservoirs. *In: Spe annual technical conference and exhibition*. Society of Petroleum Engineers, 1987.

METZ, J.; MONARD, M. C. Clustering hierárquico: uma metodologia para auxiliar na interpretação dos clusters. *In: XXV Congresso Brasileiro de computação*. São Leopoldo, RS, 2014.

MILLIGAN, G. W. Clustering validation: Results and implications for Applied analyses. *In: ARABIE, P.; HUBERT, L. J.; SOETE, G. D. (Eds.). Clustering and classification*. London: World Scientific Publishing, p. 341-375, 1996.

MINISTÉRIO DA AGRICULTURA E PECUÁRIA (MAPA). *As reservas minerais para fertilizantes no Brasil*. 2022. Disponível em: <https://www.gov.br/agricultura/pt-br/assuntos/insumos-agropecuarios/insumos-agricolas/fertilizantes/plano-nacional-de-fertilizantes/as-reservas-minerais-para-fertilizantes-no-brasil>. Acesso em: 15 mai. 23.

MITCHELL, T. *Machine Learning*. New York: McGraw-Hill, 1997.

MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. *Introduction to linear regression analysis*. New Jersey: John Wiley & Sons, v. 821. 645 p., 2012.

MOREIRA, G. C. *Análise de agrupamento aplicada à definição de domínios de estimativa para a modelagem de recursos minerais*. Dissertação (Mestrado em Engenharia de Minas, Metalurgia e de Materiais), Universidade Federal do Rio Grande do Sul, Porto Alegre, 102p., 2020.

MOTTA, E. G. M. *Definição de domínios mineralógicos em minério de ferro utilizando krigagem de indicadores*. Dissertação (Mestrado em Engenharia de Minas, Metalurgia e de Materiais), Universidade Federal do Rio Grande do Sul, Porto Alegre, 2014.

MÜLLER, A. C.; GUIDO, S. *Introduction to machine learning with Python: A guide for data scientists*. Sebastopol: O'Reilly Media, 2016.

NEWTON, M. J., GRAHAM, J. M. Spatial modelling and optimisation of geometallurgical indices. In: *The first AUSIMM international geometallurgy conference*. Brisbane, QLD: Australasian. Institute of Mining and Metallurgy, 2011.

NIQUINI, F. G. F. *Predição simultânea de produtos e rejeitos em plantas de processamento de zinco e ouro a partir das características do minério*. Tese (Doutorado em Engenharia de Minas, Metalurgia e de Materiais), Universidade Federal do Rio Grande do Sul, Porto Alegre, 244p., 2020.

OLEA, R. A. *Geostatistics for engineers and earth scientists*. Berlin: Springer Science & Business Media, 2012.

OLIVEIRA, S. M. B.; IMBERNON, R. A. L. Weathering and related REE concentration in Catalão I carbonatite complex, central Brazil. *Journal of South American Earth Sciences*, v. 11, n. 4, p. 379-388, 1998.

PACÍFICO, L. D. S. *Algoritmos de agrupamento particionais baseados na Meta-heurística de otimização por busca em grupo*. Dissertação (Mestrado), Universidade Federal de Pernambuco, 2016.

PALMIERI, M. *Modelo geológico e avaliação de recursos minerais do depósito de nióbio Morro do Padre, complexo alcalino-carbonatítico de Catalão II, GO*. Dissertação (Mestrado), Universidade de Brasília, 2011.

PEREIRA, S. G. *Inserção de dados faltantes não aleatórios para estimativa de variável geometalúrgica*. Dissertação (Mestrado), Universidade Federal do Rio Grande do Sul, 2019.

PROJETO GEOMETALURGIA RELATÓRIO FINAL DEZ. 2016. *Relatórios internos da CMOB Brasil*, 2023.

RAMOS, G. S. *Caracterização mineralógica do processo de beneficiamento do minério de nióbio da mina Boa Vista, Catalão-GO, Brasil*. Dissertação (Mestrado), Universidade de Brasília, 2022.

RIBEIRO, C. C. *Geologia, geometalurgia, controles e gênese dos depósitos de fósforo, terras raras e titânio do complexo carbonatítico Catalão I, GO*. Tese (Doutorado), Universidade de Brasília, Brasil, 2008.

ROMARY, T.; RIVOIRARD, J.; DERAISME, J.; QUINONES, C.; FREULON, X. Domaining by clustering multivariate geostatistical data. *In: ABRAHAMSEN, P.; HAUGE, R.; KOLBJORSEN, O. (Ed.). Geostatistics Oslo*. Berlim: Springer, 2012. p. 455–466, 2012.

ROSSI, M. E.; DEUTSCH, C. V. *Mineral resource estimation*. Berlim: Springer Science & Business Media. 2014.

SAAB, A. A.; PAULA, R. D. A. O mercado de fertilizantes no Brasil: diagnósticos e propostas de políticas. *Revista de política agrícola*, 17(2), p. 5-24. 2008.

SANCHES, M. K. *Aprendizado de máquina semi-supervisionado: proposta de um algoritmo para rotular exemplos a partir de poucos exemplos rotulados*. Dissertação (Mestrado em Ciências da Computação), Universidade de São Paulo, 142p., 2003.

SANCHEZ, S.; DEUTSCH, C.V. Domain delimitation with radial basis functions. *In: J. L. Deutsch (Ed.). Geostatistics lessons*. 2022. Disponível em: <https://geostatisticslessons.com>. Acesso em: 01 out. 2023.

SCHOLKOPF, B; SMOLA, A. J. *Learning With Kernels: Support Vector Machines Regularization Optimization and Beyond*. Cambridge: MIT Press, 2002.

SCRUCCA, L. Clustering multivariate spatial data based on local measures of spatial autocorrelation. *Quaderni del Dipartimento di Economia, Finanza e Statistica*, Università di Perugia, v. 20, n. 1, p. 11, 2005.

SEPULVEDA, E. E. *Quantification of uncertainty of geometallurgical variables for mine planning optimisation*. Thesis (Doctor of Philosophy, in the School of Civil, Environmental and Mining Engineering), University of Adelaide, 180p., 2018.

SHWARTZ, S. S.; DAVID, S. B. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge: Cambridge University Press, 449p., 2014.

SINCLAIR, A. J.; BLACKWELL, G. H. *Applied Mineral Inventory Estimation*. Cambridge: Cambridge University Press, 381p., 2002.

SOARES, A. Direct sequential simulation and cosimulation. *Mathematical Geology*, v. 33, n. 8, p. 911-926, 2001.

SOARES, A. *Geoestatística para ciências da Terra e do ambiente*. Instituto superior técnico. Lisboa: Editora IST Press, 214p., 2006.

TAN, P.; MICHAEL, S.; KUMAR, V. *Introduction to data mining*. New York: Pearson Education, 2006.

- TEIXEIRA, M. F. L. *Estudo de caracterização e desenvolvimento tecnológico dos materiais duros cimentados da Mina de Fosfato da Copebrás, GO*. Dissertação (Mestrado), Universidade Federal de Minas Gerais, 2019.
- TOLEDO, M. M. C., PEREIRA, V. P. A variabilidade de composição da apatita associada a carbonatitos. *Revista do Instituto Geológico*, São Paulo, 22 (1/2), p. 27-64, 2001.
- TRIOLA, M. F. *Essentials of statistics*. London: Pearson, 696 p., 2008.
- ULIANA, D. *Caracterização tecnológica do minério de fosfato do complexo alcalino de Salitre, MG-área Fوسفertil*. Tese (Doutorado), Universidade de São Paulo, 2010.
- VAPNIK, V. N. *The Nature of Statistical Learning Theory*. Berlim: Springer-Verlag, 1995.
- VENDRAMIN, L.; CAMPELLO, R. J.; HRUSCHKA, E. R. Relative clustering validity criteria: A comparative overview. *Statistical analysis and data mining: the ASA data science journal*, 3(4), p. 209-235, 2010.
- VIEIRA, M. C. A. *Metodologia para prever recuperação de zinco em planta de beneficiamento*. 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.
- WELCOX, R. R. *Applying Contemporary Statistical Techniques*. Cambridge: Academic Press, 2003.
- WILLIAMS, S. R.; RICHARDSON, J. M. Geometallurgical Mapping: A New Approach That Reduces Technical Risk. *Proceedings 36th Annual Meeting of the Canadian Mineral Processors*, Paper 16, 28p., 2004.
- WILLIAMS, S. R. A Historical perspective of the application and success of geometallurgical methodologies. In: *The second AUSIMM International geometallurgy conference, Brisbane, QLD*: Australasian Institute of Mining and Metallurgy, 2013.
- WOOLLEY, A.R.; KEMPE, D. R. C. Carbonatites: nomenclature, average chemical compositions, and element distribution. In: BELL, K. (Ed.). *Carbonatites: genesis and evolution*. London: Unwin Hyman, p. 1-14, 1989.
- XU, R.; WUNSCH, D. Survey of clustering algorithms. *IEEE Transactions on neural networks*, 16(3), p. 645-678. 2005.
- YAMAMOTO, J. K.; LANDIM, P. M. B. *Geoestatística: conceitos e aplicações*. São Paulo: Oficina de textos, 215p., 2013.
- YASOJIMA, C. T. K. *Modelo de Krigagem Automática Baseada em Agrupamento*. Tese (Doutorado em Ciências da Computação), Universidade Federal do Pará, 86p., 2020.
- ZADEH, L. A. Fuzzy sets as a basis for a theory of possibility. *Fuzzy sets and systems*, 1(1), p. 3-28, 1978.

ZHANG,

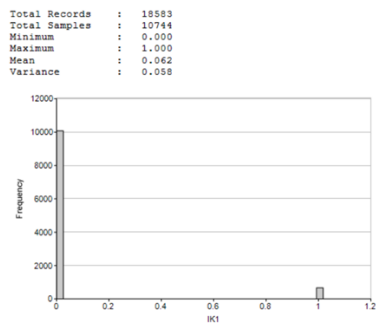
S. E.; NWAILA, G. T.; BOURDEAU, J. E.; GHORBANI, Y.; CARRANZA, E. J. M. Machine learning-based delineation of geodomain boundaries: A proof-of-concept study using data from the Witwatersrand Goldfields. *Natural Resources Research*, 32(3), p. 879-900, 2023.

ZIMMERMANN, H. J. *Fuzzy set theory and its applications*. 3rd ed. Boston/Dordrecht/London: Kluwer Academic Publishers, 429 p., 1996.

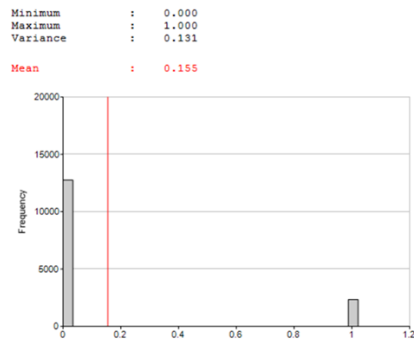
APÊNDICE

APÊNDICE A – Histograma dos indicadores das tipologias do banco de dados.

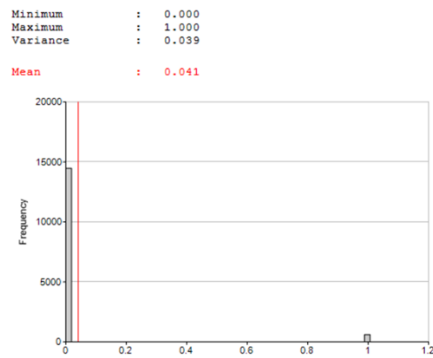
Histograma indicador 1- CB



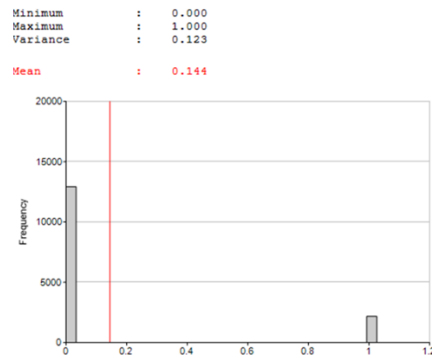
Histograma indicador 2- FO



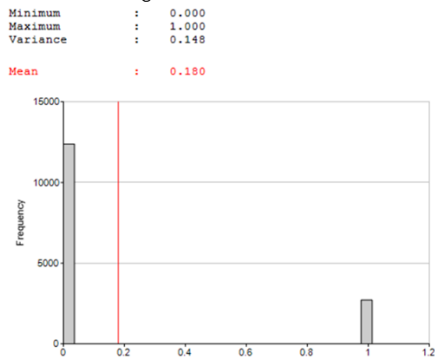
Histograma indicador 3- NL



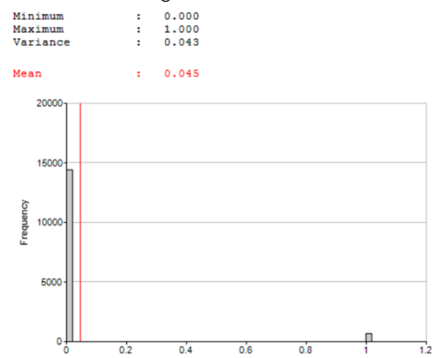
Histograma indicador 4- AL



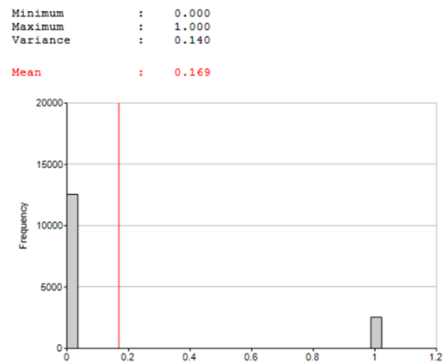
Histograma indicador 5- FL



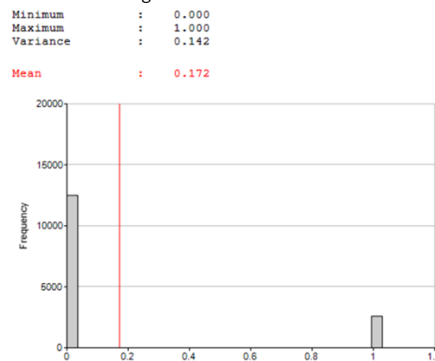
Histograma indicador 6- PI



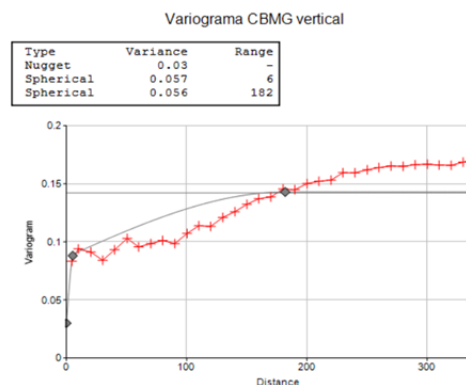
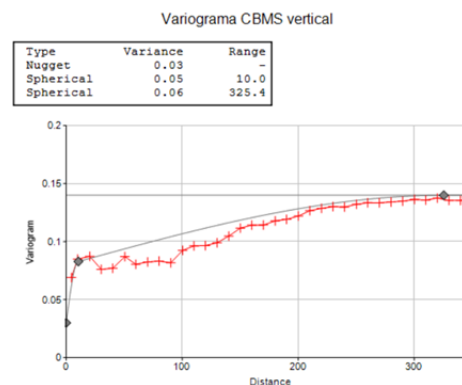
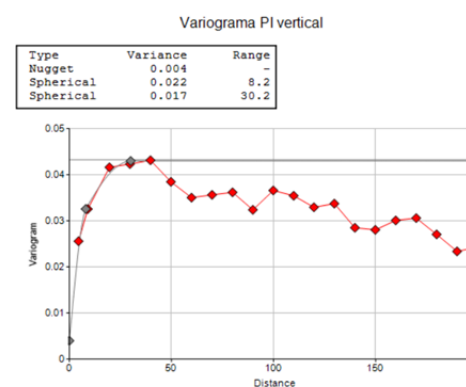
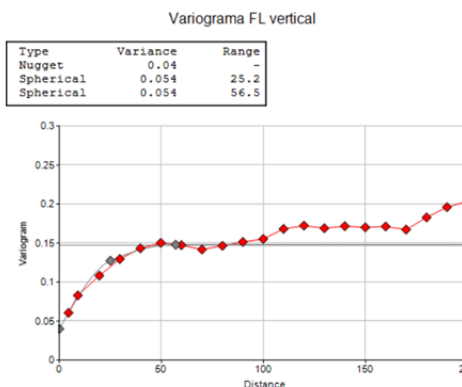
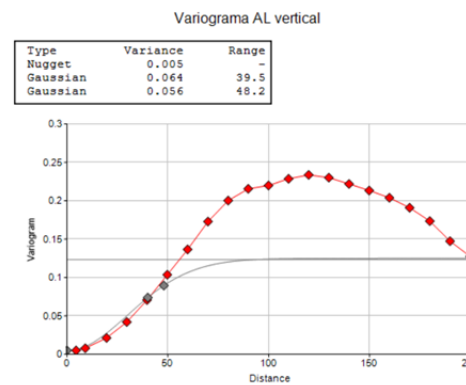
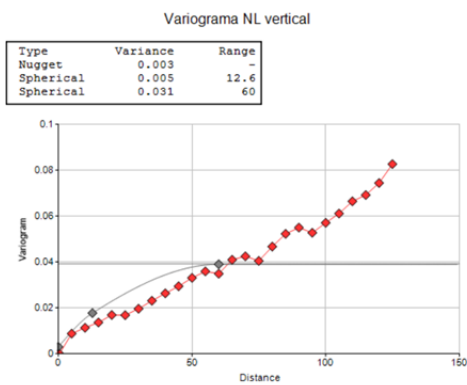
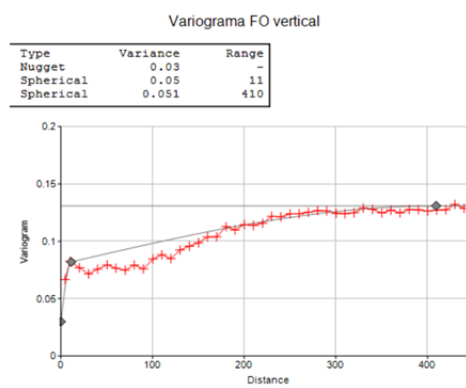
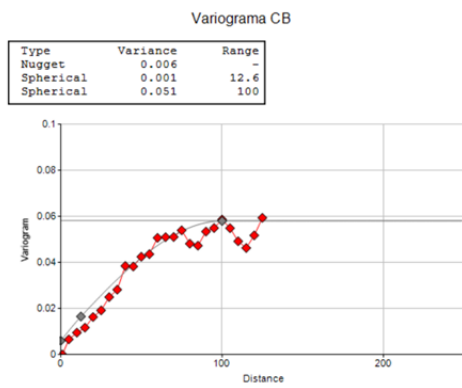
Histograma indicador 7- CBMS



Histograma indicador 8- CBMG



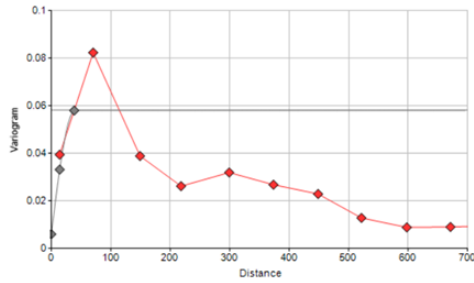
APÊNDICE B – Variogramas verticais dos indicadores das tipologias do banco de dados.



APÊNDICE C - Variogramas azimute 0° dos indicadores das tipologias.

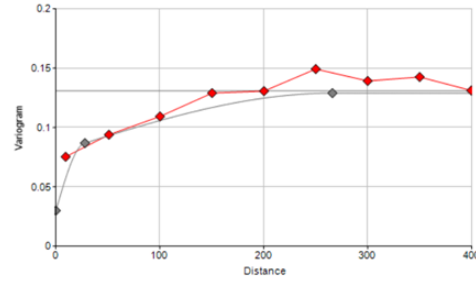
Variograma CB

Type	Variance	Range
Nugget	0.006	-
Spherical	0.001	12.6
Spherical	0.051	39.2



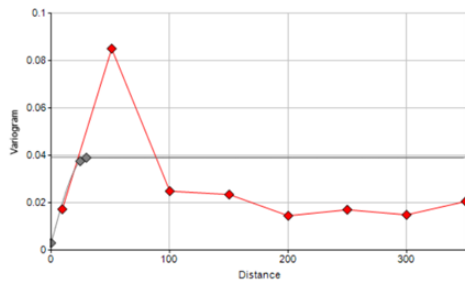
Variograma FO AZ 0°

Type	Variance	Range
Nugget	0.03	-
Spherical	0.049	27
Spherical	0.05	265.9



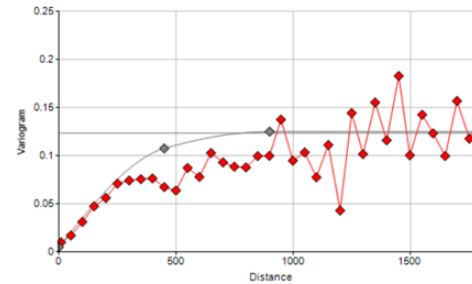
Variograma NL AZ 0°

Type	Variance	Range
Nugget	0.003	-
Spherical	0.005	25
Spherical	0.031	30



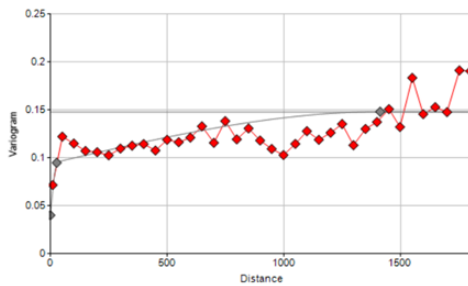
Variograma AL AZ 0°

Type	Variance	Range
Nugget	0.005	-
Spherical	0.064	450.0
Spherical	0.056	900.0



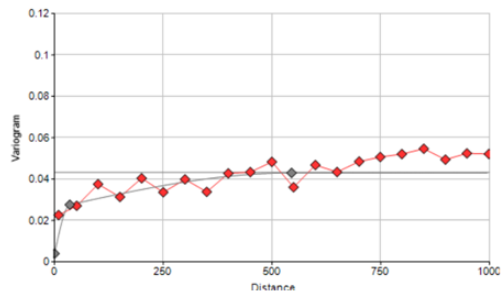
Variograma FL AZ 0°

Type	Variance	Range
Nugget	0.04	-
Spherical	0.054	30
Spherical	0.054	1417



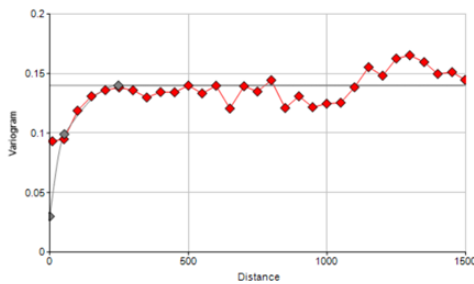
Variograma PIAZ 0°

Type	Variance	Range
Nugget	0.004	-
Spherical	0.022	35
Spherical	0.017	545



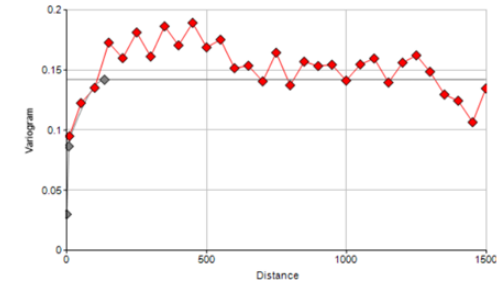
Variograma CBMS AZ 0°

Type	Variance	Range
Nugget	0.03	-
Spherical	0.05	52
Spherical	0.060	244



Variograma CBMG AZ 0°

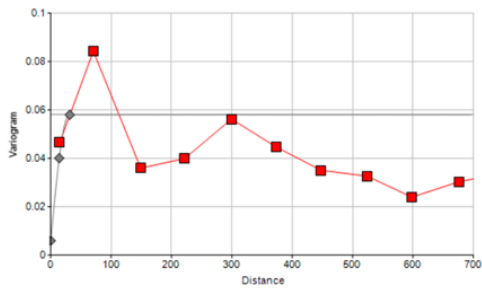
Type	Variance	Range
Nugget	0.03	-
Spherical	0.057	10
Spherical	0.055	137



APÊNDICE D - Variogramas azimute 90° dos indicadores das tipologias.

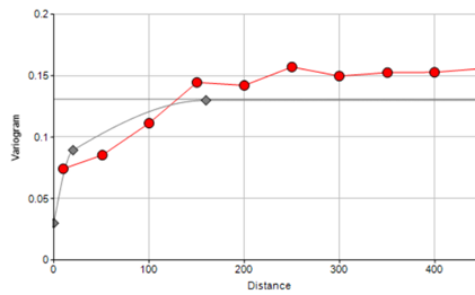
Variograma CB

Type	Variance	Range
Nugget	0.006	-
Spherical	0.001	12.6
Spherical	0.051	30



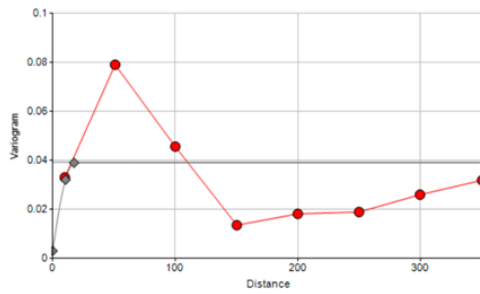
Variograma FO AZ 90°

Type	Variance	Range
Nugget	0.03	-
Spherical	0.050	20
Spherical	0.05	160



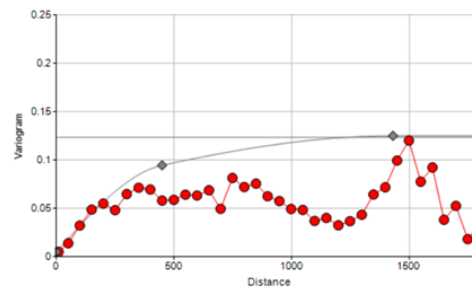
Variograma NL AZ 90°

Type	Variance	Range
Nugget	0.003	-
Spherical	0.005	9.9
Spherical	0.031	18



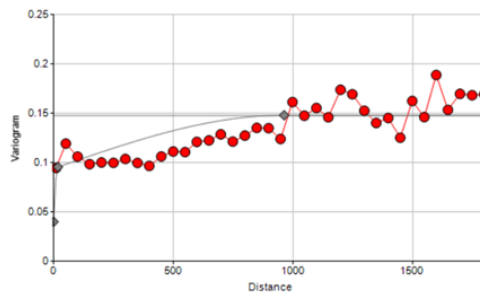
Variograma AL AZ 90°

Type	Variance	Range
Nugget	0.005	-
Spherical	0.064	450.0
Spherical	0.056	1431



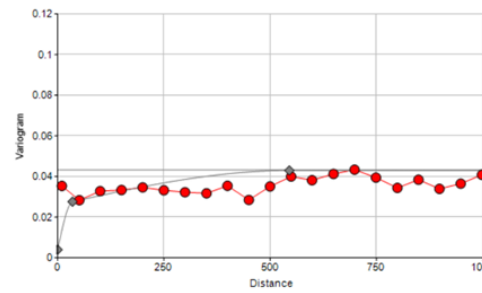
Variograma FL AZ 90°

Type	Variance	Range
Nugget	0.04	-
Spherical	0.054	15
Spherical	0.054	959.4



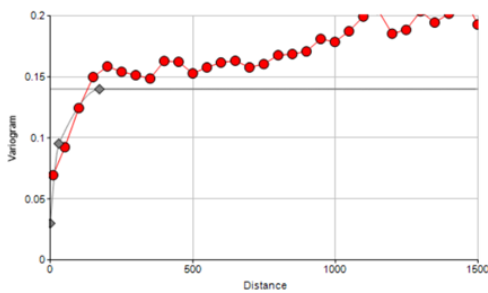
Variograma PI AZ 90°

Type	Variance	Range
Nugget	0.004	-
Spherical	0.022	35
Spherical	0.017	545



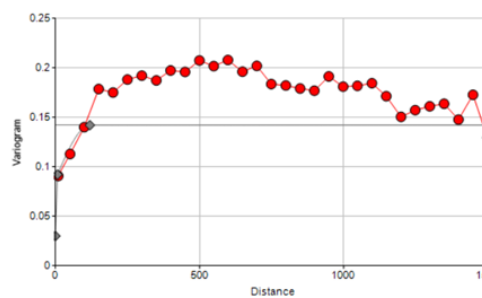
Variograma CBMS AZ 90°

Type	Variance	Range
Nugget	0.03	-
Spherical	0.05	30
Spherical	0.060	174

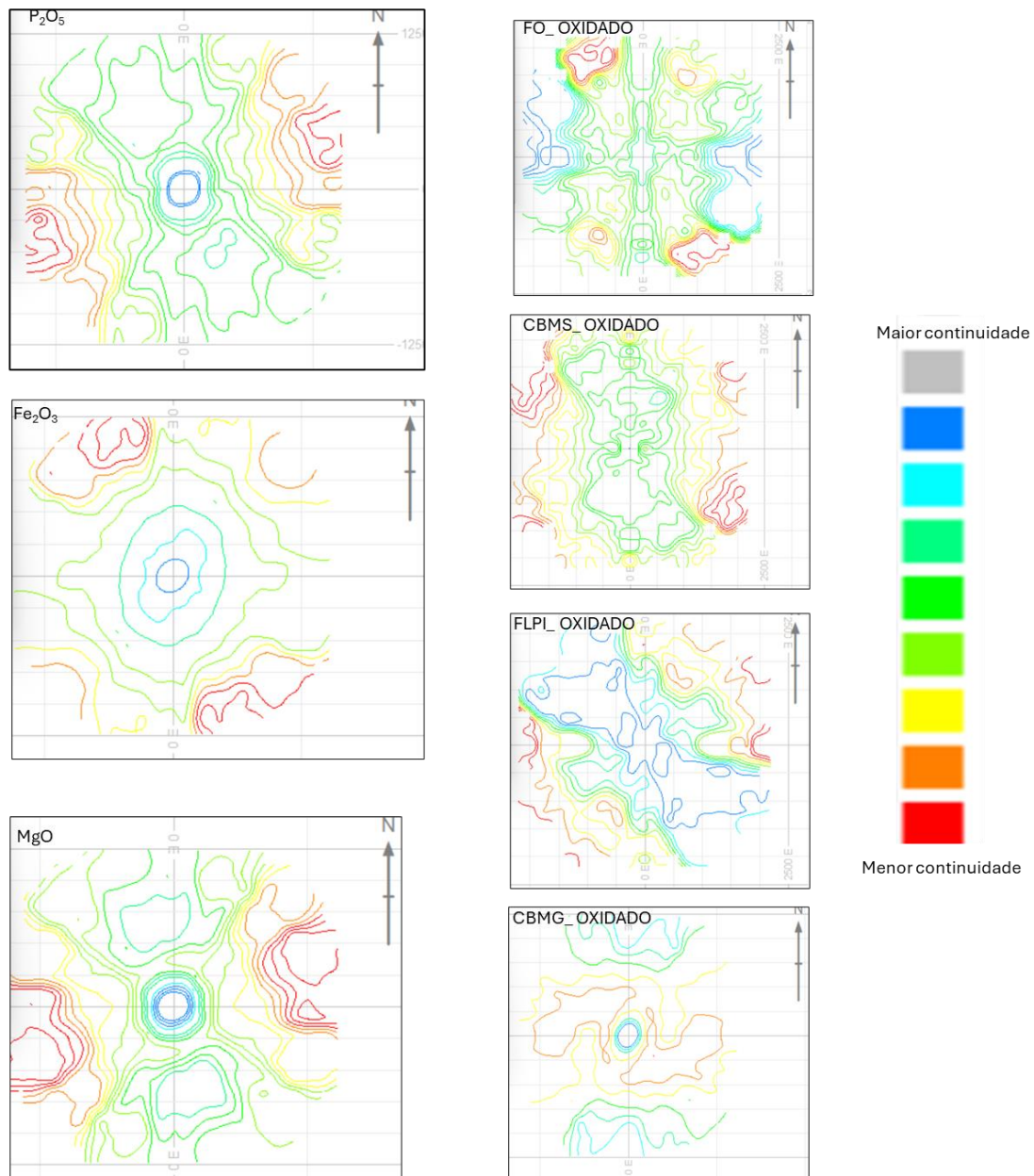


Variograma CBMG AZ 90°

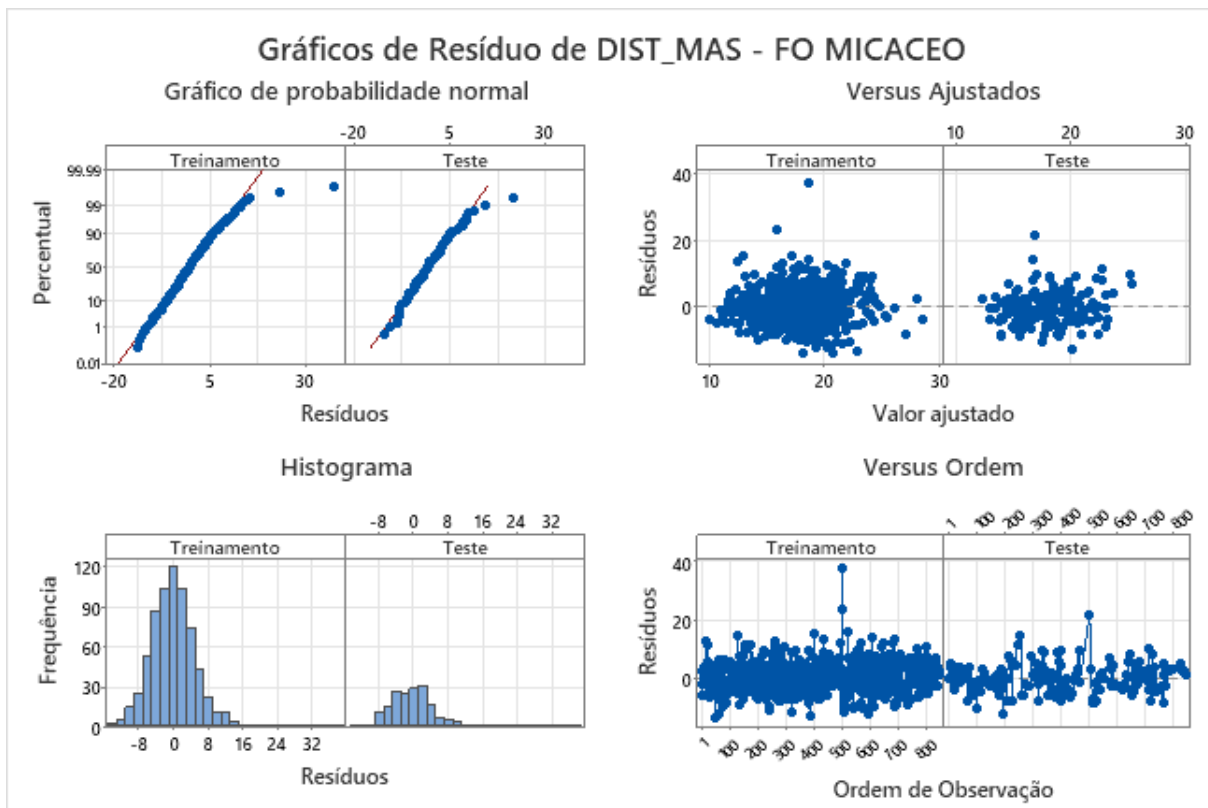
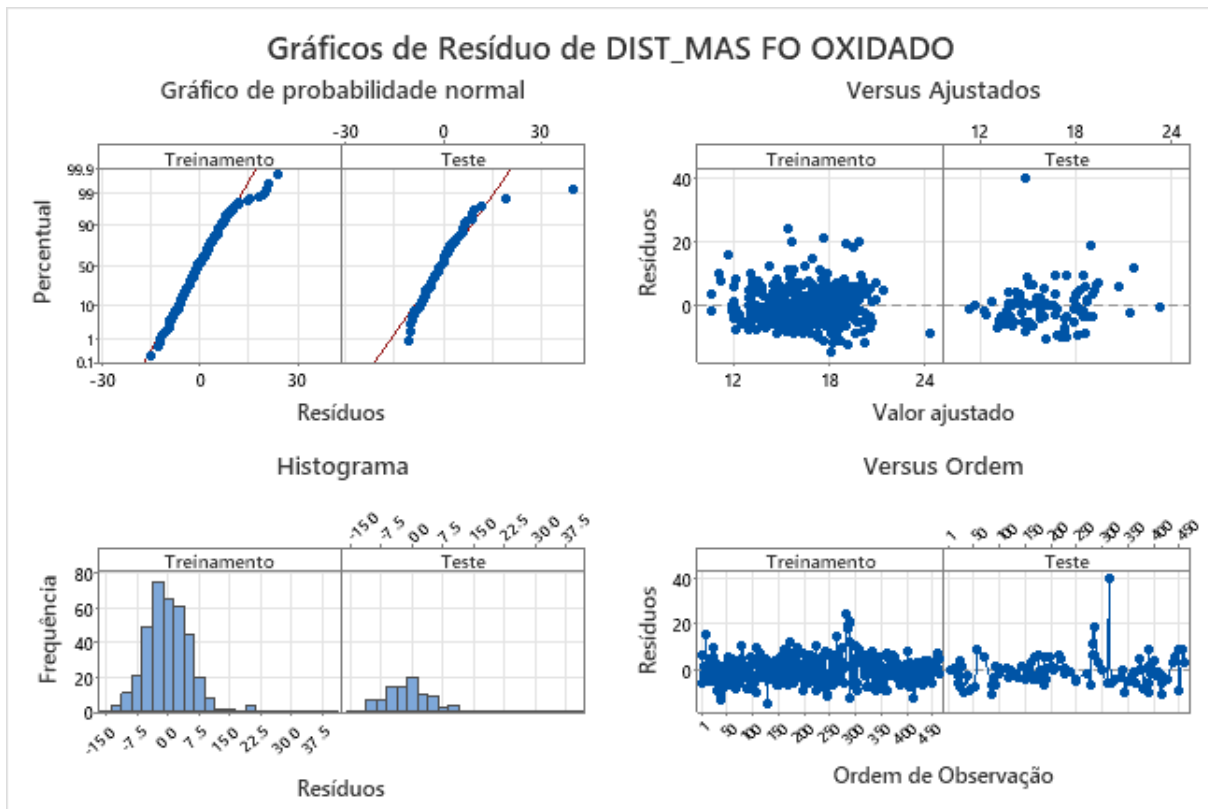
Type	Variance	Range
Nugget	0.03	-
Spherical	0.057	6
Spherical	0.055	120



APÊNDICE E - Mapa variograma à esquerda para os teores e à direita para os indicadores dos principais domínios nas zonas mineralizadas.



APÊNDICE F - Gráfico de resíduos da regressão linear.



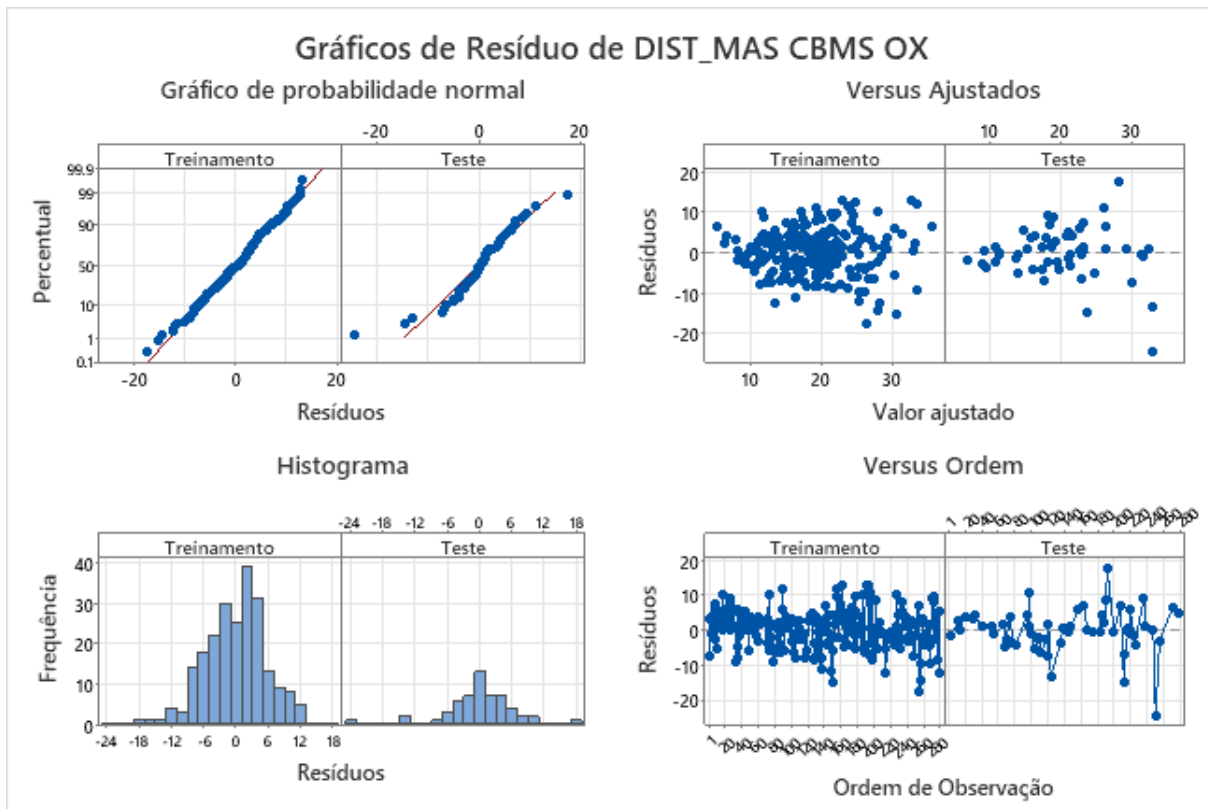
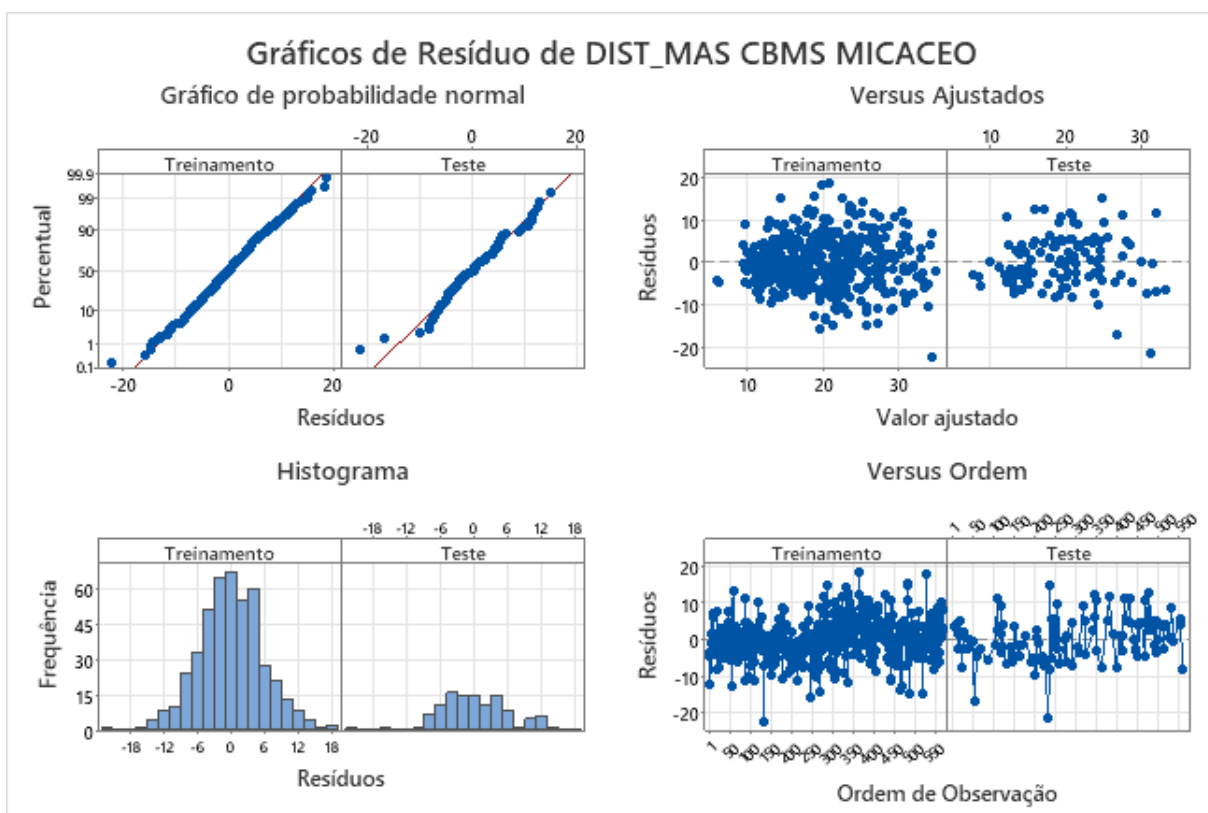


Figura 8. Gráfico de resíduos da regressão linear.



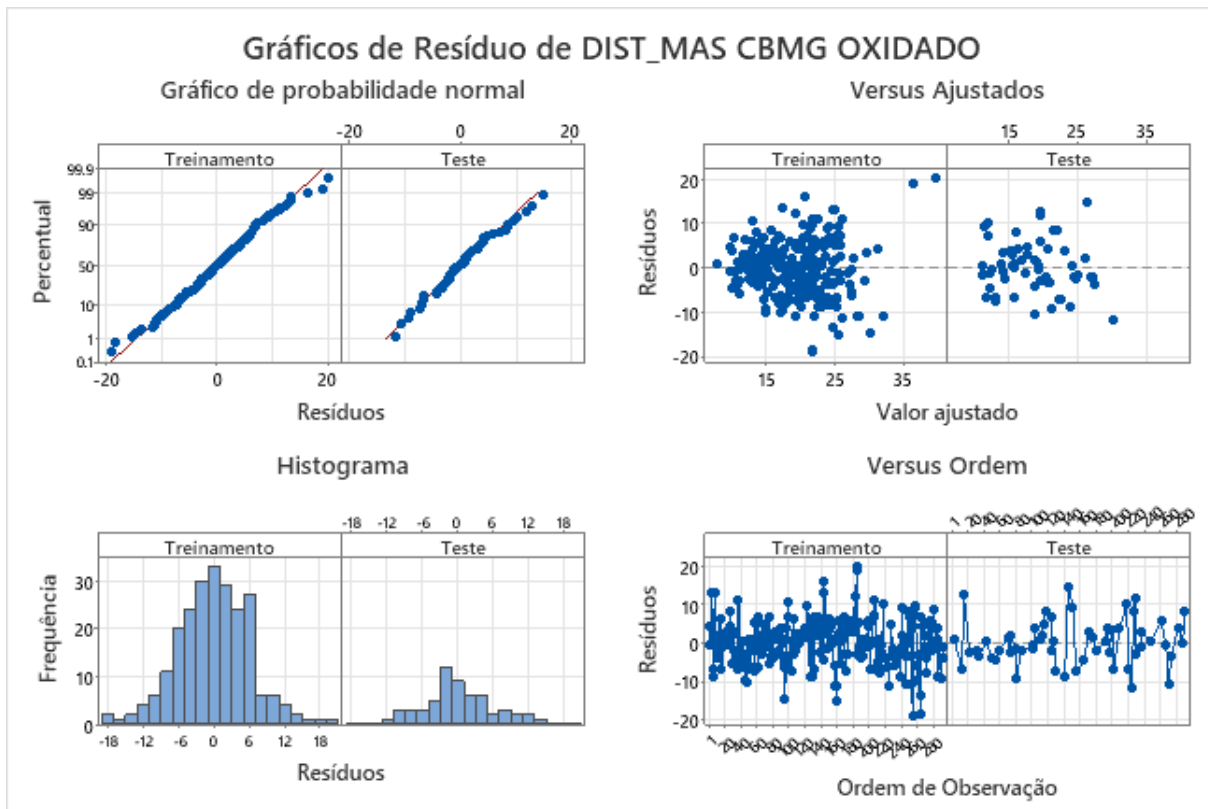
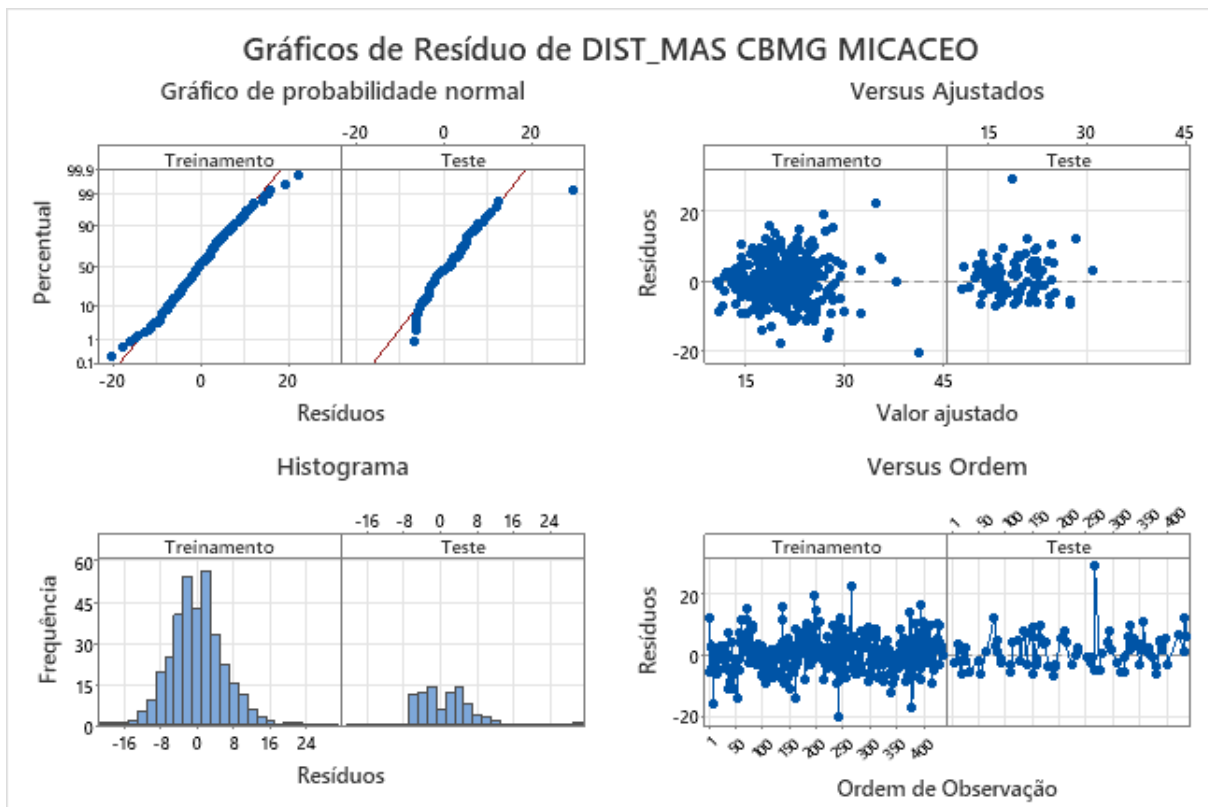
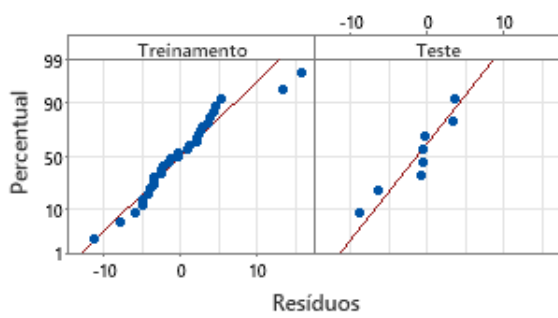


Figura 10. Gráfico de resíduos da regressão linear.

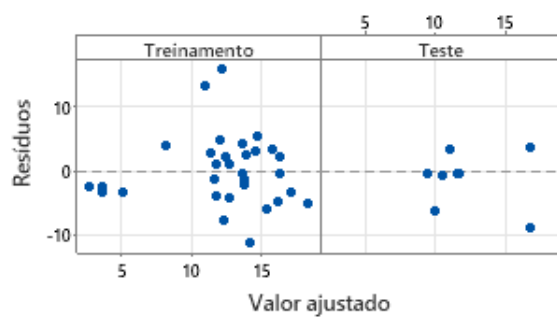


Gráficos de Resíduo de DIST_MAS FLPI OXIDADO

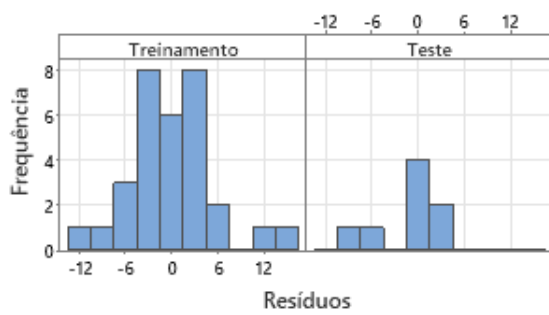
Gráfico de probabilidade normal



Versus Ajustados



Histograma

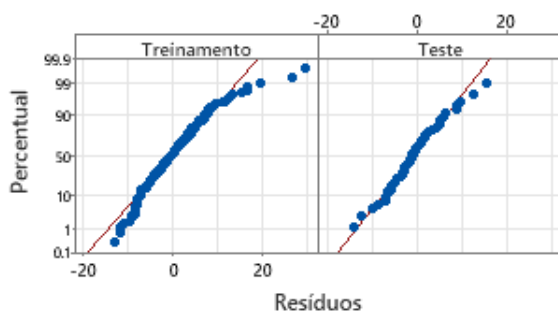


Versus Ordem



Gráficos de Resíduo de DIST_MAS FLPI MICACEO

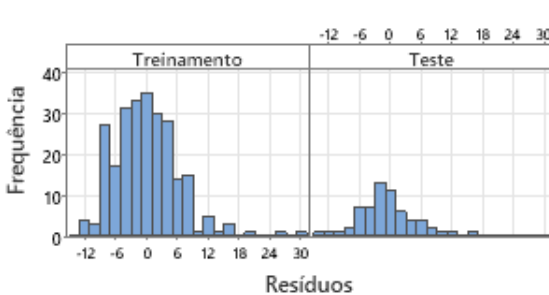
Gráfico de probabilidade normal



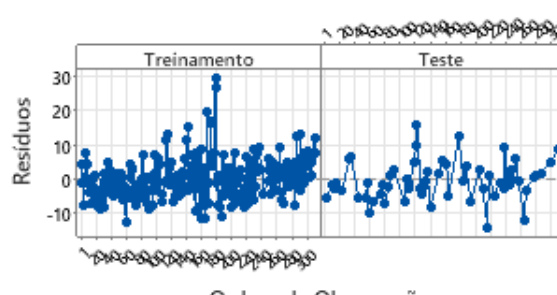
Versus Ajustados



Histograma

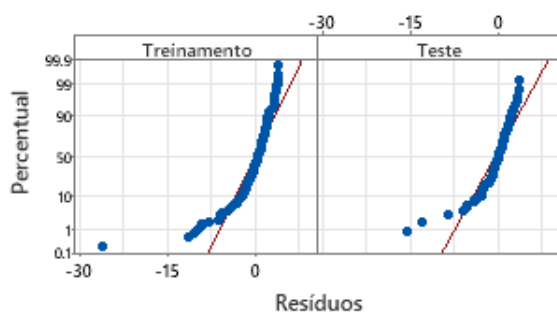


Versus Ordem

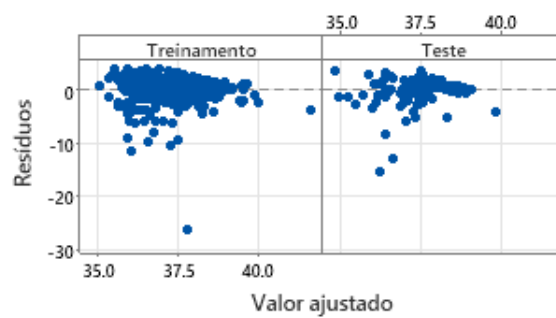


Gráficos de Resíduo de P2O5_C FO OXIDADO

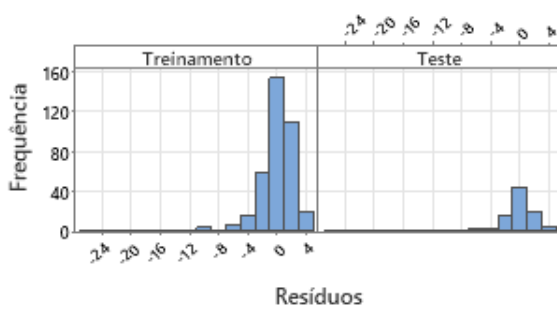
Gráfico de probabilidade normal



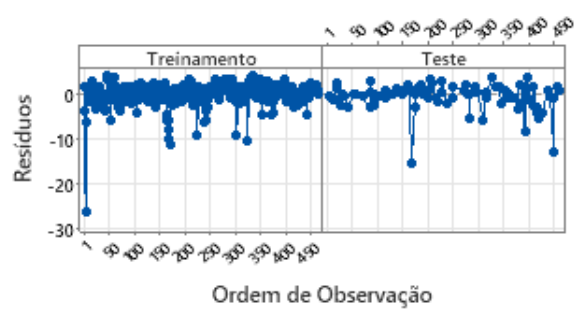
Versus Ajustados



Histograma

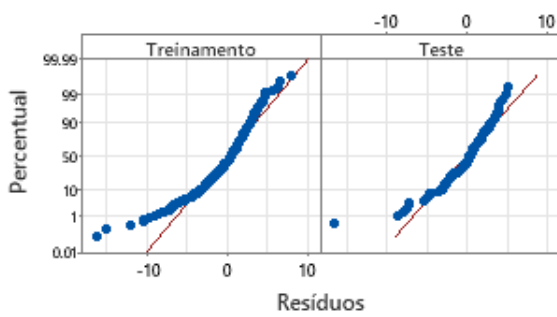


Versus Ordem

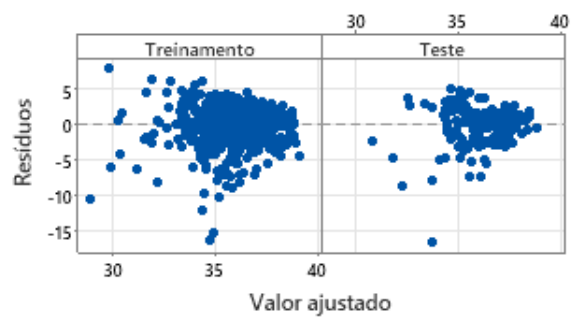


Gráficos de Resíduo de P2O5_C FO MICACEO

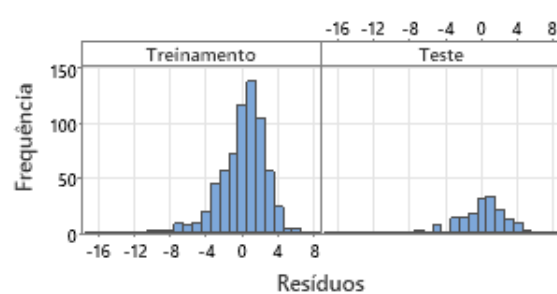
Gráfico de probabilidade normal



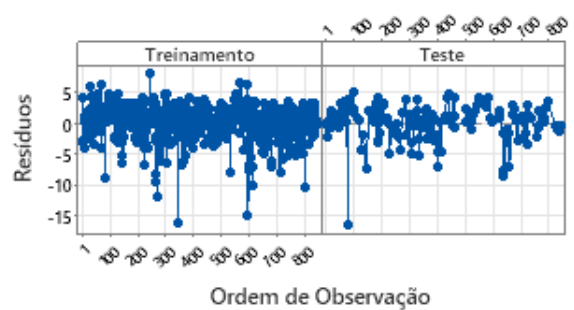
Versus Ajustados



Histograma

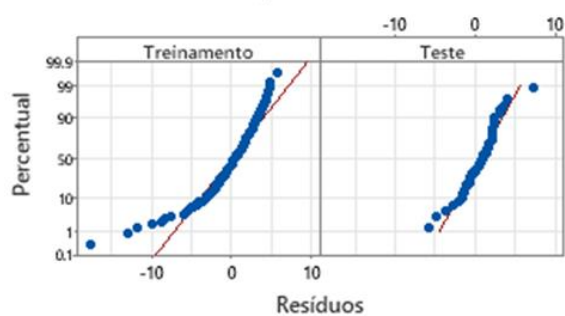


Versus Ordem

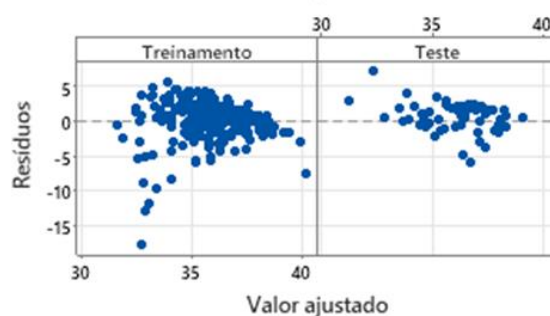


Gráficos de Resíduo de P2O5_C CBMS OXIDADO

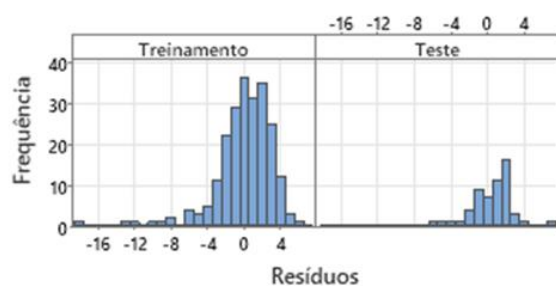
Gráfico de probabilidade normal



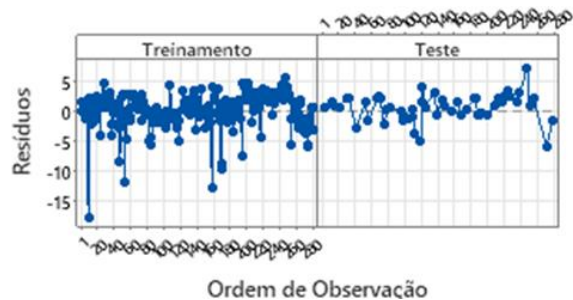
Versus Ajustados



Histograma

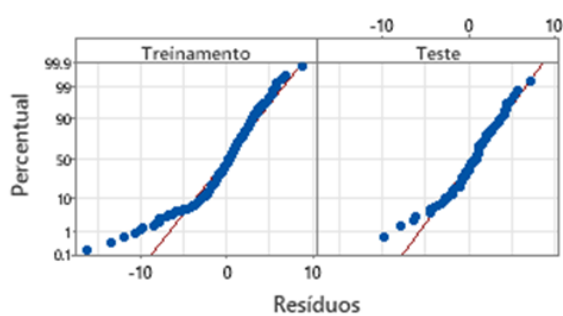


Versus Ordem

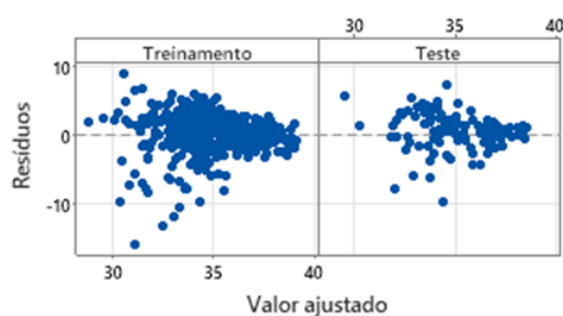


Gráficos de Resíduo de P2O5_C CBMS MICACEO

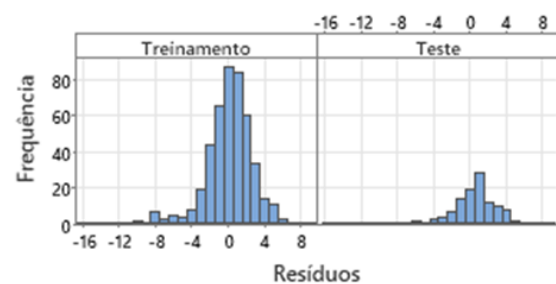
Gráfico de probabilidade normal



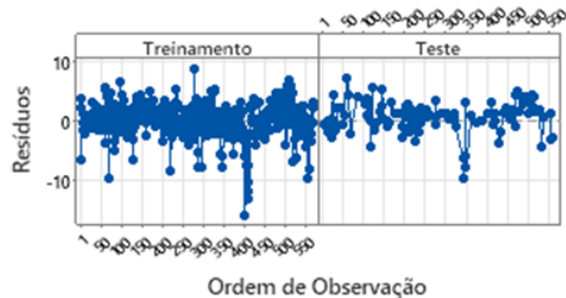
Versus Ajustados



Histograma



Versus Ordem



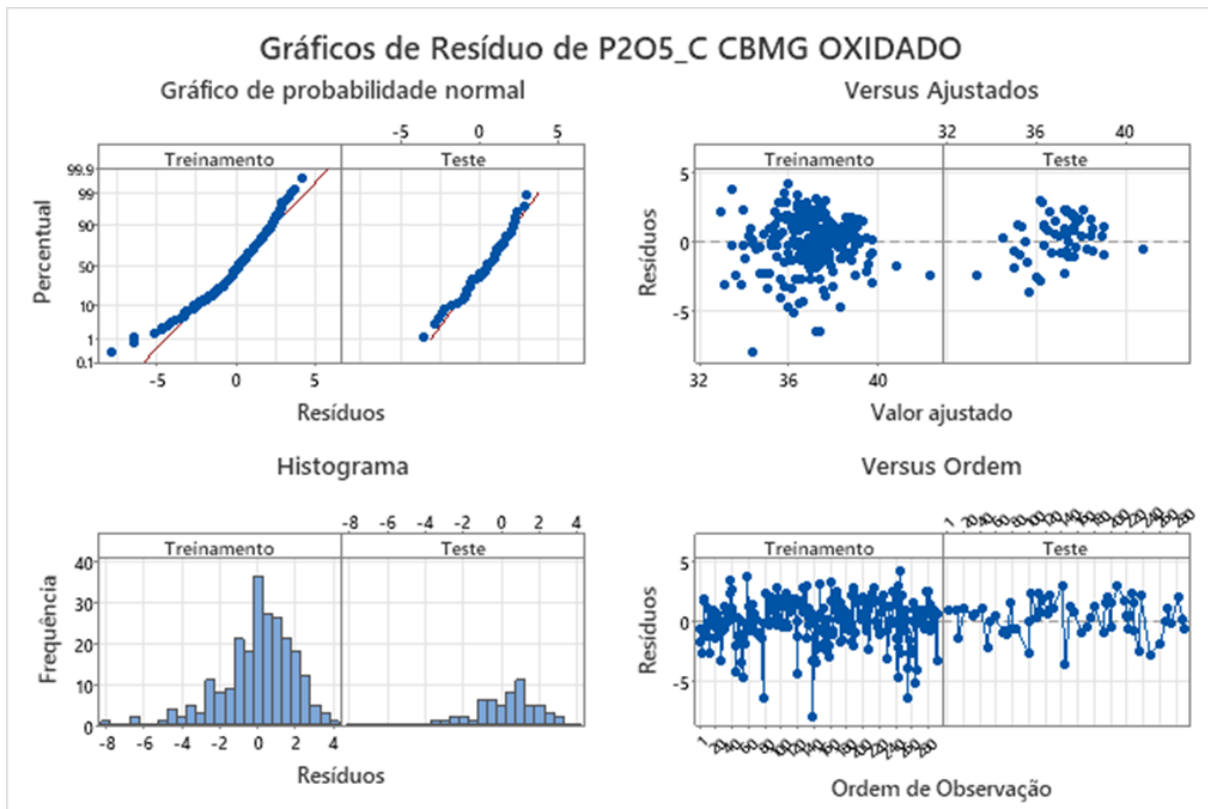


Figura 18. Gráfico de resíduos da regressão linear.

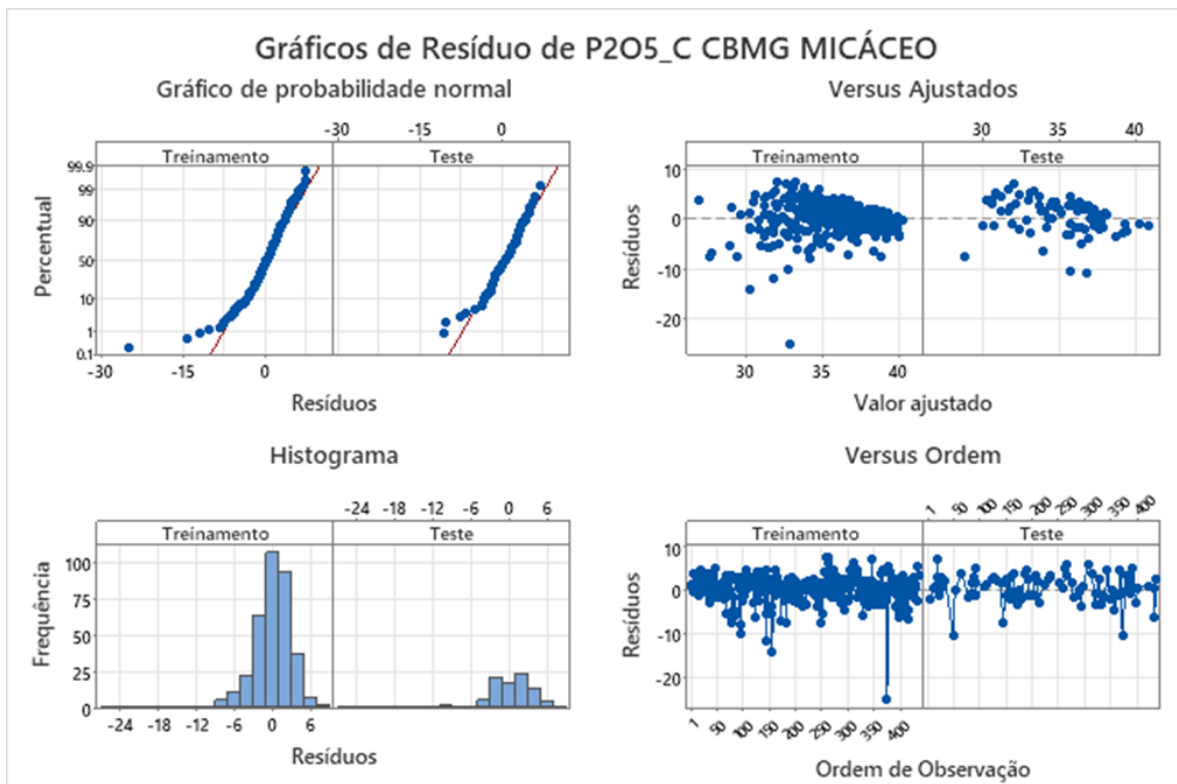


Figura 19. Gráfico de resíduos da regressão linear.

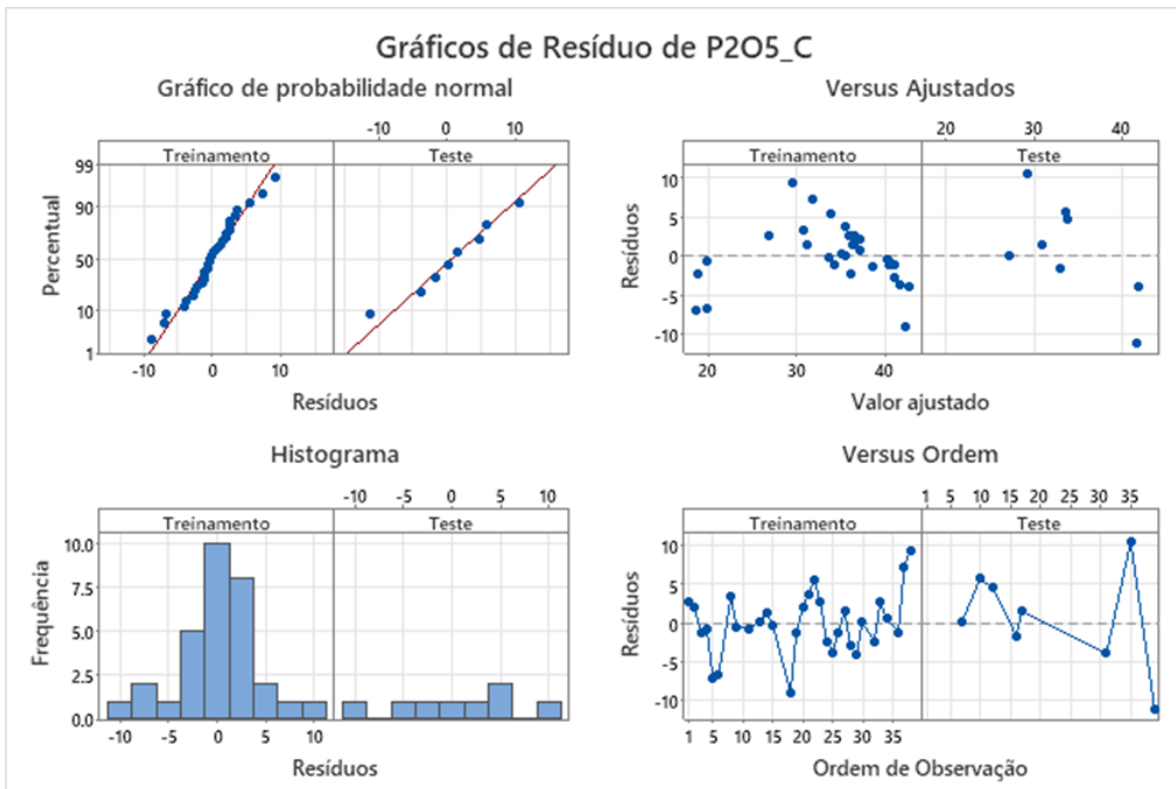


Figura 20. Gráfico de resíduos da regressão linear FLPI OXIDADO.

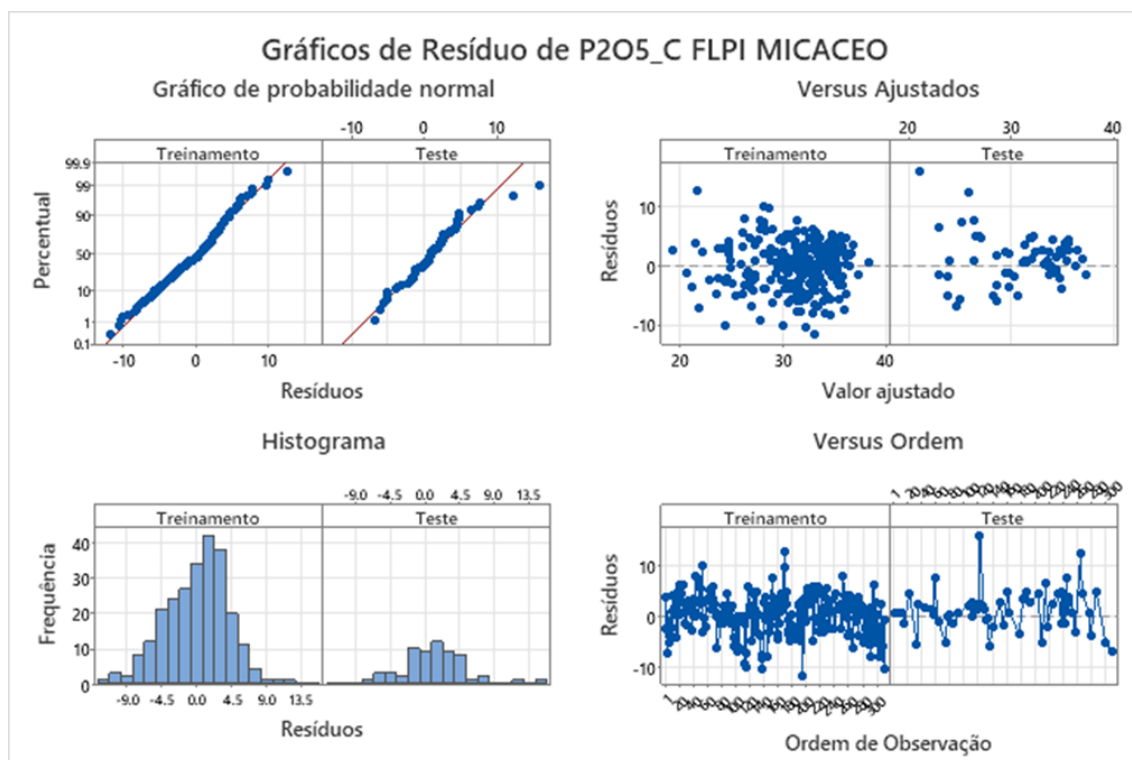


Figura 21. Gráfico de resíduos da regressão linear.

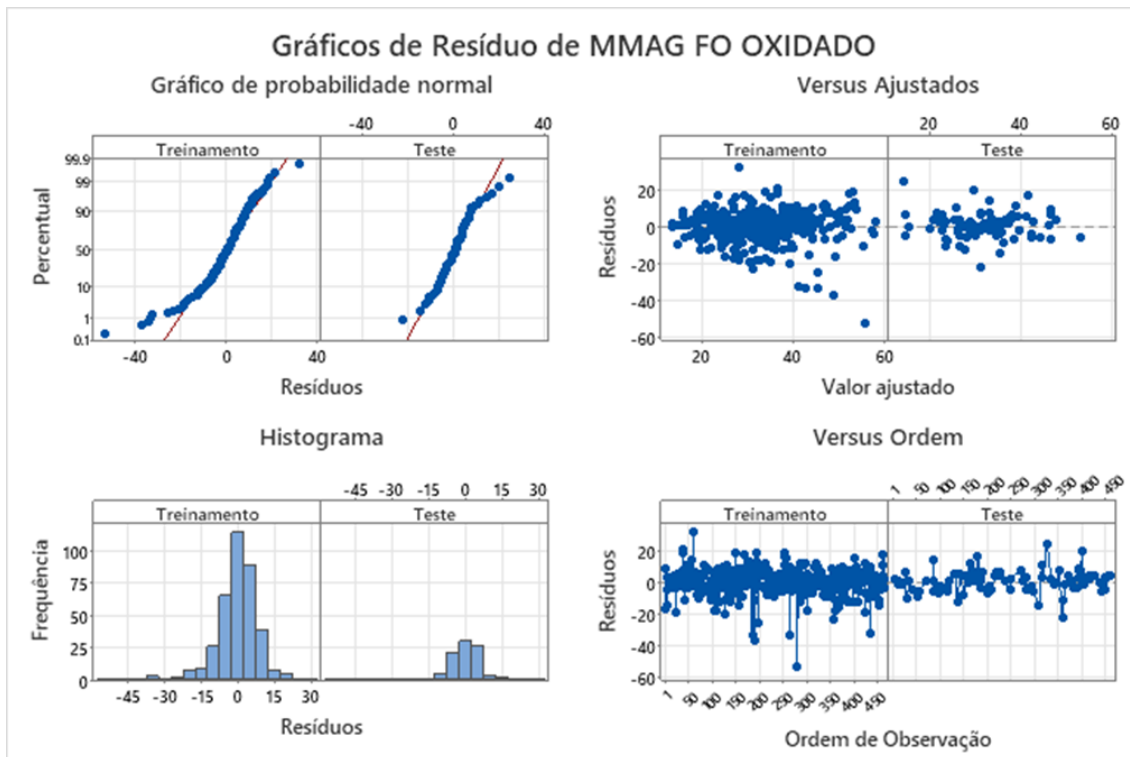


Figura 22. Gráfico de resíduos da regressão linear.

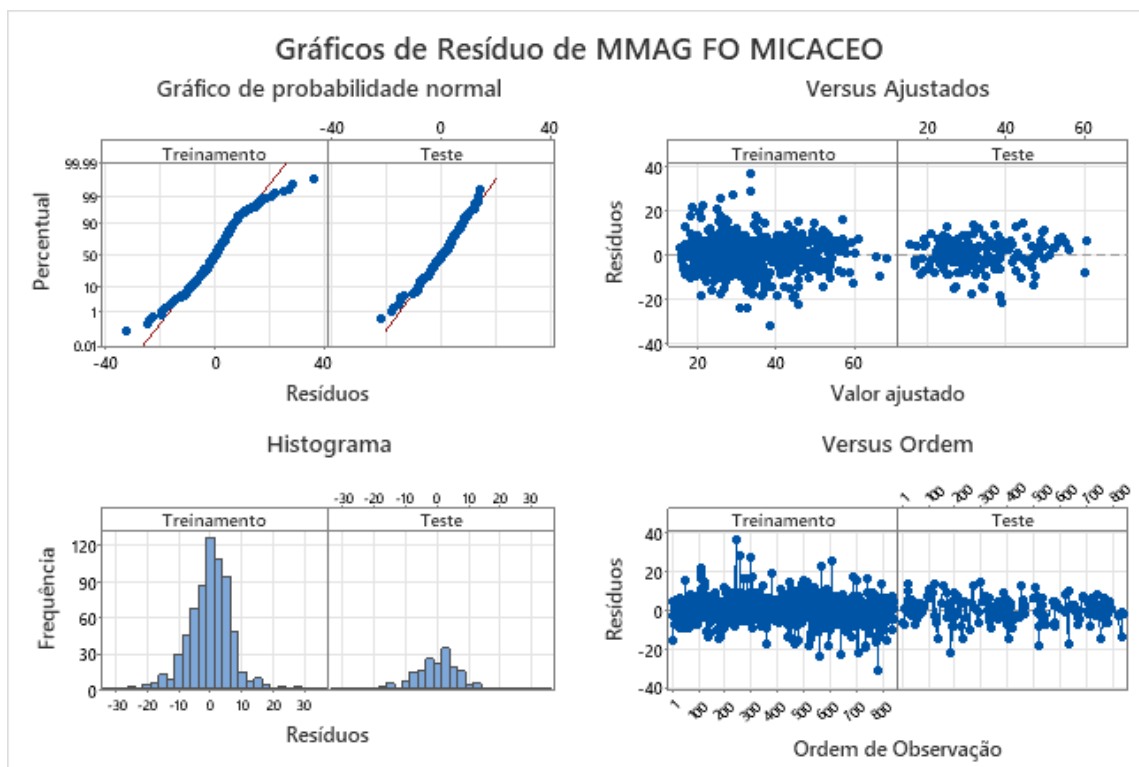


Figura 23. Gráfico de resíduos da regressão linear.

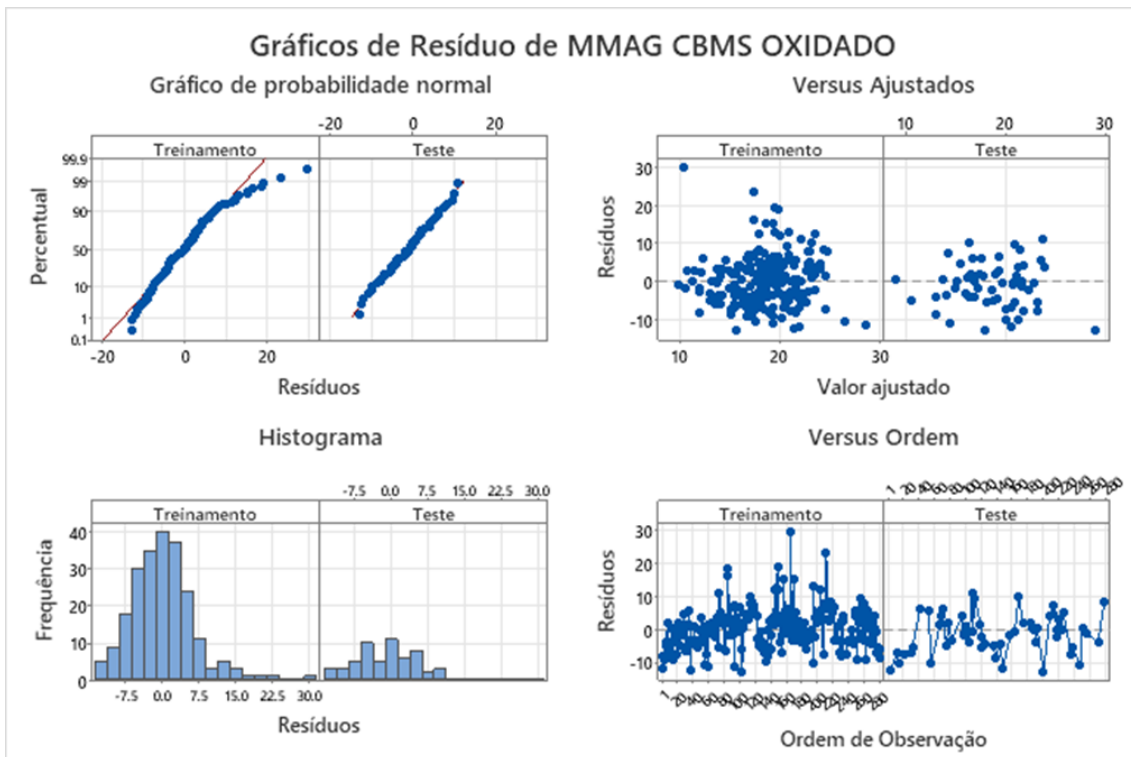


Figura 24. Gráfico de resíduos da regressão linear.

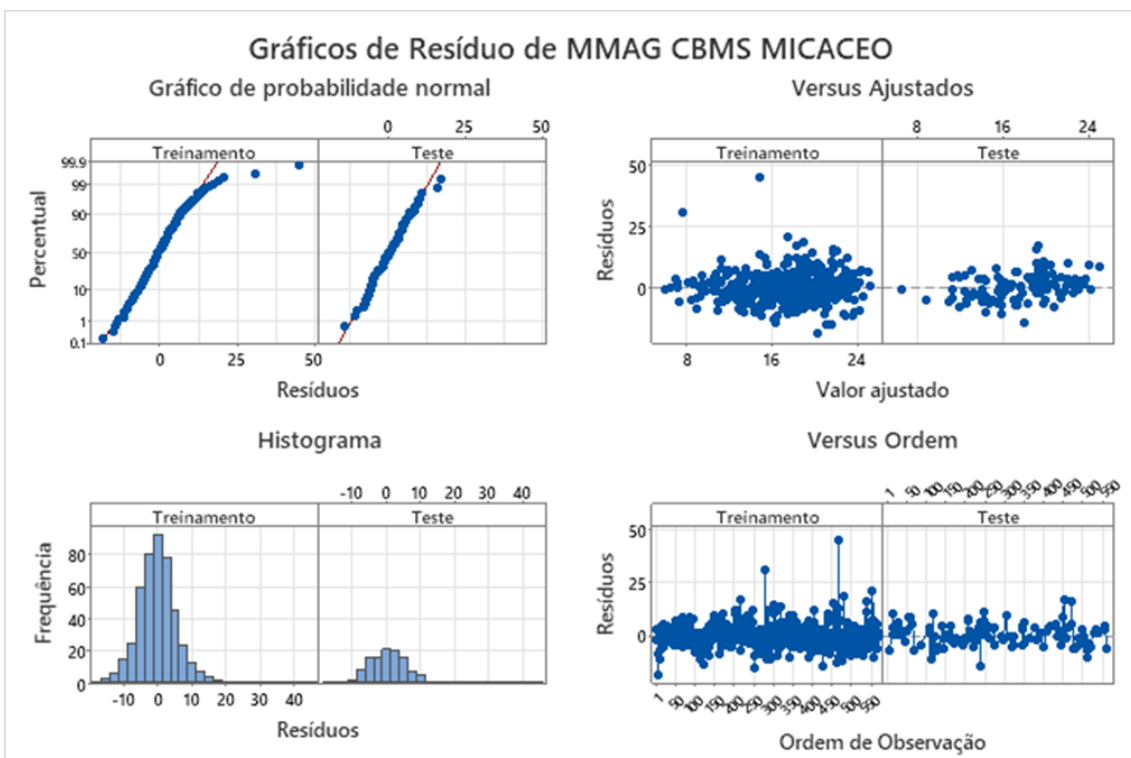


Figura 25. Gráfico de resíduos da regressão linear.

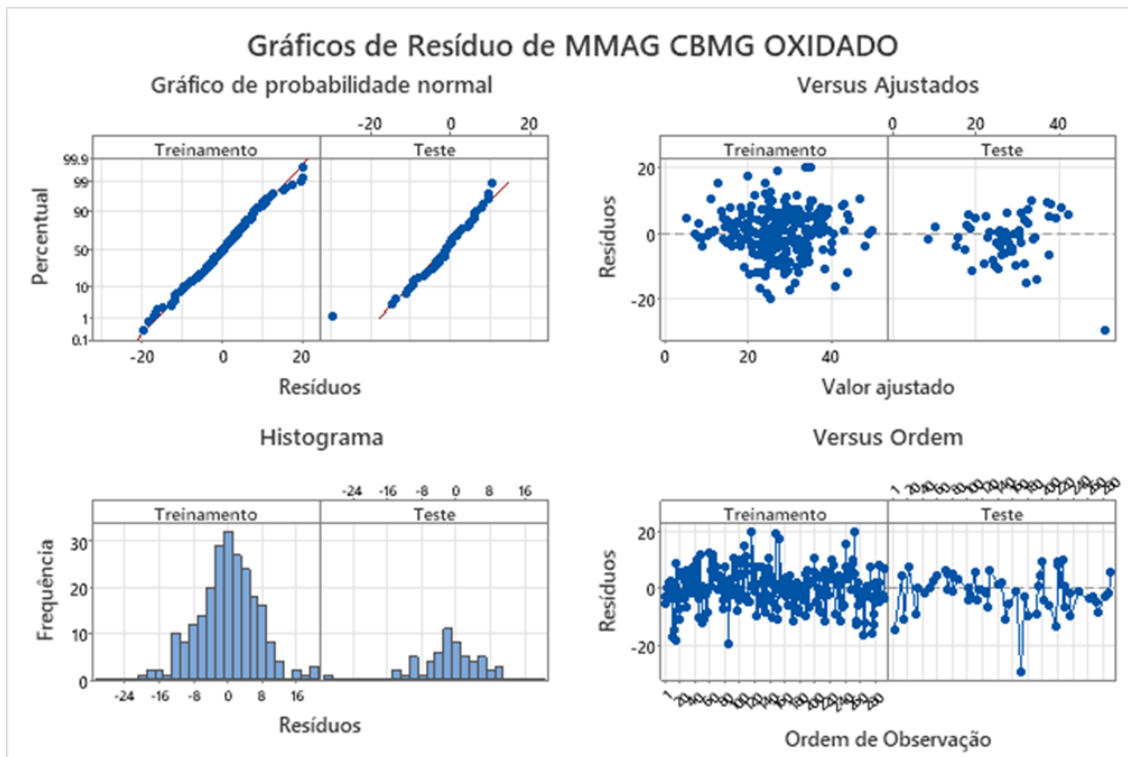


Figura 26. Gráfico de resíduos da regressão linear.

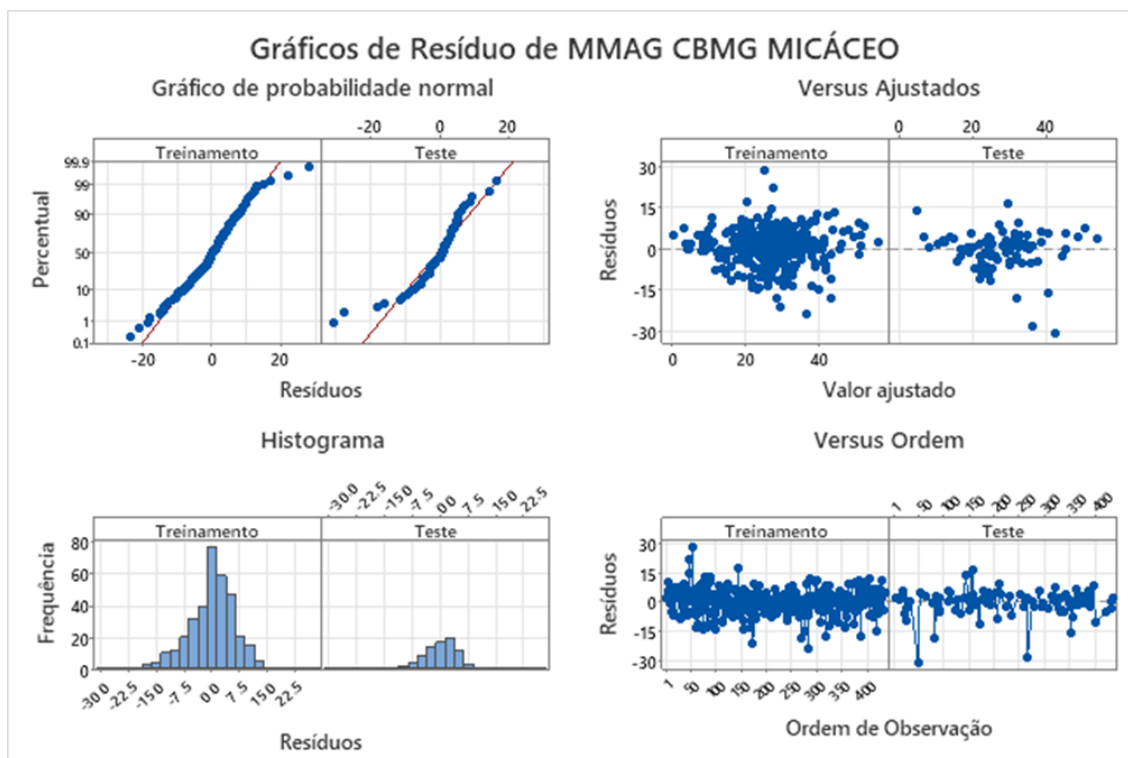


Figura 27. Gráfico de resíduos da regressão linear.

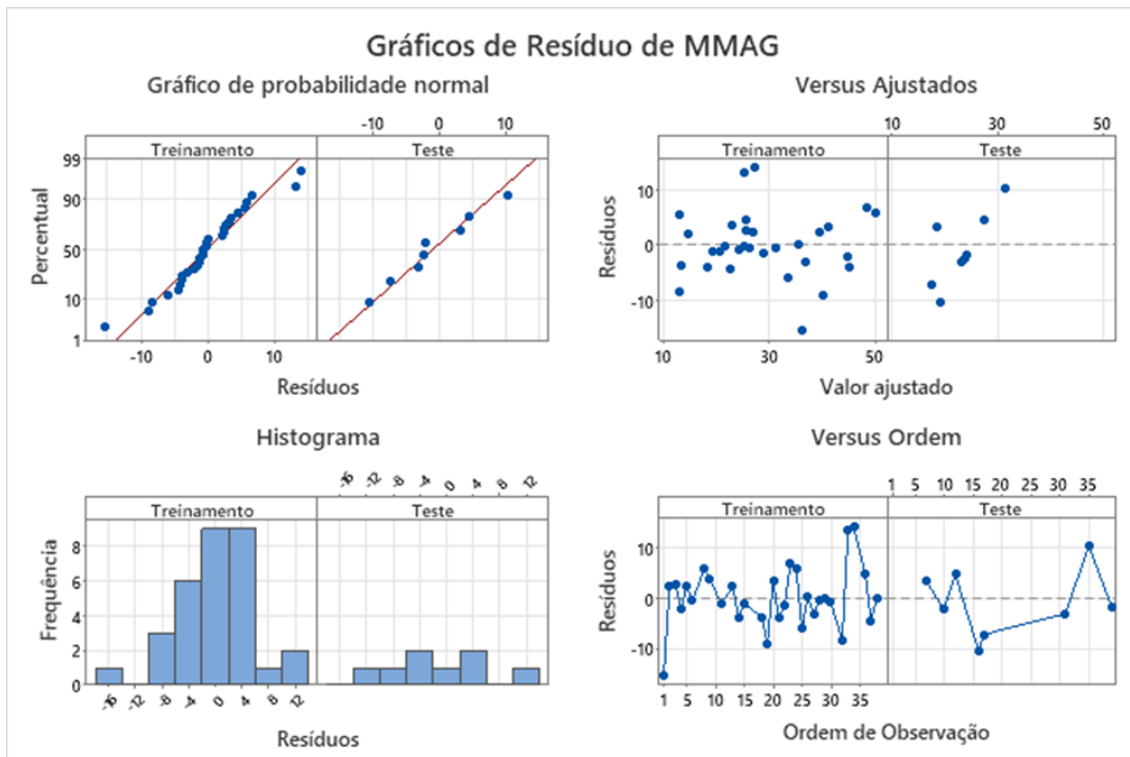


Figura 28. Gráfico de resíduos da regressão linear MMAG FLPI oxidado.

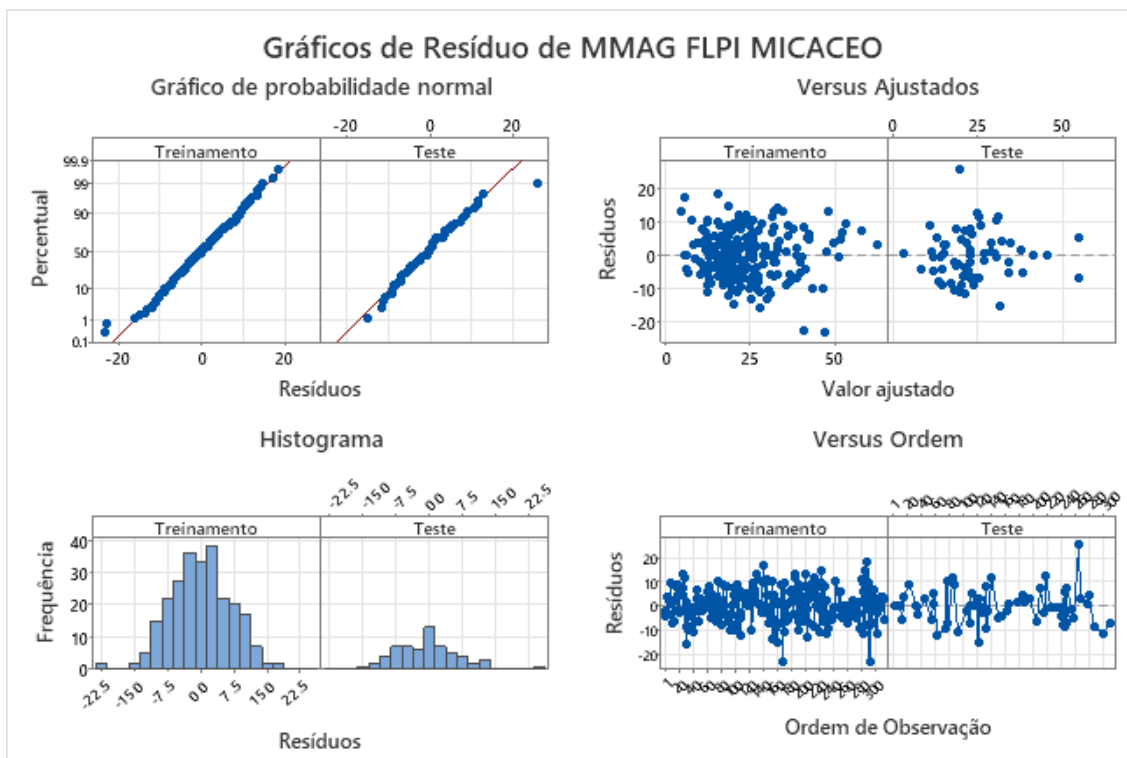
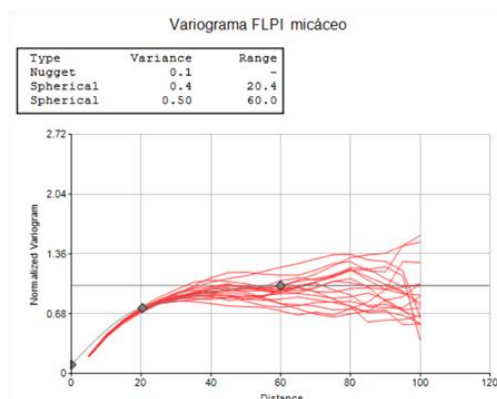
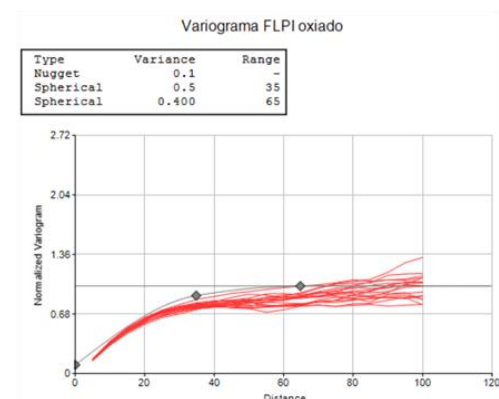
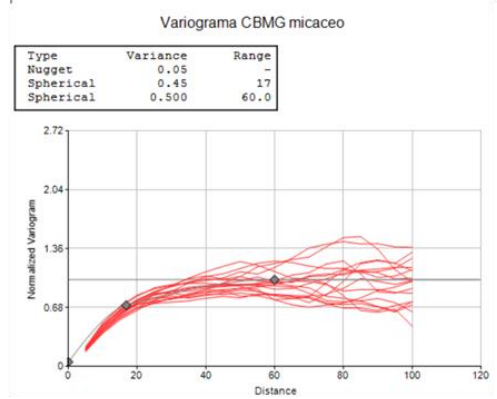
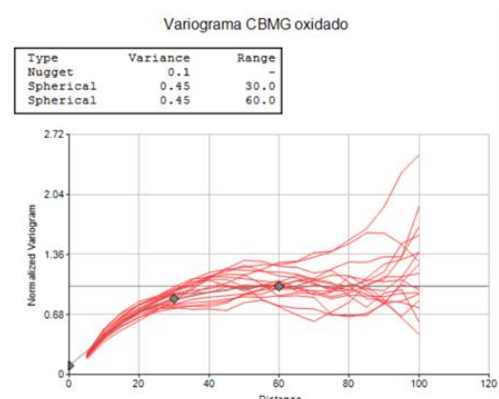
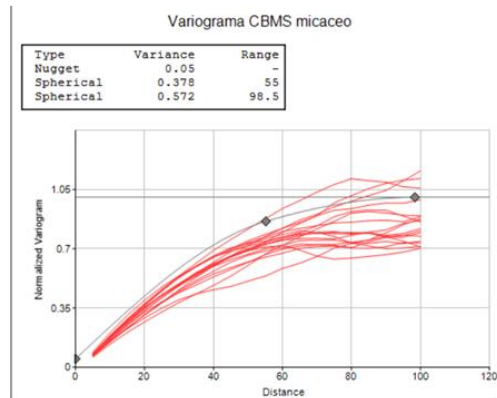
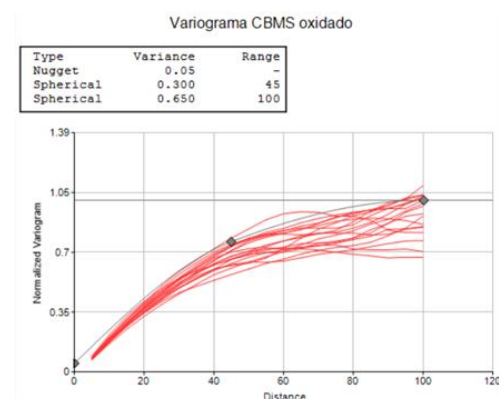
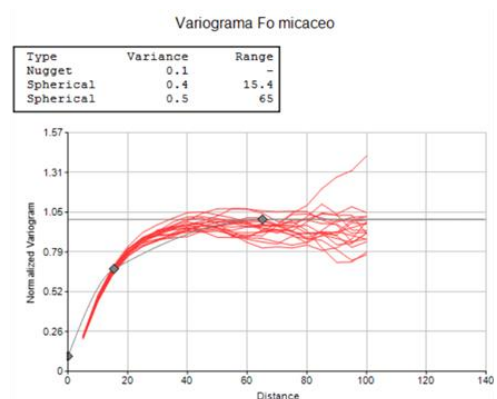
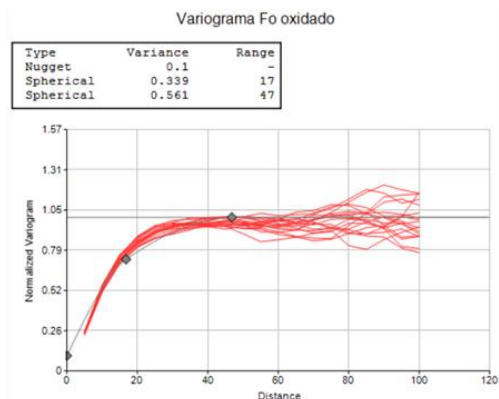


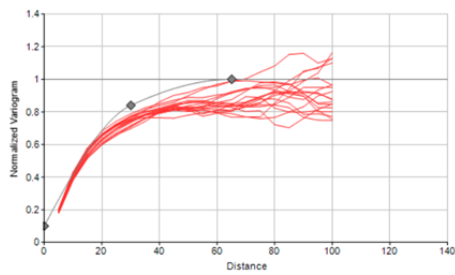
Figura 29. Gráfico de resíduos da regressão linear.

APÊNDICE G –Variogramas verticais dos cenários simulados e dos dados para os teores P2O5 do Fe2O3 e do MgO.



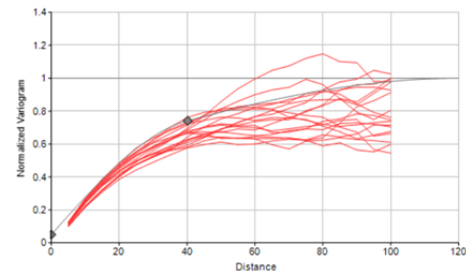
Variograma FO oxidado - FE203

Type	Variance	Range
Nugget	0.1	-
Spherical	0.450	30
Spherical	0.450	65



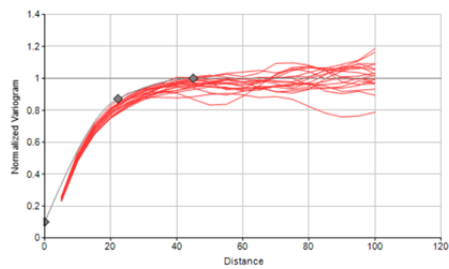
Variograma FO micaceo - FE203

Type	Variance	Range
Nugget	0.05	-
Spherical	0.45	40
Spherical	0.500	120



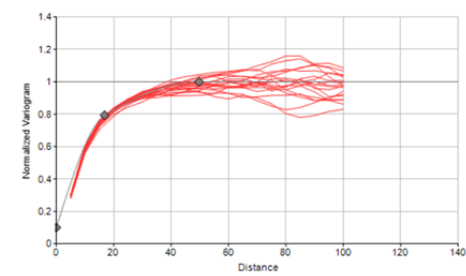
Variograma CBMS oxidado - FE203

Type	Variance	Range
Nugget	0.1	-
Spherical	0.5	22
Spherical	0.400	45



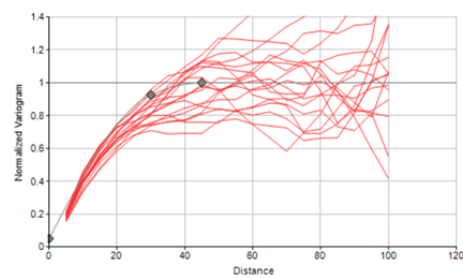
Variograma CBMS micáceo - FE203

Type	Variance	Range
Nugget	0.1	-
Spherical	0.5	17
Spherical	0.400	50



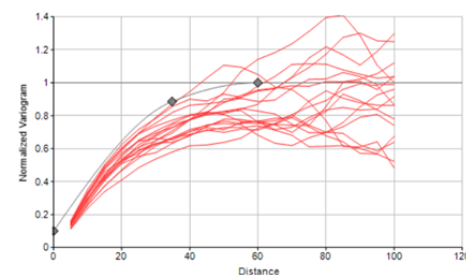
Variograma CBMG oxidado- FE203

Type	Variance	Range
Nugget	0.05	-
Spherical	0.45	30
Spherical	0.500	45



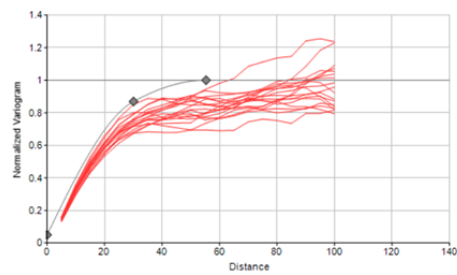
Variograma CBMG micáceo- FE203

Type	Variance	Range
Nugget	0.1	-
Spherical	0.4	35
Spherical	0.500	60.0



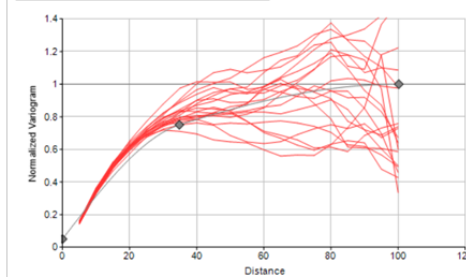
Variograma FLPI oxidado- FE203

Type	Variance	Range
Nugget	0.05	-
Spherical	0.450	30
Spherical	0.500	55



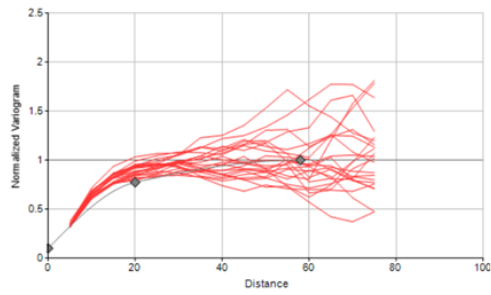
Variograma FLPI micáceo- FE203

Type	Variance	Range
Nugget	0.05	-
Spherical	0.45	35
Spherical	0.500	100



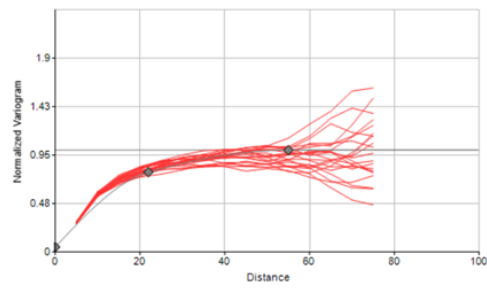
Variograma FO oxidado MGO

Type	Variance	Range
Nugget	0.1	-
Spherical	0.450	20
Spherical	0.450	58



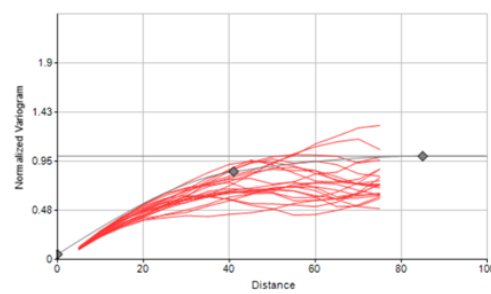
Variograma Fo micáceoMGO

Type	Variance	Range
Nugget	0.05	-
Spherical	0.45	22
Spherical	0.500	55



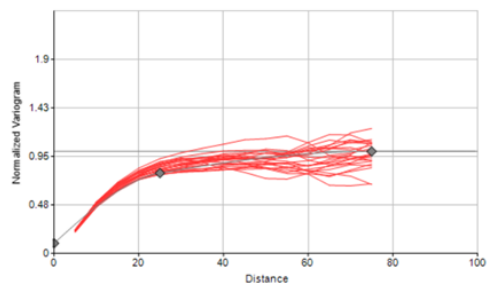
Variograma CBMS oxidado MGO

Type	Variance	Range
Nugget	0.05	-
Spherical	0.500	41
Spherical	0.45	85



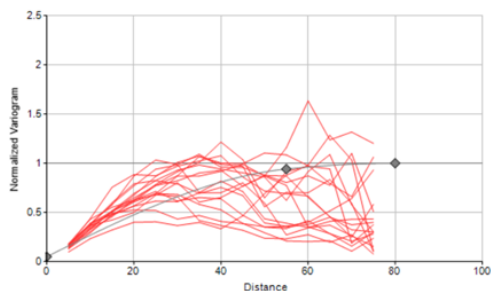
Variograma CBMS oxidado MGO

Type	Variance	Range
Nugget	0.1	-
Spherical	0.49	25
Spherical	0.41	75



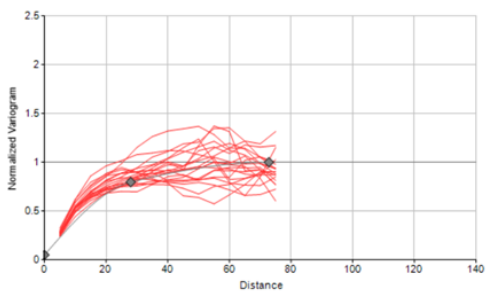
Variograma CBMG oxidado MGO

Type	Variance	Range
Nugget	0.05	-
Spherical	0.500	55
Spherical	0.45	80



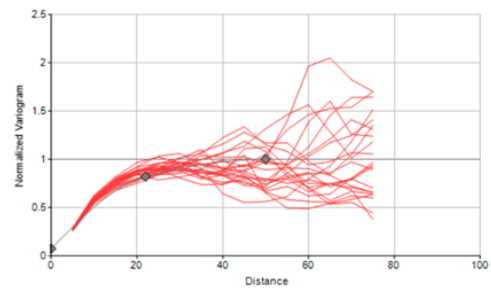
Variograma CBMG micáceo MGO

Type	Variance	Range
Nugget	0.05	-
Spherical	0.50	28
Spherical	0.45	73



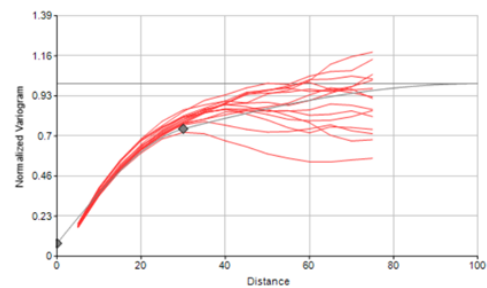
Variograma FLPI oxidado MGO

Type	Variance	Range
Nugget	0.075	-
Spherical	0.46	22
Spherical	0.465	50

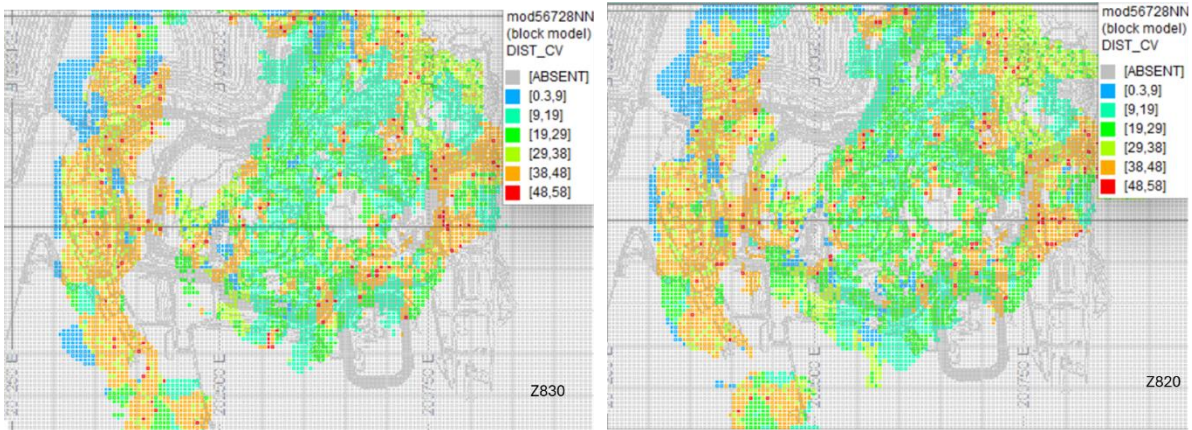


Variograma FLPI micáceo MGO

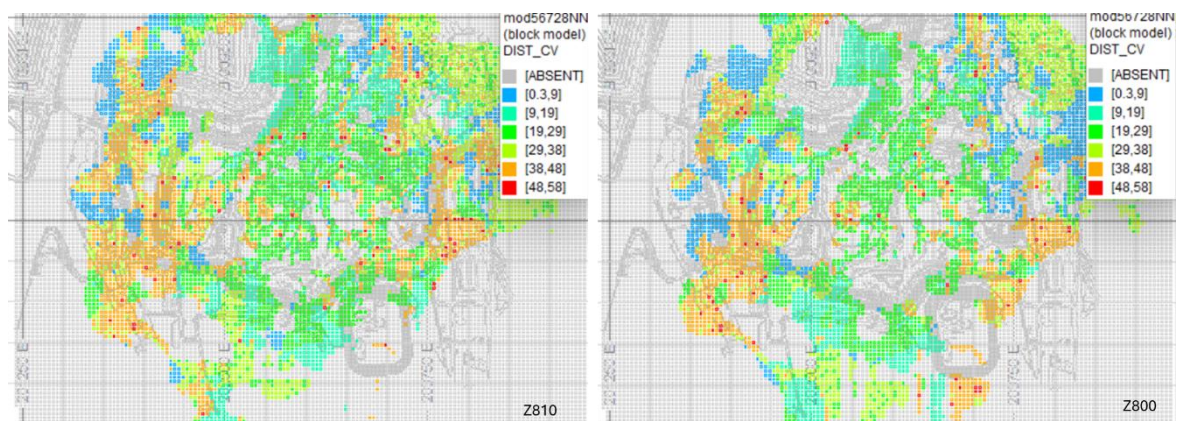
Type	Variance	Range
Nugget	0.075	-
Spherical	0.46	30
Spherical	0.465	100



APÊNDICE H – Mapas de coeficiente de variação da recuperação mássica



Coeficiente de variação da recuperação mássica na cota 830 m e 820 m.



Coeficiente de variação da recuperação mássica na cota 830 m e 820 m.