



**Programa de Pós-Graduação em Instrumentação, Controle e
Automação de Processos de Mineração (PROFICAM)
Escola de Minas, Universidade Federal de Ouro Preto (UFOP)
Associação Instituto Tecnológico Vale (ITV)**

Dissertação

**MONITORAMENTO DA QUALIDADE DE SINTER FEED ATRAVÉS
DE DADOS ESPECTRAIS ASSOCIADOS A APRENDIZADO DE
MÁQUINA – ESTUDO DE CASO: MINA DE CARAJÁS SERRA SUL
(S11D)**

Ana Cristina Pinto Silva

**Ouro Preto
Minas Gerais, Brasil**

2021

Ana Cristina Pinto Silva

**ESTIMATIVA DA QUALIDADE DE SÍNER FEED ATRAVÉS DE
ESPECTROSCOPIA DE REFLETÂNCIA E APRENDIZADO DE
MÁQUINA – ESTUDO DE CASO: MINA DE CARAJÁS SERRA SUL
(S11D)**

Dissertação apresentada ao Programa de Pós-Graduação em Instrumentação, Controle e Automação de Processos de Mineração da Universidade Federal de Ouro Preto e do Instituto Tecnológico Vale, como parte dos requisitos para obtenção do título de Mestre em Engenharia de Controle e Automação.

Orientadora: Prof. Rosa Elvira Correa Pabón, Ph.D.

Coorientador: Prof. Gustavo Pessin, Ph.D.

Ouro Preto

2021

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

S586m Silva, Ana Cristina Pinto .

Monitoramento da qualidade de SINTER FEED através de dados espectrais associados a aprendizado de máquina . [manuscrito] / Ana Cristina Pinto Silva. - 2021.

117 f.: il.: color., gráf., tab..

Orientadora: Profa. Dra. Rosa Elvira Correa Pabón.

Coorientador: Prof. Dr. Gustavo Pessin.

Dissertação (Mestrado Profissional). Universidade Federal de Ouro Preto. Programa de Mestrado Profissional em Instrumentação, Controle e Automação de Processos de Mineração. Programa de Pós-Graduação em Instrumentação, Controle e Automação de Processos de Mineração.

Área de Concentração: Engenharia de Controle e Automação de Processos Minerais.

1. Espectroscopia de Refletância. 2. Aglomeração - Sinter Feed. 3. Aprendizado de Máquina. I. Pabón, Rosa Elvira Correa. II. Pessin, Gustavo. III. Universidade Federal de Ouro Preto. IV. Título.

CDU 681.5:622.2

Bibliotecário(a) Responsável: Maristela Sanches Lima Mesquita - CRB-1716



FOLHA DE APROVAÇÃO

Ana Cristina Pinto Silva

Monitoramento da Qualidade de Sinter Feed Através de Dados Espectrais Associados a Aprendizado de Máquina

Dissertação apresentada ao Programa de Pós-Graduação em Instrumentação, Controle e Automação de Processos de Mineração (PROFICAM), Convênio Universidade Federal de Ouro Preto/Associação Instituto Tecnológico Vale - UFOP/ITV, como requisito parcial para obtenção do título de Mestre em Engenharia de Controle e Automação na área de concentração em Instrumentação, Controle e Automação de Processos de Mineração.

Aprovada em 10 de setembro de 2021

Membros da banca

Doutora - Rosa Elvira Correa Pabón - Orientadora - Instituto Tecnológico Vale
Doutor - Jefferson Rodrigo de Souza - Universidade Federal de Uberlândia
Doutora - Keyla Thayrinne Oliveira Coimbra - Universidade Estadual de Campinas
Doutor - Luciano Perdigão Cota - Instituto Tecnológico Vale

Rosa Elvira Correa Pabón, orientadora do trabalho, aprovou a versão final e autorizou seu depósito no Repositório Institucional da UFOP em 08/12/2021



Documento assinado eletronicamente por **Bruno Nazário Coelho**,
COORDENADOR(A) DE CURSO DE PÓS-GRAD EM INSTRUMENTAÇÃO



CONTROLE E AUTOMAÇÃO DE PROCESSOS DE MINERAÇÃO, em 10/12/2021, às 10:25, 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 http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0256510** e o código CRC **BDA84D8E**.

Referência: Caso responda este documento, indicar expressamente o Processo nº 23109.013043/2021-85

SEI nº
0256510

R. Diogo de Vasconcelos, 122, - Bairro Pilar Ouro Preto/MG, CEP 35400-000
Telefone: - www.ufop.br

Agradecimentos

Agradeço a orientadora Rosa Elvira Correa Pabón por todos os ensinamentos, paciência e tempo dedicado no suporte necessário a conclusão desta etapa. Agradeço também aos meus pais, irmãos, amigos, Tacila, Levi, Thaís Carrino, Fabiano Righi e Alex Hora, que foram sempre meus incentivadores e por confiarem no meu potencial.

A Vale SA, por impulsionar o desenvolvimento de seus colaboradores e a equipe do laboratório de Serra Sul pela disponibilização das bases de dados que permitiu a realização desse trabalho.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brasil (CAPES), Código de Financiamento 001; do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq); da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).

“Conheça todas as teorias, domine todas as técnicas,
mas ao tocar uma alma humana, seja apenas outra
alma humana. ”

(Carl G. Jung)

Resumo

Resumo da Dissertação apresentada ao Programa de Pós-Graduação em Instrumentação, Controle e Automação de Processos de Mineração como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

ESTIMATIVA DA QUALIDADE DE SÍNER FEED ATRAVÉS DE ESPECTROSCOPIA DE REFLETÂNCIA E APRENDIZADO DE MÁQUINA – ESTUDO DE CASO: MINA DE CARAJÁS SERRA SUL (S11D)

Ana Cristina Pinto Silva

Setembro, 2021

Orientadores: Rosa Elvira Correa Pabón

Gustavo Pessin

Essa pesquisa compreende na geração de bibliotecas espectrais e caracterização espectroscópica de síner feed, visando contribuir no aprimoramento dos métodos tradicionais utilizados na indústria mineral, para determinação de percentual de ferro e contaminantes na Mina de Carajás Serra Sul, mais conhecida como S11D. Para tanto, foram realizadas em ambiente de laboratório, leituras espectrais de amostras de síner feed de produto final e amostras preparadas, pulverizadas e secadas. As bibliotecas espectrais e sua caracterização espectroscópica foram realizadas no intervalo de 350 – 2500 nm. A biblioteca espectral gerada será integrada a dados geoquímicos como fluorescência de raio X, com o intuito de construir modelos empíricos que permitam determinar o percentual de ferro e identificar contaminantes nas amostras. Os dados produzidos deverão gerar informações que permitam: (i) identificar as bandas espectrais na assinatura do síner feed associadas ao conteúdo de ferro; (ii) identificar as bandas espectrais na assinatura do síner feed referentes aos contaminantes; (iii) gerar modelos estatísticos que permitam estimar o percentual de ferro nas amostras de síner feed; (iv) avaliar o uso de métodos de aprendizado de máquinas para estimar o teor de ferro em amostras de minério de ferro, com base em bibliotecas espectrais; (iv) espera-se que na medida em que os objetivos do projeto sejam atingidos, avaliar a possibilidade de uso dos critérios

derivados em laboratório para prever situações reais nas atividades de mineração e gerar uma nova metodologia que permita determinar o percentual de ferro e identificação de contaminantes de maneira precisa e oportuna, para a tomada de decisões e otimização nos processos produtivos.

Palavras-chave: Espectroscopia de Refletância, Sínter Feed, Percentual de Ferro, Aprendizado de Máquina.

Macrotema: Usina; **Linha de Pesquisa:** Instrumentação no Processamento de Minérios; **Tema:** Aumento de Produtividade na Obtenção de Dados Geológicos; **Área Relacionada da Vale:** Mina Carajás Serra Sul (S11D).

Abstract

Abstract of Dissertation presented to the Graduate Program on Instrumentation, Control and Automation of Mining Process as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

SINTER FEED QUALITY ESTIMATION THROUGH REFLECTANCE SPECTROSCOPY AND MACHINE LEARNING - CASE STUDY: CARAJÁS SERRA SUL MINE (S11D)

Ana Cristina Pinto Silva

September, 2021

Advisors: Rosa Elvira Correa Pabón
Gustavo Pessin

This research comprises the generation of spectral libraries and spectroscopic characterization of sinter feed, focusing on the enhancement of traditional methods used in the mineral industry to determine the percentage of iron and contaminants in Carajás Serra Sul Mine (S11D). Therefore, they were carried out in a laboratory environment, spectral measurements of final product sinter feed samples and prepared samples, dried and powdered. The spectral libraries and spectroscopic characterization were performed in the range of 350 - 2500 nm. The spectral library will be integrated with geochemical data such as X-ray fluorescence to build empirical models that allow determining the percentage of iron and identifying contaminants in the samples. The data produced should generate information that allows: (i) identify the spectral bands in the sinter feed signature associated with the iron content; (ii) identify the spectral bands in the sinter feed signature referring to the contaminants; (iii) generate statistical models that allow estimating the percentage of iron in the sinter feed samples; (iv) evaluate the use of machine learning methods to estimate the iron content in iron ore samples, based on spectral libraries; (iv) evaluate the possibility of using the derived criteria in the laboratory to predict real situations in mining activities and generate a new methodology that allows the determination of the percentage of iron and identification of contaminants in a precise and timely manner, for decision making and optimization production processes.

Keywords: Reflectance Spectroscopy, Sinter Feed, Percentage of Iron, Machine Learning.

Macrotheme: Usine; **Research Line:** Instrumentation in Mineral Processing; **Theme:** Productivity Increasing in Obtaining Geological Data; **Related Area of Vale:** Serra Sul Carajás Mine (S11D).

Lista de Figuras

Figura 1. Parâmetros espectrais obtidos dos espectros médios de cada método de amostragem.	26
Figura 2 (A). Localização - Complexo Minerador de Carajás. (B) Complexo Carajás Serra Sul – Corpo S11D.	31
Figura 3. Bens minerais exportados pelo Brasil no ano de 2018.	34
Figura 4. Características visuais dos tipos comerciais de minério de ferro: granulado, sinter feed e pellet feed.	35
Figura 5. Planta de beneficiamento da Usina Serra Sul.	36
Figura 6. Amostras representativas de sinter feed da Mina de Carajás Serra Sul.	37
Figura 7. Espectrorradiômetro FieldSpec 4 Hi-Res NG, o acessório Turntable e software RS3 utilizados para realização das leituras espectrais.	40
Figura 8: Detalhe do acondicionamento da amostra para a realização da leitura.....	41
Figura 9. Espectro de refletância bruto e corrigido através do splice correction da amostra de laboratório CN_10559.....	42
Figura 10. Espectro de refletância médio e leituras espectrais realizadas da amostra de laboratório CN-10559.	43
Figura 11 - Acima a representação do espectro da Amostra CN_10568 e abaixo o espectro normalizado da mesma amostra.	44
Figura 12 - Representação do modelo de análise de componentes principais (PCA).	46
Figura 13. Esquema de uma RNA Multilayer Perception com duas camadas ocultas e 6 neurônios cada.	48
Figura 14. Representação de uma árvore de decisão.....	49
Figura 15. Esquema de uma Random Forest com “N” conjuntos de árvore de decisão.	50
Figura 16. Representação dos hiperplanos de dados do método SVM.	52

Figura 17. Exemplos de respostas dos modelos quando ocorrem os efeitos do overfitting e underfitting comparados a um modelo ideal.	53
Figura 18. Representação do método de regressão linear.	54
Figura 19. Exemplos de gráficos Boxplot.	56
Figura 20. Fotos de amostras de sinter feed de Serra Sul como produto final (à esquerda) e preparada em laboratório (à direita), com suas respectivas identificações.	58
Figura 21. Espectros de refletância normalizados através da remoção do contínuo de amostras de sinter feed – produto final e sua equivalente pulverizada.....	60
Figura 22. Detalhe da faixa espectral onde ocorrem absorções relacionadas a óxidos e hidróxidos de ferro (600 a 1500 nm) das amostras produto e pulverizadas. Para comparação, a escala vertical é a mesma em todos os gráficos. Nota-se que as amostras CN_10585 e CN_10584 apresentam um comportamento mais hematítico, enquanto as demais amostras (CN_10564, CN_10593, CN_10556 e CN_10550) um comportamento mais goethítico. Adicionalmente, ao se comparar os espectros das amostras produto e sua equivalente pulverizada, o comportamento nesta região é semelhante.1	62
Figura 23. Detalhe da faixa espectral de contaminantes dos minérios de ferro (2100 a 2500 nm) das amostras produto e pulverizadas. Para comparação, a escala vertical é a mesma em todos os gráficos. Ao se comparar os espectros das amostras de produto e sua equivalente pulverizada, nota-se que as feições relacionadas com os contaminantes foram suavizadas nas amostras pulverizadas.	63
Figura 24. Análise dos dados pelo método PCA na faixa espectral de 380 a 2500 nm, gráfico normalizado através da remoção do contínuo das curvas espectrais das 58 amostras de Produto gerados com 3 componentes principais no intervalo de 380 a 2500 nm.....	64
Figura 25. Gráfico de dispersão dos autovetores da PC1 vs PC2 referentes aos espectros de refletância normalizados através da remoção do contínuo das amostras produto de sinter feed no intervalo de 380 a 2500 nm. As cores	

estão relacionadas com o teor de FeO e o tamanho dos círculos com o teor de SiO ₂ . Nota-se que quanto maior a percentagem de FeO, menor teor de SiO ₂	64
Figura 26. (A) Gráfico de dispersão de autovetores da PC1 vs PC2 referente aos espectros de refletância normalizados através da remoção do contínuo das amostras produto de sinter feed no intervalo de 380 à 2500 nm. As cores estão relacionadas com o teor de FeO e o tamanho dos círculos com o teor de Al ₂ O ₃ . (B) Espectro em destaque para as amostras goethíticas, com feições marcantes acima da faixa de 900 nm.....	65
Figura 27. Análise dos dados pelo método PCA na faixa espectral de 380 a 1310 nm, gráfico normalizado através da remoção do contínuo das curvas espectrais das 58 amostras de Produto gerados com 3 componentes principais no intervalo de 380 a 1310 nm.....	66
Figura 28. (A) Gráfico de dispersão de autovetores da PC1 vs PC2 referente aos espectros de refletância normalizados através da remoção do contínuo, das amostras produto de sinter feed no intervalo de 380 à 1310 nm, com foco no percentual de teor de Fe; (B) Espectros de refletância normalizados, com indicação das curvas de refletância de 3 amostras (linhas pontilhadas) que apresentaram maior dispersão em (A) com relação às demais amostras, e caracterizadas por feições de absorção abaixo da faixa de 900 nm (comportamento hematítico).....	67
Figura 29. Análise dos dados pelo método PCA na faixa espectral de 2135 a 2500 nm, gráfico normalizado através da remoção do contínuo das curvas espectrais das 58 amostras de Produto gerados com 2 componentes principais no intervalo de 2135 a 2500 nm.....	68
Figura 30. (A) Gráfico de dispersão da PC1 vs PC2 dos espectros de refletância normalizados através da remoção do contínuo, das amostras produto de sinter feed no intervalo de 2135 à 2500, com foco na relação entre percentuais de teor de Fe e CaO; (B) Espectro com destaque para a amostra CN_03_Prod na faixa de 2135 a 2500 nm.....	69

Figura 31. - - (A) Gráfico de dispersão de autovetores da PC2 vs PC3 referentes aos espectros de refletância normalizados através da remoção do contínuo, das amostras produto de sinter feed no intervalo de 2135 a 2500 nm, com foco na relação entre percentuais de teor de Fe e K ₂ O; (B) Espectros de refletância das amostras de sinter feed, com destaque para a amostra CN_14_Prod (em linha pontilhada).	70
Figura 32. Análise dos dados pelo método PCA na faixa espectral de 380 a 2500 nm, gráfico normalizado através da remoção do contínuo das curvas espectrais das 58 amostras preparadas em laboratório, gerados com 4 componentes principais no intervalo de 380 a 2500 nm.....	71
Figura 33. Gráfico de dispersão da PC1 vs PC2 dos espectros de refletância normalizados através da remoção do contínuo, das amostras de sinter feed preparadas em laboratório, no intervalo de 380 à 2500 nm, com realce de cor relativo ao percentual de Fe.	72
Figura 34. Gráfico de dispersão de autovetores da PC1 vs PC2 referente aos espectros de refletância normalizados através da remoção do contínuo das amostras sinter feed preparadas em laboratório, no intervalo de 380 a 2500 nm. As cores estão elacionadas com o teor.	73
Figura 35. Curvas espectrais de refletância normalizadas de 58 amostras de sinter feed pulverizadas, considerando-se a faixa de 380 a 1310 nm. Estes dados foram selecionados para processamento via o método PCA.	74
Figura 36. Curvas espectrais de refletância normalizadas de 58 amostras de sinter feed pulverizadas, considerando-se a faixa de 2135 a 2500 nm. Estes dados foram selecionados para processamento via o método PCA.	75
Figura 37. (A) Espectro com a amostra CN_14_Lab em destaque para (linha pontilhada) a faixa de 2135 a 2500 nm; (B) Gráfico de dispersão da PC2 vs PC4 dos espectros de refletância normalizados através da remoção do contínuo das amostras sinter feed preparadas em laboratório, no intervalo de 2135 a 2500 nm. As cores estão relacionadas com o teor de FeO e o tamanho dos círculos com o teor de K ₂ O.	76

Figura 38. Curva Espectral da amostra de produto CN_10585 e o destaque do ruído faixa abaixo de 400 nm.	79
Figura 39. Gráficos boxplot dos métodos SVM, AdaBoost e RF para as amostras de produto na faixa de 400 a 2500 nm.	82
Figura 40. Gráficos boxplot dos métodos SVM, AdaBoost e RF para as amostras de produto na faixa de 400 a 1310 nm.	84
Figura 41. Gráficos boxplot dos métodos kNN e SVM para as amostras de laboratório na faixa de 400 a 2500 nm.	86
Figura 42. Ranking dos principais atributos considerados nas análises dos modelos de aprendizado para as amostras de laboratório na faixa de 400 a 1310 nm.	88
Figura 43. Gráficos boxplot dos métodos AdaBoost, RF, RNA, Árvore de decisão e kNN para as amostras de laboratório na faixa de 400 a 1310 nm.	89

Lista de Tabelas

Tabela 1. Principais minerais de ferro que compõem os minérios de ferro	32
Tabela 2. Resultado físico-químico de laboratório das amostras coletadas em Serra Sul.....	38
Tabela 3. Métodos de Aprendizado de máquinas aplicados nas amostras de produto na faixa de 400 a 2500 nm e resultados estatísticos.	81
Tabela 4. Métodos de Aprendizado de máquinas aplicados nas amostras de produto na faixa de 400 a 1310 nm e resultados estatísticos.	83
Tabela 5. Métodos de Aprendizado de máquinas aplicados nas amostras de laboratório na faixa de 400 a 2500 nm e resultados estatísticos.....	85
Tabela 6. Métodos de Aprendizado de máquinas aplicados nas amostras de laboratório na faixa de 400 a 1310 nm e resultados estatísticos.....	87

Lista de Siglas e Abreviaturas

ASD: *Analytical Spectral Devices*.

ER: Espectroscopia de Reflectância.

IBRAM: Instituto Brasileiro de Mineração.

ITV: Instituto Tecnológico Vale.

KNN: (*K-Nearest Neighbor*).

MIR: Do inglês – *MID Infrared* – Infravermelho Médio.

MRC: Materiais de referência certificados.

NIR: Do inglês – *Near Infrared* – Infravermelho Próximo.

PCA: *Principal Component Analysis*.

RMSE: *Root Mean Square Error*.

RNA: Redes Neurais Artificiais.

RF: *Random Forest*.

ROM: *Run-of-mine*.

S11D: Mina de Carajás Serra Sul.

SVM: *Support Vector Machine*.

SWIR: Do inglês - *Shortwave Infrared* - Infravermelho de Ondas Curtas.

TCLD: Transportador de longa distância.

VNIR: Do inglês - *Visible Near Infrared* – Visível Infravermelho Próximo.

VSWIR: Do inglês – *Visible, Near Infrared and Shortwave Infrared* – Visível, Infravermelho Próximo e de Ondas Curtas

SUMÁRIO

1	INTRODUÇÃO.....	21
1.1	CONTEXTUALIZAÇÃO	21
1.2	JUSTIFICATIVA	22
1.3	OBJETIVOS	23
1.4	ESTRUTURA DO DOCUMENTO.....	23
2	REFERENCIAL TEÓRICO E FUNDAMENTAÇÃO CIENTÍFICA	25
2.1	ESPECTROSCOPIA DE REFLETÂNCIA.....	25
2.1.1	Resolução Espectral.....	26
2.1.2	Processamento e análise dos dados espectrais	27
2.2	TRABALHOS RELACIONADOS	27
3	COMPLEXO MINERAL CARAJÁS SERRA SUL – S11D	30
3.1	LOCALIZAÇÃO	30
3.2	ASPECTOS MINERALÓGICOS	30
3.3	MINÉRIO DE FERRO	33
3.4	PROCESSAMENTO DO SÍNTER FEED	35
4	MATERIAL E MÉTODOS	37
4.1	AMOSTRAS.....	37
4.2	FLUORESCÊNCIA DE RAIOS X (FRX)	39
4.3	ESPECTROSCOPIA DE REFLETÂNCIA.....	39
4.3.1	Instrumentação	39
4.3.2	Leituras Espectrais	40
4.4	PRÉ-PROCESSAMENTO DOS DADOS.....	41
4.4.1	Correção dos gaps entre os detectores – “ <i>Splice Correction</i> ”.....	40
4.4.2	Média.....	42

4.4.3	Criação da Biblioteca Espectral	43
4.5	ANÁLISE DOS DADOS ESPECTRAIS	44
4.5.1	Remoção do contínuo	44
4.5.2	Análise por componentes principais	45
4.6	MÉTODOS DE APRENDIZADO DE MÁQUINA.....	46
4.6.1	Redes Neurais Artificiais (RNA)	47
4.6.2	Arvore de Decisão	48
4.6.3	Random Forest (RF).....	49
4.6.4	Adaboosting.....	50
4.6.5	KNN (K-Nearest Neighbor)	50
4.6.6	SVM (Support Vector Machine)	51
4.7	AVALIAÇÃO DA ACURÁCIA DOS MÉTODOS DE APRENDIZADO DE MÁQUINA.....	52
4.7.1	Overfitting e Underfitting.....	52
4.7.2	Coefficiente de determinação (R^2).....	53
4.7.3	Erro médio quadrático (MSE).....	53
4.7.4	Erro percentual médio absoluto (MAPE)	54
4.7.5	Análise por regressão linear	54
4.7.6	Boxplot.....	55
4.8	SOFTWARE ORANGE CANVAS	56
5	CARACTERIZAÇÃO ESPECTRAL DE SINTER FEED.....	57
5.1	CARACTERIZAÇÃO FÍSICA	57
5.2	CARACTERIZAÇÃO ESPECTRAL.....	59
5.3	ANÁLISE DE COMPONENTES PRINCIPAIS (PCA) – AMOSTRAS BRUTAS DE SÍNTER FEED	62
5.3.1	Análise espectral na faixa de 380 a 2500 nm	62

5.3.2	Análise espectral na faixa de 380 a 1310 nm - Ferro.....	66
5.3.3	Análise espectral na faixa de 2135 a 2500 nm.....	68
5.4	ANÁLISE DE COMPONENTES PRINCIPAIS (PCA) – AMOSTRAS DE SÍNTER FEED PREPARADAS EM LABORATÓRIO.....	71
5.4.1	Análise espectral na faixa de 380 a 2500 nm.....	71
5.4.2	Análise espectral na faixa de 380 a 1310 nm.....	73
5.4.3	Análise espectral na faixa de 2135 a 2500 nm.....	74
5.5	COMPARAÇÃO ENTRE AS ANÁLISES ESPECTRAIS EM AMOSTRAS DE SÍNTER FEED DE PRODUTO E PREPARADAS EM LABORATÓRIO.....	77
6	ESTIMATIVA DA QUALIDADE DE SINTER FEED ATRAVÉS DE ESPECTROSCOPIA DE REFLETÂNCIA E APRENDIZADO DE MÁQUINA.....	78
6.1	FORMAÇÃO DA BASE DE DADOS.....	78
6.2	SELEÇÃO DAS VARIÁVEIS.....	78
6.3	AValiação de variações dos modelos.....	80
6.3.1	Dataset das Amostras de Produto.....	80
6.3.2	Dataset das Amostras Preparadas em Laboratório.....	85
7	CONCLUSÕES E CONSIDERAÇÕES FINAIS.....	90
	REFERÊNCIAS.....	92
	ANEXO 1: Exemplo de Ficha Espectral do Banco de Dados Espectrais Vale.....	99
	ANEXO 2: Passo a passo – Realização da leitura espectral das amostras e criação da biblioteca espectral.....	101

1 INTRODUÇÃO

Este capítulo busca introduzir o tema da pesquisa de maneira sucinta, de forma a apresentar os conceitos gerais como a contextualização, justificava e os objetivos do projeto. Ademais, é disposta a estrutura do texto em capítulos, com um breve descritivo dos conteúdos abordados.

1.1 CONTEXTUALIZAÇÃO

A exploração e beneficiamento do minério de ferro possui papel fundamental na economia brasileira devido ao grande volume de exportações, que faz do Brasil um dos principais fornecedores mundiais destas *comodities*. Além dos milhares de empregos que são gerados em suas plantas, que geram renda e prosperidade para as comunidades e municípios. A qualidade desse mineral é primordial para os fins de destinação na indústria siderúrgica e metalúrgica.

O valor agregado do minério depende do teor de ferro (Fe) existente, assim, como um percentual aceitável de outros contaminantes que podem estar presentes na rocha, que em níveis acima da especificação, podem não ser adequados aos processos pelos quais são destinados. Entre as alternativas de determinação do teor de Fe e seus contaminantes, temos análises laboratoriais, nas quais são colhidas amostras no processo e levadas para realização de amostragens e análises físicas e químicas, com a utilização de métodos como espectrometria de fluorescência em raio X. Porém, esses métodos tradicionais fornecem as informações requeridas para a tomada de decisão com um tempo não hábil para impactar no processo produtivo e a estratégia do negócio.

A espectroscopia de refletância é o estudo da luz como uma função do comprimento de onda que foi refletido ou disperso de um sólido, líquido ou gás (CLARK, 1999). O uso dessa tecnologia apresenta uma série de atrativos como, por exemplo, a possibilidade de análise sem a necessidade de preparação da amostra, a rapidez para a obtenção dos espectros, cálculos e apresentação dos resultados. O fato de ser um método não destrutivo, permite o uso posterior da amostra para outros ensaios ou para ser guardada como testemunho e a possibilidade do desenvolvimento de aplicações em linha através de analisadores de processo.

Nesse contexto, a presente proposta de pesquisa visa analisar a viabilidade de aplicação da espectroscopia de refletância na faixa do *VSWIR (Visible, Near Infrared and Shortwave Infrared; 350 – 2500 nm)* associada a métodos de aprendizado de máquina na caracterização do produto gerado - *sínter feed*, da Mina de Carajás Serra Sul, também conhecida como S11D,

pertencente à Vale e localizada na região sudeste do estado do Pará. Espera-se avaliar o potencial de uso destas tecnologias na determinação de percentual de Fe, e identificação de possíveis contaminantes.

1.2 JUSTIFICATIVA

Análises laboratoriais (*e.g.* fluorescência de raios X – FRX e difração de raios X – DRX) não fornecem resultados instantaneamente, de forma a contribuir para as tomadas de decisões em relação ao processo produtivo, nestes casos os resultados muitas vezes são obtidos após o minério estar a caminho de seu destino final. Sendo assim, na maioria das unidades mineradoras, a função de tais análises laboratoriais é apenas de realizar o controle da qualidade do minério já expedido.

No laboratório de Serra Sul, diferentes litologias são analisadas, de forma a atender a faixa de percentual de Fe e contaminantes para então se obter uma análise de boa precisão e exatidão. São compostas as curvas de regressão do espectrômetro com materiais de referência certificados (MRC) de alta confiabilidade, de forma a obedecer às normas que regem o controle de qualidade – ISSO 17025, que é a norma utilizada para padronização de testes laboratoriais.

Nesse sentido, torna-se necessária a exploração de tecnologias alternativas que permitam estimar, com erros aceitáveis, o percentual de Fe de maneira rápida para se determinar ações acerca da estratégia do processo produtivo. Assim, este estudo é baseado na busca de criação de bibliotecas espectrais abrangentes (diferentes amostras de *sinter feed*, teores de Fe, contaminantes, mineralogia, litologias), em paralelo com dados de mineralogia, granulometria, geoquímica, fornecidos pelos laboratórios com análises tradicionais. Com a integração destes dados, pretende-se criar modelos através de métodos de aprendizado de máquina que permitam estimar o percentual de Fe, ademais, de possibilitar a identificação de contaminantes nos produtos. Adicionalmente, será possível avaliar qual a resolução espectral e cobertura do espectro eletromagnético específica para cada estudo particular.

Pelo exposto, pretende-se avaliar a possibilidade de uso de critérios derivados em laboratório para prever situações reais e construir uma nova metodologia que auxilie na tomada de decisões e de aplicação mais genérica de ferramentas de espectroscopia de refletância e aprendizado de máquina na indústria da mineração.

1.3 OBJETIVOS

Avaliar a aplicação da espectroscopia de refletância associada a métodos de aprendizado de máquina na caracterização de sinter feed e na determinação de percentual de teor de ferro – Estudo de caso na Mina de Carajás Serra Sul, S11D.

No alcance desse objetivo contempla-se os seguintes objetivos específicos:

- Utilizar e avaliar o intervalo *VSWIR* (*Visible, Near InfraRed and Short Wave InfraRed*; 350 – 2500 nm) do espectro eletromagnético, para caracterização de amostras de sinter feed;
- Compor uma biblioteca espectral que compreenda as assinaturas espectrais no intervalo *VSWIR* das amostras coletadas na Mina de Carajás Serra Sul;
- Caracterizar as assinaturas espectrais do sinter feed, identificando as principais feições de absorção associadas a minerais de ferro e possíveis contaminantes;
- Integralizar os dados espectrais com dados geoquímicos (FRX) e avaliar possíveis padrões de agrupamento;
- Aplicar técnicas de aprendizado de máquina para estimar o conteúdo de ferro;
- Comparar as análises laboratoriais convencionais para validação do método e mensurar o grau de desvio do processo aplicado;
- Avaliar a possibilidade de uso dos procedimentos metodológicos acima listados, nas atividades rotineiras de mineração.

1.4 ESTRUTURA DO DOCUMENTO

A dissertação está estruturada em sete capítulos que permeiam nos objetivos do trabalho, bem como os métodos adotados e os próximos passos a serem desenvolvidos no estudo. O Capítulo 2 apresenta uma fundamentação teórica para embasamento da pesquisa no quesito de espectroscopia de refletância e métodos de aprendizado de máquina na caracterização de amostras de minério de ferro. O Capítulo 3 proporciona informações sobre a área de estudo e o produto estudado nesta pesquisa. No Capítulo 4 são discutidos o material e os métodos utilizados no desenvolvimento deste estudo. No Capítulo 5, diferentes aspectos relacionados com a caracterização espectral dessas amostras de sinter feed são discutidos. No Capítulo 6 são analisados e avaliados diferentes algoritmos de aprendizado de máquina na estimação do percentual de teor ferro. Ademais, o Capítulo 7, apresenta as considerações finais e sugestões

para futuras pesquisas. Adicionalmente, um exemplo de ficha espectral da Biblioteca Vale e um tutorial para realização de leituras espectrais e construção da biblioteca de dados foram elaborados e apresentados nos Anexo 1 e 2, respectivamente.

2 REFERENCIAL TEÓRICO E FUNDAMENTAÇÃO CIENTÍFICA

Neste capítulo é apresentado o referencial teórico necessário para o embasamento técnico para a realização do trabalho. São expostos conceitos relacionados a espectroscopia de refletância e algumas pesquisas desenvolvidas nessa linha.

2.1 ESPECTROSCOPIA DE REFLETÂNCIA

Quando a luz interage com algum mineral ou rocha, é preferencialmente absorvida em certos comprimentos de onda, enquanto que em outros comprimentos de onda é transmitida e refletida. A refletância é definida como a razão entre a intensidade de luz refletida pela amostra e a intensidade de luz incidente (VAN DER MEER, 2004).

Processos de transferência de carga e transição eletrônica (e.g. mudança no estado eletrônico de elétrons ligados a átomos ou moléculas) em associação a íons de metais de transição como Fe, Ti, Cr, etc., determinam, em grande parte, a posição das feições de absorção dos espectros de alguns minerais na região do VSWIR (BURNS, 1993; ADAMS, 1974, 1975). Além de processos de vibração relacionados a H₂O e OH, que também produzem feições de absorção características na faixa do SWIR (HUNT, 1977).

A posição, forma, profundidade e comprimento destas feições de absorção, são controladas pela estrutura cristalina em que estes átomos e moléculas absorventes se encontram, e pela estrutura química do material. Logo, as variáveis que caracterizam estas feições de absorção podem ser diretamente correlacionadas com a composição química e estrutura cristalina da amostra. A profundidade da feição de absorção é um indicador de quantitativo do material presente na amostra. Além disto, está também relacionado a granulometria ou tamanho das partículas, já que a luz absorvida e refletida é relacionada ao tamanho do elemento. (VAN DER MEER, 2004).

A posição da feição de absorção é definida como o comprimento de onda em que feição de absorção chega ao mínimo de refletância da curva do espectro, e a posição de uma feição de refletância é definida como comprimento de onda em que a feição chega ao máximo de refletância. Quando não é identificada nenhuma feição de absorção no espectro analisado, o valor da profundidade é próximo ou igual a zero, indicando a ausência do mineral investigado. O fator de assimetria da feição é definido por Van der Meer (2004) como a razão entre as áreas a esquerda e a direita do comprimento de onda de mínima refletância, compreendida entre os

dois ombros da feição. A inclinação de uma determinada região do espectro é calculada obtendo o coeficiente angular da reta tangente a esta região, conforme representado na Figura 1.

Figura 1. Parâmetros espectrais obtidos dos espectros médios de cada método de amostragem.



Fonte: Adaptado PRADO, 2015.

Os parâmetros espectrais são obtidos da curva polinomial que melhor se ajusta ao espectro de refletância com a remoção do contínuo.

2.1.1 Resolução Espectral

Segundo Menezes e Almeida (2012) o termo resolução espectral envolve no mínimo três parâmetros de medidas, sendo a primeira relacionada ao número de bandas que o sensor possui, em seguida a largura em comprimento de onda das bandas e por fim, as posições que as bandas estão situadas no espectro eletromagnético.

Portanto, resolução espectral é a capacidade de um sensor em sua definição de intervalos de comprimento de onda, onde quanto melhor for a resolução espectral, mais estreita será a faixa de comprimento de onda para uma determinada banda. Portanto, faixas mais largas conferem uma resolução espectral menor ao sensor. Dessa forma, um sensor possui melhor resolução espectral se ele possui maior número de bandas capaz de abranger diferentes regiões espectrais e com larguras estreitas de comprimento de onda.

Para possibilitar a abrangência de maiores faixas espectrais sem a perda de resolução, alguns instrumentos apresentam a combinação de vários sensores distintos, como no caso do espectraloradiômetro *ASD FieldSpec 4 Hi-Res* utilizado nesse estudo. Nele são utilizados 2151

canais, com 3 sensores, sendo 1 para detecção no *VNIR* (350 – 1000 nm) e 2 para o *SWIR* para cobrir respectivamente as faixas de 1001-1800 nm e 1801 - 2500 nm. Dessa forma, é possível abranger uma faixa total de 350-2500 nm com alta resolução para medições de refletância.

2.1.2 Processamento e análise dos dados espectrais

Estudos envolvendo medidas de espectroscopia de refletância fazem uso de um grande volume de dados. Nesse caso, técnicas que consigam diminuir a quantidade de dados a serem analisados (sem que ocorra perda de informação) e tornar mais eficiente a extração de informações, são indispensáveis. Para verificar qual metodologia é mais indicada, diversas técnicas serão exploradas no projeto, tais como: remoção do contínuo, análise por principais componentes, análises por regressão, entre outros. Com o desenvolvimento do projeto, outros métodos de processamento e análise de dados poderão ser utilizados.

2.2 TRABALHOS RELACIONADOS

Abaixo serão apresentados trabalhos relacionados a espectroscopia de refletância aplicada à análise quantitativa da mineralogia:

Neves (2001) utilizou a cores características dos óxidos de ferro para estabelecer um método de análise dos cinco principais óxidos de ferro de maior ocorrência, obtendo-se seus espectros de refletância difusa na região do visível, usando calibração multivariada. As estimativas dos teores de hematita e goethita pelas amplitudes de suas bandas características, mostraram uma razoável concordância com os resultados da análise química. Os resultados como um todo, mostraram que a técnica é promissora, mas que necessita mais alguns ajustes, principalmente para a estimativa de determinados óxidos de ferro em solos com mineralogia mais complexa.

No estudo de Carioca (2010), optou-se pela espectrofotometria de refletância difusa para a identificação e quantificação das fases presentes nas amostras de minérios de ferro. A partir das curvas de refletância e dos teores medidos dos minerais, e com o auxílio do programa quimiométrico PARLES, obteve-se modelos de calibração para a hematita especularita, martita, goethita magnetita, quartzo e ferro total. As melhores correlações encontradas foram obtidas no equipamento ASD portátil (350 – 2500 nm) para goethita $R^2_c = 0,96$; magnetita $R^2_c = 0,90$ e martita $R^2_c = 0,84$; e no MIR (2500 nm – 25000 nm) para especularita $R^2_c = 0,89$; ferro $R^2_c =$

0,87 e quartzo $R^2_c = 0,89$. Para a hematita a melhor correlação foi no equipamento CARY / MIR (319 – 25000) com $R^2_c = 0,91$.

A pesquisa de Teixeira (2014) desenvolveu um modelo de regressão, capaz de prever a quantidade de carbono em amostras de solo utilizando a espectroscopia na região do visível e no infravermelho próximo. Para tal, foi utilizada a técnica de aprendizado de máquinas SVM (*Support Vector Machine*). Nos experimentos realizados foram utilizados dois conjuntos de amostras de solo coletadas na região de Campos Gerais. A avaliação dos resultados teve como base os erros de previsão e os coeficientes de correlação entre os valores dos teores de carbono preditos pelo modelo. Foram encontrados coeficientes de correlação na faixa de 0,84 a 0,90 que, portanto, validou o uso da técnica como uma alternativa aos métodos convencionais de análise de carbono em solos.

No trabalho de Prado (2015) foram obtidos espectros de refletância de testemunhos de sondagem do depósito de ferro da Mina de N4WS, localizado na Província Mineral de Carajás - PA. Estes espectros foram validados utilizando análises geoquímicas de fluorescência de raio-x e lâminas delgadas. Os espectros de refletância foram coletados utilizando o espectrorradiômetro *ASD FielSpec* em nove testemunhos de sondagem. A mineralogia do depósito de ferro foi extraída dos dados espectrais utilizando a geometria (profundidade e comprimento de onda) do espectro de refletância removido do contínuo. A profundidade dessas feições de absorção é proporcional à abundância mineral e o comprimento de onda proporcional à composição química. As feições de absorção diagnosticadas de cada mineral foram utilizadas para determinar a abundância e a composição mineral e dessa forma permitem uma caracterização efetiva do minério.

Os parâmetros de abundância mineral obtidos neste trabalho, com acurácia entre 6 e 15,8% (RMSE), mostram que este método pode ser de grande utilidade como uma estimativa rápida do conteúdo de ferro e de alguns minerais relacionados ao depósito.

Já no artigo de Souza et al. (2017) foi avaliado o uso da técnica de espectroscopia de refletância difusa na região do Visível e Infravermelho próximo VNIR, para avaliar se é possível classificar os níveis de fósforo disponível no solo (P) por meio da espectroscopia. Foi utilizado o espectrorradiômetro *FieldsSpec 4* para coleta dos espectros em amostras de solo submetidas a doses diferenciadas de P. A qualidade geral dos modelos de classificação gerados foi razoável, mas a precisão foi baixa. A ausência de resposta espectral direta do P justifica

esses resultados. Dessa forma, a espectroscopia no VNIR apresenta limitações nas análises de disponibilidade desse elemento no solo, pois sua predição é dependente de correlação com outras propriedades do solo.

Em sua dissertação de mestrado, Viana (2020) fez uso de imagens hiperespectrais para desenvolvimento de métodos de aprendizado de máquina para estimar o teor de ferro em amostras de minérios de ferro. Além disso, foram necessárias investigações sobre a escolha de bandas espectrais para melhor caracterização de cada mineral presente nas amostras envolvidas no estudo.

2 COMPLEXO MINERAL CARAJÁS SERRA SUL – S11D

3.1 LOCALIZAÇÃO

A Província Mineral de Serra Sul, alvo do presente trabalho, está localizada ao Sul da Floresta Nacional de Carajás e tem Canaã dos Carajás como município referente, situado na região sudeste do estado do Pará. A reserva mineral possui uma área de 48.000 hectares e 120 km de extensão.

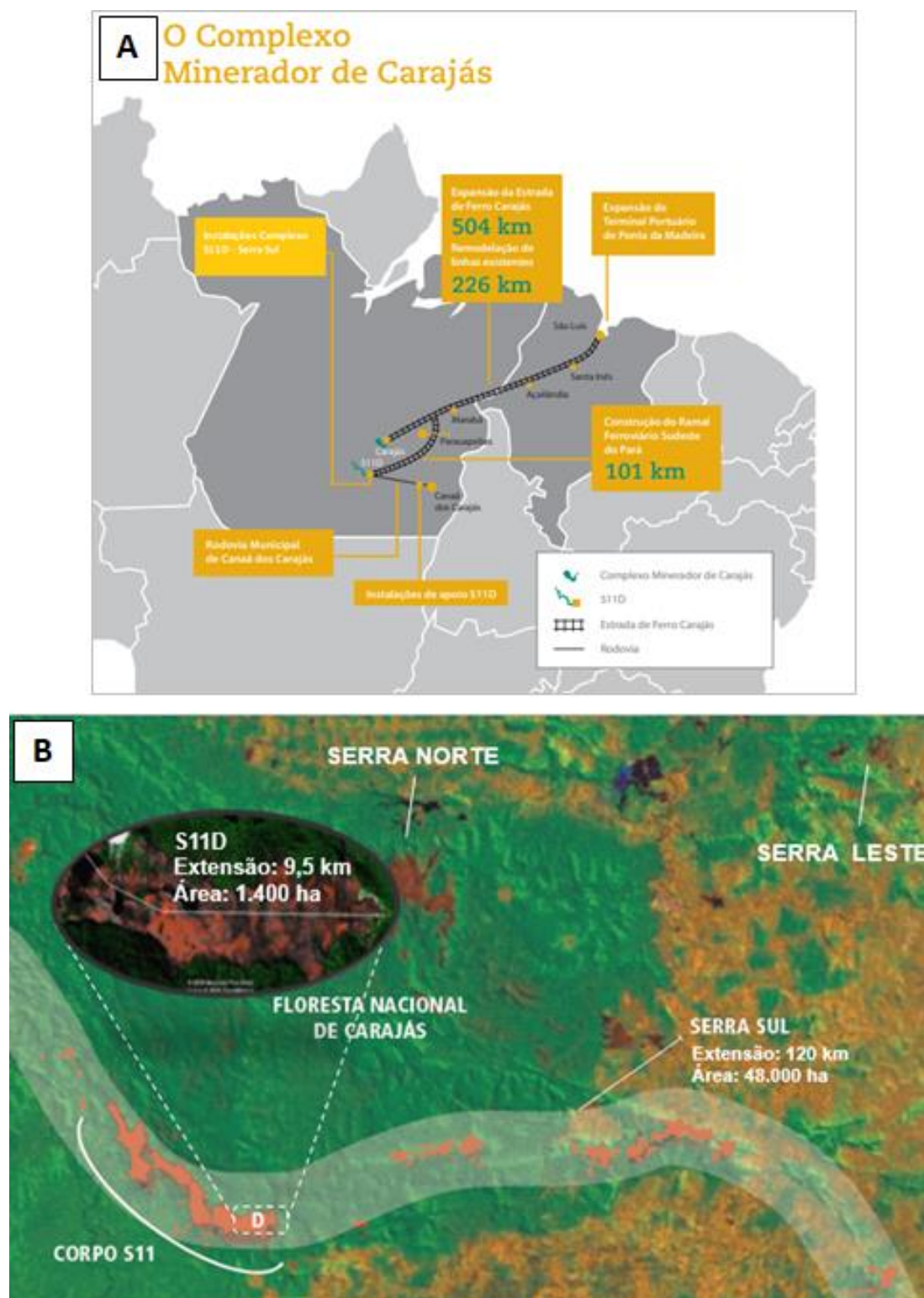
O projeto de exploração dessa área foi denominado S11D e representou a expansão da atividade de extração e processamento de minério de ferro no Complexo Minerador de Carajás, em operação desde 1985 (Figura 1A). O projeto foi nomeado pela localização (corpo mineral S11, bloco D). O potencial de extração de minério é de 10 bilhões de toneladas métricas de minério de ferro, onde bloco D contém 2,78 bilhões de toneladas de reservas a serem extraídas (Figura 1B). Após o início da operação em agosto de 2017, o projeto Ferro Carajás S11D foi denominado Complexo Mineral Carajás Serra Sul, dessa forma, vamos referenciá-lo com esta nomenclatura durante o desenvolvimento desse estudo.

3.2 ASPECTOS MINERALÓGICOS

Segundo Duarte (2019), dos vários elementos que compõem a crosta terrestre, o ferro é o quarto mais abundante, com cerca de 5% em massa, atrás apenas do oxigênio (46,6%), da sílica (26,72%) e do alumínio (8,13%).

Dependendo da composição mineralógica e química dos minérios de ferro, pode-se obter teores de ferro variados considerando diferentes fórmulas químicas de seus constituintes conforme mostrado na Tabela 1.

Figura 2 (A). Localização - Complexo Minerador de Carajás. (B) Complexo Carajás Serra Sul – Corpo S11D.



Fonte: Arquivos internos Vale.

Tabela 1. Principais minerais de ferro que compõem os minérios de ferro

Mineral	Fórmula química	% de ferro
Hematita	Fe_2O_3	69,9
Magnetita	Fe_3O_4	72,4
Goethita	FeOOH	62,9
Siderita	FeCO_3	48,3
Pirita	FeS_2	46,6
Pirrotita	Fe_{1-x}S	61,5
Ilmenita	FeTiO_3	36,8

Fonte: CARIOCA, 2010.

O elemento ferro encontra-se geralmente sob a forma de óxidos ou hidróxidos, principalmente como hematita ($\alpha\text{-Fe}_2\text{O}_3$), goethita ($\alpha\text{-FeOOH}$) e magnetita (Fe_3O_4), sendo a cor um dos atributos mais visíveis e característicos destas fases (Carioca, 2010). A variedade das cores exibidas por estes compostos resulta dos diferentes tipos de transições eletrônicas que neles ocorrem. De fato, todas as cores quentes, do amarelo na goethita ao vermelho roxo em algumas hematitas, podem ser observadas (SCHEINOST & SCHWERTMANN, 1999).

A hematita ($\alpha\text{-Fe}_2\text{O}_3$) é um óxido de ferro III que possui em média 69,9% de ferro em sua estrutura e o principal mineral de ferro dos minérios (CORNELL & SCHWERTMANN, 1996). Possui coloração avermelhada, é muito abundante em jazidas localizadas em Minas Gerais na região do quadrilátero ferrífero e no Pará, nas Minas de Carajás de Serra Norte e Serra Sul.

A goethita é um hidróxido de ferro ($\alpha\text{-FeOOH}$) que contém 62,9% de ferro em sua estrutura (CORNELL & SCHWERTMANN, 1996). Possui cor de traço amarelo, é um mineral que apresenta grandes níveis de contaminações por alumina, fósforo e outros que geralmente são prejudiciais nos processos das indústrias siderúrgicas e metalúrgicas, no qual o minério de ferro é destinado.

Já a magnetita é um óxido de ferro (Fe_3O_4) com cerca de 72,4% (CORNELL & SCHWERTMANN, 1996) e apresenta uma coloração preta. É um minério ferromagnético, com magnetismo natural e devido a essa característica, muitas vezes, dependendo de sua quantidade, pode sofrer interferências nos detectores de metais que são instalados ao longo dos

transportadores, para a remoção de sucatas provenientes do processo de beneficiamento mineral.

3.3 MINÉRIO DE FERRO

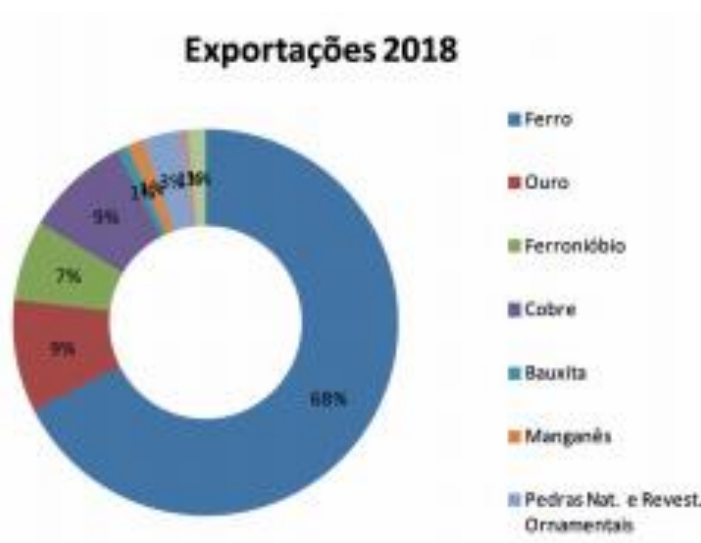
A condicionalidade econômica para aproveitamento dos minérios, está intrinsecamente ligada às condições geológicas e metalogenéticas das jazidas. Diversos fatores são influenciadores para que os empreendimentos minerários possam ser economicamente viáveis, como a mineralogia do minério, os teores de ferro, a estrutura e a textura das rochas que contêm o mineral-minério, a paragênese e toda uma série de parâmetros geológicos. (PINHEIRO, 2000).

De acordo com Quaresma (2001), o minério de ferro, em virtude de suas propriedades químico-física, é quase na sua totalidade utilizado na indústria siderúrgica, cerca de 99%. Os apenas 1% restantes são utilizados como carga na indústria de ferro-liga, cimento e eventualmente na construção de estradas.

Ainda segundo Quaresma (2009), em algumas jazidas encontradas no Brasil, como na Província Mineral de Carajás, o alto teor de ferro dispensa em alguns casos, os processos de concentração, podendo o minério ser utilizado diretamente, apenas com processos de cominuição para a adequação granulométrica. Dessa forma, já é possível atingir os padrões de qualidade, assim como dimensões uniformes e adequadas à boa operação dos processos siderúrgicos a que se destinam.

A indústria mineral se destaca por contribuir decisivamente para gerar superávits à balança comercial brasileira. Segundo relatório anual de atividades publicado pelo IBRAM (Instituto Brasileiro de Mineração), referente ao período de julho de 2018 a junho de 2019 o Brasil exportou em 2018 um volume de mais de 409 milhões de toneladas de bens minerais, e gerou divisas de US\$ FOB 29,9 bilhões. Este valor representou 12,5% das exportações totais do Brasil, e 36,6% do saldo comercial, além de empregar cerca de 195 mil trabalhadores diretamente. Do valor total de exportações, o minério de ferro correspondeu a um percentual de 68%, conforme mostra a estratificação da Figura 3.

Figura 3. Bens minerais exportados pelo Brasil no ano de 2018.



Fonte: Relatório anual de atividades – Julho 2018 a Junho 2019 – IBRAM_

Os principais produtos derivados da indústria extrativista de minério de ferro são o Granulado, o Sínter Feed e o Pellet Feed. O Granulado é o minério de ferro de maior granulometria, obtido após os processos de britagens e peneiramentos iniciais. Seu formato naturalmente aglomerado com dimensões de 50 a 6,3 mm permitem a carga direta no alto-forno. O teor de ferro do granulado possui em torno de 61%. O Sínter Feed é o minério de ferro mais utilizado pelas siderúrgicas no mundo e é obtido em fases posteriores do beneficiamento, os processos envolvidos na produção irão variar de acordo com as características geológicas da jazida como o percentual de ferro e nível de contaminantes. Possui granulometria na faixa de 6,3 a 0,150 mm e em torno de 63% de teor de ferro. Por fim, o pellet feed que é o minério mais fino com granulometrias abaixo de 0,150 mm, é obtido após o processo de flotação. Para que possa ser utilizado na siderurgia, o pellet feed deve passar pelo processo de pelotização. Em uma dosagem menor, pode também ser misturado ao sínter feed e aglomerado em usinas de sinterização. O teor de ferro do pellet feed gira em torno de 66% de ferro. A Figura 4, para fins de ilustração, mostra as características visuais de cada tipo comercial de minério de ferro.

O tipo de minério utilizado como objeto desse estudo é o Sínter Feed, com amostras coletadas da Usina de Carajás Serra Sul – PA.

Figura 4. Características visuais dos tipos comerciais de minério de ferro: granulado, sinter feed e pellet feed.



Fonte: Adaptado de CSN Mineração S.A.

3.4 PROCESSAMENTO DO SÍNTER FEED

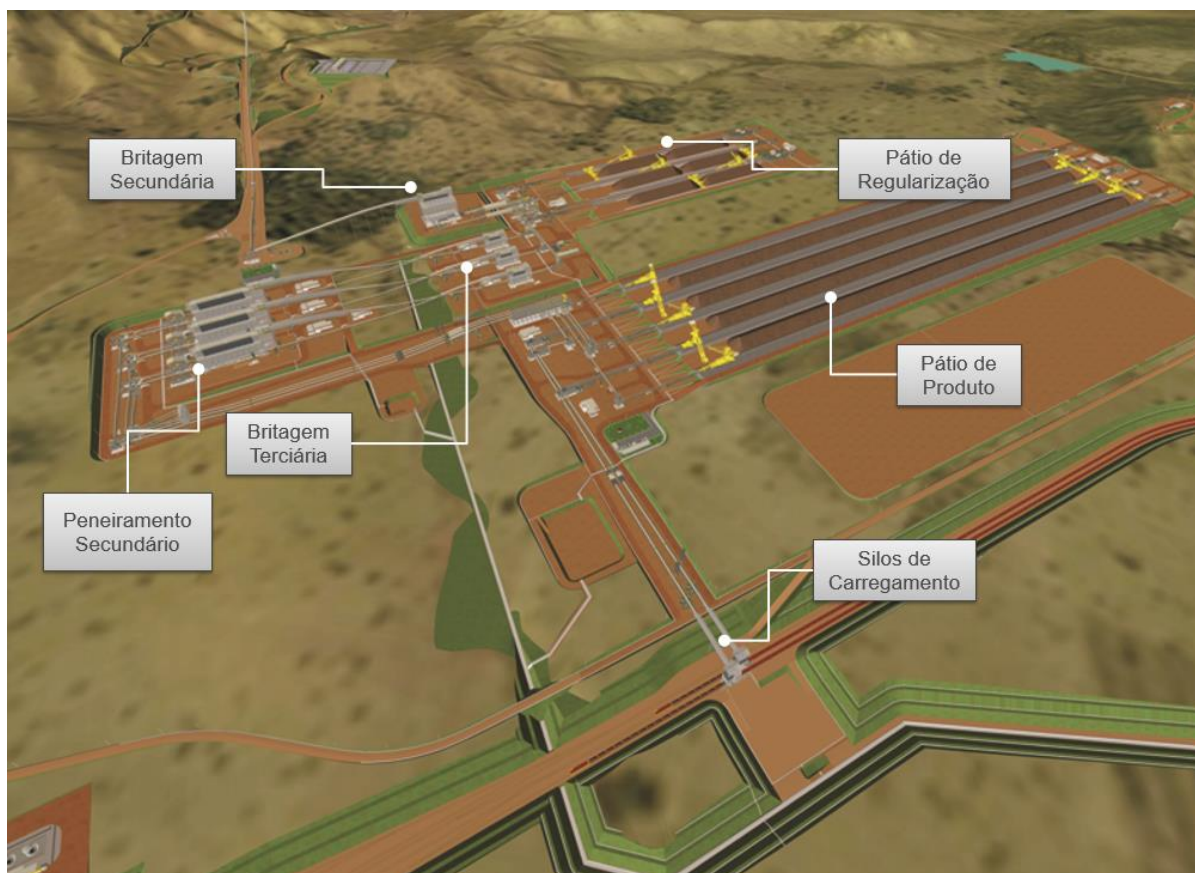
O processo de beneficiamento do sinter feed em Serra Sul é simplificado devido ao alto percentual de ferro da jazida, que gira em torno de 66%. O ROM (run-of-mine) proveniente da mina é britado através de britadores móveis localizados na frente de lavra e enviado a usina por meio de correias transportadoras de longa distância (TCLD).

O processo da usina também é simplificado e realizado em sua umidade natural, sem adição de água e por consequência sem a necessidade do uso de barragens de rejeito. O material proveniente da usina passa pela britagem secundária e em seguida é enviado ao peneiramento. O minério passante das peneiras é encaminhado ao pátio de produto e o material retido é endereçado a britagem terciária. O material re-britado na britagem terciária retorna ao peneiramento através de um circuito fechado, até atingir a granulometria necessária. A Figura 5 ilustra os processos compreendidos pela usina de beneficiamento de Serra Sul.

O pátio de regularização é utilizado como uma pilha intermediária e permite a flexibilização do processo. Dessa forma, após passar pela britagem secundária o minério tem a alternativa de ser empilhado para realização de homogeneização, antes de passar pelo peneiramento.

No pátio de produto é empilhado o material final e recuperado para os silos de carregamento, onde o sinter feed é embarcado em vagões e expedido em direção ao Porto da Madeira localizado em São Luís – MA.

Figura 5. Planta de beneficiamento da Usina Serra Sul.



Fonte: Os autores.

4 MATERIAL E MÉTODOS

4.1 AMOSTRAS

Neste estudo, foram utilizados dois grupos de amostras (Figura 6). O primeiro grupo abrange 58 amostras de sinter feed (material da frente de lavra britado), o qual possui granulometrias variadas e que corresponde ao produto final a ser expedido pela usina de beneficiamento. O segundo grupo compreende estas mesmas 58 amostras de sinter feed, porém preparadas em laboratório, quer dizer, submetidas a processos de pulverização e secagem. As amostras de sinter feed foram fornecidas pelo laboratório de Serra Sul, acompanhadas de dados de fluorescência de raios X (ênfase em elementos maiores: FeO, S, K₂O, Na₂O, CaO, MgO, P₂O₅; elementos menores: Ni, Cu, Co, Zn, P, Cl), conforme exibido na Tabela 2.

Figura 6. Amostras representativas de sinter feed da Mina de Carajás Serra Sul.



Fonte: Os autores.

Tabela 2. Resultado físico-químico de laboratório das amostras coletadas em Serra Sul.

Identificação da amostra	Fe [%]	SiO ₂ [%]	P [%]	Al ₂ O ₃ [%]	Mn [%]	TiO ₂ [%]	CaO [%]	MgO [%]	K ₂ O [%]	Identificação da amostra	Fe [%]	SiO ₂ [%]	P [%]	Al ₂ O ₃ [%]	Mn [%]	TiO ₂ [%]	CaO [%]	MgO [%]	K ₂ O [%]
CN_01_Prod	63.75	8.08	0.014	0.2	0.048	0.039	0.006	0.0254	0.0037	CN_10569_Prod	62.45	0.4	0.318	1.77	0.018	0.197	0.006	0.0254	0.007
CN_02_Prod	61.38	8.38	0.027	0.51	0.078	0.032	0.006	0.0254	0.015	CN_10573_Prod	63.94	0.89	0.205	1.52	0.011	0.2	0.006	0.0254	0.0037
CN_03_Prod	66.13	0.63	0.1	1.32	0.062	0.101	0.013	0.0254	0.0037	CN_10574_Prod	63.89	0.63	0.221	1.44	0.015	0.192	0.006	0.03	0.0037
CN_04_Prod	61.43	7.66	0.03	0.27	0.067	0.04	0.006	0.0254	0.0037	CN_10575_Prod	62.62	0.82	0.219	1.97	0.015	0.224	0.006	0.0254	0.0037
CN_05_Prod	53.29	20.08	0.025	0.54	0.035	0.025	0.006	0.031	0.022	CN_10576_Prod	61.61	0.58	0.374	1.88	0.008	0.239	0.006	0.0254	0.0037
CN_06_Prod	51.31	23.83	0.021	0.38	0.039	0.037	0.006	0.03	0.006	CN_10577_Prod	62.77	0.93	0.266	1.45	0.012	0.215	0.006	0.0254	0.0037
CN_07_Prod	67.79	0.36	0.145	0.92	0.019	0.113	0.006	0.033	0.0037	CN_10578_Prod	63.18	0.46	0.262	1.43	0.008	0.219	0.006	0.0254	0.0037
CN_08_Prod	46.52	32.62	0.016	0.23	0.048	0.028	0.006	0.056	0.0037	CN_10579_Prod	62.18	0.58	0.225	2.68	0.01	0.279	0.008	0.0254	0.0037
CN_09_Prod	52.34	23.23	0.018	0.24	0.047	0.039	0.006	0.062	0.0037	CN_10580_Prod	63.54	0.66	0.212	1.45	0.015	0.204	0.006	0.0254	0.0037
CN_10_Prod	58.14	12.42	0.025	1.45	0.071	0.031	0.006	0.027	0.012	CN_10581_Prod	62.43	0.63	0.177	2.45	0.01	0.231	0.006	0.0254	0.0037
CN_10550_Prod	60.55	0.38	0.469	3.03	0.017	0.436	0.006	0.028	0.004	CN_10582_Prod	61.98	9.65	0.012	0.41	0.112	0.042	0.006	0.1	0.0037
CN_10551_Prod	61.12	0.39	0.434	2.82	0.018	0.402	0.006	0.03	0.006	CN_10583_Prod	58.21	14.91	0.011	0.6	0.134	0.048	0.006	0.105	0.0037
CN_10552_Prod	61.76	0.4	0.457	2.49	0.018	0.407	0.006	0.027	0.007	CN_10584_Prod	41.41	39.45	0.014	0.37	0.023	0.067	0.006	0.096	0.0037
CN_10553_Prod	61.97	0.39	0.376	2.25	0.018	0.257	0.006	0.027	0.007	CN_10585_Prod	68.24	1.21	0.011	0.22	0.038	0.04	0.006	0.124	0.0037
CN_10554_Prod	61.99	0.43	0.407	2.44	0.019	0.253	0.006	0.027	0.006	CN_10757_Prod	62.78	0.56	0.446	1.9	0.01	0.126	0.006	0.043	0.009
CN_10555_Prod	61.52	0.36	0.476	2.32	0.019	0.262	0.006	0.027	0.006	CN_10758_Prod	64.11	0.58	0.258	1.55	0.01	0.124	0.007	0.048	0.008
CN_10556_Prod	60.45	0.35	0.519	2.9	0.018	0.482	0.006	0.026	0.005	CN_10759_Prod	62.92	0.58	0.353	2.27	0.01	0.19	0.007	0.05	0.009
CN_10557_Prod	62.15	0.57	0.362	2.29	0.018	0.16	0.006	0.031	0.007	CN_10760_Prod	64.37	0.63	0.364	1.33	0.015	0.084	0.007	0.057	0.012
CN_10558_Prod	60.46	0.32	0.535	2.93	0.018	0.497	0.006	0.0254	0.005	CN_10761_Prod	64.85	0.6	0.255	1.27	0.009	0.099	0.007	0.045	0.013
CN_10559_Prod	61.12	0.36	0.457	2.44	0.018	0.41	0.006	0.0254	0.006	CN_10762_Prod	65.66	0.72	0.139	0.97	0.013	0.082	0.008	0.062	0.016
CN_10560_Prod	61.13	0.34	0.463	2.42	0.018	0.404	0.006	0.026	0.007	CN_10763_Prod	65.35	0.62	0.133	0.98	0.01	0.094	0.007	0.046	0.01
CN_10561_Prod	62.56	0.45	0.317	1.94	0.02	0.19	0.006	0.028	0.007	CN_10764_Prod	64.8	0.56	0.171	1.18	0.011	0.131	0.008	0.046	0.009
CN_10562_Prod	63.34	0.47	0.293	1.61	0.019	0.19	0.006	0.028	0.007	CN_10765_Prod	66.11	0.71	0.064	0.6	0.016	0.068	0.008	0.064	0.01
CN_10563_Prod	63.26	0.48	0.262	1.69	0.02	0.245	0.006	0.034	0.008	CN_10766_Prod	65.38	0.61	0.13	1.11	0.014	0.131	0.007	0.056	0.008
CN_10564_Prod	64.15	0.67	0.205	1.47	0.022	0.176	0.006	0.035	0.01	CN_11_Prod	63.31	5.93	0.031	0.63	0.172	0.045	0.006	0.029	0.0037
CN_10565_Prod	64.44	0.49	0.231	1.29	0.02	0.139	0.006	0.034	0.007	CN_12_Prod	49.52	25.21	0.036	0.5	0.201	0.034	0.006	0.035	0.0037
CN_10566_Prod	63.56	0.59	0.252	1.58	0.02	0.167	0.006	0.027	0.011	CN_13_Prod	48.78	28.18	0.023	0.44	0.072	0.069	0.006	0.078	0.0037
CN_10567_Prod	62.06	0.51	0.347	2.41	0.019	0.418	0.006	0.027	0.008	CN_14_Prod	54.89	15.72	0.045	1.46	0.026	0.075	0.006	0.047	0.235
CN_10568_Prod	61.93	0.41	0.405	2.15	0.018	0.424	0.006	0.0254	0.007	CN_15_Prod	67.07	0.91	0.019	0.18	0.017	0.037	0.006	0.086	0.0037

Fonte: Os autores.

4.2 FLUORESCÊNCIA DE RAIOS X (FRX)

O método de análise por espectrometria de fluorescência de raios X (FRX) é uma técnica analítica muito utilizada em geoquímica, com aplicação industrial para análise de amostras de minérios, rochas, ligas, pigmentos, cerâmicas, dentre outros, para fins de qualidade. (TAVARES, 2014).

A técnica se baseia na medição das intensidades dos raios-X característicos emitidos por espectrômetro de fluorescência de raios-X. Os elementos que constituem a amostra, quando excitados por esse feixe de radiação, emitem linhas espectrais com energias características dos elementos que a constituem e cujas intensidades estão relacionadas com a concentração. Dessa forma, o processo é composto basicamente por três etapas: a excitação dos elementos que compõem a amostra, dispersão dos raio-X característicos emitidos pela amostra e a detecção desses raios-X. Dessa forma é possível mensurar os elementos que constituem a amostra, assim como seu quantitativo. (SANTOS et al., 2013).

4.3 ESPECTROSCOPIA DE REFLETÂNCIA

4.3.1 Instrumentação

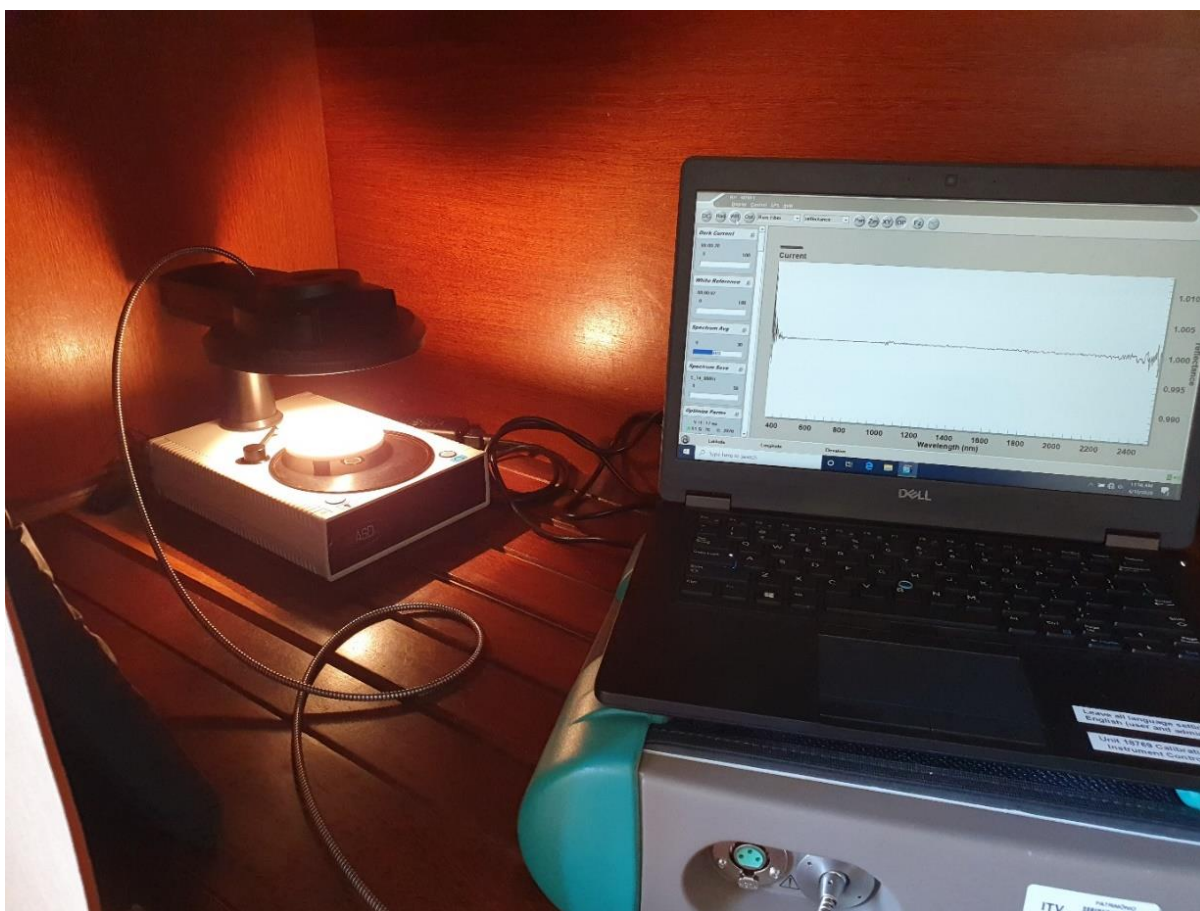
O instrumento utilizado para aquisição dos dados de refletância no intervalo *VSWIR* foi o espectrorradiômetro de alta resolução *FieldSpec 4 High Resolution Next Generation* acoplado ao equipamento de leitura *Turntable* (Figura 7), ambos do fabricante *Analytical Spectral Devices* (ASD, 2019).

O *FieldSpec 4 Hi-Res NG* detecta a radiação eletromagnética no intervalo espectral entre 350 nm até 2500 nm, com resoluções espectrais de 3 nm (@700 nm) e 6 nm (@1400 nm/2100 nm). Os intervalos de amostragem são de 1.4 nm entre 350 – 1000 nm, e 1.1 nm na faixa de 1000 nm – 2500 nm, com 2151 canais.

O *Turntable* possui uma mesa giratória ideal para obter uma média de amostras com granulometria variada.

Os dados de refletância foram gerados a partir de medidas de energia refletida pelo alvo em relação a um material de referência (*Spectralon*) em toda a faixa *VSWIR*.

Figura 7. Espectrorradiômetro FieldSpec 4 Hi-Res NG, o acessório Turntable e software RS3 utilizados para realização das leituras espectrais.



Fonte: Os autores.

4.3.2 Leituras Espectrais

A leitura dos espectros foi realizada através do *software* RS³ em conjunto com os equipamentos *Turntable* e *FieldSpec 4*. As amostras foram colocadas em placas Petri com dimensões 150 mm x 15 mm para a realização das leituras através do espectrorradiômetro (Figura 8). Cada amostra foi medida 100 vezes, considerando 50 escaneamentos. Assim, cada espectro utilizado na caracterização espectral corresponde à média de 5000 leituras.

Figura 8: Detalhe do acondicionamento da amostra para a realização da leitura.



Fonte: Os autores.

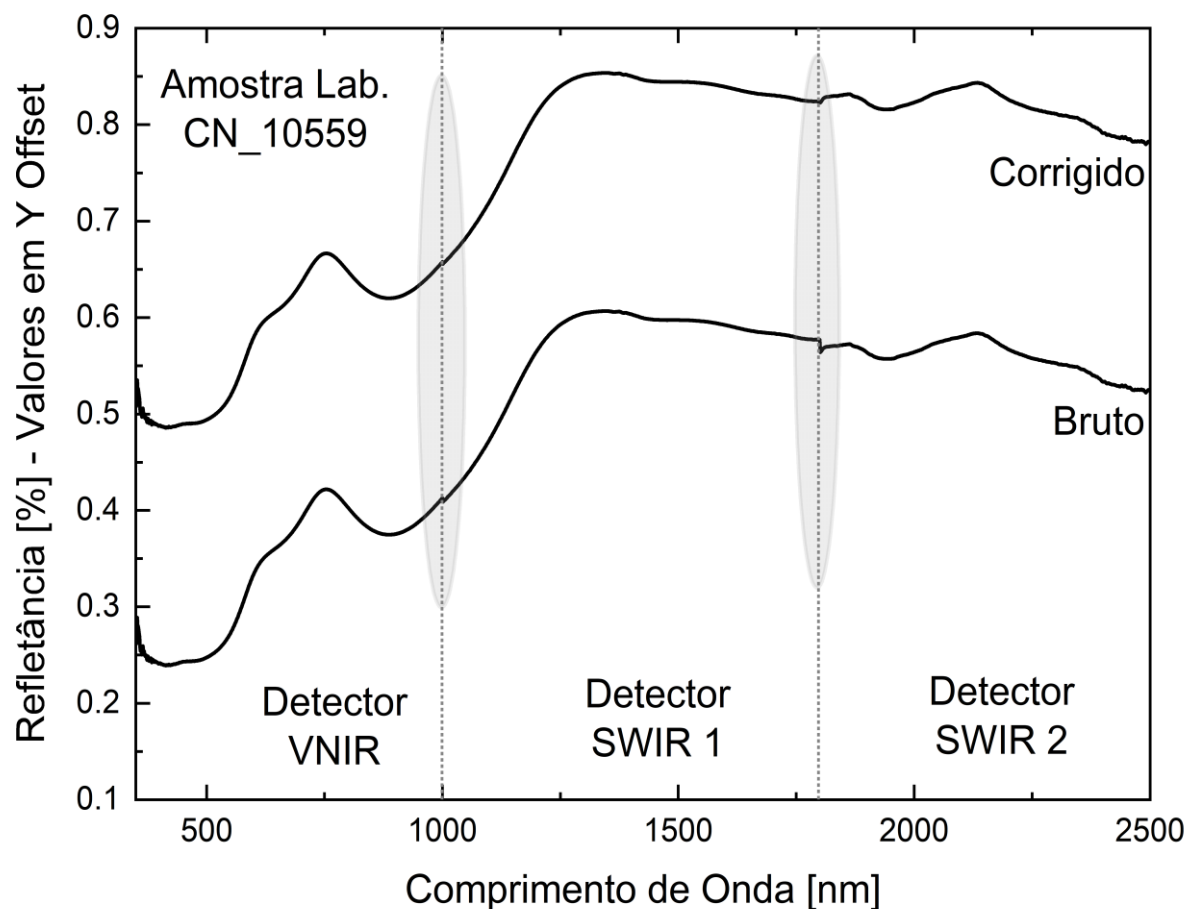
4.4 PRÉ-PROCESSAMENTO DOS DADOS

4.4.1 Correção dos gaps entre os detectores – “*Splice Correction*”

O pré-processamento de dados espectrais é essencial, visto que existem vários tipos de ruídos (e.g. ruídos associados aos detectores do sensor e circuitos eletrônicos) que podem afetar os espectros (OZAKI et al., 2006). No caso de espectros coletados com o *FieldSpec 4 Hi Res NG*, é necessário realizar a correção dos gaps, devido ao sensor possuir três diferentes detectores. Esta correção é realizada através do software *ViewSpec Pro*. Os gráficos das Figura 9, exemplificam a diferença do espectro, após a aplicação da ferramenta de remoção das lacunas entre os sensores.

Após o pré-processamento dos dados, foram elaborados, principalmente, gráficos de comprimento de onda versus fator de refletância. Os espectros das amostras foram analisados, identificando-se as principais feições características dos minerais de ferro, permitindo avaliar quais são os comprimentos de onda mais relevantes que permitiram caracterizar os minerais de ferro e contaminantes como alumina, magnésio e fósforo, presentes nas amostras de *sinter feed*.

Figura 9. Espectro de refletância bruto e corrigido através do splice correction da amostra de laboratório CN_10559.

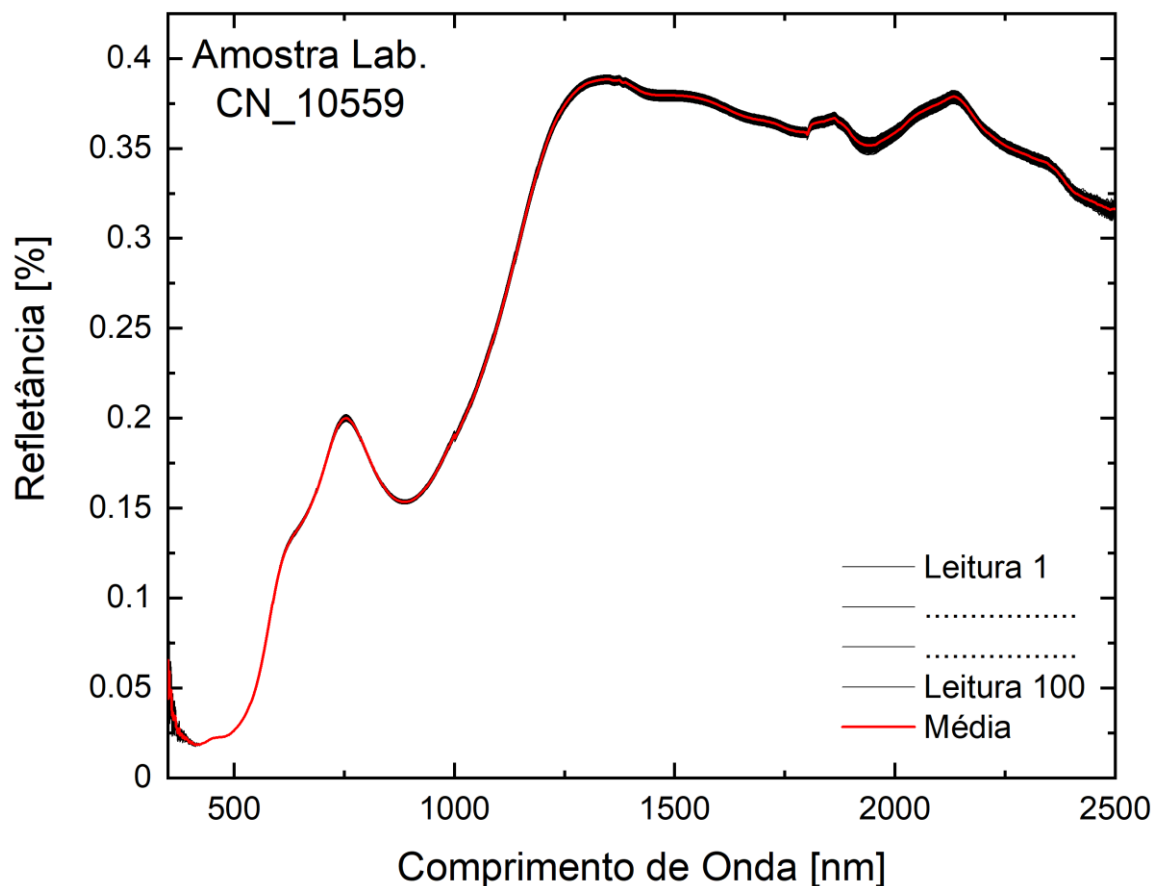


Fonte: Os autores.

4.4.2 Média

Após a correção dos gaps nos espectros, foi calculada a média das 100 leituras espectrais coletadas (Figura 10). Através da média, ruídos e artefatos podem ser reduzidos ou eliminados permitindo uma interpretação mais confiável das feições de absorção.

Figura 10. Espectro de refletância médio e leituras espectrais realizadas da amostra de laboratório CN-10559.



Fonte: Os autores.

4.4.3 Criação da Biblioteca Espectral

A biblioteca espectral consiste em organizar um banco de dados com as curvas do espectro de refletância coletadas das amostras, que correspondem as assinaturas espectrais de interesses a serem inseridas em *softwares* computacionais, os quais auxiliam na aplicação das técnicas de análise da informação espectral. Os espectros aqui coletados e analisados integrarão a Biblioteca Espectral Vale (ver exemplo no Anexo 1).

Portanto, após a realização do processamento dos dados espectrais das amostras de síntese de Serra Sul, foi utilizado o *ENVI* para criação da biblioteca espectral característica. O detalhe do passo a passo do uso do software está detalhado no Anexo 2.

4.5 ANÁLISE DOS DADOS ESPECTRAIS

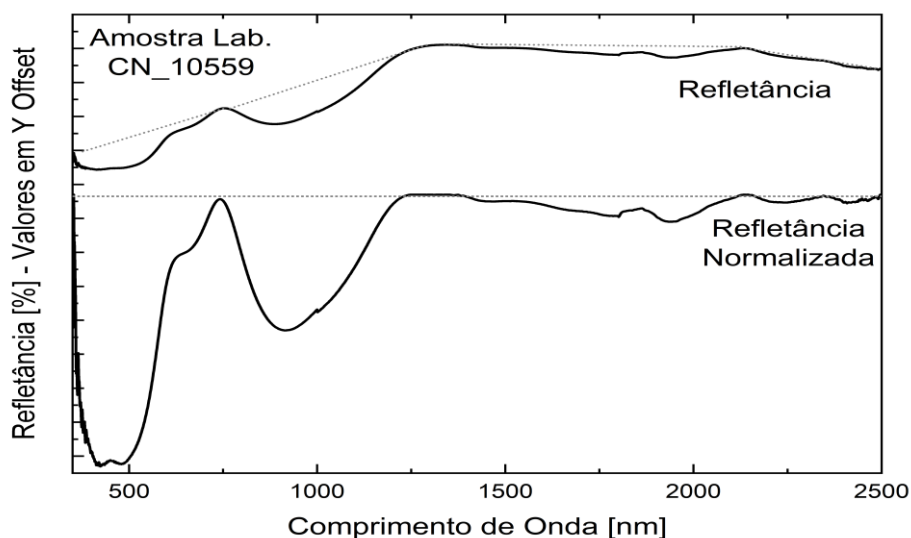
4.5.1 Remoção do contínuo

Com o intuito de realçar as feições de absorção espectrais e eliminar ou reduzir os efeitos não relacionados com as propriedades de interesse dos alvos analisados, transformações espectrais (e.g. derivadas, remoção do contínuo) são frequentemente utilizadas. A remoção do contínuo é um método que foi inicialmente utilizado para estudar a composição bioquímica de folhas secas (CLARK & ROUSH, 1984). Esse método normaliza o espectro de refletância e isola as feições de absorção de forma a permitir a comparação de bandas de absorção sob uma base comum (KOKALY, 2001).

A remoção do contínuo foi utilizada neste projeto para realçar as feições espectrais específicas que possam auxiliar na caracterização dos minerais associados ao ferro e possíveis contaminantes.

O software *ViewSpecPro* foi empregado para a criação da base de dados com remoção do contínuo. Para tal, foram utilizados os dados anteriormente pré processados pela ferramenta *Splice Correction* e realizada a remoção do contínuo pela funcionalidade *Convex Hull*. Foi criada uma base de dados com 5800 arquivos tipo “HR”, que representam a curva espectral com o contínuo removido. A Figura 11 exibe a curva espectral sem a normalização e com a aplicação da técnica do contínuo removido da amostra CN_10568 de sínter feed.

Figura 11 - Acima a representação do espectro da Amostra CN_10568 e abaixo o espectro normalizado da mesma amostra.



Fonte autores.

Portanto, a técnica da remoção do contínuo tem como objetivo a normalização dos dados. Essa normalização é importante, pois todas as informações são representadas na mesma ordem de grandeza, na escala de 0 a 1, e comparadas com o mesmo nível de relevância.

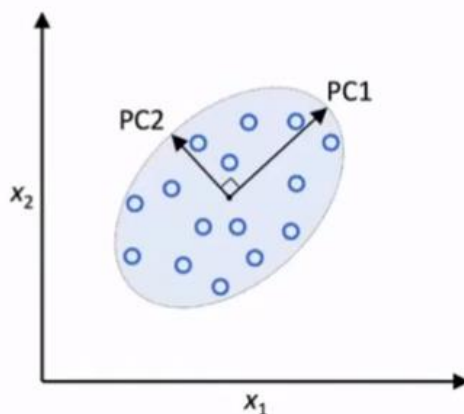
4.5.2 Análise por componentes principais

A análise de componentes principais (PCA – *Principal Component Analysis*) foi explorada com o objetivo de melhorar a extração de informações relacionadas ao conteúdo de ferro nas amostras de sinter feed. Na PCA, as variáveis (componentes) originais, que são altamente correlacionadas, são transformadas em novas variáveis e a informação de interesse é concentrada em componentes específicas. A primeira componente é aquela que agrupa a informação comum ao conjunto original de bandas. A escolha das componentes de interesse é feita a partir da análise da matriz de autovetores, a qual apresenta a contribuição de cada banda original na composição da componente principal gerada (LANDIM, 2011).

A análise de componentes principais, portanto, analisa vários dados juntos, pois os mesmos podem estar correlacionados, além de ser a maneira mais simples de fazer a ordenação das informações. A PCA reduz a dimensionalidade de dados multivariados, ou seja, são criadas poucas variáveis chaves, onde cada uma é composta dos dados originais, de forma a caracterizar o máximo possível da variação da base de dados estudada (GOTELLI & ELLISON, 2011).

O primeiro passo para transformação dos dados é realizar a normalização, de forma que todas as informações que originalmente possuem escalas diferentes, tenham o mesmo peso e sejam considerados com a mesma relevância. A Figura 12 exemplifica o conceito do PCA, onde os novos eixos PC1e PC2 representam os dados com maior variância e que melhor irão representar o conjunto de dados originais. Os dados gerados nos componentes do PCA não têm significado direto na análise dos dados e são apenas uma combinação linear dos demais atributos.

Figura 12 - Representação do modelo de análise de componentes principais (PCA).



Fonte: Adaptado de Landim, 2011.

4.6 MÉTODOS DE APRENDIZADO DE MÁQUINA

Algoritmos de aprendizado de máquina, são poderosos métodos de análise de dados utilizados no mapeamento da prospectividade mineral, pois seus resultados têm mostrado mais precisão e velocidade de resposta em comparação aos métodos estatísticos convencionais. Em alguns exemplos de aplicações, temos: o trabalho de Viana (2020), que utiliza métodos de aprendizado de máquina e imageamento hiperespectral, para estimar teores de ferro de minérios. Já Leite e de Souza Filho (2009), revisa o uso de Redes Neurais Artificiais (RNAs) aplicadas no mapeamento de potenciais minerais, para mineralizações de cobre e ouro. Na otimização da análise de dados hiperespectrais, Hu et al. (2019) apresentou o conceito de seleção de bandas espectrais para a remoção daquelas redundantes, mas com a preservação da significância física das informações estudadas. O número excessivo de bandas redundantes, aumenta a complexidade computacional, exige maior carga de processamento e custo de armazenamento de dados, onde o trabalho trouxe ganhos nesse sentido.

Chen et al. (2008) ressaltam que a relevância do uso da inteligência artificial, tem aumentado nos últimos anos nas áreas de modelagem ambiental, pelo fato da obtenção de resultados cada vez mais satisfatórios. O conceito de inteligência artificial (IA) é a reprodução da inteligência humana através de modelos cognitivos de computação. O aprendizado de máquina é uma forma de IA onde um algoritmo computacional constrói, a partir de dados conhecidos, modelos de aprendizados para a resolução de problemas (KUMAR et al., 2007). Seu objetivo é de criar modelos de *software*, que são treinados com grandes volumes de dados e usados para prever resultados, tendências e padrões (WATT et al., 2020).

Os principais algoritmos de aprendizado de máquina, devido a sua grande presença na literatura aplicados a problemas de classificação e regressão são as Redes Neurais Artificiais (RNA), as Árvores de Decisão, o *Random Forest* (RF), o *Adaboosting*, o KNN e o SVM.

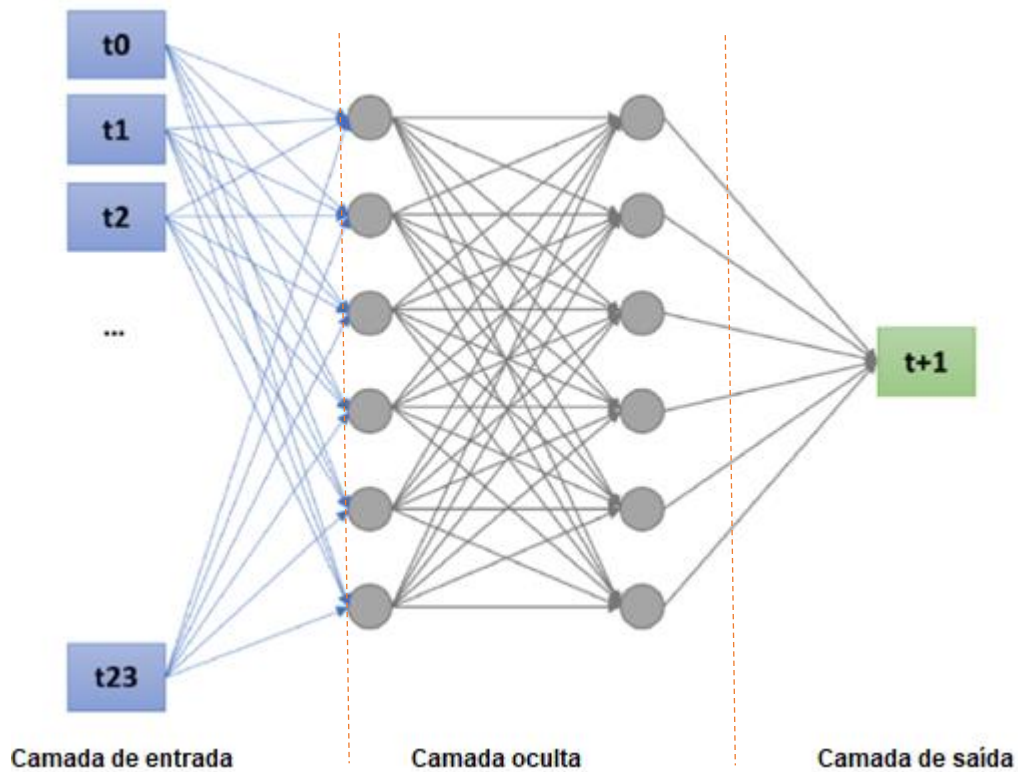
4.6.1 Redes Neurais Artificiais (RNA)

Redes Neurais Artificiais (RNA) são compostas por vários elementos computacionais que interagem através de conexões com diferentes pesos. Em analogia ao cérebro humano, as redes neurais exibem características como a capacidade de aprender padrões complexos de dados e generalizar as informações aprendidas (ZHAO et al., 2012).

O processo de aprendizagem de uma RNA é dividido em três partes, que definem como as mudanças de parâmetros ocorrem. Inicialmente, é necessário o estímulo da rede com a extração de exemplos de um ambiente. No processo intermediário, existe a modificação dos pesos através de processos iterativos, com o objetivo de minimizar o erro de saída da RNA. Por fim, a rede entrega uma resposta, que é o resultado das mudanças ocorridas.

A forma mais simples de uma RNA é o *Multilayer Perceptron* (MLP), que consiste em três camadas: a camada de entrada, a camada oculta e a camada de saída. Um *Perceptron* produz uma única saída com base em várias entradas de valor real e a RNA pode ser composta de um ou mais *Perceptrons*. A Figura 13 mostra um exemplo clássico de uma topologia MLP totalmente conectada, com camadas de entrada, ocultas e de saída.

Figura 13. Esquema de uma RNA Multilayer Perception com duas camadas ocultas e 6 neurônios cada.



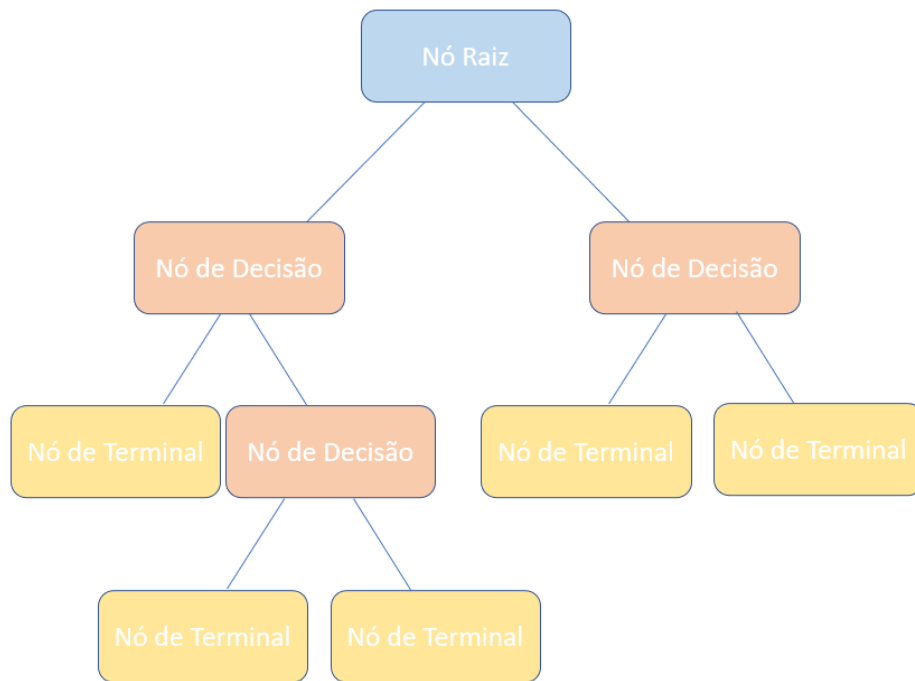
Fonte: Adaptado de SANTOS et al., 2019.

4.6.2 Árvore de Decisão

Árvore de Decisão é um método de decisão semelhante a um fluxograma que representa o processo. Possui nós que indicam a decisão tomada em um atributo e ramos, que indicam as possíveis escolhas para cada decisão (LANTZ, 2019).

Na representação da árvore de decisão, um conjunto de dados parte de um nó inicial, conhecido como nó raiz da árvore. Na sequência, são realizados testes lógicos e de acordo com o resultado adotado pelo nó, a árvore ramifica-se para um dos nós filhos e este procedimento é repetido até que o nó terminal seja alcançado, conforme exemplificado na Figura 14. A repetição deste procedimento é realizada de forma recursiva (CIAMPI et al., 2000).

Figura 14. Representação de uma árvore de decisão.



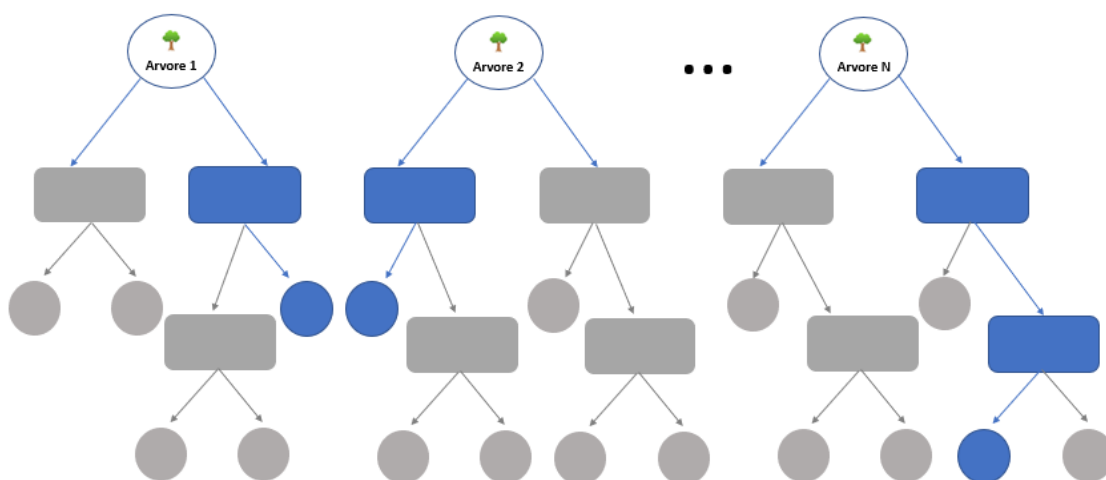
Fonte: Os autores.

4.6.3 Random Forest (RF)

Random forest (RF) são conjuntos de árvores de decisão que votam juntos em uma classificação. Cada árvore é construída por acaso e seleciona um subconjunto de recursos aleatoriamente de um subconjunto de pontos de dados. A árvore é treinada nesses pontos de dados (somente nas características selecionadas), e o restante “fora do cesto” é usado para avaliar a árvore. As RF são conhecidas por serem efetivas na prevenção do *overfitting* (CLUTER et al., 2012). Se houverem árvores suficientes, o RF não irá sobreajustar o modelo.

Este método funciona eficientemente em grandes bases de dados, pois trabalha com os dados de treino e desta forma, o algoritmo busca as melhores condições, e onde inserir cada uma dentro do fluxo (BIAU & SCORNET, 2016). A Figura 15 exibe a estrutura de uma rede *Random Forest*.

Figura 15. Esquema de uma Random Forest com “N” conjuntos de árvore de decisão.



Fonte: Os autores.

4.6.4 Adaboosting

O *Adaboosting* é um método de aprendizado de máquinas que utiliza múltiplos classificadores, baseado em uma combinação de classificadores com mais baixas acurácias, onde o resultado produz um classificador com acurácia mais elevada. Dessa forma, para o treinamento dos dados, o método induz a interação desses vários modelos de classificadores e a cada interação, o método *boosting* gera uma hipótese. Cada nova hipótese gerada, tem o objetivo de corrigir os erros provenientes da hipótese anterior. O processo é repetido até que a etapa de treinamento seja finalizada (BORGES, 2019).

O Algoritmo, que geralmente é composto por árvores de decisão, é utilizado para o treino de cada modelo classificador e conhecido como algoritmo base (FREUND et al., 1996).

4.6.5 KNN (K-Nearest Neighbor)

O algoritmo KNN (*K-Nearest Neighbor*) é um modelo simples, usado em problemas tanto de classificação quanto de regressão, que realiza a previsão de valores futuros, baseado em valores registrados passados. Nesse método, a variável “K” irá direcionar a quantidade de vizinhos, dessa forma, o algoritmo busca os valores desejados mais próximos àquele ponto, tendo por base as distâncias dos seus “K” vizinhos mais próximos (ALPAYDIN, 2020).

O cálculo para determinação dos vizinhos mais próximos, podem ser realizados utilizando diversos métodos matemáticos para cálculo de distância entre 2 pontos como (LINDEN,2009):

- Distância Euclidiana: $d(p,q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$ (1)

- Distância de Manhattan: $d(p,q) = \sum_{i=1}^n |p_i - q_i|$ (2)

- Distância de Mahalanobis: $d(p,q) = \sqrt{\sum_{i=1}^n \frac{(p_i - q_i)^2}{\pi_i^2}}$ (3)

- Distância de Chebyshev: $d(p,q) = \max_k |p_{ik} - q_{ik}|$ (4)

Na determinação dos valores a serem atribuídos a “K”, caso sejam muito baixos, serão afetados por ruídos nos dados, de forma a influenciar no resultado final obtido, que o torna mais sensível a regiões bem próximas, podendo ocorrer *overfitting*. Porém, valores muito altos, com a geração de vários vizinhos, podem ser considerados mais robusta, todavia computacionalmente custosa. (LANTZ, 2019).

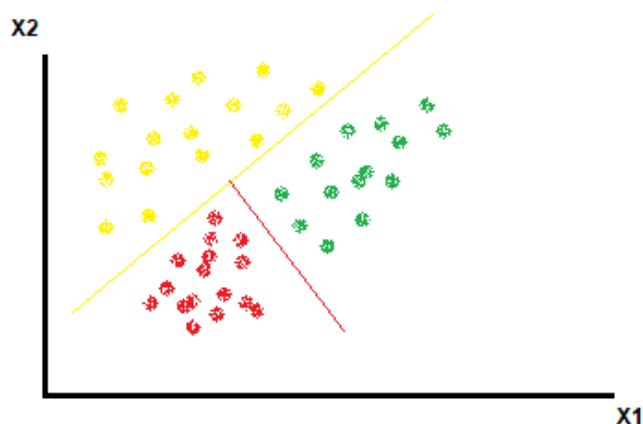
4.6.6 SVM (Support Vector Machine)

O algoritmo SVM (*Support Vector Machine*) é um método de aprendizado de máquina inicialmente concebido, com o foco na solução de problemas de classificação binária. Posteriormente, a técnica foi aprimorada e utilizada também para problemas de classificação multiclases e de regressão (FILGUEIRAS, 2014).

O conceito do SVM é a separação dos dados em classes, para tal, o algoritmo cria hiperplanos, onde cada dado pertencente a base de dados de treinamento, é plotado como um ponto no espaço n-dimensional e com referência a uma coordenada. Dessa forma, é executada a classificação, com o objetivo de encontrar o hiperplano que irá diferenciar as classes, conforme ilustrado na Figura 16 (CHEN, 2004).

O vetor suporte, define as fronteiras que melhor irá segregar as classes, ou seja, as coordenadas dos pontos limites entre classes, assim o método se torna menos eficiente quando o conjunto de dados apresentam muitos ruídos, o que faz com que as classes de destino estejam sobrepostas. O vetor suporte encontra a distância mínima da fronteira com outro vetor suporte. Com esse princípio, apenas uma solução ótima é encontrada para o problema abordado, de forma que as soluções se convergem para um mínimo global (FILGUEIRAS, 2014).

Figura 16. Representação dos hiperplanos de dados do método SVM.



Fonte: Os autores.

4.7 AVALIAÇÃO DA ACURÁCIA DOS MÉTODOS DE APRENDIZADO DE MÁQUINA

A estatística é uma ciência exata que estuda a coleta, a organização, a análise e registro de dados por amostras. Esta é uma parte da matemática aplicada que fornece dados para a coleta, organização, descrição, análise e a interpretação de dados para a utilização dos mesmos na tomada de decisão (MARÔCO, 2018).

Neste trabalho conceitos de *overfitting* e *underfitting*, e métodos estatísticos como o Coeficiente de Determinação (R^2), Erro Médio Absoluto (MAE), Erro Médio Quadrático (MSE) e Erro Percentual Absoluto Médio (MAPE) foram utilizados para se obter as relações entre as variáveis e os resultados. Dessa forma, foram avaliados os erros entre os valores que foram estimados referentes aos teores de ferro. A seguir é apresentada uma breve revisão dos métodos acima citados.

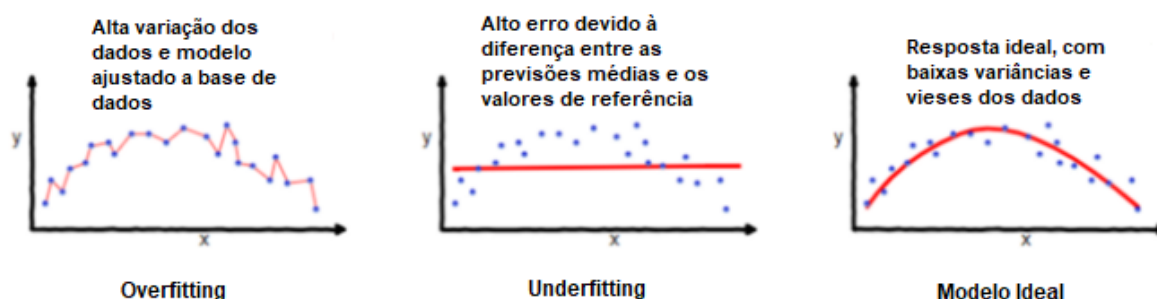
4.7.1 *Overfitting* e *Underfitting*

O termo conhecido como *overfitting* é utilizado para descrever quando um modelo se ajusta a base de dados de treino, porém se mostra ineficiente quando submetido aos dados de validação, o que o torna ineficaz na previsão de novos resultados. Dessa forma, percebe-se que o mesmo apenas decorou os dados destinados ao treinamento e ao tentar aplicar essas mesmas regras aos novos dados para predição, o resultado obtido se torna completamente fora do esperado. (MONARD & BARANAUSKAS, 2003).

Já no conceito de *underfitting*, ainda na etapa de treinamento dos dados, o modelo já não encontra uma boa relação entre os elementos, dessa forma, o resultado esperado é ruim. Isso ocorre em geral, quando a base de dados é composta por um número de dados pequenos ou pouco representativos para o modelo, assim se torna insuficiente para aplicação dos métodos de aprendizado (BROWNLEE, 2016).

A Figura 17 mostra os exemplos de respostas dos modelos quando ocorrem os efeitos do *overfitting* e *underfitting* comparados a um modelo ideal que se espera como resultado dos métodos de aprendizado de máquina.

Figura 17. Exemplos de respostas dos modelos quando ocorrem os efeitos do *overfitting* e *underfitting* comparados a um modelo ideal.



Fonte: Os autores.

4.7.2 Coeficiente de determinação (R^2)

O coeficiente de determinação, também chamado de R^2 é uma medida de ajustamento de um modelo estatístico linear generalizado, como a regressão linear, em relação aos valores observados (FIGUEIREDO FIIHO & SILVA JÚNIOR, 2009). O valor de R^2 varia entre 0 e 1, que representa o quanto o modelo consegue explicar os valores observados. Portanto, quanto maior o valor de R^2 , mais explicativo é o modelo e melhor ele se ajusta a amostra.

4.7.3 Erro médio quadrático (MSE)

Segundo Gupta et al. (2009), o erro quadrático médio é definido como sendo a média da diferença entre o valor do estimador e do parâmetro ao quadrado. O erro médio quadrático é um dos dois principais indicadores de desempenho para um modelo de previsão de regressão. Ele mede a precisão do modelo calculando a média dos quadrados das diferenças entre os valores previstos e reais.

4.7.4 Erro percentual médio absoluto (MAPE)

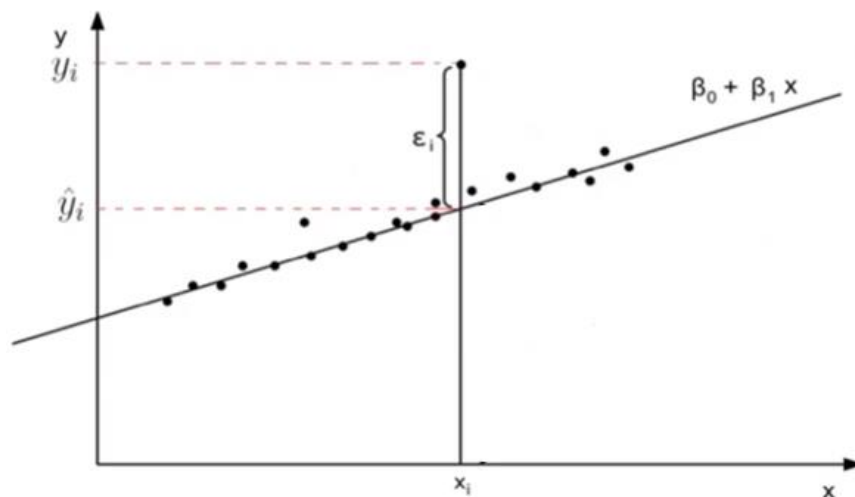
O erro percentual médio absoluto representa a média percentual da divisão entre erro de previsão e o valor real. Também conhecido como desvio percentual médio absoluto, é uma medida da precisão de previsão de um método de previsão em estatística. Na estimativa de tendência, também usado como uma função de perda para problemas de regressão no aprendizado de máquina (DE MYTTENAERE et al., 2016).

4.7.5 Análise por regressão linear

A análise por regressão possibilita encontrar uma relação razoável entre as variáveis de entrada e saída, por meio de relações empíricas (WEISBERG, 2013).

O método de regressão linear fornece a expressão de relação entre 2 variáveis através de uma reta (FIGUEIREDO FILHO & SILVA JÚNIOR, 2009). Como exibido no gráfico da Figura 18 a equação da reta que mais se aproxima os dados que estão representados por pontos, representa uma relação linear entre duas variáveis x e y . Através do método dos mínimos quadrados, a equação da reta que melhor aproxima os pontos, dever ser de forma que a soma dos quadrados dos valores das distâncias entre cada ponto e a reta, seja o menor possível.

Figura 18. Representação do método de regressão linear.



Fonte: Adaptado de Figueiredo Filho e Silva Júnior, 2009.

Índice de correlação mede o grau de relação entre as variáveis, esse coeficiente é conhecido como coeficiente de correlação de Pearson e varia entre -1 e 1, sendo que o sinal negativo representa uma correlação inversa e o sinal positivo uma correlação direta. Para valores iguais a zero significa que a correlação é nula, ou seja, inexistente, para valores quanto mais próximos de 1 (independente do sinal) o grau de aderência ao modelo é mais forte (FIGUEIREDO FILHO & SILVA JÚNIOR, 2009).

No caso do gráfico representado na figura 4 existe uma correlação linear positiva, o erro da variável menos significativa ou *outlier* dentre as amostras representadas, é mensurada por ε_i e determinado pela distância desse ponto y_i até a reta $(\beta_0 + \beta_1 x)$.

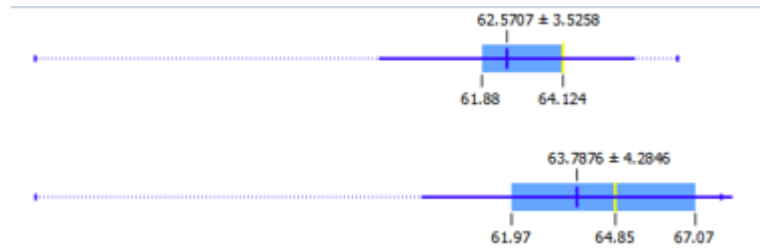
A presença de *outliers*, ou seja, dados que destoam drasticamente de todos os outros, têm a tendência de distorcer o coeficiente de correlação. Dessa forma, comprometem as estimativas nas análises dos dados, pois nessa determinada amostra considerada “fora do padrão”, pode não existir dependência estatisticamente entre as variáveis x e y , porém se considerados no cálculo de correlação podem alterar o padrão de associação entre as variáveis. Estas situações devem ser analisadas através de inspeção gráfica para que não comprometam as análises. (FIGUEIREDO FILHO & SILVA JÚNIOR, 2009).

4.7.6 Boxplot

O *boxplot* é uma ferramenta que permite avaliar a homogeneidade dos dados, ou seja, o quanto os dados estão dispersos ou agrupados. O diagrama possui uma linha transversal que indica onde está a mediada dos dados. Portanto, se a caixinha do boxplot for pequena, significa que os dados estão mais próximos da mediana, em contrapartida, se for maior, reflete que os dados são mais heterogêneos. (MUNDSTOCK et al., 2006)

A Figura 19, exemplifica duas representações de gráficos boxplot, sendo a primeira com dados mais agrupados e portanto mais homogêneos e na segunda, mais dispersos, que exibe uma caixa de tamanho maior. A linha transversal da mediada é referente ao segundo quartil e as linhas que delimitam a caixa a esquerda e a direita, são referentes ao primeiro e terceiro quartis. Já a linha que se estende a partir da caixa, indica a variabilidade dos dados fora dos quartis superior e inferior. Os dados discrepantes conhecidos como *outliers*, são aqueles que ficam localizados fora desses limites.

Figura 19. Exemplos de gráficos Boxplot.



Fonte: Os autores.

4.8 SOFTWARE ORANGE CANVAS

O *software* utilizado para simulação dos métodos de aprendizado de máquina foi o Orange Canvas, que possui código aberto e visualização dos dados através de fluxos. A escolha do Orange Canvas, foi baseada na sua característica de análise qualitativa, onde a interface gráfica permite um foco maior na análise exploratória dos dados, ao invés dos códigos de programação. Isso ocorre, uma vez que o mesmo já possui várias bibliotecas de código aberto do Python como *numpy*, *scipy* e *scikit-learn*, e assim, vários algoritmos de aprendizado de máquina disponíveis para testes e avaliação dos modelos.

Os algoritmos de aprendizado de máquina avaliados foram, as Redes Neurais Artificiais (RNA), as Árvore de Decisão, o Random Forest (RF), o *AdaBoost*, o KNN e o SVM.

5 CARACTERIZAÇÃO ESPECTRAL DE SINTER FEED

Neste Capítulo são apresentados os resultados relacionados com a caracterização espectral das amostras de Sinter Feed bruta e pulverizadas.

5.1 CARACTERIZAÇÃO FÍSICA

As características físicas das amostras (e.g. granulometria, umidade, cor, densidade) variam em função da frente de lavra de origem do material, devido ao fato de não serem homogêneas. Outros fatores também determinantes são a umidade natural, os minerais contaminantes agregados e o quão a rocha é friável, sendo que esse último atributo impacta diretamente sobre a granulometria final que o produto vai apresentar.

As amostras estudadas nesta pesquisa apresentam uma variação significativa tanto em granulometria, umidade quanto na cor. Essas características físicas estão intrinsicamente ligadas à litologia das amostras. Na **Erro! Fonte de referência não encontrada.**, foram selecionadas algumas amostras para exemplificar essas características físicas. Nota-se a mudança significativa na cor, variando desde amostras quase pretas (hematitas friáveis) até amostras avermelhadas de canga. Dessa forma, podemos notar a diversidade de características físicas, principalmente, na cor e granulometria.

A amostra identificada como CN_10563 (20) apresenta uma umidade natural elevada de 8,68%, um alto teor de ferro de 63,26%, uma coloração puxada para o vermelho e uma granulometria significativamente maior, mais grossa. Essa característica marcante da canga encontrada em Serra Sul, um minério úmido e de tom avermelhado, que foge do padrão convencional da hematita friável, que possui tonalidade mais escura e de textura mais seca. A canga é um minério não tão fácil de ser processado em usina de beneficiamento quanto a hematita, porém é tão rica quanto, em percentual de teor de ferro. Durante a preparação em laboratório, a amostra CN_10563 passou por um processo de pulverização e secagem.

Na amostra CN_010564 (**Erro! Fonte de referência não encontrada.**), a umidade natural é de 4,70% e o material se apresenta em uma granulometria menor, o que o torna um material mais friável, além de possuir uma tonalidade mais escura, característica marcante da hematita. Possui percentual de ferro de 64,15%, ou seja, um percentual pouco acima da amostra CN_10563, porém com características físicas muito diferentes. Na amostra preparada, existe

uma variação na coloração, sendo que ela torna-se mais avermelhada após passar pelos processos de secagem e pulverização.

Figura 20. Fotos de amostras de sínter feed de Serra Sul como produto final (à esquerda) e preparada em laboratório (à direita), com suas respectivas identificações.



Fonte: Os autores.

A amostra CN_010584 (**Erro! Fonte de referência não encontrada.**) foi a que apresentou o menor teor de ferro, com 41,41% e maior contaminação por sílica (SiO_2), representada por 39,45% (**Erro! Fonte de referência não encontrada.**). É uma mostra visualmente granulada e de coloração avermelhada, de baixo teor de ferro e alto teor de contaminantes para um produto final. Produtos com essas características passam por processo de blend no porto antes do embarque para o cliente final.

Já a amostra CN_010585 (**Erro! Fonte de referência não encontrada.**), dentre todas as analisadas, é a que apresentou através de resultados de laboratório, o maior percentual de

ferro referente a 68,24% (**Erro! Fonte de referência não encontrada.**). Possui coloração mais escura e baixíssima granulometria, características estas presentes na hematita friável. Assim como na preparação da amostra CN_10564, houve variação na coloração, após passar pelos processos de preparação em laboratório, tornando-se mais avermelhada.

A amostra CN_105560 (**Erro! Fonte de referência não encontrada.**) possui uma coloração intermediária e baixa granulometria. Foi observada a presença elevada de contaminantes como fósforo (P) com 0,519% e alumina (Al_2O_3) com 2,90%, além de 60,45% de teor de ferro (**Erro! Fonte de referência não encontrada.**).

A amostra CN_010550 (**Erro! Fonte de referência não encontrada.**) possui alto teor de umidade com 9,38%, granulometria intermediária, assim como coloração mais avermelhada. Dentre as amostras analisadas, é a que possui a maior presença da contaminante alumina (Al_2O_3) com 3,03% e um teor de ferro de 60,55% (**Erro! Fonte de referência não encontrada.**). Na preparação da amostra em laboratório, percebe-se variação significativa na coloração, tornando mais avermelhada.

As características físicas das amostras de sinter feed apresentam diferenças importantes relacionadas com a umidade, granulometria e cor. Nota-se que amostras mais granuladas e avermelhadas correspondem às oriundas da canga, e apresentam, principalmente, contaminantes relacionados com alumina (Al_2O_3) e sílica (SiO_2). As amostras mais ricas em ferro apresentam teor acima de 63% e a mais pobre apresentou 41,41% (**Erro! Fonte de referência não encontrada.**), o que fugiu muito da característica média das amostras avaliadas.

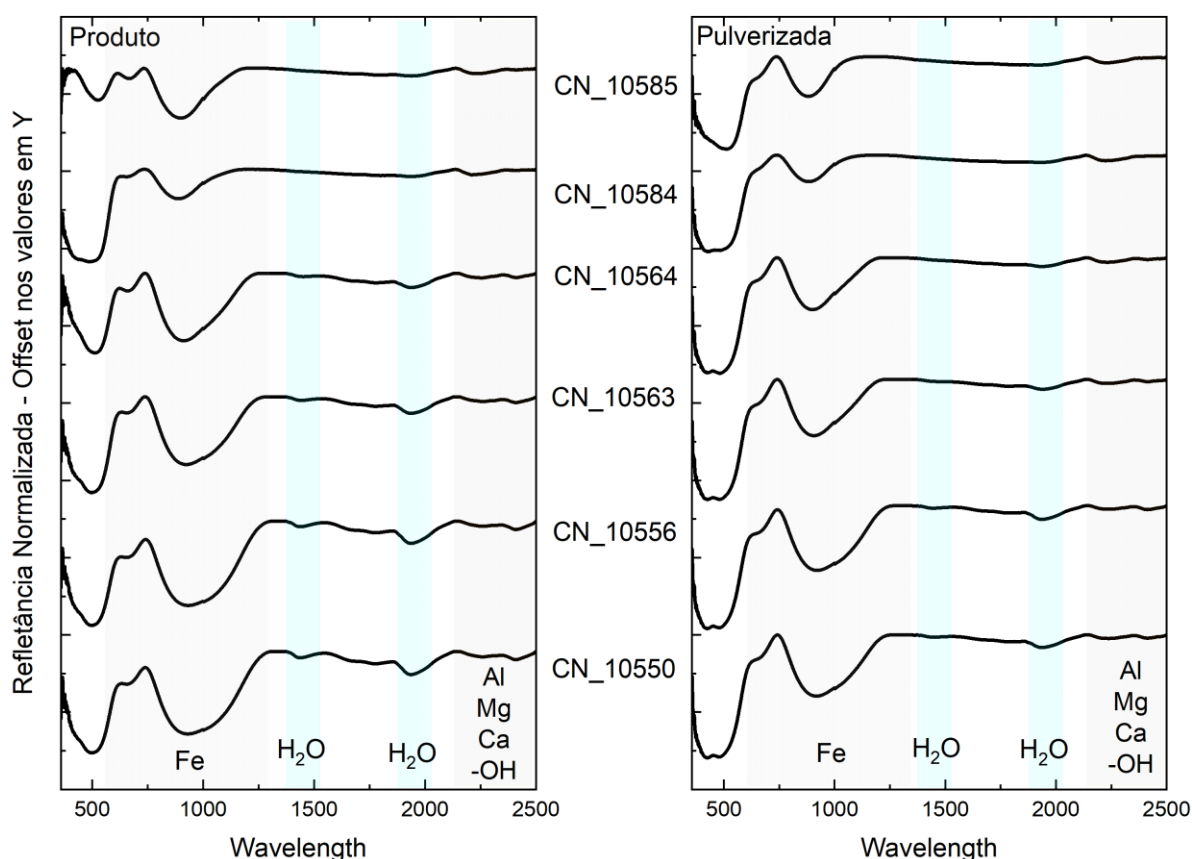
Amostras visualmente mais escuras e friáveis, ou seja, que apresentam menor granulometria como as CN_010564 e CN_010585, apresentam altos teores de ferro e mais baixos índices de contaminantes.

5.2 CARACTERIZAÇÃO ESPECTRAL

A análise quantitativa dos espectros derivados das amostras de sinter feed discutidos no item 4.1 (**Erro! Fonte de referência não encontrada.**), indica que essas amostras apresentam feições típicas de óxidos de ferro e hidroxilas, ao analisar amplamente o espectro da faixa de 350 a 2500 nm. Desta forma, o estudo nos mostra que as amostras coletadas apresentam representatividade em relação à frente de lavra de Serra Sul, de forma que nenhuma amostra

tenha saído do padrão normal de material que é expedido pela Usina de Beneficiamento, o que torna os dados significativamente aderentes.

Figura 21. Espectros de refletância normalizados através da remoção do contínuo de amostras de sinter feed – produto final e sua equivalente pulverizada.



Fonte: Os autores.

A umidade alta nas amostras produto CN_10563, CN_10564, CN_10556 e CN_10550 são refletidas nas curvas espectrais com a presença forte de água nas feições típicas próximos aos comprimentos de onda 1380 e 1950 nm (CLARK, 1999). Este comportamento não é observado nas amostras CN_10564 e CN_10585 que apresentam menor umidade (**Erro! Fonte de referência não encontrada.**). Adicionalmente, quando esses espectros do produto final são comparados com os seus equivalentes de amostra pulverizada, nota-se que as feições relacionadas com a umidade são quase totalmente atenuadas. Essa atenuação está relacionada com o processo de secagem, realizado na etapa de preparação amostral.

A amostra CN_10585 foi a que apresentou melhor resultado de laboratório, tanto em alto teor de ferro quanto no baixo percentual de contaminação (**Erro! Fonte de referência não encontrada.**), e é marcada por um espectro que se aproxima muito com o da hematita pura., marcada por feição de absorção principal em ~800 nm (HUNT, 1977). A curva espectral da amostra de produto CN_10585 apresenta o mesmo comportamento, a partir da faixa de 800 nm, relativo à amostra processada em laboratório (**Erro! Fonte de referência não encontrada.**) e, dessa forma, é possível observar que o minério oriundo dessa frente de lavra possui uma qualidade bastante satisfatória. A diferença entre os espectros é devida, principalmente, à umidade e à granulometria heterogênea na amostra do produto.

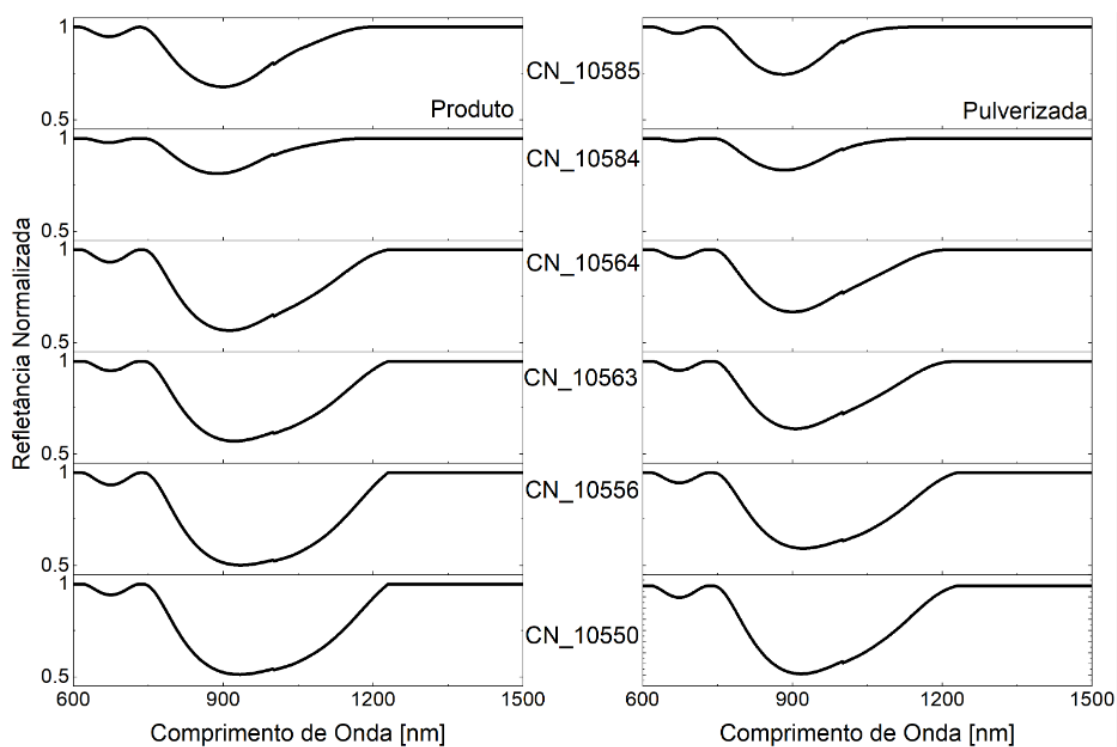
Ao observar os espectros das amostras de sínter feed preparadas em laboratórios, podemos notar que o tratamento realizado no laboratório, o comportamento espectral de todas as amostras se torna mais similar entre si, ao retirar a interferência da umidade nas curvas espectrais. Da mesma forma, o espectro da amostra de produto apresenta ruídos no início da faixa, em especial em frequências abaixo de 380 nm. Já a amostra preparada, a curva se mostrou mais uniforme (**Erro! Fonte de referência não encontrada.**).

Ao entrar no detalhe da faixa espectral no intervalo entre 600 a 1500 nm (**Erro! Fonte de referência não encontrada.**), é possível realizar algumas observações a respeito da característica de composição ferrosa das amostras.

As amostras CN_10584 e CN_10585 apresentam um comportamento espectral mais hematítico representado no espectro através de feições de absorção em 860 nm, enquanto as demais amostras um comportamento mais goethítico, com as feições deslocadas para a faixa de 910 nm. Todas as amostras apresentam pequenas feições também características do elemento Fe^{3+} na faixa de 670 nm, porém menos representativos nas amostras mais hematíticas e já bem significativas nas amostras mais goethíticas, como as CN_10563, CN_10556 e CN_10550.

De todas as amostras, a CN_10584 foi a que apresenta o menor percentual de ferro com 41,41% (**Erro! Fonte de referência não encontrada.**). Observando-se seu espectro ilustrado na **Erro! Fonte de referência não encontrada.**, é a amostra que apresenta feições mais discretas/menos profundas nos comprimentos de onda característicos da presença de minerais ferrosos em sua composição. Adicionalmente, ao se comparar os espectros das amostras produto nesse intervalo de 600 a 1500 nm e seu equivalente pulverizada, o comportamento nesta região é semelhante (**Erro! Fonte de referência não encontrada.2**).

Figura 22. Detalhe da faixa espectral onde ocorrem absorções relacionadas a óxidos e hidróxidos de ferro (600 a 1500 nm) das amostras produto e pulverizadas. Para comparação, a escala vertical é a mesma em todos os gráficos. Nota-se que as amostras CN_10585 e CN_10584 apresentam um comportamento mais hematítico, enquanto as demais amostras (CN_10564, CN_10593, CN_10556 e CN_10550) um comportamento mais goethítico. Adicionalmente, ao se comparar os espectros das amostras produto e sua equivalente pulverizada, o comportamento nesta região é semelhante.1



Fonte: Os autores.

As amostras CN_10564, CN_10563, CN_10556 e CN_10550 apresentam valores de contaminação com alumina (Al_2O_3) na faixa de 1,47% a 3,03% (Tabela 2), sendo a CN_10550 com o maior percentual (**Erro! Fonte de referência não encontrada.**). Todas as amostras

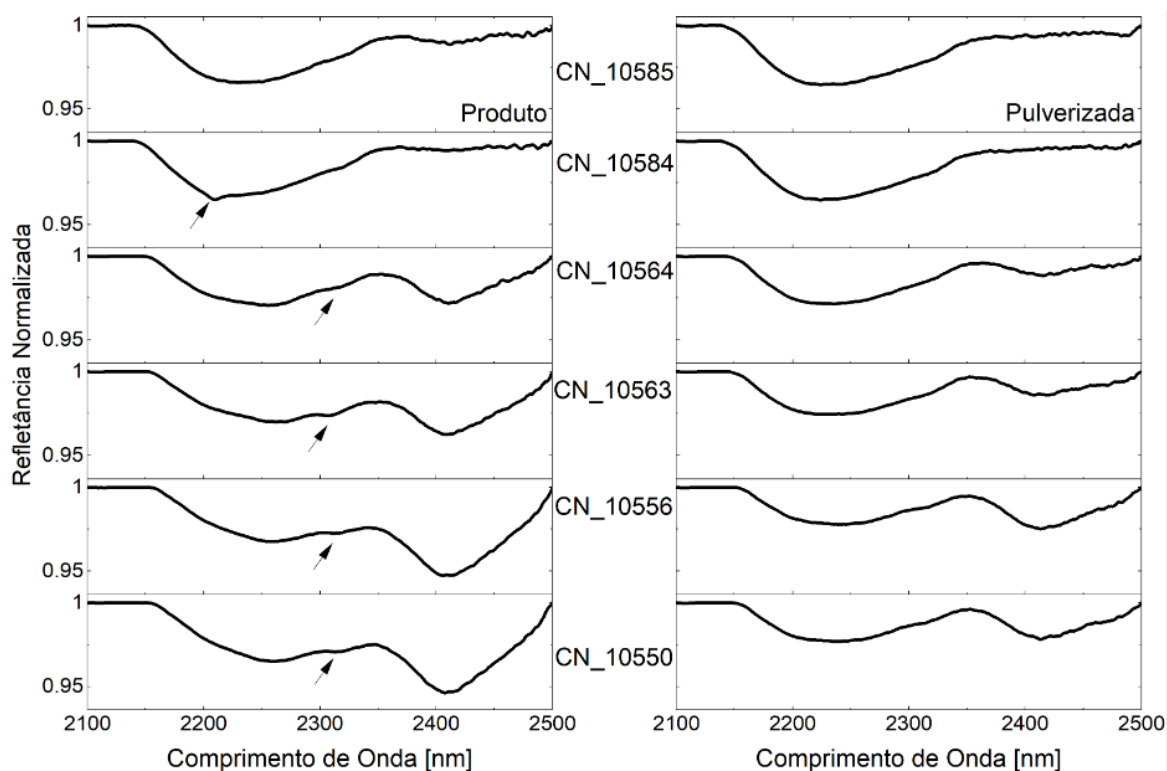
apresentam em comum, feição na faixa de 2310 nm, conforme observado na **Erro! Fonte de referência não encontrada.**

5.3 ANÁLISE DE COMPONENTES PRINCIPAIS (PCA) – AMOSTRAS BRUTAS DE SÍNTER FEED

5.3.1 Análise espectral na faixa de 380 a 2500 nm

Na primeira análise, foram considerados os espectros de refletância normalizado, através da remoção do contínuo na faixa de 380 a 2500 nm, de amostras brutas de sinter feed. O intervalo entre 350 e 380 nm foi expurgado, devido ao alto nível de ruído. Através do método PCA, com 3 componentes, é encontrada uma variância acumulada de 0,83 que mostra um valor bastante representativo, visto que quanto maior a variância, maior a diferença entre os pontos que estão sendo projetados.

Figura 23. Detalhe da faixa espectral de contaminantes dos minérios de ferro (2100 a 2500 nm) das amostras produto e pulverizadas. Para comparação, a escala vertical é a mesma em todos os gráficos. Ao se comparar os espectros das amostras de produto e sua equivalente pulverizada, nota-se que as feições relacionadas com os contaminantes foram suavizadas nas amostras pulverizadas.

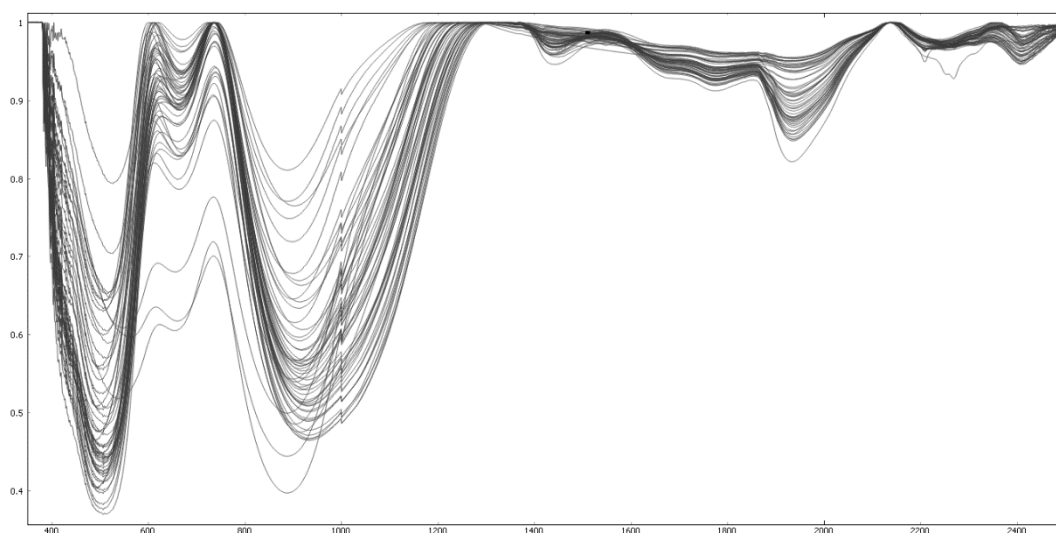


Fonte: Os autores.

O gráfico da **Erro! Fonte de referência não encontrada.** exibe as curvas espectrais das 58 amostras de produto na faixa de 380 a 2500 nm obtidas através do método PCA, é possível observar a representatividade dos dados ao considerar as 3 componentes principais.

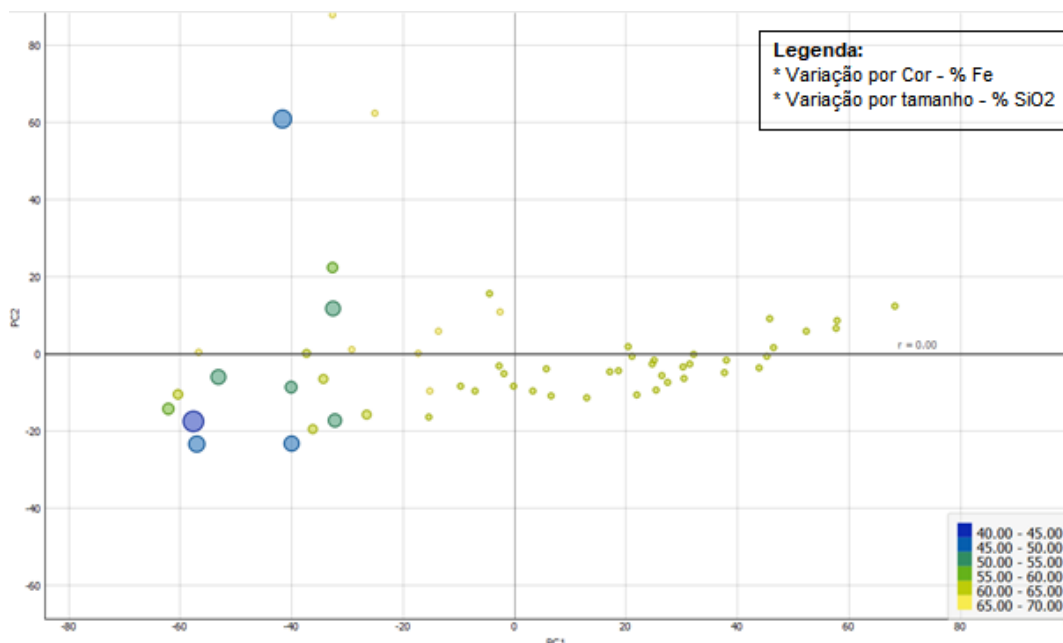
Através da análise dos gráficos de dispersão dos autovetores das amostras de sínter feed de Produto, associadas as duas primeiras principais componentes, é possível ver que quanto maior o teor de Fe da amostra, menor o percentual de SiO₂, conforme ilustrado na **Erro! Fonte de referência não encontrada.**

Figura 24. Análise dos dados pelo método PCA na faixa espectral de 380 a 2500 nm, gráfico normalizado através da remoção do contínuo das curvas espectrais das 58 amostras de Produto gerados com 3 componentes principais no intervalo de 380 a 2500 nm.



Fonte: Os autores.

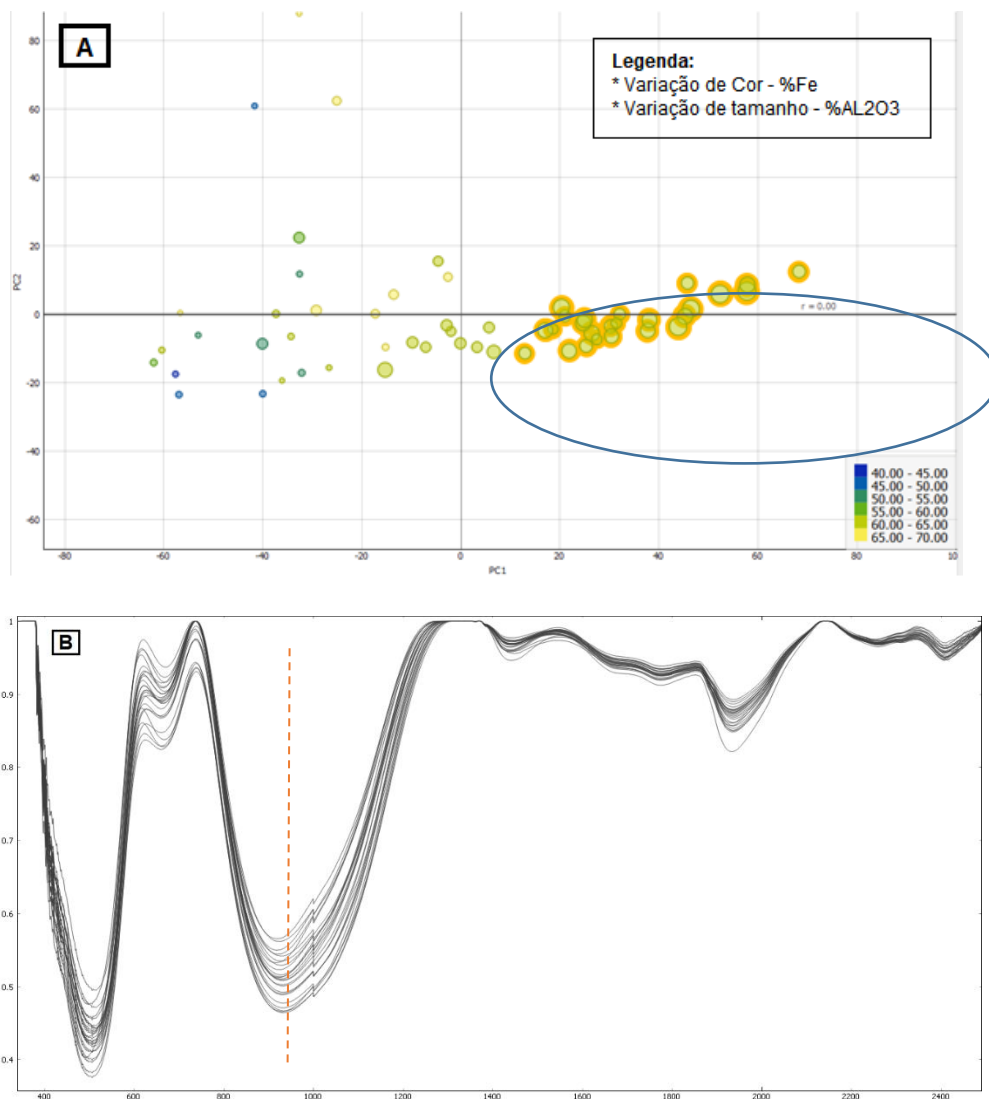
Figura 25. Gráfico de dispersão dos autovetores da PC1 vs PC2 referentes aos espectros de refletância normalizados através da remoção do contínuo das amostras produto de sínter feed no intervalo de 380 a 2500 nm. As cores estão relacionadas com o teor de FeO e o tamanho dos círculos com o teor de SiO₂. Nota-se que quanto maior a percentagem de FeO, menor teor de SiO₂



Fonte: Os autores.

As amostras que apresentam os maiores teores de Al_2O_3 , também possuem grande percentual de Fe, porém não é possível correlacionar os dados diretamente, visto que os maiores teores de ferro apresentados (**Erro! Fonte de referência não encontrada.**A), não são referentes as amostras que apresentam os maiores percentuais de Al_2O_3 . Porém, é possível notar uma relação entre o alto teor de Al_2O_3 , com amostras com características mais goethíticas, com feições de absorção centradas na faixa espectral acima de 910 nm, como destacado na **Erro! Fonte de referência não encontrada.**26(B).

Figura 26. (A) Gráfico de dispersão de autovetores da PC1 vs PC2 referente aos espectros de refletância normalizados através da remoção do contínuo das amostras produto de sinter feed no intervalo de 380 à 2500 nm. As cores estão relacionadas com o teor de FeO e o tamanho dos círculos com o teor de Al_2O_3 . (B) Espectro em destaque para as amostras goethíticas, com feições marcantes acima da faixa de 900 nm.

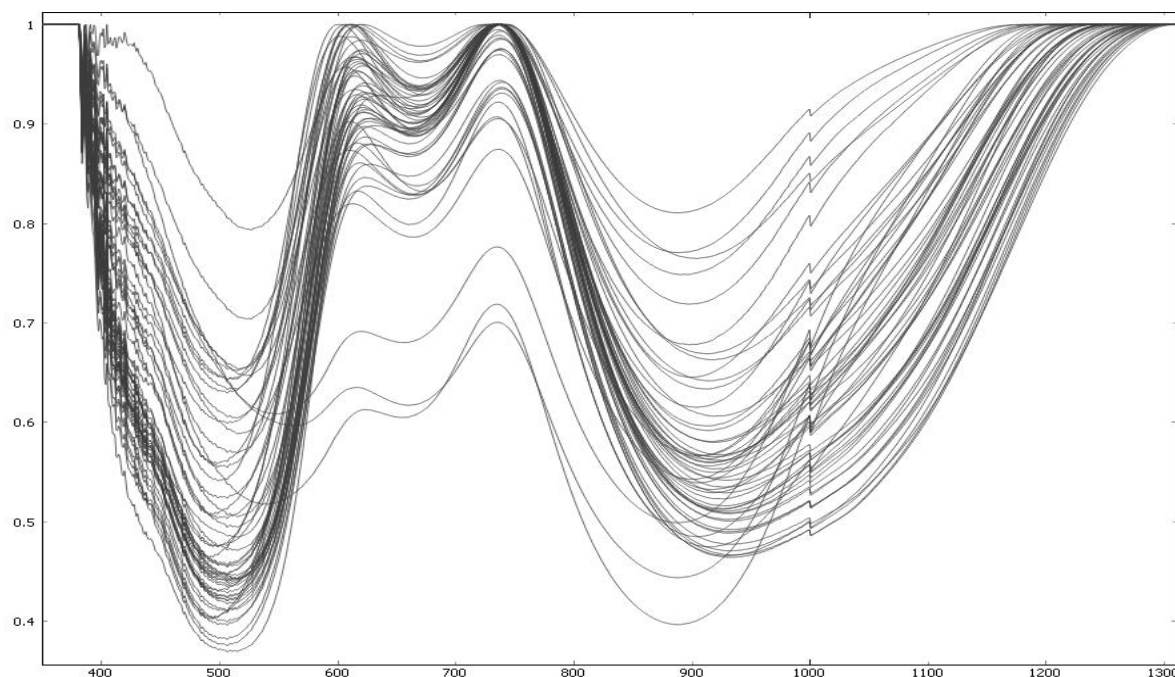


Fonte: Os autores.

5.3.2 Análise espectral na faixa de 380 a 1310 nm - Ferro

Na segunda análise, a faixa espectral considerada foi de 380 a 1310 nm, região onde as feições características do Fe são mais representativas. Nesse caso, a análise com 3 componentes principais já mostra uma variância acumulada de 0,921. O gráfico da **Erro! Fonte de referência não encontrada.** exibe as curvas espectrais obtidas através do método PCA.

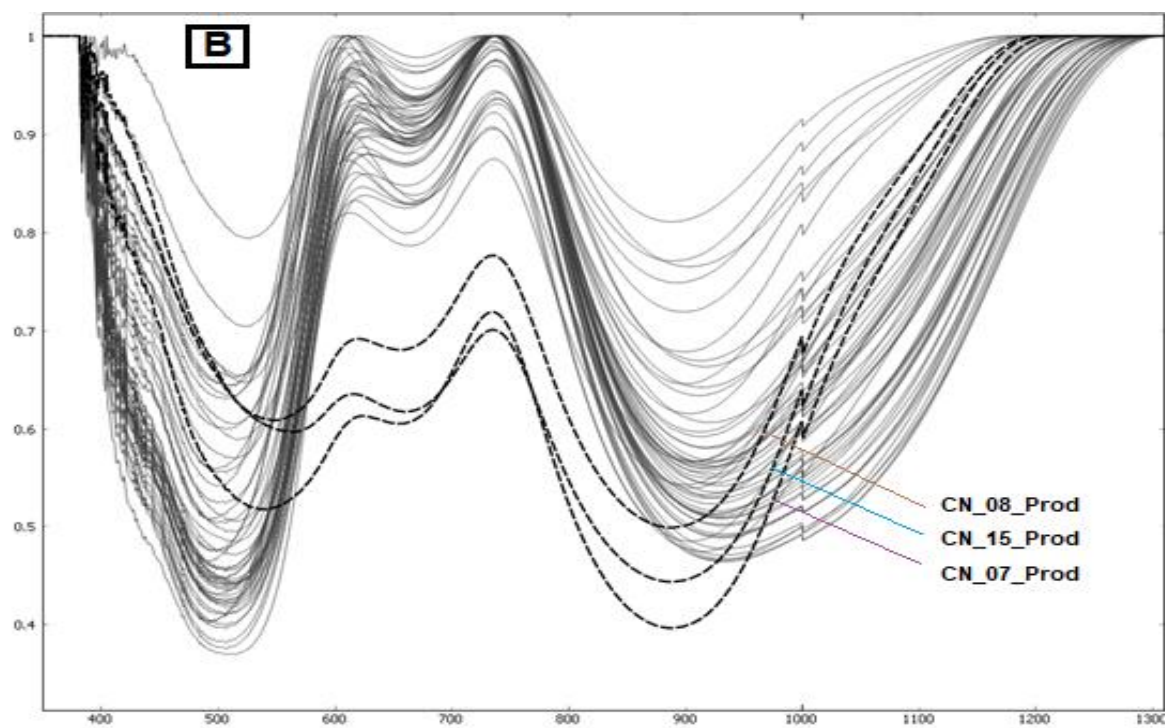
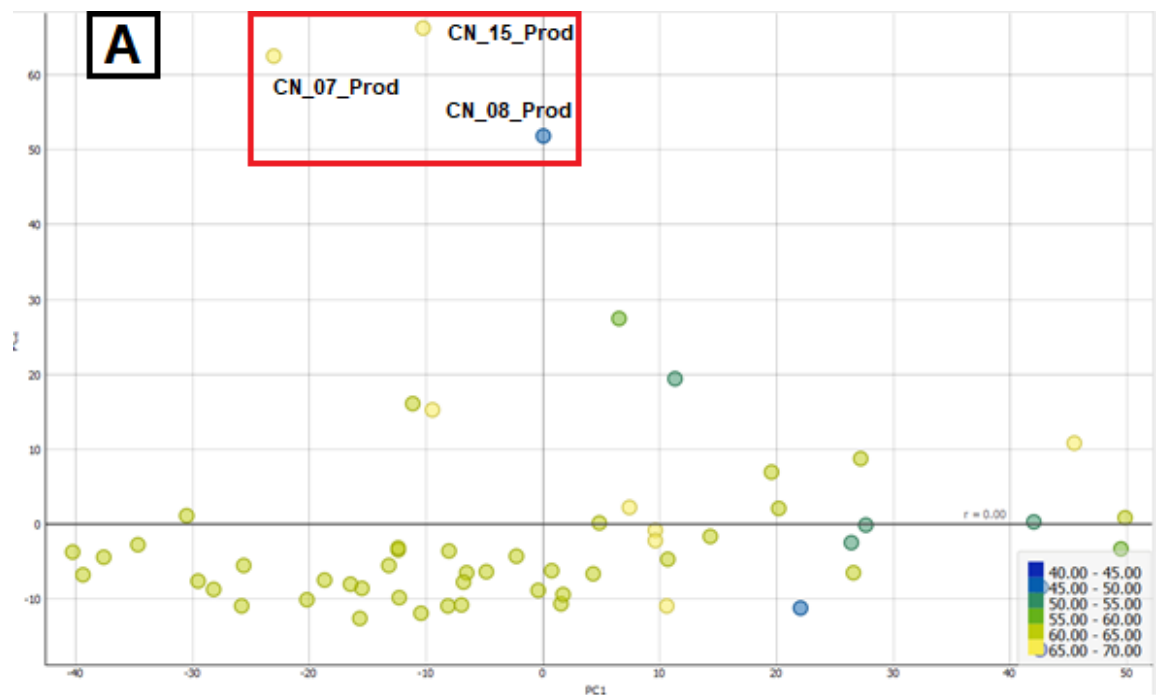
Figura 27. Análise dos dados pelo método PCA na faixa espectral de 380 a 1310 nm, gráfico normalizado através da remoção do contínuo das curvas espectrais das 58 amostras de Produto gerados com 3 componentes principais no intervalo de 380 a 1310 nm.



Fonte: Os autores.

A análise dos autovetores das PCs 1 e 2, associados com as curvas espectrais de refletância normalizadas das 58 amostras de sinter feed brutas, revela que há um maior desvio de dispersão correspondentes às amostras com características mais hematíticas (28A), ou seja, que possuem feições de absorção abaixo da faixa de 900 nm (referência de minério de ferro), conforme ilustrado na **Erro! Fonte de referência não encontrada.28 (B)**.

Figura 28. (A) Gráfico de dispersão de autovetores da PC1 vs PC2 referente aos espectros de refletância normalizados através da remoção do contínuo, das amostras produto de sinter feed no intervalo de 380 à 1310 nm, com foco no percentual de teor de Fe; (B) Espectros de refletância normalizados, com indicação das curvas de refletância de 3 amostras (linhas pontilhadas) que apresentaram maior dispersão em (A) com relação às demais amostras, e caracterizadas por feições de absorção abaixo da faixa de 900 nm (comportamento hematítico).



Fonte: Os autores.

5.3.3 Análise espectral na faixa de 2135 a 2500 nm

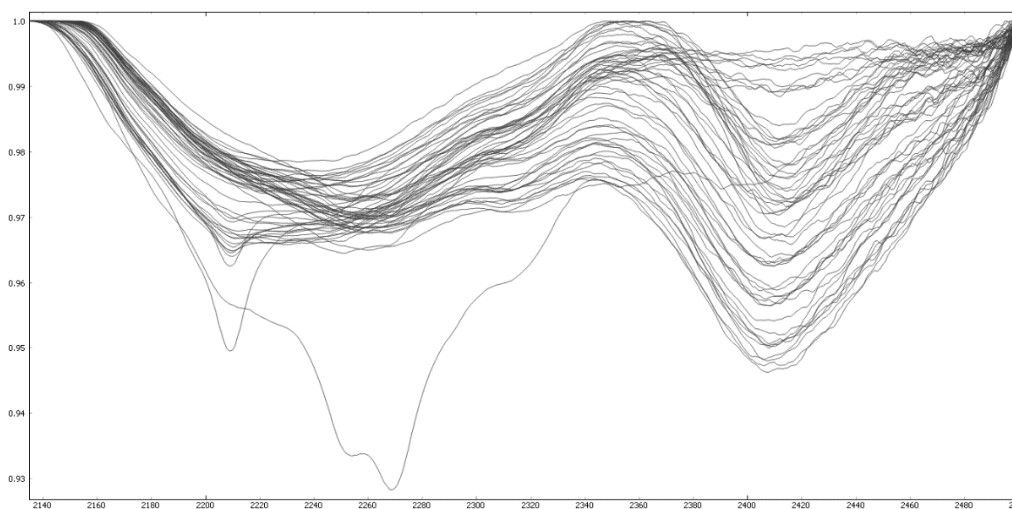
Na terceira análise, a faixa considerada foi de 2135 a 2500 nm, que é a região onde as feições de absorção características dos principais contaminantes, como Al-OH, Mn, MgO, P,

TiO₂, CaO e K₂O. Nessa faixa espectral, é possível observar que com 2 componentes temos uma variância acumulada de 0,924.

Essa variância está relacionada com a alta correlação dos dados, devido ao intervalo espectral analisado ser menor e, conseqüentemente, haver maior semelhança entre os dados. Dessa forma, nas primeiras componentes já é possível armazenar a maior parte da variância das informações. Porém, as componentes de ordem mais elevadas podem fornecer informações relacionadas com o que é menos semelhante nos gráficos, havendo, assim uma importância da análise de tais PCs na busca de agrupamentos relacionados aos contaminantes.

A **Erro! Fonte de referência não encontrada.** exibe as curvas espectrais normalizadas das amostras de sínter feed de Produto, considerando-se apenas o intervalo de 2135 a 2500 nm. Estes dados foram usados como entrada ao processamento via PCA.

Figura 29. Análise dos dados pelo método PCA na faixa espectral de 2135 a 2500 nm, gráfico normalizado através da remoção do contínuo das curvas espectrais das 58 amostras de Produto gerados com 2 componentes principais no intervalo de 2135 a 2500 nm

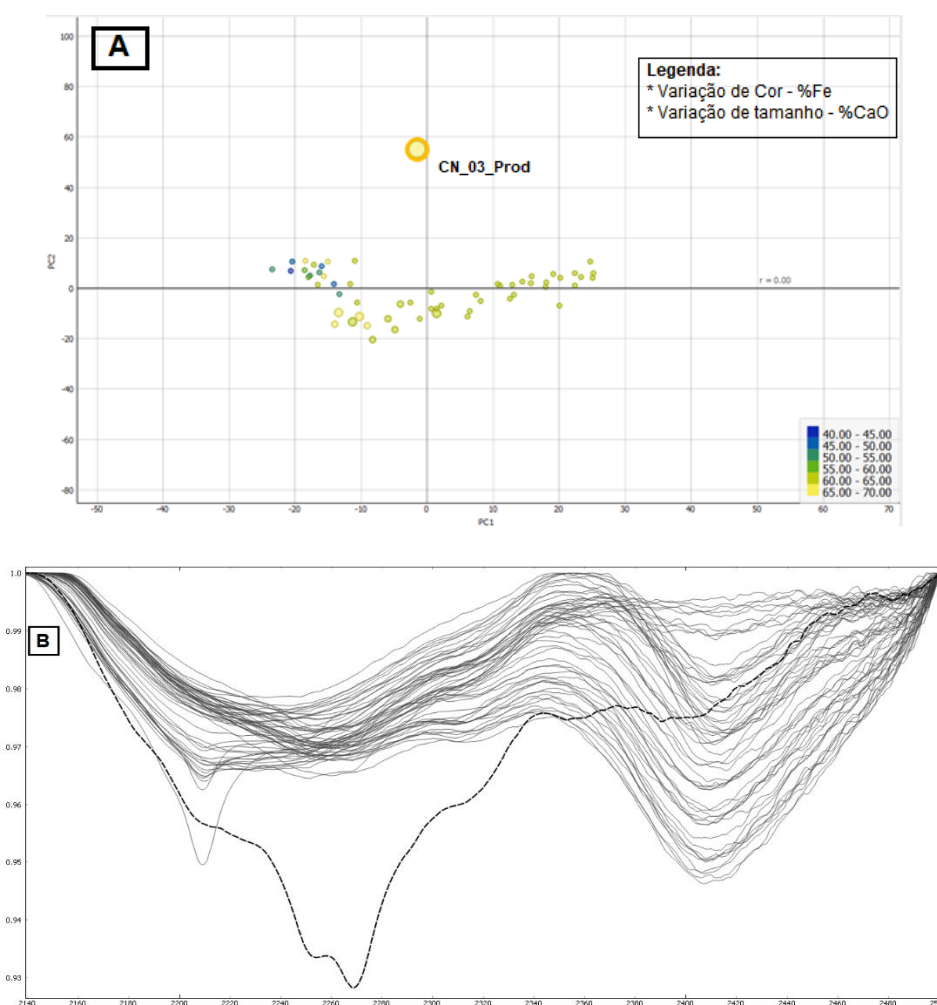


Fonte: Os autores.

No gráfico de dispersão dos autovetores das PCs 1 e 2 relacionados às curvas espectrais normalizadas das 58 amostras de sínter feed brutas (30A), é possível verificar uma maior variabilidade referente à amostra caracterizada pela presença do contaminante CaO em maior concentração e alto de Al₂O₃ (1,32%) (**Erro! Fonte de referência não encontrada.**), no caso, a amostra CN_03. Analisando-se a resposta espectral na faixa de 2135 a 2500 nm, essa

separabilidade vista nos autovetores é corroborada com o comportamento espectral da amostra CN_03 que foge completamente das características das demais curvas (**Erro! Fonte de referência não encontrada.**30B). Aparecem feições de absorção centradas em cerca de 2210, 2250, 2270 e ~2315 nm, derivadas de moléculas Al-OH presentes na estrutura da gibbsita [Al(OH)₃] (CLARK et al., 1990).

Figura 30. (A) Gráfico de dispersão da PC1 vs PC2 dos espectros de refletância normalizados através da remoção do contínuo, das amostras produto de sinter feed no intervalo de 2135 à 2500, com foco na relação entre percentuais de teor de Fe e CaO; (B) Espectro com destaque para a amostra CN_03_Prod na faixa de 2135 a 2500 nm.

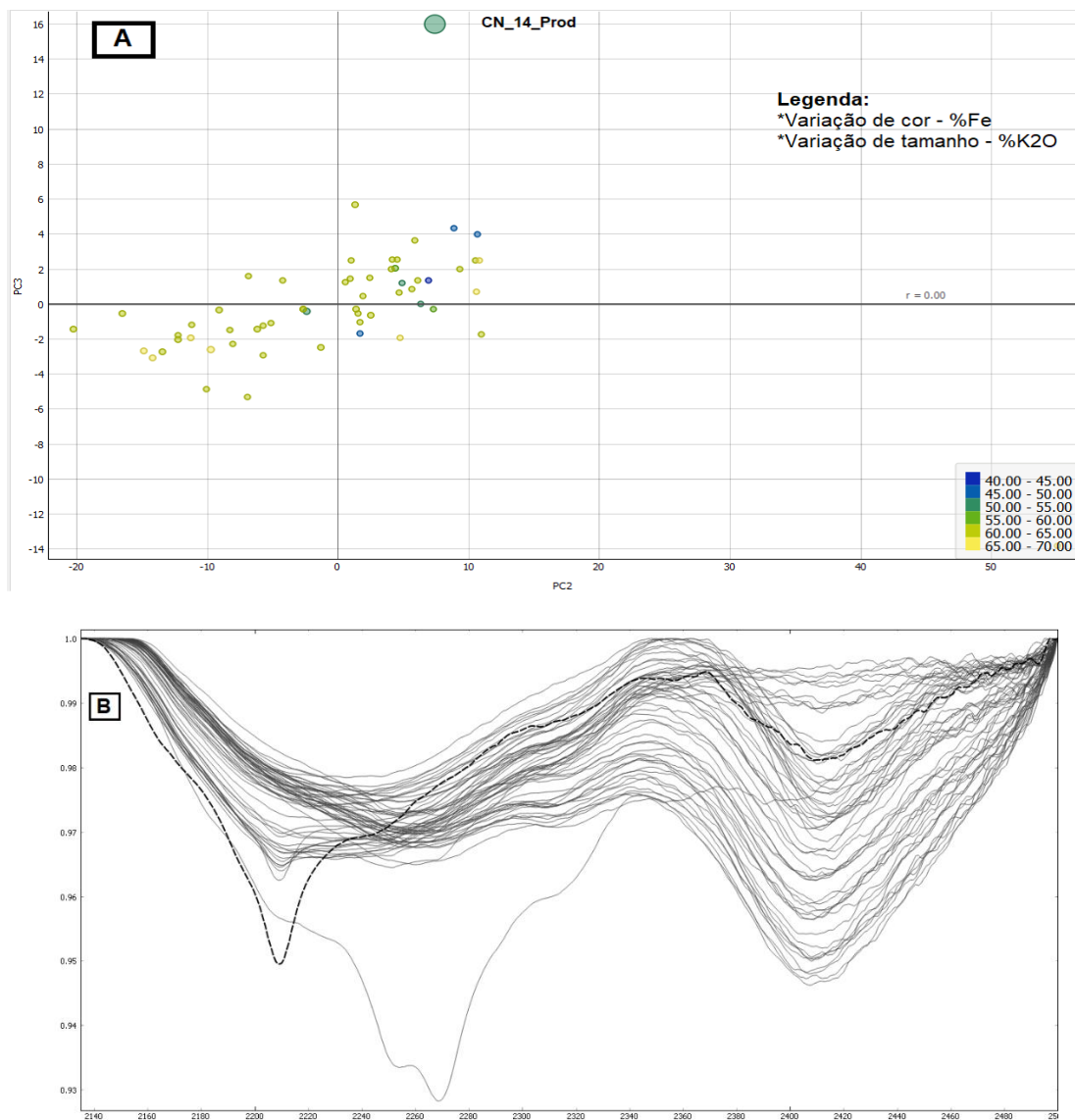


Fonte: Os autores.

Na análise do gráfico de dispersão dos autovetores das componentes PC2 e PC3, referentes às curvas espectrais na faixa de 2135 a 2500 nm, foi possível notar um comportamento que sai das características das demais amostras (31A). No gráfico das curvas espectrais apresentado na **Erro! Fonte de referência não encontrada.** (B), é possível

identificar o espectro da amostra CN_14_Prod e relacionar com a presença do contaminante K_2O em maior concentração, presente em 0,235% (**Erro! Fonte de referência não encontrada.**).

Figura 31. - (A) Gráfico de dispersão de autovetores da PC2 vs PC3 referentes aos espectros de refletância normalizados através da remoção do contínuo, das amostras produto de sinter feed no intervalo de 2135 a 2500 nm, com foco na relação entre percentuais de teor de Fe e K_2O ; (B) Espectros de refletância das amostras de sinter feed, com destaque para a amostra CN_14_Prod (em linha pontilhada).



Fonte: Os autores.

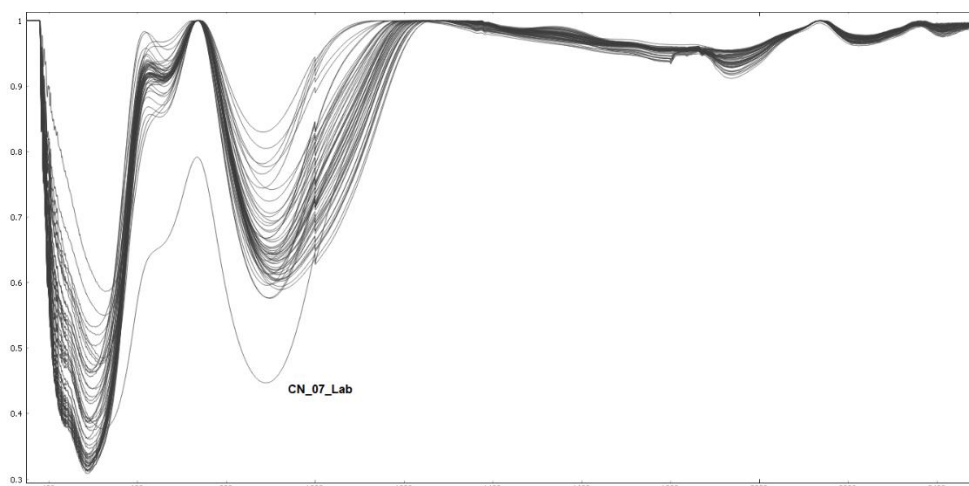
5.4 ANÁLISE DE COMPONENTES PRINCIPAIS (PCA) – AMOSTRAS DE SÍNER FEED PREPARADAS EM LABORATÓRIO

5.4.1 Análise espectral na faixa de 380 a 2500 nm

Na primeira análise, foram considerados os dados da biblioteca espectral de amostras de sínter feed pulverizadas e secadas em laboratório, na faixa de 380 a 2500 nm, ou seja, toda a faixa do espectro. Através do método PCA, é possível observar que com 4 componentes, é encontrada uma variância acumulada de 0,90.

A **Erro! Fonte de referência não encontrada.**³² exibe as curvas espectrais das amostras pulverizadas de sínter feed na faixa de 380 a 2500 nm, usadas para aplicação do método PCA.

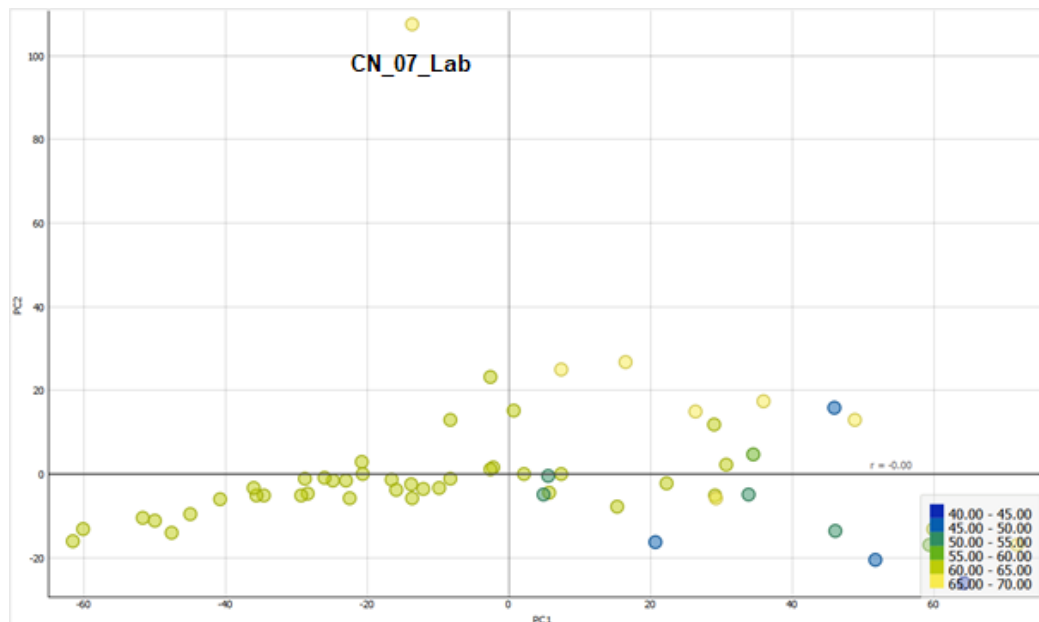
Figura 32. Análise dos dados pelo método PCA na faixa espectral de 380 a 2500 nm, gráfico normalizado através da remoção do contínuo das curvas espectrais das 58 amostras preparadas em laboratório, gerados com 4 componentes principais no intervalo de 380 a 2500 nm.



Fonte: Os autores.

Na **Erro! Fonte de referência não encontrada.**, é possível observar uma curva espectral que foge da característica das demais. A amostra CN_07_Lab possui alto teor de Fe, com 67,79% (**Erro! Fonte de referência não encontrada.**) e nenhuma correlação com outro contaminante foi encontrada de forma significativa. De acordo com o gráfico de dispersão dos autovetores das PCs 1 e 2 (**Erro! Fonte de referência não encontrada.**), observa-se que a amostra CN_07_Lab é mostrada como um outlier, ou seja, dado que destoa drasticamente dos demais. De forma equivalente à interpretação qualitativa das curvas de refletância (**Erro! Fonte de referência não encontrada.**), em que se observa maior profundidade de absorção centrada na feição de ferro férrico em cerca de aproximadamente 900 nm.

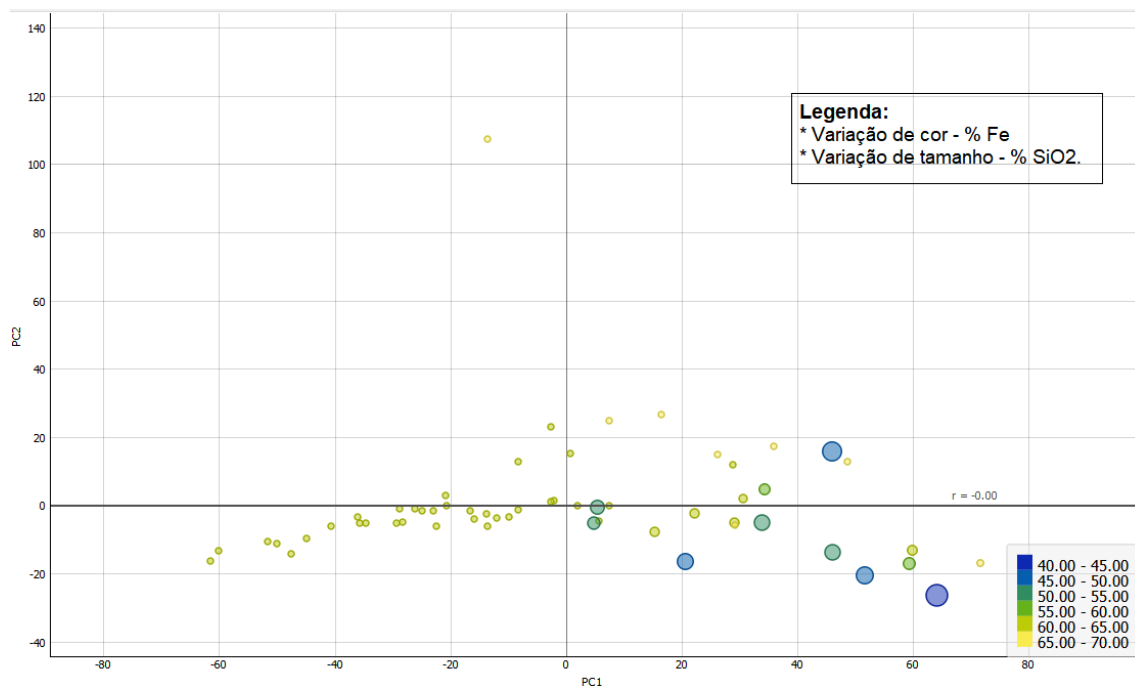
Figura 33. Gráfico de dispersão da PC1 vs PC2 dos espectros de refletância normalizados através da remoção do contínuo, das amostras de sinter feed preparadas em laboratório, no intervalo de 380 à 2500 nm, com realce de cor relativo ao percentual de Fe.



Fonte: Os autores.

Assim como na análise das amostras de sinter feed de produto, nas amostras preparadas em laboratório, através dos gráficos de dispersão de autovetores das principais componentes (**Erro! Fonte de referência não encontrada.**), é possível afirmar que quanto maior o teor de Fe da amostra, menor o percentual de SiO₂.

Figura 34. Gráfico de dispersão de autovetores da PC1 vs PC2 referente aos espectros de refletância normalizados através da remoção do contínuo das amostras sinter feed preparadas em laboratório, no intervalo de 380 a 2500 nm. As cores estão elacionadas com o teor.

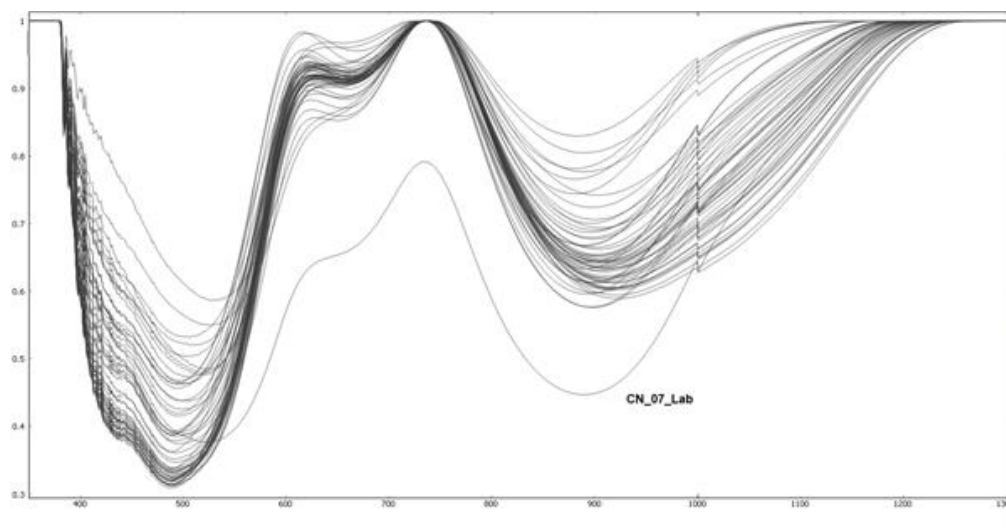


Fonte: Os autores.

5.4.2 Análise espectral na faixa de 380 a 1310 nm

A segunda análise, a faixa espectral considerada foi de 380 a 1310 nm. Nesse caso, a análise com 2 componentes já mostra uma variância acumulada de 0,91. A **Erro! Fonte de referência não encontrada.** exibe as curvas espectrais das 58 amostras de sinter feed pulverizadas usadas como dado de entrada ao processamento via PCA.

Figura 35. Curvas espectrais de refletância normalizadas de 58 amostras de sinter feed pulverizadas, considerando-se a faixa de 380 a 1310 nm. Estes dados foram selecionados para processamento via o método PCA.



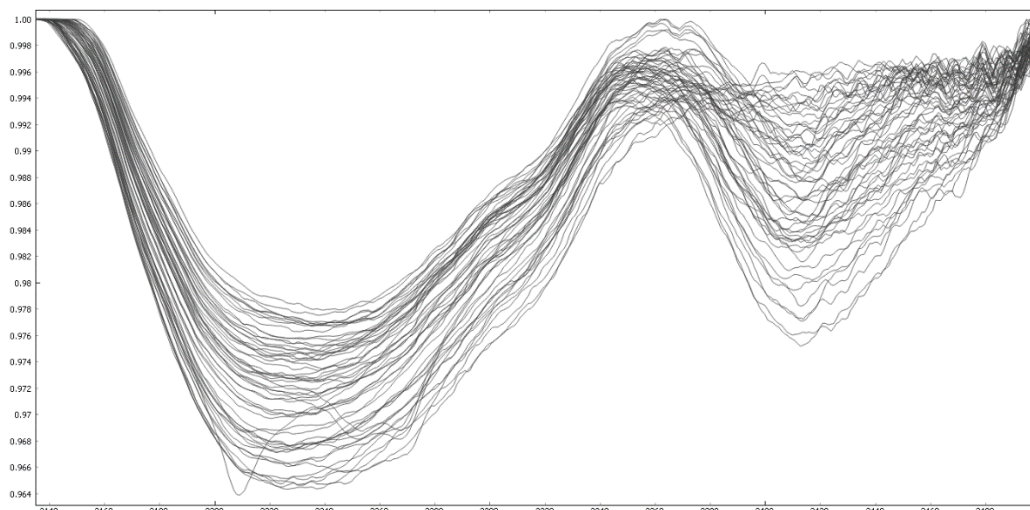
Fonte: Os autores.

Nesse intervalo espectral de 380 a 1310 nm em amostras pulverizadas e secadas em laboratório, é possível observar um comportamento espectral similar em todas as amostras e, assim como salientado na análise do item 4.3.2.1, apenas uma amostra apresentou um comportamento que foge da característica das demais. A amostra CN_07_Lab possui alto teor de Fe, com 67,79% (**Erro! Fonte de referência não encontrada.**), porém nenhuma correlação com outro contaminante foi encontrada de forma significativa. Portanto, a confirmação que esta amostra poderia ser considerada um *outlier*.

5.4.3 Análise espectral na faixa de 2135 a 2500 nm

Na terceira análise, a faixa considerada foi de 2135 a 2500 nm. Através do método PCA, com 2 componentes, temos uma variância acumulada de 0,92. A **Erro! Fonte de referência não encontrada.** apresenta as curvas espectrais normalizadas das amostras e sinter feed pulverizadas, que foram empregadas para aplicar a PCA.

Figura 36. Curvas espectrais de refletância normalizadas de 58 amostras de sinter feed pulverizadas, considerando-se a faixa de 2135 a 2500 nm. Estes dados foram selecionados para processamento via o método PCA.

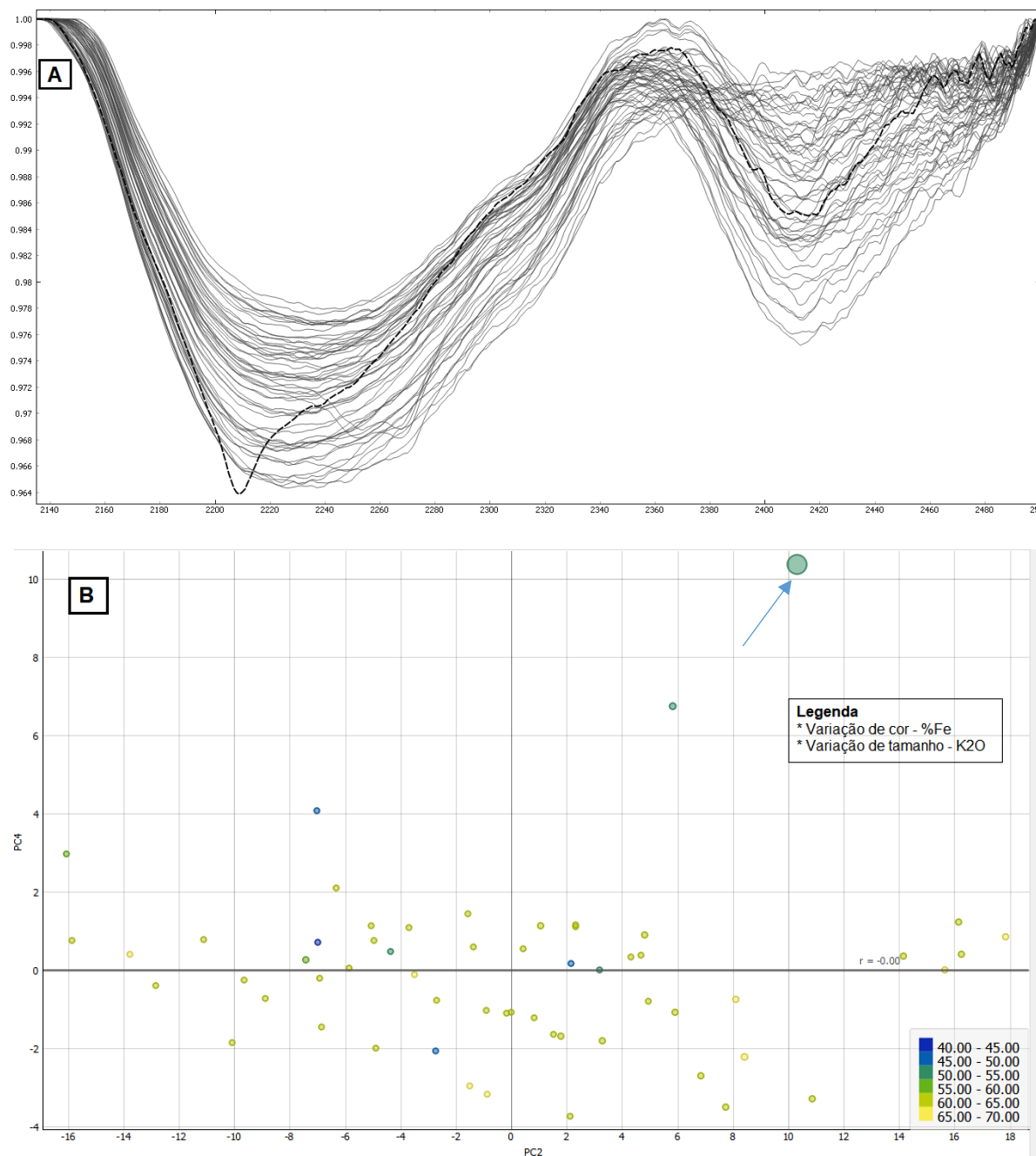


Fonte: Os autores.

Considerando a faixa espectral de 2135 a 2500 nm, foi possível notar um comportamento que foge completamente das características das demais curvas e destacado na **Erro! Fonte de referência não encontrada. (A)**. Ao analisar os autovetores das componentes principais PC2 e PC4 (**Erro! Fonte de referência não encontrada.B**) é possível verificar que a amostra de sínter feed com maior dispersão (CN_14_Lab) possui o contaminante K_2O em maior concentração, correspondente a 0.235%. Conforme podemos atestar através dos dados da **Erro! Fonte de referência não encontrada.**, esta é a amostra que possui o valor mais representativo de K_2O se comparadas às demais.

Apesar da amostra CN_14_Lab visivelmente apresentar maior teor do contaminante K_2O , não é possível correlacionar este contaminante com o teor de Fe, visto que amostras com menores teores de K, apresentaram teores variados de Fe (**Erro! Fonte de referência não encontrada.**).

Figura 37. (A) Espectro com a amostra CN_14_Lab em destaque para (linha pontilhada) a faixa de 2135 a 2500 nm; (B) Gráfico de dispersão da PC2 vs PC4 dos espectros de refletância normalizados através da remoção do contínuo das amostras sinter feed preparadas em laboratório, no intervalo de 2135 a 2500 nm. As cores estão relacionadas com o teor de FeO e o tamanho dos círculos com o teor de K₂O.



Fonte: Os autores.

5.5 COMPARAÇÃO ENTRE AS ANÁLISES ESPECTRAIS EM AMOSTRAS DE SÍNTER FEED DE PRODUTO E PREPARADAS EM LABORATÓRIO

Ao realizar um comparativo em relação aos resultados encontrados nas análises pelo método PCA das amostras brutas e de preparadas em laboratório, a correlação em relação à presença da SiO_2 e percentual de Fe é presente e perceptível de forma bem similar em ambas as análises.

Na faixa espectral de 380 a 1310 nm, onde as feições de Fe são mais representativas, as amostras preparadas em laboratório, que passam por processos de secagem e pulverização, apresentam curvas mais homogêneas em relação as amostras brutas de sinter feed. Desta forma, as amostras CN_07_Prod, CN_08_Prod e CN_15_Prod, com espectros com feições mais marcantes na faixa abaixo de 900 nm e, portanto, mais hematíticas (**Erro! Fonte de referência não encontrada. A e B**), tiveram as feições de absorção mais atenuadas. A amostra CN_07_Lab apresentou o espectro em destaque, com feição de absorção <900 nm mais profunda (**Erro! Fonte de referência não encontrada. 35**).

Em relação ao contaminante CaO, associado a alto valor de Al_2O_3 (1,32%), este foi associado à ocorrência conjunta de gibbsita $[\text{Al}(\text{OH})_3]$ observada por feições de absorção bem desenvolvidas e centradas na região de 2200-2270 nm e por separabilidade estatística multivariada a partir da aplicação da PCA (**Erro! Fonte de referência não encontrada. 30A e B**) das amostras de produto de sinter feed. Todavia, nas amostras pulverizadas, tal padrão não foi possível encontrar.

Já para o contaminante K_2O , ambas as análises espectrais e por PCA das amostras de produto e amostras preparadas em laboratório apresentaram resultado e conclusões similares; porém, em ambas as análises por PCA, somente foi possível observar a separabilidade amostral a partir da análise de autovetores de componentes principais de maior ordem (**Erro! Fonte de referência não encontrada. 31A e Erro! Fonte de referência não encontrada. 37B**).

6 ESTIMATIVA DA QUALIDADE DE SINTER FEED ATRAVÉS DE ESPECTROSCOPIA DE REFLETÂNCIA E APRENDIZADO DE MÁQUINA

6.1 FORMAÇÃO DA BASE DE DADOS

A base de dados gerados para este estudo, foi criada através das curvas do espectro de refletância coletadas das amostras, que correspondem as assinaturas espectrais de interesses, a serem inseridas no *software* de aprendizado de máquinas. Dessa forma, foi criada uma biblioteca espectral característica das amostras, através do processamento dos 100 dados de entrada referentes a cada amostra. Em seguida, foi realizada uma média das 25 leituras, no que resultou em uma base de dados com 4 espectros, alusivas a cada amostra.

A base de dados foi criada, já considerando o espectro com o contínuo removido e transformada em um arquivo do tipo ASCII. Esse formato, possibilita a leitura das informações por programas que utilizam métodos de aprendizado de máquina em seu algoritmo e tem por objetivo otimizar o processo de análise dos dados espectrais.

A mesma base de dados foi criada tanto para análise de amostras de produto, quanto para as mesmas preparadas em laboratório. O objetivo é que fosse avaliado o quanto a preparação da amostra pode ser relevante para os resultados, ou até mesmo se a técnica proposta poderá ser utilizada no sinter feed sem nenhum tipo de preparação, o que garante ganhos maiores, tanto em tempo de análise, quanto na não necessidade de preparação da amostra.

6.2 SELEÇÃO DAS VARIÁVEIS

A principal ideia do trabalho, é estimar os teores de ferro para cada conjunto de amostras, ou seja, tanto para as de produto (sinter feed), quanto para as preparadas em laboratório, a partir da relação de regressão encontrada pelos métodos de aprendizado de máquina entre o espectro e os percentuais de teor de ferro.

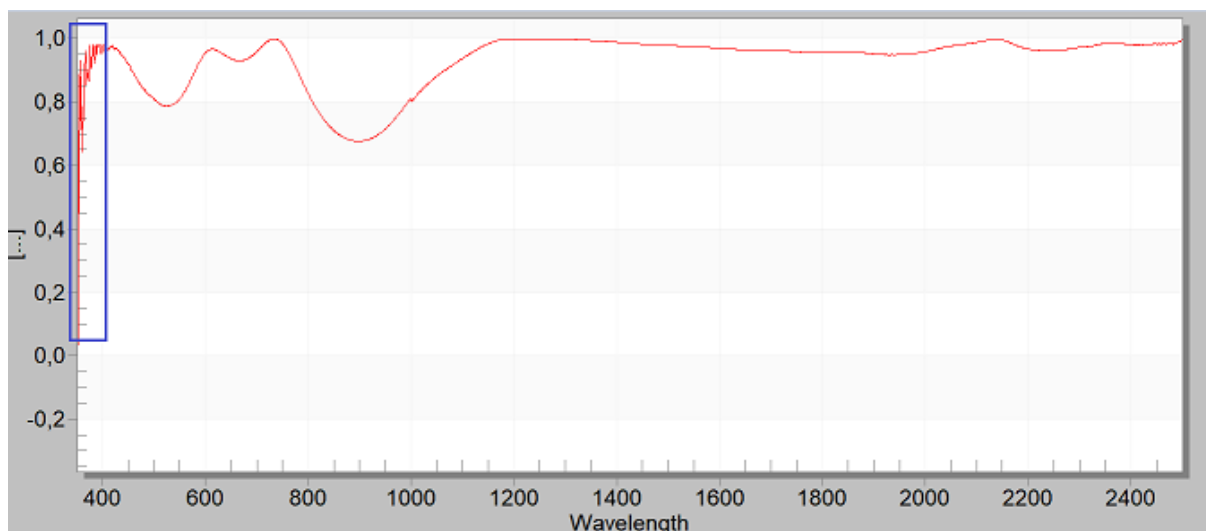
O objetivo da seleção das variáveis, é criar um modelo de forma mais representativa possível a respeito da frente de lavra de minério de ferro objeto desse estudo. Para criação dos dados de treino, foram utilizados para o aprendizado e validação do modelo, os valores de teor de ferro encontrados através de análises de laboratório, referenciados pela Tabela 2.

Do total dos dados, aproximadamente 70% foram destinados para o treinamento do modelo e 30% destinados para a validação dos métodos avaliados. As mesmas informações referentes as amostras utilizadas para treinamento e validação dos métodos, foram aplicadas em todos os estudos analisados, ou seja, os mesmos espectros considerados para as amostras de produtos, foram também adotados para as análises nas amostras preparadas em laboratório, tanto na faixa de análise total do espectro de 400 a 2500 nm, quando na faixa de 400 a 1310 nm, considerado a faixa do espectro com feições mais marcantes do ferro.

Na base de dados de 232 elementos, 156 leituras foram consideradas para dados de treinamento do modelo e 76 para validação dos métodos de aprendizado de máquina.

Como a faixa inicial do espectro apresentam muitos ruídos, foi necessário restringir o intervalo de análise de 400 a 2500 nm, de forma que esses dados abaixo da faixa de 400 nm, não interferissem negativamente na resposta dos modelos aplicados, conforme ilustrado na Figura 38.

Figura 38. Curva Espectral da amostra de produto CN_10585 e o destaque do ruído faixa abaixo de 400 nm.



Fonte: Os autores.

Além de analisar a faixa de 400 a 2500 nm, com o foco de buscar as bandas mais significativas na estimativa dos teores de ferro das amostras de minério, foi realizada a análise dentro da faixa de 400 a 1310 nm. Dessa forma, com o foco maior nas feições mais específicas do ferro, a análise poderia ser realizada com um menor volume de dados e assim requerer menos poder computacional, de forma a otimizar a análise.

6.3 AVALIAÇÃO DE VARIAÇÕES DOS MODELOS

Abaixo, são apresentados os resultados alcançados pela aplicação dos métodos de aprendizado de máquinas, tanto nas amostras de produto quanto de laboratório, com o objetivo de estimar o percentual de teor de ferro. São apresentados os resultados gerados por cada modelo, assim como a avaliação estatística de sua acurácia.

Com o intuito de associar a base de dados para validação do modelo de forma estatística, foram adotados como parâmetros, o coeficiente de determinação R^2 , o erro médio absoluto (MAE) e o erro médio quadrático (MSE). Assim, com essas métricas de avaliação, podemos observar as relações entre os dados separados para validação e os dados de teores de ferro estimados por cada modelo.

Para os modelos de aprendizados de máquinas foram considerados nas análises os parâmetros abaixo, de forma que a carga computacional fosse adequada para a quantidade de dados a serem processadas e assim não exigir muitos recursos de processamento sem ganhos mensuráveis de performance:

- Random Forest: número de árvores igual a 30;
- Redes Neurais Artificiais: número de neurônios 100 e 100 camadas ocultas;
- *AdaBoost*: Número de estimadores igual a 50;
- kNN: Número de vizinhanças igual a 5;
- SVM: Número limite de interações: 50;
- Arvore de decisão: Limite máximo de profundidade da árvore de 50.

6.3.1 Dataset das Amostras de Produto

Foram geradas 2 bases de dados distintas para a análise das amostras de produto, sendo a primeira considerando o espectro mais amplo na faixa de 400 a 2500 nm e o espectro reduzido, focado nas feições mais características do ferro no intervalo de 400 a 1310 nm.

6.3.1.1 Intervalo de 400 a 2500 nm

A Tabela 3 apresenta os resultados estatísticos encontrados para cada método. Pode-se verificar o desempenho superior dos métodos de SVM, *AdaBoost* e RF que apresentaram MSE

e MAE mais baixos e coeficiente de determinação R^2 maiores. Estes métodos apresentaram melhor performance dentre todos os modelos avaliados. Já a RNA, com valores muito inferiores, em especial de R^2 , não se torna apropriada para aplicação nesse estudo em questão.

Tabela 3. Métodos de Aprendizado de máquinas aplicados nas amostras de produto na faixa de 400 a 2500 nm e resultados estatísticos.

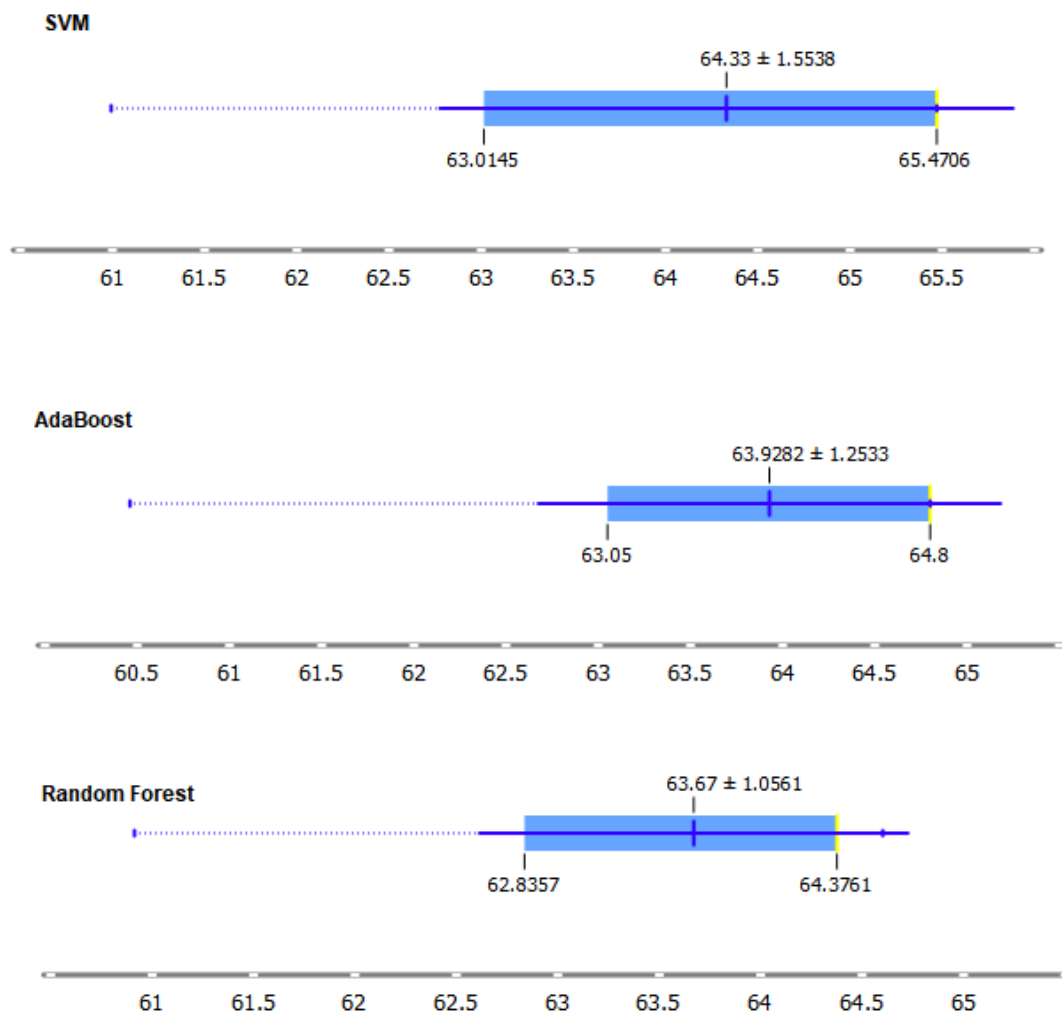
Modelo	R^2	MSE	MAE
Random Forest (RF)	0,770	0,554	0,636
Redes Neurais Artificiais (RNA)	0,266	1,770	1,114
AdaBoost	0,849	0,364	0,477
kNN	0,680	0,771	0,717
SVM	0,867	0,321	0,441
Arvore de decisão	0,557	1,067	0,809

Fonte: Os autores.

A Figura 39 mostra os gráficos *boxplot* dos principais métodos avaliados. Todos os métodos apresentaram uma tendência bem centralizada. O SVM, apesar de ter como resultado o maior coeficiente R^2 , foi o que mostrou a maior dispersão dos dados e menor intervalo interquartilico, com diferença entre o terceiro e primeiro quartis, com amplitude de 63,0145 a 65,4706.

Já o *AdaBoost*, com dados mais agrupados próximos a mediana e menor intervalo interquartilico, o que sugere ser o melhor método para aplicação para esse grupo de amostras.

Figura 39. Gráficos boxplot dos métodos SVM, AdaBoost e RF para as amostras de produto na faixa de 400 a 2500 nm.



Fonte: Os autores.

6.3.1.2 Intervalo de 400 a 1310 nm

A Tabela 4 apresenta os resultados estatísticos encontrados para cada método. Pode-se dizer os métodos SVM, *AdaBoost*, RF e Arvore de decisão, considerando os critérios analisados, são os mais eficientes, pois apresentaram MSE e MAE mais baixos e coeficiente de determinação R^2 maiores.

Assim como na análise anterior a RNA, com valores muito inferiores, em especial de R^2 , não se torna apropriada para aplicação nesse estudo em questão.

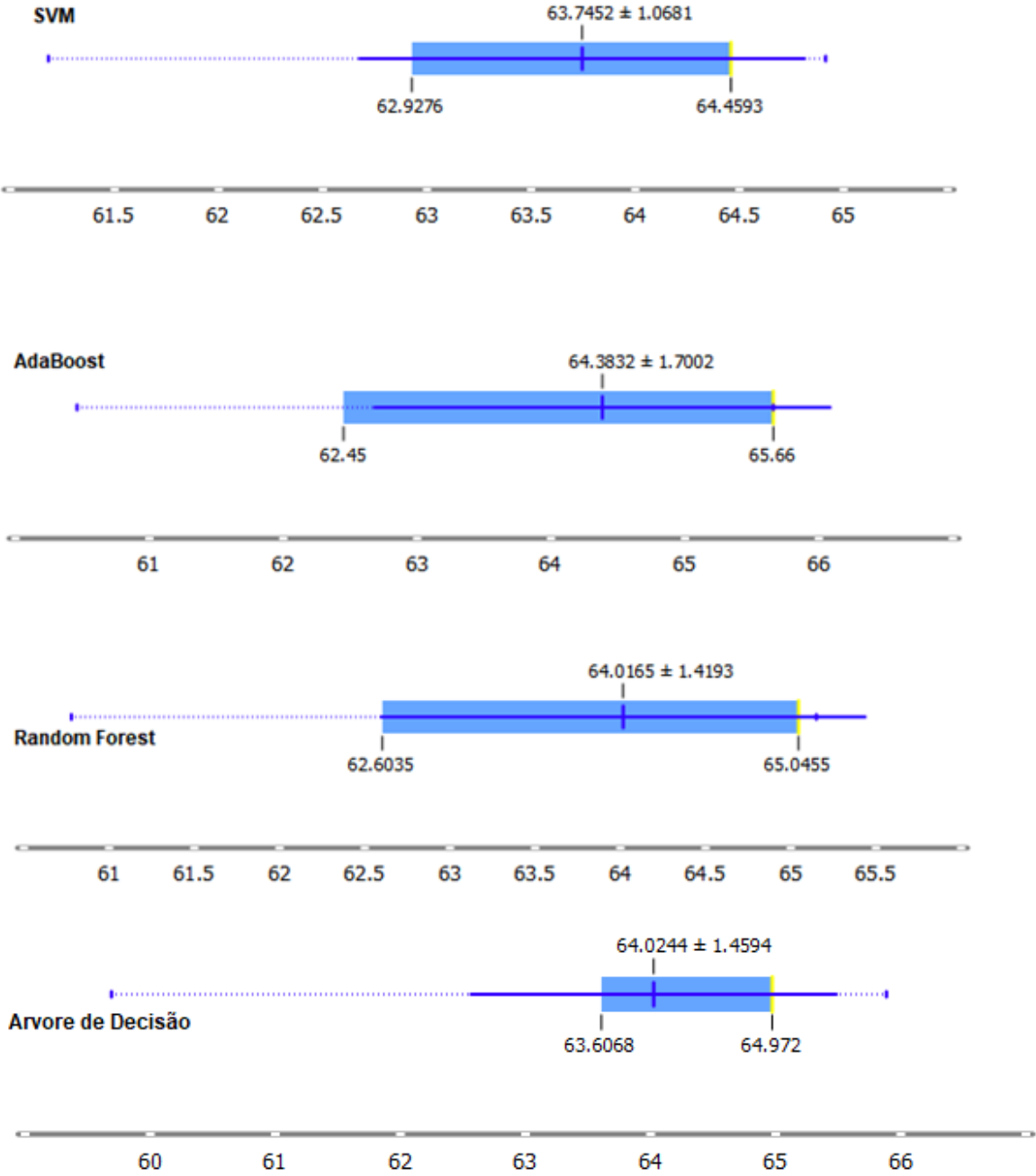
Tabela 4. Métodos de Aprendizado de máquinas aplicados nas amostras de produto na faixa de 400 a 1310 nm e resultados estatísticos.

Modelo	R2	MSE	MAE
Random Forest (RF)	0,722	0,670	0,668
Redes Neurais Artificiais (RNA)	0,340	1,590	1,030
AdaBoost	0,710	0,699	0,615
kNN	0,655	0,831	0,737
SVM	0,767	0,562	0,631
Arvore de decisão	0,715	0,687	0,686

Fonte: Os autores.

A Figura 40 exibe os gráficos *boxplot* dos principais métodos avaliados. A Árvore de decisão e o SVM foram os que mostraram o melhor agrupamento dos dados, sendo o SVM com as informações mais próximas da mediana e centralizadas. Porém, todos os métodos apresentaram resultados satisfatórios e, portanto, podem ser considerados para essa aplicação.

Figura 40. Gráficos boxplot dos métodos SVM, AdaBoost e RF para as amostras de produto na faixa de 400 a 1310 nm.



Fonte: Os autores.

6.3.2 Dataset das Amostras Preparadas em Laboratório

Assim como nas análises anteriores, foram geradas 2 bases de dados distintas para a análise das amostras de laboratório, sendo a primeira considerando o espectro mais amplo na faixa de 400 a 2500 nm e o espectro reduzido, focado nas feições mais características do ferro no intervalo de 400 a 1310 nm.

6.3.2.1 Intervalo de 400 a 2500 nm

A Tabela 5 mostra os resultados estatísticos encontrados para cada método. Foram observados pioras nas respostas dos modelos em relação as análises em cima das amostras de produto de sínter feed. Podemos verificar o desempenho superior do método kNN, que apresentou MSE e MAE mais baixos e coeficiente de determinação R^2 maior, que o torna mais indicado dentre os modelos avaliados. Os demais métodos, todos com valores muito inferiores, se tornam inapropriados para aplicação nesse estudo em questão.

Tabela 5. Métodos de Aprendizado de máquinas aplicados nas amostras de laboratório na faixa de 400 a 2500 nm e resultados estatísticos.

Modelo	R^2	MSE	MAE
Random Forest (RF)	0,129	16,763	0,129
Redes Neurais Artificiais (RNA)	-0,920	36,971	-0,920
AdaBoost	-0,179	22,697	-0,179
kNN	0,610	7,512	0,610
SVM	0,428	11,009	0,428
Arvore de decisão	-0,924	37,036	-0,924

Fonte: Os autores.

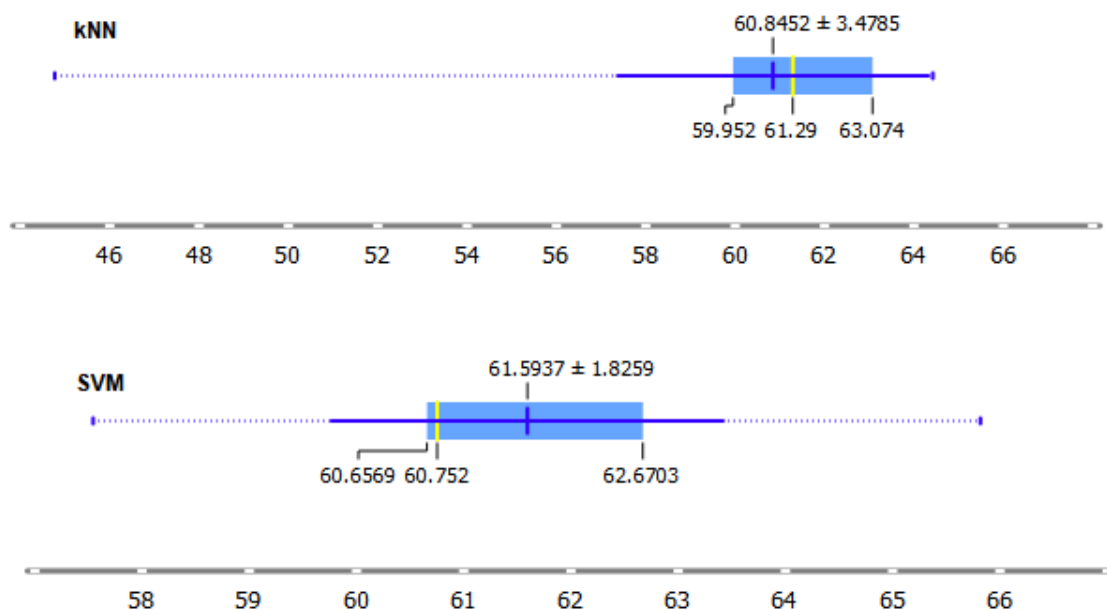
A possibilidade desse modelo não ter apresentado boa performance, pode ser atribuída a muitas características espectrais redundantes devido à forte correlação entre as bandas de onda adjacentes. Isso é intensificado devido a preparação dos dados em laboratório, nos processos

de pulverização e secagem, que torna a amostra mais homogênea. Nesse caso, a análise é complexa e precisa ser simplificada, selecionando os recursos espectrais mais relevantes e assim restringindo a banda a região do espectro onde as feições do ferro são mais marcantes.

A Figura 41, mostra os gráficos *boxplot* dos principais métodos avaliados. O kNN, apesar de ter apresentado o maior coeficiente R^2 , foi o que mostrou a maior dispersão dos dados, menor intervalo interquartílico, com diferença entre o terceiro e primeiro quartis, com amplitude de 59,952 a 63,074.

Já o SVM, com dados mais agrupados próximos a mediana, mais centralizados e menor intervalo interquartílico, apresentou melhor desempenho no *boxplot*, apesar de ter apresentado um coeficiente R^2 de 0,428, não satisfatório para validação do modelo.

Figura 41. Gráficos *boxplot* dos métodos kNN e SVM para as amostras de laboratório na faixa de 400 a 2500 nm.



Fonte: Os autores.

6.3.2.2 Intervalo de 400 a 1310 nm

A Tabela 6 apresenta os resultados estatísticos encontrados para cada método. Podemos verificar o desempenho superior dos métodos de AdaBoost, RF, RNA, Árvores de decisão e kNN, que apresentaram MSE e MAE mais baixos e coeficiente de determinação R^2 acima de

0.9. Dessa forma, todos esses métodos se mostraram bastante eficientes para a aplicação em questão. Além disso, podemos dar ênfase ao *AdaBoost*, com valor R^2 de 1.00 e com MSE e MAE próximos de 0, pois a diferença entre os valores estimados e reais (obtidos em análises de laboratório) é mínima. Situação que pode indicar que o modelo possa ter “decorado” os dados e não ser o mais indicado para essa aplicação.

O método com a pior performance foi o SVM, o que não o torna apropriado para aplicação nesse estudo em comparado com os demais, apesar de ter conseguido um resultado de coeficiente R^2 de 0,718.

Tabela 6. Métodos de Aprendizado de máquinas aplicados nas amostras de laboratório na faixa de 400 a 1310 nm e resultados estatísticos.

Modelo	R^2	MSE	MAE
Random Forest (RF)	0,990	0,201	0,263
Redes Neurais Artificiais (RNA)	0,989	0,215	0,360
AdaBoost	1,000	0,000	0,000
kNN	0,971	0,562	0,380
SVM	0,718	5,429	1,109
Arvore de decisão	0,979	0,405	0,359

Fonte: Os autores.

Um dos fatores que contribuíram para boa performance da análise dos métodos de aprendizado de máquina das amostras de laboratório na faixa de 400 a 1310, são referentes aos dados considerados como principais atributos na análise, como mostra a Figura 42. Os valores apresentados, correspondem a faixa espectral com feições muito significativas de ferro na casa de 890 nm.

Figura 42. Ranking dos principais atributos considerados nas análises dos modelos de aprendizado para as amostras de laboratório na faixa de 400 a 1310 nm.

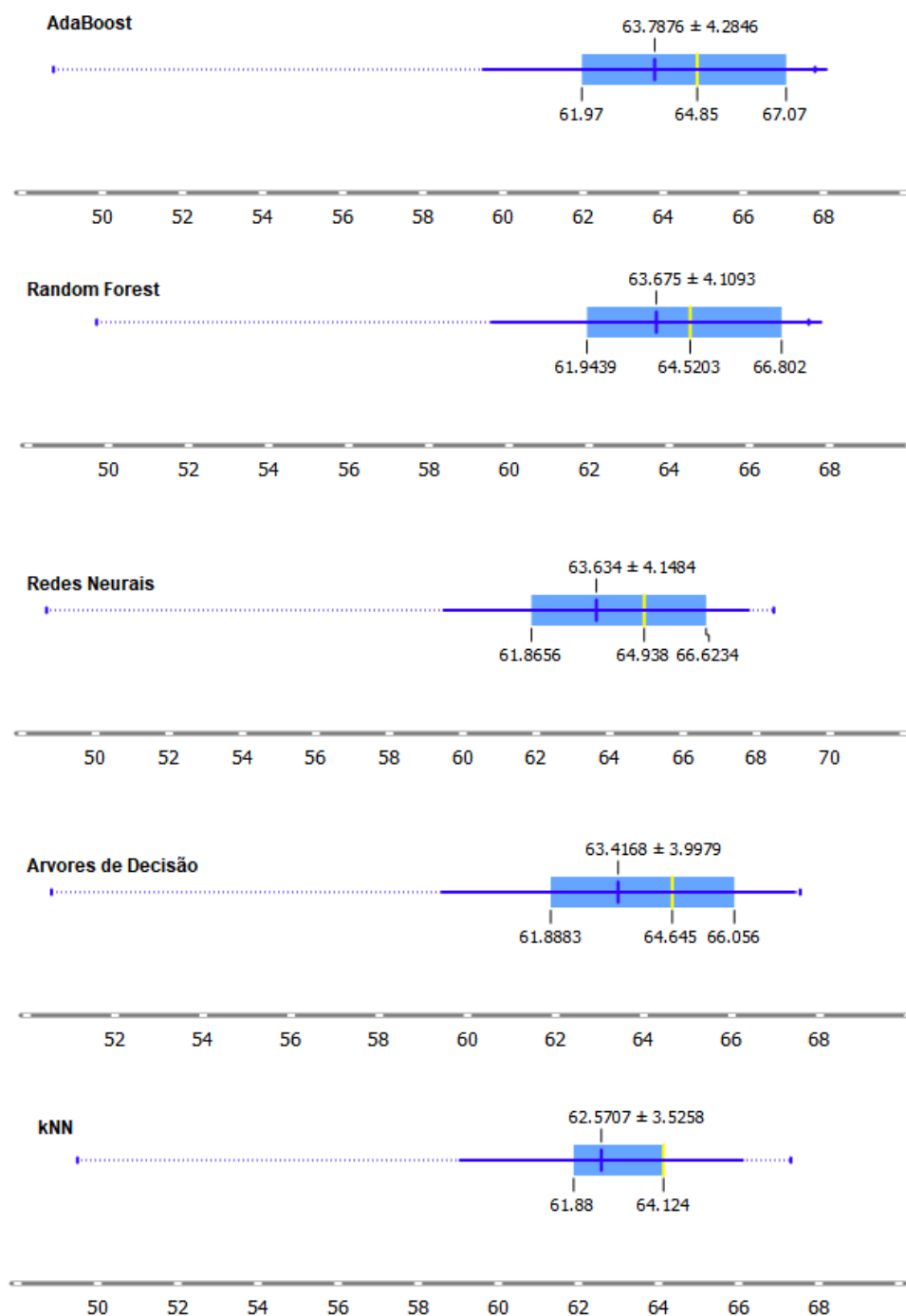
#	Univar. reg.	RReliefF
N 890		
N 891		
N 889		
N 888		
N 892		
N 893	110.768	0.277
N 887	110.745	0.276
N 894	110.725	0.278
N 895	110.639	0.278
N 886	110.635	0.276
N 885	110.521	0.276
N 896	110.497	0.278
N 897	110.407	0.278
N 884	110.373	0.275
N 898	110.328	0.279
N 883	110.210	0.275
N 899	110.205	0.279
N 882	110.065	0.275
N 900	109.999	0.279
N 881	109.904	0.274

Fonte: Os autores.

A Figura 43, mostra os gráficos *boxplot* dos principais métodos avaliados. É possível observar comportamento muito semelhante dos métodos RF, RNA e Árvore de decisão na semelhança entre dispersão dos dados e intervalo interquartílico.

Já o kNN, apresentou menor intervalo interquartílico, com diferença entre o terceiro e primeiro quartis, com amplitude de 61,88 a 64,124, portanto se destacando entre os métodos que apresentaram melhor coeficiente R^2 .

Figura 43. Gráficos boxplot dos métodos AdaBoost, RF, RNA, Árvore de decisão e kNN para as amostras de laboratório na faixa de 400 a 1310 nm.



Fonte: Os autores.

7 CONCLUSÕES E CONSIDERAÇÕES FINAIS

No âmbito desta pesquisa, foi possível determinar a resposta espectral de amostras de *sinter feed* no intervalo *VSWIR*. A caracterização das feições de absorção das amostras de produto de *sinter feed*, apresentam melhores resultados, justamente pelo fato de possuírem maiores variações de cores e granulometria, fundamental para que se permita realizar algum tipo de caracterização física e espectral. Contudo, é difícil ainda prever qual seria o impacto da não preparação da amostra nos processos e quantificação de teores, em especial aqueles relacionados com os contaminantes.

A espectroscopia de refletância no intervalo *VSWIR* utilizada na análise de produtos de minério de ferro, mais especificamente *sinter feed*, demonstrou um ganho ao retirar a interferência da umidade no espectro, porém ao mesmo tempo, as curvas de refletância passaram a se assemelhar muito, o que tornou um fator dificultante para a determinação de contaminantes.

Para os métodos de aprendizado de máquinas aplicados como o SVM e *AdaBoost* nos testes realizados nas amostras de produto, com a abrangência do espectro de 400 a 2500 nm, foram obtidos coeficientes de determinação R^2 acima de 0,8. Dessa forma, os resultados se mostram bastante satisfatórios para aplicação em casos nos quais não é necessário um fator de precisão extremamente elevado. Principalmente pelo fato de que, ao se tornar possível uma análise dos dados nas amostras de produto de *sinter feed* sem a necessidade de nenhuma preparação. Os ganhos são grandes em relação ao tempo em que os resultados serão adquiridos e também no que tange ao custo, uma vez que não seria necessário passar por todo o processo de preparação da amostra em laboratório e por todos os processos, materiais e equipamentos envolvidos.

O melhor resultado foi obtido na aplicação dos métodos de aprendizado de máquina nas amostras preparadas em laboratório, considerando o espectro de 400 a 1310 nm, o qual foi possível o modelo focar na faixa espectral nas quais as características do ferro são mais marcantes. Como estas amostras passaram por processos de pulverização e secagem em laboratório, a interferência de outros fatores, como por exemplo a umidade presente no material, foi minimizada. Para esse caso, vários modelos, como *AdaBoost*, RF, RNA, Árvores de decisão e kNN, que apresentaram coeficiente de determinação R^2 acima de 0,9 e ofereceram respostas satisfatórias para estimativa de percentual de teor de ferro.

Por outro lado, os testes realizados nas amostras de produto, com a abrangência do espectro de 400 a 2500 nm, foram obtidos bons resultados, ao levar em consideração que para as análises, não foram necessárias nenhum processo de preparação da amostra. Assim, temos ganhos no tempo de resposta, uma vez que a amostra pode ser colhida diretamente do processo produtivo, além de custos inerentes aos processos de preparação em laboratório. Diante disto, métodos como o SVM e *AdaBoost* com coeficientes de determinação R^2 acima de 0,8, podem ser interessantes para aplicação em os casos nos quais não é necessário um fator de precisão extremamente elevado.

Diante dos resultados dessa pesquisa, cria-se a possibilidade da aplicação dos métodos de aprendizado de máquinas diretamente em amostras de produto. Os dados, métodos e acervo de conhecimento produzido nesse estudo, podem se tornar viáveis para testes de forma operacional no monitoramento da qualidade de *sinter feed* em correias transportadoras, em operações de beneficiamento de minério de ferro.

Para efeito de continuidade deste trabalho, uma análise interessante, seria o foco na estimativa de teores dos outros elementos que compõem uma amostra de minério de ferro e que atuam como contaminantes, como por exemplo: sílica, alumina, manganês, óxido de cálcio, fósforo, dentre outros.

REFERÊNCIAS

ABU-MOSTAFA, Yaser S.; MAGDON-ISMAIL, Malik; LIN, Hsuan-Tien. **Learning from data**. New York, NY, USA: AMLBook, 2012.

ADAMS, John B. Visible and near-infrared diffuse reflectance spectra of pyroxenes as applied to remote sensing of solid objects in the solar system. **Journal of Geophysical Research**, v. 79, n. 32, p. 4829-4836, 1974.

ADAMS, John B. Interpretation of visible and near-infrared diffuse reflectance spectra of pyroxenes and other rock-forming minerals. **Infrared and Raman spectroscopy of lunar and terrestrial minerals**, p. 91-116, 1975.

ALPAYDIN, Ethem. **Introduction to machine learning**. MIT press, 2020.

BIAU, G., SCORNET, E. A random forest guided tour. **Test**, v. 25, n. 2, pp. 197–227, 2016.

BORGES, FABBIO AS et al. **Metodologia baseada no algoritmo adaboost combinado com rede neural para localização do distúrbio de afundamento de tensão**. In: Congresso Brasileiro de Automática-CBA. 2019.

BROWNLEE, J. Overfitting and underfitting with machine learning algorithms. **Machine Learning Mastery**, v. 21, 2016.

BURNS, Roger G.; BURNS, Roger George. **Mineralogical applications of crystal field theory**. Cambridge university press, 1993.

CARIOCA, A. C. **Caracterização de minério de ferro por espectroscopia de refletância difusa**. (Dissertação de Mestrado). UFOP, Escola de Minas, Ouro Preto, 2010.

CHEN, Nianyi. **Support vector machine in chemistry**. World Scientific, 2004.

CHEN, S. H., JAKEMAN, A. J., NORTON, J. P. Artificial intelligence techniques: an introduction to their use for modelling environmental systems. **Mathematics and computers in simulation**, v. 78, n. 2-3, pp. 379–400, 2008.

CIAMPI, Antonio; ZIGHED, Djamel A.; CLECH, Jérémy. **Trees and induction graphs for multivariate response**. In: European Conference on Principles of Data Mining and Knowledge Discovery. Springer, Berlin, Heidelberg, 2000. p. 359-366.

CLARK, R. N. Spectroscopy of Rocks and Minerals, and Principles of Spectroscopy. In: RENCZ, A. N. e RYERSON, R. (Ed.). **Manual of Remote Sensing, Remote Sensing for the Earth Sciences**. 3rd. New York: John Wiley & Sons, cap. 1, p.728, 1999.

CLARK, R. N.; ROUSH, T. L. Reflectance Spectroscopy - Quantitative-Analysis Techniques for Remote-Sensing Applications. **Journal of Geophysical Research**, v. 89, n. Nb7, p. 6329-6340, 1984.

CLARK, R. N. et al. High spectral resolution reflectance spectroscopy of minerals. **Journal of Geophysical Research**, v. 95, n. B8. 1990

CUTLER, A., CUTLER, D. R., STEVENS, J. R. “**Random forests**”. Em: Ensemble machine learning, Springer, pp. 157–175, 2012.

CSN Mineração S.A. **Portifólio de Produtos e Serviços – Minério de Ferro**. Disponível em: http://www.csn.com.br/conteudo_pti.asp?idioma=0&tipo=61062&conta=45. Acesso em Maio 2020.

CORNELL, R. H.; SCHWERTMANN, U. **The iron-oxides**. VCH Publ. 1996.

DE MYTTENAERE, A., GOLDEN, B., LE GRAND, B. et al. Mean absolute percentage error for regression models. **Neurocomputing**, v. 192, pp. 38–48, 2016.

DE MYTTENAERE, A., GOLDEN, B., LE GRAND, B. et al. Mean absolute percentage error for regression models. **Neurocomputing**, v. 192, pp. 38–48, 2016.

DUARTE, Hélio A. Ferro - Um elemento químico estratégico que permeia história, economia e sociedade. **Quím. Nova**, São Paulo , v. 42, n. 10, p. 1146-1153, Oct. 2019 . Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-40422019005001146&lng=en&nrm=iso>. Acesso em Maio de 2020.

FIGUEIREDO FILHO, D. B., SILVA JÚNIOR, J. A. D. **Desvendando os Mistérios do Coeficiente de Correlação de Pearson (r)**. 2009.

FILGUEIRAS, Paulo Roberto et al. **Regressão por vetores de suporte aplicado na determinação de propriedades físico-químicas de petróleo e biocombustíveis**. 2014.

FREUND, Yoav et al. **Experiments with a new boosting algorithm**. In: icml. 1996. p. 148-156.

GOTELLI, NICHOLAS J.; ELLISON, AARON M. **Princípios de estatística em ecologia**. Artmed Editora, 2011.

GUPTA, H. V.; KLING, H.; YILMAZ, K. K. et al. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. **Journal of hydrology**, v. 377, n. 1-2, pp. 80–91, 2009.

HU, P., LIU, X., CAI, Y., et al. **Band selection of hyperspectral images using multiobjective optimization-based sparse self-representation**. IEEE Geoscience and Remote Sensing Letters, v. 16, n. 3, pp. 452–456, 2019.

HUNT, G.R.; **Electromagnetic radiation**: the communication link in remote sensing. In: Siegal, B.S.; Gillespie, A.R. Remote sensing in geology. New York: J. Wiley & Sons, p. 5-45, 1980.

HUNT, G.R., 1977. Spectral signatures of particulate minerals in the visible and near-infrared. **Geophysics**, v. 42, 501–513.

IBRAM, Instituto Brasileiro de Mineração. Relatório anual de atividades – Julho 2018 a Junho 2019. **Portal da Mineração**. 2019. Disponível em: <http://portaldamineracao.com.br/ibram/wp-content/uploads/2019/07/relatorio-anual-2018-2019.pdf>. Acesso em Maio 2020.

JENKINS, Ron. **X-Ray Fluorescence Spectrometry**. Second edition. New York: Wiley, 1999.

KOKALY, R. F. Investigating a physical basis for spectroscopic estimates of leaf nitrogen concentration. **Remote Sensing of Environment**. v. 75, n. 2, p. 153-161, Feb 2001.

KUMAR, A., MEHTA, S., VIJAYKEERTHY, D. **An Introduction to Adversarial Machine Learning**. In: International Conference on Big Data Analytics, pp. 293–299. Springer, 2017.

LANDIM, P. M. B. **Análise estatística de dados geológicos multivariados**. Editora Oficina de Textos, 1º edição, São Paulo. 2011.

LANTZ, Brett. **Machine learning with R: expert techniques for predictive modeling**. Packt publishing ltd, 2019.

LEITE, E. P., DE SOUZA FILHO, C. R. Artificial neural networks applied to mineral potential mapping for copper-gold mineralizations in the Carajás Mineral Province, Brazil. **Geophysical Prospecting**, v. 57, n. 6, pp. 1049–1065, 2009.

LINDEN, Ricardo. **Técnicas de agrupamento. Revista de Sistemas de Informação da FSMA**, v. 4, n. 4, p. 18-36, 2009.

MARÔCO, J. **Análise Estatística com o SPSS Statistics**. 7ª edição. ReportNumber, Lda, 2018.

MENESES, Paulo Roberto; ALMEIDA, T. de. **Introdução ao processamento de imagens de sensoriamento remoto**. Universidade de Brasília, Brasília, 2012.

MILTON, E. J. et al. Progress in field spectroscopy. **Remote Sensing of Environment**, v. 113, p. S92-S109, Sep 2009.

MONARD, Maria Carolina; BARANAUSKAS, José Augusto. Conceitos sobre aprendizado de máquina. **Sistemas inteligentes-Fundamentos e aplicações**, v. 1, n. 1, p. 32, 2003.

MONARD, Maria Carolina; BARANAUSKAS, José Augusto. Conceitos sobre aprendizado de máquina. **Sistemas inteligentes-Fundamentos e aplicações**, v. 1, n. 1, p. 32, 2003.

MUNDSTOCK, Elsa et al. **Introdução à Análise Estatística utilizando o SPSS 13.0**. Cadernos de Matemática e Estatística Série B. Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, 2006.

NEUMANN, Reiner; SCHNEIDER, Cláudio Luiz; ALCOVER NETO, Arnaldo. **Parte II: Caracterização tecnológica de minérios**. CETEM/MCT, 2010.

NEVES, A.A. **Quantificação de óxidos de ferro por espectroscopia de reflectância difusa**. Tese de Doutorado – UFV. Viçosa, 2001.

OZAKI, Y.; MCCLURE, W. F.; CHRISTY, A. A. **Near-Infrared Spectroscopy in Food Science and Technology**. Wiley, 2006.

PINHEIRO, J. C. F. **A Mineração Brasileira de Ferro e a Reestruturação do Setor Siderúrgico**. Tese de Doutorado – UNICAMP. Campinas, SP, 2000.

PRADO, E. M. G., **Espectrorradiometria de reflectância aplicada à análise quantitativa da mineralogia do depósito N4WS, Província Mineral de Carajás, Pará, Brasil.** UNB, Brasília, 2015. (Dissertação de Mestrado).

QUARESMA, L. F. **Ferro. Balanço Mineral Brasileiro.** 2001. Disponível em: <http://antigo.anm.gov.br/portal/dnpm/paginas/balanco-mineral/arquivos/balanco-mineral-brasileiro-2001-ferro>. Acesso em Junho de 2020.

QUARESMA, L. F. Perfil da mineração de ferro. Produto 09 - Minério de ferro. Relatório técnico 18. Secretaria de Geologia, Mineração e Transformação Mineral- SGM. **Ministério de Minas e Energia – MME.** Agosto, 2009. Disponível em : http://www.mme.gov.br/documents/36108/448620/P09_RT18_Perfil_da_Mineraçao_de_Ferro.pdf/362a62a6-04db-b739-98d0-03f7ff9f749f?version=1.0. Acesso em Junho de 2020.

REZENDE, Solange Oliveira. **Sistemas inteligentes: fundamentos e aplicações.** Editora Manole Ltda, 2003.

SANTOS, Elenir Souza et al. Espectrometria de Fluorescência de Raios-X na determinação de espécies químicas. **Enciclopédia biosfera**, v. 9, n. 17, p. 3413-3432, 2013.

SANTOS, U. J. L., PESSIN, G., DA COSTA, C. A., et al. AgriPrediction: A proactive internet of things model to anticipate problems and improve production in agricultural crops. **Computers and electronics in agriculture**, v. 161, pp. 202–213, 2019.

SAVITZKY, A.; GOLAY, M. J. E. Smoothing + Differentiation of Data by Simplified Least Squares Procedures. **Analytical Chemistry**, v. 36, n. 8, p. 1627-&, 1964.

SCHEINOST, A. C. e SCHUWERTMANN, U. Color identification of iron oxides na hidroxysulfates: use and limitations. **Soil Science Society of America**. 1999, v.63 (5), p.1463-1471.

SOUSA, M.F.; COUTINHO, M. A. N.; AMARAL, L.R.; FRANCO, H. C. J. **Capacidade da espectroscopia VIS-NIR em diferenciar níveis de fósforo no solo.** Apresentado no XLVI Congresso Brasileiro de Engenharia Agrícola – CONBEA. Maceió - AL, Brasil, 2017.

TEIXEIRA, S. **Determinação de modelo de estimativa de teores de carbono em solos utilizando máquina de vetor de suporte e reflectância espectral.** (Dissertação de Mestrado). UEPG, Ponta Grossa, 2014.

VAN DER MEER, F., 2004. Analysis of spectral absorption features in hyperspectral imagery. **Int. J. Earth Observ. Geoinf.** 5 (1), 55– 68.

VIANA, A. O. **Métodos de aprendizado de máquina e imageamento hiperespectral aplicados na estimação dos teores de ferro de minérios.** (Dissertação de Mestrado). UFOP, ITV, Ouro Preto, 2020.

WASAN, Siri Krishan; BHATNAGAR, Vasudha; KAUR, Harleen. The impact of data mining techniques on medical diagnostics. **Data Science Journal**, v. 5, p. 119-126, 2006.

WATT, J., BORHANI, R., KATSAGGELOS, A. **Machine learning refined:** foundations, algorithms, and applications. Cambridge University Press, 2020.

WEISBERG, S. **Applied Linear Regression.** Wiley, 2013. ISBN 9781118594858.

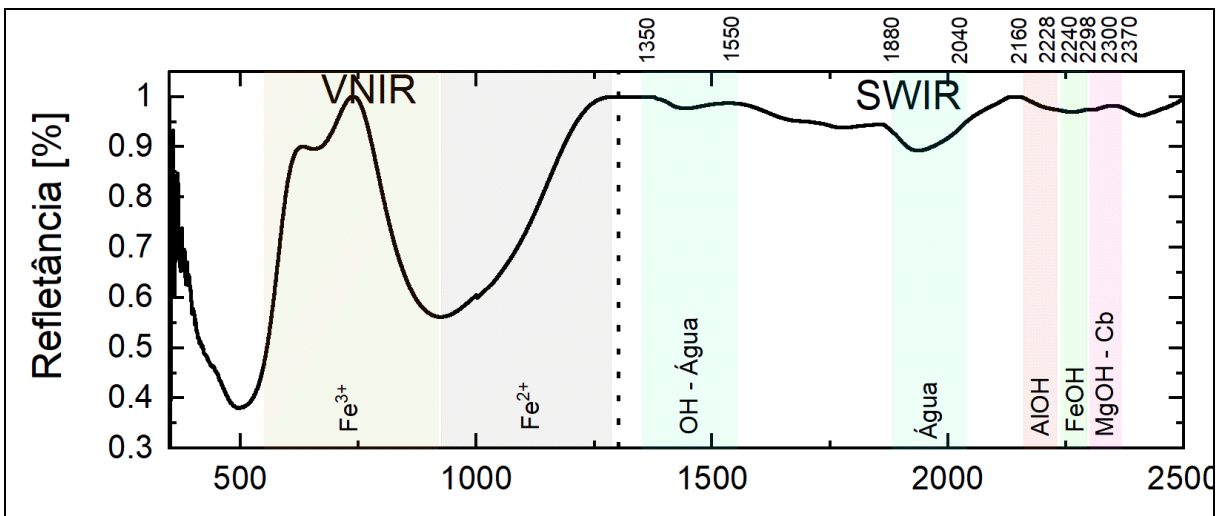
WILLMOTT, C. J., MATSUURA, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. **Climate research**, v. 30, n. 1, pp. 79–82, 2005.

ZHAO, C.; GAO, Y.; HE, J. et al. Recognition of driving postures by multiwavelet transform and multilayer perceptron classifier. **Engineering Applications of Artificial Intelligence**, v. 25, n. 8, pp. 1677–1686, 2012.

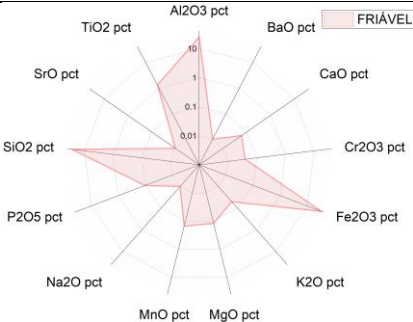
ANEXO

ANEXO 1- Exemplo de Ficha Espectral do Banco de Dados Espectrais Vale

CN-01	
Descrição: Amostra de produto de Sinter Feed	
Equipamento: FieldSpec 4 Hi-Res NG	
Intervalo Espectral: 350 -2500 nm	
Localização: Mina Serra Sul (S11D), PA- Brasil	
Responsáveis: Ana Cristina Silva e Rosa Correa Pabón	



WAVELENGTH (NM)	DEPTH	WIDTH (NM)	AREA	ASYMMETRY
514	56.040	188.454	10561.006	-0.345
900	24.281	206.253	5008.093	0.123
2208	22.439	74.273	1666.597	0.115
1414	9.974	59.174	590.224	-0.241
1915	5.152	156.929	808.544	0.130

	Fe [%]	63.75	TiO2 [%]	0.039
	SiO2 [%]	8.08	CaO [%]	0.006
	P [%]	0.014	MgO [%]	0.0254
	Al2O3 [%]	0.2	K2O [%]	0.0037
	Mn [%]	0.048		

ANEXO 2- Passo a passo – Realização da leitura espectral das amostras e criação da biblioteca espectral

PASSO A PASSO – REALIZAÇÃO DA LEITURA DAS AMOSTRAS E CRIAÇÃO DA BIBLIOTECA ESPECTRAL

1 PREPARAÇÃO DAS AMOSTRAS

O primeiro passo do processo, foi a separação das amostras coletadas de Serra Sul para a realização da leitura dos espectros através do espectro-radiômetro. A figura 1 mostra algumas amostras de produto junto com suas devidas identificações. Da mesma forma, foram também realizadas leituras de amostras de material preparado em laboratório, conforme figura 2, para fins de comparações futuras.

Figura 1: Amostras de Sinter Feed.



Fonte: Autores.

Figura 2: Amostras de Sinter Feed Processada em Laboratório.



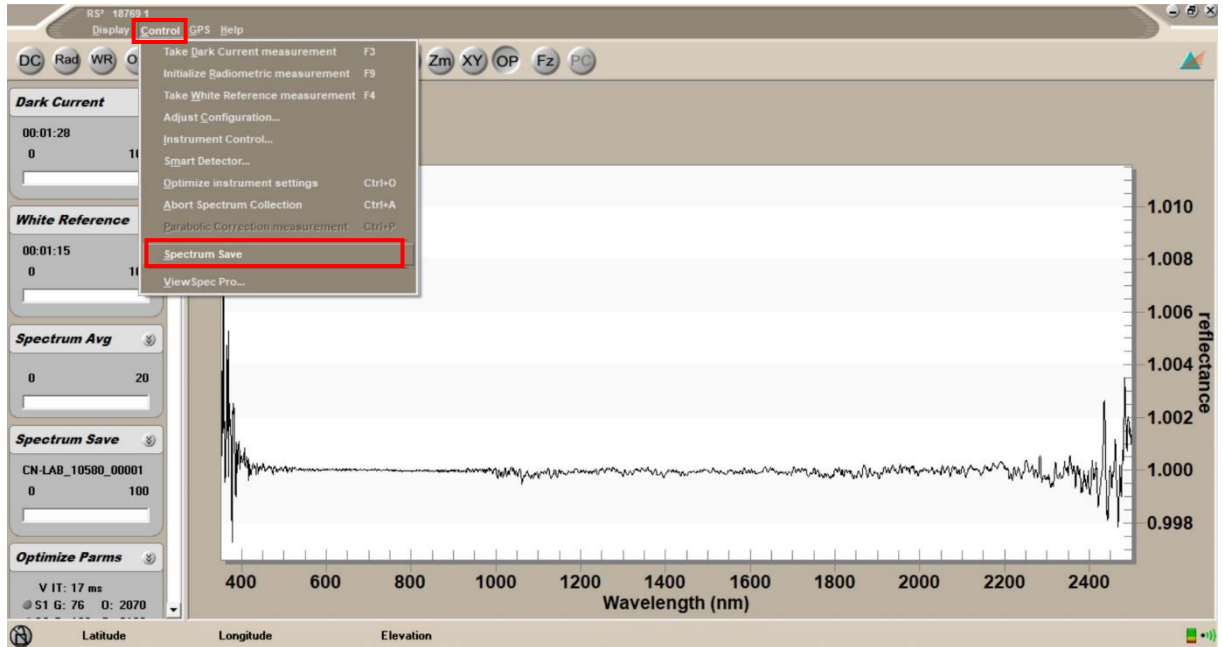
Fonte: Autores.

3 LEITURA DAS AMOSTRAS

Na sequência, foram realizadas as leituras das amostras através do software RS³ em conjunto com os equipamentos *Turntable* e *FieldSpec 4*.

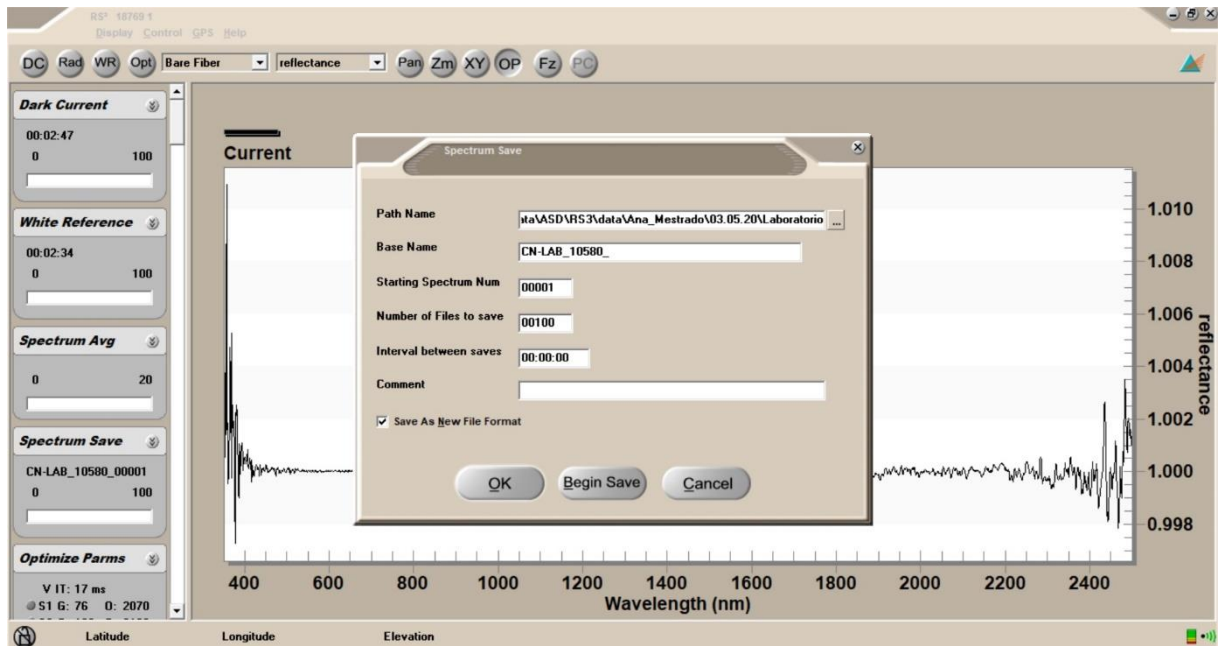
Nas figuras 3 e 4 são ilustrados o processo de obtenção do espectro da respectiva amostra. No *software* é inserido o caminho onde os arquivos serão salvos, o nome e o número de arquivos que serão salvos referentes a aquela leitura.

Figura 3: Tela do software RS³



Fonte: Autores.

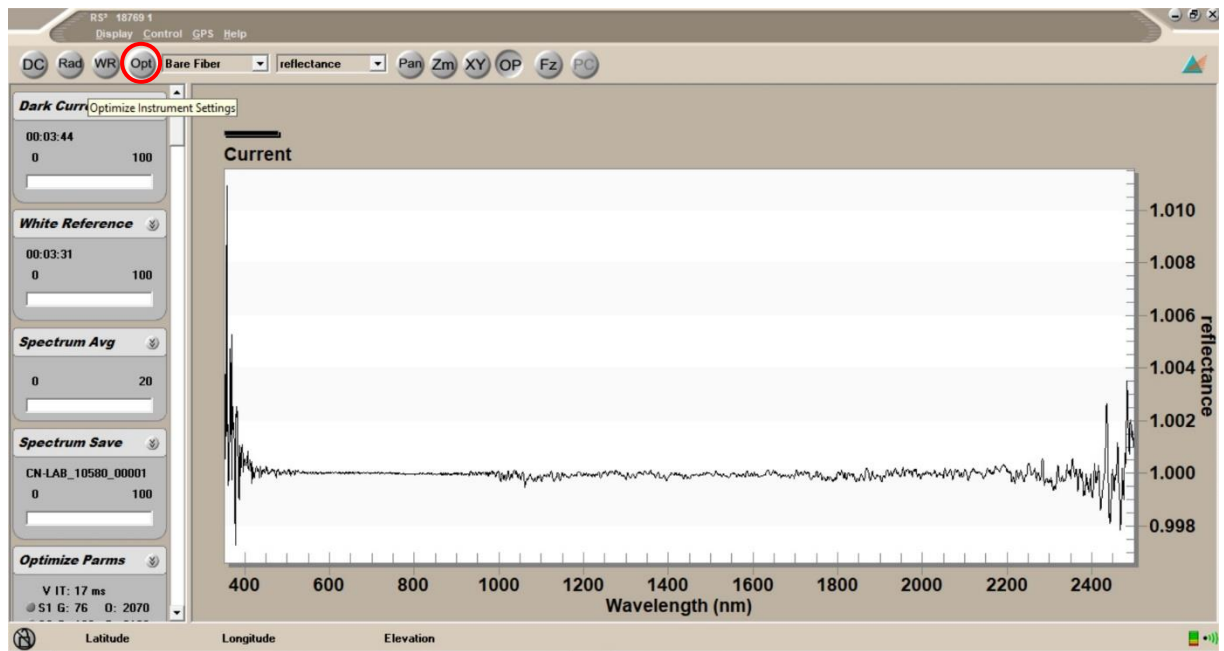
Figura 4: Tela de parâmetros do software



Fonte: Autores.

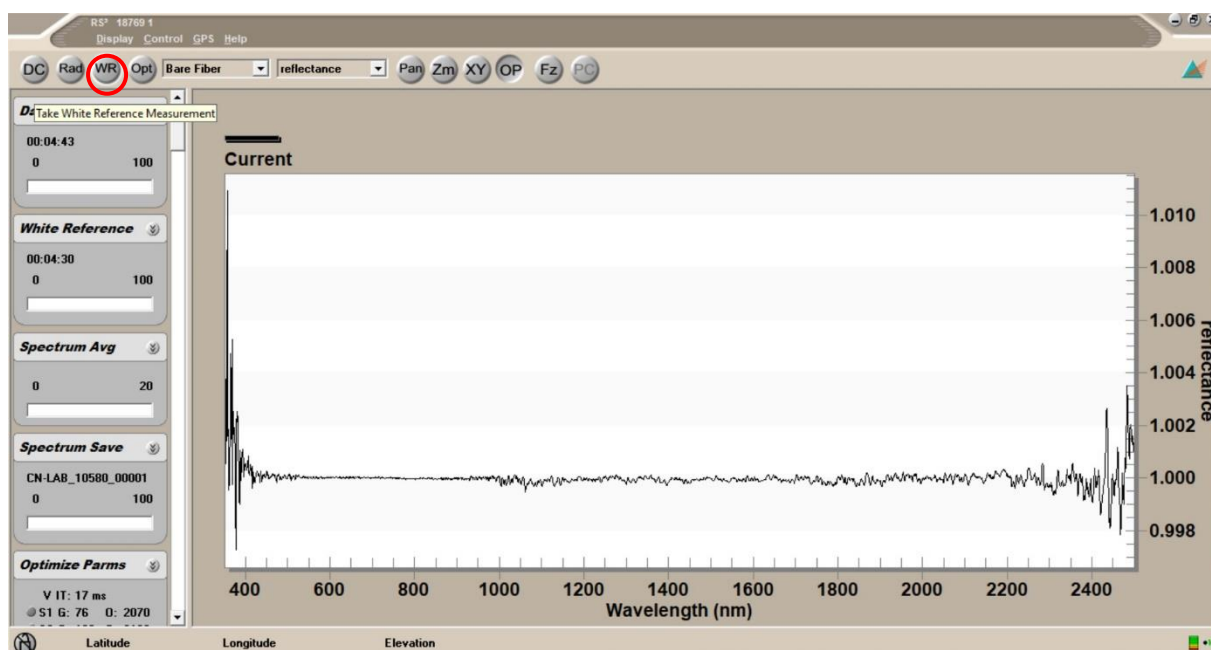
Após atribuir os nomes e o caminho onde os arquivos serão salvos, é necessário realizar a otimização do instrumento. Para isso, é realizada a calibração do equipamento com a utilização de uma placa branca de referência conhecida como *Spectralon*, que apresenta refletância uniforme ao longo de todos os comprimentos de onda e próxima a 100%. Os passos são ilustrados nas figuras 5 e 6.

Figura 5: Tela de Medição da Referência Branca



Fonte: Autores.

Figura 6: Tela de Medição da Referência Branca



Fonte: Autores.

Após a medição da reflectância da placa de referência, que representa de forma indireta a medida da irradiância E da fonte, visto que o material de composição da placa possibilita uma reflectância próxima a 100%, é possível iniciar a medida da radiância da amostra. A figura 7 ilustra o processo de realização da calibração com a utilização da placa *Spectralon*.

Figura 7: Realização da leitura da placa *Spectralon* no *Turntable*.



Fonte: Autores.

A partir desse ponto, são realizadas as leituras de cada amostra preparada para a composição da base de dados. Durante o processo, caso ocorra alguma interferência de luminosidade no ambiente, ou algo que possa comprometer a qualidade da leitura, é realizada novamente a calibração do instrumento através da placa *Spectralon*, de forma a garantir ao máximo possível a confiabilidade dos dados.

4 PROCESSAMENTO DOS DADOS NO VIEWSPEC PRO

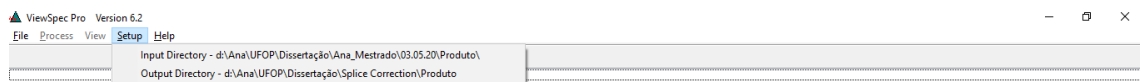
3.1 CORREÇÃO DAS AMOSTRAS

Após realizada a leitura dos espectros de todas as amostras pelo *software RS³* em conjunto com os equipamentos *Turntable* e *FieldSpec 4* é utilizado o *software ViewSpecPro* para realizar o pré processamento dos dados.

O primeiro passo é realizar a correção dos dados, com a criação do espectro contínuo para a ajuste dos *gaps* entre os sensores do *FieldSpec 4* que cobrem toda a faixa de faixa espectral entre 350 – 2500 nm.

No *ViewSpecPro* é necessário apontar a pasta onde estão os dados de origem dos arquivos com extensão *.asd* criados no *RS³* e também a nova pasta, onde serão salvos os arquivos com o espectro contínuo, conforme figura 8.

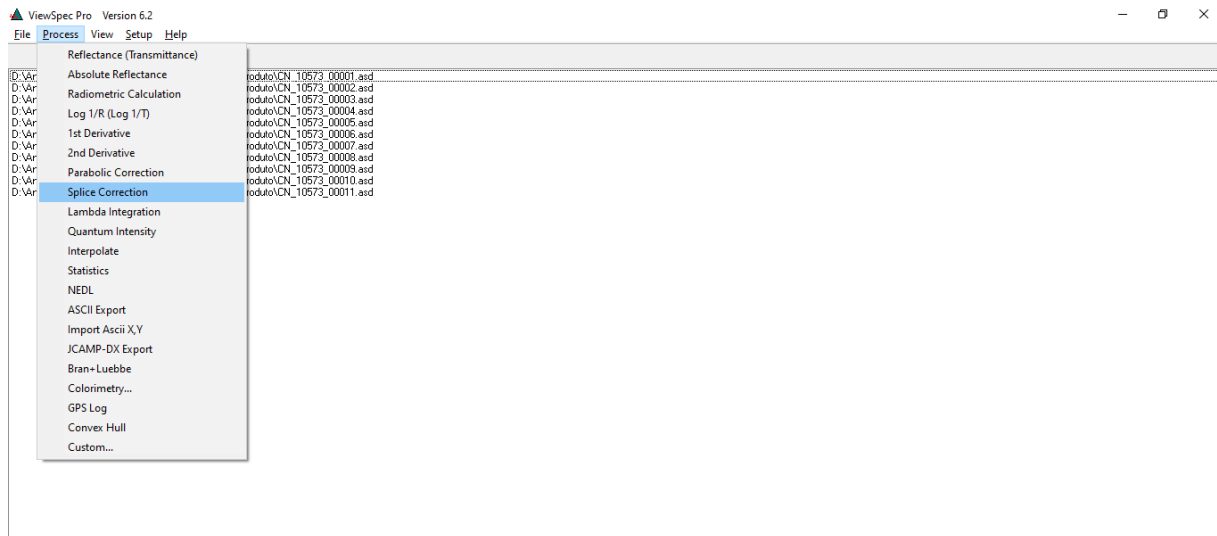
Figura 8: Tela do *software ViewSpecPro*



Fonte: Autores.

Para a realização da correção é apontado o local onde os arquivos de origem foram salvos através do menu “*File*” e na sequência “*Open*”. Após selecionar os arquivos a serem trabalhados, executar o procedimento no menu “*Process*” e “*Splice Correction*” conforme ilustrado na figura 9.

Figura 9: Ferramenta “*Splice Correction*”

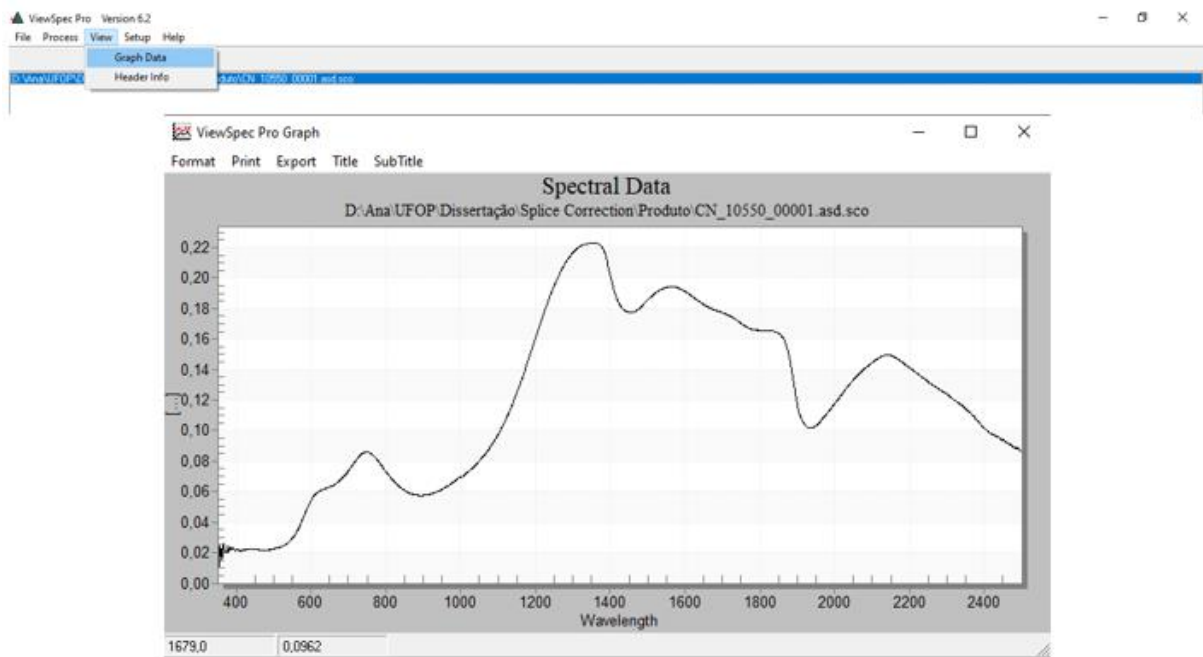


Fonte: Autores.

Na sequência será aberta a caixa de texto “*Splice Correct Gap*” onde foi mantido o valor padrão em 5. Após o processamento dos dados, é realizada a correção com a criação de novos arquivos com extensão “.sco”.

Finalizado o processo, é possível visualizar no *ViewSpecPro* o aspecto da curva espectral gerada por cada arquivo através no menu *View – Graph Data*, conforme exibido na figura 10.

Figura 10: Geração do gráfico da curva espectral da leitura da amostra.

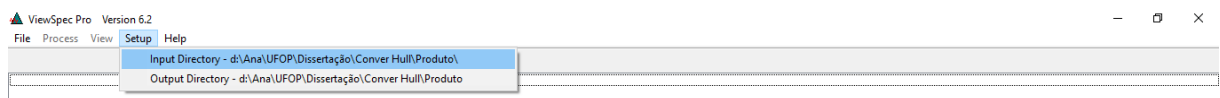


Fonte: Autores.

3.2 REMOÇÃO DO CONTÍNUO

Para criação dos dados com remoção do contínuo, foi utilizada a ferramenta “*Convex Hull*”. Inicialmente, no menu “*Setup*” foram direcionadas as pastas de origem dos dados, que no caso são os dados processados anteriormente pelo *Splice Correct Gap*” e a pasta onde serão salvos os arquivos com o contínuo removido, conforme exibido na figura 11.

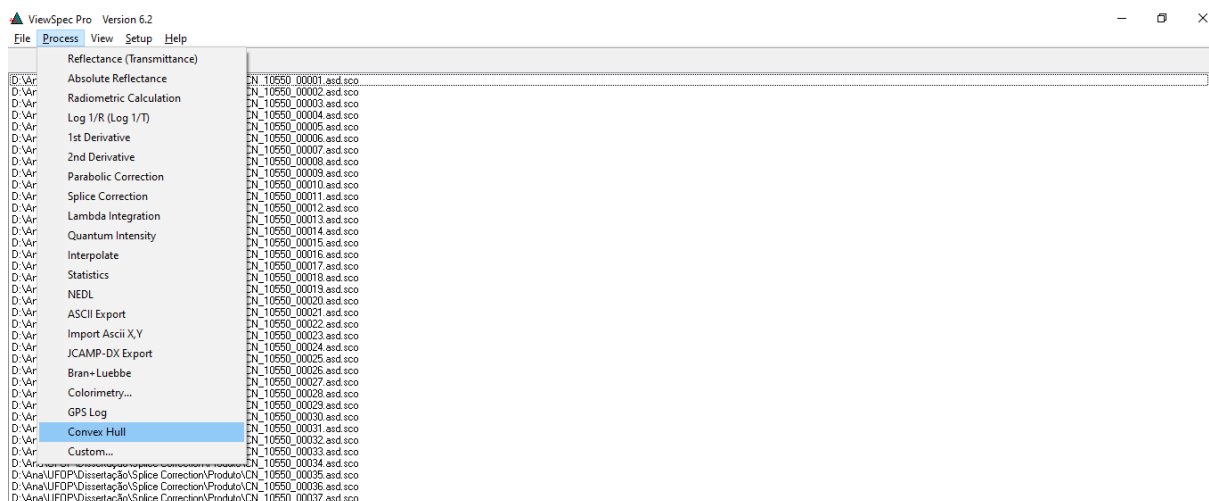
Figura 11: Seleção de dados do *software ViewSpecPro*



Fonte: Autores.

Após seleção dos dados de origem, no menu “*Process*” é aplicado a remoção do contínuo através do menu *Process - Convex Hull*, conforme exibido na figura 12.

Figura 12: Remoção do Contínuo

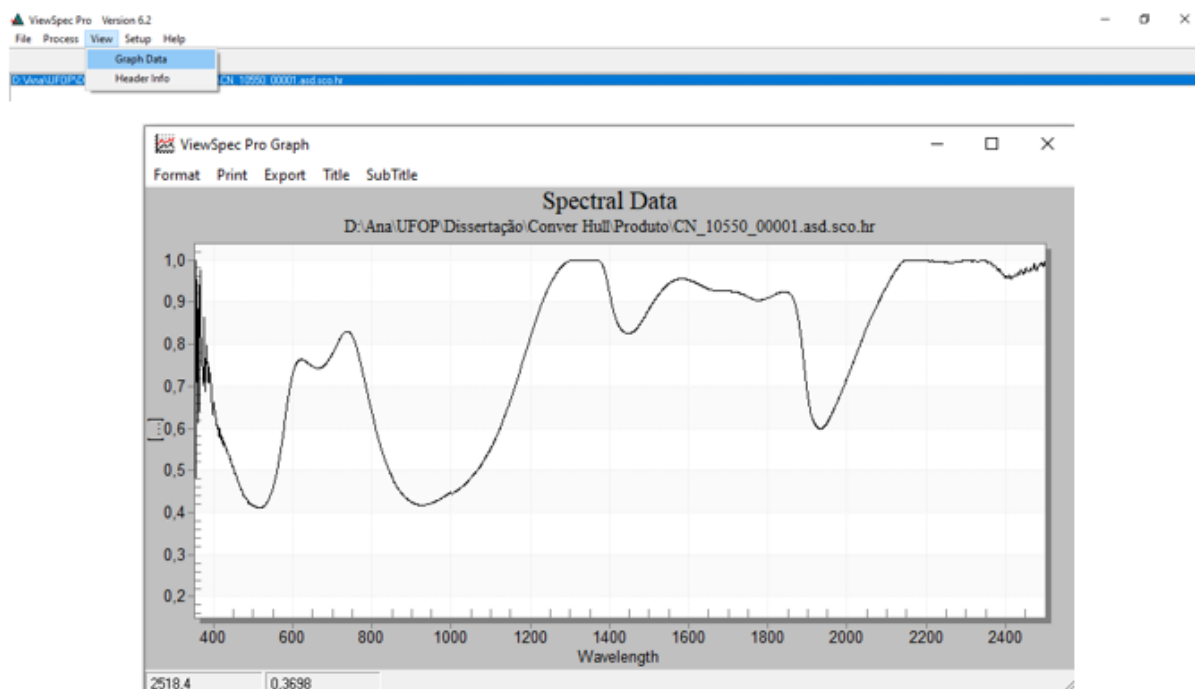


Fonte: Autores.

Foi criada uma base de dados com arquivos tipo “HR”, que representam o contínuo removido.

Finalizado o processo, é possível visualizar no *ViewSpecPro* o aspecto da curva espectral gerada com a remoção do contínuo de cada arquivo, através no menu *View – Graph Data*, conforme exibido na figura 13.

Figura 13: Geração do gráfico da curva espectral com contínuo removido.

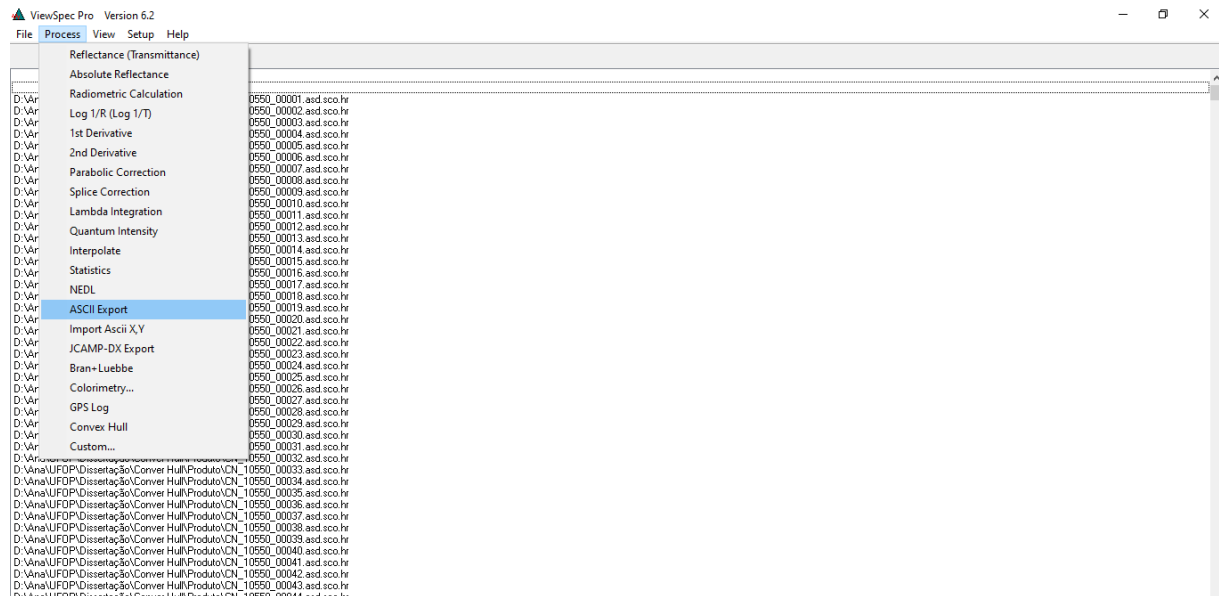


Fonte: Autores.

3.3 BASE DE DADOS ASCII

É necessário transformar as informações geradas no processo do contínuo removido em arquivos com extensão tipo ASCII, para que permita a leitura das informações pelo programa que processará os dados em *Machine Learning*. Esta conversão também é realizada no *ViewSpecPro* pelo menu *Process – ASCII Export*, conforme figura 14.

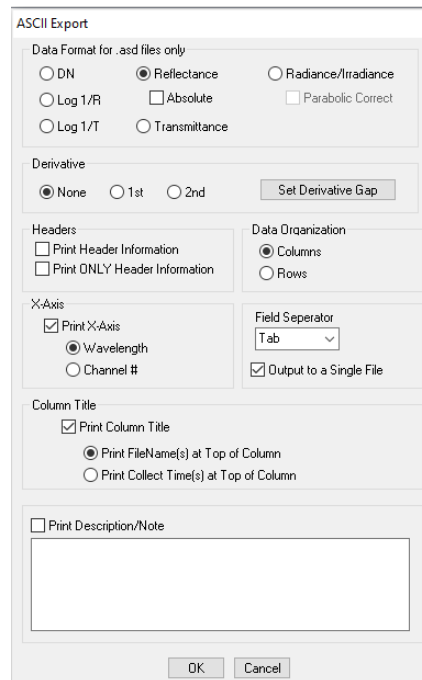
Figura 14: Menu *ASCII Export*



Fonte: Autores.

É importante se atentar em criar os dados com dados de reflectância e em um único arquivo, conforme exibido na figura 15 o detalhamento do menu *ASCII Export*.

Figura 15: Detalhe do menu *ASCII Export*

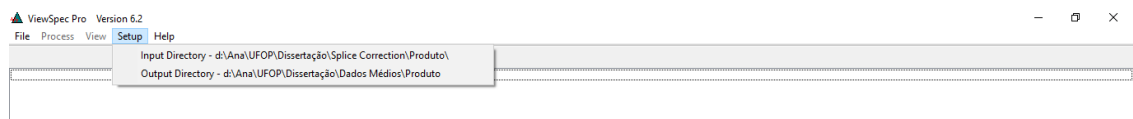


Fonte: Autores.

3.4 ARQUIVO COM A MÉDIA DAS LEITURAS

Foram criados arquivos com a média das leituras realizadas com os dados salvos pela ferramenta “*Splice Correction*”. Desta forma, foi gerado um arquivo único, com a média das 100 leituras colhidas para cada amostra e que será utilizada na criação da biblioteca espectral. Para tal, uma nova pasta foi criada para receber os dados, conforme exibido na figura 16.

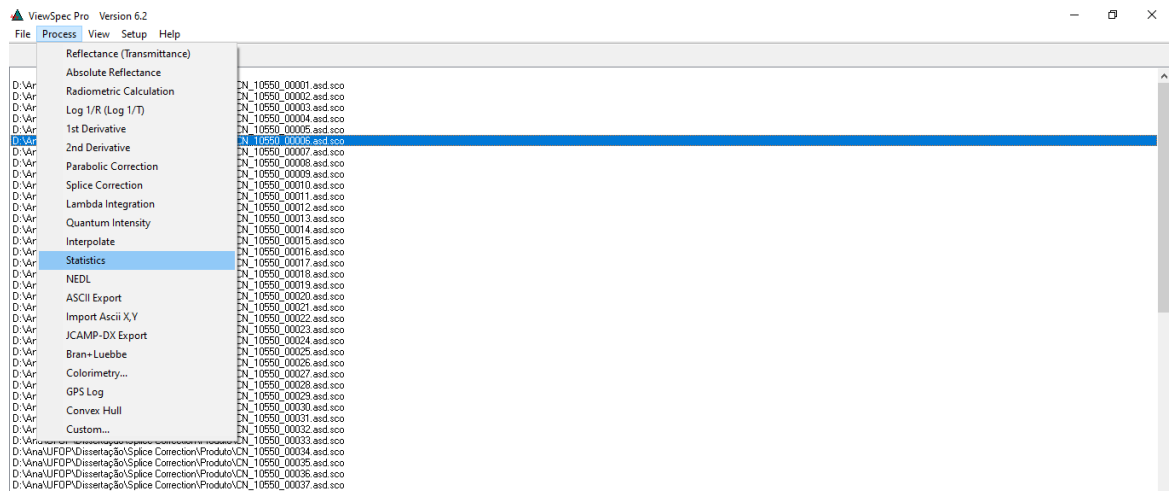
Figura 16: Seleção dos diretórios.



Fonte: Autores.

Após a seleção dos 100 arquivos de leitura referentes a amostra, no *ViewSpecPro* foi utilizado menu *Process – Statitics* para o cálculo das médias, conforme exibido na figura 17.

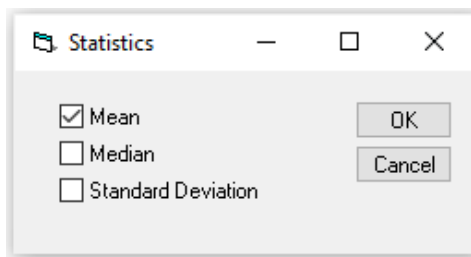
Figura 17: Seleção da função *Statistics*.



Fonte: Autores.

Dentro do menu *Statistics* é possível calcular a média, mediana e desvio padrão dos dados selecionados, como nosso objetivo é o cálculo da média das leituras, utilizamos a opção *Mean* como mostra a figura 18.

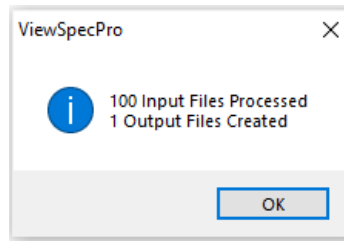
Figura 18: Função *Mean* do menu *Statistics*.



Fonte: Autores.

O *ViewSpecPro* criou 48 arquivos com extensão “.mn” através do processamento dos 100 dados de entrada referentes a cada amostra, ao final do processo é exibida uma janela de conclusão do processo, como a exibida na figura 19.

Figura 19: Término da criação do arquivo.



Fonte: Autores.

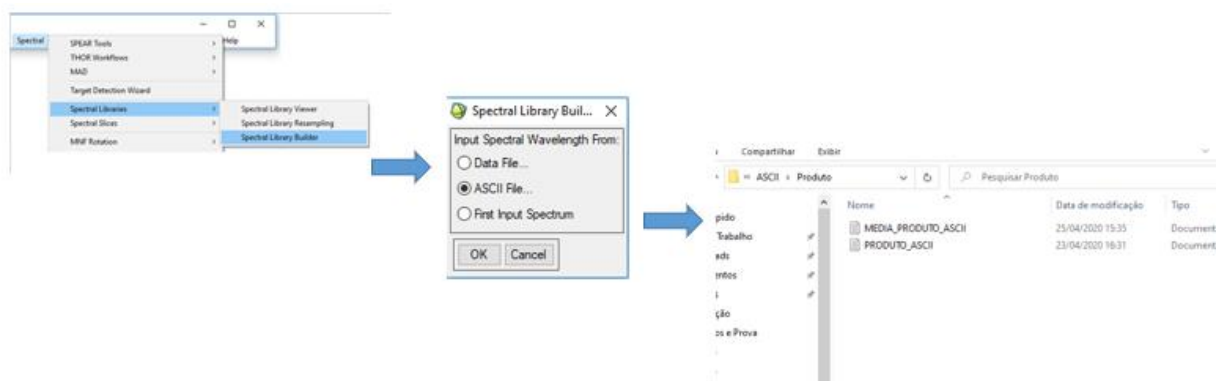
Como os dados médios serão utilizados para a criação da biblioteca espectral no *ENVI* é necessária a criação de um arquivo tipo ASCII, que possibilitará a leitura das informações. Da mesma forma que no procedimento anterior, esta conversão também é realizada no *ViewSpecPro* no menu *Process – ASCII Export*.

5 CRIAÇÃO DA BIBLIOTECA ESPECTRAL

Para a criação da biblioteca espectral foi utilizado o *software ENVI Classic*. As figuras abaixo sequenciadas de 18 a 20, mostram o passo a passo do processo. Foram utilizadas como base de dados a média das leituras espectrais de cada amostra de produto de sínther feed coletadas em Serra Sul exportadas através de um arquivo com extensão tipo ASCII.

No *ENVI Classic* através do menu *Spectral – Spectral Libraries – Spectral Library Builder* é selecionado o local onde estão salvos os arquivos de entrada utilizados para formação da biblioteca, conforme ilustrado na figura 20.

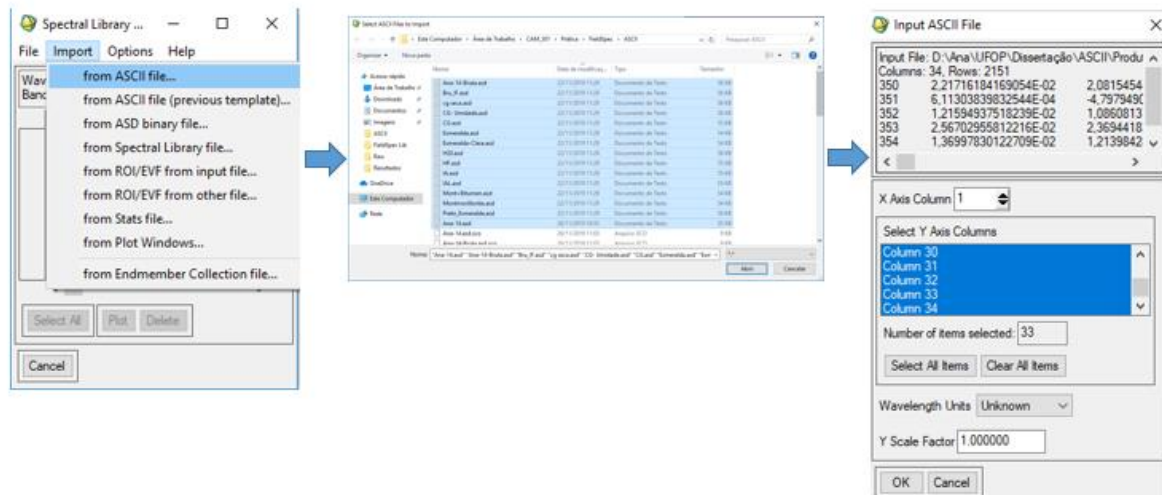
Figura 20: Passo 1 - Menu *Spectral Library Builder* seleção da entrada de dados.



Fonte: Autores.

Na sequência os dados são importados através da seleção do arquivo e processados pelo *ENVI* para criação da biblioteca, conforme exibido figura 21.

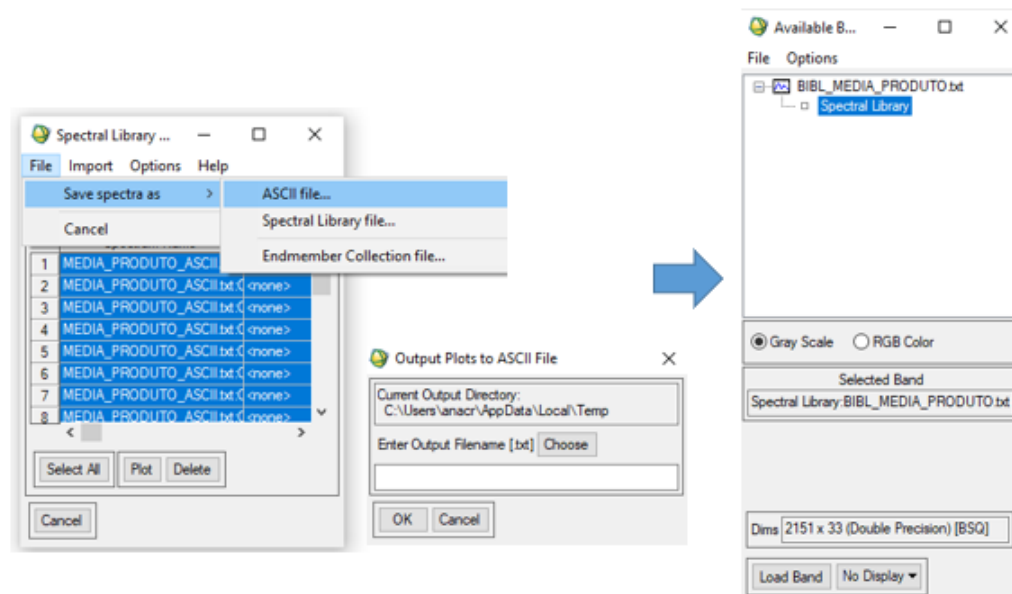
Figura 21: Passo 2 - Menu *Spectral Library Builder* para criação da biblioteca.



Fonte: Autores.

Por fim, a biblioteca pode ser salva tanto em formato ASCII ou através de um arquivo específico de biblioteca espectral no caminho onde for direcionado, conforme figura 22.

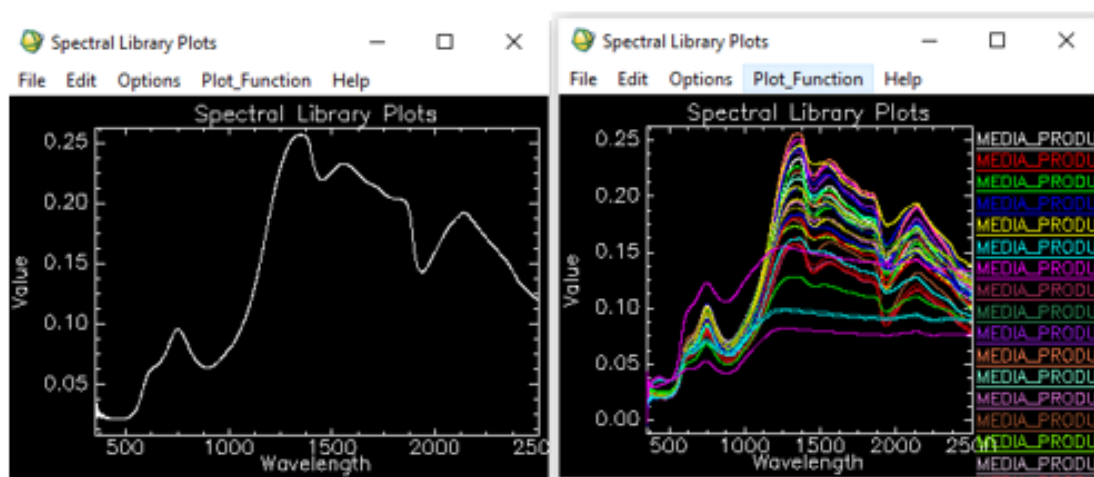
Figura 22: Passo 3 – Salvando a biblioteca criada.



Fonte: Autores.

A biblioteca pode ser visualizada em modo gráfico através do menu *Spectral – Spectral Libraries – Spectral Library Viewer*. É selecionada a biblioteca onde é possível plotar curvas individuais ou o conjunto de dados como um todo, como mostrado na figura 23.

Figura 23: Gráficos das curvas espectrais.



Fonte: Autores.