



Universidade do Estado do Rio de Janeiro

Centro de Tecnologia e Ciências

Instituto de Química

Raquel Escrivani Guedes

Modelagem da previsão do rendimento do bio-óleo com base nos parâmetros de processo de pirólise e composição centesimal, elementar e lignocelulósica da biomassa em diferentes reatores usados na indústria química

Rio de Janeiro

2024

Raquel Escrivani Guedes

Modelagem da previsão do rendimento do bio-óleo com base nos parâmetros de processo de pirólise e composição centesimal, elementar e lignocelulósica da biomassa em diferentes reatores usados na indústria química

Tese apresentada, como requisito final para a obtenção do título de Doutor, ao Programa de Pós-Graduação em Engenharia Química, da Universidade do Estado do Rio de Janeiro.

Área de Concentração: Processos Químicos, Petróleo e Meio Ambiente.

Orientadores: Prof. Dr. Aderval Severino Luna

Prof. Dr. Alexandre Rodrigues Torres

Rio de Janeiro

2024

CATALOGAÇÃO NA FONTE
UERJ / REDE SIRIUS / BIBLIOTECA CTC/Q

G924 Guedes, Raquel Escrivani.

Modelagem da previsão do rendimento do bio-óleo com base nos parâmetros de processo de pirólise e composição centesimal, elementar e lignocelulósica da biomassa em diferentes reatores usados na indústria química. – 2024.

238 f.

Orientador (a): Aderval Severino Luna
Alexandre Rodrigues Torres

Tese (Doutorado) – Universidade do Estado do Rio de Janeiro.
Instituto de Química.

1. Pirólise – Teses. 2. Bio-óleo – Teses. 3. Biomassa – Teses. I. Luna, Aderval Severino. II. Torres, Alexandre Rodrigues. III. Universidade do Estado do Rio de Janeiro. Instituto de Química. V. Título.

CDU 66.092-977

Autorizo, apenas para fins acadêmicos e científicos, a reprodução total ou parcial desta dissertação, desde que citada a fonte.

 Documento assinado digitalmente
RAQUEL ESCRIVANI GUEDES
Data: 17/06/2024 08:41:20-0300
Verifique em <https://validar.iti.gov.br>

17/06/2024

Assinatura

Data

Raquel Escrivani Guedes

Modelagem da previsão do rendimento do bio-óleo com base nos parâmetros de processo de pirólise e composição centesimal, elementar e lignocelulósica da biomassa em diferentes reatores usados na indústria química

Tese apresentada, como requisito final para a obtenção do título de Doutor, ao Programa de Pós-Graduação em Engenharia Química, da Universidade do Estado do Rio de Janeiro.
Área de Concentração: Processos Químicos, Petróleo e Meio Ambiente.

Aprovada em 24 abril de 2024

Banca Examinadora:

Documento assinado digitalmente

 **ADERVAL SEVERINO LUNA**
Data: 05/06/2024 21:13:47-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Aderval Severino Luna (Orientador)
Programa de Pós-graduação em Engenharia Química - UERJ

Documento assinado digitalmente

 **ALEXANDRE RODRIGUES TORRES**
Data: 05/06/2024 12:22:20-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Alexandre Rodrigues Torres (Orientador)
Faculdade de Tecnologia - UERJ

Documento assinado digitalmente

 **ARGIMIRO RESENDE SECCHI**
Data: 02/06/2024 19:22:02-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Argimiro Resende Secchi
Programa de Engenharia Química - COPPE/UFRJ

Documento assinado digitalmente

 **MARISA FERNANDES MENDES**
Data: 31/05/2024 11:25:16-0300
Verifique em <https://validar.iti.gov.br>

Prof.^a Dr.^a. Marisa Fernandes Mendes
Departamento de Engenharia Química - UFRRJ

Documento assinado digitalmente

 **ERIKA CHRISTINA ASHTON NUNES CHRISMAN**
Data: 29/05/2024 12:21:24-0300
Verifique em <https://validar.iti.gov.br>

Prof.^a Dr.^a. Erika Christina Asthon Nunes Chrisman
Escola de Química - UFRJ

Documento assinado digitalmente

 **JOSE LICARION PINTO SEGUNDO NETO**
Data: 06/06/2024 13:42:07-0300
Verifique em <https://validar.iti.gov.br>

José Licarion Pinto Segundo Neto
Programa de Pós-graduação em Engenharia Química – UERJ

Rio de Janeiro

2024

AGRADECIMENTOS

A Deus por ter me dado a oportunidade de realizar este trabalho e por ter me dado força nos momentos de dúvida e incerteza.

À minha família, que sempre esteve ao meu lado. Foi uma jornada difícil, mas seu apoio, amor e incentivo, foram essenciais para que eu pudesse superar todos os desafios.

Ao meu marido, Bruno, pelo apoio, carinho, companheirismo e incentivo durante todo o doutorado.

Aos meus orientadores, Alexandre R. Torres e Aderval S. Luna, pela amizade, disponibilidade, paciência e pela confiança e dedicação ao me orientar por novos caminhos para que esse trabalho fosse concluído.

Aos meus amigos da FAT, em especial ao José Luís e Camila, pelas palavras de incentivo e por tornarem os dias no laboratório mais leves.

Ao corpo docente e aos funcionários do programa de pós graduação em Engenharia Química da UERJ e da FAT-UERJ.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de doutorado.

RESUMO

GUEDES, Raquel Escrivani. **Modelagem da previsão do rendimento do bio-óleo com base nos parâmetros de processo de pirólise e composição centesimal, elementar e lignocelulósica da biomassa em diferentes reatores usados na indústria química.** 2024. 236 f. Tese (Doutorado em Engenharia Química) – Instituto de Química, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2024.

A pirólise é um processo com alto potencial de conversão de biomassa e o bio-óleo, produto líquido da pirólise, que ganha destaque por possuir inúmeras aplicações. O rendimento de bio-óleo é afetado pela composição da biomassa e parâmetros de processo. Um desafio da pirólise é desenvolver técnicas para interpretar o grande volume de dados experimentais disponíveis que analisam a influência das variáveis no rendimento dos produtos obtidos. Esta pesquisa tem como objetivo aplicar diferentes métodos de regressão multivariada, linear e não linear, para prever a produção de bio-óleo obtido em pirólise de biomassa realizada em reatores operando em modo contínuo e em batelada. Para a modelagem, desenvolveu-se um banco de dados contendo resultados experimentais de 220 artigos disponíveis na literatura. A modelagem foi realizada com oito diferentes combinações de reatores e variáveis. Os parâmetros do processo analisados foram a taxa de aquecimento, temperatura, tempo de reação, tamanho de partícula e tempo de residência do vapor. As variáveis relacionadas à composição das biomassas incluíram a composição elementar, composição centesimal e composição lignocelulósica. Quatorze modelos de regressão foram testados. Para o reator em batelada, o modelo *Cubist* forneceu o melhor resultado de previsão de bio-óleo com RMSE de 0,92%, MAE de 0,73% e R^2 de 0,99. Este melhor resultado foi obtido utilizando-se, como preditores, todas as variáveis estudadas e somente dados com observações completas. Para o reator contínuo, o melhor resultado foi obtido aplicando-se o modelo Árvores Extremamente Aleatórias também ao conjunto completo de preditores com observações completas. As métricas de desempenho obtidas foram um RMSE de 2,15%, MAE de 1,75% e R^2 de 0,96. Ambos os modelos mostraram um excelente desempenho para previsão de rendimento de bio-óleo para reatores em batelada e contínuo. A modelagem também foi realizada utilizando-se a imputação de dados faltantes por k-NN e imputação múltipla, os resultados mostraram que não houve ganho de eficiência de previsão dos modelos utilizando os conjuntos de dados contendo os valores imputados. O método de enxame de partículas foi utilizado, nos melhores modelos obtidos na modelagem, para o estudo de otimização dos parâmetros de processo da pirólise da casca de arroz. Para o reator batelada o máximo de rendimento de bio-óleo previsto foi de 52,9% em temperatura de 605 °C, tamanho de partícula de 7,4 mm, tempo de residência do vapor de 1400 s e taxa de aquecimento de 300 °C/min. Para reator contínuo o rendimento máximo previsto pelo modelo foi de 52,1% em temperatura de 513 °C, tamanho de partícula de 0,41 mm e tempo de residência do vapor de 0,6 s.

Palavras-chave: pirólise; rendimento de bio-óleo; modelos de regressão; reator batelada; reator contínuo.

ABSTRACT

GUEDES, Raquel Escrivani. **Modeling bio-oil yield prediction based on pyrolysis process parameters and biomass's proximate, elemental, and lignocellulosic composition in different reactors used in the chemical industry.** 2024. 236 f. Tese (Doutorado em Engenharia Química) – Instituto de Química, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2024.

The bio-oil yield, the liquid product of pyrolysis, is affected by biomass composition and process parameters. This research applies different multivariate regression methods, linear and non-linear, to predict bio-oil production, exploring various combinations of predictors in reactors operating in continuous and batch mode in biomass pyrolysis. A database with experimental data from 220 articles in the literature was developed for the modeling. Modeling was performed on eight case studies with different variables. Process parameters analyzed were heating rate, temperature, reaction time, particle size, and vapor residence time. The variables related to the composition of the analyzed biomass were elemental composition, proximate composition, and concentration of lignocellulosic materials. Fourteen multivariate regression models were applied to model the bio-oil yield. The Cubist model provided the best bio-oil prediction results for the batch reactor with RMSE of 0.92%, MAE of 0.73%, and R^2 of 0.99. This best result was obtained using, as predictors, all the variables studied and only data with complete observations. The best result for the continuous reactor was obtained with the Extremely Randomized Trees model and a full set of predictors with complete observations. The performance metrics obtained were an RMSE of 2.15%, MAE of 1.75% and R^2 of 0.96. Both models showed excellent performance for bio-oil yield prediction for batch and continuous reactors. The modeling was also carried out using the imputation of missing data by k-NN and multiple imputation; the results showed no gain in the prediction efficiency of the models using the data sets containing the imputed values. The particle swarm method was used, in the best models obtained in modeling, to study the optimization of process parameters for rice husk pyrolysis. For the batch reactor, the maximum predicted bio-oil yield was 52.9% at a temperature of 605 °C, particle size of 7.4 mm, vapor residence time of 1400s and heating rate of 300 °C/min. For a continuous reactor, the maximum yield predicted by the model was 52.1% at a temperature of 513 °C, particle size of 0.41 mm and vapor residence time of 0.6 s.

Keywords: pyrolysis; bio-oil yield; regression models; batch reactor; continuous reactor.

LISTA DE FIGURAS

Figura 1 - Processos de conversão e seus produtos	23
Figura 2 - Esquema do processo de pirólise de biomassa	25
Figura 3 - Aplicações dos produtos da pirólise	26
Figura 4 - Fatores que influenciam no processo.....	30
Figura 5 - Produtos obtidos da degradação da lignina.....	43
Figura 6 - Produtos obtidos da degradação da celulose	43
Figura 7 - Produtos obtidos da degradação da hemicelulose.....	44
Figura 8 - Estrutura simples da ANN	54
Figura 9 - SVM para o problema de regressão linear.....	57
Figura 10 - Exemplo de árvore de regressão com 10 nós terminais.....	61
Figura 11 - Metodologia de imputação múltipla	76
Figura 12 - Distribuição das variáveis por rendimento de bio-óleo	96
Figura 13 - Matriz de correlação entre variáveis de processo e a produção de bio-óleo.....	98
Figura 14 - Matriz de correlação entre composição e produção de bio-óleo	100
Figura 15 - Matriz de correlação entre composição lignocelulósica e rendimento.	101
Figura 16 - Rendimento de bio-óleo em diferentes composições da biomassa.....	101
Figura 17 - Tipos de reatores de pirólise presentes no banco de dados.....	103
Figura 18 - Rendimento de bio-óleo por tipo de reator e regime de operação.	103
Figura 19 - Distribuição das variáveis por regime de operação dos reatores.	105
Figura 20 - Distribuição dos dados antes e após o pré-processamento do caso I.....	107
Figura 21 - RMSE para os modelos dos casos I, II, III e IV	109
Figura 22 - MAE para os modelos dos casos I, II, III e IV	109
Figura 23 - R^2 para os modelos dos casos I, II, III e IV	110

Figura 24 - Importância das variáveis para os modelos obtidos nos casos I, II, III e IV	111
Figura 25 - Valores observados e preditos de rendimento de bio-óleo para os melhores modelos dos casos I, II, III e IV	113
Figura 26 - RMSE para os modelos dos casos V, VI, VII e VIII.....	117
Figura 27 - MAE para os modelos dos casos V, VI, VII e VIII.....	117
Figura 28 - R ² para os modelos dos casos V, VI, VII e VIII.....	118
Figura 29 - Valores observados e preditos de rendimento de bio-óleo para os melhores modelos dos casos V, VI, VII e VIII	119
Figura 30 - Importância das variáveis para os melhores modelos obtidos nos casos V, VI, VII e VIII.....	120
Figura 31 - Distribuição dos dados observados e imputados por k-NN- Caso I.....	124
Figura 32 - Densidade dos dados observados e imputados por k-NN - Caso I.....	124
Figura 33 - Distribuição dos dados observados e imputados por RF- Caso I	125
Figura 34 - Densidade dos dados observados e imputados por RF - Caso I.....	126
Figura 35 - Distribuição dos dados observados e imputados por Midastouch - Caso I	126
Figura 36 - Densidade dos dados observados e imputados por Midastouch - Caso I.....	127
Figura 37 - Densidade dos dados observados e imputados por Midastouch - Caso III.....	130
Figura 38 - Densidade dos dados observados e imputados por k-NN - Caso VII.....	136
Figura 39 - Densidade dos dados observados e imputados por RF - Caso VII.....	137
Figura 40 - Densidade dos dados observados e imputados por Midastouch - Caso VII.....	137
Figura 41 - Superfície de resposta para otimização do rendimento de bio-óleo no caso IV..	141
Figura 42 - Superfície de resposta para otimização do rendimento de bio-óleo no caso VIII	142

LISTA DE TABELAS

Tabela 1 - Composição de algumas biomassas	22
Tabela 2 - Diferentes tipos de pirólise.....	26
Tabela 3 - Comparação entre as propriedades de bio-óleo e combustível fóssil.....	28
Tabela 4 - Amostra do banco de dados.....	80
Tabela 5 - Parâmetros de ajuste, método usado e os pacotes necessários.....	88
Tabela 6 - Pacotes e funções do pré-processamento e modelagem.....	92
Tabela 7 - Variáveis analisadas nos estudos de caso.....	93
Tabela 8 - Variáveis analisadas nas pesquisas presentes no banco de dados	94
Tabela 9 - Resumo das variáveis de processo e composição da biomassa.....	95
Tabela 10 - Número de observações, variáveis, biomassas e outliers nos estudos	106
Tabela 11 - Resultados da regressão realizada por LazyPredict para o caso I.....	108
Tabela 12 - Ferramentas de diagnóstico para o melhor modelo do reator em batelada	115
Tabela 13 - Ferramentas de diagnóstico para o melhor modelo em reator contínuo	122
Tabela 14 - Dados faltantes em cada variável	123
Tabela 15 - Ferramentas de diagnóstico da modelagem com dados imputados -Caso I.....	128
Tabela 16 - Ferramentas de diagnóstico da modelagem com dados imputados-Caso II.....	129
Tabela 17- Ferramentas de diagnóstico da modelagem com dados imputados-Caso III	131
Tabela 18 - Ferramentas de diagnóstico da modelagem com dados imputados-Caso IV	132
Tabela 19 - Ferramentas de diagnóstico para os melhores modelos obtidos em dados completos e imputados em reator batelada.....	132
Tabela 20 - Ferramentas de diagnóstico da modelagem com dados imputados-Caso V	133
Tabela 21 - Ferramentas de diagnóstico da modelagem com dados imputados-Caso VI.....	134
Tabela 22 - Ferramentas de diagnóstico da modelagem com dados imputados-Caso VII.....	135
Tabela 23 - Ferramentas de diagnóstico da modelagem com dados imputados-Caso VIII ...	138

Tabela 24 - Ferramentas de diagnóstico para os melhores modelos obtidos em dados completos e imputados em reator contínuo	139
Tabela 25 - Composição média da casca de arroz utilizada na otimização.....	140
Tabela 26 - Faixa de otimização e melhores parâmetros para otimização do caso IV.....	140
Tabela 27 - Faixa de otimização e melhores parâmetros para otimização do caso VIII.....	141

LISTA DE ABREVIATURAS E SIGLAS

ANN	Rede Neural Artificial
CART	Árvores de Regressão
ExtraTree	Árvores Extremamente Aleatórias
GBM	Gradient Boosting Machine
k-NN	k-vizinhos mais próximos
KS	Kennard-Stone
LASSO	Least Absolute Shrinkage and Selection
MAE	Erro médio de estimativa
MAR	Dados faltantes aleatórios
MARS	Multivariate Adaptive Regression Splines
MCAR	Dados faltantes completamente aleatórios
MLR	Regressão Linear Múltipla
NMAR	Dados faltantes não aleatórios
OLS	Mínimos Quadrados Ordinários
PMM	Correspondência Média Preditiva
PSO	Otimização por enxame de partícula
R^2	Coefficiente de determinação
Rede	Rede Elástica
RF	Floresta Aleatória
RMSE	Raiz do erro médio quadrático
SSE	Soma dos Quadrados dos Resíduos
SVM	Máquina de Vetores de Suporte
SVM/Br	Máquina de Vetores de Suporte com função de base radial
SVM/Pl	Máquina de Vetores de Suporte com função polinomial

SUMÁRIO

	INTRODUÇÃO	16
1	OBJETIVO	19
1.1	Objetivos específicos	19
2	REVISÃO BIBLIOGRÁFICA	21
2.1	Biomassa	21
2.2	Processos de conversão	23
2.3	Pirólise	25
2.4	Bio-óleo	28
2.5	Fatores que influenciam na pirólise	30
2.5.1	<u>Temperatura</u>	31
2.5.2	<u>Tamanho da partícula</u>	33
2.5.3	<u>Taxa de aquecimento</u>	35
2.5.4	<u>Tempo de residência do vapor</u>	36
2.5.5	<u>Tempo de reação</u>	38
2.5.6	<u>Composição da biomassa</u>	39
2.5.7	<u>Interação entre as variáveis</u>	45
3	MODELAGEM	48
3.1	Quimiometria	48
3.1.1	<u>Modelos Lineares</u>	49
3.1.1.1	Regressão linear múltipla.....	49
3.1.1.2	Ridge	51
3.1.1.3	LASSO	52
3.1.1.4	Rede Elástica.....	53
3.1.2	<u>Modelos não lineares</u>	53
3.1.2.1	Redes Neurais Artificiais	53
3.1.2.2	Spline de Regressão Adaptativa Multivariada.....	54
3.1.2.3	k-Vizinhos mais próximos	56
3.1.2.4	Máquina de Vetores de Suporte	57
3.1.2.5	Regras e árvores de decisão	60
3.1.2.5.1	Árvores de decisão	62
3.1.2.5.2	Cubist	63

3.1.2.5.3	Floresta Aleatória.....	65
3.1.2.5.4	Árvores Extremamente Aleatórias	66
3.1.2.5.5	Gradient Boosting Machine	67
3.2	Estudos de modelagem do processo de pirólise.....	68
3.3	Imputação	73
3.3.1	<u>Imputação por k-vizinhos mais próximos.....</u>	77
3.3.2	<u>Imputação por Floresta Aleatória.....</u>	77
3.3.3	<u>Imputação por Midastouch.....</u>	78
4	MATERIAIS E MÉTODOS	79
4.1	Descrição do banco de dados utilizado.....	79
4.2	Análise exploratória dos dados	82
4.2.1	<u>Valores faltantes.....</u>	83
4.2.2	<u>Seleção de preditores</u>	83
4.2.3	<u>Teste de normalidade</u>	84
4.2.4	<u>Teste de Kruskal-Wallis.....</u>	85
4.3	Pré-processamento e transformação dos dados	85
4.4	Algoritmo de seleção de amostras	86
4.5	Aplicação dos modelos	87
4.6	Imputação	88
4.7	Ferramentas de Diagnóstico.....	89
4.8	Otimização	90
4.9	Software e pacotes.....	91
4.10	Descrição dos estudos de caso de modelagem.....	92
5	RESULTADOS E DISCUSSÃO	94
5.1	Análise exploratória dos dados	94
5.2	Modelagem de dados completos.....	105
5.2.1	<u>Estudo de caso I</u>	106
5.2.2	<u>Estudo de caso II.....</u>	110
5.2.3	<u>Estudo de caso III.....</u>	113
5.2.4	<u>Estudo de caso IV</u>	114
5.2.5	<u>Estudo de caso V.....</u>	116
5.2.6	<u>Estudo de caso VI</u>	118
5.2.7	<u>Estudo de caso VII.....</u>	120

5.2.8	<u>Estudo de caso VIII</u>	121
5.3	Modelagem de dados imputados	122
5.3.1	<u>Estudo de caso I</u>	123
5.3.2	<u>Estudo de caso II</u>	128
5.3.3	<u>Estudo de caso III</u>	129
5.3.4	<u>Estudo de caso IV</u>	131
5.3.5	<u>Estudo de caso V</u>	133
5.3.6	<u>Estudo de caso VI</u>	134
5.3.7	<u>Estudo de caso VII</u>	135
5.3.8	<u>Estudo de caso VIII</u>	138
5.4	Otimização	139
	CONCLUSÃO	143
	REFERÊNCIAS	145
	ANEXO A - Biomassas presentes no banco de dados e o número de observações.....	162
	ANEXO B – Referências bibliográficas dos estudos presentes no banco de dados.....	174
	ANEXO C – Valores de lambda da transformação box-cox para os casos de I a VIII.....	189
	ANEXO D – Resultado do lazyprediction para os dados de teste dos casos de I a VIII.....	190
	ANEXO E - gráficos de distribuição e densidade dos valores imputados e observados para os estudos de caso I a VIII.....	199

INTRODUÇÃO

A demanda por energia está crescendo a cada dia juntamente com a rápida expansão da população e o desenvolvimento econômico (Sharma; Sheth, 2015). Devido ao impacto negativo do uso de combustíveis fósseis no meio ambiente, a busca por recursos renováveis aumenta em todo o mundo (Goyal; Seal; Saxena, 2008). Neste cenário, a biomassa se apresenta como uma das melhores soluções para a substituição de recursos fósseis em diversas aplicações como produção de calor, energia e combustíveis para transporte, produção de produtos químicos e biomateriais (Bridgwater, 2003).

O desenvolvimento de novas tecnologias de conversão está entre as principais vantagens da utilização da biomassa como fonte de energia. Além disso, a produção de biomassa pode gerar empregos, contribuir para a recuperação de terras degradadas e aumentar a biodiversidade. Outra vantagem é que o consumo de resíduos agrícolas para a produção de energia reduz os danos ambientais causados pela agroindústria (Gerçel, 2002; Mckendry, 2002).

A pirólise, decomposição térmica que ocorre na ausência de oxigênio, tem sido aceita como um método potencial para conversão de biomassa em produtos de melhor qualidade e conteúdo energético devido ao custo e facilidade de operação (Bridgwater, 2003). Nesse contexto, o bio-óleo de pirólise ganha destaque por ser um produto que possui inúmeras aplicações.

O bio-óleo, líquido produzido na pirólise, é composto por uma mistura complexa de substâncias orgânicas com inúmeras aplicações, sendo utilizado diretamente como combustível de caldeiras e fornos ou processado para produzir combustíveis e outras substâncias de interesse econômico (Bennett; Helle; Duff, 2009; Bridgwater, 2012; Ikura; Stanciulescu; Hogan, 2003).

Nos últimos anos, muitas pesquisas envolvendo o estudo da pirólise da biomassa foram realizadas na tentativa de otimizar o processo de obtenção de bio-óleo (Guedes; Luna; Torres, 2018). Os estudos mostram que os parâmetros do processo e a composição da biomassa afetam fortemente o rendimento e a composição dos produtos de pirólise, isso ocorre devido às várias reações químicas que são favorecidas em cada condição.

Apesar de um grande número de pesquisas avaliar as influências das variáveis no processo de pirólise de biomassa, os dados obtidos nos diferentes estudos são frequentemente

conflitantes, não permitindo determinar de forma clara seus efeitos. Assim, um desafio imposto pela pirólise é o desenvolvimento de técnicas que ajudem a interpretar os dados disponíveis para o aprimoramento do processo.

Paralelamente, a modelagem é uma ferramenta amplamente usada na análise multivariada de dados. A regressão simula relacionamentos entre os dados de uma variável desejada e suas variáveis preditoras, incluídas em um sistema. Devido à sua importância, uma grande variedade de modelos de regressão tem sido proposta na literatura. Cada um deles tem diferentes suposições sobre a natureza das variáveis e suas relações. O método de regressão adequado pode ser selecionado usando conhecimento prévio sobre a relação da resposta e seus preditores. No entanto, tais informações geralmente são desconhecidas e a escolha do modelo é obtida de forma indireta (Rendall; Pereira; Reis, 2016).

Os modelos estatísticos ajudam a avaliar um resultado sem realizar experimentos adicionais e, portanto, podem ampliar o conhecimento para o processo de pirólise usando os dados já disponíveis na literatura. Cao, Xin e Yuan (2016) usaram o modelo de Redes Neurais Artificiais (ANN) e o modelo de Máquina de Vetores de Suporte (SVM) ajustado por mínimos quadrados para prever a produção de carvão obtido da pirólise de esterco bovino. Sun et al. (2016) estudaram o modelo ANN para simular e prever a seletividade e o rendimento dos produtos gasosos da pirólise da serragem de pinus. Os efeitos da temperatura operacional, tamanho de partícula de biomassa e da velocidade espacial foram investigados. Merdun e Sezgin (2018) utilizaram ANN para modelar os rendimentos dos produtos de pirólise de biomassa analisando as variáveis de processo presentes no reator em batelada. Chen et al. (2018) aplicaram ANN e SVM para modelar a produção de bio-óleo em reatores contínuos. Zhu, Li e Wang (2019) sugeriram que o modelo Floresta Aleatória poderia prever, com precisão, o rendimento e o conteúdo de carbono no carvão de acordo com as características da biomassa e as condições de pirólise. Tang et al. (2020) aplicaram os métodos de Regressão Linear Múltipla e Floresta Aleatória para prever o rendimento e os teores de hidrogênio no bio-óleo com base nas composições de biomassa e nas condições do processo em um reator em batelada. Os resultados desses estudos fornecem uma boa referência para uma expansão do uso de modelagem no processo de pirólise, visando uma análise ampla do processo e a obtenção de modelos com melhor exatidão de predição.

Neste contexto, esta pesquisa foca na aplicação de diferentes modelos de regressão multivariada, linear e não linear, para prever a produção de bio-óleo, explorando diferentes combinações de preditores em reator de pirólise operando de forma contínua e em batelada.

Para dar suporte à modelagem, um banco com dados experimentais coletados de 220 artigos disponíveis na literatura foi desenvolvido. A modelagem foi realizada em oito estudos de caso. Os parâmetros do processo analisados foram temperatura, tempo de residência de vapor, tamanho de partícula, taxa de aquecimento e tempo de reação. As variáveis relacionadas à composição da biomassa analisada foram composição elementar (carbono, oxigênio, hidrogênio e nitrogênio), composição centesimal (umidade, cinzas, carbono fixo e volátil) e composição lignocelulósica (celulose, lignina e hemicelulose).

1 OBJETIVO

Há um grande volume de resíduos de biomassa, de diferentes setores industriais, que podem ser uma importante fonte de recurso renovável para a substituição dos recursos fósseis. Diversos estudos experimentais mostram a alta capacidade do processo de pirólise para a produção de bio-óleo a partir de biomassa, produto este que pode ser utilizado como fonte de energia. Os estudos mostram que o processo de pirólise é fortemente dependente das condições de processo e da matéria prima utilizada. Porém, apesar do grande número de dados disponíveis, a relação entre as variáveis que afetam a pirólise e o rendimento de produto obtido não é clara, devido à complexidade do processo e às interações entre as variáveis. Assim, é de grande importância, para aumentar a eficiência do processo, encontrar as melhores condições operacionais para a produção de bio-óleo. Neste contexto, esta pesquisa tem como objetivo a aplicação de ferramentas quimiométricas de regressão multivariada em dados de pirólise de biomassa, provenientes da literatura, com o objetivo de desenvolver modelos, com alta capacidade preditiva, para previsão de rendimento do bio-óleo partir das condições operacionais que afetam o sistema e da composição da biomassa. O modelo é uma ferramenta muito útil para prever o rendimento de bio-óleo que será obtido nas condições e com a biomassa que se deseja utilizar ou para se obter as melhores condições para aumentar a eficiência da pirólise sem a necessidade de realização de trabalho experimental muitas vezes demorado e caro.

1.1 Objetivos específicos

- ✓ Construir e organizar um banco de dados, reunindo dados experimentais disponíveis na literatura, com informações de processo de pirólise em diferentes biomassas nas mais diversas condições operacionais e configurações de reatores;

- ✓ Realizar análises de correlação entre as variáveis de maior importância na pirólise para uma melhor compreensão de como a composição da matéria-prima e as variáveis de processo interferem no rendimento do bio-óleo;

- ✓ Investigar diferentes modelos de regressão para verificar a possibilidade de prever o rendimento de bio-óleo sem a necessidade de realização de experimentos;
- ✓ Estudar a combinação de diferentes variáveis de processo e de composição da biomassa para verificar a combinação apropriada para a previsão, ou seja, a combinação de preditores que levam a um maior ajuste do modelo aos dados;
- ✓ Realizar análise de imputação com diferentes metodologias, devido a presença de dados faltantes no banco, para verificar se a inclusão de dados imputados na modelagem melhora os resultados de performance dos modelos desenvolvidos;
- ✓ Realizar um estudo de otimização, utilizando os modelos com maior capacidade preditiva, visando comparar os resultados de rendimento previsto de bio-óleo em reator batelada e contínuo, utilizando-se uma biomassa de mesma composição.

2 REVISÃO BIBLIOGRÁFICA

2.1 Biomassa

Biomassa é o termo usado para o material orgânico que se origina a partir de plantas e animais (Mckendry, 2002). Para Goyal, Seal e Saxena (2008), as fontes de biomassa incluem madeira e seus resíduos, culturas agrícolas e seus subprodutos, resíduos sólidos urbanos, resíduos de processamento de alimentos, plantas aquáticas e algas. A utilização da biomassa como fonte de energia depende principalmente da disponibilidade desses recursos em uma determinada região.

Em relação à composição elementar, a biomassa consiste principalmente de carbono (C) e oxigênio (O) e menores concentrações de hidrogênio (H) e nitrogênio (N). A composição aproximada da biomassa descreve a concentração de cinza (Cz), material volátil (V), carbono fixo (CF) e umidade (U). Alguns tipos de biomassa carregam proporções significativas de espécies inorgânicas. A concentração das cinza pode variar de menos de 1% em madeiras macias a 15% em biomassa herbácea e resíduos agrícolas (Yaman, 2004). Além disso, a biomassa contém altos teores de material volátil e menores concentrações de carbono fixo. Como pode-se observar na Tabela 1 a composição pode variar em um mesmo tipo de biomassa, como no caso do bagaço de cana e a torta de soja. As variações na composição da biomassa ocorrem devido a diferentes espécies de plantas utilizadas, composição do solo, variações do clima e a condições de cultivo em que essa matéria-prima é produzida (Gent et al., 2017).

As matérias-primas de biomassa são de natureza lignocelulósica, ou seja, são compostas por celulose, hemicelulose, lignina e pequenas quantidades de extrativos (como taninos e resinas) (Isahak et al., 2012). Cada um dos três compostos lignocelulósicos tem uma faixa de temperatura específica na qual se torna mais suscetível à degradação térmica. Assim, a composição da biomassa é uma consideração importante no desenvolvimento de processos térmicos para a produção de combustíveis e produtos químicos derivados deste material (Schorr; Muinonen; Nurminen, 2012).

Tabela 1- Composição de algumas biomassas

Biomassa	C*	H*	O*	N*	CF**	V**	Cz**	U	Referência
Bagaço de cana	47,3	6,1	45,7	0,8	9,5	83,0	7,5	6,0	Xu et al., 2011
Bagaço de cana	44,9	5,9	49,0	0,2	11,9	84,8	3,3	5,4	Varma e Mondal, 2017
Torta de soja	55,9	6,6	28,2	9,3	15,7	78,1	6,1	8,4	Uzun, Pütün e Pütün, 2006
Torta de soja	41,7	6,2	45,0	7,1	16,5	76,6	7,0	9,5	Şensöz e Kynar, 2006
Casca de palma	49,7	5,3	44,9	0,1	22,1	75,5	2,4	11	Abnisa et al., 2011
Palha de arroz	43,2	5,6	48,8	2,1	1,6	88,8	9,5	6,8	Jung, Kang e Kim, 2008
Folha de cana	48,9	6,5	44,4	0,2	9,2	84,7	6,1	6,7	Pattiya, Sukkasi e Goodwin, 2012
Casca de arroz	40,0	4,3	55,3	0,5	16,1	69,7	14	8,5	Vieira et al., 2020
Torta de nim	47,4	4,4	44,3	3,9	8,57	85,8	5,7	12	Dhanavath et al., 2019

* - % em base seca livre de cinza; ** - % em base seca.

Fonte: O autor, 2022.

Nos últimos anos, tem havido um crescente interesse, em nível mundial, na utilização de biomassa como fonte renovável de energia. Segundo Mckendry (2002), há várias razões para esta situação. Em primeiro lugar, a evolução tecnológica promete a aplicação de biomassa com menor custo e com maior eficiência de conversão. Além disso, a produção de biomassa estimula a geração de empregos e contribuiu positivamente para o meio ambiente.

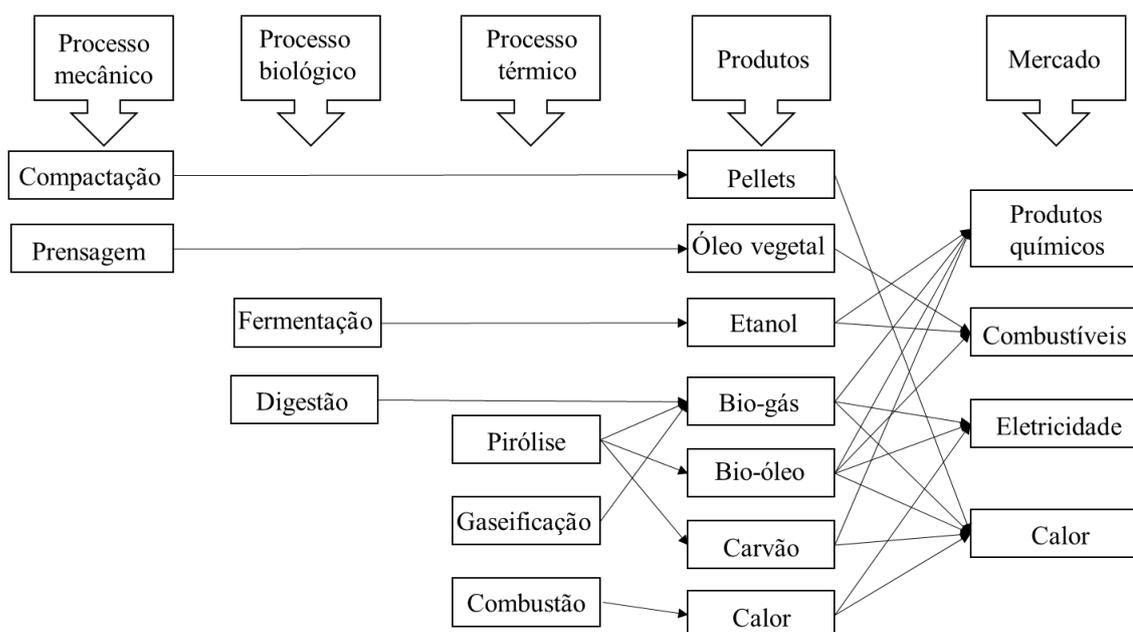
Devido aos problemas ambientais relacionados à utilização de combustíveis de origem fóssil, muitos estudos têm sido realizados na busca por alternativas que possibilitem a diminuição do uso desses materiais. Como alternativa para reduzir os impactos ambientais devido a utilização destes combustíveis, a biomassa residual, proveniente da agroindústria, pode substituir os recursos fósseis em muitas aplicações como produção de energia, produtos químicos (aditivos alimentares, surfactantes, solventes orgânicos, fertilizantes) e biomateriais (Collard; Blin, 2014).

2.2 Processos de conversão

A biomassa pode ser convertida em energia sendo submetida a diferentes processos, sendo eles térmicos, biológicos ou mecânicos. O processamento biológico é geralmente muito seletivo e produz um pequeno número de produtos com um rendimento elevado utilizando catalisadores biológicos. Dentre os processos bioquímicos temos a fermentação, para conversão de açúcar em etanol, e a digestão para produção de biogás. Os processos mecânicos não alteram o estado físico da biomassa, mas aumentam seu valor comercial. Entre eles pode-se citar a compactação de resíduos na forma de *pellets*, moagem ou picagem de palha e extração mecânica do óleo (Bridgwater, 2006). A conversão térmica gera produtos múltiplos e complexos em curto tempo de reação. Algumas vezes utiliza-se catalisadores para melhorar a qualidade do produto obtido (Bridgwater, 2012).

Entre os processos de conversão de biomassa, os térmicos são uma alternativa promissora para muitas aplicações de energia. Estes processos consistem na conversão de biomassa pela ação do calor. Os processos térmicos mais comuns são a combustão, a gaseificação e a pirólise (Bridgwater, 2003). A Figura 1 resume os principais processos de conversão e seus produtos.

Figura 1 - Processos de conversão e seus produtos



Fonte: Adaptado de Bridgwater, 2012.

A combustão é um processo termoquímico no qual o carbono e o hidrogênio, presentes na biomassa, reagem com o oxigênio formando CO_2 , água e liberando energia. A tecnologia está disponível no mercado e há muitos exemplos de trabalhos bem-sucedidos em toda a América do Norte e Europa, onde utiliza-se, frequentemente, resíduos florestais, agrícolas e industriais. Porém, a eficiência na geração de energia tende a ser baixa, tipicamente 15% para unidades pequenas e até 30% para as unidades maiores e mais modernas. Além disso, as emissões de monóxido de carbono, devido à queima incompleta, e cinzas geradas no processo ainda são problemas a serem contornados (Bridgwater, 2003).

Na gaseificação a biomassa é convertida em uma mistura de gases, além de pequenas quantidades de carvão, cinzas e componentes condensáveis, por oxidação parcial em altas temperaturas (800-900 °C) (Goyal; Seal; Saxena, 2008). O gás produzido é composto de monóxido de carbono, dióxido de carbono, hidrogênio, metano e nitrogênio, e pode ser produzido a partir de biomassa por oxidação parcial com ar, oxigênio ou vapor (Bridgwater, 2003).

O gás produzido a partir de oxigênio ou vapor possui um poder calorífico médio, sendo mais adequado para a síntese de combustíveis para transportes e produtos químicos. Na geração de eletricidade, não há evidência dos benefícios da utilização de gases com maior poder calorífico, sendo assim, utiliza-se o gás gerado a partir do ar, que possui um poder calorífico baixo devido a diluição com o nitrogênio (Bridgwater, 2003).

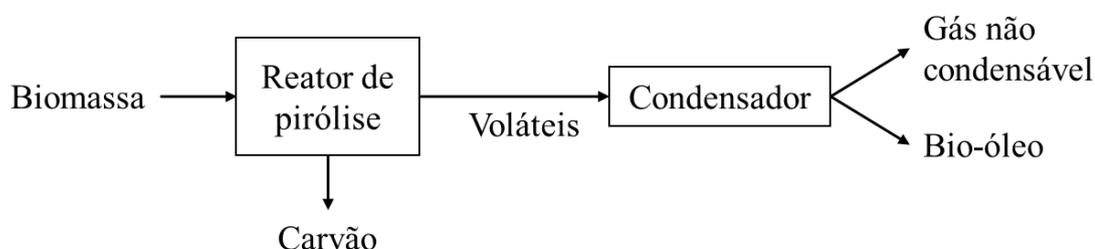
Apesar da tecnologia de gaseificação da biomassa ter sido demonstrada com sucesso, ela é relativamente cara em comparação com a energia fóssil. Além disso, mais pesquisas são necessárias para melhorar a pureza dos produtos. Por exemplo, o desenvolvimento da produção de hidrocarbonetos sintéticos por gaseificação é dificultado pelo elevado teor de contaminantes presentes nos gases produzidos por biomassa (Collard; Blin, 2014).

A terceira alternativa de conversão térmica é a pirólise que consiste na destruição térmica da biomassa na ausência de oxigênio, o que leva à produção de bio-óleo, bio-gás e carvão (Goyal; Seal; Saxena, 2008).

2.3 Pirólise

De acordo com a Agência Nacional de Energia Elétrica (2008), a pirólise é o mais simples e mais antigo processo de conversão de biomassa em outro produto de melhor qualidade e conteúdo energético. O processo consiste na decomposição térmica da biomassa, em atmosfera isenta de oxigênio, em carvão, gás e bio-óleo cujas proporções variam de acordo com a natureza da biomassa e com as condições nas quais o procedimento é realizado. A Figura 2 mostra um esquema do processo de pirólise de biomassa.

Figura 2 - Esquema do processo de pirólise de biomassa

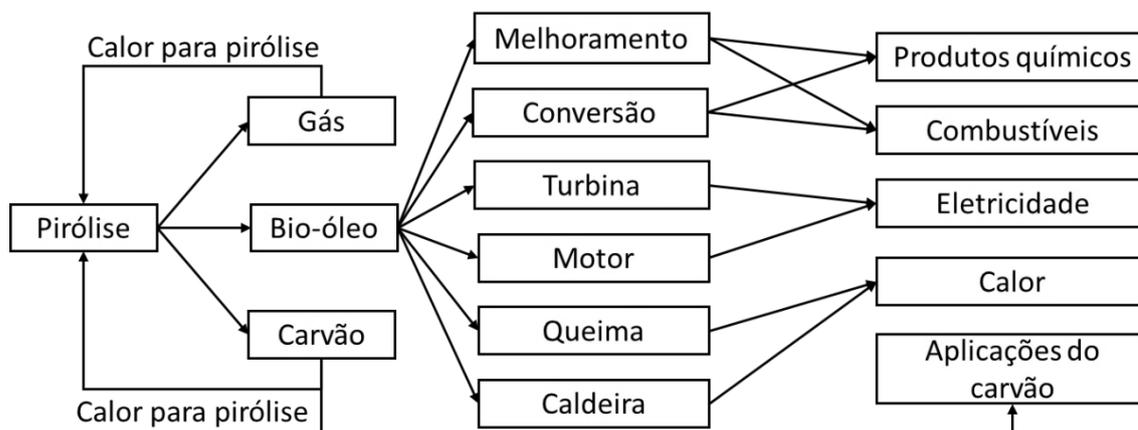


Fonte: O autor, 2020.

O carvão, produto sólido da pirólise, é composto principalmente de carbono (91-63%) e oxigênio (30-10%) (Angin, 2013; Bartoli *et al.*, 2016; Chang *et al.*, 2016; Park *et al.*, 2014; Pütün *et al.*, 2005; Quan; Gao; Song, 2016; Razuan *et al.*, 2010; Wu *et al.*, 2016; Yanik *et al.*, 2007). Em função de seu alto conteúdo de carbono pode ser usado para gerar calor para o processo, na preparação de carvão ativo (Sricharoenchaikul *et al.*, 2008; Venderbosch; Prins, 2010) e para melhorar a fertilidade do solo (Lehmann; Gaunt; Rondon, 2006). Além disso, devido a uma grande área superficial, o carvão pode ser usado para adsorção de poluentes, sendo que os processos de ativação física ou química podem melhorar este recurso (Lee *et al.*, 2013). O biogás, fração de gases não condensáveis do processo, é composto principalmente de dióxido de carbono, metano, monóxido de carbono, hidrogênio e hidrocarbonetos leves (Chang *et al.*, 2016; Fu *et al.*, 2011; Ly *et al.*, 2015; Park. *et al.*, 2014; Ucar; Ozkan, 2008; Wu *et al.*, 2016). Pode ser usado como combustível para fins de combustão industrial, como fonte de calor para o próprio processo de pirólise (Goyal; Seal; Saxena, 2008) ou em turbinas a gás para produzir eletricidade (Demirbaş, 2001). O bio-óleo é o líquido produzido na pirólise e tem a vantagem de ser facilmente armazenado e transportado. Pode ser utilizado

para gerar calor e energia ou como um intermediário para produção de biocombustíveis e produtos químicos (Bennett; Helle; Duff, 2009; Bridgwater, 2012; Ikura; Stanciulescu; Hogan, 2003). As aplicações dos produtos da pirólise são mostradas na Figura 3.

Figura 3 - Aplicações dos produtos da pirólise



Fonte: Adaptada de Bridgwater, 2012

Segundo Bridgwater (2012) praticamente qualquer tipo de biomassa pode ser considerado para a pirólise, sendo que mais de 100 tipos diferentes já foram testados. O processo de pirólise é dividido em três tipos, dependendo das condições operacionais utilizadas: pirólise lenta, rápida e *flash*. Além disso, as proporções e a composição dos produtos de pirólise também dependem das condições utilizadas. A Tabela 2 mostra os diferentes tipos de pirólise e as condições de operação utilizadas em cada um.

Tabela 2 - Diferentes tipos de pirólise

Tipo	Parâmetros	Rendimento de produto (%)		
		Bio-óleo	Carvão	Bio-gás
Lenta	Temperatura baixa (400 °C) e elevado tempo de residência	30	35	35
Rápida	Temperatura moderada (500 °C) e curto tempo de residência	75	12	13
<i>Flash</i>	Altas temperaturas (800 °C) e curto tempo de residência	5	10	85

Fonte: Adaptada de Bridgwater, 2012.

A pirólise lenta ocorre em uma temperatura de processo mais baixa, menor taxa de aquecimento e tempos de residência mais longos, o que favorece a produção de carvão. A

pirólise *flash* é o processo em que a temperatura e taxa de aquecimento são mais elevadas, o que beneficia a produção de bio-gás. Neste caso, devido ao rápido aquecimento, a biomassa deve possuir um menor tamanho de partícula. A pirólise rápida favorece a formação de bio-óleo e ocorre a uma temperatura moderada, curtos tempos de residência de vapor na zona quente de reação e rápido arrefecimento dos vapores (Bridgwater, 2012).

A pirólise é aplicada a milhares de anos para a produção de carvão vegetal, mas apenas nos últimos 30 anos que a pirólise rápida se tornou de grande interesse. Isto porque o processo leva a um alto rendimento de líquido (Bridgwater, 2012). A tecnologia de pirólise rápida pode ter custos de investimento relativamente baixos, alta eficiência em comparação com outros processos e aceitabilidade ambiental (Bridgwater, 2012; Jahirul *et al.*, 2012; Maity, 2015). O processo de pirólise rápida, aplicado à produção de bio-óleo, compreende três fases principais:

- ✓ A recepção da biomassa, armazenamento e pré-tratamento;
- ✓ A conversão de biomassa para formação de bio-óleo;
- ✓ A conversão deste produto líquido primário, por processamento, refino ou limpeza, a um produto final comercializável, tais como eletricidade, calor, biocombustíveis e produtos químicos (Bridgwater, 2012).

Durante as últimas décadas, diferentes reatores foram testados para o processo de pirólise com o objetivo de atender aos requisitos de transferência rápida de calor (Mohan; Pittman; Steele, 2006). Os projetos de reatores de pirólise incluem: leito fixo; leito fluidizado; leito fluidizado circulante; ablativo; micro-ondas; cone giratório; vácuo e reator solar. Este último possui a vantagem de usar fontes de energia renováveis no aquecimento (Joardder *et al.*, 2014). A escolha da tecnologia influencia nos parâmetros operacionais desses reatores, como o tamanho das partículas da matéria-prima, a taxa de transferência de calor e o tempo de residência do sólido e dos voláteis (Jahirul *et al.*, 2012).

O aprimoramento da pirólise para obtenção de bio-óleo a partir de biomassa continua sendo tema de grande interesse para estudos. Os parâmetros do processo afetam fortemente o rendimento e a composição dos produtos da pirólise devido às diferentes reações químicas que são favorecidas em cada condição (Isahak *et al.*, 2012). Deste modo, um dos principais focos das pesquisas é a determinação das melhores condições de pirólise para se obter maiores conversões de bio-óleo em cada biomassa de interesse.

2.4 Bio-óleo

O líquido de pirólise é referido por muitos nomes, incluindo óleo de pirólise, bio-óleo, óleo biocombustível, óleo de madeira, alcatrão pirolenhoso e ácido pirolenhoso. O bio-óleo é composto por uma complexa mistura de substâncias orgânicas como hidrocarbonetos aromáticos, fenóis, cetonas, ésteres, éter, açúcares, aminas, álcoois, furanos e água obtida por reações químicas e pela umidade da biomassa. O líquido possui coloração marrom escuro e se aproxima da biomassa em sua composição elementar (Isahak *et al.*, 2012).

As massas molares das substâncias que formam o bio-óleo variam de muito baixo, 18 g/mol (água) para tão alto quanto 5000 g/mol ou mais (ligninas pirolíticas). A massa molar média de diferentes bio-óleos varia de 370-1000 g/mol. Mais de 300 substâncias orgânicas já foram identificadas em diferentes bio-óleos (Qiang; Wen-Zhi; Xi-Feng, 2009).

As composições químicas dos bio-óleos são determinadas por diversos fatores, tais como o tipo de biomassa, o pré-tratamento de matéria-prima (tamanho das partículas e umidade), condições de pirólise (temperatura, taxa de aquecimento, tempo de residência) e condensação. Assim, bio-óleos produzidos a partir de materiais diferentes e por diferentes condições de pirólise podem diferir muito uns dos outros (Qiang; Wen-Zhi; Xi-Feng, 2009).

A composição química do bio-óleo é totalmente diferente dos combustíveis derivados do petróleo, o que resulta na grande diferença nas propriedades entre eles. A Tabela 3 mostra uma comparação entre as propriedades do bio-óleo de pirólise e óleo combustível com os dados obtidos no trabalho de Zhang *et al.* (2007).

Tabela 3 - Comparação entre as propriedades de bio-óleo e combustível fóssil

Propriedades	Bio-óleo	Combustível fóssil
Umidade (%)	15-30	0,1
pH	2,5	-
Poder calorífico superior (MJ/Kg)	16-19	40
Viscosidade (cP)	40-100	180
Sólidos (%)	0,2-1	1
Composição elementar (%)		
Carbono	54-58	85
Hidrogênio	5,5-7	11
Oxigênio	35-40	1
Nitrogênio	0-0,2	0,3
Cinza	0-0,1	0,1

Fonte: Adaptada de Zhang *et al.*, 2007

O líquido contém um teor de água que pode variar de cerca de 15% para 30 - 50%, dependendo da biomassa e das condições de operação. Se o teor em água do líquido de pirólise excede um certo limite (geralmente 30%), o líquido separa-se em duas fases. Este tipo de separação de fase ocorre, geralmente, quando o teor de umidade da biomassa é muito elevado ($> 15\%$) (Oasmaa; Peacocke, 2010). A presença de água diminui o poder calorífico, mas, por outro lado, reduz a viscosidade e aumenta a fluidez, o que é bom para a combustão do bio-óleo em motores (Zhang *et al.*, 2007). A viscosidade do bio-óleo pode variar em uma ampla faixa. Em Shen *et al.* (2009), o bio-óleo produzido a partir do Eucalipto possui uma viscosidade variando de 30 a 91 mPa·s apenas alterando as condições do processo. Garcia-Peres *et al.* (2008) também observaram esta variação com a mesma biomassa, sendo que o bio-óleo com menor viscosidade contém uma maior concentração de água em sua composição.

O bio-óleo possui quantidade substancial de ácidos carboxílicos, tais como ácidos acético e fórmico, o que leva a baixos valores de pH. Em muitos estudos o valor de pH do bio-óleo se encontra na faixa de 2-4 (Abnisa *et al.*, 2013; Asadullah *et al.*, 2013; Bakar; Titiloye, 2013; Bordoloi *et al.*, 2015; Choudhury *et al.*, 2014; Ingram *et al.*, 2008; Kim *et al.*, 2011; Lazzari *et al.*, 2016; Oasmaa *et al.*, 2010; Paenpong; Pattiya, 2016; Pattiya; Suttibak, 2012; Tsai; Lee; Chang, 2006; Varma *et al.*, 2019; Varma; Mondal, 2017; Wu *et al.*, 2016; Yin *et al.*, 2013). Devido a corrosividade deve-se tomar maior cuidado com os materiais utilizados para seu armazenamento (Zhang *et al.*, 2007).

Observa-se que algumas propriedades afetam negativamente a qualidade do bio-óleo como, alta viscosidade, incompatibilidade com combustíveis convencionais, corrosividade, alto teor de oxigênio e instabilidade (Bridgwater, 2012). Assim, um processo de melhoramento é necessário antes da sua aplicação como combustível.

Muitos pesquisadores avaliam os processos de melhoramento do bio-óleo que pode ser de forma física, química ou catalítica. Os processos de melhoramento incluem adição de solvente (Wenting *et al.*, 2014), hidrotratamento (Zhang *et al.*, 2005), emulsificação (Ikura; Stanciulescu; Hogan, 2003), craqueamento com zeólita (Park *et al.*, 2007; Vitolo *et al.*, 2001) e craqueamento em produtos químicos (Bennett; Helle; Duff, 2009; Isahak *et al.*, 2012; Zhang *et al.*, 2007).

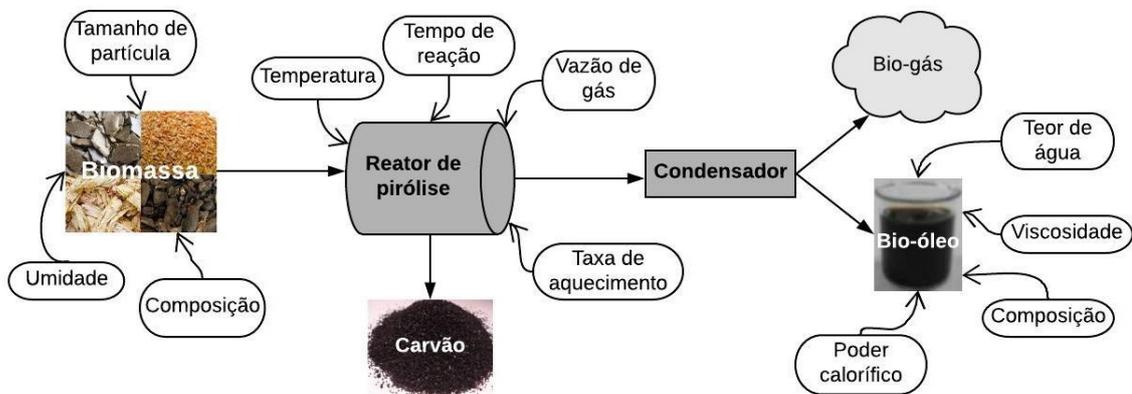
O mercado do bio-óleo é bastante promissor pois desperta interesse do setor industrial devido a aplicações em diferentes áreas como: produção de solventes, adesivos, fibras sintéticas, fármacos, resinas, cosméticos, entre outras. Além disso, pode substituir o óleo

combustível ou diesel em muitas aplicações incluindo caldeiras, fornos, motores e turbinas para geração de energia elétrica (Bridgwater, 2012).

2.5 Fatores que influenciam na pirólise

Durante a pirólise, a biomassa passa por reações primárias e secundárias envolvendo mecanismo de transferência de calor e de massa. Nas reações primárias ocorre a decomposição da celulose, hemicelulose e lignina com a formação de produtos primários e intermediários por reação de desidratação e carbonização. Posteriormente, os produtos intermediários sofrem reações de decomposição e volatilização. Os produtos obtidos são muito sensíveis a variações operacionais e composição da matéria prima (Akhtar; Amin, 2012). A Figura 4 mostra as principais variáveis do processo que influenciam no bio-óleo obtido.

Figura 4 - Fatores que influenciam no processo



Fonte: O autor, 2022.

Para uma mesma biomassa pode-se encontrar diferenças significativas no rendimento e na composição do bio-óleo devido à variação das condições operacionais impostas. Uma grande variedade de biomassas e condições de operação já foram testadas no processo de pirólise. De acordo com a literatura, as variáveis como, taxa de aquecimento, tamanho de partícula da biomassa, tempo de residência do gás, temperatura de reação, tempo de reação e a composição da biomassa afetam fortemente o produto final.

2.5.1 Temperatura

A utilização da temperatura elevada na pirólise fornece o calor necessário para a decomposição da matéria-prima. Ou seja, a temperatura gera energia para quebrar as ligações da biomassa.

Tsai, Lee e Chang (2007) observaram um aumento no rendimento de óleo de 11,26% para 35,92% quando a temperatura de pirólise da casca de arroz mudou de 400 °C para 500 °C, respectivamente. No entanto, acima de 500 °C, o rendimento aumentou a uma taxa menor com o aumento da temperatura e atingiu 40% em 800 °C. Lazzari *et al.* (2016) variaram a temperatura na pirólise da semente de manga entre 450 °C e 650 °C. O rendimento máximo de bio-óleo obtido foi de 38,8% na temperatura mais alta. O mesmo comportamento foi observado por Schröder (2004) na pirólise de madeira de faia com a temperatura variando de 280 a 525 °C.

O aumento da temperatura causa um efeito positivo no rendimento de líquidos, mas em temperaturas muito altas, o efeito oposto é observado na maioria dos estudos analisados. Isso acontece porque, em temperaturas muito elevadas, ocorre o craqueamento secundário dos voláteis, o que resulta em um maior rendimento de gás (Isahak *et al.*, 2012). Jung, Kang e Kim (2008) mostraram este comportamento usando serragem de bambu no reator de pirólise de leite fluidizado. O aumento da temperatura de 350 °C para 405 °C aumentou o rendimento de bio-óleo de 56% para 72%. Enquanto o aumento da temperatura para 510 °C reduziu o rendimento para 61%. Já utilizando a palha de arroz os pesquisadores evidenciaram que na temperatura de 445 °C o maior rendimento de líquido é alcançado (68%) e o aumento ou redução da temperatura causa uma variação negativa no rendimento. Abnisa *et al.* (2011) utilizando casca de palma, relataram que o aumento da temperatura da pirólise de 400 °C para 500 °C levou a um aumento no rendimento de bio-óleo de 36,8% para 46,1%. No entanto, o rendimento diminuiu de 42,9% para 35,6% quando a temperatura foi de 600 °C para 800 °C. No estudo realizado com casca de arroz por Alvarez *et al.* (2014), o mesmo efeito da temperatura foi observado, ou seja, houve um aumento no rendimento do bio-óleo quando a temperatura foi elevada de 400 °C (66%) para 450 °C (70%) e, uma posterior redução com a elevação da temperatura para 600 °C (65%). Em relação ao rendimento do carvão a elevação da temperatura de 400 para 450 °C levou a uma redução deste de 32% para 25%, a qual não se alterou a uma temperatura de 600 °C, segundo os autores o alto rendimento do carvão se dá

devido ao alto teor de cinza na biomassa utilizada (12,9%). Já em relação à produção de gás, um aumento gradativo foi verificado juntamente com a elevação da temperatura de operação devido ao craqueamento secundário ocorrido nos vapores de pirólise.

Numerosos estudos experimentais discutem o papel da temperatura no rendimento de bio-óleo. Muitos estudos mostram que a temperatura para obtenção de maiores rendimentos de líquidos está na faixa de 450-550 °C, mas esses valores variam com a biomassa utilizada e demais variáveis empregadas no processo (Sharma; Sheth, 2015).

Diferentes temperaturas para obter o máximo rendimento de líquido foram encontradas na pirólise de diversas biomassas, como resíduo da palma (72,4%) a 500 °C (Abdullah; Gerhauser, 2008); torta de nim, (40,2%) e *Cynara cardunculus L.* (56,23%) a 400 °C (Encinar; González; González, 2000; Volli; Singh, 2012); bagaço de azeitona (46,3%) a 600 °C (Encinar *et al.*, 1998); bagaço de cana (56%) a 475 °C (Islam; Parveen; Haniu, 2010); mandioca (63,23%) e caule de mandioca (61,39%) a 472 °C e 469 °C, respectivamente (Pattiya; Suttibak, 2012); torta de casca de pinhão-manso (48%) a 470 °C (Kim *et al.*, 2013); casca de pistache (20,5%) a 550 °C (Apaydin-Varol; Pütün; Pütün, 2007); álamo (69%) a 455 °C (Makibar *et al.*, 2015).

Pattiya (2011) mostrou que, para rizoma e caule de mandioca, as temperaturas para maiores rendimentos de líquidos foram de 475 °C e 510 °C, respectivamente, nas mesmas condições de pirólise. Isto indica que a composição da biomassa interfere na temperatura ideal do processo. Biswas *et al.* (2017) acreditam que cada biomassa tem o potencial de fornecer um rendimento máximo de bio-óleo a uma temperatura específica de pirólise.

Além do rendimento, a qualidade do líquido também varia com a temperatura. Segundo Huang *et al.* (2014), o bio-óleo obtido em baixas temperaturas é composto de alcenos, alcanos, ácidos graxos de cadeia longa, ésteres, nitrilas alifáticas e amidas. Uma temperatura de pirólise elevada favorece o craqueamento das espécies alifáticas e a formação de aromáticos, resultando em um bio-óleo com menor razão H/C e maior conteúdo de organonitrogenados. Alvarez *et al.* (2014) observaram que o aumento da temperatura reduz a concentração de cetonas e ácidos e aumenta os fenóis, mas o conteúdo de água, nitrogenados, éteres e aldeídos não são afetados.

O conteúdo de carbono e hidrogênio são indicadores essenciais da qualidade do bio-óleo. Para produtos líquidos de alta qualidade, o teor de carbono e hidrogênio deve ser alto; por outro lado, um alto teor de oxigênio reduz o valor calórico do bio-óleo (Pütün; Ateş; Pütün, 2008). Na pirólise lenta da palha de trigo e da palha de aveia, Ateş e Isikdag (2008)

mostraram uma redução no teor de oxigênio enquanto as porcentagens de carbono e hidrogênio aumentaram com a temperatura. Ly *et al.* (2016) também observaram o mesmo comportamento na pirólise de *Saccharina japonica*.

2.5.2 Tamanho da partícula

A transferência de calor durante a pirólise é prejudicada pelo fato de a biomassa ser um mau condutor de calor. Sendo assim, o tamanho da partícula influencia no rendimento do bio-óleo produzido e é importante para minimizar os problemas de transferência de calor (Akhtar; Amin, 2012).

Shen *et al.* (2009) observaram um aumento na produção de bio-óleo, de 12–14%, com uma redução da partícula de eucalipto de 1,5 mm para 0,3 mm, no reator de leito fluidizado. Kang *et al.* (2006) também observaram o aumento no rendimento de bio-óleo com a redução do tamanho de partículas em pirólise de pinho em reator de leito fluidizado. O mesmo ocorreu em Garg, Anand e Kumar (2016) onde uma redução no rendimento do líquido foi observada com o aumento no tamanho de partícula de sementes de acácia em reator de leito fixo, sendo que o rendimento máximo de bio-óleo (~ 32%) foi obtido com diâmetro de partícula <0,4 mm. Um resultado oposto foi encontrado por Abnisa *et al.* (2011) que estudaram a pirólise da casca de palma em reator de leito fixo e observaram que o aumento no tamanho de partícula leva a um aumento no rendimento de bio-óleo. Sendo os rendimentos obtidos com as partículas menores (0,5 mm) e maiores (2 mm) de 27% e 45,8%, respectivamente. O mesmo ocorreu em Li *et al.* (2004) no qual o rendimento de bio-óleo de caroço de damasco, em reator de leito fluidizado, aumentou de 17,8% para 48,3% variando-se o tamanho da partícula de 0,25 mm para 1,25 mm.

Em geral, partículas menores são preferidas para pirólise devido ao seu aquecimento mais rápido e uniforme. Além disso, a taxa de aquecimento com partículas maiores é baixa devido à condutividade térmica baixa da biomassa. Assim, o calor pode fluir apenas para uma certa profundidade no tempo disponível da pirólise e os núcleos das partículas maiores se tornam carbonizados ou podem não ser completamente decompostos, resultando num aumento do rendimento de carvão e uma diminuição na produção de líquidos. No entanto, as partículas microscópicas podem reduzir o rendimento de líquido na pirólise pois a

decomposição da biomassa ocorre muito rapidamente, sobrando tempo suficiente para que ocorra reações secundárias nos vapores e, conseqüentemente, aumento do rendimento de gás (Islam; Parveen; Haniu, 2010). Além disso, o material muito fino pode ser removido da zona de reação antes da conclusão do processo (Pattiya; Suttibak, 2012).

Experiências de pirólise realizadas por Pütün, Apaydm e Pütün (2004), Şensöz e Kaynar (2006) e Varma e Mondal (2017) usando palha de arroz, bagaço de soja e bagaço de cana, respectivamente, sugeriram que o tamanho das partículas tem pouco efeito sobre o rendimento do produto no reator de leito fixo. Onay e Koçkar (2004) obtiveram um maior rendimento de bio-óleo (68%) usando um tamanho de partícula na faixa de 0,6-1,25 mm. O uso de partículas maiores ou menores causou um efeito negativo no rendimento líquido da pirólise de colza em reator de leito fixo. Islam, Parveen e Haniu (2010) observaram o mesmo comportamento no bagaço de cana. A produção máxima de bio-óleo foi obtida com um tamanho de partícula entre 0,6 mm e 1,18 mm analisando uma faixa de 0-4,5 mm. Raja *et al.* (2010) investigaram a pirólise *flash* de torta de pinhão manso em reator de leito fluidizado. O tamanho das partículas variou de 0,6 a 1,18 mm; o rendimento máximo de óleo foi obtido com um tamanho de partícula de 1,0 mm. Pattiya e Suttibak (2012) obtiveram maior rendimento líquido (64%) com um tamanho de partícula de rizoma de mandioca entre 0,25 e 0,42 mm em uma faixa de 0 a 0,6 mm.

Alguns estudos analisaram a mesma faixa de tamanho de partícula encontrando diferentes resultados de rendimento. Onay, Beis e Koçkar (2001) variando o tamanho de partícula de 0,224 a 1,8 mm concluíram que um tamanho de partícula entre 0,6 e 1,8 mm é adequado para obter maior rendimento de líquido em pirólise de colza, em reator de leito fixo. Pütün, Apaydin e Pütün (2002), variando o tamanho de partícula de 0,224 a 1,8 mm na pirólise da torta de soja, obtiveram um rendimento máximo de óleo (30,23%) entre 0,85-1,25 mm. Ateş, Pütün e Pütün (2004) também investigaram a faixa de tamanho de partícula de 0,224-1,8 mm, mas usando o talo de gergelim em reator de leito fixo. O rendimento de óleo teve um aumento na faixa de 0,425-0,85 mm. Isso mostra que o tamanho de partícula, usado para otimizar o rendimento do bio-óleo, varia dependendo do tipo de biomassa e demais condições impostas. As informações conflitantes encontradas na literatura sobre o tamanho ideal da biomassa dificultam a generalização desse parâmetro para o sistema de pirólise.

O tamanho da partícula de biomassa também influencia na composição do produto obtido. Onay e Koçkar (2004) estudaram a influência dos diferentes tamanhos de partícula de colza na pirólise em reator de leito fixo com taxa de aquecimento de 30 °C/min, vazão de N₂

de 100 cm³/min e temperatura de 550 °C, os autores observaram que o aumento no tamanho da partícula causa um aumento na concentração de substâncias orgânicas mais voláteis e uma redução de substâncias de massa molar elevada como açúcares, fenóis e furanos no bio-óleo.

2.5.3 Taxa de aquecimento

A taxa de aquecimento do processo é outro fator importante na decomposição da biomassa. Na pirólise rápida as taxas de aquecimento são geralmente mais elevadas e reduzem as limitações de transferência de calor e massa, resultando em rendimentos máximos de bio-óleo.

Pütün *et al.* (2007a) compararam diferentes taxas de aquecimento no rendimento do produto de pirólise, utilizando resíduos de tabaco em reator de leito fixo. O rendimento de óleo foi de 27% com a taxa de aquecimento de 7 °C/min. O aumento da taxa até 300 °C/min resultou em um aumento de 10% na produção do líquido. Uzun, Pütün e Pütün (2006) observaram que uma maior taxa de aquecimento aumenta a produção de bio-óleo na pirólise da torta de soja. O rendimento máximo alcançado foi de 42,83% com uma taxa de aquecimento de 700 °C/min, a uma temperatura de pirólise de 550 °C. O aumento da taxa de aquecimento de 5 para 700 °C/min resultou num aumento de rendimento de bio-óleo de 23,36%.

Resultados contrários também foram observados na literatura. Şensöz, Demiral e Gerçel (2006), utilizando taxa de aquecimento de 10 °C/min, obtiveram rendimentos de bio-óleo cerca de 8,4% maiores do que a 50 °C/min na pirólise do bagaço de azeitona no reator de leito fixo. Mishra e Mohanty (2018) obtiveram um menor rendimento de líquido com a elevação da taxa de aquecimento de 80 °C/min (49,5%) para 120 °C/min (40,4%) na pirólise de semente de nim. Onay, Beis e Koçkar (2001) analisaram taxas de aquecimento de 100 °C/min, 300 °C/min e 800 °C/min na pirólise de colza. O rendimento do óleo foi cerca de 58% mais elevado empregando-se a taxa de aquecimento de 300 °C/min, em comparação com a de 100 °C/min. O aumento da taxa de aquecimento para 800 °C/min não alterou o rendimento do bio-óleo, permanecendo constante a um nível máximo de 68%. A mesma tendência foi observada na pirólise de talo de gergelim onde, Ateş, Pütün e Pütün (2004) realizaram experiências com taxas de 100 a 700 °C/min. O rendimento de óleo atingiu um

valor máximo de 36,22% com uma taxa de aquecimento de 500 °C/min. Este valor foi de cerca de 9,5% maior do que obtido em 100 °C/min. Em taxa de aquecimento de 700 °C/min o aumento no rendimento do bio-óleo ocorreu em quantidades insignificantes. Tsai, Lee e Chang (2007) estudaram a pirólise da casca de arroz no reator de leito fixo. A taxa de aquecimento foi alterada de 100 °C/min até 500 °C/min. O rendimento de óleo atingiu seu máximo com uma taxa de aquecimento de 200 °C/min e permaneceu constante em taxas mais altas.

Alguns pesquisadores observaram que a taxa de aquecimento mais elevada afeta a temperatura ideal para a obtenção de maior rendimento de bio-óleo. Por exemplo, Debdoubi *et al.* (2006) pirolisaram esparto com diferentes taxas de aquecimento em um intervalo de temperatura de 400-700 °C. Em taxas de 50 °C/min e 150 °C/min o maior rendimento de bio-óleo foi alcançado a 500 °C, dando um rendimento de 45% e 57%, respectivamente. No entanto, quando a taxa de 250 °C/min foi utilizada, o rendimento máximo de óleo foi alcançado em 550°C (68,5%).

Şensöz e Angin (2008) estudaram o efeito da taxa de aquecimento de 10 °C/min, 30 °C/min e 50 °C/min na composição do bio-óleo da torta de cártamo. Os resultados das análises mostraram que a composição química do bio-óleo obtido em diferentes condições é semelhante. Uzun, Pütün e Pütün (2006) obtiveram uma redução acentuada do teor de água no bio-óleo com o aumento da taxa de aquecimento, obtendo um bio-óleo de melhor qualidade. O aquecimento rápido inibiu reações de desidratação secundária, causando uma redução no teor de água no bio-óleo. O teor de oxigênio no produto líquido de pirólise é também reduzido em taxas de aquecimento mais elevadas. O que sugere que a formação de gases contendo oxigênio, como CO₂ ou CO, aumenta em altas taxas de aquecimento (Akhtar; Amin, 2012).

2.5.4 Tempo de residência do vapor

Durante o processo de pirólise da biomassa é formada uma quantidade significativa de vapor. O vapor de pirólise pode sofrer reações secundárias de craqueamento térmico, polimerização e condensação, o que leva a uma diminuição na produção de bio-óleo (Akhtar; Amin, 2012; Heo *et al.*, 2010). Assim, a remoção rápida dos vapores da zona de reação é necessária para minimizar essas reações. O nitrogênio é o gás mais usado para purgar os

vapores produzidos na pirólise por ser um gás inerte, barato e de fácil acesso (Tripathi; Sahu; Ganesan, 2016). Um maior fluxo de gás no processo leva a um menor tempo de residência do vapor na zona quente de pirólise. O tempo de residência do vapor de pirólise é calculado pelo volume do reator dividido pela vazão do gás inerte utilizado. Assim, um menor tempo de residência do vapor de pirólise no processo é conseguido com uma maior vazão de gás.

Onay (2007) comparou o rendimento de bio-óleo da semente de girassol obtida nas condições estáticas e na vazão de gás de $100 \text{ cm}^3/\text{min}$ em reator de leito fixo. O rendimento de óleo aumentou cerca de 23% com a utilização do gás.

Islam, Parveen e Haniu (2010) estudando o tempo de residência do vapor na pirólise do bagaço de cana-de-açúcar, em reator de leito fixo, observaram que o aumento do tempo de 5 s para 20 s diminuiu o rendimento de líquido e carvão enquanto o rendimento do bio-gás aumentou. A mesma tendência foi observada por Heidari *et al.* (2014) em reator de leito fluidizado. Um efeito oposto foi encontrado por Garg, Anand e Kumar (2016) que observaram uma redução contínua no rendimento de bio-óleo de 44% para 30% com o aumento da vazão de gás de $100 \text{ cm}^3/\text{min}$ para $400 \text{ cm}^3/\text{min}$ na pirólise de sementes de acácia em reator de leito fixo.

Vários estudos que analisaram a vazão de N_2 na pirólise mostraram que o aumento do fluxo de gás inerte causa uma melhora no rendimento de bio-óleo. No entanto, uma vazão de gás muito alta pode reduzir o rendimento de líquido devido a incompleta condensação dos vapores, o que aumenta o rendimento do gás no processo. Além disso, em altas vazões de gás parte da biomassa pode ser removida do reator antes que a pirólise seja concluída (Biradar; Subramanian; Dastidar, 2014; Sulaiman; Abdullah, 2011). Com isso, muitas vezes, uma vazão de gás intermediário é recomendada para a obtenção de alto rendimento de líquido.

Acikgoz, Onay e Kockar (2004) avaliaram a vazão de N_2 na pirólise da linhaça e observaram que o rendimento do produto líquido aumentou 3% quando a vazão foi de $50 \text{ cm}^3/\text{min}$ (55%) para $100 \text{ cm}^3/\text{min}$ (58%). No entanto, quando a vazão aumentou para $200 \text{ cm}^3/\text{min}$ e $400 \text{ cm}^3/\text{min}$ observou-se uma redução no rendimento do bio-óleo, de 47% e 48%, respectivamente, e ocorreu um aumento no rendimento de gás. Abnisa *et al.* (2011), estudando a pirólise da casca de palma, também observaram que o aumento da vazão de N_2 de 1 L/min para 2 L/min aumentou o rendimento de óleo, de 40,5% para 46%. Porém, o rendimento reduziu com um novo aumento da vazão para 3 L/min (38%), 4 L/min (34%) e 5 L/min (26%), e um maior rendimento de bio-gás foi obtido. O mesmo comportamento foi verificado em Raja *et al.* (2010) onde o aumento de $1,25 \text{ m}^3/\text{h}$ para $1,75 \text{ m}^3/\text{h}$ de N_2 elevou o

rendimento de bio-óleo de 37,78% para 64,25% e o rendimento do bio-gás foi reduzido de 58,8% para 31,86% na pirólise de torta de pinhão manso em reator de leito fluidizado. Um acréscimo ainda maior no fluxo, de 1,75 m³/h para 2,4 m³/h, reduziu o rendimento de bio-óleo para 30,5%, e o rendimento de gás aumentou para 60%. Varma *et al.* (2019) também conseguiram melhores resultados de rendimento de líquido com uma vazão de gás intermediária, ou seja, de 0,1 L/min, estudando esta variável na faixa de 0,05 L/min a 0,2 L/min na pirólise de serragem de cedro. Os trabalhos mostram a sensibilidade do rendimento com as variações na vazão do gás inerte e a necessidade de otimizar esta variável.

Pütün *et al.* (2001) analisaram a influência do tempo de residência do vapor em diferentes biomassas. Em *Euphorbia rigida*, o maior rendimento líquido foi obtido a 400 cm³/min de gás inerte, enquanto no bagaço de girassol e na casca de avelã foi de 100 cm³/min e 200 cm³/min, respectivamente. Isso indica que a composição da biomassa também interfere no tempo de residência ideal do vapor no processo.

Segundo Akhtar e Amin (2012) altas temperaturas em tempos de residência relativamente mais longos são adequados para a produção de bio-óleo livre de oxigenados. Sendo assim, a otimização dos tempos de residência do vapor, considerando as outras variáveis, pode ajudar a obter um bio-óleo com maior rendimento e melhor qualidade.

2.5.5 Tempo de reação

O tempo de reação é o tempo em que a biomassa é mantida na temperatura de pirólise especificada. Em operações de processo em batelada, o tempo de reação deve ser suficiente para alcançar o resultado desejado no processo.

Tsai, Lee e Chang (2007), no processo de pirólise rápida de casca de arroz, observaram que aumentando o tempo de reação de 1 min para 2 min, o rendimento de bio-óleo aumentou. No entanto, os valores observados diminuiriam ligeiramente com um tempo de reação maior. Segundo os autores isso ocorreu devido a possíveis reações de gasificação e/ou craqueamento térmico dos produtos de pirólise. Tsai, Lee e Chang (2006) observaram o mesmo comportamento na pirólise do bagaço de cana-de-açúcar, palha de arroz e casca de coco. Em Açikalin, Karaca e Bolat (2012), a casca de pistache foi pirolisada em tempo de reação de 10 a 50 min. O rendimento do produto líquido foi ligeiramente afetado quando o

tempo de reação variou de 10 min (52,96%) a 20 min (53,08%). Outros incrementos do tempo resultaram em uma diminuição contínua do rendimento, chegando a 50,13% em 50 min. O valor constante do rendimento de carvão indicou que a conversão estava completa com apenas 10 min de reação. Sricharoenchaikul *et al.* (2008) também investigaram a influência do tempo de reação na pirólise em reator de leito fixo. O efeito desta variável na distribuição dos produtos de pirólise do resíduo de pinhão manso não foi significativo. No menor tempo, 15 min, utilizado no trabalho, a maioria das reações da pirólise estavam completas.

Islam *et al.* (2005) estudaram o tempo de reação de 20 min a 60 min na pirólise de resíduo de papel e conseguiram aumentar o rendimento do bio-óleo de 32% para 45% com o tempo aumentando de 20 min para 40 min. A partir deste tempo o rendimento de líquidos não se alterou. O mesmo comportamento foi observado por Abnisa *et al.* (2011) que conduziram experimentos com tempo de reação variando de 30 a 150 min em casca de palma. O rendimento de bio-óleo aumentou continuamente com tempos de reação de 30 min a 60 min, porém, após 60 min o rendimento de líquido permaneceu constante. No entanto, quanto maior o tempo de reação, obviamente maior é o custo do processo.

2.5.6 Composição da biomassa

Outro parâmetro que afeta o processo de pirólise é a composição de biomassa utilizada. A biomassa consiste em lignina, hemicelulose, celulose e uma pequena fração de matéria inorgânica cuja porcentagem varia nos diferentes tipos de biomassa. A mesma biomassa pode ter diferenças na composição devido a diferentes condições de solo, idade ou plantio (Omar *et al.*, 2011). A composição e o rendimento dos produtos de pirólise dependem dessas variações de composição.

A biomassa com maiores teores de celulose e hemicelulose geralmente produz um maior rendimento de bio-óleo do que a biomassa com alto teor de lignina. Casoni *et al.* (2015), estudando celulose e biomassas tratadas com ácido e fungos, mostraram que o maior rendimento de bio-óleo é obtido na pirólise da celulose, ou seja, isenta de lignina. Além disso, o tratamento da biomassa com fungos aumentou o rendimento líquido em comparação com a biomassa bruta, pois a atividade enzimática dos cogumelos levou à degradação da estrutura da

lignina, tornando a biomassa mais propensa ao ataque térmico. O tratamento com ácido reduziu o rendimento líquido pois o ácido protege as fibras de celulose contra o calor.

A decomposição da lignina é difícil devido a maior estabilidade estrutural, o que leva a um maior rendimento de carvão (Akhtar; Amin, 2012). Entretanto, o uso de altas taxas de aquecimento e temperaturas pode favorecer a degradação desse material, resultando em maior rendimento líquido. Shen *et al.* (2010) mostraram que o aumento da temperatura de 475 °C para 650 °C na pirólise de lignina aumentou o rendimento de bio-óleo de 32% para 43,2% utilizando um reator em batelada. Liu *et al.* (2017) também estudaram a pirólise dos componentes lignocelulósicos e, neste caso, a temperatura do processo foi analisada. Pode-se observar que para a celulose e hemicelulose os melhores rendimentos ocorreram na temperatura de 450 °C enquanto para a lignina o rendimento foi aumentado de 19% para 23% quando a temperatura foi alterada de 450 °C para 600 °C.

Segundo Qu *et al.* (2011), a celulose é mais volátil que a hemicelulose. Portanto, o material com o maior teor de celulose leva a um maior rendimento de óleo. Wang *et al.* (2011) realizaram a pirólise dos três componentes isolados a 540 °C e os rendimentos obtidos foram de 81,4%, 44,2% e 21,8% para celulose, hemicelulose e lignina, respectivamente. Porém, isso não foi observado por Quan, Gao e Song (2016) ao investigarem o comportamento dos três componentes mais importantes da biomassa na pirólise a 500 °C. A pirólise da celulose produziu 18,67% de líquido, enquanto a pirólise da hemicelulose produziu 30,83% de líquido e a lignina gerou um pequeno rendimento, 0,5%. Kim *et al.* (2013) observaram um maior rendimento de bio-óleo utilizando uma biomassa com maior teor de celulose e hemicelulose. O estudo foi realizado com casca de pinhão manso (PM), torta de albúmen da palma (TP) e cacho vazio do fruto da palma (FP). Os autores observaram um maior teor de celulose na FP (59,7%) enquanto nos casos de PM e TP esses percentuais foram de 36,64% e 30,59%, respectivamente. Já para os teores de lignina, o maior percentual foi encontrado na PM (39,61%) e na TP e FP esses valores foram de apenas 20,41% e 18,2%, respectivamente. Um teor muito reduzido de hemicelulose foi observado na PM (4,82%) enquanto na FP e TP esses valores foram de 22,1% e 30,64%. Após a pirólise, que ocorreu a uma temperatura de 478 °C, os rendimentos de bio-óleo foram de 54%, 52% e 48% na FP, TP e PM, respectivamente. Yanik *et al.* (2007) não encontraram diferença significativa no rendimento dos produtos da pirólise usando espiga de milho, palha e talos de orégano. A quantidade de carvão formado na espiga de milho não foi maior, embora esta biomassa tivesse

um teor de lignina maior em comparação com a palha e o caule. Esse resultado pode estar associado a altas taxas de aquecimento utilizadas no estudo.

A influência do tipo de biomassa utilizada no processo também pode ser observada no estudo realizado por Xu *et al.* (2011) onde utilizou-se a cana de açúcar e o bagaço da cana de açúcar como biomassas de pirólise sob as mesmas condições de processo, sendo estas: temperatura de reação de 400 °C, tempo de residência de 2 s e tamanho de partícula de 0,21-0,85 mm. Os rendimentos obtidos para a cana e o bagaço foram de 58,1% e 55,5% de bio-óleo, 29,4% e 28,8% de carvão e 17,3% e 16,7% de bio-gás, respectivamente. Abnisa *et al.* (2013) também estudaram a pirólise de diferentes biomassas, folha, tronco e haste de palma, sob mesmas condições de operação, sendo: temperatura de 500 °C, vazão de N₂ de 2 L/min e tamanho de partícula de 1-2 mm. Os teores de celulose, hemicelulose e lignina encontrados foram de 34,44%, 23,94% e 35,89% para o tronco; 50,33%, 23,18% e 21,7% para a haste e 32,49%, 22,97% e 26% para a folha, respectivamente. Uma grande variação no teor de cinzas também foi observada, de 5,27%, 0,24% e 12,32% para tronco, haste e folha, respectivamente. Os rendimentos de bio-óleo foram maiores para tronco (40,87%) e haste (43,5%) enquanto para a folha o rendimento foi de apenas 16,58%. O rendimento de bio-óleo mais elevado pode ter ocorrido devido ao elevado teor de celulose e hemicelulose em tronco e haste. Observa-se também que o menor rendimento de bio-óleo foi obtido na biomassa com o teor de cinza mais elevado.

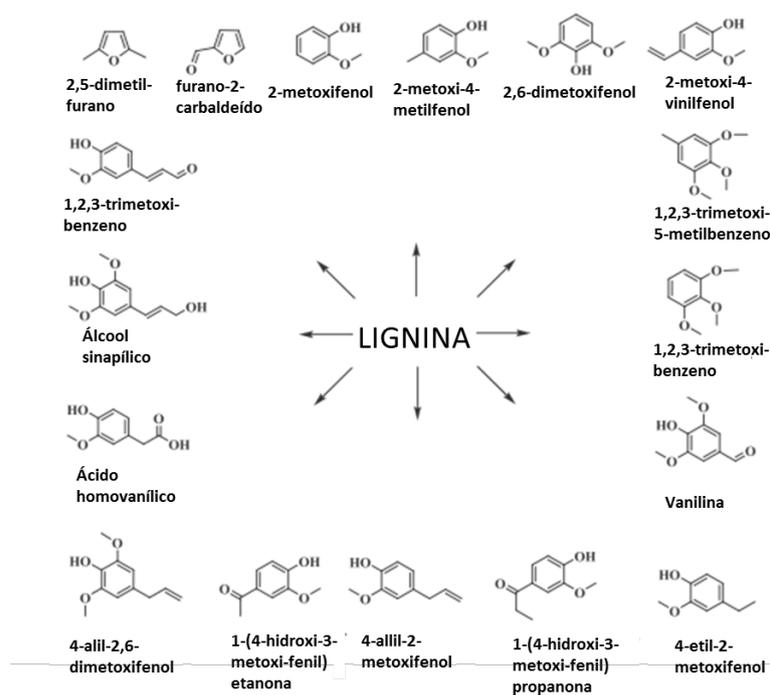
Os teores de matéria volátil e carbono fixo também são indicadores de rendimentos do produto de pirólise. Em geral, a biomassa com alta concentração de material volátil produz grandes quantidades de bio-óleo e bio-gás, enquanto que o carbono fixo aumenta a produção de carvão (Jahirul *et al.*, 2012). Jung, Kang e Kim (2008) mostraram um maior rendimento de bio-óleo na biomassa com maior conteúdo de material volátil na pirólise de palha de arroz e serragem de bambu.

A cinza também tem um papel significativo na determinação dos rendimentos de produtos de pirólise de biomassa. Um alto teor de cinzas contribui para diminuir a produção de bio-óleo e aumentar a produção de carvão (Venderbosch; Prins, 2010). Abdullah e Gerhauser (2008) mostraram que a redução do teor de cinzas, de 5,36% para 1,03%, aumentou a produção de bio-óleo de 34,71% para 61,34%, respectivamente. Lee *et al.* (2013) verificaram, em diferentes biomassas, que a produção de bio-óleo foi menor na biomassa com maior teor de cinzas, o que levou a um maior rendimento de carvão. Pattiya e Suttibak (2012) também observaram um maior rendimento de bio-óleo na biomassa com maior quantidade de

matéria volátil e menor teor de cinzas. Razuan *et al.* (2010), Fu *et al.* (2010), Maddi, Viamajala e Varanasi (2011), Pattiya, Sukkasi e Goodwin (2012) e Abnisa *et al.* (2013) observaram o mesmo comportamento. Alvarez *et al.* (2015) realizaram a pirólise de lodo de esgoto e biomassa lignocelulósica. A principal diferença na distribuição dos produtos é o alto rendimento de carvão produzido com o lodo de esgoto devido ao seu alto teor de cinzas. Lim *et al.* (2016) observaram que a alta concentração de matéria volátil, carbono, nitrogênio e menores teores de umidade, cinza, carbono fixo, hidrogênio, oxigênio e enxofre na biomassa resultam em maior rendimento de bio-óleo.

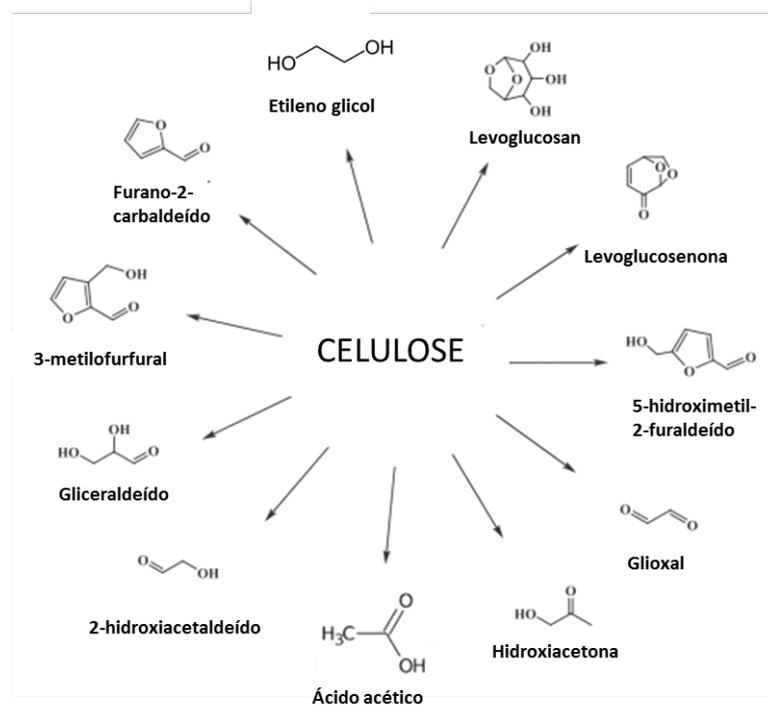
A diferença estrutural entre os principais componentes da biomassa também causa variação na composição dos produtos de pirólise. Stefanidis *et al.* (2014) estudando a pirólise da celulose, hemicelulose e lignina, separadamente, mostraram que os principais produtos líquidos da celulose são açúcares, principalmente levoglucosano e, em baixas concentrações, fenóis simples, cetonas, aldeídos e álcoois. O bio-óleo da lignina é composto de fenóis complexos com alta massa molar, quase exclusivamente. Os principais produtos do bio-óleo da hemicelulose são cetonas, fenóis, ácidos e aldeídos. Chang *et al.* (2016), estudando pirólises de diferentes biomassas, descobriram que o fenol é um componente principal do bio-óleo da casca de palma, obtido principalmente a partir da decomposição do álcool p-cumarílico obtido da lignina. O ácido acético foi o principal produto do bio-óleo da palha de trigo e bio-óleo de serragem de pinus, como resultado do craqueamento dos componentes de celulose e hemicelulose das matérias-primas. Algumas das substâncias presentes no bio-óleo derivados de lignina, celulose e hemicelulose são mostrados nas Figuras 5, 6 e 7.

Figura 5 - Produtos obtidos da degradação da lignina



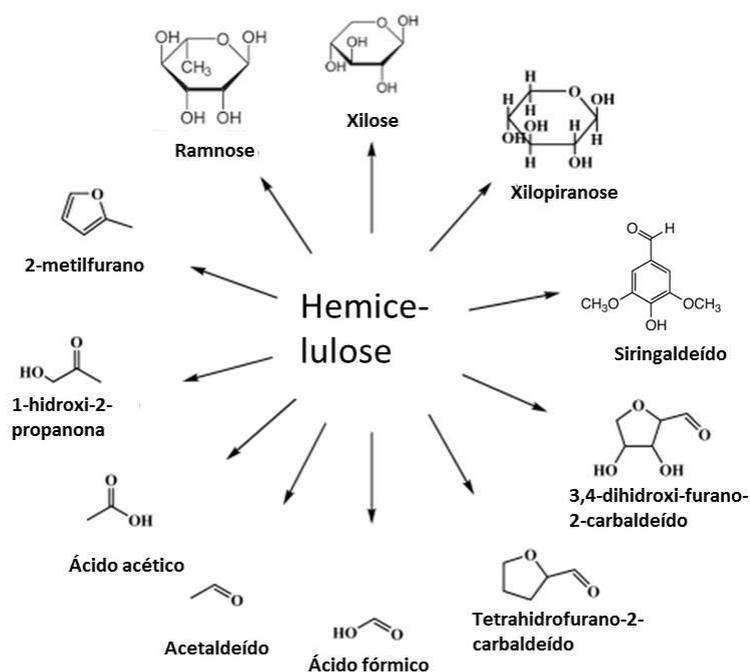
Fonte: Adaptado de Kanaujia *et al.*, 2014.

Figura 6 - Produtos obtidos da degradação da celulose



Fonte: Adaptado de Kanaujia *et al.*, 2014.

Figura 7 - Produtos obtidos da degradação da hemicelulose



Fonte: Adaptado de Kanaujia *et al.*, 2014.

A caracterização do bio-óleo é necessária para determinar a melhor aplicação deste produto. Um bio-óleo com maior porcentagem de fenol pode ser aplicado em resinas fenólicas para a produção de produtos químicos. O bio-óleo com uma quantidade significativa de ácidos carboxílicos de cadeia longa, cetonas e hidrocarbonetos é mais atraente para uso como combustível líquido após o processo de melhoramento (Lazzari *et al.*, 2016).

Outro fator crítico no processo de pirólise é a umidade presente na biomassa. O bio-óleo contém quantidades variáveis de água, dependendo da biomassa e de como foi produzido (Bridgwater, 2012). A água presente no líquido é o resultado da umidade nas matérias-primas e das reações ocorridas durante o processo de pirólise (Cardoso; Machado; Caramão, 2016; Tsai; Lee; Chang, 2006). Uma especificação padrão de material de alimentação é um máximo de 10% de umidade da biomassa para minimizar o teor de água no óleo. Um alto teor de umidade no bio-óleo pode reduzir o valor calorífico, melhorar a estabilidade e diminuir a viscosidade do bio-óleo. A viscosidade de um combustível líquido desempenha um papel significativo no projeto e na operação do sistema de injeção de combustível, bem como nas propriedades de combustão do combustível (Qiang; Xu Lai; Xi Feng, 2008).

2.5.7 Interação entre as variáveis

Apesar do grande número de pesquisas disponíveis sobre pirólise, poucos estudiosos analisaram o efeito combinado dos parâmetros no processo. Na maior parte dos artigos, o efeito dos parâmetros é analisado individualmente, mantendo sempre as outras variáveis do processo constantes ou, muitas vezes, não especificadas.

Isa *et al.* (2011) empregaram um planejamento composto central usando a metodologia de superfície de resposta para estudar o efeito de variáveis de processo no rendimento de bio-óleo obtido da casca de arroz em um reator de leito fixo. As variáveis foram analisadas nas faixas de 300 a 700 °C, 25 a 125 °C/min, 0,3 a 1,5 mm e 1 a 5 min para temperatura de pirólise, taxa de aquecimento, tamanho de partícula e tempo de reação, respectivamente. O resultado da análise mostrou que a produção de bio-óleo foi afetada por todas as variáveis. Sendo um efeito positivo observado em temperatura e taxa de aquecimento, com um maior efeito da temperatura no processo. O tamanho de partícula e tempo de reação exerceram um efeito negativo no rendimento de bio-óleo. Os efeitos das interações entre as variáveis não foram significativos na pirólise.

Ellens e Brown (2012) analisaram a pirólise de carvalho em reator de queda livre. Foram analisados temperatura (450 a 650 °C), tamanho de partícula (0,2 a 0,6 mm), vazão de gás (1 a 5 L/min) e taxa de alimentação de biomassa (1 a 2 kg/h). Um planejamento composto central foi realizado para a obtenção de dados para posterior otimização do processo. Não foram observados efeitos significativos da vazão de gás e das interações das variáveis no processo. A temperatura e a taxa de alimentação tiveram um efeito negativo no rendimento de bio-óleo, enquanto o aumento do tamanho da partícula levou a um maior rendimento.

Kiliç, Pütün e Pütün (2014) também não encontraram efeitos significativos das interações entre as variáveis no estudo da pirólise de *Euphorbia rígida* em reator de leito fixo. As faixas analisadas foram de 300 a 700 °C, 100 a 500 °C/min e 50 a 250 mL/min para a temperatura, taxa de aquecimento e vazão de N₂, respectivamente. Neste caso todas as variáveis tiveram efeitos principais significativos no processo.

Um resultado diferente foi obtido por Saikia *et al.* (2018) estudando a pirólise de *Saccharum ravannae L.* também em reator de leito fixo. Neste caso um planejamento foi utilizado para estudar as variáveis temperatura, na faixa de 300 a 550 °C, taxa de aquecimento, na faixa de 20 a 60 °C/min e vazão de N₂, na faixa de 70 a 250 mL/min. A

temperatura e a taxa de aquecimento tiveram um efeito positivo no rendimento de bio-óleo. Além disso, as interações entre vazão de N₂ com temperatura e taxa de aquecimento também foram significativas no rendimento do líquido.

Abnisa, Wan Daud e Sahu (2011) relataram o efeito de quatro parâmetros de processo, sendo temperatura (400 a 800 °C), vazão de N₂ (1 a 5 L/min), tamanho de partícula (0,5 a 2,5 mm) e tempo de reação (30 a 150 min), na otimização da pirólise de casca de palma em reator de leito fixo. Entre os efeitos principais somente o tempo de reação não foi significativo no rendimento de bio-óleo. A vazão de N₂ foi encontrada como o fator mais significativo na otimização da produção de bio-óleo, tendo um efeito negativo no processo. O aumento da temperatura também levou a uma redução no rendimento de líquidos. Em relação as interações, apenas a interação entre a temperatura e o tamanho das partículas não foi significativa para a pirólise.

Dhanavath *et al.* (2019) realizaram a pirólise de torta de semente de nim em reator de leito fixo para estudar os efeitos da temperatura (450 a 575 °C), tempo de operação (30 a 60 min) e vazão de N₂ (0,1 a 0,5 L/min) no rendimento de líquido. A metodologia de superfície de resposta baseada no planejamento Box-Behnken foi usada para determinar as condições operacionais que maximizam o rendimento. A análise dos resultados mostrou que as três variáveis experimentais foram significantes. Sendo que todas possuem um efeito positivo no produto final. Da mesma forma, a interação entre temperatura e tempo de retenção também indicou efeito significativo no rendimento do bio-óleo.

Singh, Chakraborty e Mondal (2020) investigaram a pirólise de *Acacia nilotica* torrada em um reator de leito fixo. O processo foi realizado utilizando-se um planejamento composto central a fim de obter o máximo rendimento de óleo de pirólise. As variáveis independentes controladas nos experimentos foram a temperatura (400 a 600 °C), tempo de operação (30 a 90 min), taxa de aquecimento (15 a 40 °C/min) e vazão de N₂ (40 a 80 mL/min). A análise mostrou que o rendimento de óleo de pirólise aumentou com a temperatura e com a taxa de aquecimento e diminuiu com o aumento do tempo de operação. Sendo que a interação entre temperatura e taxa de aquecimento e do tempo de operação com a taxa de aquecimento também foram significativas no processo. Mostrando, mais uma vez, que a interação das variáveis é importante para o processo.

O planejamento composto central também foi empregado em Arazo *et al.* (2017) para otimizar a pirólise de lodo de esgoto, em reator de leito fluidizado, analisando os efeitos da temperatura (425 a 550 °C), tamanho de partícula (0,2 a 1 mm) e tempo de residência do

vapor (1 a 2 s) no rendimento, no poder calorífico e no teor de umidade do bio-óleo produzido. Todos os efeitos individuais e de interações foram considerados significativos para o rendimento. O tamanho de partícula e temperatura tiveram um efeito positivo, sendo um maior efeito observado em tamanho de partícula. No poder calorífico, somente a interação do tempo de residência com a temperatura tem um efeito insignificante no processo. Analisando o teor água no bio-óleo obtido, apenas a temperatura foi significativa no processo, tendo um efeito negativo no produto.

Assim como nos trabalhos daqueles pesquisadores que analisam as variáveis do processo individualmente, os estudos de otimização mostram diferenças na importância de cada variável no processo de pirólise, indicando a influência de outros parâmetros como tipo de reator e composição da biomassa. Além disso, os pesquisadores de otimização também estão focados na obtenção de maiores rendimentos de bio-óleo, não analisando a influência das interações das variáveis nas propriedades do produto, como acidez, viscosidade, teor de oxigênio e composição que são de extrema importância para a aplicabilidade do produto.

3 MODELAGEM

3.1 Quimiometria

A quimiometria é uma sub área da química analítica que faz uso de dados de processos químicos com fundamentação matemática e estatística para, dentre outras aplicações, projetar ou selecionar procedimentos e experimentos ideais e para extrair o máximo de informações relevantes dos dados (Massart *et al.*, 1997). O termo quimiometria foi inicialmente usado na década de 1970 e tem evoluído como uma especialidade da química à medida que a necessidade de métodos estatísticos e matemáticos avançados aumenta com a sofisticação da instrumentação e dos processos químicos (Hopke, 2003).

Com o avanço das técnicas computacionais, a quimiometria se tornou uma ferramenta importante para análises rápidas de dados e para o desenvolvimento de produtos em um menor tempo. Essa ferramenta tem a capacidade de analisar e modelar uma ampla variedade de tipos de dados (Sarker; Nahar, 2015).

Muitos processos são complexos, com interações químicas, físicas e biológicas. Assim, grandes conjuntos de dados multivariados são necessários para aplicação de análises apropriadas, transformando os dados em resultados mais compactos e interpretáveis, que podem servir como base para descrições do sistema e auxiliar no desenvolvimento de modelos matemáticos para estimar os efeitos prováveis de atividades futuras ou estratégias de controle (Hopke, 2020).

Na análise de processos, a quimiometria é aplicada com sucesso para calibração multivariada, reconhecimento de padrões, modelagem e monitoramento de processos multivariados (Potyrailo, 2001). O objetivo da modelagem é extrair de informações brutas uma estimativa exata (Alexopoulos, 2010). Modelagem refere-se ao desenvolvimento de expressões matemáticas que descrevem o comportamento de uma variável de interesse, ou seja, a modelagem tenta descrever uma relação estatística entre uma variável de resposta (y) e as variáveis preditoras (x_i) (Rawlings; Pantula; Dickey, 1998). Existem vários tipos de modelos de regressão. Se a resposta for contínua e aproximadamente normal, um modelo linear geralmente se enquadra na classe de modelos que descreve bem essa resposta. Porém, muitas vezes, esse tipo de algoritmo não é satisfatório para modelar a resposta desejada, sendo

necessário a utilização de modelos mais complexos (Rawlings; Pantula; Dickey, 1998). Os modelos selecionados para a aplicação nesta pesquisa são divididos em cinco grupos: (1) Modelos Lineares (Regressão Linear Múltipla (MLR), Ridge, LASSO e Rede Elástica (Rede)); (2) Modelos não lineares (Rede Neural Artificial (ANN), *Multivariate Adaptive Regression Splines* (MARS), k-vizinhos mais próximos (k-NN), Máquinas de Vetores de Suporte com função de base radial (SVM/Br) e Máquinas de Vetores de Suporte com função polinomial (SVM/Pl)); (3) Árvores de Regressão (CART); (4) Regras de decisão (*Cubist*) e (5) Métodos Ensemble (Floresta Aleatória (RF), Árvores Extremamente Aleatórias (ExtraTree) e *Gradient Boosting Machine* (GBM)).

3.1.1 Modelos Lineares

3.1.1.1 Regressão linear múltipla

A regressão linear múltipla (MLR) é um dos métodos de regressão mais antigos e simples e é usada para estabelecer relações lineares entre variáveis independentes e a variável dependente influenciada por elas (Djuris; Ibric; Djuric, 2013). MLR pode ser escrito na Equação 1:

$$y_i = b_0 + b_1x_{i1} + b_2x_{i2} + \dots + b_kx_{ik} + e_i, \quad i = (1, 2, \dots, n) \quad (1)$$

onde y_i representa a resposta para a i -ésima amostra, b_0 representa o intercepto estimado, b_k representa o coeficiente estimado para o k -ésimo preditor, x_{ik} representa o valor do k -ésimo

preditor para a i -ésima amostra e e_i representa a fração de dados originais que não pode ser explicada pelo modelo. O número de parâmetros do modelo é $p = k + 1$.

O modelo é linear nos parâmetros e pode ou não ser linear nas variáveis x . Assim, o modelo pode incluir, por exemplo, termo quadrático, cúbico ou interações entre preditores que podem ser adicionados aos preditores originais.

MLR também pode ser escrito na forma de matriz, como na Equação 2:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2)$$

onde a resposta é escrita como vetor \mathbf{Y} ($n \times 1$); \mathbf{X} é a matriz do modelo ($n \times p$); $\boldsymbol{\beta}$ é o vetor dos parâmetros do modelo ($p \times 1$) e $\boldsymbol{\varepsilon}$ é o vetor de erros ($n \times 1$).

Na forma básica do MLR geralmente assume-se que $\boldsymbol{\varepsilon} \sim \text{NID}(0, \sigma^2)$, ou seja, os erros são normalmente e independentemente distribuídos com média zero e variância constante (Ferre, 2009; Vittinghoff *et al.*, 2007).

A estimativa dos parâmetros de regressão no modelo de Regressão Linear Múltipla é feita, geralmente, usando mínimos quadrados ordinários (OLS). Na abordagem OLS, $\boldsymbol{\beta}$ é estimado de forma que a soma dos quadrados dos resíduos (SSE-Equação 3) seja a menor possível.

$$SSE_{\text{ols}} = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p x_{ij}\beta_j \right)^2 \quad (3)$$

A solução dos parâmetros de regressão OLS é dada pela Equação 4:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}, \quad (4)$$

onde $\hat{\boldsymbol{\beta}}$ é o valor de $\boldsymbol{\beta}$ que minimiza SSE (Vittinghoff *et al.*, 2007). Sob suposições padrão, os coeficientes produzidos pela regressão de mínimos quadrados ordinários têm baixa variância (Hoerl; Kennard, 1970). Se não houver multicolinearidade entre as variáveis e o número de preditores for menor que o tamanho da amostra, é provável que o estimador OLS simples produza bons resultados (Melkumova; Shatskikh, 2017).

Quando o modelo ajusta excessivamente os dados ou quando existem problemas de colinearidade, as estimativas dos parâmetros de regressão linear podem ficar infladas. O

controle das estimativas dos parâmetros pode ser realizado adicionando-se uma penalidade à soma do quadrado dos resíduos (Kuhn; Johnson, 2013). Os modelos Ridge, LASSO e Rede Elástica foram propostos para a regularização dos parâmetros.

3.1.1.2 Ridge

Hoerl e Kennard (1970) propuseram que problemas nas estimativas de parâmetros por regressão OLS poderiam ser melhorados adicionando o parâmetro de Ridge (λ_2) à diagonal da matriz $X^T X$, formando uma nova matriz. Assim, a expressão usada para encontrar os coeficientes é mostrada na Equação 5:

$$\tilde{\beta} = (X^T X + \lambda_2 I)^{-1} X^T y, \quad (5)$$

onde I denota a matriz de identidade e $\lambda_2 \geq 0$ é um parâmetro de regularização que controla a quantidade de encolhimento.

Para uma regressão Ridge, a função perda a ser minimizada é a soma dos quadrados residuais contendo a penalidade imposta (Equação 6).

$$SSE_{ridge} = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda_2 \sum_{j=1}^p \beta_j^2 \quad (6)$$

Uma penalidade de segunda ordem (ℓ_2) é usada, a qual é igual ao quadrado dos coeficientes. Quando $\lambda_2 = 0$, a regressão Ridge é igual à regressão de mínimos quadrados. Porém, à medida que λ_2 aumenta, os parâmetros estimados diminuem para valores próximos a 0. Mesmo que algumas estimativas de parâmetros se tornem desprezíveis, o modelo não define os valores como 0 absoluto para nenhum valor da penalidade (Tibshirani, 1996). Portanto, aumentar o valor de λ_2 tende a reduzir as magnitudes dos coeficientes, mas não resulta na exclusão de nenhum deles. Isso pode criar um desafio na interpretação do modelo gerado a partir de um conjunto de dados com um grande número de preditores (James *et al.*, 2013).

3.1.1.3 LASSO

Tibshiranit (1996) propôs a técnica chamada de LASSO (*Least Absolute Shrinkage and Selection*). Este modelo usa uma penalidade semelhante à regressão de Ridge com diferenças sutis, mas importantes. No caso de LASSO, usa-se a penalidade ℓ_1 , onde o parâmetro λ_1 é multiplicado pela norma do vetor β (Equação 7).

$$SSE_{\text{lasso}} = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda_1 \sum_{j=1}^p |\beta_j| \quad (7)$$

Um efeito positivo do LASSO, ao contrário da regressão Ridge, é o fato de que o modelo obtido possui estimativas de coeficiente iguais a zero quando λ_1 é suficientemente grande. Conseqüentemente, o LASSO realiza a seleção de variáveis. Como resultado, os modelos gerados a partir do LASSO são, geralmente, mais fáceis de interpretar (James *et al.*, 2013). Para conjuntos de dados onde o número de preditores é menor que a amostra e há altas correlações entre os preditores, a regressão Ridge possui um melhor desempenho do que a regressão LASSO (Tibshiranit, 1996).

Segundo Zou e Hastie (2005) o LASSO possui algumas limitações. Para número de preditores maior que a amostra ($p > n$) o LASSO seleciona no máximo n variáveis antes de saturar. Este é um recurso limitante para um método de seleção de variável. Além disso, se houver um grupo de variáveis entre as quais as correlações de pares são muito altas o LASSO tende a selecionar apenas uma variável do grupo e não se importa com qual delas é selecionada.

Ambas regressões, Ridge e LASSO produzem um conjunto de estimativas de coeficientes cujos valores dependem de λ . A escolha do valor de λ é uma etapa crítica para ambos os métodos (Pereira; Basto; Silva, 2016). A validação cruzada é uma forma simples de determinar o melhor valor para esse parâmetro de ajuste.

3.1.1.4 Rede Elástica

A regressão de Rede Elástica é um híbrido de LASSO e Ridge, que tem se mostrado uma ferramenta valiosa para ajuste de modelo e seleção de atributos. Zou e Hastie (2005) propuseram esta técnica de regularização com o objetivo de encontrar um método que possa imitar o método de seleção de variáveis e que tenha um melhor desempenho de predição do que o LASSO. Este modelo combina os dois tipos de penalidades, conforme Equação 8:

$$SSE_{ridge} = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda_2 \sum_{j=1}^p \beta_j^2 + \lambda_1 \sum_{j=1}^p |\beta_j| \quad (8)$$

A vantagem deste modelo é que faz simultaneamente a seleção automática de variáveis e a redução contínua, podendo selecionar grupos de variáveis correlacionadas.

3.1.2 Modelos não lineares

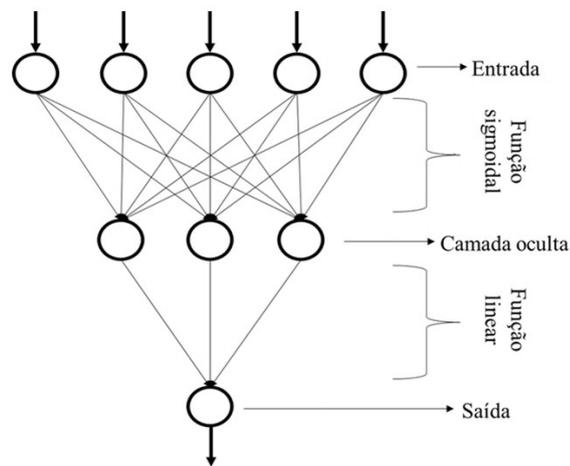
3.1.2.1 Redes Neurais Artificiais

A Rede Neural Artificial (ANN) é uma técnica de regressão não linear inspirada por teorias sobre como o cérebro funciona. Ela é projetada para simular o processo de tomada de decisão humana por meio de algoritmos, explorando estruturas ocultas nos dados (Yang; Yang, 2014). As redes neurais reúnem seu conhecimento detectando os padrões e relacionamentos nos dados (Agatonovic-Kustrin; Beresford, 2000). A resposta é modelada por um conjunto intermediário de camadas ocultas, cuja função é proporcionar complexidade e não linearidade para o modelo. Cada unidade oculta é uma combinação linear de algumas ou todas as variáveis preditoras. No entanto, essa combinação linear é normalmente transformada

por uma função não linear, sendo a função de transferência, frequentemente, uma função tipo sigmóide (Hallinan, 2013). Uma combinação linear conecta as unidades ocultas ao resultado.

Existem vários tipos de redes neurais. No *Feedforward*, a arquitetura de rede neural mais simples (Figura 8), as informações fluem unidirecionalmente da camada de entrada para a camada de saída, através da camada oculta. Em modelos mais complexos, existem arquiteturas com loops indo em ambas as direções entre as camadas ou arquiteturas com mais de uma camada de unidades ocultas.

Figura 8 - Estrutura simples da ANN



Fonte: Adaptado de Kuhn e Johnson, 2013

A ANN é afetada pela alta correlação entre as variáveis preditoras. Uma abordagem para mitigar esse problema é realizar um pré-processamento para filtrar os preditores que estão associados a altas correlações. Alternativamente, uma técnica de extração de recursos, como a análise de componentes principais, pode ser usada antes da modelagem para eliminar correlações (Kuhn; Johnson, 2013).

3.1.2.2 Spline de Regressão Adaptativa Multivariada

Spline de Regressão Adaptativa Multivariada (*Multivariate Adaptive Regression Spline* - MARS) foi introduzida por Friedman (1991). É uma técnica de regressão multivariada não paramétrica que pode explicar a relação não linear complexa entre as variáveis preditoras e a resposta e, automaticamente, escolher as variáveis que são importantes para o processo de

modelagem. O MARS é adequado para problemas de alta dimensão e não requer nenhuma suposição a priori sobre a relação entre as variáveis dependentes e independentes.

O modelo de regressão é construído ajustando funções de base a intervalos distintos das variáveis independentes, ou seja, o MARS usa uma classe específica de funções de base como preditores no lugar das variáveis de entrada originais (Andrés *et al.*, 2011; Shahsavani; Tarantola; Ratto, 2010). O método MARS usa funções de base *splines* lineares para cada entrada x_j conforme Equações 9 e 10 (Hastie; Tibshiranit; Friedman, 2008):

$$(x - t)_+ = \begin{cases} x - t, & \text{se } x > t, \\ 0, & \text{caso contrário,} \end{cases} \quad (9)$$

$$(t - x)_+ = \begin{cases} t - x, & \text{se } x < t, \\ 0, & \text{caso contrário.} \end{cases} \quad (10)$$

onde t é uma constante nomeada nó (valor limite), $t \in \{x_{1j}, x_{2j}, \dots, x_{nj}\}$, $j = 1, 2, \dots, p$, x é a variável de entrada e y é a variável de saída. A forma geral do modelo MARS pode ser representada pela Equação 11 (Mehdizadeh; Behmanesh; Khalili, 2017):

$$\hat{y} = \beta_0 + \sum_{j=1}^M \beta_j B_j(x) \quad (11)$$

onde M é o número de termos, B_j é a j -ésima função de base, que pode ser uma única função ou um produto de duas ou mais funções de base, e β_j s são os coeficientes estimados minimizando o SSE.

Este modelo funciona no procedimento *forward* e *backward* para construir o modelo final. Inicialmente implementa-se um procedimento de avanço para adicionar todas as funções de base possíveis ao modelo (Kisi; Parmar, 2016). Nesta fase, gera-se um modelo muito complexo e com sobreajuste, possuindo baixa capacidade de predição. Na próxima etapa, as funções base que possuem menor contribuição no aumento do SSE são removidas usando um procedimento *backward*, gerando um melhor modelo para cada número de termos. Para determinar quais funções de base devem ser incluídas no modelo, o MARS utiliza validação cruzada generalizada (Andrés *et al.*, 2011). O número de termos a ser removido pode ser definido manualmente ou tratado como um parâmetro de ajuste determinado por reamostragem. Outro parâmetro de ajuste associado ao modelo MARS é o grau dos recursos

que são adicionados ao modelo. Assim, o MARS pode construir modelos onde os recursos substitutos dos preditores originais envolvem vários preditores de uma vez (Kuhn; Johnson, 2013).

O MARS apresenta como vantagens a seleção de variáveis, fácil entendimento da relação entre preditores e resposta e requer pouco pré-processamento do conjunto de dados. As variáveis fortemente correlacionadas podem dificultar a interpretação, mas não afetam drasticamente o modelo (Kuhn; Johnson, 2013).

3.1.2.3 k-Vizinhos mais próximos

O método k-vizinhos mais próximos (k-NN) é usado para problemas de classificação e regressão. A abordagem prevê uma nova amostra usando as k amostras mais próximas do conjunto de treinamento. Sua construção é baseada exclusivamente em amostras individuais deste conjunto de dados. Para prever uma nova amostra para regressão, o método identifica os k-NNs dessa amostra no espaço do preditor. A resposta prevista para a nova amostra é, então, a média das respostas dos k vizinhos (Kuhn; Johnson, 2013).

O método k-NN depende de como é definida a distância entre as amostras. A distância euclidiana é a métrica mais comumente usada e é definida pela Equação 12.

$$d(a, b) = \left(\sum_{j=1}^p (x_{aj} - x_{bj})^2 \right)^{\frac{1}{2}}, \quad (12)$$

onde x_a e x_b são duas amostras individuais.

Existem muitas outras métricas de distância, como a distância de Manhattan, que é uma métrica comum usada para amostras com preditores binários, e as distâncias de Tanimoto e Hamming que são mais apropriadas para tipos específicos de preditores (Kuhn; Johnson, 2013).

O método requer apenas a escolha do número de vizinhos (k) a ser considerado ao fazer a regressão. Normalmente, k é escolhido como o valor que minimiza o erro usando alguns dados de validação independente ou por procedimentos de validação cruzada (Ebbels,

2007). O número de vizinhos pode ter uma grande influência no poder preditivo. Uma alta variância do modelo ocorre com pequenos valores de k . Por outro lado, o aumento de k pode aumentar o viés (Kuhn; Johnson, 2013).

Apesar de sua simplicidade, o método tem uma base teórica sólida e muitas vezes pode superar métodos muito mais sofisticados. Suas desvantagens são o alto custo computacional em tempo e memória, uma vez que todas as amostras são analisadas para encontrar os vizinhos mais próximos. O cálculo da distância se torna mais difícil à medida que o conjunto de dados de treinamento aumenta (Gonzalez-Sanchez; Frausto-Solis; Ojeda-Bustamante, 2014).

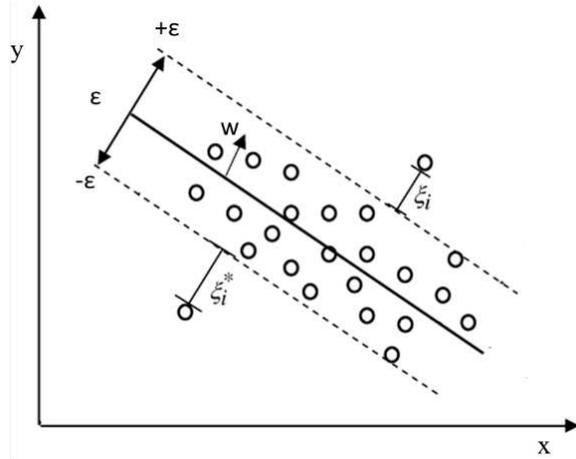
O método k -NN pode ter um desempenho preditivo ruim quando a estrutura do preditor local não é relevante para a resposta. Portanto, a remoção destes preditores é uma etapa de pré-processamento importante para o k -NN. Outra abordagem para aumentar a precisão da previsão é ponderar a contribuição do vizinho, para a previsão de uma nova amostra, com base em sua distância. Nesse caso, a amostra de treinamento mais próxima da nova amostra tem maior contribuição para a resposta prevista (Kuhn; Johnson, 2013).

3.1.2.4 Máquina de Vetores de Suporte

A Máquina de Vetores de Suporte (SVM) é uma técnica de modelagem poderosa e altamente flexível. O SVM corresponde a um método linear em um espaço de recursos de alta dimensão não linearmente relacionado ao espaço de entrada. O SVM utiliza o princípio de minimização de risco estrutural, que limita a taxa de erro nos dados de teste. Assim, o SVM possui um bom desempenho de generalização em uma ampla gama de problemas (Min; Lee, 2005).

Esta ferramenta foi desenvolvida inicialmente para resolver problemas de classificação e adaptada para regressão. Suponha os dados de treinamento $\{(x_1, y_1), \dots, (x_i, y_i)\} \subset X \times \mathbb{R}$. Na técnica de regressão ε -insensível o objetivo é encontrar um hiperplano ótimo que tenha no máximo ε desvio dos alvos realmente obtidos y_i , para todos os dados de treinamento, e ao mesmo tempo tenha o mínimo de oscilações possível (Smola; Scholkopf, 2004). A Figura 9 retrata a situação graficamente.

Figura 9 - SVM para o problema de regressão linear.



O SVM para regressão é formulado como minimização do seguinte problema de otimização da Equação 13 (Schölkopf; Smola, 2002):

$$\text{Min. } f(x_i, \xi_i, \xi_i^*) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \quad (13)$$

$$\text{Sujeito a: } \begin{cases} y_i - f(x_i) \leq \varepsilon + \xi_i \\ f(x_i) - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

Onde ξ_i e ξ_i^* são variáveis de folga que medem o desvio das amostras de treinamento fora da zona insensível a ε e C é o parâmetro de penalidade que determina o *trade-off* entre erro de treinamento e a complexidade do modelo. Em outras palavras, a constante $C > 0$ determina a compensação entre a planura de $f(x)$ e o valor até o qual os desvios maiores que ε são tolerados. Isso corresponde a lidar com o chamado função de perda ε -insensível $|\xi|_\varepsilon$ descrita pela Equação 14 (Smola; Scholkopf, 2004):

$$|\xi|_\varepsilon := \begin{cases} 0, & \text{se } |\xi| \leq \varepsilon \\ |\xi| - \varepsilon, & \text{caso contrário,} \end{cases} \quad (14)$$

Na previsão SVM linear, a estimativa de parâmetro para uma nova amostra pode ser escrita como função de um conjunto de parâmetros α_i e dos pontos de dados do conjunto de treinamento, conforme Equação 15 (Smola; Scholkopf, 2004):

$$f(x) = (w \cdot x + b) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \langle x_i, x \rangle + b \quad (15)$$

Onde α_i e α_i^* são multiplicadores de Lagrange ($\alpha_i \geq 0$, $\alpha_i^* \leq C$).

Para alguma porcentagem das amostras do conjunto de treinamento, os parâmetros α_i serão exatamente zero, indicando que eles não têm impacto na equação de predição. Então, apenas um subconjunto de pontos de dados do conjunto de treinamento, onde $\alpha \neq 0$, são necessários para a previsão. Essas amostras são chamadas de vetores de suporte (Kuhn; Johnson, 2013).

Em muitos problemas de regressão os dados possuem relações não lineares. Nesses casos, os SVM tentam encontrar um hiperplano ótimo que minimize o erro de predição dos dados usando a função de transformação não linear, $K(x_i, x) = \Phi(x_i) \cdot \Phi(x)$, conhecida como função kernel (Ergun *et al.*, 2010). Dessa maneira, a equação acima pode ser reescrita de forma mais geral na Equação 16:

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x_i, x) + b \quad (16)$$

Existem vários tipos de funções kernel que podem ser usadas para generalizar o modelo de regressão e englobar funções não lineares dos preditores, entre elas temos as descritas na Equações 17, 18 e 19:

$$\text{Função polinomial} = (\Phi(x_i'x) + 1)^{\text{degree}} \quad (17)$$

$$\text{Função base radial} = \exp(-\sigma \|x_i - x\|^2) \quad (18)$$

$$\text{Função tangente hiperbólica} = \tanh(\Phi(x_i'x) + 1) \quad (19)$$

Onde Φ e σ são parâmetros de escala.

A escolha da função kernel depende do problema. Algumas das funções do kernel possuem parâmetros extras. Assim, o desempenho do SVM depende de um conjunto de parâmetros, sendo eles: o parâmetro C, o tipo de kernel e os parâmetros presentes em cada kernel (Min; Lee, 2005).

3.1.2.5 Regras e árvores de decisão

Os modelos baseados em árvore e regras geram um conjunto de condições que são fáceis de interpretar e implementar. Esses modelos podem lidar com muitos tipos de preditores (esparços, distorcidos, contínuos, categóricos etc.) sem a necessidade de pré-processamento. Além disso, esses modelos não exigem o conhecimento prévio do relacionamento dos preditores com a resposta. Por outro lado, os modelos baseados em árvores ou regras únicas têm pontos fracos, como instabilidade e baixo desempenho preditivo. Para combater esses problemas, foram desenvolvidos métodos *Ensemble* que combinam muitas árvores (ou modelos baseados em regras) em um único modelo. Estes conjuntos tendem a ter um desempenho preditivo muito melhor que modelos com uma única árvore ou regra (Kuhn; Johnson, 2013).

Os modelos baseados em árvore consistem de uma ou mais instruções *if-then* para os preditores, que particionam os dados. Dentro dessas partições, um modelo é usado para prever o resultado. Por exemplo, uma árvore muito simples pode ser definida abaixo, onde existem duas divisões dos dados em três nós terminais, ou seja, três regiões onde a resposta é prevista por um único número:

```
if Predictor A  $\geq$  2, then
    | if Predictor B  $\geq$  100 then Outcome = 1.5
    | else Outcome = 6
else Outcome = 3
```

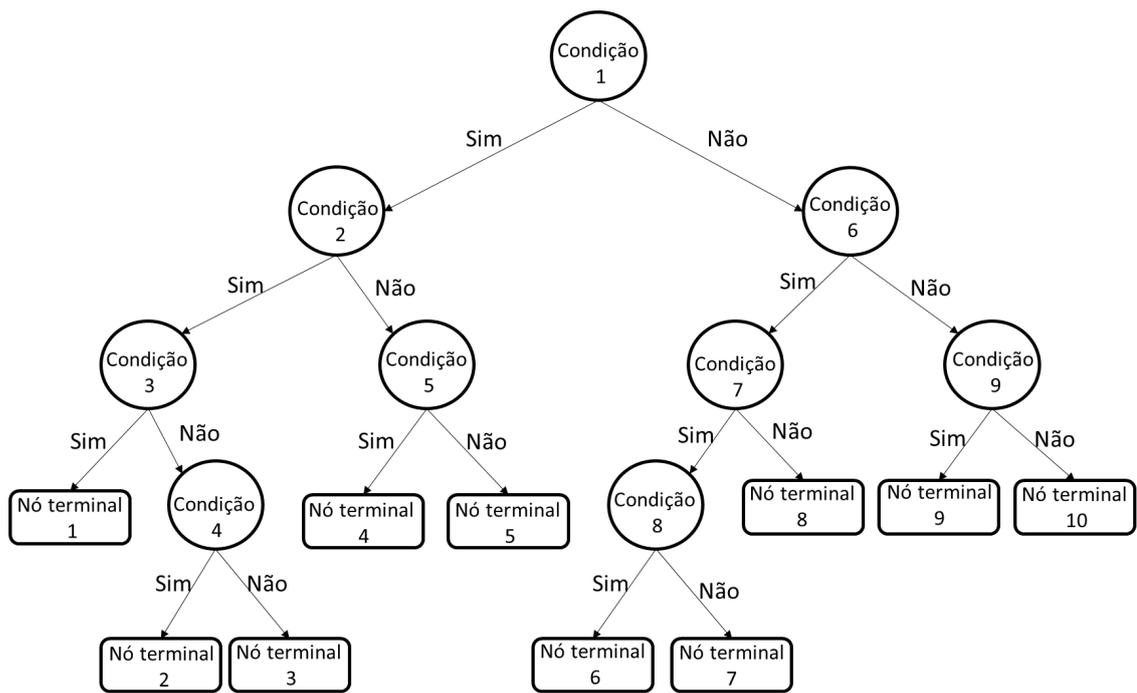
Uma regra é definida como um caminho distinto em uma árvore. Ou seja, a regra é um conjunto de condições *if-then* reduzidas a condições independentes. Para o exemplo acima, haveria três regras:

```
if Predictor A  $\geq$  2 and Predictor B  $\geq$  100 then Outcome = 1.5
if Predictor A  $\geq$  2 and Predictor B < 100 then Outcome = 6
if Predictor A < 2 then Outcome = 3
```

A Figura 10 mostra uma árvore de regressão com 10 nós terminais. Neste caso, há um total de 10 regras, ou seja, 10 caminhos possíveis de serem percorridos nesta árvore. As regras

podem ser simplificadas ou podadas, reduzindo assim a complexidade do modelo. Podem ser removidas regras inteiras, ou algumas das condições que definem a regra, que sejam redundantes ou que não contribuem para o modelo.

Figura 10 - Exemplo de árvore de regressão com 10 nós terminais



Fonte: Adaptado de Kuhn e Johnson, 2013.

3.1.2.5.1 Árvores de decisão

Existem muitas técnicas para construir árvores de regressão. Uma das técnicas mais antigas e mais utilizadas é a metodologia de árvore de classificação e regressão conhecida como CART. Neste método os dados são particionados em grupos menores que são mais homogêneos em relação à resposta. Para atingir a homogeneidade do resultado, o modelo de árvores determina o preditor a ser dividido, o valor da divisão para particionar os dados e a complexidade da árvore (Kuhn; Johnson, 2013). No modelo, o conjunto de dados, S , é dividido em grupos S_1 e S_2 de modo que a soma do quadrado dos erros seja minimizada (Equação 20):

$$SSE = \sum_{i \in S_1} (y_i - \bar{y}_1)^2 + \sum_{i \in S_2} (y_i - \bar{y}_2)^2, \quad (20)$$

onde \bar{y}_1 e \bar{y}_2 são as médias dos resultados do conjunto de treinamento dentro dos grupos S_1 e S_2 , respectivamente.

A árvore cresce dividindo-se, sucessivamente, os nós em nós-filhos direitos e esquerdos. Os nós são gerados até que nenhuma divisão adicional seja possível ou um determinado critério de terminação seja satisfeito. Devido à natureza de divisão recursiva das árvores de regressão, esse método também é conhecido como particionamento recursivo (Nayak *et al.*, 2018).

O tamanho da árvore é um parâmetro de ajuste que rege a complexidade do modelo. Uma árvore muito grande pode causar um sobreajuste dos dados, enquanto uma árvore pequena pode não capturar o relacionamento da resposta com os preditores. Existem regras de parada simples para declarar nós terminais. Normalmente, um nó é declarado terminal quando seu tamanho cai abaixo de um valor limite. Outra abordagem para determinar o número de ramos é usar um algoritmo de poda. Isso envolve o crescimento de uma grande árvore de decisão inicialmente e, em seguida, os ramos são podados usando o que é conhecido como poda de complexidade de custo (Landau; Barthel, 2010). Nesse processo é usado um parâmetro de ajuste, chamado de parâmetro de complexidade, que penaliza a taxa de erro usando o tamanho da árvore. O objetivo é encontrar o tamanho de árvore que tenha a menor SSE penalizada. Alternativamente, o modelo pode ser ajustado escolhendo o valor do

parâmetro de complexidade associado ao menor valor de RMSE. O parâmetro de ajuste governa a compensação entre o tamanho da árvore e sua adequação aos dados (Hastie; Tibshirani; Friedman, 2008; Kuhn; Johnson, 2013).

Os algoritmos de partição recursiva apresentam como vantagens a capacidade de modelar relacionamentos complexos em um formato simples, a interpretabilidade dos resultados e a flexibilidade do algoritmo (Winters-Miner *et al.*, 2015). Uma limitação do modelo é que cada nó terminal usa a média das respostas do conjunto de treinamento naquele nó para predição. Assim, esses modelos podem não fazer um bom trabalho de previsão de amostras cujos resultados verdadeiros são extremamente altos ou baixos (Kuhn; Johnson, 2013).

3.1.2.5.2 Cubist

O *Cubist* é um modelo baseado em regra proposto com base no modelo de árvore M5 de Quinlan (1992, 1993). O modelo cria inicialmente uma estrutura de árvore, onde as folhas possuem um modelo linear multivariado. Ao contrário do modelo CART, o *Cubist* prevê uma nova amostra usando um modelo, ao invés de uma média simples da resposta.

O critério de divisão dos dados, para formar as ramificações, é baseado em tratar o desvio padrão dos dados de treino que atingem o nó como uma medida de erro naquele nó, e calcular a redução esperada do erro com a subsequente divisão destes dados. O cálculo da redução esperada do erro é definida pela Equação 21 (Neupane *et al.*, 2021):

$$Redução = sd(S) - \sum_{i=1}^P \frac{n_i}{n} sd(S_i) \quad (21)$$

onde S é o conjunto dos dados de treino que atingem o nó, sd é o desvio padrão, S_1, \dots, S_P representam os subconjuntos P dos dados após a divisão e n_i é o número de amostras na partição i . Para a construção da árvore, a divisão dos dados associada a uma maior redução esperada do erro é escolhida e um modelo linear é criado. O processo de crescimento da árvore continua ao longo dos ramos até que não haja mais melhorias na redução do erro ou

não haja mais amostras suficientes para divisão. Uma vez que a árvore esteja totalmente desenvolvida, haverá um modelo linear para cada nó na árvore.

Embora inicialmente crie uma estrutura de árvore, o *Cubist* transforma cada ramificação da árvore em uma regra. Dessa forma a árvore é reduzida a um conjunto de regras, que inicialmente são caminhos do topo da árvore até a base. As regras são, então, eliminadas por meio de poda e/ou combinadas para simplificação.

Para a previsão de uma nova amostra, se o conjunto de preditores satisfizer determinada regra, a previsão é feita usando o modelo presente no nó terminal da árvore referente à ramificação que gera este conjunto de regras (Zhou *et al.*, 2019). Porém um processo de suavização, entre os modelos presentes ao longo desta ramificação, é realizado para melhorar a precisão da predição. A suavização ocorre levando-se em consideração a previsão do modelo presente no nó anterior da árvore. Dessa forma, o modelo *Cubist* emprega uma combinação dos modelos para a previsão (Kuhn; Johnson, 2013), como descrito na Equação 22:

$$\hat{y}_{par} = \alpha \hat{y}_{(k)} + (1 - \alpha) \hat{y}_{(p)}, \quad (22)$$

onde α é o coeficiente de suavização, $\hat{y}_{(k)}$ é a previsão do modelo atual e $\hat{y}_{(p)}$ é a previsão do modelo pai, localizado acima dele na árvore. Sejam $\varepsilon_{(k)}$ e $\varepsilon_{(p)}$ os resíduos dos modelos atual e modelo pai, respectivamente, o procedimento de suavização determina a covariância entre os resíduos dos modelos e a variância da diferença entre os resíduos. O coeficiente de suavização usado pelo *Cubist* é então (Equação 23):

$$\alpha = \frac{Var[\varepsilon_{(p)}] - Cov[\varepsilon_{(k)}, \varepsilon_{(p)}]}{Var[\varepsilon_{(p)} - \varepsilon_{(k)}]}, \quad (23)$$

De acordo com (Zhou *et al.*, 2019) a principal vantagem do *Cubist* é a possibilidade de criação de grupos (*committees*) de modelos, dessa forma uma série de árvores com pesos ajustados são produzidas para estabelecer o modelo *Cubist*. Nesse método, as novas amostras são previstas usando cada modelo e a previsão final é a média simples das previsões individuais dos modelos. Além disso, o *Cubist* tem a capacidade de ajustar a previsão do modelo com amostras do conjunto de treinamento usando os vizinhos mais próximos (Kuhn;

Johnson, 2013). Para a aplicação dessa ferramenta, os parâmetros números de *committees* e número de vizinhos devem ser ajustados.

3.1.2.5.3 Floresta Aleatória

A Floresta Aleatória (RF) é um método em conjunto (*Ensemble*) que permite modelar problemas de regressão e classificação. Este método foi introduzido por Breiman (2001) e é baseado em árvores de decisão, combinado com ideias de agregação e *bootstrap* (Genuer *et al.*, 2017).

Uma Floresta Aleatória, como o nome sugere, é um conjunto de árvores com cada árvore crescendo a partir de uma seleção de amostras obtidas por *bootstrap*. Na regressão, a previsão da nova amostra é a média, não ponderada, dos valores obtidos em cada árvore do conjunto.

No *bootstrap* as amostras para treinar as árvores são selecionadas aleatoriamente do conjunto de dados originais com reposição. Dessa maneira, o conjunto de amostras obtidas possui o mesmo tamanho que o conjunto de dados original, porém, algumas amostras serão representadas várias vezes na seleção do *bootstrap*, enquanto outras não serão selecionadas, são as amostras conhecidas por *out-of-bag*. As amostras *out-of-bag* são extremamente úteis para estimar o erro de generalização e a importância da variável (Kuhn; Johnson, 2013).

Para um vetor $x = (x_1, \dots, x_p)$, que representa as variáveis de entrada, e uma variável y , que representa a resposta, o objetivo da RF é encontrar uma função de predição $f(x)$ para prever y . As RFs utilizam um grupo de “aprendizes básicos” que são combinados para fornecer a função de predição ($f(x)$). Para isso o modelo utiliza, como aprendizes, árvores de regressão, $h_j(x, \Theta_j)$ onde Θ_j é uma coleção de variáveis aleatórias e é independente para $j = 1, \dots, J$ (Equação 24) (Breiman, 2001; Cutler; Cutler; Stevens, 2011):

$$f(x) = \frac{1}{J} \sum_{j=1}^J h_j(x, \Theta_j) \quad (24)$$

A construção das árvores na Floresta Aleatória tem duas diferenças principais em relação às árvores CART. Primeiro, cada árvore é ajustada a uma amostra de *bootstrap* independente em vez de todo conjunto de dados originais como acontece no CART. Em segundo lugar, ao dividir um nó, a melhor divisão é encontrada em um subconjunto selecionado aleatoriamente de m variáveis preditoras em vez de todos os preditores, de forma independente em cada nó. Além disso, as árvores não passam pelo processo de poda (Cutler; Cutler; Stevens, 2011).

Existem três parâmetros que podem ser ajustados para fornecer maior precisão na Floresta Aleatória, sendo o número de preditores selecionados aleatoriamente em cada divisão, ($mtry$); tamanho da árvore, medido pelo menor tamanho de nó para divisão ou o número máximo de nós terminais; o número de árvores na floresta ($ntree$). Quanto maior a floresta, mais carga computacional será necessária para treinar e construir o modelo (Kuhn; Johnson, 2013).

Do ponto de vista computacional, a Floresta Aleatória é atraente porque é relativamente rápida de treinar e prever; possui uma estimativa de erro de generalização; pode ser usada diretamente para problemas de alta dimensão e pode ser implementada em paralelo (Cutler; Cutler; Stevens, 2011).

3.1.2.5.4 Árvores Extremamente Aleatórias

Árvores Extremamente Aleatórias (ExtraTree), propostas por Geurts, Ernst e Wehenkel (2006), é um método em conjunto baseado em árvore para problemas de classificação e regressão.

O ExtraTree é uma variante do RF com mais randomização em cada etapa. O algoritmo randomiza as escolhas de variáveis e pontos de corte enquanto divide o nó da árvore (Sharma *et al.*, 2019). Ao contrário do RF, onde as variáveis são divididas com base na divisão ótima e o conjunto de treinamento é escolhido por *bootstrap*, os critérios de divisão do ExtraTree são aleatórios e todo o conjunto de treinamento é considerado (Geurts; Ernst; Wehenkel, 2006). A inclusão de mais uma etapa de randomização reduz a variância da Floresta Aleatória e fornece uma melhor compensação entre viés e variância. Além da exatidão, o principal ponto forte do algoritmo resultante é a eficiência computacional.

3.1.2.5.5 Gradient Boosting Machine

O *Gradient Boosting Machine* (GBM) é um algoritmo de aprendizado de máquina extremamente popular produzido por Friedman (2002), que engloba tanto a classificação quanto a regressão.

Assim como na Floresta Aleatória, a previsão do GBM é baseada em um conjunto de modelos fracos, os aprendizes base. As árvores de regressão costumam ser mais usadas como o aprendiz no GBM por possuírem excelentes características. Primeiro, elas têm a flexibilidade de se tornarem modelos fracos simplesmente restringindo sua profundidade. Em segundo lugar, árvores podem ser facilmente adicionadas para gerar uma previsão. Em terceiro, as árvores podem ser geradas muito rapidamente (Kuhn; Johnson, 2013).

Ao contrário das Florestas Aleatórias, no GBM as árvores contribuem de forma desigual para o modelo final e dependem das árvores anteriores, com cada árvore aprendendo e melhorando com a anterior (Boehmke; Greenwell, 2019).

A ideia principal do GBM é adicionar novos modelos, ou seja, novas árvores, em sequência, com o objetivo de encontrar um modelo aditivo que reduza o erro de previsão da iteração anterior. O algoritmo é normalmente inicializado com uma estimativa da resposta, como sua média no conjunto de treinamento. Então, o resíduo é calculado e um modelo é então ajustado ao resíduo, minimizando a função perda. O modelo atual é adicionado ao modelo anterior e o procedimento continua por um número especificado de iterações.

Quando a árvore de regressão é usada como o aluno básico, o GBM tem dois parâmetros de ajuste: profundidade da árvore e número de iterações, ou seja, o número de árvores incluídas sequencialmente no modelo. Em geral, árvores muito profundas tendem a causar sobreajuste. Por outro lado, se a estrutura for muito simples, o modelo não consegue capturar as interações entre as variáveis de entrada. Para Hastie, Tibshirani e Friedman (2008) as árvores com números de nós terminais entre 4 e 8 produzem bons resultados para o modelo. Já o número de iterações pode ser obtido por dados de validação ou validação cruzada.

O GBM é suscetível ao sobreajuste. Para evitar esse ajuste excessivo aos dados, uma taxa de aprendizagem (λ) pode ser usada para restringir o processo de aprendizagem. Nesse caso, apenas uma fração do valor previsto atual é adicionada ao valor predito da iteração anterior. Geralmente, quanto menor λ , mais exato o modelo pode ser, porém mais árvores são

necessárias na sequência (Boehmke; Greenwell, 2019; Landry *et al.*, 2016). O algoritmo do GBM pode ser generalizado conforme as Equações 25, 26, 27 e 28 (Mayrink, 2016):

$$\hat{y}_i^{(0)} = f_0(x_i) = \bar{y} \quad (25)$$

$$\hat{y}_i^{(1)} = f_0(x_i) + \lambda f_1(x_i) = \hat{y}_i^{(0)} + \lambda f_1(x_i) \quad (26)$$

$$\hat{y}_i^{(2)} = f_0(x_i) + \lambda f_1(x_i) + \lambda f_2(x_i) = \hat{y}_i^{(1)} + \lambda f_2(x_i) \quad (27)$$

...

$$\hat{y}_i^{(M)} = \sum_{m=0}^M f_m(X) = \hat{y}_i^{(M-1)} + \lambda f_M(X) \quad (28)$$

na qual $\hat{y}_i^{(M)}$ representa o valor estimado da resposta para a i-ésima observação do conjunto de treinamento, após a m-ésima iteração do algoritmo.

3.2 Estudos de modelagem do processo de pirólise

Sun *et al.* (2016) utilizaram o modelo Rede Neural Artificial (ANN) para simular e prever a seletividade de produtos gasosos na pirólise de serragem de pinus. Os dados experimentais foram obtidos em reator em batelada. As variáveis controladas no processo, ou seja, variáveis preditoras do modelo, foram a temperatura de operação, tamanho de partícula da biomassa e velocidade espacial sobre os produtos da pirólise. As variáveis de saída estudadas foram a seletividade de H₂, CO, CH₄ e CO₂ no gás de pirólise. A modelagem mostrou que o modelo, utilizando uma função de transferência sigmoide na camada oculta com sete neurônios, ofereceu uma boa concordância entre os resultados experimentais e preditos no processo de pirólise.

Neste estudo, a temperatura apresentou uma maior importância na seletividade de todos os produtos gasosos. O tamanho de partícula foi identificado como o segundo parâmetro de maior importância na seletividade de H₂, CO e CH₄.

Cao, Xin e Yuan (2016) realizaram uma modelagem para previsão da produção de carvão obtido da pirólise de esterco de gado. Para a obtenção dos dados foram realizados experimentos em reator batelada, sendo obtido um conjunto 33 amostras. Nesta pesquisa,

cinco parâmetros foram considerados, sendo eles, temperatura de pirólise, taxa de aquecimento, tempo de retenção, teor de umidade e massa da amostra.

O modelo de Rede Neural Artificial (ANN) e o modelo de Máquina de Vetores de Suporte de mínimos quadrados (LS-SVM) foram desenvolvidos. Os parâmetros estatísticos dos modelos mostraram que o modelo LS-SVM possui uma maior capacidade preditiva do rendimento de carvão, com R^2 de 0,96 e RMSE de 0,36 g, enquanto para o modelo ANN esses valores foram de 0,80 e 0,83 g, respectivamente.

Chen *et al.* (2018) estudaram modelos de previsão para a distribuição de produtos de pirólise de diferentes biomassas. Foram coletados dados experimentais da literatura com um total de 175 amostras para o estudo do bio-óleo e 166 amostras para carvão e biogás. Os dados foram obtidos de reatores operando em leito fluidizado borbulhante. As variáveis de entrada referentes às propriedades da biomassa incluíram umidade, cinzas, carbono, hidrogênio, nitrogênio e oxigênio. As variáveis de tamanho de partícula, número de fluidização e temperatura de pirólise foram as variáveis de processo analisadas neste estudo. O número de fluidização é o número de vezes em que a velocidade mínima de fluidização é aplicada no sistema.

Para a previsão foram desenvolvidos modelos de Rede Neural Artificial e Máquina de Vetor de Suporte com função base radial (SVM/Br). Para a predição do rendimento de carvão o R^2 do modelo ANN foi de 0,83. Os valores de MAE e RMSE foram de 2,53 e 3,35, respectivamente. O algoritmo SVM/Br mostrou um desempenho superior com um valor de R^2 de 0,90, MAE de 1,73 e RMSE de 2,54.

Assim como na modelagem do rendimento de carvão, o modelo SVM/Br também apresentou melhor capacidade preditiva comparado ao ANN no estudo com biogás. As métricas de desempenho obtidas para o ANN foram R^2 de 0,82, RMSE de 4,50 e MAE de 3,16, enquanto no modelo SVM/Br esses valores foram superiores, alcançando o valor de 0,89 para R^2 , 3,34 para RMSE e 2,53 para MAE.

Na avaliação da capacidade de previsão do rendimento de bio-óleo também verificou-se a superioridade do modelo SVM/Br. O R^2 do modelo ANN foi de 0,80 enquanto no modelo SVM/Br esse valor foi de 0,89. Adicionalmente, os valores de MAE e RMSE também confirmaram a superioridade de predição do modelo SVM/Br com valores de 2,25 e 2,89, respectivamente.

Os resultados encontrados para a previsão do rendimento dos produtos de pirólise foram aproximadamente semelhantes, porém, o modelo obtido para o rendimento de bio-óleo

teve um melhor ajuste nos dados. A precisão de previsão dos modelos ANN e SVM pode ser melhorada por meio da correta determinação dos parâmetros, como a função kernel utilizada no modelo SVM. Outro parâmetro que pode melhor controlado é o tempo de residência do vapor na zona de reação. Um maior número de fluidização significa menor tempo de residência do vapor, porém, se um mesmo número de fluidização for utilizado em reatores de diferentes tamanhos o tempo de residência do vapor não será o mesmo nos dois processos. Portanto, devido a importância da variável tempo de residência do vapor no processo de pirólise, a utilização desta variável em substituição ao número de fluidização poderia levar a modelos com melhor capacidade preditiva.

Merdun e Sezgin (2018) reuniram dados de estudos experimentais de pirólise publicados na Turquia entre os anos de 2001 e 2015. O banco de dados utilizado para modelar o rendimento dos produtos de pirólise incluiu 72 amostras com 11 variáveis, sendo elas umidade, volátil, carbono fixo, cinzas, carbono, hidrogênio, oxigênio, nitrogênio, poder calorífico superior, taxa de aquecimento e temperatura de pirólise.

Dois tipos de arquiteturas de Rede Neural Artificial foram utilizados, a rede *feed-forward* e a rede *cascade-forward*, para modelar rendimentos de carvão, bio-óleo e biogás. Em ambos os métodos de ANN, a abordagem de modelagem hierárquica foi utilizada adicionando variáveis uma a uma ao modelo para determinar como a variância na variável de saída pode ser explicada pela adição de uma ou mais variáveis de entrada. As variáveis foram adicionadas na ordem como aparecem anteriormente gerando, para cada estrutura de rede, 11 modelos para cada uma das três respostas estudadas.

Os resultados mostraram que ambos os métodos utilizados não conseguiram prever de forma satisfatória o rendimento de carvão e gás, sendo os valores de R^2 inferiores a 0,5 em todos os modelos obtidos. Para a modelagem do bio-óleo, o modelo rede *feed-forward* contendo todas as variáveis preditoras apresentou o melhor desempenho, com RMSE de 5,71 e R^2 de 0,81. O modelo rede *cascade-forward*, com as mesmas variáveis, apresentou um desempenho inferior, com RMSE e R^2 de 6,73 e 0,70, respectivamente. Segundo os autores, uma melhora no desempenho dos modelos pode ser alcançada com um maior conjunto de dados experimentais e com a utilização de diferentes variáveis de entrada.

Zhu, Li e Wang (2019) desenvolveram modelos de previsão de rendimento e teor de carbono do carvão com base em dados de pirólise de biomassa lignocelulósica. Os dados experimentais de rendimento (245 amostras) e teor de carbono (128 amostras) usados para construir e testar modelos de previsão foram coletados de 19 publicações encontradas na

literatura. As biomassas presentes no banco consistiam em resíduos agrícolas e florestais e culturas energéticas.

Os fatores que influenciam no carvão obtido por pirólise de biomassa foram definidos por: condições de pirólise (taxa de aquecimento, temperatura e tempo de residência); componentes lignocelulósicos (lignina, celulose e hemicelulose); composição elementar da biomassa (teor de carbono, hidrogênio, oxigênio e nitrogênio); tamanho de partícula e teor de cinza. Para cada resposta analisada foram desenvolvidos três modelos de Floresta Aleatória com diferentes combinações de variáveis preditoras, sendo eles: modelo desenvolvido com todas as variáveis estudadas; modelo excluindo as variáveis de composição elementar; e modelo excluindo as variáveis de composição lignocelulósicas.

Os resultados sugeriram que a exatidão de predição aumentou quando todas as variáveis preditoras foram introduzidas nos modelos desenvolvidos em ambas as respostas. Neste caso, o R^2 e o RMSE foram de 0,85 e 3,40, respectivamente, para o rendimento de carvão e de 0,84 e 5,81, respectivamente, para teor de carbono. O modelo onde não foram utilizadas as variáveis de composição lignocelulósica para a previsão do rendimento de carvão obteve a pior performance, com RMSE de 3,97 e R^2 de 0,80. Porém, para o teor de carbono no carvão, o modelo em que a composição elementar não foi utilizada alcançou a pior performance, sendo, neste caso, o RMSE de 6,94 e o R^2 de 0,75.

As variáveis foram classificadas por sua importância nos modelos com melhores performances em ambas as respostas. Os resultados indicaram que a temperatura de pirólise foi o fator de maior influência tanto na previsão do rendimento do carvão (56%) quanto no teor de carbono (38%). No geral, as condições de pirólise forneceram cerca de 65% de contribuição para o rendimento do carvão, o que representou uma proporção maior do que na previsão do teor de carbono (53%). Para as características da biomassa, a composição lignocelulósica foi mais importante do que a composição elementar para prever, com precisão, o rendimento de carvão e o inverso foi encontrado para o teor de carbono.

Tang *et al.* (2020) utilizaram ferramentas de aprendizado de máquina para prever o rendimento e o teor de hidrogênio do bio-óleo com base nas composições da biomassa e nas condições do processo de pirólise. Os métodos empregados foram a Regressão Linear Múltipla (MLR) e Floresta Aleatória (RF).

Para a modelagem, um banco foi estruturado utilizando-se informações experimentais obtidas na literatura. Os conjuntos de dados utilizados continham 264 amostras para previsão do rendimento de bio-óleo e 137 amostras para determinação do teor de hidrogênio. Os

bancos incluíram experimentos realizados em reatores de leito fixo com biomassas oriundas de plantas herbáceas, biomassa lignocelulósica e algas. As variáveis de processos utilizadas como preditoras, em todos os modelos, foram a temperatura, taxa de aquecimento, tamanho de partícula e a vazão de nitrogênio. Além disso, foram analisadas variáveis de composição aproximada (cinza, carbono fixo e voláteis) e elementar (carbono, hidrogênio, nitrogênio e oxigênio) de forma separada na modelagem, ou seja, foram desenvolvidos modelos contendo variáveis de processo e composição elementar ou composição aproximada.

Os modelos MLR apresentaram um desempenho inferior aos modelos RF, o que mostrou que as variáveis de resposta não possuem relações lineares com as variáveis preditoras. Para a predição do rendimento de bio-óleo, o modelo RF, contendo as variáveis de composição aproximada e de processo, obteve um melhor resultado de predição com R^2 de 0,92 e RMSE de 2,13. Enquanto para o modelo incluindo variáveis de composição elementar os resultados foram de 0,87 e 3,05, para R^2 e RMSE, respectivamente. Um resultado oposto foi observado para a predição do teor de hidrogênio, onde o modelo incluindo variáveis de processo e composição elementar apresentou melhor performance com de R^2 de 0,84 e RMSE de 0,56.

A análise de influência das variáveis, para os modelos com melhores desempenhos, mostrou que a composição da biomassa exerce grande importância tanto no rendimento (60,5%) quanto no teor de hidrogênio do bio-óleo (76,7%). O teor de carbono fixo foi a variável mais importante para o modelo que prevê o rendimento de bio-óleo enquanto o teor de hidrogênio foi a variável de maior importância para predição do hidrogênio presente no bio-óleo. Entre as variáveis de processo, a temperatura e a taxa de aquecimento foram as de menor importância para a previsão do rendimento e do teor de hidrogênio, respectivamente.

Para ambas as respostas a variável vazão de nitrogênio apresentou uma baixa influência nos modelos obtidos. Isso pode ter ocorrido pois a vazão de N_2 controla o tempo de permanência do vapor na zona de reação, porém, para a determinação deste tempo é necessário conhecer o volume do reator utilizado.

Ullah *et al.* (2021) estudaram a predição do rendimento de bio-óleo de pirólise utilizando RF, SVM, MLR e CART. Para a modelagem foram coletados um total de 263 pontos de dados sobre rendimento de bio-óleo de 20 artigos publicados na literatura. Todos os experimentos foram realizados em reator de leito fixo com condições determinadas de taxa de aquecimento, temperatura, tamanho de partícula e vazão de nitrogênio. O banco de dados incluiu biomassas caracterizadas por sua composição elementar (carbono, hidrogênio,

nitrogênio e oxigênio) e por sua composição aproximada (carbono fixo, matéria volátil e cinzas).

Cada modelo de aprendizado de máquinas foi treinado para dois casos, primeiro usando a análise aproximada e as condições de pirólise e, segundo usando a análise elementar e as condições de pirólise. Uma abordagem de seleção de recursos baseada em algoritmo genético foi utilizada e, com base nos recursos selecionados, os quatro métodos de aprendizagem de máquinas foram investigados.

O desempenho do MLR não foi satisfatório, tanto na análise envolvendo a composição aproximada quanto na análise utilizando composição elementar, em comparação com RF, SVM e CART. A razão para o baixo desempenho do MLR em relação a outros modelos mostra o comportamento não linear do rendimento com as variáveis.

Para os modelos não lineares o modelo Floresta Aleatória obteve um melhor desempenho na predição do rendimento de bio-óleo em ambas as combinações de variáveis. No estudo utilizando composição aproximada o R^2 de treino foi de 0,88 e o RMSE de 2,78, enquanto os dados obtidos no conjunto de teste apresentaram um melhor resultado, com R^2 0,98 e RMSE de 1,06, respectivamente. Para o modelo utilizando composição elementar os R^2 foram de 0,92 e 0,99 para treino e teste respectivamente, enquanto os valores de RMSE foram de 1,96 para treino e 0,56 para teste. Um menor erro obtido no conjunto de teste pode indicar que o conjunto não foi obtido apropriadamente. O conjunto de teste pode ser muito pequeno ou não foi amostrado da mesma distribuição que seu conjunto de treinamento.

Apesar do alto ajuste dos dados ao modelo Floresta Aleatória descrito pelos autores, os dados de comparação de valores previstos por estes modelos e valores obtidos de diferentes fontes da literatura revelam que o modelo possui uma baixa capacidade preditiva.

3.3 Imputação

Muitos bancos de dados são afetados pelo problema de valores ausentes. Os dados podem estar ausentes por fatores como erro na entrada dos dados, medições incorretas e erros de equipamento. Em muitas áreas de aplicação, é comum encontrar bancos de dados com até 50% ou mais de valores ausentes. Estes dados são uma questão desafiadora para os pesquisadores pois, pode gerar problemas como perda de informação e eficiência,

complicações no manuseio e na análise dos dados e viés resultante de diferenças entre dados observados e ausentes (Barnard; Meng, 1999).

A análise com banco de dados contendo valores faltantes normalmente segue duas abordagens, os dados são removidos do banco ou imputados. A exclusão dos casos ausentes para obter dados completos pode causar a perda de informações importantes, portanto, esta solução é aplicável quando há uma pequena quantidade de dados faltantes em comparação com o número total de registros. Para bancos com uma grande quantidade de dados ausentes é mais viável considerar a imputação destes dados, porém, deve-se considerar que a imputação pode levar a viés (Farhangfar; Kurgan; Pedrycz, 2007).

A abordagem aos dados depende não apenas da quantidade de dados faltantes, mas também da identificação do tipo de dado ausente. O mecanismo de dados faltantes é geralmente classificado como (Little; Rubin, 2019):

- ✓ Dados faltantes completamente aleatórios (MCAR) - quando a probabilidade de um registro de valor ausente para uma variável não depende dos dados observados ou dos dados ausentes. No MCAR não há processo associativo que cause a ausência de dados, ou seja, a ocorrência de valores ausentes é uniformemente aleatória em toda a matriz de dados.

- ✓ Dados faltantes aleatórios (MAR) - quando a probabilidade de um registro ter um valor ausente para uma variável depende dos dados observados. Assim, uma variável com dados faltantes depende de uma variável com dados observados.

- ✓ Dados faltantes não aleatórios (NMAR) - quando a probabilidade de um registro ter um valor ausente para uma variável depende do valor do dado faltante. Por exemplo, pessoas que não preenchem a renda anual em pesquisas se a renda exceder um determinado valor.

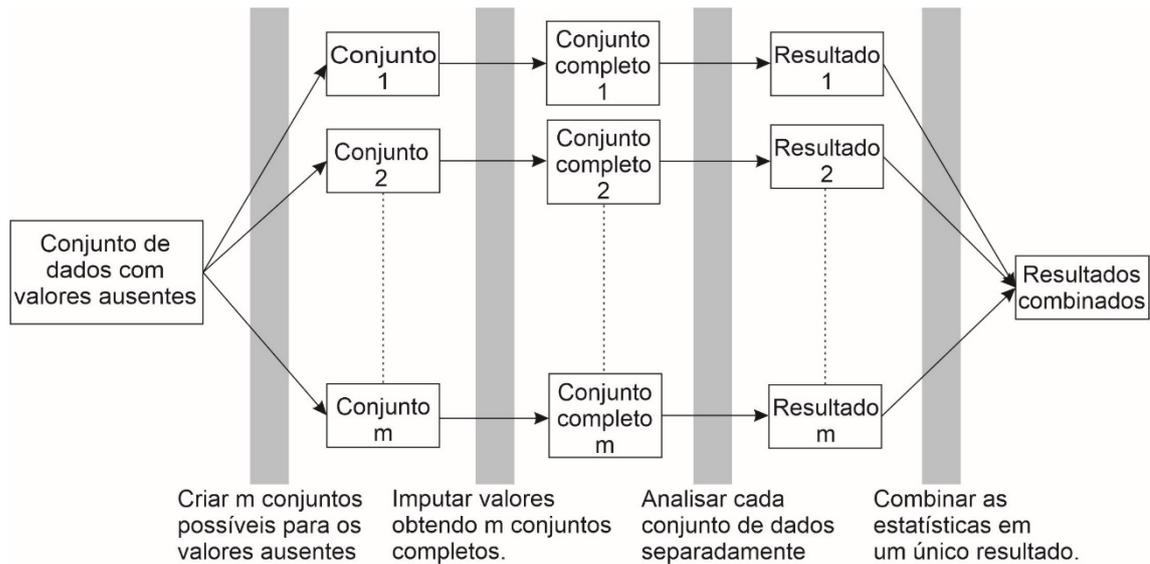
O tipo de mecanismo presente nos valores faltantes vem da suposição dos pesquisadores com base em sua compreensão da geração de dados e do processo de falta de dados. A suposição sobre a estrutura de valores ausentes tem implicações fundamentais para o processo de imputação. A suposição de que o mecanismo de dados faltantes é MCAR leva a uma perda de exatidão resultante da redução do tamanho, mas não gera problemas na análise e interpretação de resultados, entretanto esse mecanismo não é facilmente encontrado. Normalmente, dados ausentes são considerados ausentes aleatoriamente (MAR). Quando os valores ausentes são MAR, a exclusão das amostras faltantes resulta em viés. Isso ocorre porque os dados ausentes, não são uma subamostra aleatória de todos os dados. As técnicas de imputação múltipla, por outro lado, são ideais para lidar com dados MAR pois produzem dados eficientes e imparciais (Allison, 2001; Cranmer; Gill, 2012).

Com dados NMAR, as ferramentas de exclusão e imputação produzem resultados estatísticos tendenciosos, pois os dados ausentes são diferentes dos dados observados. Este mecanismo influencia a imputação múltipla, pois não há informações observadas a partir das quais construir um processo de imputação para preencher os valores ausentes. Os dados MNAR são chamados de não-ignoráveis, dado que o mecanismo gerador de ausência deve ser modelado para obtenção de estimativas de parâmetros imparciais (Allison, 2001; Cranmer; Gill, 2012).

A imputação é o processo de substituir os valores ausentes por estimativas. Vários métodos já foram desenvolvidos e usados com sucesso para lidar de forma eficiente com os dados ausentes. Os métodos para a imputação podem ser divididos em dois ramos: imputação única e imputação múltipla. Exemplos de métodos de imputação únicas são: substituição pela média, última observação transportada e imputação de vizinhos mais próximos. As estimativas obtidas por imputação única podem ser inconsistentes se o mecanismo de dados perdidos não for MCAR (Faisal; Tutz, 2021).

Um dos métodos mais populares para lidar com valores ausentes é a imputação múltipla. É uma técnica para gerar mais de um valor plausível para cada valor ausente nos dados. A imputação múltipla é um método que incorpora a incerteza no processo de imputação. Este método é composto por três estágios: estágio de imputação, no qual os dados faltantes são imputados M vezes; estágio de análise, no qual cada conjunto de dados completo é analisado usando uma técnica de dados completos; e a última etapa, na qual os resultados da análise são combinados para produzir um resultado. Sendo y , por exemplo, coeficiente de determinação, esta estatística é estimada M vezes e sua média, \bar{y}_m , é calculada para criar uma única estimativa de y (Cranmer; Gill, 2012). Uma visão geral gráfica do processo é mostrada na Figura 11.

Figura 11 - Metodologia de imputação múltipla



Fonte: Adaptado de Cranmer; Gill, 2012.

O principal atrativo da imputação múltipla é a sua eficiência, mesmo quando o número de imputações M é reduzido. Rubin (1987) mostrou que o ganho de eficiência de imputação reduz rapidamente após as primeiras imputações. Para dados onde a fração de dados faltantes é grande, a imputação múltipla com M modesto pode não ser completamente satisfatória, porém uma repetição de cinco imputações eleva a eficiência para mais de 90%.

O desenvolvimento dos métodos de imputação é estimulado pela necessidade de melhorar a precisão dos valores imputados. Os métodos mais simples, como imputação pela média, são computacionalmente baratos. Enquanto que métodos baseados em modelos de aprendizado de máquina podem melhorar a qualidade da imputação, mas exigem um esforço computacional maior (Farhangfar; Kurgan; Pedrycz, 2007).

Ascher *et al.* (2022) utilizaram a média como método de imputação em dados faltantes de composição de biomassa e resíduos e condições operacionais da gaseificação para aplicação em diferentes modelos de aprendizado de máquina. Os resultados mostraram que o método de imputação utilizado não foi eficiente e métodos como imputação por aprendizagem de máquinas devem ser consideradas para garantir o desempenho ideal na modelagem. Belden *et al.* (2023) utilizaram o método k-NN para imputação de dados em estudo de pirólise de plástico. Neste caso, as variáveis imputadas foram, tamanho de partícula, taxa de aquecimento e volume de biomassa alimentada no reator. A maior porcentagem de dados faltantes foi 28%. A modelagem mostrou que a inclusão dos dados obtidos por imputação beneficiou a

modelagem da previsão do rendimento de óleo no processo, em comparação à utilização somente de dados completos.

Diferentes métodos de imputação, incluindo métodos mais simples, como imputação pela média, a métodos que utilizam modelos de aprendizagem de máquinas, como o algoritmo de Floresta Aleatória, foram testados para a imputação de dados de pirólise desta pesquisa. Os três que apresentaram melhores resultados para a modelagem são descritos a seguir.

3.3.1 Imputação por k-vizinhos mais próximos.

A imputação por k-NN é uma técnica não paramétrica popular que tem sido amplamente utilizada para imputação simples. Neste método os valores ausentes no conjunto de dados podem ser restaurados usando os vizinhos mais próximos das características desse objeto.

Se considerarmos o valor A , ausente no experimento 1, esse método encontra k outros valores, presentes no experimento 1, mais semelhante a A . Uma média dos valores dos k valores mais próximos é então usada como uma estimativa para o valor ausente (Troyanskaya *et al.*, 2001).

A maioria dos métodos de imputação assume que os dados são pelo menos MAR, se não MCAR, assim como o método do k-NN. Ao usar os vizinhos mais próximos para imputar dados, deve-se definir quantos vizinhos serão considerados, ou seja, k , qual medida de distância será utilizada e como esses vizinhos mais próximos serão usados para obter o valor imputado (Tutz; Ramzan, 2015).

3.3.2 Imputação por Floresta Aleatória

Para o processo de imputação pelo algoritmo RF, suponha uma matriz de dados Y , em que Y_j é a j -ésima coluna das variáveis e p é o número de variáveis, Y_j^{obs} são os dados observados e Y_j^{mis} são os dados ausentes na j -ésima coluna. Primeiramente preenche-se as imputações iniciais por sorteios aleatórios de Y_j^{obs} , gerando uma matriz de dados Y' . Em seguida, são extraídas k amostras geradas por *bootstrap* a partir de Y' . Para cada amostra uma

árvore é ajustada. Isto resulta em k árvores, onde cada folha inclui um subconjunto de Y_j^{obs} que serão chamados doadores. Para cada árvore deve-se determinar a folha doadora para os membros de Y_j^{mis} , isto resulta em k folhas com doadores por membro de Y_j^{mis} . Em seguida o valor dos doadores é selecionado aleatoriamente para serem imputados. Este processo é realizado até alcançar um determinado número de iterações (Brancaglioni, 2023).

3.3.3 Imputação por Midastouch

O Midastouch é um método de imputação múltipla que usa uma adaptação da Correspondência Média Preditiva (PMM), em que doadores candidatos, ou seja, valores observados, têm diferentes probabilidades de serem utilizados na imputação (Näf *et al.*, 2023). Os valores imputados pelo método são considerados plausíveis, pois são valores observados no conjunto de dados. Contudo, a qualidade das imputações depende, em grande parte, da disponibilidade de doadores próximos (Gaffert; Meinfelder; Bosch, 2015).

Considere Y uma variável contendo valores faltantes (Y_{mis}) e valores observados (Y_{obs}). Para a imputação dos dados, os valores de Y_{obs} , chamados doadores, são utilizados para estimar n_{obs} (número de valores observados) conjuntos de parâmetros pelo princípio *leave-one-out*. Usando os parâmetros de regressão obtidos a partir dos dados observados, os valores previstos são calculados para os doadores e para os valores faltantes. Em seguida, calcula-se a distância entre todos os Y_{obs} preditos e o Y_{mis} . Finalmente, os valores de Y_{mis} são imputados usando os valores de Y_{obs} cujos seus valores previstos são mais próximos do valor predito faltante (Gaffert; Meinfelder; Bosch, 2015).

4 MATERIAIS E MÉTODOS

4.1 Descrição do banco de dados utilizado

O banco de dados contém informações sobre a pirólise em diversas condições do processo e foi desenvolvido para ajudar a explicar a importância das variáveis no rendimento e na composição dos produtos. O conjunto de dados foi desenvolvido a partir de dados experimentais obtidos de 220 pesquisas de pirólise de biomassa disponíveis na literatura. O banco de dados inclui artigos publicados desde 1984 até estudos recentes na área.

A Tabela 4 apresenta uma amostra do banco criado. Estruturar essas informações foi desafiador, pois a maioria dos dados de resultados experimentais são apresentados na forma de gráficos nas referências originais, o que dificulta a leitura dos valores com exatidão.

Tabela 4 - Amostra do banco de dados

Biomassa	Rendimento de bio-óleo (%)	Lignina (BS*)	Umidade (%)	Voláteis (BS*)	Cinza (BS*)	Carbono (DAF**)	Temp (°C)	Tam. partícula (mm)	Taxa de aque. (°C/min)	Regime de operação	Referência
Palha de arroz	51,00	20,70	6,80	88,84	9,55	43,25	471	5,00	-	Contínuo	Jung; Kang; Kim (2008)
Bambu	56,00	24,81	7,30	98,06	1,83	46,90	350	0,60	-	Contínuo	Jung; Kang; Kim (2008)
Bambu	68,00	24,81	7,30	98,06	1,83	46,90	410	0,85	-	Contínuo	Jung; Kang; Kim (2008)
Bambu	64,00	24,81	7,30	98,06	1,83	46,90	414	0,85	-	Contínuo	Jung; Kang; Kim (2008)
Celulose	62,20	0,00	-	-	0,00	42,27	500	-	-	Batelada	Stefanidis <i>et al.</i> (2014)
Hemicelulose	33,00	0,00	-	-	6,36	44,93	500	-	-	Batelada	Stefanidis <i>et al.</i> (2014)
Lignina	42,50	100	-	-	2,28	64,89	500	-	-	Batelada	Stefanidis <i>et al.</i> (2014)
Palha de arroz	48,50	-	13,61	-	11,04	50,93	500	-	200	Batelada	Tsai; Lee; Chang (2006)
Palha de arroz	50,00	-	13,61	-	11,04	50,93	500	-	300	Batelada	Tsai; Lee; Chang (2006)
Palha de arroz	49,00	-	13,61	-	11,04	50,93	500	-	400	Batelada	Tsai; Lee; Chang (2006)
Palha de arroz	51,00	-	13,61	-	11,04	50,93	500	-	500	Batelada	Tsai; Lee; Chang (2006)
Torta de soja	61,80	-	8,38	78,15	6,14	55,89	550	0,64	300	Batelada	Uzun; Pütün; Pütün (2006)
Torta de soja	60,60	-	8,38	78,15	6,14	55,89	550	0,64	700	Batelada	Uzun; Pütün; Pütün (2006)
Bagaço de girassol	60,20	-	5,90	78,43	6,16	49,60	400	0,64	7	Batelada	Pütün <i>et al.</i> (2001)
Casca de avelã	48,00	-	7,00	80,65	0,75	51,30	400	0,64	7	Batelada	Pütün <i>et al.</i> (2001)
Casca de pistache	35,00	13,82	7,39	83,07	1,45	49,98	500	1,75	100	Batelada	Pütün <i>et al.</i> (2007b)
Casca de pistache	42,00	13,82	7,39	83,07	1,45	49,98	500	1,75	300	Batelada	Pütün <i>et al.</i> (2007b)
Casca de pistache	41,00	13,82	7,39	83,07	1,45	49,98	500	1,75	500	Batelada	Pütün <i>et al.</i> (2007b)
Casca de pistache	38,00	13,82	7,39	83,07	1,45	49,98	500	1,75	700	Batelada	Pütün <i>et al.</i> (2007b)
Talo de gergelim	48,46	-	8,70	81,93	7,26	52,43	550	0,32	300	Batelada	Ateş; Pütün; Pütün (2004)
Talo de gergelim	50,20	-	8,70	81,93	7,26	52,43	500	0,64	300	Batelada	Ateş; Pütün; Pütün (2004)
Talo de gergelim	52,40	-	8,70	81,93	7,26	52,43	550	0,64	300	Batelada	Ateş; Pütün; Pütün (2004)
Talo de gergelim	44,00	-	8,70	81,93	7,26	52,43	700	0,64	300	Batelada	Ateş; Pütün; Pütün (2004)
Talo de gergelim	55,70	-	8,70	81,93	7,26	52,43	550	0,64	500	Batelada	Ateş; Pütün; Pütün (2004)

* - % em base seca; ** - % em base seca livre de cinza.

O banco de dados foi dividido em três partes principais, sendo elas, características da biomassa, como tipo (por exemplo, casca de arroz) e composição (por exemplo, composição elementar, teor de cinzas, teor de umidade, teor de lignina, poder calorífico), condições de processo (por exemplo, tipo de reator, taxa de aquecimento, temperatura, tipo de pirólise) e rendimento e composições dos produtos. No banco estão incluídas 170 variáveis, sendo 36 relacionadas à caracterização da biomassa, 37 relacionadas às condições do processo e 96 relacionadas aos produtos. Mais da metade das variáveis que analisam os produtos, 51 variáveis, estão ligadas às características do bio-óleo. Do restante, 27 mostram as propriedades do carvão e 18 do biogás produzido.

Diferentes formas de calcular o rendimento do bio-óleo foram encontradas na literatura. O bio-óleo contém quantidades variáveis de água, variando de cerca de 15% a um limite superior de cerca de 30-50%, dependendo da biomassa, do processo e da coleta do produto (Bridgwater, 2012). Um alto teor de umidade no bio-óleo pode levar à separação de fases, sendo obtida uma fase aquosa e uma fase oleosa. Em algumas pesquisas da literatura, o bio-óleo foi considerado a fase oleosa do líquido produzido na pirólise (Duman *et al.*, 2011; Pütün; Ateş; Pütün, 2008; Şensöz; Kaynar, 2006; Yanik *et al.*, 2007). Outros, separaram o líquido obtido na pirólise em orgânicos e água por extração com solvente, sendo a porcentagem dos orgânicos o rendimento do bio-óleo (Garg; Anand; Kumar, 2016; Pärpärilä *et al.*, 2014; Tsai; Lee; Chang, 2006; Ucar; Ozkan, 2008). Para padronização do banco de dados, o rendimento do bio-óleo, neste estudo, foi considerado como líquido total produzido no processo de pirólise. Além disso, as variáveis de composição aproximada (umidade, cinzas, carbono fixo e volátil) e concentração de materiais lignocelulósicos da biomassa foram transformadas para % em base seca e as variáveis de composição elementar (carbono, oxigênio, hidrogênio e nitrogênio) foram transformadas para % em base seca livre de cinza.

Além da lignina, celulose, hemicelulose e suas misturas, outros 177 tipos diferentes de biomassa foram analisados. A Tabela do Anexo A mostra as várias biomassas estudadas no processo de pirólise nos artigos revisados e o número de observações de cada uma no banco. Além disso, o Anexo B apresenta as referências bibliográficas dos estudos presentes no banco.

As biomassas com mais observações no banco de dados são cascas de arroz, casca de palma e torta de pinhão manso que foram analisadas em 26, 8, 10 pesquisas, respectivamente. A casca de arroz é um resíduo agrícola amplamente disponível pois, o arroz é considerado um alimento básico na maioria dos países, com uma produção global estimada em 2020 de 509,2

milhões de toneladas (Food and Agriculture Organization, 2020a). Assim, os subprodutos estão disponíveis em grande quantidade para fornecer uma fonte de energia alternativa.

Um dos principais problemas no processamento do óleo de palma e de pinhão manso é o gerenciamento dos resíduos gerados durante os processos. A *Jatropha curcas* foi identificada como uma das culturas energéticas mais adequada para a produção de biodiesel e está atraindo atenção em todo o mundo como fonte de energia renovável, porém, a produção de biodiesel a partir de sementes de pinhão manso gera uma grande quantidade de torta, resíduo do processo, que não pode ser utilizada na alimentação devido a sua toxicidade. A produção de óleo de palma disparou nos últimos anos, resultando em uma produção estimada de 18,1 milhões de toneladas de semente de palma em 2020 (Food and Agriculture Organization, 2020b). O resíduo da palma tem um alto potencial de ser transformado em energia renovável, mas, geralmente, não é aproveitado e muitas vezes é queimado sem o aproveitamento adequado (Abnisa *et al.*, 2011). Assim, é necessário um gerenciamento eficiente dos resíduos para melhorar a eficiência do processo e a economia na produção de óleo.

4.2 Análise exploratória dos dados

Na análise exploratória do banco utilizou-se ferramentas de visualização que ajudam a esclarecer a estrutura geral dos dados. As ferramentas gráficas utilizadas foram histogramas, gráficos tipo boxplot e gráficos de dispersão. Essas ferramentas são úteis para verificação de tendência central dos dados, assimetria da distribuição, presença de outliers e presença de correlação (Li Vigni; Durante; Cocchi, 2013). Os gráficos são importantes na análise dos dados devido à capacidade do cérebro humano de obter uma interpretação mais direta e confiável de semelhanças, tendências e correlações por meio de imagens (Li Vigni; Durante; Cocchi, 2013).

Testes estatísticos foram utilizados para a preparação dos dados para a modelagem. Foram realizados testes de normalidade, seleção de preditores com testes de correlação e variância, pré-processamento dos dados e seleção de amostras com a remoção de amostras com valores faltantes.

4.2.1 Valores faltantes

Em todos os casos analisados, alguns preditores não possuem valores para uma determinada amostra. Para lidar com este problema duas abordagens foram utilizadas, sendo elas, realização da modelagem apenas com dados completos e modelagem dos dados após a aplicação de diferentes técnicas de imputação. A imputação foi realizada para verificar se há ganho na eficiência dos modelos com a inclusão das informações adicionais.

4.2.2 Seleção de preditores

A seleção de preditores é uma etapa importante no processo de modelagem. Muitas vezes, nem todas as variáveis disponíveis têm importância para a previsão. A remoção destas variáveis irrelevantes pode levar a uma melhoria no modelo.

Alguns preditores podem ter apenas alguns valores que ocorrem com frequências muito baixas. Esses “preditores de variância quase zero” podem ter um único valor para a grande maioria das amostras. Para alguns modelos, essas variáveis podem ter pouco efeito nos cálculos. Nesse caso, esses dados foram descartados do banco (Kuhn; Johnson, 2013).

Outro método utilizado para a seleção das variáveis foi a análise de correlação dos preditores. Variáveis com correlações fortes levam ao problema de multicolinearidade. A multicolinearidade pode fazer com que a solução seja mal condicionada e o modelo resultante tenha grande variância de predição (Qin, 1997).

A correlação entre duas variáveis, em termos matemáticos, é a força da relação matemática entre dois conjuntos de dados. O coeficiente de correlação reduz essa relação a um número que podemos interpretar (Metcalf; Casey, 2016). Em termos de força de relacionamento, o valor do coeficiente de correlação varia entre +1 e -1, sendo estas as correlações positiva e negativa mais fortes, respectivamente. Quando o valor do coeficiente de correlação aproxima-se de 0, a relação entre as duas variáveis é fraca. Se dois preditores estiverem altamente correlacionados, isso implica que eles estão medindo as mesmas informações subjacentes. Assim, a remoção de uma dessas variáveis gera benefícios, como a redução do tempo computacional e da complexidade do modelo (Kuhn; Johnson, 2013).

Como as variáveis predictoras do bio-óleo não possuem uma distribuição normal e há a presença de valores atípicos, utilizou-se a correlação de Spearman para a obtenção dos coeficientes de correlação entre os preditores. A abordagem utilizada para lidar com a colinearidade foi a remoção de preditores com correlação acima de 0,85. A função usada, "findCorrelation" do pacote Caret, pesquisa a matriz de correlação e retorna um vetor indicando as variáveis que possuem correlações acima do limite estabelecido (R Core Team, 2022). Para escolher qual dos preditores com alta correlação remover dos modelos, primeiro realizou-se a modelagem com todas as alternativas de remoção, ou seja, foram gerados os modelos com a remoção de uma ou da outra variável com alta correlação. O conjunto de variáveis, contendo uma das variáveis correlacionadas, que levou a um melhor resultado de predição foi escolhido como o mais apropriado para seguir no estudo.

4.2.3 Teste de normalidade

A distribuição normal dos dados é um pré-requisito para muitos procedimentos estatísticos. Existem diferentes métodos usados para testar a normalidade dos dados, incluindo métodos numéricos e visuais (Mishra *et al.*, 2019). O exame gráfico dos dados é importante, mas a interpretação de qualquer gráfico é subjetiva. Portanto, na maioria dos casos, o exame gráfico de uma distribuição deve ser complementado por um teste estatístico formal (Looney; Hagan, 2007; Yap; Sim, 2011).

Existe um número significativo de testes de normalidade disponíveis na literatura. O teste utilizado nesta pesquisa foi o Shapiro-Wilk, um teste multivariado amplamente utilizado e que possui um bom poder estatístico para uma ampla variedade de distribuições não normais (Alva; Estrada, 2009; Looney; Hagan, 2007; Razali; Wah, 2011).

O teste de Shapiro-Wilk verifica, em um conjunto de dados, a semelhança com a distribuição normal. No teste, a hipótese nula supõe que os dados tenham distribuição normal; portanto, um valor de p suficientemente pequeno ($<0,05$) indica desvio da normalidade multivariada (Tacq, 2010).

4.2.4 Teste de Kruskal-Wallis

O teste não paramétrico de Kruskal-Wallis foi aplicado para comparar três ou mais grupos independentes durante a análise exploratória do banco de pirólise. O teste é uma alternativa quando os pressupostos necessários pelo teste Análise de Variância não são atendidos, tendo em vista que o teste de Kruskal-Wallis não necessita da suposição de normalidade e homocedasticidade. Neste teste a hipótese nula afirma que todas as os grupos possuem uma mesma distribuição contra a hipótese alternativa de que pelo menos um grupo possui uma distribuição distinta das demais (Ostertagová; Ostertag; Kováč, 2014).

Quando o teste de Kruskal-Wallis leva a resultados de $p < 0,05$, então pelo menos um dos grupos difere significativamente dos demais. O teste não identifica onde ocorrem as diferenças. Portanto, é necessário realizar comparações múltiplas para detectar quais pares de grupos podem ser considerados diferentes. Neste caso utilizou-se o teste de Dunn, apropriado para grupos com tamanhos diferentes (Elliott; Hynan, 2011).

4.3 Pré-processamento e transformação dos dados

Após a análise exploratória, o pré-processamento incluindo centralização e escalonamento, transformação Box-Cox e aplicação de sinal espacial, foi aplicado aos dados. Essas manipulações foram utilizadas para minimizar as diferenças de magnitude e variabilidade de cada atributo (Andrade *et al.*, 2020).

A centralização e escalonamento são recomendadas quando os preditores possuem uma grande variação na magnitude dos dados ou quando possuem diferentes unidades. Na centralização e escalonamento, os dados de cada variável são centrados na média, ou seja, subtrai-se de cada variável a média das amostras dessa variável, e depois divididos pelo desvio padrão da variável. Dessa forma, cada variável tem média zero e desvio padrão um.

Outra etapa de pré-processamento que se mostrou importante após a análise exploratória é a transformação de dados. As transformações podem reduzir a assimetria distributiva e induzir os dados à normalidade, constituindo-se em uma importante ferramenta de preparação de dados para análises estatísticas. Existem muitos tipos de transformações de

dados. Algumas das mais tradicionais incluem: adição de constantes, raiz quadrada, conversão em escalas logarítmicas e aplicação de transformações trigonométricas, como transformações de onda senoidal (Osborne, 2010). Métodos estatísticos podem ser usados para identificar e aplicar a transformação apropriada. Nesta pesquisa utilizou-se o método definido por Box e Cox (1964) que usa estimativa de máxima verossimilhança para determinar o parâmetro de transformação (λ) para a conversão dos dados, conforme Equação 29.

$$x^* = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \log(x) & \text{if } \lambda = 0 \end{cases} \quad (29)$$

onde x é a variável preditora original e x^* é o dado transformado.

Draper e Cox (1969) mostraram que o procedimento de transformação para normalidade com base no Box-Cox pode ajudar a regularizar os dados, mesmo quando a normalidade não é alcançada.

Outra técnica utilizada no pré-processamento foi a transformação dos dados pela aplicação do sinal espacial. Esse procedimento foi realizado devido a verificação da presença de outliers nas amostras. A técnica é conceitualmente simples e requer esforços computacionais limitados, pois eles se resumem a uma transformação dos dados em seus sinais espaciais (Serneels; De Nolf; Van Espen, 2006). Este procedimento projeta os valores do preditor em uma esfera multidimensional. Matematicamente, cada amostra é dividida por sua norma ao quadrado (Equação 30):

$$x_{ij}^* = \frac{x_{ij}}{\sum_{j=1}^p x_{ij}^2} \quad (30)$$

4.4 Algoritmo de seleção de amostras

Após o pré-processamento as amostras foram divididas em conjunto de treinamento, que é utilizado para a aplicação dos modelos, contendo 75% dos dados, e conjunto de teste,

que tem como objetivo avaliar a capacidade preditiva dos modelos, com 25% dos dados. Foi utilizado o algoritmo Kennard-Stone (KS) para a divisão dos dados.

O algoritmo de Kennard-Stone (KS) é um dos mais conhecidos algoritmos para seleção de amostras. O algoritmo permite selecionar amostras com uma distribuição uniforme sobre o espaço preditor. O algoritmo inicia-se selecionando, para o conjunto de teste, as duas amostras com a maior distância de Mahalanobis entre si, no espaço X. Para cada uma das amostras que permaneceram, calcula-se a distância em relação às amostras já selecionadas. Então, a amostra com a maior distância é retida. Este procedimento é repetido até que o número desejado de amostras seja selecionado (Kennard; Stone, 1969).

4.5 Aplicação dos modelos

Inicialmente treze modelos foram selecionados para modelagem, sendo eles: MLR, Ridge, LASSO, Rede, ANN, MARS, k-NN, SVM/Br, SVM/Pl, CART, *Cubist*, RF e GBM.

Após a aplicação dos modelos citados acima o pacote LazyPredict, presente no módulo homônimo da linguagem Python, foi utilizado para verificar se os modelos utilizados estavam entre os mais apropriados para a predição do bio-óleo. Esta ferramenta treina 42 diferentes modelos em uma única linha de código e fornece as figuras de mérito de cada um destes modelos. Assim, esta biblioteca ajuda a entender quais modelos funcionam melhor sem nenhum ajuste de parâmetro e que, provavelmente, serão adequados para a modelagem. Após a aplicação do LazyPredict o modelo Árvores Extremamente Aleatórias (ExtraTree) foi incluído no conjunto de modelos utilizados para a modelagem do rendimento de bio-óleo.

Exceto o modelo de Regressão Linear Múltipla, todos os modelos foram criados com a função de 'train' do pacote Caret disponível no software R. A Tabela 5 mostra os parâmetros de ajuste para cada modelo, o método utilizado e os pacotes necessários para a modelagem. Os parâmetros foram ajustados utilizando uma reamostragem 10-fold com cinco repetições.

Tabela 5 - Parâmetros de ajuste, método usado e os pacotes necessários

Modelo	Parâmetros de ajuste	Método	Pacotes necessários
MLR	-	lm	olsrr
Ridge	lambda	ridge	elasticnet
LASSO	fraction	lasso	elastic net
Rede	lambda, fraction	enet	elasticnet
ANN	decay, size	nnet	nnet
MARS	nprune, degree	earth	earth
k-NN	k	knn	Caret
SVM/Br	sigma, c	svmRadial	kernlab
SVM/Pl	c, degree, scale	svmPoly	kernlab
CART	maxdepth	rpart2	rpart
<i>Cubist</i>	committees, neighbors	<i>Cubist</i>	<i>Cubist</i>
RF	mtry	rf	randomForest
ExtraTree	numRandomCuts, mtry	extraTrees	extraTrees
GBM	ntrees, interaction.depth, shrinkage, n.minobsinnode	gbm	gbm, plyr

lambda = Queda de peso; fraction= Fração de solução completa; decay = Queda de peso; size= Número de neurônios; nprune= Número de termos; degree= grau; k= Número de vizinhos; c= Custo; maxdepth= Profundidade máxima da árvore; committees =Número de comitês; neighbors= Número de instâncias; mtry= Preditores selecionados aleatoriamente; numRandomCuts=número de cortes aleatórios; ntrees= Número de iterações de Boosting; interaction.depth= Profundidade máxima da árvore; shrinkage= Encolhimento; n.minobsinnode= Tamanho mínimo do nó terminal.

Fonte: O autor, 2023.

4.6 Imputação

Após a modelagem de dados completos, ou seja, com a remoção de dados faltantes, a regressão foi realizada também com dados obtidos por diferentes técnicas de imputação, sendo elas: k-NN e imputação múltipla com os modelos RF e Midastouch.

A imputação k-NN foi realizada com a função `knnImpute` do pacote `Caret`, disponível no software R. Para a imputação dos dados foram identificados 5 vizinhos mais próximos ($k=5$) da amostra a ser imputada. Utilizou-se a distância euclidiana e o valor imputado foi a média dos vizinhos mais próximos encontrados.

As imputações utilizando os algoritmos RF e Midastouch foram implementadas utilizando o pacote MICE no software R. Foram realizadas cinco imputações em cada método, ou seja, M=5. Para o algoritmo RF o número de árvores do modelo foi 10, sendo o padrão do pacote utilizado. Shah *et al.* (2014) observaram pouca diferença na utilização de 10 e 100 árvores em estudos utilizando o mesmo algoritmo.

4.7 Ferramentas de Diagnóstico

Nenhum modelo domina todos os outros em todos os conjuntos de dados possíveis. Ou seja, em um determinado conjunto de dados, um método específico pode funcionar melhor, mas esse método pode não ser tão bom em outro conjunto de dados semelhante (James *et al.*, 2013).

Para comparar a eficácia de diferentes técnicas de análise multivariada, os critérios de seleção de modelo são necessários para minimizar o erro de predição e auxiliar na escolha de modelos para resultados ideais. Existem muitas maneiras diferentes de avaliar a qualidade de um modelo de regressão (Boucher *et al.*, 2015). Três estatísticas de verificação coeficiente de determinação (R^2), raiz do erro médio quadrático (RMSE) e erro médio de estimativa (MAE), descritas nas Equações 31, 32 e 33, respectivamente, foram empregadas para analisar os resultados preditivos dos modelos (Mehdizadeh; Behmanesh; Khalili, 2017; Zhou *et al.*, 2019).

$$R^2 = \frac{\sum_{i=1}^n (y_i - \bar{y}_i)^2 - \sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (31)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (32)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (33)$$

Nas quais: n é o número total de dados usados; \hat{y}_i e y_i são a resposta prevista e observada, respectivamente, da i -ésima observação e \bar{y}_i é a média de todas as respostas observadas.

O coeficiente de determinação (R^2) caracteriza a proporção da variabilidade em um conjunto de dados que é contabilizado pelo modelo estatístico. O intervalo de R^2 está entre 0 e 1. R^2 próximo de 1 indica um melhor ajuste do modelo aos dados.

O RMSE mede o desvio entre os valores previstos por um estimador e os valores realmente observados. O RMSE tem um valor igual ou superior a 0, sendo que, um menor valor de RMSE indica uma melhor performance do modelo. Enquanto o valor do erro igual a 0 representa um ajuste estatisticamente perfeito entre os dados observados e preditos.

O MAE é a média de todos os erros absolutos, ou seja, é a média da diferença entre os valores observados e preditos, e é outra medida útil amplamente empregada em avaliações de modelos. Um menor valor de MAE indica que o modelo tem uma boa capacidade preditiva (James *et al.*, 2013; Sousa *et al.*, 2007; Ullah *et al.*, 2021; Zhou *et al.*, 2016).

Existem diferentes técnicas de diagnóstico que podem ser utilizadas para avaliar a qualidade dos dados imputados. Neste estudo os gráficos de densidade foram utilizados para comparar dados observados e imputados e identificar onde as imputações diferem dos valores observados. As diferenças na distribuição não sinalizam necessariamente um problema com as imputações. No entanto, diferenças significativas entre os dados imputados e os observados podem sugerir um problema na imputação realizada (Abayomi, 2008).

4.8 Otimização

A otimização envolve encontrar o mínimo ou o máximo de uma função específica chamada função objetivo, definida em um determinado espaço. Um algoritmo de otimização não linear livre de derivadas foi implementado para encontrar o ponto ótimo, ou seja, o maior rendimento de bio-óleo, nos melhores modelos obtidos em estudos de pirólise em batelada e em reator contínuo. Esta etapa foi realizada para demonstrar uma utilização prática da modelagem na pirólise de biomassa. Além disso, pode-se comparar os resultados da variação das condições de processo em reatores com diferentes regimes operacionais e mesma composição de biomassa.

A otimização por enxame de partícula (PSO) foi desenvolvida por Kennedy e Eberhart (1995). É um método meta-heurístico inspirado no comportamento de organismos sociais em grupos, como bandos de pássaros, cardumes de peixes e teoria da enxameação. Este método simula a interação entre os membros do grupo para compartilhar informações.

O algoritmo busca a função objetivo ajustando as trajetórias de pássaros individuais (partículas). A localização de uma partícula no enxame é influenciada por sua melhor localização e pela melhor localização do grupo durante a trajetória da partícula. Cada indivíduo possui uma determinada posição e uma velocidade de deslocamento, que são ajustadas a cada iteração do algoritmo (Koziel; Yang, 2011; Martínez; Cao, 2019). O método PSO pode encontrar a solução global ótima, mas requer muitas partículas em um processo ótimo, o que reduz a velocidade de computação (Liu; Yang, 2012).

A etapa de otimização do rendimento de bio-óleo foi realizada utilizando-se o modelo que apresentou maior capacidade preditiva de bio-óleo. Neste estudo somente as variáveis de processo foram otimizadas para na pirólise da casca de arroz. Esta biomassa foi escolhida por apresentar um maior número de observações no banco de dados. A faixa de estudo das variáveis de processo otimizadas está dentro dos limites utilizados para obtenção do modelo. O número de partículas no enxame utilizado foi 100. De acordo com Piotrowski, Napiorkowski e Piotrowska (2020) na maioria das aplicações com PSO o tamanho da população utilizado varia entre 20 e 50 partículas. Porém, se o tamanho de população não for conhecido usar de 70-100 partículas é a escolha mais segura.

4.9 Software e pacotes

Todas as análises do banco e os modelos de regressão, com ajuste de parâmetros, foram implementadas no *R Statistical Software*, versão 4.1.3 (R Core Team, 2022), um software gratuito para a criação de gráficos e análises estatísticas. A Tabela 6 mostra os pacotes e suas funções utilizadas em cada etapa do pré-processamento.

O pacote LazyPredict (Pandala, 2020), módulo da linguagem Python, foi utilizado para fazer uma varredura em 42 diferentes modelos de regressão para verificar a necessidade de inclusão de novos algoritmos na modelagem.

Tabela 6 - Pacotes e funções do pré-processamento e modelagem

Etapa	Pacote	Função	Referência
Análise exploratória			
Teste de alta correlação	Caret	findCorrelation	Kuhn, 2021
Teste de variância	Caret	nearZeroVar	Kuhn, 2021
Teste de normalidade	mvnrmtest	mshapiro.test	Jarek, 2012
Teste de Kruskal-Wallis	rstatix	kruskal.test	Kassambara, 2023
Pré-processamento	Caret	preProcess	Kuhn, 2021
Identificação de outliers	mvoutlier	pcout	Filzmoser; Gschwandtner, 2018
Aplicação do sinal espacial	Caret	spatialSign	Kuhn, 2021
Divisão dos dados em teste e treinamento	prospectr	kenStone	Stevens; Ramirez-Lopez, 2021
Imputação			
k-NN	Caret	knnImpute	Kuhn, 2021
Floresta Aleatória	mice	rf	Van Buuren; Oudshoorn, 2011
Midastouch	mice	midastouch	Van Buuren; Oudshoorn, 2011
Modelagem			
Aplicação dos modelos	Caret	train	Kuhn, 2021
	Lazypredict	LazyRegressor	Pandala, 2020
Figuras de mérito	Caret	postResample	Kuhn, 2021
Otimização			
PSO	PSO	psoptim	Bendtsen, 2022

Fonte: O autor, 2023.

4.10 Descrição dos estudos de caso de modelagem

Vários reatores diferentes foram analisados para atingir alto rendimento de bio-óleo. O banco de dados contém um total de oito tipos de reatores. O processo de modelagem de bio-óleo foi dividido de acordo com o regime de operação do reator utilizado, sendo batelada ou contínuo. O conjunto de dados possui 1906 observações de reatores em batelada e 711 reatores operando continuamente.

O regime operacional, usado no reator, altera as variáveis do processo que influenciam a pirólise. A variável tempo de reação, ou seja, tempo de duração do processo na temperatura

especificada, está presente apenas na modelagem do processo realizado em um reator operando em batelada. Além disso, em reatores contínuos, a biomassa é alimentada no reator à temperatura de operação. Portanto, a variável de taxa de aquecimento não é usada. Por outro lado, a variável da taxa de alimentação é um fator importante em reatores de alimentação contínua (Heo *et al.*, 2010; Jung; Kang; Kim, 2008; Kim; Jung; Kim, 2010; Wu *et al.*, 2016) e sua influência está relacionada à configuração do reator utilizado (Asadullah *et al.*, 2013).

Para reatores operando em batelada, a variável tempo de residência do gás foi obtida dividindo-se o volume do reator (L) pela vazão do gás inerte (L/s).

A modelagem foi realizada em oito estudos de caso diferentes. A Tabela 7 mostra as variáveis analisadas em cada caso. Os estudos de caso I e V analisaram apenas variáveis de processo em reatores em batelada e contínuos, respectivamente. Os demais estudos incluíram variáveis de composição aproximada, elementar e lignocelulósica da biomassa.

Tabela 7 - Variáveis analisadas nos estudos de caso

Variáveis	Caso I	Caso II	Caso III	Caso IV	Caso V	Caso VI	Caso VII	Caso VIII
Regime de operação do reator								
Batelada	✓	✓	✓	✓				
Contínuo					✓	✓	✓	✓
Variáveis de processo								
Temperatura	✓	✓	✓	✓	✓	✓	✓	✓
Tamanho de partícula	✓	✓	✓	✓	✓	✓	✓	✓
Residência do vapor	✓	✓	✓	✓	✓	✓	✓	✓
Taxa de aquecimento	✓	✓	✓	✓				
Tempo de reação	✓	✓	✓	✓				
Composição aproximada e elementar								
Cinza		✓		✓		✓		✓
Umidade		✓		✓		✓		✓
Voláteis		✓		✓		✓		✓
Carbono fixo		✓		✓		✓		✓
Carbono		✓		✓		✓		✓
Hidrogênio		✓		✓		✓		✓
Oxigênio		✓		✓		✓		✓
Nitrogênio		✓		✓		✓		✓
Composição lignocelulósica								
Celulose			✓	✓			✓	✓
Lignina			✓	✓			✓	✓
Hemicelulose			✓	✓			✓	✓

Fonte: O autor, 2023.

5 RESULTADOS E DISCUSSÃO

5.1 Análise exploratória dos dados

O banco de dados contém um total de 2617 observações, com dados ausentes em quase todas as variáveis. A Tabela 8 descreve a quantidade de artigos presentes na base de dados que analisaram cada uma das variáveis consideradas como as mais importantes no processo de pirólise.

Tabela 8 - Variáveis analisadas nas pesquisas presentes no banco de dados

Variáveis	Número de pesquisas que analisaram as variáveis
Temperatura	162
Tempo de residência do vapor	62
Tamanho de partícula	44
Taxa de aquecimento	38
Composição da biomassa	35
Tempo de reação	11

Fonte: O autor, 2023.

A maioria das pesquisas presentes na base de dados (73,3%) analisou a importância da temperatura no processo. As demais variáveis foram analisadas por apenas 28% dos pesquisadores. A temperatura é a variável com maior influência no rendimento do bio-óleo, sendo o parâmetro mais crítico do processo, o que explica a maior atenção dada a esta variável.

A Tabela 9 resume os atributos e o número de observações de cada uma das variáveis utilizadas na modelagem descrita nesta pesquisa. A produção de bio-óleo varia de 0,02% a 90% no banco de dados, com média de 43,95%. O menor rendimento foi obtido com a casca de arroz, em reator em batelada, na maior temperatura analisada (1220 °C), o que favorece a formação do biogás. O maior rendimento de líquido, 90%, foi obtido na pirólise da celulose, em reator contínuo, a uma temperatura de 600 °C. Esses dados mostram a importância das variáveis consideradas nesta modelagem.

Tabela 9 - Resumo das variáveis de processo e composição da biomassa

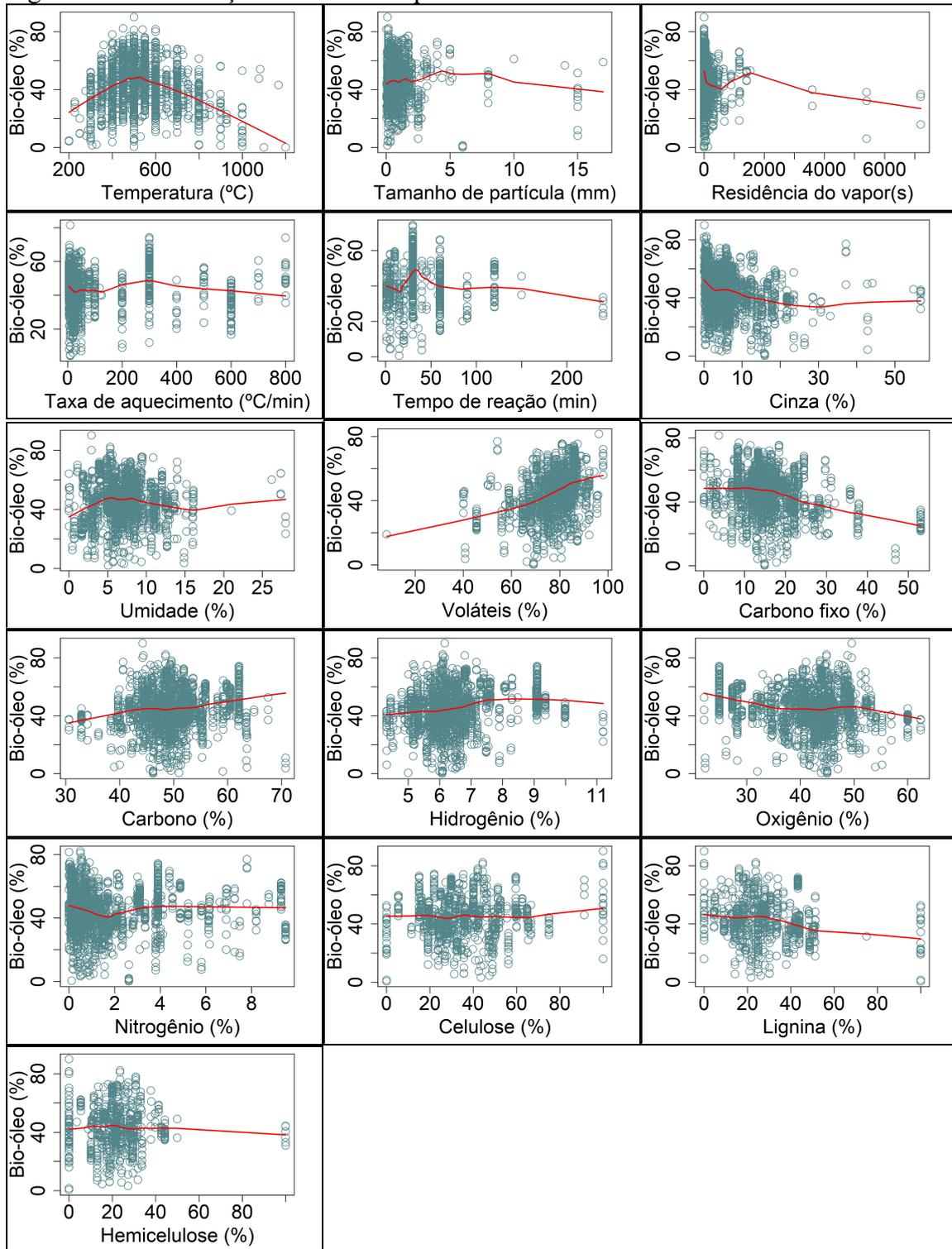
Variável	Min. ¹	1º Qu. ²	Mediana	Média	3º Qu. ³	Max. ⁴	Obs. ⁵
Rendimento de bio-óleo (%)	0,02	35,50	44,80	43,95	52,73	90,00	2373
Processo							
Temperatura (°C)	200	450	500	527,20	585,20	1200	2602
Tamanho de partícula (mm)	0,03	0,42	0,73	1,10	1,29	20,00	1805
Residência do vapor (s)	0,05	3,07	24,00	142,06	117,13	7200	1482
Taxa de aquecimento (°C/min)	5,00	10,00	30,00	103,10	80,00	800,00	1422
Tempo de reação (min)	1,00	20,00	30,00	38,06	60,00	240,00	1014
Composição							
Cinza (%)	0,00	1,55	4,28	6,00	6,98	56,80	2372
Umidade (%)	0,00	5,34	6,90	7,14	8,50	78,10	2027
Voláteis (%)	8,19	72,94	78,75	77,27	83,43	98,06	1944
Carbono fixo (%)	0,11	11,98	15,55	16,74	20,01	53,02	1907
Carbono (%)	30,60	46,09	49,27	49,57	52,10	70,73	2254
Hidrogênio (%)	4,31	5,81	6,20	6,44	6,79	11,20	2250
Oxigênio (%)	21,71	38,93	42,90	41,86	45,80	62,44	2187
Nitrogênio (%)	0,00	0,43	0,93	1,79	2,35	9,49	2120
Celulose (%)	0,00	25,00	32,67	36,37	47,50	100	1141
Lignina (%)	0,00	16,54	26,15	28,61	40,00	100	960
Hemicelulose (%)	0,00	16,66	21,60	21,47	28,18	100	877

Min. - Valor mínimo; 1º Qu. - Valor do primeiro quadrante; 3º Qu. - Valor do terceiro quadrante; Max. - Valor máximo; Obs - Número de observações no banco de dados.

Fonte: O autor, 2023.

As variáveis de processo e de composição que apresentaram um maior número de observações no banco de dados foram a temperatura de pirólise e teor de cinzas. Um grande número de observações com teor de cinzas se justifica pois, de acordo com a literatura, as cinzas têm uma influência importante no rendimento dos produtos da pirólise. A obtenção de um menor rendimento de bio-óleo é alcançada em biomassa com alta concentração de cinzas, o que leva a uma maior produção de carvão (Abdullah; Gerhauser, 2008; Abnisa *et al.*, 2013; Alvarez *et al.*, 2015; Fu *et al.*, 2010; Lee *et al.*, 2013; Pattiya; Sukkasi; Goodwin, 2012; Pattiya; Suttibak, 2012; Razuan *et al.*, 2010). Essa tendência é observada na Figura 12, que mostra a distribuição das variáveis em relação ao rendimento do bio-óleo.

Figura 12 - Distribuição das variáveis por rendimento de bio-óleo



Fonte: O autor, 2023.

A temperatura foi analisada entre 200 a 1200 °C, com média de 527,2 °C. De acordo com a literatura, altos rendimentos de bio-óleo foram alcançados em temperaturas próximas a 500 °C (Angin, 2013; Asadullah *et al.*, 2007; Garg; Anand; Kumar, 2016; Paenpong; Pattiya,

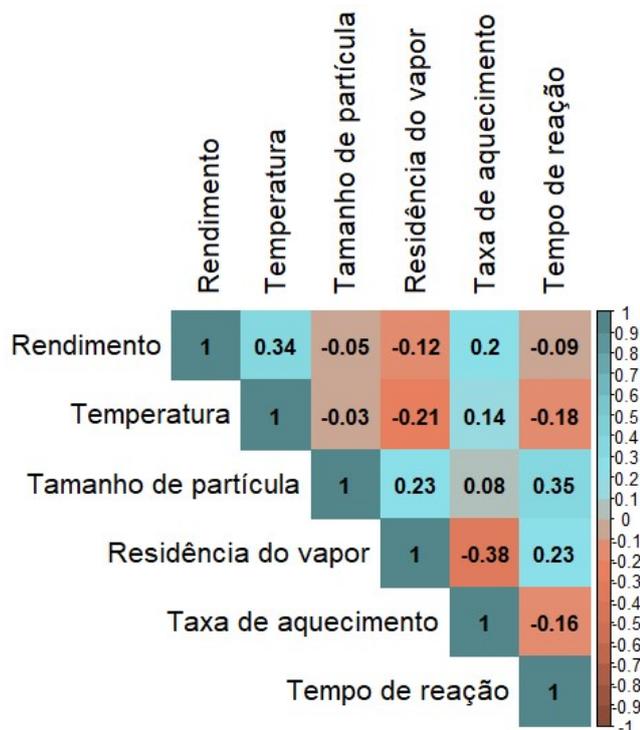
2016; Pütün; Apaydm; Pütün, 2004; Sharma; Sheth, 2015; Tsai; Lee; Chang, 2007; Varma; Mondal, 2017), o que pode ser confirmado na Figura 12. As médias do tempo de residência do vapor e do tamanho das partículas no estudo foram 142,1 s e 1,1 mm, respectivamente. A maior concentração dos estudos ocorreu em valores menores da faixa analisada, o que é confirmado pela mediana.

A composição da biomassa foi amplamente analisada na pirólise (Akhtar; Amin, 2012; Casoni *et al.*, 2015; Jung; Kang; Kim, 2008; Kim *et al.*, 2013; Omar *et al.*, 2011; Părpăriță *et al.*, 2014; Qu *et al.*, 2011; Quan; Gao; Song, 2016). As concentrações de cinzas, umidade, voláteis e carbono fixo variaram de 0-56,8%, 0-78,1%, 8,19-98,06% e 0,11-53,02%, respectivamente. As biomassas analisadas apresentaram, em geral, altos valores de material volátil e baixa umidade e teor de carbono fixo (Figura 12). A média da composição elementar das biomassas foi 49,57%, 6,44%, 41,86% e 1,79% para carbono, hidrogênio, oxigênio e nitrogênio, respectivamente.

Embora os estudos presentes na literatura relatem a grande importância dos teores de celulose, lignina e hemicelulose no processo, poucas pesquisas analisam a influência desses componentes. Assim, uma quantidade menor de observações dessas variáveis está presente no banco de dados. As abordagens experimentais para determinar essas propriedades são caras e demoradas (Xing *et al.*, 2019). A concentração de material lignocelulósico varia de 0 a 100% no banco de dados, para os três componentes, ou seja, os componentes puros também foram analisados para a obtenção do bio-óleo. A concentração média de celulose é maior do que hemicelulose e lignina na biomassa, o que pode favorecer a formação de líquidos.

Análises de correlação entre o rendimento de bio-óleo e as variáveis analisadas também foram realizadas para confirmar a tendência dos dados. A Figura 13 mostra uma correlação positiva da temperatura de pirólise e da taxa de aquecimento com o rendimento do bio-óleo, esse comportamento também foi observado por Arazo *et al.* (2017), Ellens; Brown, (2012) e Isa *et al.* (2011). Processos que utilizam maior taxa de aquecimento e temperatura, como a pirólise rápida e flash, reduzem as limitações de transferência de calor e massa, resultando em rendimentos altos de bio-óleo, mas a influência de outros fatores, como a composição da biomassa, pode alterar esse comportamento.

Figura 13 - Matriz de correlação entre variáveis de processo e a produção de bio-óleo.



Fonte: O autor, 2023.

O rendimento do bio-óleo tem uma correlação negativa fraca com a tempo de residência do vapor, tamanho de partícula e tempo de reação do processo. Em geral, é aceito que um curto tempo de residência do vapor é adequada para obter melhores rendimentos de bio-óleo, mas resultados opostos também são encontrados na literatura. Abnisa, Wan Daud e Sahu (2011) estudando a otimização do processo de pirólise de resíduos de casca de palmeira encontraram um efeito negativo do aumento da vazão de gás no processo de pirólise com um coeficiente de -1,57. Jourabchi, Gan e Ng (2014) observaram o aumento do rendimento do bio-óleo da pirólise de torta de pinhão manso quando a velocidade do N₂ reduziu de 6,7 cm/s para 0,0078 cm/s.

É necessário um tempo mínimo de reação para obter o resultado desejado no processo. Porém, um tempo de reação muito longo leva a uma redução no rendimento de bio-óleo (Bartoli *et al.*, 2016; Tsai; Lee; Chang, 2007). Isa *et al.* (2011) mostraram que um curto tempo de reação é apropriado para obter rendimentos mais elevados de líquidos. No modelo proposto, para a otimização da pirólise da casca de arroz em um reator de leito fixo, o coeficiente obtido para o tempo de reação foi -2,58, evidenciando o efeito negativo do parâmetro no processo. Abnisa, Wan Daud e Sahu (2011) descobriram que o tempo é um fator

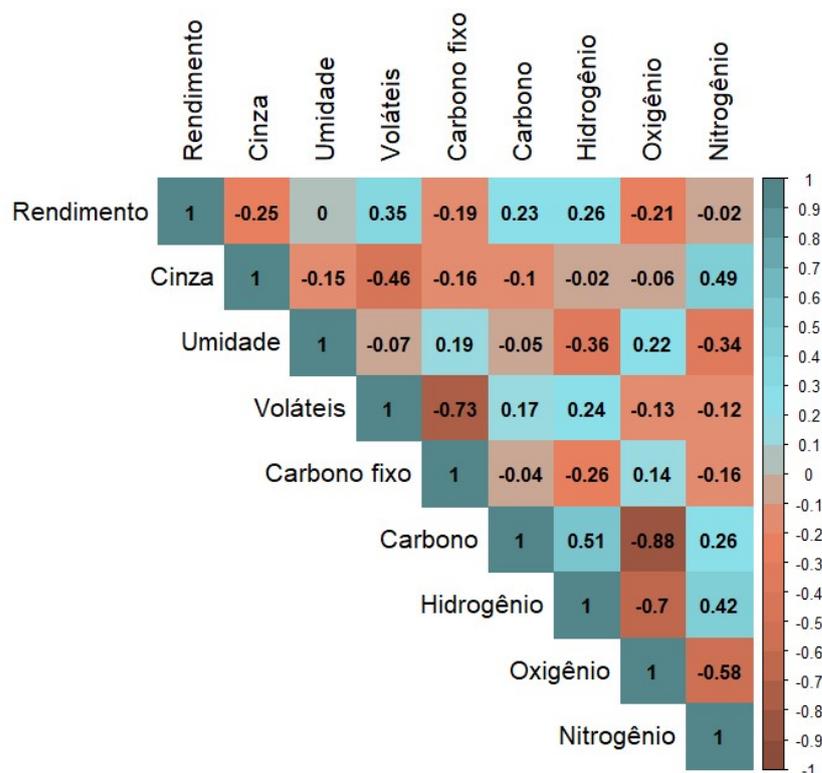
insignificante na otimização da produção de bio-óleo, mas as interações entre o tempo de reação e a temperatura são significativas na pirólise da casca da palma.

A taxa de aquecimento da biomassa com maior tamanho de partícula é menor devido à sua baixa condutividade térmica, o que resulta em uma baixa transferência de calor. Assim, uma biomassa de maior tamanho pode não ser completamente decomposta, o que aumenta a produção do carvão e diminui a produção de líquidos e gases (Biradar; Subramanian; Dastidar, 2014; Islam; Parveen; Haniu, 2010; Kang *et al.*, 2006). Uma redução no rendimento do bio-óleo com o aumento do tamanho das partículas foi observada por Ateş, Pütün e Pütün (2004); Beis, Onay e Koçkar (2002); Ellens e Brown (2012); Isa *et al.* (2011); Kang *et al.* (2006); Shen, *et al.* (2009). Esta tendência está relacionada a outras variáveis do processo como a velocidade do gás de arraste. Apesar de um menor tamanho de partícula favorecer o rendimento de bio-óleo, deve-se tomar precauções para que essa biomassa não seja arrastada do processo juntamente com os vapores formados.

A Figura 14 mostra os coeficientes de correlação de Spearman entre a composição elementar e aproximada da biomassa e o rendimento do bio-óleo. Pode-se verificar uma correlação negativa entre o rendimento e os teores de carbono fixo e de cinza e uma correlação positiva com o teor de voláteis. Assim, o maior rendimento de líquido é obtido com menor concentração de carbono fixo e cinza na biomassa, enquanto que um maior teor de voláteis aumenta o rendimento deste produto na pirólise (Abdullah; Gerhauser, 2008; Abnisa *et al.*, 2013; Fu *et al.*, 2010; Jung; Kang; Kim, 2008; Lee *et al.*, 2013; Lim *et al.*, 2016; Omar *et al.*, 2011; Pattiya; Sukkasi; Goodwin, 2012; Pattiya; Suttibak, 2012; Razuan *et al.*, 2010; Venderbosch; Prins, 2010). Nenhuma correlação foi observada entre a umidade e o teor de nitrogênio da biomassa com a produção de bio-óleo.

Em relação à composição elementar, pode-se observar uma correlação negativa entre rendimento de bio-óleo e teor de oxigênio e uma correlação positiva do produto com as variáveis teor de carbono e de hidrogênio. Esta mesma tendência foi descrita por Chouhan e Sarma (2013), Tang *et al.* (2020) e Ullah *et al.* (2021).

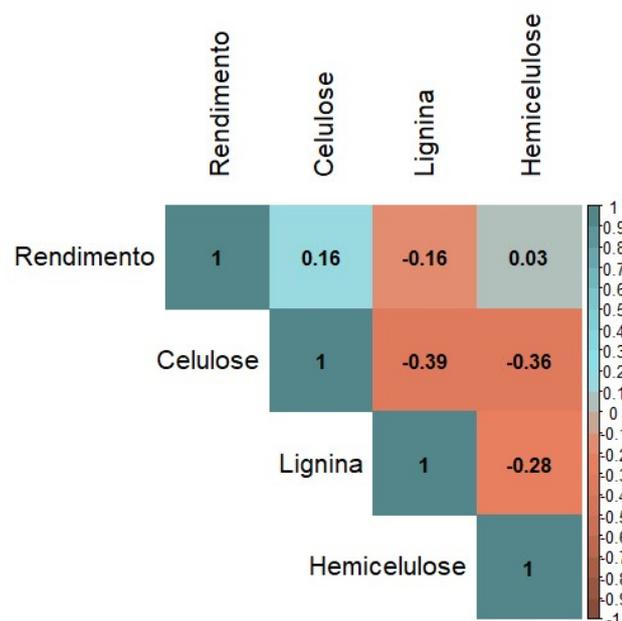
Figura 14 - Matriz de correlação entre composição e produção de bio-óleo



Fonte: O autor, 2023.

A composição lignocelulósica da biomassa também interfere no rendimento e na composição dos produtos da pirólise. A Figura 15 sugere uma correlação positiva fraca entre o conteúdo de celulose e o rendimento do bio-óleo. Por outro lado, um aumento na concentração de lignina na composição da biomassa acarreta uma redução no rendimento do líquido. A lignina tem uma maior estabilidade estrutural, o que dificulta sua decomposição (Akhtar; Amin, 2012). Esses resultados estão de acordo com a revisão da literatura, mas as correlações entre esses componentes e o rendimento do bio-óleo são fracas. Não foi observada correlação entre o rendimento do bio-óleo e o teor de hemicelulose.

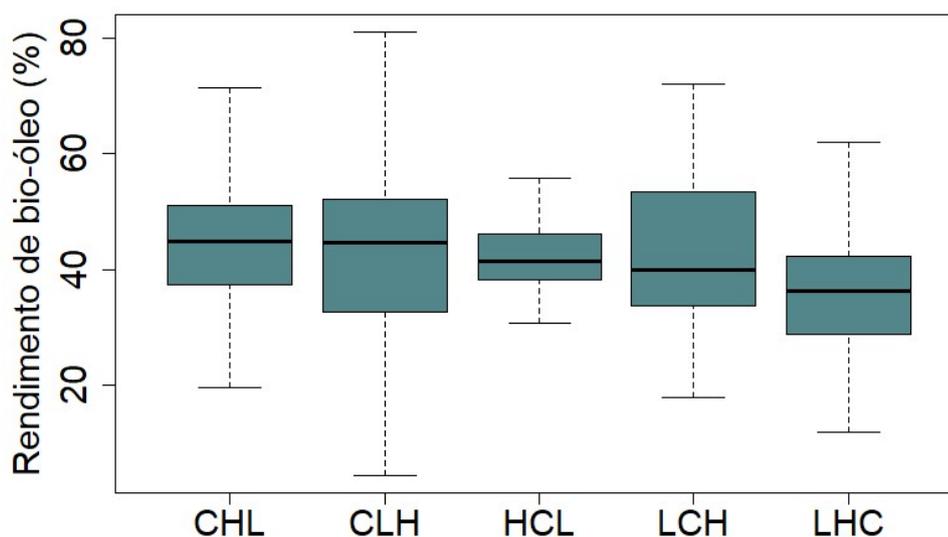
Figura 15 - Matriz de correlação entre composição lignocelulósica e rendimento.



Fonte: O autor, 2023.

Devido a importância da composição lignocelulósica na pirólise, uma análise mais aprofundada da influência destes compostos na produção de bio-óleo é mostrada na Figura 16.

Figura 16 - Rendimento de bio-óleo em diferentes composições da biomassa



C = Celulose; L = Lignina; H = Hemicelulose.
 CHL=C > H > L; CLH=C > L > H; HCL=H > C > L; LCH=L > C > H;
 LHC=L > H > C.

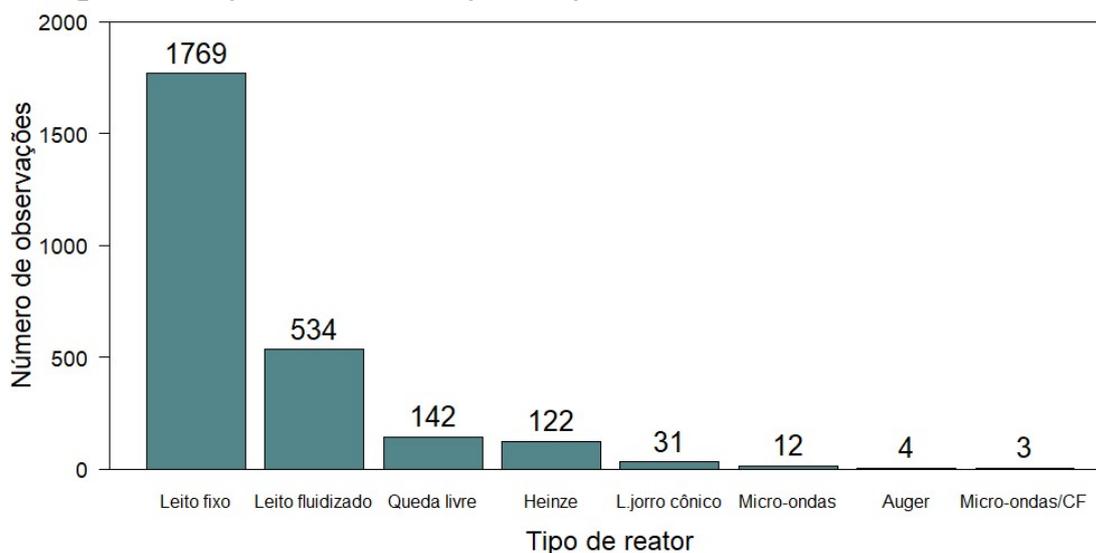
Fonte: O autor, 2023.

O teste de Kruskal-Wallis foi aplicado aos grupos analisados na Figura 16 e mostrou uma diferença significativa entre os grupos em relação ao rendimento de bio-óleo. A biomassa com maior concentração de celulose e menor concentração de lignina (CHL) levou a um maior rendimento de bio-óleo, com uma média de 45,17% em comparação ao grupo contendo maior concentração de lignina e menor concentração de celulose (LHC), com média de 36,43% de rendimento de bio-óleo. O aumento da lignina na biomassa causou um efeito negativo no rendimento. Porém, pode-se observar, que outros fatores estão influenciando no processo, com biomassa contendo altos teores de celulose produzindo um baixo rendimento de bio-óleo. Wang *et al.* (2011) mostraram que hemicelulose, celulose e lignina não agem independentemente durante a pirólise. Assim, as características de pirólise da biomassa não podem ser previstas apenas com base no comportamento térmico dos três componentes, devido à alta influência das interações entre os componentes e as condições do processo na distribuição dos produtos.

A lignina se decompõe em uma faixa de temperatura mais extensa em comparação com a celulose e a hemicelulose, que se degradam rapidamente em faixas de temperatura mais estreitas. A taxa e a extensão da decomposição de cada um desses componentes dependem dos parâmetros como, temperatura e taxa de aquecimento (Bridgwater, 1999). Portanto, em geral, as pesquisas utilizando as mesmas condições de processo mostram um maior rendimento de bio-óleo em biomassas com alta concentração de celulose, mas as condições podem ser ajustadas para que uma biomassa com alta concentração de lignina produza alto rendimento de bio-óleo. Como Demirbas (2006), que utilizando uma biomassa com concentração de celulose de 49,4%, na temperatura de 427 °C, obteve um rendimento de 38,3%, enquanto que na biomassa contendo 50,8% de lignina o rendimento foi de 39,4%, na temperatura de 527 °C. Shen *et al.* (2010) conseguiram um aumento no rendimento da pirólise de lignina de 32% para 43,2% aumentando a temperatura de 475 °C para 650 °C.

Vários reatores diferentes são utilizados para o processo de pirólise. O banco de dados contém oito tipos de reatores. A Figura 17 mostra o número de ocorrências dos reatores presente no banco. A grande maioria das amostras do banco de dados foram obtidas em pirólise em reator de leito fixo com 1769 observações.

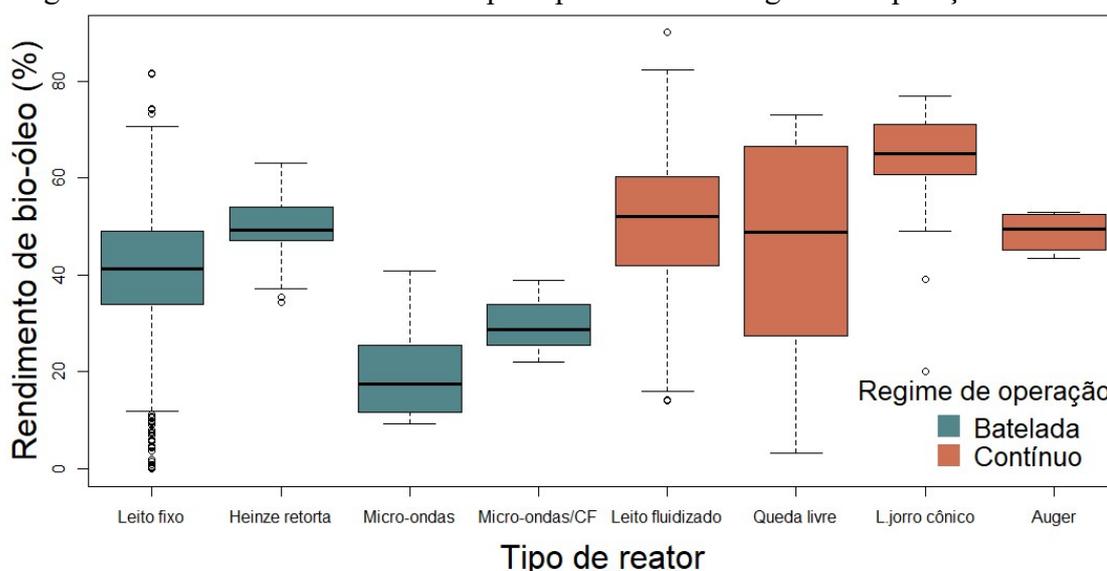
Figura 17 - Tipos de reatores de pirólise presentes no banco de dados.



Fonte: O autor, 2023.

Os reatores de Leito fixo, Heinze retorta, Micro-ondas e Micro-ondas com coluna de fracionamento foram utilizados operando em regime de batelada e os demais reatores trabalharam com alimentação contínua. Com isso, o banco de dados contém 1906 observações de reatores em batelada e 711 reatores operando continuamente. A Figura 18 mostra o rendimento de bio-óleo em relação ao tipo de reator e ao regime de operação.

Figura 18 - Rendimento de bio-óleo por tipo de reator e regime de operação.



Fonte: O autor, 2023.

A análise de Kruskal-Wallis mostrou que os tipos de reatores possuem diferenças significativas em relação ao rendimento de bio-óleo. O reator de leito de jorro cônico difere

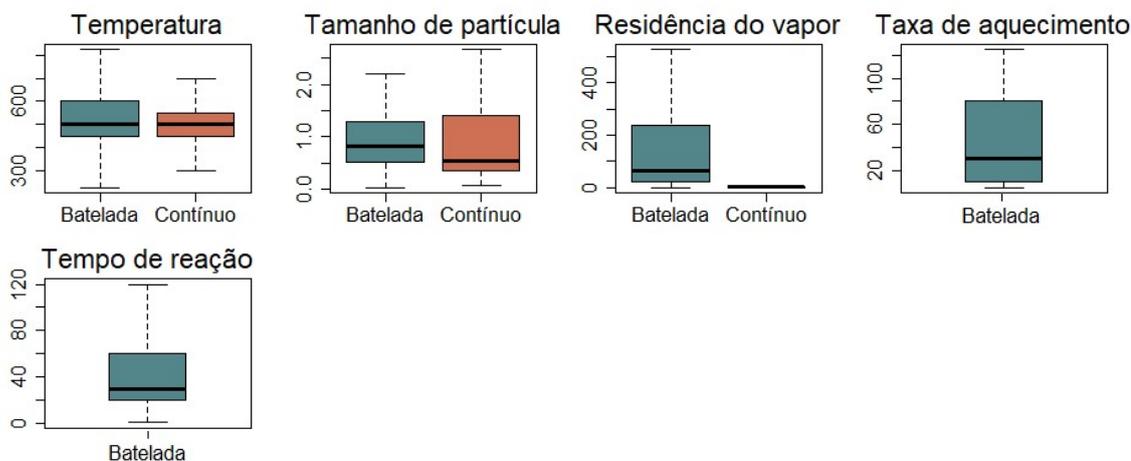
estatisticamente de todos os demais reatores, sendo o reator que levou a maior produção de bio-óleo com uma média de 63,5%. Makibar *et al.* (2015) obtiveram um rendimento de líquido de 69% a 455 °C na pirólise de *Populus nigra*, utilizando reator de leito de jorro cônico. Já Alvarez *et al.* (2014) produziram um rendimento de bio-óleo de 70% a 450 °C na pirólise da casca de arroz, neste mesmo tipo de reator. Os reatores de leito fluidizado e queda livre também produziram altos rendimentos de bio-óleo. Nos reatores de queda livre foi observada uma maior variabilidade nos resultados.

A análise de comparação realizada com o teste de Dunn mostrou que o rendimento de bio-óleo obtido nos reatores que operam em batelada possui diferença significativa em relação aos reatores contínuos. Em geral, os reatores operando em regime de batelada alcançaram menores rendimentos de bio-óleo. Sendo os melhores resultados obtidos em reatores com Heinze retorta. A pirólise em micro-ondas obteve um menor rendimento de líquido. Bartoli *et al.* (2016) usaram um micro-ondas para pirólise de *Arundo donax*. Um máximo de 40,9% de bio-óleo foi obtido no processo. A utilização de coluna de fracionamento na saída do reator em micro-ondas ocasionou um aumento do tempo de residência dos materiais no interior do forno. Com isso, ocorreu um aumento das reações de craqueamento dos vapores levando a uma menor produção de bio-óleo (na ordem de 2–7%).

A produção de bio-óleo foi favorecida por altas taxas de aquecimento e tempos de residência curtos, características dos reatores contínuos. Os reatores contínuos obtiveram uma média de rendimento superior (49,86%) em relação aos reatores em batelada (41,48%).

O tipo de regime de operação dos reatores altera as variáveis utilizadas em cada processo. A Figura 19 mostra a distribuição das variáveis por regime de operação. Nesta figura os *outliers* foram removidos para uma melhor visualização da faixa de maior concentração dos dados. Pode-se observar que as variáveis, taxa de aquecimento e tempo de reação estão presentes somente nos reatores em batelada. Além disso, a Figura 19 mostra que o reator em batelada foi analisado em uma faixa mais ampla de temperatura. Por outro lado, os reatores contínuos apresentaram uma maior variabilidade nos tamanhos de partículas analisados e trabalharam com baixos tempos de residência do vapor. A Figura indica, também, que as variáveis não seguem uma distribuição normal.

Figura 19 - Distribuição das variáveis por regime de operação dos reatores.



Fonte: O autor, 2023.

O teste de normalidade de Shapiro-Wilk foi aplicado aos oito subconjuntos de dados utilizados para a modelagem do rendimento de bio-óleo. O teste mostrou que, em todos os casos, os dados não são normalmente distribuídos e a transformação Box-Cox foi aplicada para induzir a normalidade. Além disso, a análise de variância mostrou que não há baixa variância dos dados, em nenhum dos casos analisados. Os dados foram centralizados e escalonados, dessa forma cada variável tem média 0 e desvio padrão 1.

A análise visual e estatística mostrou a presença de outliers em todos os casos. A transformação de sinal espacial foi aplicada para eliminar os efeitos dos outliers na modelagem sem remover essas amostras do conjunto analisado.

5.2 Modelagem de dados completos

A análise exploratória dos dados mostrou uma grande quantidade de dados ausentes. Depois de remover esses valores, o maior subconjunto de dados foi o caso V, com 358 observações restantes. Os casos contendo os compostos lignocelulósicos da biomassa como preditores do bio-óleo (III, IV, VII e VIII) resultaram em conjuntos de dados menores para a modelagem, isso ocorreu devido ao menor número de observações de celulose, lignina e hemicelulose no banco de dados. A Tabela 10 mostra o número de observações e variáveis restantes usadas para modelar o rendimento do bio-óleo em cada estudo de caso após o pré-

processamento dos dados, além do número de diferentes biomassas presentes e do número de outliers identificados.

Tabela 10 - Número de observações, variáveis, biomassas e outliers nos estudos

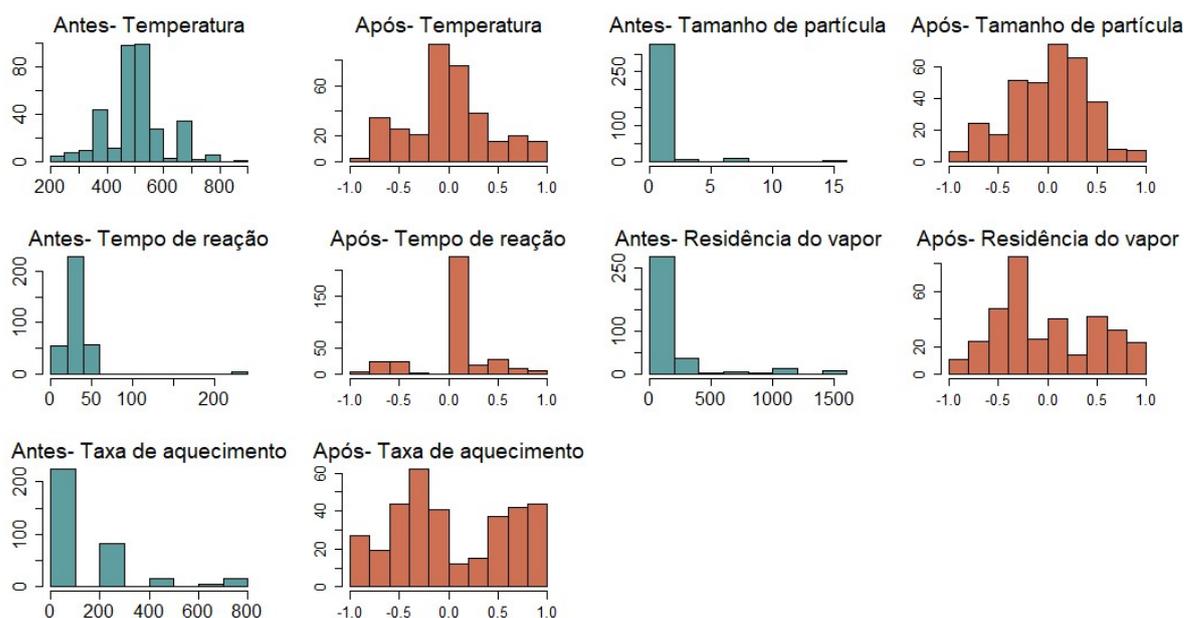
Variável	Caso I	Caso II	Caso III	Caso IV	Caso V	Caso VI	Caso VII	Caso VIII
Observações	343	305	93	80	358	225	114	70
Conjunto de treino	258	229	70	60	269	169	86	53
Conjunto de teste	85	76	23	20	89	56	28	17
Número de variáveis	5	12	7	14	3	10	6	10
Número de biomassas	35	30	14	13	47	22	19	8
Outliers	121	104	38	24	115	75	25	15

Fonte: O autor, 2023.

5.2.1 Estudo de caso I

No caso I, apenas variáveis de processo foram analisadas no rendimento do bio-óleo obtido em reator operando em batelada. O teste de correlação de Spearman não mostrou correlação forte entre as variáveis preditoras. A Figura 20 mostra a distribuição dos dados antes e após o pré-processamento. Uma melhora na assimetria dos dados pode ser observada após a transformação Box-Cox. Os valores de lambda utilizados na transformação Box-Cox, em todos os estudos de caso, estão presentes no Anexo C.

Figura 20 - Distribuição dos dados antes e após o pré-processamento do caso I



Fonte: O autor, 2023.

Após o pré-processamento dos dados os treze modelos propostos inicialmente para a modelagem foram utilizados. Além disso, como uma forma de verificar a existência de algum outro modelo promissor para a modelagem do rendimento de bio-óleo, os mesmos dados foram utilizados no programa Python com a ferramenta LazyPrediction. Esta ferramenta realiza uma modelagem simultânea com 42 modelos de regressão mostrando parâmetros como RMSE e R^2 para cada modelo.

A Tabela 11 mostra o resultado do LazyPredict para os dados de teste do caso I. O resultado obtido para os demais casos encontra-se no Anexo D. Os modelos que apareceram entre os cinco melhores na análise do LazyPredict foram testados no software R com ajuste de parâmetros. Esta análise foi realizada para os oito estudos de caso e os resultados mostraram que apenas o modelo Florestas Extremamente Aleatórias deveria ser incluído à listagem de modelos previamente selecionados.

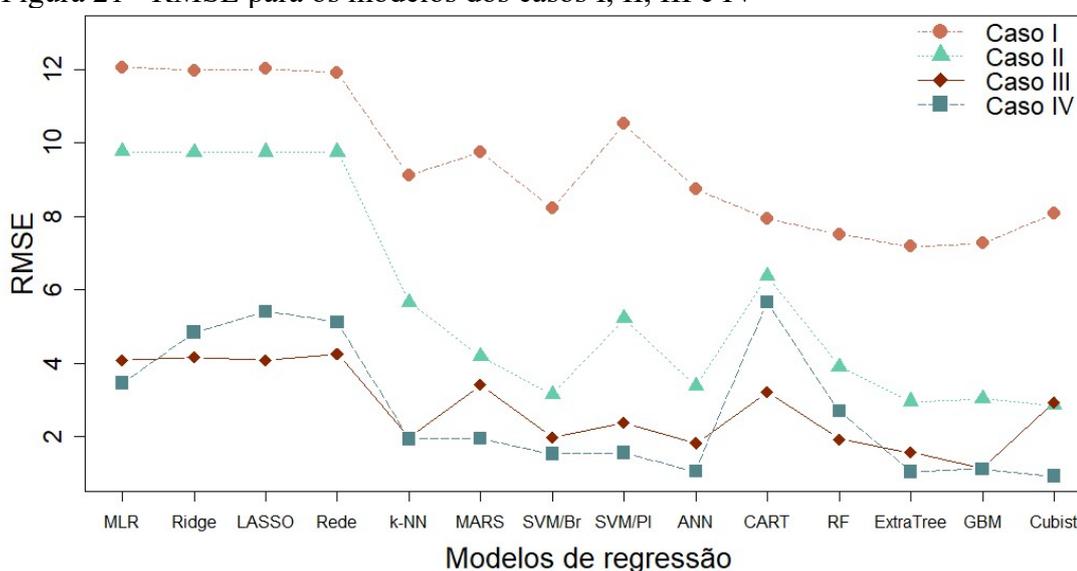
Tabela 11 - Resultados da regressão realizada por LazyPredict para o caso I

Modelo	R ²	RMSE (%)	Tempo de corrida (s)
RandomForestRegressor	0,71	7,20	0,23
GradientBoostingRegressor	0,71	7,27	0,20
HistGradientBoostingRegressor	0,68	7,57	0,14
LGBMRegressor	0,67	7,71	0,13
XGBRegressor	0,65	7,91	0,14
BaggingRegressor	0,63	8,11	0,04
ExtraTreeRegressor	0,63	8,14	0,01
ExtraTreesRegressor	0,62	8,27	0,16
AdaBoostRegressor	0,60	8,53	0,11
DecisionTreeRegressor	0,58	8,66	0,01
KNeighborsRegressor	0,54	9,06	0,02
NuSVR	0,24	11,70	0,02
OrthogonalMatchingPursuitCV	0,23	11,74	0,03
LassoCV	0,20	12,00	0,06
LarsCV	0,20	12,01	0,05
LassoLarsCV	0,20	12,01	0,02
LassoLarsIC	0,20	12,01	0,02
SVR	0,20	12,02	0,01
SGDRegressor	0,19	12,07	0,01
TransformedTargetRegressor	0,19	12,07	0,01
LinearRegression	0,19	12,07	0,01
Lars	0,19	12,07	0,03
Ridge	0,19	12,07	0,01
RidgeCV	0,19	12,08	0,01
ElasticNetCV	0,19	12,08	0,07
OrthogonalMatchingPursuit	0,19	12,11	0,02
PoissonRegressor	0,18	12,12	0,03
BayesianRidge	0,18	12,13	0,04
LassoLars	0,18	12,18	0,01
Lasso	0,18	12,18	0,03
LinearSVR	0,16	12,33	0,01
ElasticNet	0,14	12,42	0,02
GammaRegressor	0,13	12,55	0,03
HuberRegressor	0,12	12,55	0,02
TweedieRegressor	0,12	12,56	0,01
GeneralizedLinearRegressor	0,12	12,56	0,01
DummyRegressor	-	13,67	0,01
RANSACRegressor	-	13,97	0,14
PassiveAggressiveRegressor	-	16,13	0,01
MLPRegressor	-	20,86	0,38
KernelRidge	-	47,89	0,03

Fonte: O autor, 2023.

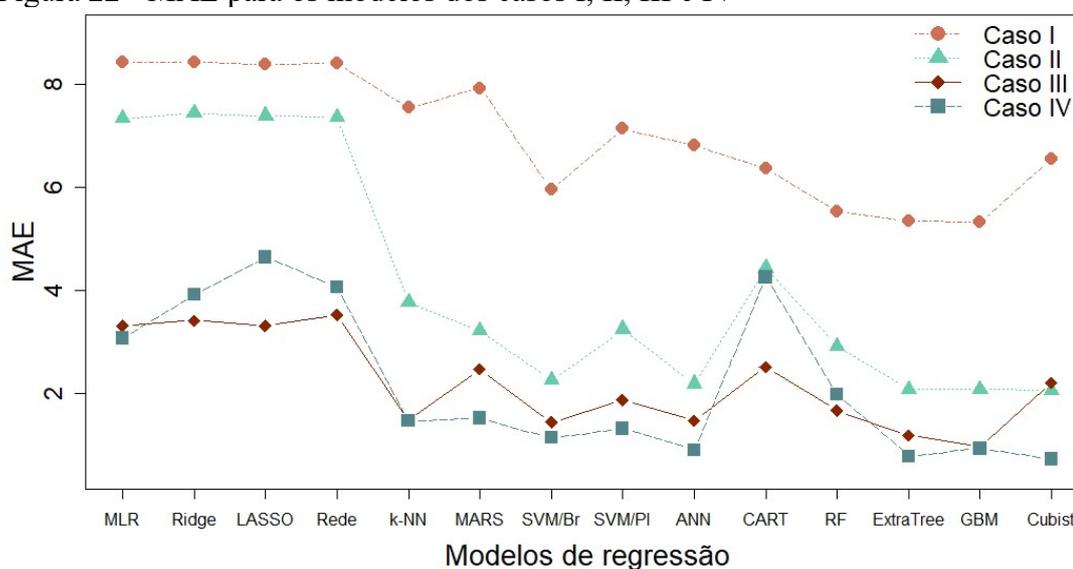
Os modelos de regressão multivariada foram aplicados ao conjunto de treinamento com 258 observações. As métricas para avaliação dos modelos são apresentadas nas Figuras 21, 22 e 23. O modelo ExtraTree apresentou o melhor ajuste aos dados com RMSE e MAE de 7,18% e 5,34%, respectivamente. O R^2 foi de 0,73, indicando que o modelo explicou 73% da variação dos dados. Cao, Xin e Yuan (2016) obtiveram um R^2 de 0,96 usando o modelo de máquina de vetores de suporte de mínimos quadrados para prever o rendimento do carvão vegetal obtido por pirólise de esterco bovino usando como variáveis preditoras, temperatura, taxa de aquecimento, tempo de reação e umidade de biomassa em reator batelada.

Figura 21 - RMSE para os modelos dos casos I, II, III e IV

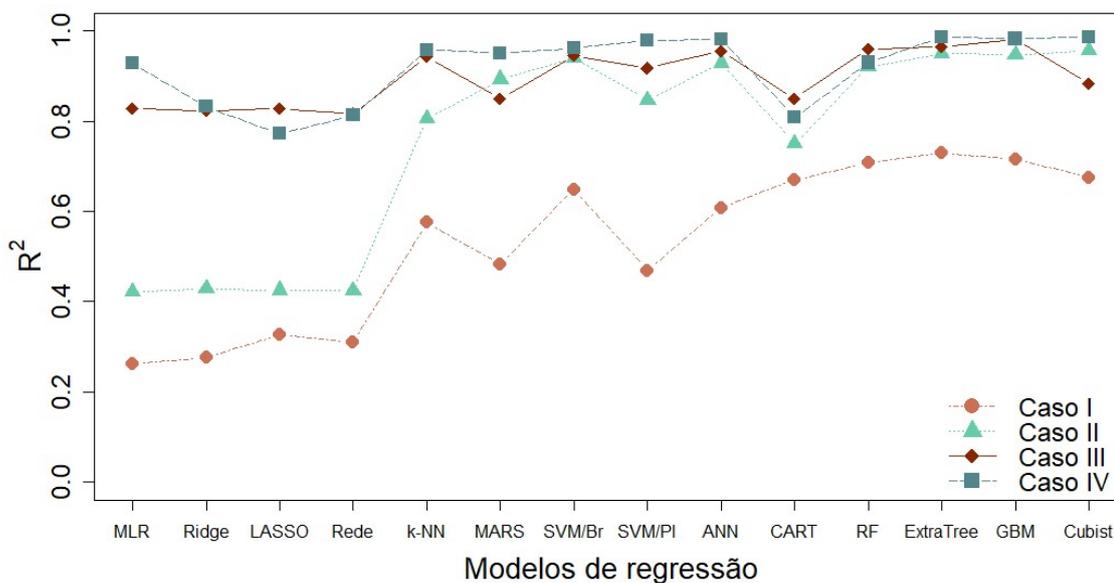


Fonte: O autor, 2023.

Figura 22 - MAE para os modelos dos casos I, II, III e IV



Fonte: O autor, 2023.

Figura 23 - R^2 para os modelos dos casos I, II, III e IV

Fonte: O autor, 2023.

Como pode ser observado nas Figuras 21, 22 e 23, os modelos lineares apresentaram uma baixa precisão de predição para o rendimento do bio-óleo, indicando que as variáveis não possuem relação linear. Esse resultado foi consistente com o obtido por Tang *et al.* (2020).

A modelagem empregada permite analisar a importância das variáveis nos modelos aplicados. A Figura 24 mostra que, para o reator em batelada, analisando apenas as variáveis do processo (Caso I), as variáveis temperatura e taxa de aquecimento possuem uma maior importância para o modelo ExtraTree. Também pode ser observado que o tempo de residência do vapor e tamanho de partícula são os preditores menos significativos para o modelo. Esse resultado justifica que, para reatores em batelada, exista um número mais significativo de estudos que analisa a temperatura como a variável mais crítica do processo.

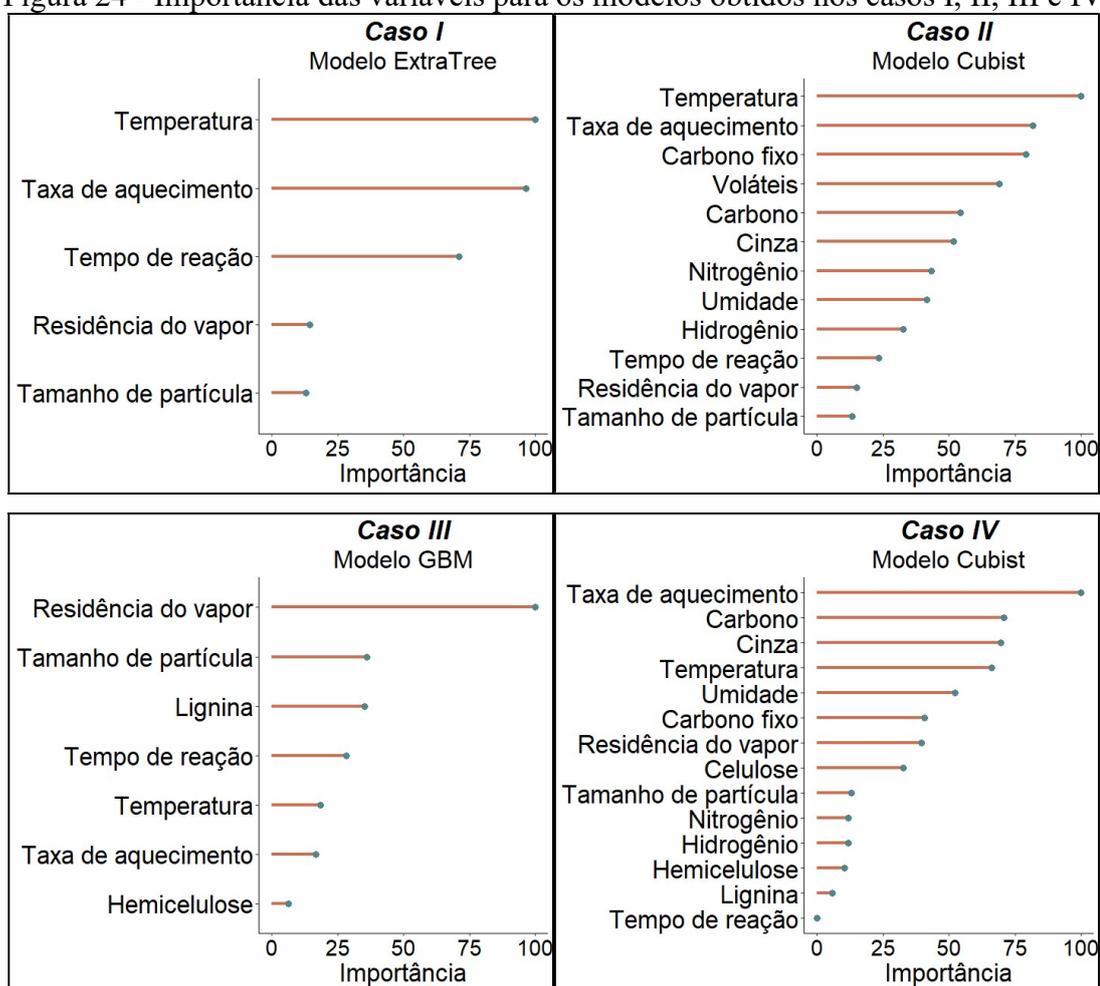
5.2.2 Estudo de caso II

O estudo de caso II inclui a composição elementar e aproximada da biomassa ao conjunto analisado no caso I. A análise exploratória mostrou que a maioria dos dados faltantes está relacionada a variáveis de processo. As variáveis de tempo de residência de vapor e tempo de reação contêm 938 e 892 amostras com dados ausentes em observações relacionadas

ao reator em batelada, respectivamente. Com a remoção destas amostras o conjunto final analisado possui 305 observações completas.

O teste de correlação de Spearman mostrou forte correlação entre as variáveis teor de oxigênio e teor de carbono, com coeficiente de correlação de 0,91. A variável carbono foi mantida na modelagem por apresentar melhores resultados nas métricas dos modelos em comparação com a variável oxigênio. Além disso, a variável oxigênio é obtida indiretamente na análise elementar da biomassa. Assim, seu cálculo pode carregar possíveis erros que ocorreram na obtenção dos demais componentes. O pré-processamento dos dados resultou em melhora na assimetria dos dados.

Figura 24 - Importância das variáveis para os modelos obtidos nos casos I, II, III e IV



Fonte: O autor, 2023.

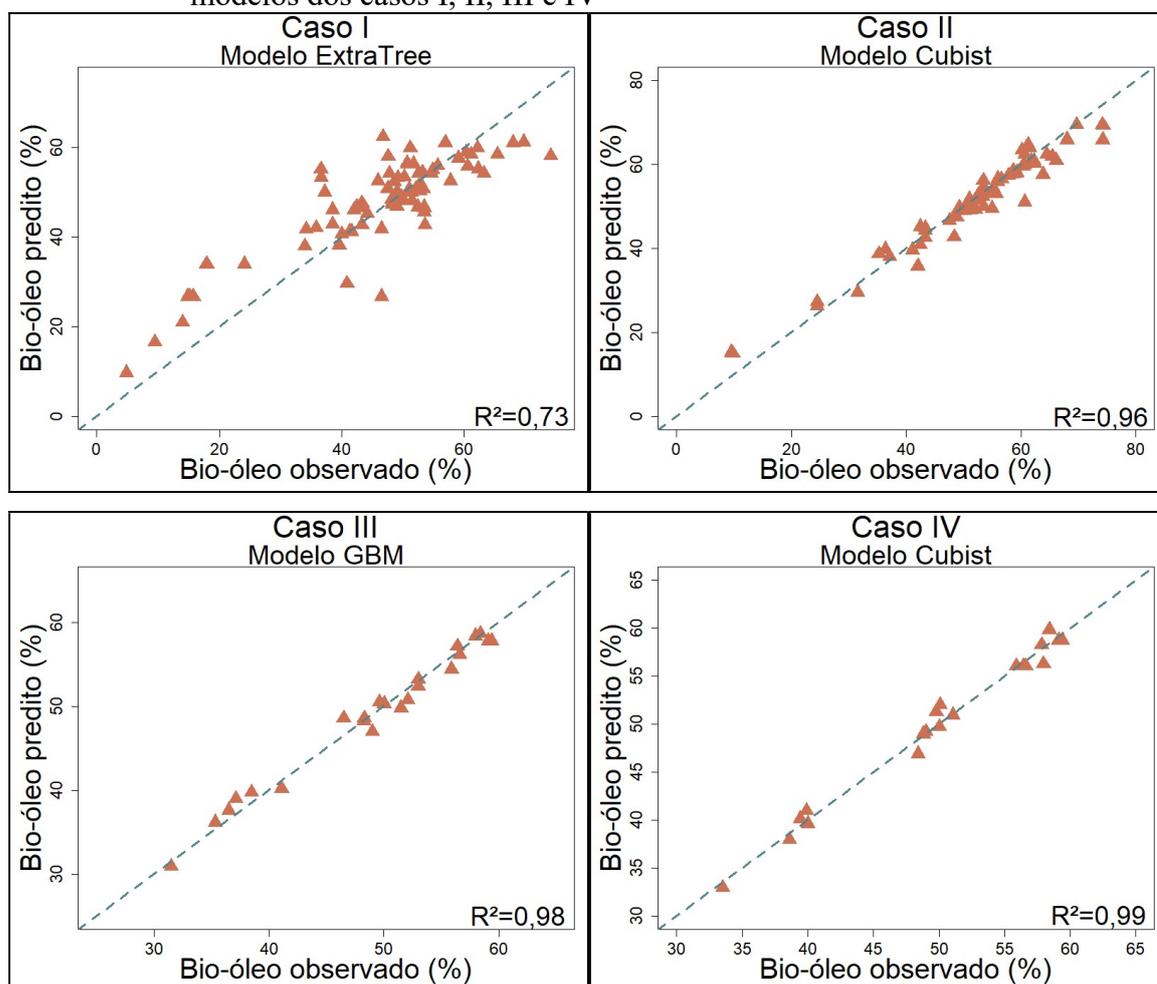
Os modelos de regressão multivariada foram aplicados a 229 observações dos dados de teste. O método *Cubist* apresentou menor RMSE (2,85%) e MAE (2,06%), sendo o modelo com melhor desempenho de predição. O coeficiente de determinação foi de 0,96. Os modelos

GBM e ExtraTree também apresentaram bons resultados de predição, ambos com R^2 de 0,95. Esse resultado foi superior ao obtido por Merdun e Sezgin (2018) que utilizaram a Rede Neural Artificial para modelar o rendimento do bio-óleo em amostras obtidas em um reator batelada. Os autores obtiveram R^2 de 0,81 e RMSE de 5,71. Tang *et al.* (2020) aplicaram o modelo Floresta Aleatória para produção de bio-óleo usando como variáveis a temperatura, taxa de aquecimento, tamanho de partícula, vazão de nitrogênio e composição elementar ou aproximada. Para o modelo utilizando as variáveis de processo e composição elementar o R^2 obtido foi de 0,87 e o RMSE de 3,05. Para a composição aproximada, o coeficiente de determinação e o RMSE foram de 0,92 e 2,13, respectivamente.

A Figura 24 mostra que os preditores temperatura e taxa de aquecimento têm maior importância para a predição no modelo *Cubist*, assim como para o modelo ExtraTree do caso I. A composição da biomassa se mostrou importante para o modelo, justificando sua inclusão na modelagem.

Na Figura 25, é possível observar um melhor ajuste dos dados do estudo II em relação ao caso I. Os resultados mostram, mais uma vez, que a inclusão de variáveis de composição beneficia a modelagem da produção do bio-óleo.

Figura 25 - Valores observados e preditos de rendimento de bio-óleo para os melhores modelos dos casos I, II, III e IV



Fonte: O autor, 2023.

5.2.3 Estudo de caso III

No estudo de caso III, os materiais lignocelulósicos da biomassa, ou seja, lignina, celulose e hemicelulose foram incluídas como variáveis preditoras do rendimento do bio-óleo em comparação com as variáveis analisadas no caso I. O conjunto de dados desta modelagem é menor, em relação aos outros casos já analisados, devido ao baixo número de observações das variáveis utilizadas, como observado na Tabela 9. Do total de 1906 observações sobre os reatores em batelada a variável hemicelulose contém 1218 dados faltantes. As variáveis lignina e celulose estão presentes em apenas 758 e 928 amostras, respectivamente. Assim, o conjunto contendo apenas observações completas foi reduzido para 93 observações.

O teste de correlação de Spearman mostrou um coeficiente de correlação de 0,89 entre a celulose e a hemicelulose, o que é considerado uma forte correlação. A variável hemicelulose foi mantida da modelagem pois, os modelos incluindo esta variável apresentam melhor desempenho de predição.

Após o pré-processamento dos dados, o teste de normalidade Shapiro-Wilk mostrou que os preditores não seguem uma distribuição normal. Porém, como nos casos I e II, houve melhora da assimetria.

O modelo de regressão que apresentou maior precisão de predição foi o GBM, com maior coeficiente de determinação e menores RMSE e MAE, de 0,98, 1,15% e 0,97%, respectivamente. Pode-se observar que, em relação ao caso I, houve melhora na precisão de predição com a adição das variáveis hemicelulose e lignina como preditores de bio-óleo. Além disso, a utilização de composição lignocelulósica levou a melhores resultados se comparado a utilização da composição aproximada e elementar da biomassa (Caso II).

Conforme mostrado nas Figuras 21, 22 e 23 e caso III, houve uma melhoria nos modelos lineares usando variáveis de composição para prever o rendimento do bio-óleo. No entanto, os modelos não lineares continuaram a ter um melhor desempenho.

A Figura 24 mostra que a variável residência do vapor possui uma ampla importância para o modelo GBM. Em relação às variáveis lignocelulósicas, a hemicelulose é a variável de menor importância para o modelo.

Na Figura 25, referente aos casos III e IV, é possível observar um menor número de observações presentes no conjunto de teste. No entanto, existe um bom ajuste entre as variáveis previstas e observadas pelos modelos.

5.2.4 Estudo de caso IV

Com o aprimoramento dos resultados de predição com a inclusão de variáveis de composição e materiais lignocelulósicos da biomassa na modelagem, foi realizado o estudo de caso IV incluindo todas as variáveis dos estudos anteriores para prever o rendimento do bio-óleo.

A análise de correlação de Spearman mostrou forte correlação entre oxigênio e carbono e entre carbono fixo e voláteis, com coeficientes de correlação de 0,88 e 0,94,

respectivamente. Os preditores teor de oxigênio e teor de voláteis foram removidos para a obtenção de variáveis não correlacionadas, a remoção obedeceu ao critério estipulado de remover variáveis que reduzem o desempenho dos modelos. O conjunto de dados completo após o pré-processamento contém 80 observações para a modelagem de bio-óleo com 14 variáveis preditoras.

Conforme mostrado pelas métricas dos modelos (Figuras 21, 22 e 23), os modelos *Cubist* e ExtraTree apresentaram o melhor desempenho na regressão de dados. O modelo *Cubist* mostrou-se ligeiramente superior com RMSE, MAE e R^2 de 0,92%, 0,73% e 0,99, respectivamente.

Em relação à importância das variáveis o tempo de reação não se mostrou importante para o *Cubist*, podendo ser removido da modelagem, o que reduz a complexidade do modelo. Além disso as composições de cinza e carbono da biomassa apresentaram ser tão importante quanto a temperatura empregada. Tang *et al.* (2020) analisando variáveis de processo e composição aproximada, observaram maior importância das variáveis de composição para a produção de bio-óleo. O carbono fixo foi uma variável importante para o modelo Floresta Aleatória. Além disso, a temperatura foi a variável menos importante para o modelo. Merdun e Sezgin (2018) mostraram que a temperatura e a taxa de aquecimento, na modelagem de produção de bio-óleo usando composição elementar e aproximada da biomassa, não melhoram o desempenho do modelo Rede Neural Artificial.

A Tabela 12 mostra as ferramentas de diagnóstico dos melhores modelos para reatores em batelada. O uso de todas as variáveis no processo de modelagem (Caso IV) resultou em um modelo com um pequeno erro de previsão de produção de bio-óleo. Porém, considerando a maior escassez de dados lignocelulósicos na literatura e a maior dificuldade na análise para a obtenção destes dados, a modelagem apenas com as variáveis de processo e variáveis de composição aproximada e elementar, ou seja, variáveis utilizadas no estudo de caso II, é suficiente para prever o rendimento do bio-óleo com alta precisão de predição.

Tabela 12 - Ferramentas de diagnóstico para o melhor modelo do reator em batelada

Estudo de caso	Modelo	RMSE	MAE	R^2
Case I	ExtraTree	7,18	5,34	0,73
Case II	<i>Cubist</i>	2,85	2,06	0,96
Case III	GBM	1,15	0,97	0,98
Case IV	<i>Cubist</i>	0,92	0,73	0,99

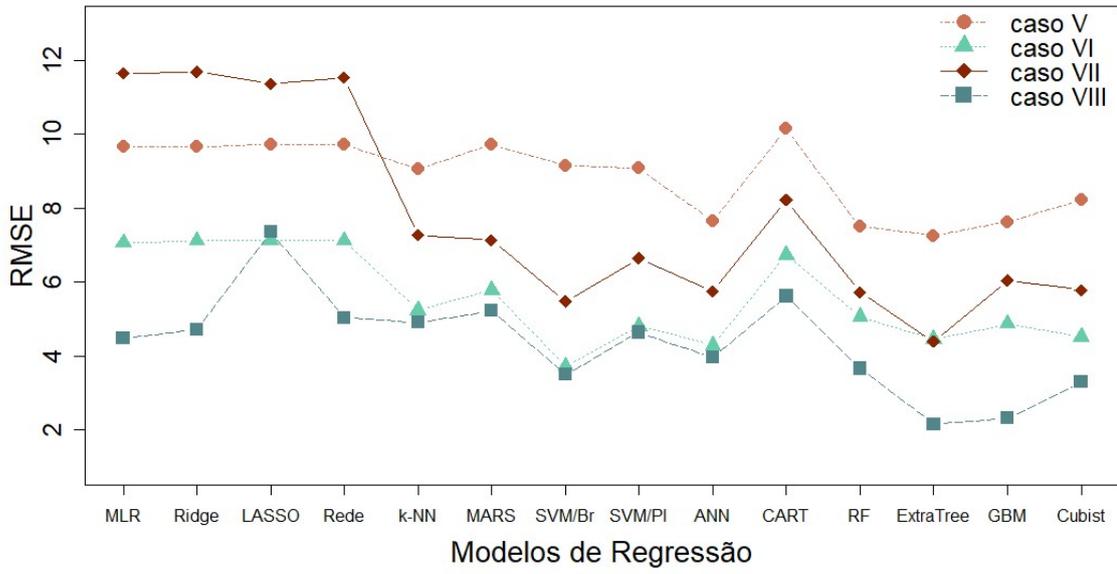
Fonte: O autor, 2023.

5.2.5 Estudo de caso V

O estudo de caso V contém apenas variáveis de processo em reatores contínuos. Nesse caso, apenas três variáveis preditoras, temperatura, tamanho de partícula e tempo de residência do vapor, foram usadas para modelar o rendimento do bio-óleo. O conjunto completo de dados analisados contém 358 observações, 269 usadas para treinamento e 89 para teste. As análises realizadas não mostraram forte correlação entre os preditores. Além disso, o teste de Shapiro-Wilk mostrou que nenhuma variável segue a distribuição normal dos dados. A transformação Box-Cox levou a uma melhora acentuada na assimetria dos dados.

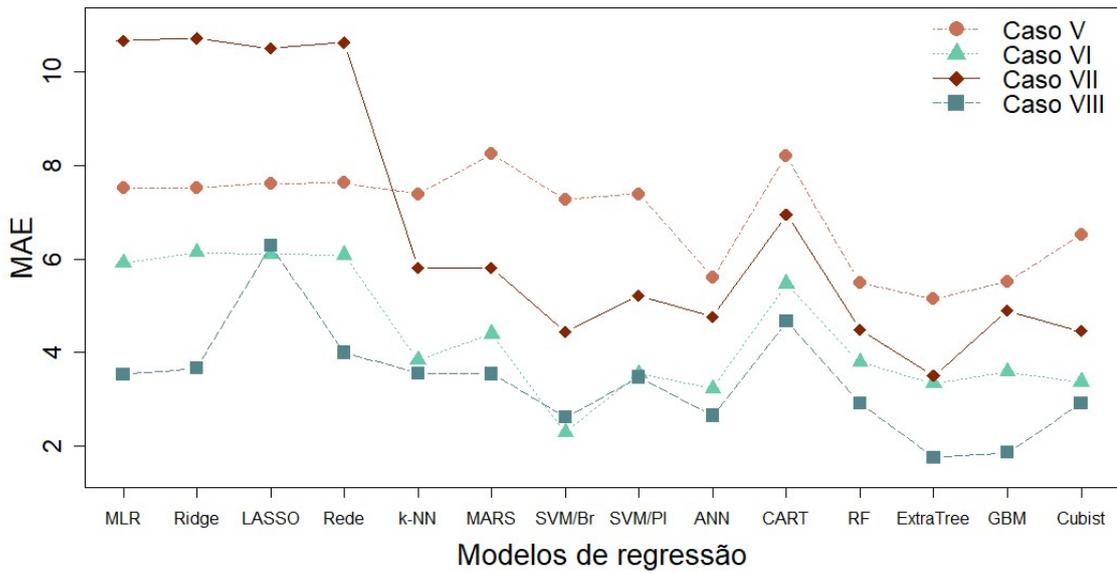
As Figuras 26, 27 e 28 mostram as ferramentas de diagnóstico para os estudos de caso envolvendo os reatores operando continuamente. Pode-se observar, para o caso V, que o modelo Florestas Extremamente Aleatórias obteve o melhor ajuste na regressão, com RMSE, MAE e R^2 de 7,26%, 5,14% e 0,50, respectivamente. Os resultados mostram que o modelo explica apenas 50% da variabilidade dos dados, isso se deve, provavelmente, ao baixo número de variáveis preditoras utilizado. As Figuras que apresentam as métricas dos modelos (Figuras 26, 27 e 28) também mostraram que a Regressão Linear Múltipla e os demais métodos lineares de regressão apresentaram desempenho mais limitado do que os modelos não lineares, comprovando que o rendimento de bio-óleo não tem uma relação linear simples com as variáveis. Dos modelos não lineares o CART apresentou uma menor capacidade preditiva do rendimento de bio-óleo.

Figura 26 - RMSE para os modelos dos casos V, VI, VII e VIII

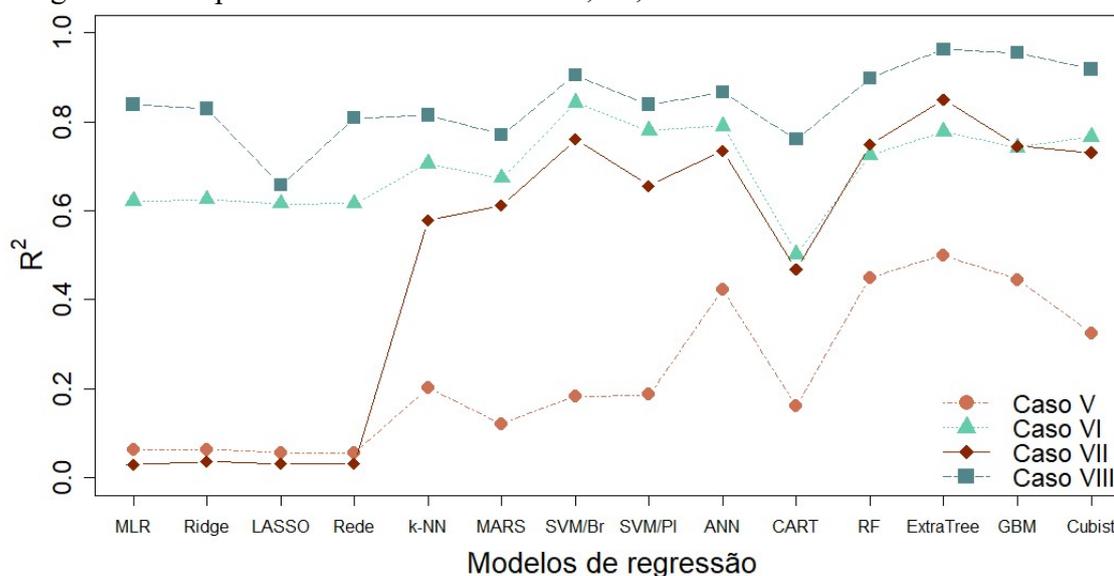


Fonte: O autor, 2023.

Figura 27 - MAE para os modelos dos casos V, VI, VII e VIII



Fonte: O autor, 2023.

Figura 28 - R^2 para os modelos dos casos V, VI, VII e VIII

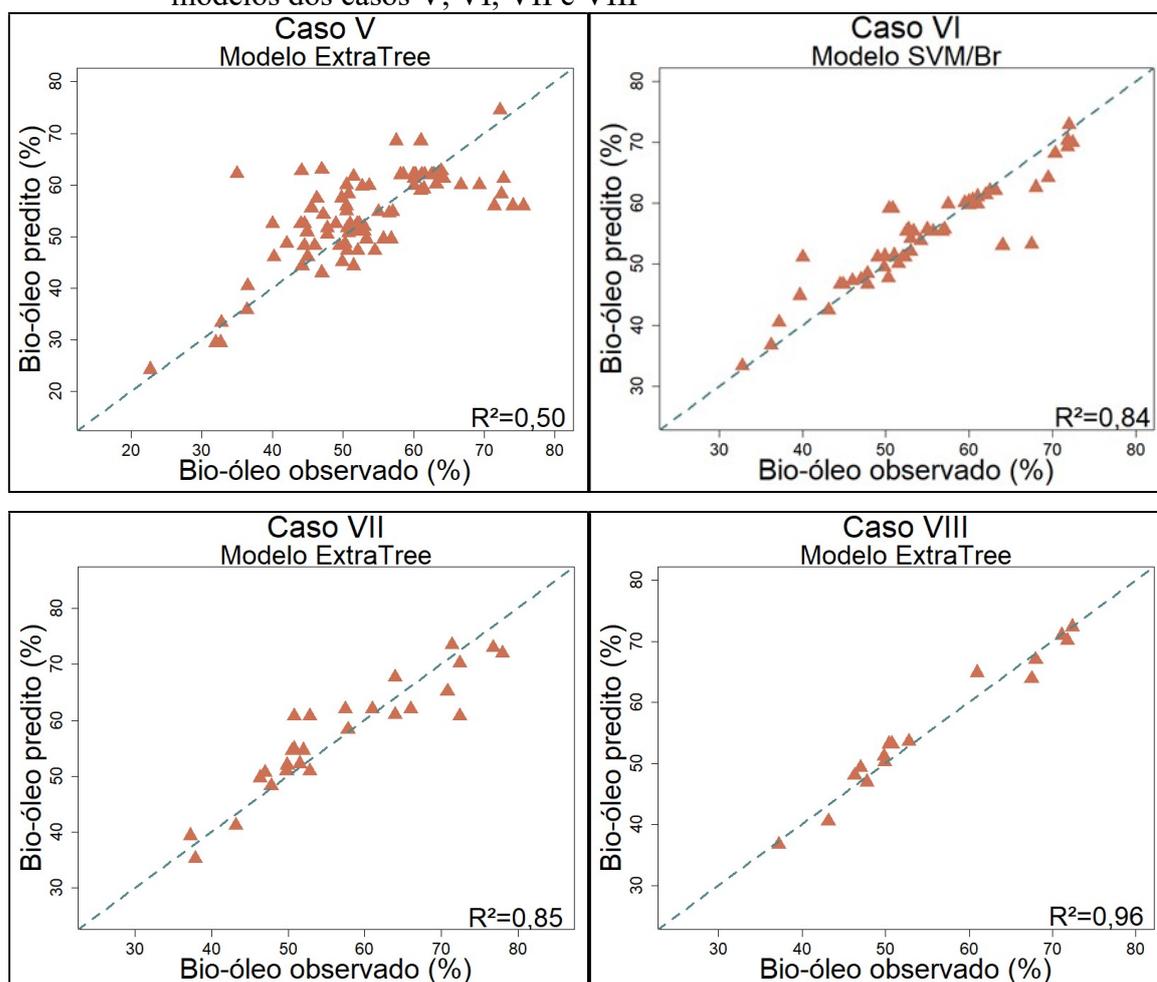
Fonte: O autor, 2023.

5.2.6 Estudo de caso VI

No caso VI, envolvendo variáveis de processo e composição de biomassa, o pré-processamento mostrou que os dados dos preditores não apresentam distribuição normal. Além disso, como no estudo de caso II, houve alta correlação entre as variáveis carbono e oxigênio, com coeficiente de correlação de 0,88. Nesse caso, a variável oxigênio também foi retirada da modelagem. O conjunto completo usado para a modelagem ficou com 225 observações com dez variáveis predictoras para a produção de bio-óleo.

O modelo SVM com função de base radial foi o modelo que apresentou o melhor resultado de predição, conforme visto nas Figuras 26, 27 e 28, apresentando um RMSE, MAE e R^2 de 3,73%, 2,28% e 0,84, respectivamente. Observa-se que o estudo de caso VI apresentou um menor erro de predição em relação ao estudo de caso V. Com isso, o desempenho da modelagem foi superior, mostrando que a inclusão das variáveis de composição da biomassa beneficia a previsão do rendimento de bio-óleo. O desempenho do modelo pode ser visualizado com o gráfico de valores previstos e observados no conjunto de teste, mostrado na Figura 29.

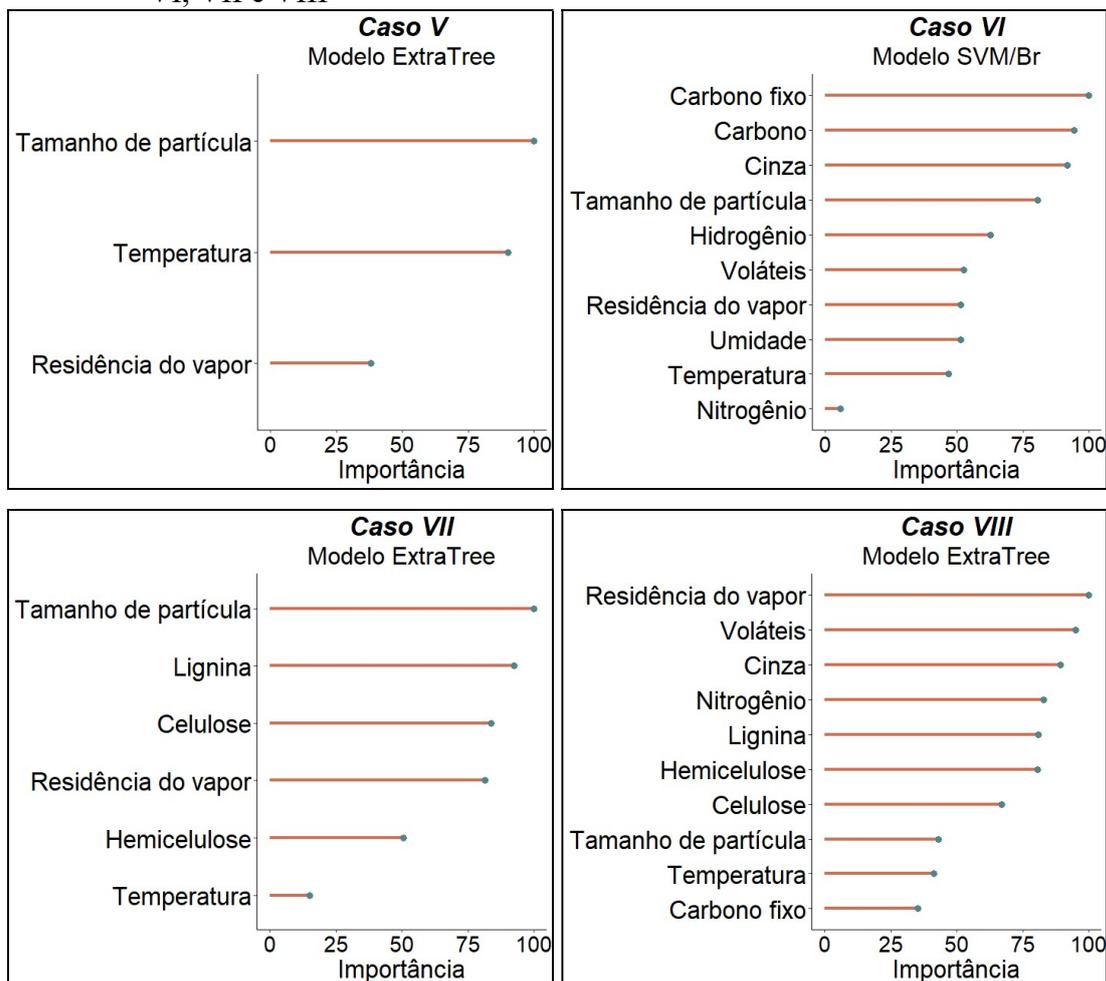
Figura 29 - Valores observados e preditos de rendimento de bio-óleo para os melhores modelos dos casos V, VI, VII e VIII



Fonte: O autor, 2023.

A Figura 30 mostra a importância das variáveis para os melhores modelos obtidos nos reatores contínuos. Para o caso VI pode-se observar que as variáveis de composição são de maior importância para o SVM/Br, em comparação às variáveis de processo.

Figura 30 - Importância das variáveis para os melhores modelos obtidos nos casos V, VI, VII e VIII



Fonte: O autor, 2023.

5.2.7 Estudo de caso VII

Neste estudo, o conjunto de modelagem incluiu, após o pré-processamento, 75 amostras para obtenção dos modelos e 25 amostras para testar o melhor desempenho do modelo. O teste de Shapiro-Wilk mostrou que nenhuma variável segue uma distribuição normal de dados. Além disso, não foi encontrada alta correlação entre as variáveis.

O modelo ExtraTree apresentou um RMSE, MAE e R^2 de 4,40%, 3,50% e 0,85, respectivamente, sendo o melhor resultado de predição para os modelos testados, conforme pode ser visto nas Figuras 26, 27 e 28. A utilização das variáveis lignocelulósicas na modelagem resultou em uma melhora no modelo em relação à utilização apenas de variáveis de processo (caso V). Porém, comparando com o caso VI, a utilização de variáveis de

composição aproximada e elementar resultou em um menor erro de predição de bio-óleo. A Figura 30 mostra que a concentração de lignina e celulose são variáveis importantes para o modelo ExtraTree, em reatores contínuos.

5.2.8 Estudo de caso VIII

Este estudo utilizou todas as variáveis previamente analisadas na modelagem. Devido ao número mais significativo de preditores, foi obtido um conjunto completo de apenas 70 amostras para aplicação dos modelos de regressão. O teste de Spearman mostrou alta correlação entre os preditores oxigênio com carbono e hidrogênio, voláteis com hidrogênio e oxigênio, além de correlação entre umidade e tempo de residência do vapor. Com isso as variáveis carbono, oxigênio, umidade e hidrogênio foram removidas da modelagem, seguindo o critério de manter as variáveis que resultam em modelos com maior capacidade preditiva de bio-óleo. A transformação Box-Cox reduziu a assimetria das variáveis predictoras.

O modelo de melhor desempenho para a predição foi o modelo ExtraTree, com RMSE e MAE de 2,15%, 1,75% e 0,96, respectivamente. O R^2 foi superior comparando-se com os outros casos de reator contínuo, indicando que o modelo explicou 96% da variabilidade dos dados. O modelo GBM também apresentou bons resultados de desempenho com um coeficiente de correlação igual ao modelo ExtraTree, porém com um erro um pouco superior, sendo RMSE de 2,33% e MAE de 1,86%. Chen *et al.* (2018) aplicaram ANN e SVM/Br para modelar o rendimento de bio-óleo em reatores contínuos usando composição elementar, umidade, cinzas, tamanho de partícula, número de fluidização e temperatura como preditores. De acordo com os resultados o modelo SVM/Br teve um melhor desempenho de predição com RMSE de 2,89, R^2 de 0,90 e MAE de 2,25. Portanto o modelo desenvolvido neste estudo teve um melhor desempenho do que o encontrado na literatura com o mesmo objetivo.

De acordo com a Figura 30, todas as variáveis utilizadas são importantes para a modelagem do caso VIII. De um modo geral, a temperatura não se apresenta como uma das variáveis mais importantes para a modelagem de rendimento de bio-óleo em reatores contínuos. Nos casos VI, VII e VIII, por exemplo, as demais variáveis de processo, ou seja, tempo de residência do vapor e tamanho de partícula, apresentam maior importância para os modelos do que a temperatura do processo.

A Tabela 13 resume as ferramentas de diagnóstico utilizadas para calcular a precisão da predição para os melhores modelos de cada caso para reatores contínuos.

Tabela 13 - Ferramentas de diagnóstico para o melhor modelo em reator contínuo

Estudo de caso	Modelo	RMSE	MAE	R ²
Caso V	ExtraTree	7,26	5,14	0,50
Caso VI	SVM/Br	3,73	2,28	0,84
Caso VII	ExtraTree	4,40	3,50	0,85
Caso VIII	ExtraTree	2,15	1,75	0,96

Fonte: O autor, 2023.

Os resultados sugerem que o modelo ExtraTree pode modelar a previsão do rendimento do bio-óleo para o reator contínuo usando as variáveis de processo, composição aproximada, elementar e lignocelulósica da biomassa, pois apresentou excelente desempenho com um maior coeficiente de determinação e baixos valores de RMSE e MAE. O melhor ajuste dos valores previstos e observados nos dados de teste do caso VIII, em relação aos demais estudos de caso, pode ser visto na Figura 29.

5.3 Modelagem de dados imputados

Uma grande quantidade de valores faltantes é observada no banco de dados obtido. A Tabela 14 mostra o número de observações faltantes para cada variável considerada nesta pesquisa. A variável temperatura possui apenas oito observações faltantes para os dados em reator batelada e seis observações faltantes para reator contínuo. Por outro lado, as variáveis de composição lignocelulósica possuem muitos dados faltantes, em ambos os tipos de reatores. A presença de dados faltantes ocorre devido à não disponibilidade dos valores para determinada variável nos estudos contidos no banco.

Para o estudo de imputação, as observações com valores faltantes para a resposta da modelagem, ou seja, rendimento de bio-óleo, foram removidas previamente do banco. Pode-se observar, na Tabela 14, que os estudos de caso dos reatores em batelada possuem um total de 1673 observações, enquanto os reatores contínuos possuem 700 observações.

Tabela 14 - Dados faltantes em cada variável

	Casos I a IV	Casos V a VIII
Número total de observações	1673	700
Quantidade de observações faltantes		
Hemicelulose	1096	513
Lignina	1044	490
Celulose	891	479
Residência do vapor	780	196
Tempo de reação	740	-
Tamanho de partícula	626	167
Taxa de aquecimento	463	-
Carbono fixo	408	212
Umidade	390	69
Voláteis	382	203
Nitrogênio	304	123
Oxigênio	273	104
Hidrogênio	210	104
Carbono	206	104
Cinza	132	67
Temperatura	8	6

Fonte: O autor, 2023.

A imputação foi realizada nos oito estudos de caso descritos anteriormente. Após a imputação, todos os modelos empregados na análise dos dados completos foram também aplicados nos dados imputados a fim de comparar se a imputação dos dados leva a uma melhora na modelagem.

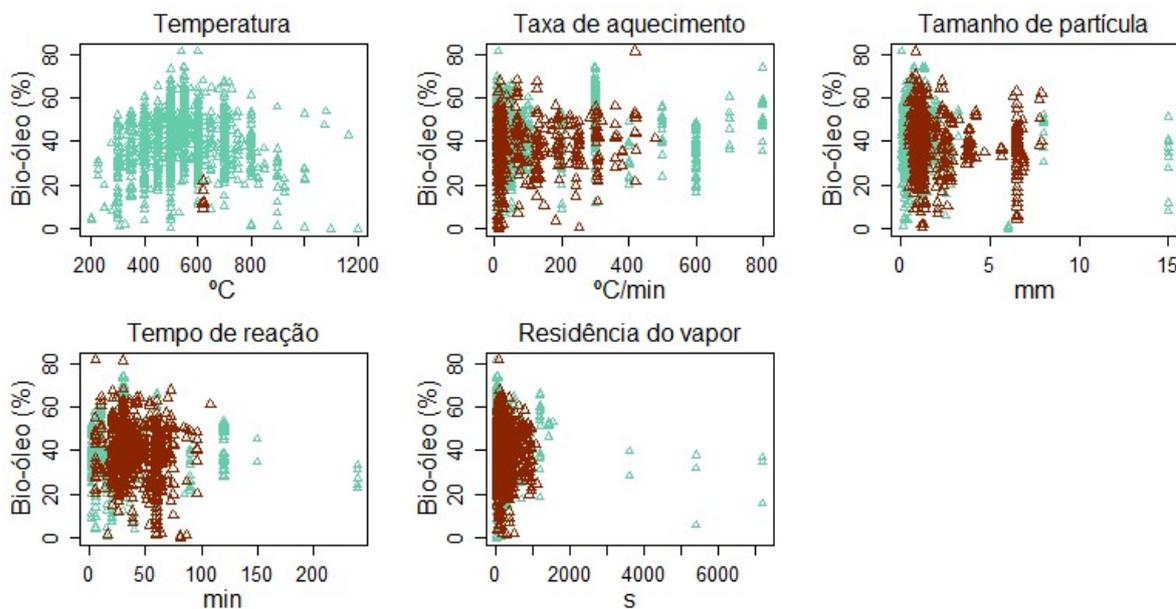
5.3.1 Estudo de caso I

A maior parte dos valores ausentes no caso I está relacionada a tempo de residência do vapor no processo, com 46,6% de dados faltantes. Esta variável nem sempre é disponibilizada nos estudos, pois muitos pesquisadores fornecem somente a vazão de gás inerte no processo. O tempo de reação também apresenta uma alta porcentagem de dados faltantes, visto que, esta variável, é pouco avaliada nos processos de pirólise.

A Figura 31 mostra os dados observados (pontos verdes) e imputados (pontos marrons) de cada variável utilizando-se a imputação por k-NN. Pode-se observar que os dados imputados não estão presentes em todas as faixas de valores que contêm dados observados. Por exemplo, para a temperatura os dados foram imputados somente no valor de 620 °C. Nos

gráficos de densidade (Figura 32) pode-se observar que os valores foram imputados na faixa de maior concentração de dados das variáveis.

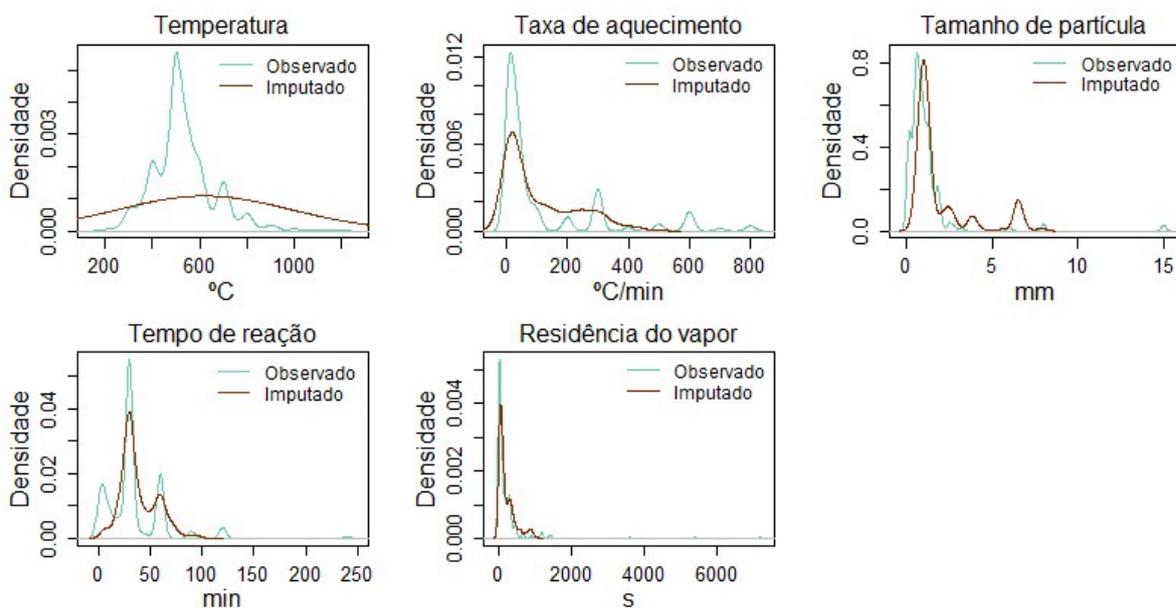
Figura 31 - Distribuição dos dados observados e imputados por k-NN- Caso I



● -Valores observados; ● -Valores imputados

Fonte: O autor, 2023.

Figura 32 - Densidade dos dados observados e imputados por k-NN - Caso I

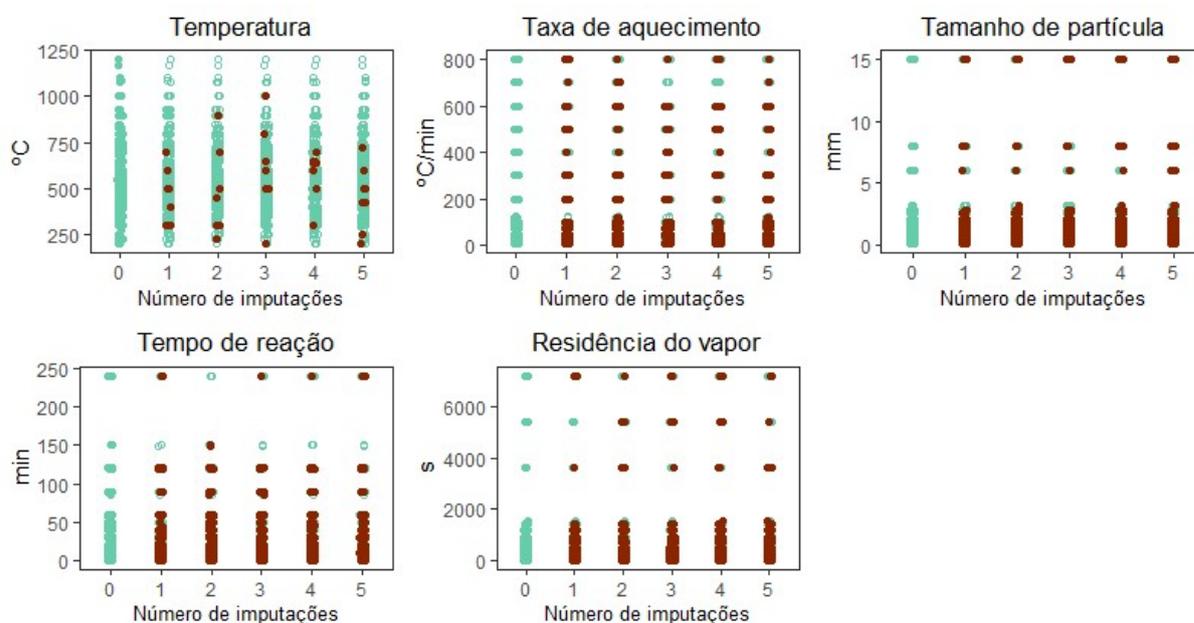


Fonte: O autor, 2023.

Outra técnica de imputação utilizada foi a imputação por RF. A imputação foi repetida cinco vezes e a média das figuras de mérito de cada modelo obtido na modelagem foi

calculada. A Figura 33 mostra os diferentes conjuntos de dados imputados comparados aos dados observados, sendo que os pontos verdes indicam os dados observados e os pontos marrons indicam os dados imputados. Em comparação à imputação k-NN pode-se observar uma maior distribuição dos dados imputados na faixa de estudo. Por exemplo, para a temperatura, no método k-NN houve a imputação dos dados em apenas um valor (Figura 31) enquanto no método RF os dados imputados para a temperatura variaram de 300 °C a 700 °C, apenas no primeiro conjunto completo obtido. Já no conjunto 3 a variação foi de 200 °C a 1000 °C. Isso ocorreu devido à imputação múltipla incluir a incerteza da imputação nos resultados (Rubin, 1987).

Figura 33 - Distribuição dos dados observados e imputados por RF- Caso I

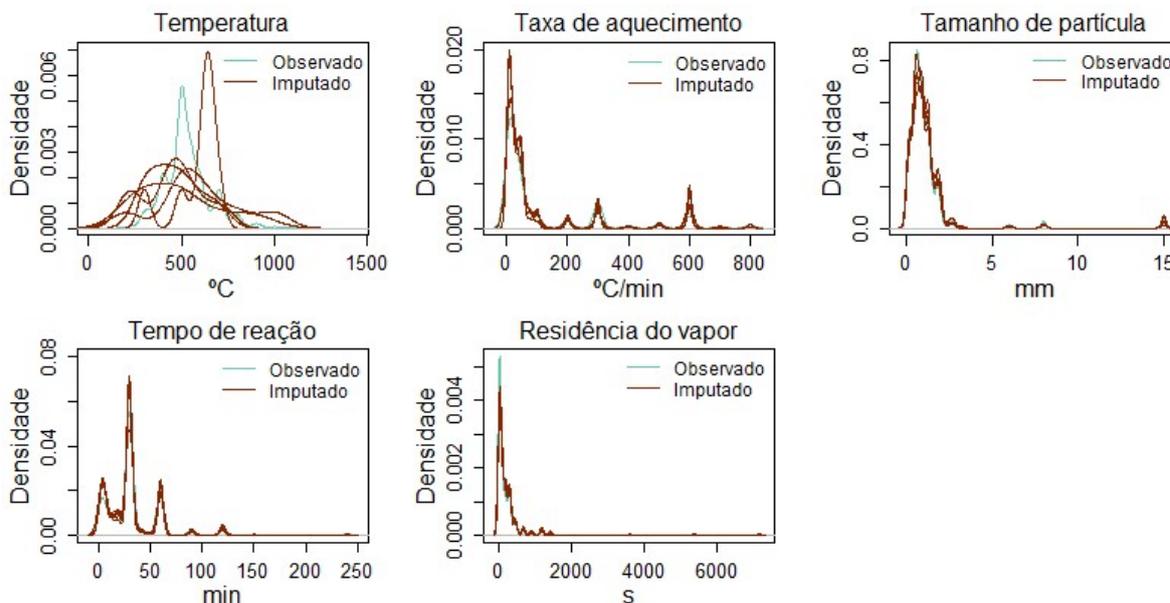


● -Valores observados; ● -Valores imputados

Fonte: O autor, 2023.

Os gráficos de densidade dos valores observados e imputados com RF (Figura 34) mostram uma maior precisão da imputação em relação à distribuição dos dados imputados por k-NN. Para a temperatura, onde tem-se apenas oito valores a serem imputados, os dados ficam mais dispersos por toda a faixa disponível para a imputação. Porém, para as demais variáveis, o modelo de imputação segue a distribuição dos dados observados.

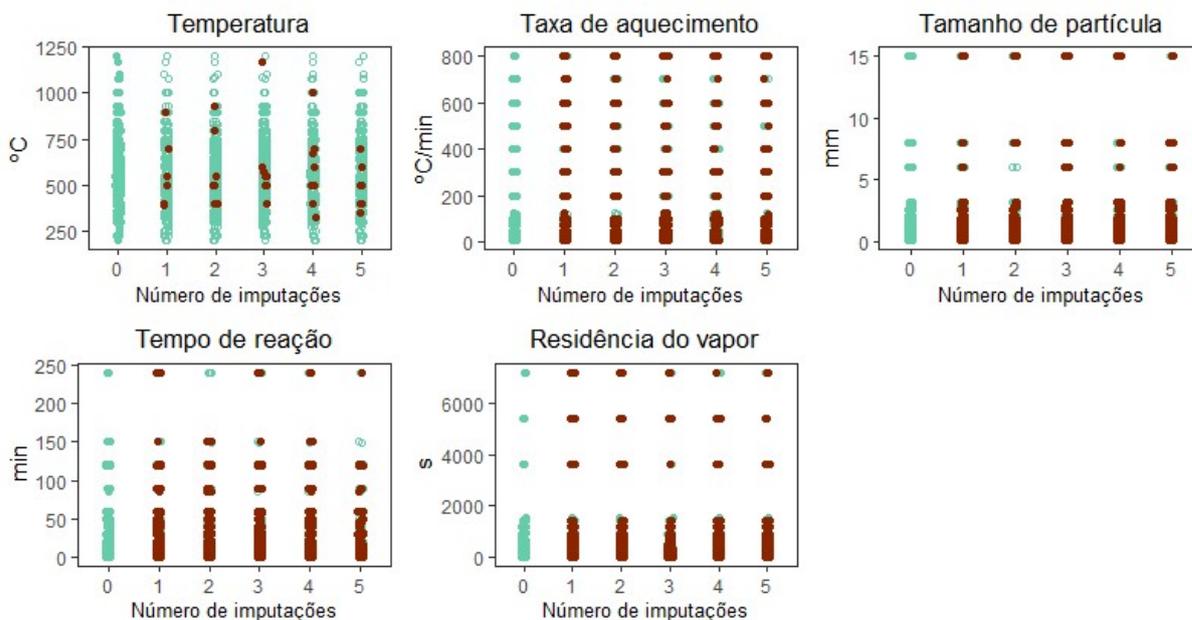
Figura 34 - Densidade dos dados observados e imputados por RF - Caso I



Fonte: O autor, 2023.

A imputação pelo método Midastouch também foi realizada com cinco repetições. A Figura 35 mostra os resultados obtidos em cada repetição. Pode-se observar diferentes valores imputados para todas as variáveis. Assim como na imputação RF, os dados imputados pelo método Midastouch são melhores distribuídos na faixa de estudo das variáveis.

Figura 35 - Distribuição dos dados observados e imputados por Midastouch - Caso I

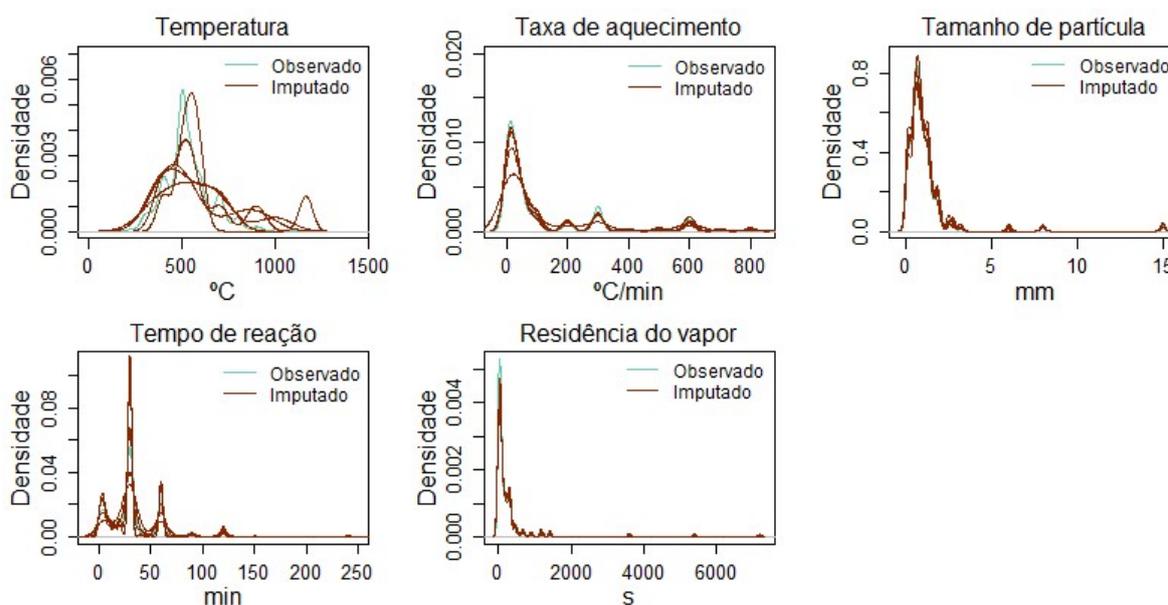


● -Valores observados; ● -Valores imputados

Fonte: O autor, 2023.

Na Figura 36, que mostra a densidade dos dados observados e imputados por Midastouch, pode-se observar que o método conseguiu imputar dados com distribuição bem próxima à distribuição dos dados observados. Porém, com uma variação maior entre os cinco conjuntos obtidos, quando comparado ao método RF.

Figura 36 - Densidade dos dados observados e imputados por Midastouch - Caso I



Fonte: O autor, 2023.

A análise exploratória dos dados foi realizada em todos os conjuntos de dados obtidos após a imputação e em todos os estudos de caso (I ao VIII). Com isso, as variáveis altamente correlacionadas foram identificadas e removidas. Em todos estudos de caso o teste de Shapiro-Wilk identificou a não normalidade dos dados e a transformação Box-Cox foi utilizada. Além disso, o sinal espacial foi aplicado devido a presença de *outliers*. Após o pré-processamento, a modelagem foi realizada com os mesmos parâmetros utilizados nos estudos de caso com dados completos.

Para o estudo de caso I a análise mostrou que não há forte correlação entre as variáveis e os dados não apresentam baixa variância em nenhum dos conjuntos analisados. As figuras de mérito dos modelos obtidos estão apresentadas na Tabela 15.

Comparando-se a modelagem com os dados obtidos nas diferentes imputações pode-se verificar que o melhor resultado foi alcançado com o modelo RF obtido a partir dos dados imputados por Floresta Aleatória, neste caso os valores de RMSE e MAE foram de 6,64% e 5,01%, respectivamente. Para a imputação por Midastouch o melhor modelo para predição de

bio-óleo foi o ExtraTree com RMSE, MAE e R^2 de 7,33%, 5,70% e 0,56, respectivamente. Este resultado foi inferior aos obtidos após a imputação por k-NN e por RF. Apesar do maior número de observações utilizadas no estudo com dados imputados os modelos obtidos apresentaram erro de predição elevado e baixo coeficiente de determinação.

Tabela 15 - Ferramentas de diagnóstico da modelagem com dados imputados -Caso I

Modelo	k-NN			RF			Midastouch		
	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2
MLR	12,42	9,43	0,01	11,16	8,76	0,03	11,34	8,91	0,01
Ridge	12,43	9,43	0,01	11,17	8,76	0,02	11,34	8,91	0,01
LASSO	12,43	9,43	0,00	11,18	8,76	0,02	11,36	8,90	0,01
Rede	12,43	9,43	0,00	11,18	8,76	0,02	11,35	8,90	0,01
k-NN	9,38	7,22	0,44	7,97	6,25	0,47	8,81	6,79	0,36
MARS	10,20	7,45	0,36	8,36	6,51	0,41	9,07	7,01	0,33
SVM/Br	8,66	6,22	0,53	7,25	5,43	0,56	9,32	6,98	0,35
SVM/Pl	11,54	8,35	0,18	8,99	6,73	0,33	9,45	7,11	0,27
ANN	9,79	7,38	0,38	7,64	5,99	0,51	9,36	7,26	0,29
CART	10,35	7,86	0,33	8,70	6,68	0,37	9,59	7,44	0,25
RF	7,29	5,20	0,65	6,64	5,01	0,63	7,47	5,75	0,54
ExtraTree	7,36	5,25	0,64	6,69	5,04	0,62	7,33	5,70	0,56
GBM	8,28	6,27	0,57	7,08	5,52	0,58	8,14	6,31	0,45
<i>Cubist</i>	8,21	6,30	0,57	6,95	5,48	0,59	8,69	6,69	0,38

Fonte: O autor, 2023.

5.3.2 Estudo de caso II

O caso II analisou dados de processo de pirólise em reator batelada juntamente com a composição elementar e aproximada da biomassa. O banco utilizado incluiu um total de treze variáveis preditoras do rendimento de bio-óleo. As variáveis com maior número de dados faltantes estão relacionadas às condições do processo de pirólise, sendo a residência do vapor com 46,62% e tempo de reação com 44,23% de dados faltantes.

Assim como no estudo de caso I, as imputações foram realizadas para os três algoritmos analisados. A distribuição dos valores imputados e observados estão de acordo com os resultados encontrados no estudo de caso I, dessa forma, os gráficos de densidade e distribuição não estão incluídos nos resultados, mas podem ser verificados no Anexo E para os estudos de caso de I a VIII.

A análise exploratória, realizada após a imputação com os três métodos empregados, mostrou uma forte correlação entre as variáveis oxigênio e carbono. A variável oxigênio foi removida por levar a melhores resultados na modelagem.

Os parâmetros de mérito dos modelos, mostrados na Tabela 16, indicam que o melhor modelo para o rendimento de bio-óleo foi o SVM com função base radial, obtido com a imputação realizada por RF, sendo o RMSE de 3,67%, MAE de 2,65% e R^2 de 0,90. Para os dados imputados por Midastouch o melhor modelo também foi o SVM/Br, porém o erro de predição foi um pouco superior ao modelo obtido por imputação RF, sendo RMSE, MAE e R^2 de 4,09%, 2,85% e 0,88, respectivamente. Os modelos obtidos por imputação k-NN resultaram em maiores erros de predição.

Tabela 16 - Ferramentas de diagnóstico da modelagem com dados imputados-Caso II

Modelo	k-NN			RF			Midastouch		
	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2
MLR	11,75	9,03	0,23	11,73	9,25	0,19	12,18	9,56	0,15
Ridge	11,74	9,06	0,22	11,69	9,21	0,19	12,18	9,55	0,15
LASSO	11,75	9,03	0,23	11,76	9,27	0,18	12,22	9,59	0,15
Rede	11,76	9,02	0,23	11,71	9,23	0,19	12,22	9,59	0,15
k-NN	4,60	3,08	0,88	5,70	4,24	0,77	6,37	4,86	0,73
MARS	10,16	7,95	0,41	8,71	6,91	0,47	9,45	7,40	0,43
SVM/Br	5,73	3,42	0,82	3,67	2,65	0,90	4,09	2,85	0,88
SVM/Pl	10,33	7,53	0,40	9,41	7,01	0,41	9,85	7,35	0,37
ANN	7,11	5,26	0,72	9,22	7,12	0,43	10,13	7,89	0,35
CART	9,74	7,52	0,47	9,36	7,32	0,42	10,62	8,37	0,30
RF	4,92	3,41	0,88	4,84	3,72	0,87	5,28	4,06	0,84
ExtraTree	4,93	3,47	0,88	5,29	4,12	0,87	5,80	4,57	0,84
GBM	4,43	3,03	0,89	5,04	3,88	0,83	5,01	3,79	0,83
<i>Cubist</i>	4,63	3,12	0,88	4,30	3,17	0,87	4,50	3,34	0,86

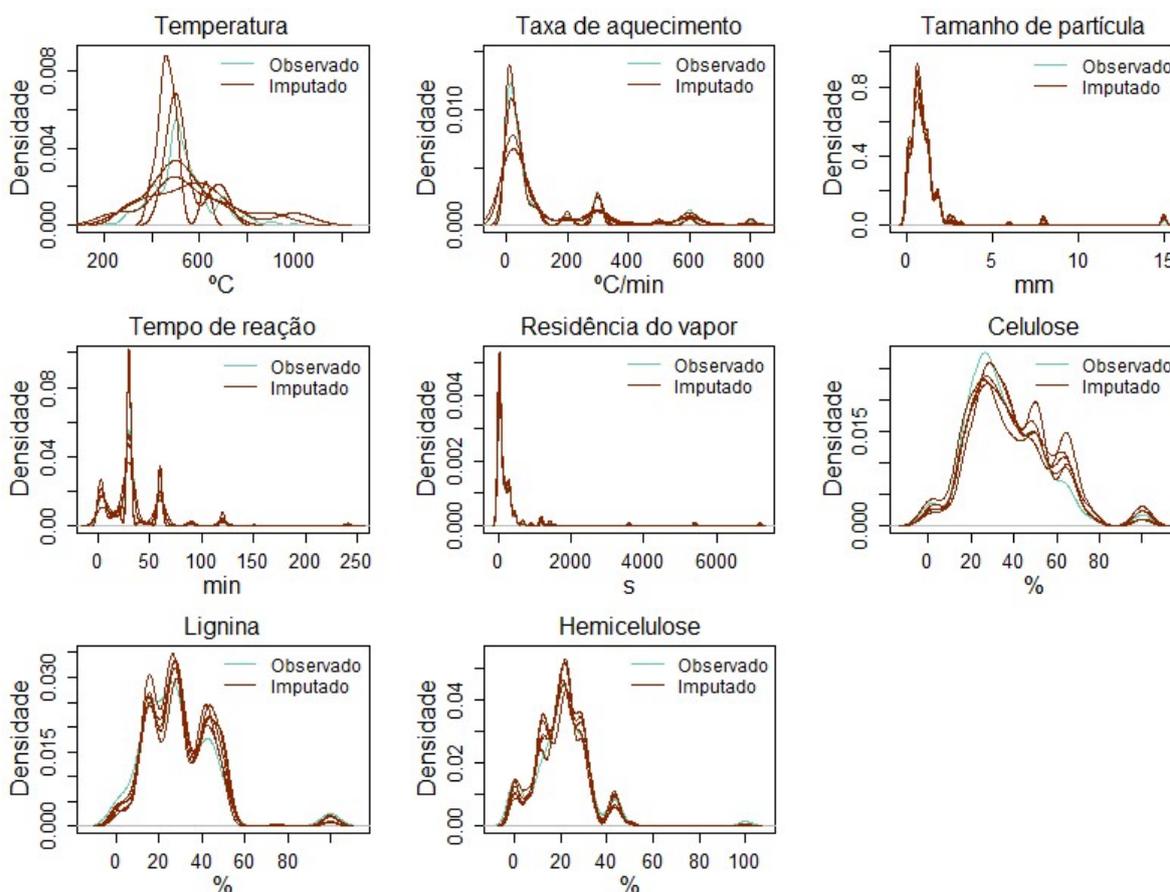
Fonte: O autor, 2023.

5.3.3 Estudo de caso III

Os dados do caso III incluíram as variáveis de composição lignocelulósicas da biomassa e variáveis de processo. Na Tabela 14 pode-se observar que as variáveis lignocelulósicas possuem uma grande quantidade de dados faltantes no banco. A hemicelulose, por exemplo, possui 1096 observações ausentes, o que corresponde a 65,51%

do total do banco. A porcentagem de dados ausentes de lignina e celulose corresponde a 62,40% e 53,26%, respectivamente. A Figura 37 mostra que, para as variáveis lignocelulósicas, ocorre uma maior variação na densidade dos dados observados e imputados por Midastouch. Isso mostra a maior incerteza sobre os dados a serem imputados. O mesmo ocorreu com os dados imputador por RF.

Figura 37 - Densidade dos dados observados e imputados por Midastouch - Caso III



Fonte: O autor, 2023.

A análise exploratória realizada após a imputação não indicou forte correlação entre as variáveis. Ao contrário dos dados completos, que apresentou forte correlação entre celulose e a hemicelulose. A Tabela 17 mostra que o melhor modelo obtido com a modelagem foi o ExtraTree incluindo os dados imputados por k-NN. O modelo conseguiu explicar 75% da variância dos dados, com RMSE e MAE de 6,46% e 4,49%.

Os modelos obtidos após a imputação RF resultaram em erros de predição próximos aos obtidos após a imputação k-NN. Porém, os dados oriundos do algoritmo Midastouch resultaram em modelos com menor capacidade preditiva.

Tabela 17- Ferramentas de diagnóstico da modelagem com dados imputados-Caso III

Modelo	k-NN			RF			Midastouch		
	RMSE	MAE	R ²	RMSE	MAE	R ²	RMSE	MAE	R ²
MLR	12,34	9,41	0,04	10,75	8,52	0,14	10,97	8,68	0,04
Ridge	12,33	9,39	0,04	10,76	8,52	0,14	10,97	8,68	0,04
LASSO	12,34	9,41	0,04	10,78	8,53	0,14	10,98	8,68	0,04
Rede	12,32	9,39	0,04	10,79	8,53	0,14	10,98	8,68	0,04
k-NN	8,02	6,14	0,61	7,81	6,08	0,52	9,13	7,18	0,33
MARS	10,11	7,62	0,37	9,12	7,07	0,33	9,50	7,42	0,25
SVM/Br	7,19	4,71	0,68	8,05	6,15	0,57	9,84	7,54	0,40
SVM/Pl	11,29	8,36	0,21	8,52	6,46	0,43	9,13	6,97	0,31
ANN	9,15	7,13	0,48	8,88	7,08	0,43	10,33	8,07	0,27
CART	10,73	8,33	0,28	9,47	7,26	0,29	9,89	7,62	0,20
RF	6,55	4,64	0,74	6,62	5,08	0,67	7,59	5,85	0,54
ExtraTree	6,46	4,49	0,75	6,83	5,34	0,69	7,88	6,13	0,61
GBM	7,48	5,56	0,66	6,99	5,40	0,61	8,11	6,35	0,46
<i>Cubist</i>	7,24	5,41	0,69	6,50	5,04	0,66	7,39	5,74	0,54

Fonte: O autor, 2023.

5.3.4 Estudo de caso IV

O caso IV englobou todas as variáveis consideradas para a previsão do rendimento do bio-óleo em reator batelada. Assim como no caso III, as variáveis lignocelulósicas, com grande quantidade de dados faltantes estão presentes no caso IV.

A imputação foi realizada com os três métodos propostos e as análises de pré-processamento foram realizadas. O teste de correlação de Spearman indicou alta correlação entre oxigênio e carbono nos dados obtidos em todos os métodos de imputação. Vale lembrar que a análise dos dados completos, ou seja, removendo os dados faltantes, mostrou uma alta correlação também entre as variáveis carbono fixo e voláteis.

A variável oxigênio foi removida da modelagem com os dados imputados e os resultados estão apresentados na Tabela 18.

O modelo SVM/Br obtido com os dados imputados por k-NN apresentou um menor erro de predição com RMSE de 4,40% e MAE de 2,66%. O modelo SVM/Br também apresentou o melhor resultado para as demais imputações, porém com um maior erro de predição de rendimento de bio-óleo.

Tabela 18 - Ferramentas de diagnóstico da modelagem com dados imputados-Caso IV

Modelo	k-NN			RF			Midastouch		
	RMSE	MAE	R ²	RMSE	MAE	R ²	RMSE	MAE	R ²
MLR	12,18	9,54	0,20	11,28	8,83	0,20	11,20	8,77	0,22
Ridge	12,19	9,52	0,19	11,24	8,79	0,20	11,21	8,78	0,22
LASSO	12,18	9,54	0,20	11,37	8,90	0,20	11,23	8,80	0,22
Rede	12,19	9,52	0,19	11,33	8,87	0,20	11,24	8,81	0,22
k-NN	4,48	2,75	0,89	6,74	5,08	0,69	6,40	4,80	0,72
MARS	9,56	7,29	0,50	8,50	6,68	0,48	8,56	6,74	0,50
SVM/Br	4,40	2,66	0,90	4,59	3,43	0,85	4,62	3,45	0,86
SVM/Pl	9,45	6,53	0,52	9,01	6,67	0,45	9,34	6,87	0,41
ANN	5,76	4,41	0,82	9,42	7,47	0,41	10,06	7,78	0,32
CART	10,10	7,64	0,44	9,70	7,60	0,35	10,01	7,93	0,34
RF	5,24	3,45	0,87	5,77	4,45	0,80	6,10	4,64	0,79
ExtraTree	4,99	3,45	0,89	6,43	4,97	0,80	6,78	5,18	0,79
GBM	4,64	3,11	0,89	5,40	4,11	0,80	5,76	4,39	0,78
<i>Cubist</i>	5,07	3,54	0,87	4,91	3,67	0,83	5,02	3,80	0,83

Fonte: O autor, 2023.

A Tabela 19 mostra um resumo dos melhores modelos obtidos com os dados completos e com a imputação para os quatro estudos de caso que envolvem reator em batelada. Pode-se verificar que o aumento dos dados por imputação não beneficiou a modelagem do rendimento de bio-óleo. O modelo SVM/Br, utilizando as variáveis do caso II e incluindo dados imputados por RF, apresentou maior capacidade preditiva entre os estudos com dados imputados. Porém, este modelo apresentou um maior erro e menor coeficiente de correção quando comparado com o modelo de dados completos obtido com as mesmas variáveis, modelo *Cubist*. Comparando-se os modelos obtidos nos estudos de caso onde o número de dados faltantes é superior a 65%, ou seja, casos III e IV, pode-se verificar uma piora acentuada na modelagem utilizando os dados imputados.

Tabela 19 - Ferramentas de diagnóstico para os melhores modelos obtidos em dados completos e imputados em reator batelada

Estudo de caso	Modelo	RMSE	MAE	R ²	Imputação
Dados completos					
Caso I	ExtraTree	7,18	5,34	0,73	-
Caso II	<i>Cubist</i>	2,85	2,06	0,96	-
Caso III	GBM	1,15	0,97	0,98	-
Caso IV	<i>Cubist</i>	0,92	0,73	0,99	-
Dados imputados					
Caso I	RF	6,64	5,01	0,63	RF
Caso II	SVM/Br	3,67	2,65	0,90	RF
Caso III	ExtraTree	6,46	4,49	0,75	k-NN
Caso IV	SVM/Br	4,40	2,66	0,90	k-NN

Fonte: O autor, 2023.

5.3.5 Estudo de caso V

O estudo de caso V incluiu apenas variáveis de processo que influenciam na pirólise realizada de forma contínua. Dessa forma, são apenas três variáveis utilizadas na modelagem do rendimento de bio-óleo. Os reatores contínuos apresentam 700 observações com resultado para a variável rendimento de bio-óleo. Entre as variáveis utilizadas, o total de dados faltantes é de 196 para o tempo de residência de vapor, 167 para tamanho de partícula e 6 para temperatura do processo.

O resultado da modelagem dos dados imputados estão presentes na Tabela 20. O modelo ExtraTree apresentou o maior coeficiente de determinação, explicando 78% da variabilidade dos dados imputados por k-NN. O erro de predição obtido para este modelo foi de 7,71% para RMSE e de 5,20% para MAE.

Tabela 20 - Ferramentas de diagnóstico da modelagem com dados imputados-Caso V

Modelo	k-NN			RF			Midastouch		
	RMSE	MAE	R ²	RMSE	MAE	R ²	RMSE	MAE	R ²
MLR	15,75	12,80	0,09	14,04	11,26	0,11	13,74	11,00	0,04
Ridge	15,76	12,82	0,09	14,04	11,27	0,11	13,74	11,00	0,04
LASSO	15,85	12,87	0,09	14,04	11,26	0,11	13,73	10,99	0,04
Rede	15,86	12,86	0,09	14,06	11,27	0,10	13,73	10,99	0,04
k-NN	12,68	10,18	0,41	11,27	9,15	0,42	11,83	9,85	0,28
MARS	13,92	11,36	0,29	11,88	9,61	0,36	12,36	10,16	0,22
SVM/Br	13,30	10,14	0,35	11,13	8,46	0,44	12,02	9,52	0,27
SVM/Pl	16,22	12,65	0,06	12,93	10,21	0,27	12,52	10,01	0,20
ANN	12,54	10,10	0,42	10,79	8,65	0,47	11,68	9,39	0,30
CART	11,85	9,72	0,51	10,77	8,62	0,47	11,75	9,29	0,31
RF	7,88	5,60	0,78	7,60	5,75	0,74	8,74	6,56	0,60
ExtraTree	7,71	5,20	0,78	7,45	5,57	0,75	8,68	6,38	0,61
GBM	9,11	7,06	0,72	8,94	7,29	0,64	10,95	9,10	0,38
<i>Cubist</i>	10,89	8,51	0,58	9,71	7,57	0,57	10,75	8,59	0,40

Fonte: O autor, 2023.

5.3.6 Estudo de caso VI

O estudo de caso VI inclui as variáveis de composição elementar e aproximada da biomassa ao caso anterior. A Tabela 14 mostra que a variável com maior número de dados faltantes para este estudo é o teor de carbono fixo com 212 valores ausentes, ou seja, 30,29% do total de observações.

Assim como na análise dos dados completos, foi observada uma alta correlação entre as variáveis oxigênio e carbono nos dados obtidos em todas as imputações realizadas. A variável oxigênio foi removida da modelagem e as figuras de mérito para todos os modelos estão presentes na Tabela 21.

Tabela 21 - Ferramentas de diagnóstico da modelagem com dados imputados-Caso VI

Modelo	k-NN			RF			Midastouch		
	RMSE	MAE	R ²	RMSE	MAE	R ²	RMSE	MAE	R ²
MLR	13,92	11,51	0,21	12,29	10,11	0,25	11,66	9,49	0,30
Ridge	13,90	11,49	0,22	12,20	10,04	0,26	11,62	9,49	0,30
LASSO	13,97	11,57	0,21	12,41	10,26	0,24	11,68	9,52	0,29
Rede	13,98	11,58	0,21	12,38	10,24	0,24	11,65	9,53	0,30
k-NN	5,18	3,50	0,89	5,77	4,26	0,83	5,63	4,12	0,83
MARS	11,29	8,50	0,49	8,79	6,94	0,60	8,76	6,82	0,58
SVM/Br	7,94	3,56	0,75	4,08	2,82	0,91	3,85	2,72	0,92
SVM/Pl	11,08	6,99	0,51	8,88	6,47	0,59	9,15	6,58	0,54
ANN	10,12	6,76	0,59	6,57	5,11	0,77	7,81	6,15	0,67
CART	6,80	4,89	0,81	9,36	7,16	0,56	8,69	6,95	0,60
RF	4,34	3,11	0,93	4,92	3,76	0,88	4,72	3,48	0,89
ExtraTree	4,50	3,22	0,92	5,05	3,96	0,88	4,93	3,77	0,89
GBM	4,03	2,86	0,94	4,73	3,58	0,88	4,78	3,60	0,88
<i>Cubist</i>	4,60	3,17	0,91	4,85	3,65	0,88	4,56	3,43	0,89

Fonte: O autor, 2023.

Pode-se observar que o melhor modelo para a previsão do rendimento de bio-óleo foi o modelo SVM/Br obtido com os dados imputados por Midastouch. Este modelo apresentou RMSE, MAE e R² de 3,85%, 2,72% e 0,92, respectivamente. O mesmo modelo obtido por RF, também apresentou uma boa capacidade preditiva para o rendimento de bio-óleo, porém com um erro um pouco superior. Para a modelagem utilizando-se os dados obtidos por imputação k-NN o modelo GBM apresentou um bom resultado.

5.3.7 Estudo de caso VII

O estudo de caso VII incluiu as variáveis de composição lignocelulósica da biomassa e variáveis de processo em reator contínuo. Como pode ser verificado na Tabela 14, há uma grande quantidade de dados faltantes para esta composição utilizada. O teor de celulose possui 513 observações faltantes do total de 700 observações deste estudo de caso. Isso corresponde a 73,3% das observações. Para a lignina e celulose a porcentagem de dados faltantes é de 70% e 68,4% do total de dados, respectivamente.

A análise exploratória dos dados após as imputações não mostrou forte correlação entre as variáveis. A comparação entre os modelos obtidos com os diferentes métodos de imputação está presente na Tabela 22.

Tabela 22 - Ferramentas de diagnóstico da modelagem com dados imputados-Caso VII

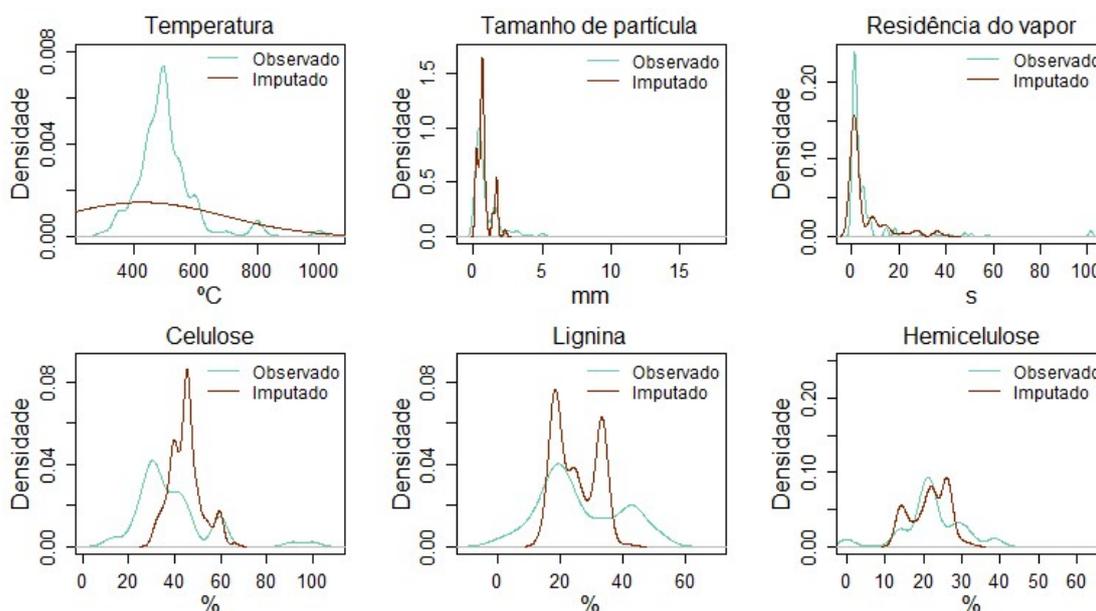
Modelo	k-NN			RF			Midastouch		
	RMSE	MAE	R ²	RMSE	MAE	R ²	RMSE	MAE	R ²
MLR	15,40	12,47	0,03	14,53	11,98	0,02	15,03	12,48	0,08
Ridge	15,29	12,37	0,04	14,58	12,02	0,02	15,05	12,50	0,07
LASSO	15,39	12,47	0,03	14,59	12,01	0,02	15,03	12,48	0,08
Rede	15,25	12,33	0,04	14,62	12,04	0,02	15,04	12,49	0,07
k-NN	5,60	3,88	0,87	10,77	8,65	0,41	12,21	10,16	0,40
MARS	13,61	10,10	0,25	12,34	9,98	0,24	12,31	9,85	0,39
SVM/Br	9,33	5,08	0,66	8,74	6,58	0,57	12,81	10,24	0,46
SVM/Pl	16,07	12,27	0,05	13,19	10,75	0,11	13,32	10,73	0,26
ANN	8,77	5,72	0,70	10,56	8,12	0,40	11,54	9,29	0,43
CART	9,84	7,35	0,61	11,58	9,23	0,33	13,51	10,75	0,29
RF	5,25	3,90	0,91	8,71	6,66	0,62	9,08	7,37	0,67
ExtraTree	5,19	3,78	0,90	8,79	6,84	0,64	9,46	7,92	0,70
GBM	5,13	3,74	0,90	8,96	7,09	0,58	10,70	8,69	0,55
<i>Cubist</i>	6,27	4,35	0,84	8,94	7,00	0,60	9,73	8,01	0,61

Fonte: O autor, 2023.

Pode-se observar, na Tabela acima, que a imputação por k-NN resultou em dados que levaram à obtenção de modelos de regressão com maior capacidade preditiva. O modelo GBM obtido neste método de imputação, alcançou um coeficiente de correlação de 0,90 e RMSE e MAE de 5,13% e 3,74%, respectivamente. É importante observar que os modelos oriundos de dados imputados com RF e Midastouch, métodos mais robustos, tiveram resultados muito inferiores aos modelos obtidos pela imputação k-NN. O modelo GBM, por

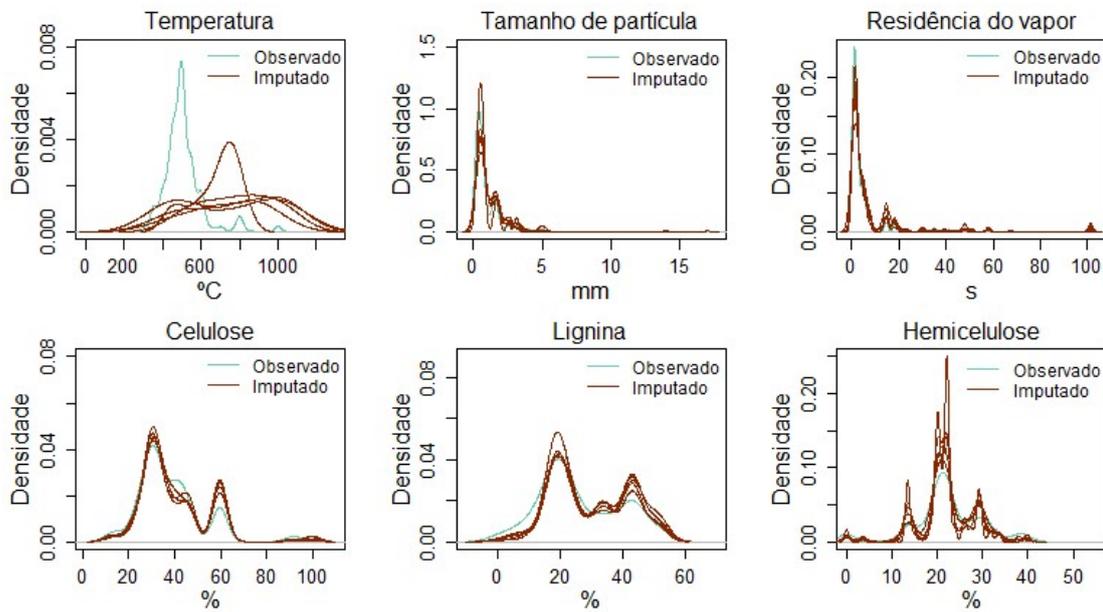
exemplo, explica somente 58% da variabilidade dos dados com imputação por RF e 55% quando os dados foram obtidos por Midastouch. Além disso, os gráficos de densidade mostram que os valores obtidos por imputação múltipla apresentam uma melhor distribuição sobre os valores já presentes no banco em comparação aos valores imputados por k-NN (Figuras 38, 39 e 40). Portanto, os resultados nos mostram que a imputação k-NN resultou em modelos com menor erro de predição, porém, a imputação não respeita a estatística distributiva dos dados, o que indica que a imputação não foi bem sucedida, principalmente para as variáveis com maior quantidade de valores faltantes, ou seja, celulose, lignina e hemicelulose.

Figura 38 - Densidade dos dados observados e imputados por k-NN - Caso VII



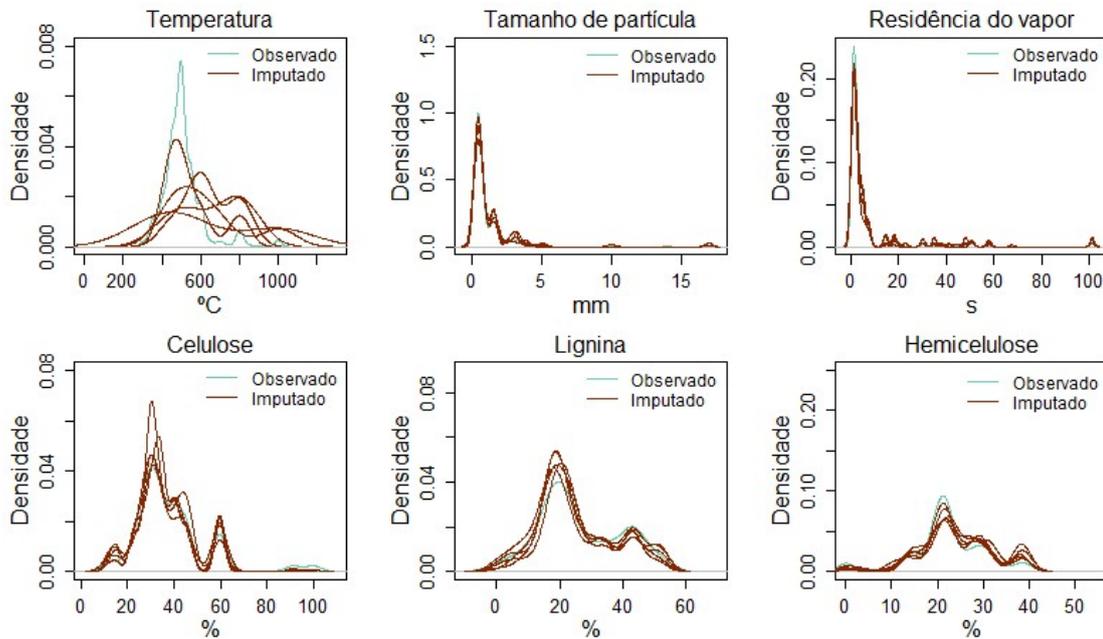
Fonte: O autor, 2023.

Figura 39 - Densidade dos dados observados e imputados por RF - Caso VII



Fonte: O autor, 2023.

Figura 40 - Densidade dos dados observados e imputados por Midastouch - Caso VII



Fonte: O autor, 2023.

5.3.8 Estudo de caso VIII

O caso VIII analisou todas as variáveis consideradas importantes para a predição do rendimento de bio-óleo em reatores contínuos. Assim como no caso VII, as variáveis de composição lignocelulósica, que contém alta porcentagem de dados faltantes, também estão presentes neste estudo.

O pré-processamento dos dados, realizado após a imputação, mostrou uma alta correlação entre as variáveis oxigênio e carbono nos dados obtidos por todos os algoritmos de imputação. Assim, para a modelagem a variável oxigênio foi removida dos preditores. O resultado, apresentado na Tabela 23, mostra que a imputação k-NN obteve dados que geraram modelos com melhores performances. O modelo GBM apresentou um melhor desempenho, com RMSE e MAE de 3,61% e 2,54%, respectivamente. Apesar do bom resultado obtido para o k-NN o mesmo problema do caso VII pode ter ocorrido devido à presença de variáveis com uma alta porcentagem de dados faltantes. Os gráficos de densidade, para os três métodos de imputação, podem ser comparados no Anexo E.

Comparando-se as imputações múltiplas, o melhor resultado de modelagem foi obtido com os dados oriundos de Midastouch, onde o modelo SVM/Br apresentou um RMSE de 5,70% e MAE de 4,11%.

Tabela 23 - Ferramentas de diagnóstico da modelagem com dados imputados-Caso VIII

Modelo	k-NN			RF			Midastouch		
	RMSE	MAE	R ²	RMSE	MAE	R ²	RMSE	MAE	R ²
MLR	13,32	10,74	0,32	12,16	10,18	0,34	12,03	9,77	0,39
Ridge	13,30	10,70	0,32	12,16	10,21	0,34	12,04	9,80	0,39
LASSO	13,36	10,76	0,32	12,48	10,47	0,31	12,17	9,91	0,38
Rede	13,35	10,76	0,32	12,38	10,41	0,32	12,17	9,92	0,38
k-NN	6,28	3,98	0,85	8,62	6,60	0,65	8,43	6,49	0,68
MARS	11,46	8,41	0,49	9,60	7,80	0,56	9,83	7,69	0,58
SVM/Br	7,62	3,34	0,78	5,88	4,20	0,84	5,70	4,11	0,85
SVM/Pl	10,20	6,13	0,62	9,27	7,24	0,59	9,65	7,36	0,57
ANN	6,03	3,53	0,86	9,38	7,33	0,60	9,89	7,62	0,58
CART	10,35	7,97	0,59	9,73	7,59	0,56	10,60	8,02	0,52
RF	4,12	2,86	0,94	6,57	4,96	0,81	6,44	4,98	0,83
ExtraTree	4,30	2,90	0,93	7,05	5,48	0,82	7,05	5,62	0,84
GBM	3,61	2,54	0,95	6,59	4,92	0,79	6,36	4,88	0,82
<i>Cubist</i>	5,82	3,55	0,87	6,58	4,86	0,79	6,24	4,74	0,82

Fonte: O autor, 2023.

A comparação da modelagem dos dados completos e por imputação é mostrada na Tabela 24. Para os casos V e VI os modelos obtidos por dados completos e por imputação alcançaram resultados muito próximos no erro de predição, sendo que os modelos obtidos por dados completos apresentaram um desempenho um pouco superior. O RMSE e MAE foram de 7,26% e 5,14%, respectivamente, para o modelo de dados completos no caso V e de 7,71% e 5,20%, respectivamente, para os dados imputados por k-NN.

Comparando-se os casos VII e VIII, onde a porcentagem de dados a ser imputada é alta, acima de 70%, os modelos obtidos dos dados completos apresentaram melhor desempenho de predição de rendimento de bio-óleo.

Tabela 24 - Ferramentas de diagnóstico para os melhores modelos obtidos em dados completos e imputados em reator contínuo

Estudo de caso	Modelo	RMSE	MAE	R ²	Imputação
Dados completos					
Caso V	ExtraTree	7,26	5,14	0,50	-
Caso VI	SVM/Br	3,73	2,28	0,84	-
Caso VII	ExtraTree	4,40	3,50	0,85	-
Caso VIII	ExtraTree	2,15	1,75	0,96	-
Dados imputados					
Caso V	ExtraTree	7,71	5,20	0,78	k-NN
Caso VI	SVM/Br	3,85	2,72	0,92	Midastouch
Caso VII	GBM	5,13	3,74	0,90	k-NN
Caso VIII	GBM	3,61	2,54	0,95	k-NN

Fonte: O autor, 2023.

5.4 Otimização

Após constatada a alta capacidade preditiva dos modelos, foi realizada uma etapa de otimização do rendimento de bio-óleo. A otimização foi realizada utilizando-se os modelos que apresentaram os menores erros de predição nos diferentes reatores, ou seja, modelo *Cubist* obtido no estudo de caso IV e modelo ExtraTree obtido no caso VIII, ambos resultantes da modelagem dos dados completos.

Para a análise foi utilizada como biomassa a casca de arroz, por apresentar um maior número de observações no banco de dados. Esta biomassa foi analisada em 26 estudos presentes no banco, com um total de 160 observações. Neste estudo, foi utilizada a composição média das cascas de arroz presentes no banco de dados e as variáveis do processo

foram ajustadas para obter maior rendimento de bio-óleo. A composição média da casca de arroz é mostrada na Tabela 25.

Tabela 25 - Composição média da casca de arroz utilizada na otimização

Variável	Composição (%)
Celulose	32,80
Lignina	19,42
Hemicelulose	24,14
Umidade	7,78
Carbono fixo	14,79
Cinza	16,38
Voláteis	68,63
Carbono	47,36
Hidrogênio	6,17
Nitrogênio	1,00

Fonte: O autor, 2023.

As variáveis de processo foram otimizadas somente dentro da faixa de valores disponível no estudo de modelagem de cada estudo de caso. No estudo em batelada, a faixa analisada foi entre 300-700 °C para temperatura, 0,6-8 mm para tamanho de partícula, 6-1400 s para tempo de residência de vapor e 5-300 °C/min para taxa de aquecimento (Tabela 26). A variável tempo de reação não foi utilizada na otimização por não apresentar importância para o modelo *Cubist*, conforme mostrado na Figura 22.

Tabela 26 - Faixa de otimização e melhores parâmetros para otimização do caso IV

Variável	Faixa de otimização	Valor para máximo rendimento (Rendimento=52,9%)
Temperatura (°C)	300-700	605
Tamanho de partícula (mm)	0,6-8	7,40
Residência do vapor (s)	6-1400	1400
Taxa de aquecimento (°C/min)	5-300	300

Fonte: O autor, 2023.

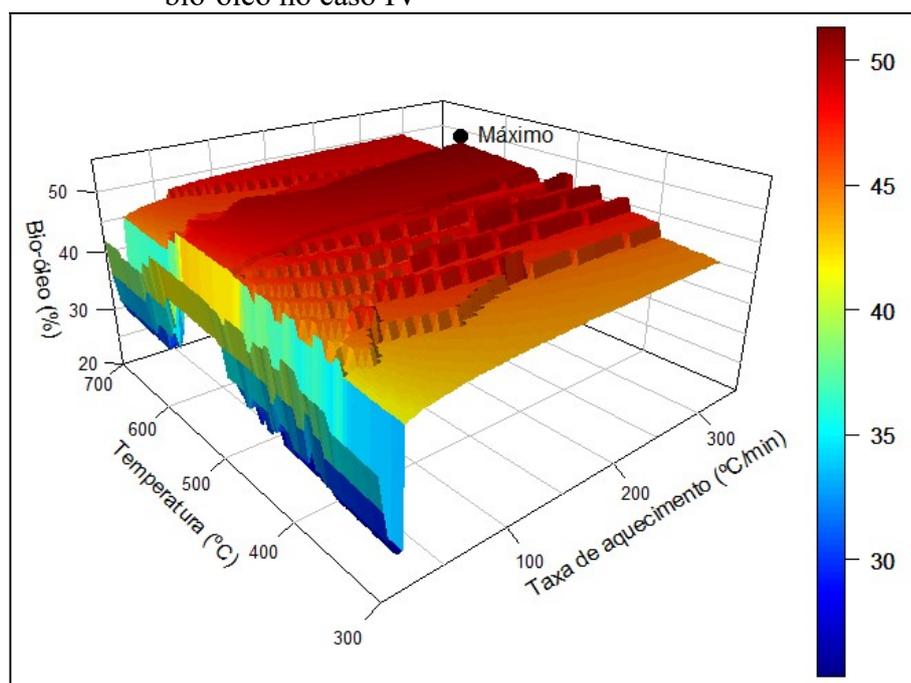
O rendimento máximo de bio-óleo obtido pelo método PSO foi de 52,9% utilizando temperatura de 605 °C, tamanho de partícula de 7,4 mm, tempo de residência de vapor de 1400 s e taxa de aquecimento de 300 °C/min.

Um elevado teor de cinzas na composição da casca de arroz pode ser observado na Tabela 25. A redução desse componente pode aumentar o rendimento de bio-óleo no processo, pois a cinza tem um efeito negativo na resposta, conforme mostrado em Guedes, Luna e Torres (2018). Um maior teor de cinzas na biomassa reduz o rendimento de bio-óleo

devido aos componentes inorgânicos presentes nas cinzas. O sódio e o potássio têm um impacto significativo na redução de líquidos (Venderbosch; Prins, 2010).

A Figura 41 mostra o ponto ótimo encontrado no PSO na superfície obtida com as duas variáveis mais importantes para o modelo, ou seja, temperatura do processo e taxa de aquecimento.

Figura 41 - Superfície de resposta para otimização do rendimento de bio-óleo no caso IV



Fonte: O autor, 2023.

A Tabela 27 apresenta as faixas de estudo das variáveis de processo utilizadas no estudo de otimização do reator contínuo. A temperatura foi analisada na faixa de 350 a 600 °C, tamanho de partícula na faixa de 0,3 a 10 mm e tempo de residência do vapor na faixa de 0,5 a 45 s.

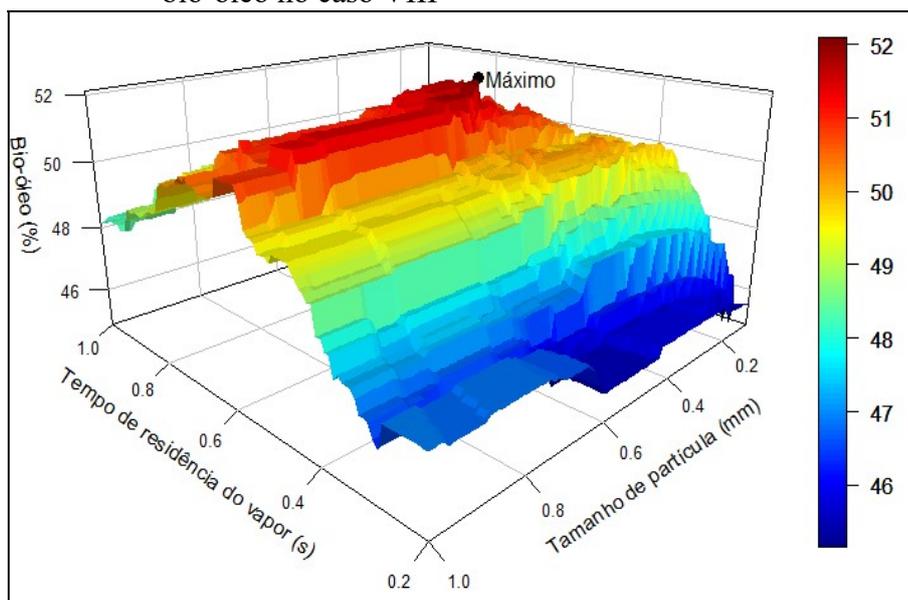
Tabela 27 - Faixa de otimização e melhores parâmetros para otimização do caso VIII

Variável	Faixa de otimização	Valor para máximo rendimento (Rendimento=52,1%)
Temperatura (°C)	350-600	513
Tamanho de partícula (mm)	0,3-10	0,41
Residência do vapor (s)	0,5-45	0,60

Fonte: O autor, 2023.

O rendimento máximo do bio-óleo foi de 52,1% na temperatura de 513 °C, com tamanho de partícula de 0,41 mm e tempo de residência de vapor de 0,60 s. A superfície obtida com as variáveis tempo de residência do vapor e temperatura, Figura 42, mostra o ponto ótimo. O modelo indica maior rendimento de líquido em temperatura intermediária, menor tamanho de partícula e baixo tempo de residência de vapor.

Figura 42 - Superfície de resposta para otimização do rendimento de bio-óleo no caso VIII



Fonte: O autor, 2023.

A análise de otimização dos parâmetros de processo dos casos IV e VIII mostrou que o reator batelada pode alcançar um rendimento de bio-óleo tão alto quanto o reator contínuo, se operado nas condições corretas de processo. O reator de leito fixo deve trabalhar a uma temperatura mais elevada, tamanho de partícula maior, maior tempo de residência de vapor e altas taxas de aquecimento para alcançar um resultado próximo ao reator de leito fluidizado.

CONCLUSÃO

Um banco de dados contendo dados de 220 trabalhos experimentais sobre pirólise, com um total de 2617 observações foi desenvolvido. O banco de dados pode constituir uma ferramenta útil para a realização de estudos envolvendo a pirólise de biomassa não somente para a análise de bio-óleo, mas também podendo englobar estudos de bio-gás e carvão. Podendo ser utilizado para pesquisas de rendimento e composição dos produtos de pirólise.

A partir da análise exploratória do banco observou-se que a pirólise se trata de um processo complexo e influenciado por parâmetros como temperatura, taxa de aquecimento, tempo de residência do vapor, tamanho de partícula, tempo de reação e composição da biomassa. Além disso, evidenciou-se um importante efeito das interações dessas variáveis no processo.

A temperatura é o parâmetro mais analisado na pirólise da biomassa. Sendo que temperaturas intermediárias de pirólise geralmente maximizam o rendimento do bio-óleo. Uma análise geral da correlação dos dados mostrou que as variáveis de processo e as variações composicionais na biomassa alteram o rendimento do óleo de pirólise. Uma concentração superior de voláteis e inferior de cinzas e carbono fixo aumentam o rendimento do bio-óleo. Nenhuma correlação foi observada entre a umidade da biomassa e a produção de bio-óleo.

O principal reator utilizado nos processos de pirólise é o reator de leito fixo, que opera em batelada, porém, um rendimento médio de bio-óleo superior, de 49,86%, é produzido nos reatores contínuos, onde a produção de bio-óleo é favorecida por altas taxas de transferência de calor e curto tempo de residência.

Diferentes preditores foram usados para modelar o processo de pirólise em reatores operando em batelada e de modo contínuo. Para o reator em batelada, o modelo *Cubist* apresentou alto desempenho de predição com RMSE de 0,92%, MAE de 0,73% e coeficiente de determinação de 0,99. Este resultado foi obtido utilizando as variáveis de processo, composição elementar, aproximada e lignocelulósica da biomassa. Além disso, os estudos de caso II e III, que usaram diferentes combinações de variáveis de composição de biomassa, mostraram alta precisão de previsão. Com isso, é possível realizar a modelagem de acordo com diferentes variáveis preditoras disponíveis para o estudo.

Os modelos de regressão multivariada aplicados em um reator contínuo apresentaram desempenho de predição inferior, quando comparados aos modelos aplicados aos reatores em batelada. No entanto, a modelagem usando todas as variáveis preditoras testadas neste estudo resultou no modelo ExtraTree com alta precisão de predição, com um RMSE de 2,15%, MAE de 1,75% e um coeficiente de determinação de 0,96.

A modelagem mostrou que, além das variáveis de processo, as variáveis de composição são de grande importância para o rendimento do bio-óleo. Para modelos com melhor precisão de previsão, a temperatura, que é a variável mais discutida na literatura, não é a variável mais importante.

Diferentes métodos de imputação foram utilizados devido a presença de dados faltantes no banco. Os resultados da modelagem não mostraram uma melhora na capacidade preditiva dos modelos com a utilização do maior número de amostras, obtido com a imputação de dados.

A análise de otimização do rendimento de bio-óleo da casca de arroz em diferentes reatores de pirólise mostrou a importância da modelagem para a escolha dos melhores parâmetros de processo. Além disso, a otimização mostrou que o reator em batelada, amplamente utilizado para pirólise, pode atingir rendimentos de bio-óleo tão altos quanto os reatores contínuos, se operado nas condições corretas.

REFERÊNCIAS

- ABAYOMI, K.; GELMAN, A.; LEVY, M. Diagnostics for multivariate imputations. **Journal of the Royal Statistical Society: Series C (Applied Statistics)**, 57(3), 273-291, 2008.
- ABDULLAH, N.; GERHAUSER, H. Bio-oil derived from empty fruit bunches. **Fuel**, v. 87, n. 12, p. 2606-2613, 2008.
- ABNISA, F. *et al.* Utilization of oil palm tree residues to produce bio-oil and bio-char via pyrolysis. **Energy Conversion and Management**, v. 76, p. 1073-1082, 2013.
- ABNISA, F. *et al.* Utilization possibilities of palm shell as a source of biomass energy in Malaysia by producing bio-oil in pyrolysis process. **Biomass and Bioenergy**, v. 35, n. 5, p. 1863-1872, 2011.
- ABNISA, F.; WAN DAUD, W. M. A.; SAHU, J. N. Optimization and characterization studies on bio-oil production from palm shell by pyrolysis using response surface methodology. **Biomass and Bioenergy**, v. 35, n. 8, p. 3604-3616, 2011.
- AÇIKALIN, K.; KARACA, F.; BOLAT, E. Pyrolysis of pistachio shell: Effects of pyrolysis conditions and analysis of products. **Fuel**, v. 95, p. 169-177, 2012.
- ACIKGOZ, C.; ONAY, O.; KOCKAR, O. M. Fast pyrolysis of linseed: Product yields and compositions. **Journal of Analytical and Applied Pyrolysis**, v. 71, n. 2, p. 417-429, 2004.
- AGATONOVIC-KUSTRIN, S.; BERESFORD, R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. **Journal of Pharmaceutical and Biomedical Analysis**, v. 22, n. 5, p. 717-727, 2000.
- AKHTAR, J.; AMIN, N. S. A review on operating parameters for optimum liquid oil yield in biomass pyrolysis. **Renewable and Sustainable Energy Reviews**, v. 16, n. 7, p. 5101-5109, 2012.
- ALEXOPOULOS, E. C. Introduction to multivariate regression analysis. **Hippokratia**, v. 14, p. 23-28, 2010.
- ALLISON, P. D. **Missing Data**. Thousand Oaks: Sage Publications, 2001.
- ALVA, J. A. V.; ESTRADA, E. G. A generalization of Shapiro-Wilk's test for multivariate normality. **Communications in Statistics-Theory and Methods**, v. 38, n. 11, p. 1870-1883, 2009.
- ALVAREZ, J. *et al.* Bio-oil production from rice husk fast pyrolysis in a conical spouted bed reactor. **Fuel**, v. 128, p. 162-169, 2014.
- ALVAREZ, J. *et al.* Fast co-pyrolysis of sewage sludge and lignocellulosic biomass in a conical spouted bed reactor. **Fuel**, v. 159, p. 810-818, 2015.
- ANDRADE, B. M. *et al.* Comparison of the performance of multiclass classifiers in chemical

data: Addressing the problem of overfitting with the permutation test. **Chemometrics and Intelligent Laboratory Systems**, 2020.

ANDRÉS, J. *et al.* Bankruptcy forecasting: A hybrid approach using fuzzy c-means clustering and multivariate adaptive regression splines (MARS). **Expert Systems with Applications**, v. 38, n. 3, p. 1866-1875, 2011.

AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA (ANEEL). **Atlas de Energia Elétrica. Cap. 3. Biomassa**. 3 ed. Brasília. 2008.

ANGIN, D. Effect of pyrolysis temperature and heating rate on biochar obtained from pyrolysis of safflower seed press cake—Bioresource **Technology**, v. 128, p. 593-597, 2013.

APAYDIN-VAROL, E.; PÜTÜN, E.; PÜTÜN, A. E. Slow pyrolysis of pistachio shell. **Fuel**, v. 86, n. 12-13, p. 1892-1899, 2007.

ARAZO, R. O. *et al.* Bio-oil production from dry sewage sludge by fast pyrolysis in an electrically-heated fluidized bed reactor. **Sustainable Environment Research**, v. 27, n. 1, p. 7-14, 2017.

ASADULLAH, M. *et al.* Production and detailed characterization of bio-oil from fast pyrolysis of palm kernel shell. **Biomass and Bioenergy**, v. 59, p. 316-324, 2013.

ASADULLAH, M. *et al.* Production of bio-oil from fixed bed pyrolysis of bagasse. **Fuel**, v. 86, p. 2514-2520, 2007.

ASCHER, S. *et al.* Interpretable machine learning to model biomass and waste gasification. **Bioresource Technology**, v. 364, 2022.

ATEŞ, F.; IŞIKDAĞ, M. A. Evaluation of the role of the pyrolysis temperature in straw biomass samples and characterization of the oils by GC/MS. **Energy and Fuels**, v. 22, p. 1936-1943, 2008.

ATEŞ, F.; PÜTÜN, E.; PÜTÜN, A. E. Fast pyrolysis of sesame stalk: Yields and structural analysis of bio-oil. **Journal of Analytical and Applied Pyrolysis**, v. 71, n. 2, p. 779-790, 2004.

BAKAR, M. S. A.; TITIOYE, J. O. Catalytic pyrolysis of rice husk for bio-oil production. **Journal of Analytical and Applied Pyrolysis**, v. 103, p. 362-368, 2013.

BARNARD, J.; MENG, X. L. Applications of multiple imputation in medical studies: from AIDS to NHANES. **Statistical Methods in Medical Research**, p. 17-36, 1999.

BARTOLI, M. *et al.* Production of bio-oils and bio-char from *Arundo donax* through microwave assisted pyrolysis in a multimode batch reactor. **Journal of Analytical and Applied Pyrolysis**, v. 122, p. 479-489, 2016.

BEIS, S. H.; ONAY, O.; KOÇKAR, O. M. Fixed-bed pyrolysis of safflower seed: Influence of pyrolysis parameters on product yields and compositions. **Renewable Energy**, v. 26, n. 1, p. 21-32, 2002.

BELDEN, E. R. *et al.* Machine Learning Predictions of Oil Yields Obtained by Plastic Pyrolysis and Application to Thermodynamic Analysis. **ACS Engineering Au**, v. 3, p. 91-101, 2023.

BENDTSEN, C. **Pso**: Particle Swarm Optimization. R package version 1.0.4., 2022.

BENNETT, N. M.; HELLE, S. S.; DUFF, S. J. B. Extraction and hydrolysis of levoglucosan from pyrolysis oil. **Bioresource Technology**, v. 100, n. 23, p. 6059-6063, 2009.

BIRADAR, C. H.; SUBRAMANIAN, K. A.; DASTIDAR, M. G. Production and fuel quality upgradation of pyrolytic bio-oil from *Jatropha Curcas* de-oiled seed cake. **Fuel**, v. 119, p. 81-89, 2014.

BISWAS, B. *et al.* Pyrolysis of agricultural biomass residues: Comparative study of corn cob, wheat straw, rice straw and rice husk. **Bioresource Technology**, v. 237, p. 57-63, 2017.

BOEHMKE, B.; GREENWELL, B. **Hands-On Machine Learning with R**. 1st. ed. New York: Chapman and Hall/CRC, 2019.

BORDOLOI, N. *et al.* Pyrolysis of *Mesua ferrea* and *Pongamia glabra* seed cover: Characterization of bio-oil and its sub-fractions. **Bioresource Technology**, v. 178, p. 83-89, 2015.

BOUCHER, T. F. *et al.* A study of machine learning regression methods for major elemental analysis of rocks using laser-induced breakdown spectroscopy. **Spectrochimica Acta - Part B Atomic Spectroscopy**, v. 107, p. 1-10, 2015.

BOX, G. E. P.; COX, D. R. An Analysis of Transformations. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 26, n. 2, p. 211-243, 1964.

BRANCAGLIONI, V. A. **Comparação de métodos de imputação para dados de pecuária de precisão**. 2023. 56 f. Tese (Doutorado em Ciências) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2023.

BREIMAN, L. Random Forests. **Machine Learning**, v. 45, p. 5-32, 2001.

BRIDGWATER, T. Review Biomass for energy. **Journal of the Science of Food and Agriculture**, v. 86, p. 1755-1768, 2006.

BRIDGWATER, A. V. Principles and practice of biomass fast pyrolysis processes for liquids. **Journal of Analytical and Applied Pyrolysis**, v. 51, n. 2, p. 3-22, 1999.

BRIDGWATER, A. V. Renewable fuels and chemicals by thermal processing of biomass. **Chemical Engineering Journal**, v. 91, n. 3, p. 87-102, 2003.

BRIDGWATER, A. V. Review of fast pyrolysis of biomass and product upgrading. **Biomass and Bioenergy**, v. 38, p. 68-94, 2012.

CAO, H.; XIN, Y.; YUAN, Q. Prediction of biochar yield from cattle manure pyrolysis via least squares support vector machine intelligent approach—**Bioresource Technology**, v. 202,

p. 158-164, 2016.

CARDOSO, C. A. L.; MACHADO, M. E.; CARAMÃO, E. B. Characterization of bio-oils obtained from pyrolysis of bocaiuva residues. **Renewable Energy**, v. 91, p. 21-31, 2016.

CASONI, A. I. *et al.* Pyrolysis of sunflower seed hulls for obtaining bio-oils. **Bioresource Technology**, v. 177, p. 406-409, 2015.

CHANG, G. *et al.* The lignin pyrolysis composition and pyrolysis products of palm kernel shell, wheat straw, and pine sawdust. **Energy Conversion and Management**, v. 124, p. 587-597, 2016.

CHEN, X. *et al.* Prediction of product distribution and bio-oil heating value of biomass fast pyrolysis. **Chemical Engineering and Processing - Process Intensification**, v. 130, p. 36-42, 2018.

CHOUDHURY, N. D. *et al.* Pyrolysis of jute dust: Effect of reaction parameters and analysis of products. **Journal of Material Cycles and Waste Management**, v. 16, p. 449-459, 2014.

CHOUHAN, A.; SARMA, A. Critical Analysis of Process Parameters for Bio-oil Production via Pyrolysis of Biomass: A Review. **Recent Patents on Engineering**, v. 7, n. 2, p. 98-114, 2013.

COLLARD, F. X.; BLIN, J. A review on pyrolysis of biomass constituents: Mechanisms and composition of the products obtained from the conversion of cellulose, hemicelluloses and lignin. **Renewable and Sustainable Energy Reviews**, v. 38, p. 594-608, 2014.

CRANMER, S. J.; GILL, J. We have to be discrete about this: A non-parametric imputation technique for missing categorical data—British **Journal of Political Science**, v. 43, n. 2, p. 425-449, 2012.

CUTLER, A.; CUTLER, D. R.; STEVENS, J. R. Random Forests. **Machine Learning**, v. 45, n. 1, p. 157-176, 2011.

DEBDOUBI, A. *et al.* The effect of heating rate on yields and compositions of oil products from esparto pyrolysis. **International Journal of Energy Research**, v. 30, n. 15, p. 1243-1250, 2006.

DEMIRBAS, A. Biomass Resource Facilities and Biomass Conversion Processing for Fuels and Chemicals. **Energy Convers. Manage.**, v. 42, p. 1357-1378, 2001.

DEMIRBAS, A. Effect of temperature on pyrolysis products from four nut shells. **Journal of Analytical and Applied Pyrolysis**, v. 76, p. 285-289, 2006.

DHANAVATH, K. N. *et al.* Optimization of process parameters for slow pyrolysis of neem press seed cake for liquid and char production. **Journal of Environmental Chemical Engineering**, v. 7, 2019.

DJURIS, J.; IBRIC, S.; DJURIC, Z. Chemometric methods application in pharmaceutical products and processes analysis and control. In: DJURIS, J. **Computer-Aided Applications**

in **Pharmaceutical Technology**. Cambridge: UK, p. 57-90, 2013.

DRAPER, N. R.; COX, D. R. On Distributions and Their Transformation to Normality. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 31, n. 3, p. 472-476, 1969.

DUMAN, G. *et al.* The slow and fast pyrolysis of cherry seed. **Bioresource Technology**, v. 102, p. 1869-1878, 2011.

EBBELS, T. M. D. Non-linear Methods for the Analysis of Metabolic Profiles. **The Handbook of Metabonomics and Metabolomics**, p. 201-226, 2007.

ELLENS, C. J.; BROWN, R. C. Optimization of a free-fall reactor for the production of fast pyrolysis bio-oil. **Bioresource Technology**, v. 103, p. 374-380, 2012.

ELLIOTT, A. C.; HYNAN, L. S. A SAS® macro implementation of a multiple comparison post hoc test for a Kruskal-Wallis analysis. **Computer Methods and Programs in Biomedicine**, v. 102, p. 75-80, 2011.

ENCINAR, J. M. *et al.* Pyrolysis/gasification of agricultural residues by carbon dioxide in the presence of different additives: influence of variables. **Fuel Processing Technology**, v. 55, n. 3, p. 219-233, 1998.

ENCINAR, J. M.; GONZÁLEZ, J. F.; GONZÁLEZ, J. Fixed-bed pyrolysis of *Cynara cardunculus* L. Product yields and compositions. **Fuel processing technology**, v. 68, n. 3, p. 209-222, 2000.

ERGUN, B. *et al.* Data filtering with support vector machines in geometric camera calibration. **Optics Express**, v. 18, n. 3, p. 1927, 2010.

FAISAL, S.; TUTZ, G. Multiple imputation using nearest neighbor methods. **Information Sciences**, v. 570, p. 500-516, 2021.

FOOD AND AGRICULTURE ORGANIZATION (FAO). **Crop Prospects and Food Situation- Quarterly Global Report**. Rome, 2020a.

FOOD AND AGRICULTURE ORGANIZATION (FAO). **Biannual Report on Global Food Markets: Food Outlook**. Rome, 2020b.

FARHANGFAR, A.; KURGAN, L. A.; PEDRYCZ, W. A novel framework for imputation of missing values in databases. **IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans**, v. 37, n. 5, p. 692-709, 2007.

FERRE, J. Regression Diagnostics. In: BROWN, S. D.; TAULER, R.; WALCZAK, B. **Comprehensive Chemometric: Chemical and Biochemical Data Analysis**, Oxford: Elsevier, 2009, p. 33-89.

FILZMOSER, P.; GSCHWANDTNER, M. **Mvoutlier: Multivariate Outlier Detection Based on Robust Methods**. R package version 2.0.9. 2018.

FRIEDMAN, J. H. Multivariate adaptive regression splines. **Annals of Statistics**, v. 19, n. 1, p. 1-141, 1991.

FRIEDMAN, J. H. Stochastic gradient boosting. **Computational Statistics and Data Analysis**, v. 38, n. 4, p. 367-378, 2002.

FU, P. *et al.* Effect of temperature on gas composition and char structural features of pyrolyzed agricultural residues. **Bioresource Technology**, v. 102, n. 17, p. 8211-8219, 2011.

FU, P. *et al.* FTIR study of pyrolysis products evolving from typical agricultural residues. **Journal of Analytical and Applied Pyrolysis**, v. 88, p. 117-123, 2010.

GAFFERT, P.; MEINFELDER, F.; BOSCH, V. Towards an MI-proper Predictive Mean Matching. **Discussion Paper**, p. 1-14, 2015.

GARCIA-PEREZ, M. *et al.* Fast pyrolysis of oil mallee woody biomass: Effect of temperature on the yield and quality of pyrolysis products. **Industrial and Engineering Chemistry Research**, v. 47, p. 1846-1854, 2008.

GARG, R.; ANAND, N.; KUMAR, D. Pyrolysis of babool seeds (*Acacia nilotica*) in a fixed bed reactor and bio-oil characterization. **Renewable Energy**, v. 96, p. 167-171, 2016.

GENT, S. *et al.* Introduction to Feedstocks. **Theoretical and Applied Aspects of Biomass Torrefaction**, p. 17-39, 2017.

GENUER, R. *et al.* Random Forests for Big Data. **Big Data Research**, v. 9, p. 28-46, 2017.

GERÇEL, H. F. The production and evaluation of bio-oils from the pyrolysis of sunflower-oil cake. **Biomass and Bioenergy**, v. 23, p. 307-314, 2002.

GEURTS, P.; ERNST, D.; WEHENKEL, L. Extremely randomized trees. **Machine Learning**, v. 63, n. 1, p. 3-42, 2006.

GONZALEZ-SANCHEZ, A.; FRAUSTO-SOLIS, J.; OJEDA-BUSTAMANTE, W. Predictive ability of machine learning methods for massive crop yield prediction. **Spanish Journal of Agricultural Research**, v. 12, n. 2, p. 313-328, 2014.

GOYAL, H. B.; SEAL, D.; SAXENA, R. C. Bio-fuels from thermochemical conversion of renewable resources: A review. **Renewable and Sustainable Energy Reviews**, v. 12, n. 2, p. 504-517, 2008.

GUEDES, R. E.; LUNA, A. S.; TORRES, A. R. Operating parameters for bio-oil production in biomass pyrolysis: A review. **Journal of Analytical and Applied Pyrolysis**, v. 129, p. 134-149, 2018.

HALLINAN, J. S. Computational Intelligence in the Design of Synthetic Microbial Genetic Systems. **Methods in Microbiology**, v. 40, p. 1-37, 2013.

HASTIE, T.; TIBSHIRANIT, R.; FRIEDMAN, J. **The elements of statistical learning**. 2.ed. New York: Springer, 2008.

HEIDARI, A. *et al.* Effect of process conditions on product yield and composition of fast pyrolysis of *Eucalyptus grandis* in fluidized bed reactor. **Journal of Industrial and Engineering Chemistry**, v. 20, p. 2594-2602, 2014.

HEO, H. S. *et al.* Fast pyrolysis of rice husk under different reaction conditions. **Journal of Industrial and Engineering Chemistry**, v. 16, n. 1, p. 27-31, 2010.

HOERL, A. E.; KENNARD, R. W. Ridge Regression: Biased Estimation for Nonorthogonal Problems. **Technometrics**, v. 12, n. 1, p. 69-82, 1970.

HOPKE, P. K. Environmental Chemometrics. In: BROWN, S. D.; TAULER, R.; WALCZAK, B. **Comprehensive Chemometric: Chemical and Biochemical Data Analysis**, 2. ed. Oxford: Elsevier, 2020. p. 69-85.

HOPKE, P. K. The evolution of chemometrics. **Analytica Chimica Acta**, v. 500, n. 1-2, p. 365-377, 2003.

HUANG, X. *et al.* Influences of pyrolysis conditions in the production and chemical composition of the bio-oils from fast pyrolysis of sewage sludge. **Journal of Analytical and Applied Pyrolysis**, v. 110, p. 353-362, 2014.

IKURA, M.; STANCIULESCU, M.; HOGAN, E. Emulsification of pyrolysis derived bio-oil in diesel fuel. **Biomass and Bioenergy**, v. 24, n. 3, p. 221-232, 2003.

INGRAM, L. *et al.* Pyrolysis of wood and bark in an auger reactor: Physical properties and chemical analysis of the produced bio-oils. **Energy and Fuels**, v. 22, p. 614-625, 2008.

ISA, K. M. *et al.* Thermogravimetric analysis and optimizing bio-oil yield from fixed-bed pyrolysis of rice husk using response surface methodology (RSM). **Industrial Crops and Products**, v. 33, n. 2, p. 481-487, 2011.

ISAHAK, W. N. R. W. *et al.* A review on bio-oil production from biomass by using pyrolysis method. **Renewable and Sustainable Energy Reviews**, v. 16, n. 8, p. 5910-5923, 2012.

ISLAM, M. N. *et al.* Pyrolytic oil from fixed bed pyrolysis of municipal solid waste and its characterization. **Renewable Energy**, v. 30, n. 3, p. 413-420, 2005.

ISLAM, M. R.; PARVEEN, M.; HANIU, H. Properties of sugarcane waste-derived bio-oils obtained by fixed-bed fire-tube heating pyrolysis. **Bioresource Technology**, v. 101, n. 11, p. 4162-4168, 2010.

JAHIRUL, M. I. *et al.* Biofuel production through biomass pyrolysis. A technological review. **Energies**, v. 5, n. 12, p. 4952-5001, 2012.

JAMES, G. *et al.* **An introduction to statistical learning**. 1. ed. New York: Springer, 2013.

JAREK, S. **Mvnormtest: Normality test for multivariate variables**. R package version 0.1-9., 2012.

JOARDDER, M. U. H. *et al.* Solar Assisted Fast Pyrolysis: A Novel Approach of Renewable

Energy Production. **Journal of Engineering (United Kingdom)**, v. 2014, 2014.

JOURABCHI, S. A.; GAN, S.; NG, H. K. Pyrolysis of Jatropha curcas pressed cake for bio-oil production in a fixed-bed system-**Energy Conversion and Management**, v. 78, p. 518-526, 2014.

JUNG, S. H.; KANG, B. S.; KIM, J. S. Production of bio-oil from rice straw and bamboo sawdust under various reaction conditions in a fast pyrolysis plant equipped with a fluidized bed and a char separation system. **Journal of Analytical and Applied Pyrolysis**, v. 82, p. 240-247, 2008.

KANAUIA, P. K. *et al.* Review of analytical strategies in the production and upgrading of bio-oils derived from lignocellulosic biomass. **Journal of Analytical and Applied Pyrolysis**, v. 105, p. 55 - 74, 2014.

KANG, B. S. *et al.* Fast pyrolysis of radiata pine in a bench scale plant with a fluidized bed: Influence of a char separation system and reaction conditions on the production of bio-oil. **Journal of Analytical and Applied Pyrolysis**, v. 76, n. 1-2, p. 32 - 37, 2006.

KASSAMBARA, A. **rstatix: Pipe-Friendly Framework for Basic Statistical Tests**. R package version 0.7.2., 2023.

KENNARD, R. W.; STONE, L. A. Computer-Aided Design of Experiments. **Technometrics**, v. 11, n. 1, p. 137-148, 1969.

KENNEDY, J.; EBERHART, R. Particle swarm optimization. **Proceedings of the International Conference on Neural Networks**.: IEEE, 1995. p. 1942 - 1948.

KILIÇ, M.; PÜTÜN, E.; PÜTÜN, A. E. Optimization of Euphorbia rigida fast pyrolysis conditions by using response surface methodology. **Journal of Analytical and Applied Pyrolysis**, v. 110, p. 163 - 171, 2014.

KIM, K. H. *et al.* Investigation of physicochemical properties of biooils produced from yellow poplar wood (*Liriodendron tulipifera*) at various temperatures and residence times. **Journal of Analytical and Applied Pyrolysis**, v. 92, p. 2 - 9, 2011.

KIM, S. J.; JUNG, S. H.; KIM, J. S. Fast pyrolysis of palm kernel shells: Influence of operation parameters on the bio-oil yield and the yield of phenol and phenolic compounds. **Bioresource Technology**, v. 101, p. 9294 - 9300, 2010.

KIM, S. W. *et al.* Bio-oil from the pyrolysis of palm and Jatropha wastes in a fluidized bed. **Fuel Processing Technology**, v. 108, p. 18 - 124, 2013.

KISI, O.; PARMAR, K. S. Application of least square support vector machine and multivariate adaptive regression spline models in long-term prediction of river water pollution. **Journal of Hydrology**, v. 534, p. 104 - 112, 2016.

KOZIEL, S.; YANG, X. **Computational Optimization, Methods and Algorithms**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011. (Studies in Computational Intelligence).v. 356

KUHN, M. **Caret: Classification and Regression Training. R package version 6.0-86., 2021.**

KUHN, M.; JOHNSON, K. **Applied predictive modeling.** New York: Springer, 2013.

LANDAU, S.; BARTHEL, S. Recursive partitioning. **International Encyclopedia of Education**, v. 1, p. 383 - 389, 2010.

LANDRY, M. *et al.* Probabilistic gradient boosting machines for GEFCom2014 wind forecasting. **International Journal of Forecasting**, v. 32, n. 3, p. 1061 - 1066, 2016.

LAZZARI, E. *et al.* Production and chromatographic characterization of bio-oil from the pyrolysis of mango seed waste. **Industrial Crops and Products**, v. 83, p. 529 - 536, 2016.

LEE, Y. *et al.* Comparison of biochar properties from biomass residues produced by slow pyrolysis at 500°C. **Bioresource Technology**, v. 148, p. 196 - 201, 2013.

LEHMANN, J.; GAUNT, J.; RONDON, M. Bio-char sequestration in terrestrial ecosystems - A review. **Mitigation and Adaptation Strategies for Global Change**, v. 11, n. 2, p. 403 - 427, 2006.

LI, S. *et al.* Fast pyrolysis of biomass in free-fall reactor for hydrogen-rich gas. **Fuel Processing Technology**, v. 85, n. 8, p. 1201 - 1211, 2004.

LI VIGNI, M.; DURANTE, C.; COCCHI, M. Exploratory Data Analysis. **Data Handling in Science and Technology**, v.28, p. 55-126, 2013.

LIM, C. H. *et al.* Novel input-output prediction approach for biomass pyrolysis. **Journal of Cleaner Production**, v. 136, p. 51-61, 2016.

LITTLE, R. J. A.; RUBIN, D. B. **Statistical Analysis with Missing Data.** 3. ed. New York: Wiley, 2019.

LIU, A.; YANG, M. T. A new hybrid nelder-mead particle swarm optimization for coordination optimization of directional overcurrent relays. **Mathematical Problems in Engineering**, v. 2012, p. 1-18, 2012.

LIU, Z. *et al.* Cellulose-lignin and xylan-lignin interactions on the formation of lignin-derived phenols in pyrolysis oil. **BioResources**, v. 12, n. 3, p. 4958-4971, 2017.

LOONEY, S. W.; HAGAN, J. L. Statistical Methods for Assessing Biomarkers and Analyzing Biomarker Data. **Handbook of Statistics**, v. 27, p. 27-65, 2007.

LY, H. V. *et al.* Fast pyrolysis of macroalga *Saccharina japonica* in a bubbling fluidized-bed reactor for bio-oil production. **Energy**, v. 93, p. 1436-1446, 2015.

LY, H. V. *et al.* Fast pyrolysis of *Saccharina japonica* alga in a fixed-bed reactor for bio-oil production. **Energy Conversion and Management**, v. 122, p. 526-534, 2016.

MADDI, B.; VIAMAJALA, S.; VARANASI, S. Comparative study of pyrolysis of algal biomass from natural lake blooms with lignocellulosic biomass. **Bioresource Technology**, v. 102, p. 11018-11026, 2011.

MAITY, S. K. Opportunities, recent trends and challenges of integrated biorefinery: Part II. **Renewable and Sustainable Energy Reviews**, v. 43, p. 1446-1466, 2015.

MAKIBAR, J. *et al.* Performance of a conical spouted bed pilot plant for bio-oil production by poplar flash pyrolysis. **Fuel Processing Technology**, v. 137, p. 283-289, 2015.

MARTÍNEZ, C. M.; CAO, D. *i*Horizon-Enabled Energy Management For Electrified Vehicles. 1. ed. Amsterdam: Elsevier, 2019.

MASSART, D. L. *et al.* **Handbook of Chemometrics and Qualimetrics: part A**. 1. ed. Amsterdam: Elsevier, 1997.

MAYRINK, V. T. M. **Avaliação do algoritmo Gradient Boosting em aplicações de previsão de carga elétrica a curto prazo**. 2016. 91 f. Dissertação (Mestrado em Modelagem Computacional) - Universidade Federal de Juiz de Fora, Juiz de Fora, 2016.

MCKENDRY, P. Energy production from biomass (part 1): overview of biomass. **Bioresource Technology**, v. 83, n. 83, p. 37-64, 2002.

MEHDIZADEH, S.; BEHMANESH, J.; KHALILI, K. Using MARS, SVM, GEP and empirical equations for estimation of monthly mean reference evapotranspiration. **Computers and Electronics in Agriculture**, v. 139, p. 103-114, 2017.

MELKUMOVA, L. E.; SHATSKIKH, S. Y. Comparing Ridge and LASSO estimators for data analysis. **Procedia Engineering**, v. 201, p. 746-755, 2017.

MERDUN, H.; SEZGIN, I. V. Modelling of pyrolysis product yields by artificial neural networks. **International Journal of Renewable Energy Research**, v. 8, n. 2, p. 1178-1188, 2018.

METCALF, L.; CASEY, W. Introduction to data analysis. **Cybersecurity and Applied Mathematics**, p. 43-65, 2016.

MIN, J. H.; LEE, Y. C. Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. **Expert Systems with Applications**, v. 28, n. 4, p. 603-614, 2005.

MISHRA, P. *et al.* Descriptive statistics and normality tests for statistical data. **Annals of Cardiac Anaesthesia**, v. 22, n. 1, p. 67-72, 2019.

MISHRA, R. K.; MOHANTY, K. Thermocatalytic conversion of non-edible Neem seeds towards clean fuel and chemicals. **Journal of Analytical and Applied Pyrolysis**, v.134, p. 83-92, 2018.

MOHAN, D.; PITTMAN, C. U.; STEELE, P. H. Pyrolysis of wood/biomass for bio-oil: A critical review. **Energy and Fuels**, v. 20, n. 3, p. 848-889, 2006.

NÄF, B. J. *et al.* Imputation Scores. **Annals of Applied Statistics**, v. 17, n. 3, p. 2452-2472, 2023.

NAYAK, S. *et al.* A recursive partitioning approach to investigating correlates of self-rated health: The CARDIA Study. **SSM - Population Health**, v. 4, p. 178-188, 2018.

NEUPANE, A. *et al.* Development of data-driven models for wind speed forecasting in Australia. **Predictive Modelling for Energy Management and Power Systems Engineering**, p. 143-190, 2021.

OASMAA, A. *et al.* Fast pyrolysis bio-oils from wood and agricultural residues - energy and Fuels, v. 24, p. 1380-1388, 2010.

OASMAA, A.; PEACOCKE, C. **Properties and fuel use of biomass-derived fast pyrolysis liquids. A guide.** Finland: VTT, 2010.

OMAR, R. *et al.* Characterization of empty fruit bunch for microwave-assisted pyrolysis. **Fuel**, v. 90, n. 4, p. 1536-1544, 2011.

ONAY, O. Influence of pyrolysis temperature and heating rate on the production of bio-oil and char from safflower seed by pyrolysis, using a well-swept fixed-bed reactor. **Fuel Processing Technology**, v. 88, p. 523-531, 2007.

ONAY, Ö.; BEIS, S. H.; KOÇKAR, Ö. M. Fast pyrolysis of rape seed in a well-swept fixed-bed reactor. **Journal of Analytical and Applied Pyrolysis**, v. 59, p. 995-1007, 2001.

ONAY, O.; KOÇKAR, O. M. Fixed-bed pyrolysis of rapeseed (*Brassica napus* L.). **Biomass and Bioenergy**, v. 26, n. 3, p. 289-299, 2004.

OSBORNE, J. W. Improving your data transformations: Applying the Box-Cox transformation. **Practical Assessment, Research and Evaluation**, v. 15, n. 12, 2010.

OSTERTAGOVÁ, E.; OSTERTAG, O.; KOVÁČ, J. Methodology and Application of the Kruskal-Wallis Test. **Applied Mechanics and Materials**, V. 611, 2014.

PAENPONG, C.; PATTIYA, A. Effect of pyrolysis and moving-bed granular filter temperatures on the yield and properties of bio-oil from fast pyrolysis of biomass. **Journal of Analytical and Applied Pyrolysis**, v. 119, p. 40-51, 2016.

PANDALA, S. R. **Lazy Predict - Lazy Predict Documentation.** Release 0.2.12, 2022. Disponível em: <<https://lazypredict.readthedocs.io/en/latest/>>. Acesso em: 07 de maio de 2023.

PARK, H. J. *et al.* Conversion of the Pyrolytic Vapor of Radiata Pine over Zeolites. **Journal of Industrial and Engineering Chemistry**, v. 13, n. 2, p. 182-189, 2007.

PARK, J. *et al.* Slow pyrolysis of rice straw: Analysis of products properties, carbon and energy yields. **Bioresource Technology**, v. 155, p. 63-70, 2014.

PÄRPÄRIŤĂ, E. *et al.* Pyrolysis behaviors of various biomasses. **Polymer Degradation and**

Stability, v. 100, n. 1, p. 1-9, 2014.

PATTIYA, A. Bio-oil production via fast pyrolysis of biomass residues from cassava plants in a fluidized-bed reactor. **Bioresource Technology**, v. 102, p. 1959-1967, 2011.

PATTIYA, A.; SUKKASI, S.; GOODWIN, V. Fast pyrolysis of sugarcane and cassava residues in a free-fall reactor. **Energy**, v. 44, p. 1067-1077, 2012.

PATTIYA, A.; SUTTIBAK, S. Production of bio-oil via fast pyrolysis of agricultural residues from cassava plantations in a fluidized-bed reactor with a hot vapor filtration unit. **Journal of Analytical and Applied Pyrolysis**, v. 95, p. 227-235, 2012.

PEREIRA, J. M.; BASTO, M.; SILVA, A. F. The Logistic Lasso and Ridge Regression in Predicting Corporate Failure. **Procedia Economics and Finance**, v. 39, n. November 2015, p. 634-641, 2016.

PIOTROWSKI, A. P.; NAPIORKOWSKI, J. J.; PIOTROWSKA, A. E. Population size in Particle Swarm Optimization. **Swarm and Evolutionary Computation**, v. 58, p. 100718, 2020.

POTYRAILO, R. A. On-Line Measurement. In: JURGEN, K. H. *et al.* **Encyclopedia of Materials: Science and Technology**. Amsterdam: Elsevier, 2001, p. 6401-6411.

PÜTÜN, A. E. *et al.* Bio-oil from olive oil industry wastes: Pyrolysis of olive residue under different conditions. **Fuel Processing Technology**, v. 87, n. 1, p. 25-32, 2005.

PÜTÜN, A. E. *et al.* Comparison between the “slow” and “fast” pyrolysis of tobacco residue. **Industrial Crops and Products**, v. 26, p. 307-314, 2007a.

PÜTÜN, A. E. *et al.* Production of biocrudes from biomass in a fixed-bed tubular reactor: Product yields and compositions. **Fuel**, v. 80, p. 1371-1378, 2001.

PÜTÜN, A. E. *et al.* Rapid and slow pyrolysis of pistachio shell: effect of pyrolysis conditions on the product yields and characterization of the liquid product. **International Journal of Energy Research**, v. 31, n. 5, p. 506-514, 2007b.

PÜTÜN, A. E.; APAYDM, E.; PÜTÜN, E. Bio-oil production from pyrolysis and steam pyrolysis of soybean-cake: product yields and composition—**Energy**, v. 27, n. 7, p. 703-713, 2002.

PÜTÜN, A. E.; APAYDM, E.; PÜTÜN, E. Rice straw as a bio-oil source via pyrolysis and steam pyrolysis. **Energy**, v. 29, p. 2171-2180, 2004.

PÜTÜN, E.; ATEŞ, F.; PÜTÜN, A. E. Catalytic pyrolysis of biomass in inert and steam atmospheres. **Fuel**, v. 87, n. 6, p. 815-824, 2008.

QIANG, L.; WEN-ZHI, L.; XI-FENG, Z. Overview of fuel properties of biomass fast pyrolysis oils. **Energy Conversion and Management**, v. 50, n. 5, p. 1376-1383, 2009.

QIANG, L.; XU LAI, Y.; XI FENG, Z. Analysis on chemical and physical properties of bio-

oil pyrolyzed from rice husk. **Journal of Analytical and Applied Pyrolysis**, v. 82, p. 191-198, 2008.

QIN, S. J. Neural Networks for Intelligent Sensors and Control — Practical Issues and Some Solutions. **Neural Systems for Control**, p. 213-234, 1997.

QU, T. *et al.* Experimental Study of Biomass Pyrolysis Based on Three Major Components: Hemicellulose, Cellulose, and Lignin. **Industrial & Engineering Chemistry Research**, v. 50, n. 18, p. 10424-10433, 2011.

QUAN, C.; GAO, N.; SONG, Q. Pyrolysis of biomass components in a TGA and a fixed-bed reactor: Thermochemical behaviors, kinetics, and product characterization. **Journal of Analytical and Applied Pyrolysis**, v. 121, p. 84-92, 2016.

QUINLAN, J. R. Combining Instance-Based and Model-Based Learning. **Machine Learning Proceedings**, v. 93, p. 236-243, 1993.

QUINLAN, J. R. Learning with continuous classes. **Australian Joint Conference on Artificial Intelligence**, v. 92, p. 343-348, 1992.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. R Foundation for Statistical Computing, 2022.

RAJA, S. A. *et al.* Flash pyrolysis of jatropha oil cake in electrically heated fluidized bed reactor. **Energy**, v. 35, p. 2819-2823, 2010.

RAWLINGS, J. O.; PANTULA, S. G.; DICKEY, D. A. **Applied Regression Analysis: A Research Tool**. 2. ed. New York: Springer-Verlag, 1998.

RAZALI, N. M.; WAH, Y. B. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. **Journal of Statistical Modeling and Analytics**, v. 2, n. 1, p. 21-33, 2011.

RAZUAN, R. *et al.* Pyrolysis and combustion of oil palm stone and palm kernel cake in fixed-bed reactors. **Bioresource Technology**, v. 101, n. 12, p. 4622-4629, 2010.

RENDALL, R.; PEREIRA, A.; REIS, M. An extended comparison study of large-scale data-driven prediction methods based on variable selection, latent variables, penalized regression, and machine learning. **Computer Aided Chemical Engineering**, v. 38, p. 1629-1634, 2016.

RUBIN, D. B. **Multiple Imputation for Nonresponse in Surveys**. New York: Wiley Interscience, 1987.

SAIKIA, R. *et al.* Pyrolysis and kinetic analyses of a perennial grass (*Saccharum ravannae* L.) from north-east India: Optimization through response surface methodology and product characterization. **Bioresource Technology**, v. 253, p. 304-314, 2018.

SARKER, S. D.; NAHAR, L. Applications of High-Performance Liquid Chromatography in the Analysis of Herbal Products. **Evidence-Based Validation of Herbal Medicine**, p. 406-425, 2015.

SCHÖLKOPF, B.; SMOLA, A. J. **Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond**. London: MIT Press, 2002.

SCHORR, C.; MUINONEN, M.; NURMINEN, F. Torrefaction of biomass. **Chemical Engineering**, v. 117, n. 7, 2012.

SCHRÖDER, E. Experiments on the pyrolysis of large beechwood particles in fixed beds. **Journal of Analytical and Applied Pyrolysis**, v. 71, p. 669-694, 2004.

ŞENSÖZ, S.; ANGIN, D. Pyrolysis of safflower (*Charthamus tinctorius* L.) seed press cake in a fixed-bed reactor: Part 2. Structural characterization of pyrolysis bio-oils. **Bioresource Technology**, v. 99, p. 5498-5504, 2008.

ŞENSÖZ, S.; DEMIRAL, I.; GERÇEL, H. F. Olive bagasse (*Olea europea* L.) pyrolysis. **Bioresource Technology**, v. 97, n. 3, p. 429-436, 2006.

ŞENSÖZ, S.; KAYNAR, I. Bio-oil production from soybean (*Glycine max* L.); Fuel properties of Bio-oil. **Industrial Crops and Products**, v. 23, n. 1, p. 99-105, 2006.

SERNEELS, S.; DE NOLF, E.; VAN ESPEN, P. J. Spatial sign preprocessing: A simple way to impart moderate robustness to multivariate estimators. **Journal of Chemical Information and Modeling**, v. 46, n. 3, p. 1402-1409, 2006.

SHAH, A. D. *et al.*, Comparison of Random Forest and Parametric Imputation Models for Imputing Missing Data Using MICE: A CALIBER Study. **American Journal of Epidemiology**, v.179, n. 6, p. 764-774, 2014.

SHAHSAVANI, D.; TARANTOLA, S.; RATTO, M. Evaluation of MARS modeling technique for sensitivity analysis of model output. **Procedia - Social and Behavioral Sciences**, v. 2, n. 6, p. 7737-7738, 2010.

SHARMA, J. *et al.* Multi-layer intrusion detection system with ExtraTrees feature selection, extreme learning machine ensemble, and softmax aggregation. **Eurasip Journal on Information Security**, v. 2019, n. 1, 2019.

SHARMA, R.; SHETH, P. N. Thermo-Chemical Conversion of Jatropha Deoiled Cake: Pyrolysis vs. Gasification. **International Journal of Chemical Engineering and Applications**, v. 6, n. 5, p. 376-380, 2015.

SHEN, D. K. *et al.* The pyrolytic degradation of wood-derived lignin from pulping process. **Bioresource Technology**, v. 101, p. 6136-6146, 2010.

SHEN, J. *et al.* Effects of particle size on the fast pyrolysis of oil mallee woody biomass. **Fuel**, v. 88, n. 10, p. 1810-1817, 2009.

SINGH, S.; CHAKRABORTY, J. P.; MONDAL, M. K. Pyrolysis of torrefied biomass: Optimization of process parameters using response surface methodology, characterization, and comparison of properties of pyrolysis oil from raw biomass. **Journal of Cleaner Production**, v. 272, 2020.

SMOLA, A. J.; SCHOLKOPF, B. A tutorial on support vector regression. **Statistics and Computing**, v. 14, p. 199-222, 2004.

SOUSA, S. I. V. *et al.* Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. **Environmental Modelling and Software**, v. 22, n. 1, p. 97-103, 2007.

SRICHAROENCHAIKUL, V. *et al.* Preparation and characterization of activated carbon from the pyrolysis of physic nut (*Jatropha curcas* L.) waste. **Energy and Fuels**, v. 22, p. 31-37, 2008.

STEFANIDIS, S. D. *et al.* A study of lignocellulosic biomass pyrolysis via the pyrolysis of cellulose, hemicellulose and lignin. **Journal of Analytical and Applied Pyrolysis**, v. 105, p. 143-150, 2014.

STEVENS, A.; RAMIREZ-LOPEZ, L. **An introduction to the prospectr package**. R package Vignette., 2021.

SULAIMAN, F.; ABDULLAH, N. Optimum conditions for maximizing pyrolysis liquids of oil palm empty fruit bunches. **Energy**, v. 36, p. 2352-2359, 2011.

SUN, Y. *et al.* Pyrolysis products from industrial waste biomass based on a neural network model. **Journal of Analytical and Applied Pyrolysis**, v. 120, p. 94-102, 2016.

TACQ, J. Multivariate Normal Distribution. In: PETERSON, P.; MCGAW, B.; BAKER, E. **International encyclopedia of education**. 3. ed. Amsterdam: Elsevier, 2010. p. 332-338.

TANG, Q. *et al.* Prediction of Bio-oil Yield and Hydrogen Contents Based on Machine Learning Method: Effect of Biomass Compositions and Pyrolysis Conditions. **Energy and Fuels**, v. 34, p. 11050-11060, 2020.

TIBSHIRANIT, R. Regression Shrinkage and Selection via the Lasso. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 58, n. 1, p. 267-288, 1996.

TRIPATHI, M.; SAHU, J. N.; GANESAN, P. Effect of process parameters on production of biochar from biomass waste through pyrolysis: A review. **Renewable and Sustainable Energy Reviews**, v. 55, p. 467-481, 2016.

TROYANSKAYA, O. *et al.* Missing value estimation methods for DNA microarrays. **Bioinformatics**, v. 17, n. 6, p. 520-525, 2001.

TSAI, W. T.; LEE, M. K.; CHANG, Y. M. Fast pyrolysis of rice husk: Product yields and compositions. **Bioresource Technology**, v. 98, n. 1, p. 22-28, 2007.

TSAI, W. T.; LEE, M. K.; CHANG, Y. M. Fast pyrolysis of rice straw, sugarcane bagasse and coconut shell in an induction-heating reactor. **Journal of Analytical and Applied Pyrolysis**, v. 76, n. 2, p. 230-237, 2006.

TUTZ, G.; RAMZAN, S. Improved methods for the imputation of missing data by nearest neighbor methods. **Computational Statistics and Data Analysis**, v. 90, p. 84-99, 2015.

UCAR, S.; OZKAN, A. R. Characterization of products from the pyrolysis of rapeseed oil cake. **Bioresource Technology**, v. 99, n. 18, p. 8771-8776, 2008.

ULLAH, Z. *et al.* A comparative study of machine learning methods for bio-oil yield prediction - A genetic algorithm-based features selection—**Bioresource Technology**, v. 335, 2021.

UZUN, B. B.; PÜTÜN, A. E.; PÜTÜN, E. Fast pyrolysis of soybean cake: Product yields and compositions. **Bioresource Technology**, v. 97, n. 4, p. 569-576, 2006.

VAN BUUREN, S.; OUDSHOORN, K. G. mice: Multivariate Imputation by Chained Equations in R. **Journal of Statistical Software**, v. 45, p. 1-67, 2011.

VARMA, A. K. *et al.* Pyrolysis of wood sawdust: Effects of process parameters on products yield and characterization of products. **Waste Management**, v. 89, p. 224-235, 2019.

VARMA, A. K.; MONDAL, P. Pyrolysis of sugarcane bagasse in semi-batch reactor: Effects of process parameters on product yields and characterization of products—**Industrial Crops and Products**, v. 95, p. 704-717, 2017.

VENDERBOSCH, R. H.; PRINS, W. Fast pyrolysis technology development. **Biofuels, Bioproducts and Biorefining**, v. 4, n. 2, p. 178-208, 2010.

VIEIRA, F. R. *et al.* Optimization of slow pyrolysis process parameters using a fixed bed reactor for biochar yield from rice husk. **Biomass and Bioenergy**, v. 132, 2020.

VITOLO, S. *et al.* Catalytic upgrading of pyrolytic oils over HZSM-5 zeolite: Behaviour of the catalyst when used in repeated upgrading-regenerating cycles. **Fuel**, v. 80, n. 1, p. 17-26, 2001.

VITTINGHOFF, E. *et al.* Linear and Non-Linear Regression Methods in Epidemiology and Biostatistics. **Handbook of Statistics**. v. 27, p. 148-186, 2007.

VOLLI, V.; SINGH, R. K. Production of bio-oil from de-oiled cakes by thermal pyrolysis. **Fuel**, v. 96, p. 579-585, 2012.

WANG, S. *et al.* Influence of the interaction of components on the pyrolysis behavior of biomass. **Journal of Analytical and Applied Pyrolysis**, v. 91, p. 183-189, 2011.

WENTING, F. *et al.* Influence of methanol additive on bio-oil stability. **International Journal of Agricultural and Biological Engineering**, v. 7, n. 3, p. 83-92, 2014.

WINTERS-MINER, L. A. *et al.* Prediction in Medicine - The Data Mining Algorithms of Predictive Analytics. **Practical Predictive Analytics and Decisioning Systems for Medicine**, p. 239-259, 2015.

WU, S. R. *et al.* Comparison of oil-tea shell and Douglas-fir sawdust for the production of bio-oils and chars in a fluidized-bed fast pyrolysis system. **Fuel**, v. 175, p. 57-63, 2016.

XING, J. *et al.* Estimating biomass major chemical constituents from ultimate analysis using a

random forest model. **Bioresource Technology**, v. 288, 2019.

XU, R. *et al.* Bio-oil production by flash pyrolysis of sugarcane residues and post-treatments of the aqueous phase. **Journal of Analytical and Applied Pyrolysis**, v. 91, n. 1, p. 263-272, 2011.

YAMAN, S. Pyrolysis of biomass to produce fuels and chemical feedstocks. **Energy Conversion and Management**, v. 45, n. 5, p. 651-671, 2004.

YANG, Z. R.; YANG, Z. Artificial Neural Networks. **Comprehensive Biomedical Physics**, v. 6, p. 1-17, 2014.

YANIK, J. *et al.* Fast pyrolysis of agricultural wastes: Characterization of pyrolysis products. **Fuel Processing Technology**, v. 88, n. 10, p. 942-947, 2007.

YAP, B. W.; SIM, C. H. Comparisons of various types of normality tests. **Journal of Statistical Computation and Simulation**, v. 81, n. 12, p. 2141-2155, 2011.

YIN, R. *et al.* Characterization of bio-oil and bio-char obtained from sweet sorghum bagasse fast pyrolysis with fractional condensers. **Fuel**, v. 112, p. 96-104, 2013.

ZHANG, Q. *et al.* Review of biomass pyrolysis oil properties and upgrading research. **Energy Conversion and Management**, v. 48, n. 1, p. 87-92, 2007.

ZHANG, S. *et al.* Upgrading of liquid fuel from the pyrolysis of biomass. **Bioresource Technology**, v. 96, n. 5, p. 545-550, 2005.

ZHOU, J. *et al.* Feasibility of Random-Forest Approach for Prediction of Ground Settlements Induced by the Construction of a Shield-Driven Tunnel. **International Journal of Geomechanics**, v. 17, n. 6, 2016.

ZHOU, J. *et al.* Random forests and cubist algorithms for predicting shear strengths of rockfill materials. **Applied Sciences (Switzerland)**, v. 9, n. 8, p. 1-16, 2019.

ZHU, X.; LI, Y.; WANG, X. Machine learning prediction of biochar yield and carbon contents in biochar based on biomass characteristics and pyrolysis conditions. **Bioresource Technology**, v. 288, n. April, p. 121527, 2019.

ZOU, H.; HASTIE, T. Regularization and variable selection via the elastic net. **Journal of the Royal Statistical Society. Series B: Statistical Methodology**, v. 67, n. 2, p. 301-320, 2005.

ANEXO A - Biomassas presentes no banco de dados e o número de observações

Biomassa	Observações	Referências
Casca de arroz	160	Alvarez <i>et al.</i> , 2014; Bakar Titiloye, 2013; Biswas <i>et al.</i> , 2017; Di Blasi <i>et al.</i> , 1999; Fu <i>et al.</i> , 2011; Guo <i>et al.</i> , 2011; Heo <i>et al.</i> , 2010b; Hsu; Huang; Kuo, 2015; Huang <i>et al.</i> , 2016; Isa <i>et al.</i> , 2011; Lim <i>et al.</i> , 2016; Maddi; Viamajala; Varanasi, 2011; Qiang; Xu Lai; Xi Feng, 2008; Raveendran; Ganesh, 1996; Raveendran; Ganesh; Khilar, 1995, 1996; Tsai; Lee; Chang, 2007; Vieira <i>et al.</i> , 2020; Wang <i>et al.</i> , 2016; Williams; Nugranad, 2000; Worasuwanarak; Sonobe; Tanthapanichakoon, 2007; Zhai <i>et al.</i> , 2015; Zhang <i>et al.</i> , 2015; Zhao; Li, 2016; Zheng, 2007; Zheng <i>et al.</i> , 2006
Casca de palma	107	Abnisa <i>et al.</i> , 2011; Abnisa; Wan Daud; Sahu, 2011; Asadullah <i>et al.</i> , 2013; Chang <i>et al.</i> , 2016; Islam; Zailani; Nasir Ani, 1999; Kim.; Jung; Kim, 2010; Kim <i>et al.</i> , 2013; Lee <i>et al.</i> , 2013b
Torta de pinhão manso	103	Biradar; Subramanian; Dastidar, 2014; Figueiredo <i>et al.</i> , 2011; Jourabchi; Gan; Ng, 2014; Kanaujia <i>et al.</i> , 2016; Kim, 2016; Kim <i>et al.</i> , 2013; Raja <i>et al.</i> , 2010; Sharma; Sheth, 2015; Sricharoenchaikul <i>et al.</i> , 2008; Sricharoenchaikul; Atong, 2009
Pinho	83	Aguado <i>et al.</i> , 2000; Aho <i>et al.</i> , 2008; Alvarez <i>et al.</i> , 2015a; Amutio <i>et al.</i> , 2012; Chang <i>et al.</i> , 2016; Demirbaş, 2002a; Garcia-Perez <i>et al.</i> , 2008; Ingram <i>et al.</i> , 2008; Kang <i>et al.</i> , 2006; Kim <i>et al.</i> , 2012; Oasmaa <i>et al.</i> , 2010; Oudenhoven <i>et al.</i> , 2016; Şensöz; Can, 2002a, 2002b; Song; Tahmasebi; Yu, 2014; Wang <i>et al.</i> , 2005; Williams; Besler, 1996
Semente de colza	79	Onay, 2003; Onay; Beis; Koçkar, 2001; Onay; Koçkar, 2003, 2004; Sánchez <i>et al.</i> , 2009; Şensöz; Angin; Yorgun, 2000
Bagaço de cana	73	Asadullah <i>et al.</i> , 2007; Islam; Parveen; Haniu, 2010; Lee <i>et al.</i> , 2013b; Scott; Plskorz; Radleln, 1985; Tsai; Lee; Chang, 2006; Varma; Mondal, 2017; Xu <i>et al.</i> , 2011; Zanzi; Sjöström; Björnbom, 1996
Palha de arroz	68	Biswas <i>et al.</i> , 2017; Fu <i>et al.</i> , 2010; Fu <i>et al.</i> , 2011; Jung; Kang; Kim, 2008; Lee <i>et al.</i> , 2013b; Park <i>et al.</i> , 2014; Pütün; Apaydm; Pütün, 2004; Raveendran; Ganesh; Khilar, 1996; Tsai; Lee; Chang, 2006; Worasuwanarak; Sonobe; Tanthapanichakoon, 2007

Biomassa	Observações	Referências
Madeira de carvalho	62	Ellens; Brown, 2012; Figueiredo <i>et al.</i> , 1989; Ingram <i>et al.</i> , 2008; Valenzuela-Calahorro <i>et al.</i> , 1987
Madeira de faia	55	Balci; Doğu; Yücel, 1993; Beaumont; Schwob, 1984; Demirbas, 2007; Di Blasi; Branca, 2001; Schröder, 2004; Wang <i>et al.</i> , 2005
Casca de avelã	53	Balci; Doğu; Yücel, 1993; Bonelli; Cerrella; Cukierman, 2003; Demirbas, 2006, 2007; Demirbaş, 1999; Güllü, 2003; Özçimen; Ersoy-Meriçboyu, 2010; Pütün <i>et al.</i> , 2001; Pütün; Özean; Pütün, 1999
Sabugo de milho	52	Ateş; Işıkdağ, 2009; Biswas <i>et al.</i> , 2017; Demirbas, 2004; Maddi; Viamajala; Varanasi, 2011; Mullen <i>et al.</i> , 2010; Raveendran; Ganesh, 1996; Raveendran; Ganesh; Khilar, 1996; Worasuwannarak; Sonobe; Tanthapanichakoon, 2007; Yanik <i>et al.</i> , 2007; Zhang <i>et al.</i> , 2009
Resíduo de papel	47	Islam <i>et al.</i> , 2005; Li; Zhang; Zhuang, 2005
Torta de soja	47	Pütün; Apaydin; Pütün, 2002; Şensöz; Kaynar, 2006; Uzun; Pütün; Pütün, 2006
Semente de cártamo	46	Beis; Onay; Koçkar, 2002; Onay, 2007
Casca de nozes	42	Demirbas, 2006; Onay; Beis; Koçkar, 2004
Resíduo de azeitona	38	Encinar <i>et al.</i> , 1998; Pütün <i>et al.</i> , 2005; Şensöz; Demiral; Gerçel, 2006; Zanzi; Sjöström; Björnbom, 2002
Resíduo de uva	38	Demiral; Ayan, 2011; Di Blasi <i>et al.</i> , 1999; Encinar <i>et al.</i> , 1998
Lodo de esgoto	37	Alvarez <i>et al.</i> , 2015a, 2015b; Arazo <i>et al.</i> , 2017; Fonts <i>et al.</i> , 2009; Huang <i>et al.</i> , 2014; Menéndez; Inguanzo; Pis, 2002
Serragem de cedro	35	Krishna <i>et al.</i> , 2016; Varma <i>et al.</i> , 2019
Palha de trigo	34	Ateş; Işıkdağ, 2008; Biswas <i>et al.</i> , 2017; Chang <i>et al.</i> , 2016; Di Blasi <i>et al.</i> , 1999; Fahmi <i>et al.</i> , 2008; Oudenhoven <i>et al.</i> , 2016; Scott; Plskorz; Radleln, 1985; Yanik <i>et al.</i> , 2007; Zanzi; Sjöström; Björnbom, 2002

Biomassa	Observações	Referências
Torta de cártamo	34	Angin, 2013; Şensöz; Angin, 2008b, 2008a
Casca de pistache	33	Açikalin; Karaca; Bolat, 2012; Apaydin-Varol; Pütün; Pütün, 2007; Pütün <i>et al.</i> , 2007b
Eucalipto	32	Garcia-Perez <i>et al.</i> , 2008; Heidari <i>et al.</i> , 2014; Oasmaa <i>et al.</i> , 2010; Părpăriță <i>et al.</i> , 2014; Shen <i>et al.</i> , 2009
Lignina	32	French; Czernik, 2010; Liu <i>et al.</i> , 2017; Nunn <i>et al.</i> , 1985a; Quan; Gao; Song, 2016; Raveendran; Ganesh; Khilar, 1996; Sharma <i>et al.</i> , 2004; Shen <i>et al.</i> , 2010; Stefanidis <i>et al.</i> , 2014; Wang <i>et al.</i> , 2011; Worasuwannarak; Sonobe; Tanthapanichakoon, 2007
Agulha de pinheiro	31	Mandal <i>et al.</i> , 2018
Mandioca	30	Paenpong; Pattiya, 2016; Pattiya, 2011; Pattiya; Sukkasi; Goodwin, 2012; Pattiya; Suttibak, 2012; Suttibak; Sriprateep; Pattiya, 2012
Torta de algodão	30	Özbay <i>et al.</i> , 2001
Casca de azeitona	29	Demirbas, 2004, 2007; Di Blasi <i>et al.</i> , 1999
Bagaço de girassol	28	Gerçel, 2002; Pütün <i>et al.</i> , 2001
Cacho vazio de palma	28	Abdullah; Gerhauser, 2008; Kim <i>et al.</i> , 2013; Sulaiman; Abdullah, 2011
Caule de mandioca	28	Pattiya, 2011; Pattiya; Sukkasi; Goodwin, 2012; Pattiya; Suttibak, 2012
Semente de <i>Cascabela thevetia</i>	28	Mishra; Muraraka; Mohanty, 2020
Alga <i>Saccharina japonica</i>	26	Ly <i>et al.</i> , 2015, 2016
Madeira de acácia torrificada	26	Ly <i>et al.</i> , 2015, 2016
Semente de nim	26	Mishra; Mohanty, 2018; Nayan; Kumar; Singh, 2013

Biomassa	Observações	Referências
Casca de coco	25	Fagbemi; Khezami; Capart, 2001; Lee <i>et al.</i> , 2013b; Raveendran; Ganesh; Khilar, 1995; Tsai; Lee; Chang, 2006
Celulose	23	Casoni <i>et al.</i> , 2015; French; Czernik, 2010; Liu <i>et al.</i> , 2017; Quan; Gao; Song, 2016; Raveendran; Ganesh; Khilar, 1996; Scott <i>et al.</i> , 1988; Scott; Plskorz; Radleln, 1985; Stefanidis <i>et al.</i> , 2014; Wang <i>et al.</i> , 2011; Worasuwannarak; Sonobe; Tanthapanichakoon, 2007
<i>Euphorbia rigida</i>	23	Kiliç; Pütün; Pütün, 2014; Pütün <i>et al.</i> , 2001
Madeira	23	Di Blasi <i>et al.</i> , 1999; Fagbemi; Khezami; Capart, 2001; Gray; Corcoran; Gavalas, 1985; Maddi; Viamajala; Varanasi, 2011
Resíduo de chá	23	Demirbas, 2004; Güllü, 2003
Palha e caule de algodão	21	Pütün, 2002
Talo de milho	21	Ali <i>et al.</i> , 2014a; Fu <i>et al.</i> , 2010; Fu <i>et al.</i> , 2011
Resíduo de tabaco	20	Pütün <i>et al.</i> , 2007a
Torta de nim	19	Dhanavath <i>et al.</i> , 2019; Volli; Singh, 2012
Casca de algodão	17	Çağlar; Demirbaş, 2000
Talo de gergelim	17	Ateş; Pütün; Pütün, 2004
<i>Cynara cardunculus</i>	16	Encinar; González; González, 2000
Polpa de damasco	16	Özbay <i>et al.</i> , 2008
Celulose e lignina	15	Liu <i>et al.</i> , 2017; Raveendran; Ganesh; Khilar, 1996
Madeira de bétula	15	Thunman <i>et al.</i> , 2001; Yu <i>et al.</i> , 1997; Zanzi; Sjöström; Björnbom, 1996, 2002
Torta de karanja	15	Chutia; Kataki; Bhaskar, 2014
Casca de melaleuca	14	Wu <i>et al.</i> , 2016

Biomassa	Observações	Referências
Serragem de Douglas-fir	14	Wu <i>et al.</i> , 2016
Casca de amendoim	13	Bonelli; Cerrella; Cukierman, 2003; Raveendran; Ganesh, 1996; Raveendran; Ganesh; Khilar, 1995, 1996
Pó de juta	13	Choudhury <i>et al.</i> , 2014
Serragem de Abeto	13	Demirbas, 2007; Dufour <i>et al.</i> , 2009; Părpăriță <i>et al.</i> , 2014; Thunman <i>et al.</i> , 2001
Serragem de carpino	13	Morali; Yavuzel; Şensöz, 2016
Bambu	12	Hernandez-Mena; Pécora; Beraldo, 2014; Jung; Kang; Kim, 2008; Wang <i>et al.</i> , 2005
Casca de amêndoa	12	Balci; Doğu; Yücel, 1993; Demirbas, 2006
Casca de carpino	12	Morali; Şensöz, 2015
Hemicelulose e lignina	12	Liu <i>et al.</i> , 2017
Semente de acácia	12	Garg; Anand; Kumar, 2016
Torta de colza	12	Özçimen; Karaosmanoğlu, 2004; Ucar; Ozkan, 2008
Casca de pinheiro	11	Ingram <i>et al.</i> , 2008; Şensöz, 2003
Casca de semente de girassol	11	Casoni <i>et al.</i> , 2015; Demirbas, 200
Palha de milho	11	Agblevor; Besler; Wiseloge, 1995; Mullen <i>et al.</i> , 2010; Scott; Plskorz; Radleln, 1985
Hemicelulose	10	Liu <i>et al.</i> , 2017; Quan; Gao; Song, 2016; Raveendran; Ganesh; Khilar, 1996; Stefanidis <i>et al.</i> , 2014; Wang <i>et al.</i> , 2011; Worasuwannarak; Sonobe; Tanthapanichakoon, 2007
Madeira de poplar	10	Agblevor; Besler; Wiseloge, 1995; Kim <i>et al.</i> , 2011
Serragem de populus nigra	10	Makibar <i>et al.</i> , 2015
Topo de cana	10	Pattiya; Sukkasi; Goodwin, 2012; Xu <i>et al.</i> , 2011

Biomassa	Observações	Referências
Caroço de damasco	9	Li <i>et al.</i> , 2004; Özçimen; Ersoy-Meriçboyu, 2010
Casca de uva	9	Xu <i>et al.</i> , 2009
Lemna minor	9	Muradov <i>et al.</i> , 2010
Panicum virgatum	9	Agblevor; Besler; Wiseloge, 1995; Boateng <i>et al.</i> , 2007; Fahmi <i>et al.</i> , 2008; Imam; Capareda, 2012; Mullen; Boateng, 2008
Talo de algodão	9	Ali <i>et al.</i> , 2014b; Zheng; Yi; Wang, 2008
Madeira de álamo trémulo e bétula	8	Zanzi; Sjöström; Björnbom, 1996
Madeira de abeto sem casca	8	Adam <i>et al.</i> , 2006
Madeira de álamo-trémulo e poplar	8	Scott; Plskorz; Radlein, 1985
Resíduo de madeira	8	Bridgwater; Meier; Radlein, 1999; Horne; Williams, 1996
Talo de tabaco	8	Demirbaş, 2002a
Torta de mostarda	8	Volli; Singh, 2012
Undaria pinnatifida	8	Bae <i>et al.</i> , 2011
Casca de cereja	7	Duman <i>et al.</i> , 2011
Geodae-Uksae 1	7	Lee <i>et al.</i> , 2013a
Semente de cereja	7	Duman <i>et al.</i> , 2011
Casca e semente de uva	6	Xu <i>et al.</i> , 2009
Celulose, lignina e hemicelulose	6	Stefanidis <i>et al.</i> , 2014; Wang <i>et al.</i> , 2011
Folha de cana	6	Pattiya; Sukkasi; Goodwin, 2012; Zanzi; Sjöström; Björnbom, 1996

Biomassa	Observações	Referências
Madeira de <i>Liquidambar styraciflua</i>	6	Nunn <i>et al.</i> , 1985b
Palha de aveia	6	Ateş; Işikdağ, 2008
Pinho lavado com ácido	6	Oudenhoven <i>et al.</i> , 2016
Resíduo de móveis	6	Heo <i>et al.</i> , 2010a
Semente de mamona	6	Singh; Shadangi, 2011
Torta de amendoim	6	Agrawalla; Kumar; Singh, 2011
Torta de gergelim	6	Volli; Singh, 2012
Casca de arroz e Serragem de madeira	5	Zheng <i>et al.</i> , 2006
Casca de castanha de caju	5	Das; Ganesh, 2003
Casca de álamo-trémulo e poplar	5	Scott; Plskorz; Radleln, 1985
Alga <i>Chlorella protothecoides</i> heterotróficas	5	Miao; Wu, 2004
<i>Euphorbia macroclada</i>	5	Tuncel; Gerçel, 2004
Madeira de <i>Acer rubrum</i>	5	Scott; Plskorz; Radleln, 1985
Madeira de Quebracho	5	Della Rocca <i>et al.</i> , 1999; Zanzi; Sjöström; Björnbom, 1996
Palha de algodão	5	Fu <i>et al.</i> , 2011

Biomassa	Observações	Referências
Ponta de palha	5	Fagbemi; Khezami; Capart, 2001
Resíduo de cana de açúcar	5	Xu <i>et al.</i> , 2011
Semente de karanja	5	Nayan; Kumar; Singh, 2012
Serragem de madeira	5	Zheng <i>et al.</i> , 2006
Casca de semente de pau-ferro	4	Bordoloi <i>et al.</i> , 2015
Casca de sementes de karanja	4	Bordoloi <i>et al.</i> , 2015
<i>Alga Laminaria japônica</i>	4	Bae <i>et al.</i> , 2011
Licor preto	4	Demirbaş, 2002b
Musgo	4	Scott; Plskorz; Radleln, 1985
Palha de leguminosa	4	Li <i>et al.</i> , 2004
<i>Alga Porphyra tenera</i>	4	Bae <i>et al.</i> , 2011
<i>Alga Scenedesmus dimorphus</i>	4	Bordoloi <i>et al.</i> , 2016
Semente de romã	4	Uçar; Karagöz, 2009
Amêndoa da semente de manga	3	Lazzari <i>et al.</i> , 2016
Casca de macaúba	3	Cardoso; Machado; Caramão, 2016
Casca interna de pinhão manso	3	Manurung <i>et al.</i> , 2009; Murata <i>et al.</i> , 2012

Biomassa	Observações	Referências
Endocarpo de macaúba	3	Cardoso; Machado; Caramão, 2016
Fibra de macaúba	3	Cardoso; Machado; Caramão, 2016
Lignito	3	Song; Tahmasebi; Yu, 2014
<i>Lyngbya sp</i>	3	Maddi; Viamajala; Varanasi, 2011
Madeira de acácia	3	Reza <i>et al.</i> , 2019; Singh; Chakraborty; Mondal, 2020
Madeira de álamo-trémulo	3	French; Czernik, 2010
Polpa de pêsego	3	Özbay <i>et al.</i> , 2008
Semente de palma	3	Razuan <i>et al.</i> , 2010
Tegumento da semente de manga	3	Lazzari <i>et al.</i> , 2016
Torta de palma	3	Razuan <i>et al.</i> , 2010
Fibra de coco	2	Raveendran; Ganesh, 1996; Raveendran; Ganesh; Khilar, 1996
Capim amarelo	2	Fahmi <i>et al.</i> , 2008; Oasmaa <i>et al.</i> , 2010
Casca externa de pinhão manso	2	Murata <i>et al.</i> , 2012
Caule de <i>Arundo Donax</i>	2	Bartoli <i>et al.</i> , 2016
Celulose e hemicelulose	2	Stefanidis <i>et al.</i> , 2014; Wang <i>et al.</i> , 2011
Alga <i>Chlorella protothecoides</i>	2	Miao; Wu, 2004; Miao; Wu; Yang, 2004
Feno	2	Oasmaa <i>et al.</i> , 2010
Folhas de <i>Arundo Donax</i>	2	Bartoli <i>et al.</i> , 2016

Biomassa	Observações	Referências
Galho de pinhão manso	2	Murata <i>et al.</i> , 2012
Madeira de Leucena	2	Raveendran; Ganesh, 1996; Raveendran; Ganesh; Khilar, 1996
madeira de peroba branca	2	Della Rocca <i>et al.</i> , 1999
Rizoma de <i>Arundo Donax</i>	2	Bartoli <i>et al.</i> , 2016
Semente de uva	2	Özçimen; Ersoy-Meriçboyu, 2010; Părpăriță <i>et al.</i> , 2014
Farinha de sagu, capim napier e casca de arroz	1	Lim <i>et al.</i> , 2016
Bagaço de sorgo	1	Yin <i>et al.</i> , 2013
Brassica rapa	1	Părpăriță <i>et al.</i> , 2014
Capim Napier	1	Lim <i>et al.</i> , 2016
Capim napier casca de arroz	1	Lim <i>et al.</i> , 2016
Casca de árvore	1	Lee <i>et al.</i> , 2013b
Casca de carvalho	1	Ingram <i>et al.</i> , 2008
Casca externa da avelã	1	Özçimen; Ersoy-Meriçboyu, 2010
Caule de alfafa	1	Mullen; Boateng, 2008
Caule de alfafa com flor	1	Mullen; Boateng, 2008
Alga <i>Cladophora</i>	1	Maddi; Viamajala; Varanasi, 2011
<i>Dactylis glomerata</i>	1	Fahmi <i>et al.</i> , 2008
Farinha de sagu e capim Napier	1	Lim <i>et al.</i> , 2016

Biomassa	Observações	Referências
Farinha de sagu e casca de arroz	1	Lim <i>et al.</i> , 2016
Farinha de sagu	1	Lim <i>et al.</i> , 2016
<i>Festuca arundinacea</i>	1	Fahmi <i>et al.</i> , 2008
Folha da palma	1	Abnisa <i>et al.</i> , 2013
Fruto de pinhão manso	1	Figueiredo <i>et al.</i> , 2011
Girassol	1	Sánchez <i>et al.</i> , 2009
Gramma	1	Părpăriță <i>et al.</i> , 2014
Haste da folha da palma	1	Abnisa <i>et al.</i> , 2013
Lodo de óleo de palma	1	Thangalazhy-Gopakumar <i>et al.</i> , 2015
<i>Lolium perene</i>	1	Fahmi <i>et al.</i> , 2008
Madeira de salgueiro	1	Fahmi <i>et al.</i> , 2008
Meio da folha da palma	1	Abnisa <i>et al.</i> , 2013
Alga <i>Microcystis aeruginosa</i>	1	Miao; Wu; Yang, 2004
Palha de cevada	1	Oasmaa <i>et al.</i> , 2010
Palha de colza	1	Oasmaa <i>et al.</i> , 2010
Palha de trigo lavada ácido/minerais	1	Oudenhoven <i>et al.</i> , 2016
Palha de trigo lavada com ácido	1	Oudenhoven <i>et al.</i> , 2016
Pellets de palha	1	Oudenhoven <i>et al.</i> , 2016
Pinha	1	Părpăriță <i>et al.</i> , 2014

Biomassa	Observações	Referências
Pinho lavado com ácido/minerais	1	Oudenhoven <i>et al.</i> , 2016
Resíduo de eucalipto	1	Amutio <i>et al.</i> , 2015
Resíduo de floresta estocada e fresca	1	Oasmaa <i>et al.</i> , 2010
Talo de orégano	1	Yanik <i>et al.</i> , 2007
Tronco da palma	1	Abnisa <i>et al.</i> , 2013
Tronco de árvore	1	Lee <i>et al.</i> , 2013b

ANEXO B – Referências bibliográficas dos estudos presentes no banco de dados

ABDULLAH, N.; GERHAUSER, H. Bio-oil derived from empty fruit bunches. **Fuel**, v. 87, n. 12, p. 2606-2613, 2008.

ABNISA, F. *et al.* Utilization of oil palm tree residues to produce bio-oil and bio-char via pyrolysis. **Energy Conversion and Management**, v. 76, p. 1073-1082, 2013.

ABNISA, F. *et al.* Utilization possibilities of palm shell as a source of biomass energy in Malaysia by producing bio-oil in pyrolysis process. **Biomass and Bioenergy**, v. 35, n. 5, p. 1863-1872, 2011.

ABNISA, F.; WAN DAUD, W. M. A.; SAHU, J. N. Optimization and characterization studies on bio-oil production from palm shell by pyrolysis using response surface methodology. **Biomass and Bioenergy**, v. 35, n. 8, p. 3604-3616, 2011.

AÇIKALIN, K.; KARACA, F.; BOLAT, E. Pyrolysis of pistachio shell: Effects of pyrolysis conditions and analysis of products. **Fuel**, v. 95, p. 169-177, 2012.

ADAM, J. *et al.* In situ, catalytic upgrading of biomass-derived fast pyrolysis vapors in a fixed bed reactor using mesoporous materials. **Microporous and Mesoporous Materials**, v. 96, n. 1-3, p. 93-101, 2006.

AGBLEVOR, F. A.; BESLER, S.; WISELOGEL, A. E. Fast Pyrolysis of Stored Biomass Feedstocks. **Energy and Fuels**, v. 9, n. 4, p. 635-640, 1995.

AGRAWALLA, A.; KUMAR, S.; SINGH, R. K. Pyrolysis of groundnut de-oiled cake and characterization of the liquid product. **Bioresource Technology**, v. 102, n. 22, p. 10711-10716, 2011.

AGUADO, R. *et al.* Pyrolysis of sawdust in a conical spouted bed reactor. Yields and product composition. **Industrial and Engineering Chemistry Research**, v. 39, p. 1925-1933, 2000.

AHO, A. *et al.* Catalytic pyrolysis of woody biomass in a fluidized bed reactor: Influence of the zeolite structure. **Fuel**, v. 87, n. 12, p. 2493-2501, 2008.

ALI, N. *et al.* Effect of Temperature on the Bio Oil Yield from Pyrolysis of Maize Stalks in Fluidized Bed Reactor. **Journal of Pakistan Institute of Chemical Engineers**, v. 42, n. 2, p. 79-86, 2014a.

ALI, N. *et al.* Fast Pyrolysis Of Pakistani Cotton Stalks In Fluidized Bed Reactor: Design And Preliminary Results. **Life Science Journal**, v. 11, n. 7, p. 203, 2014b.

ALVAREZ, J. *et al.* Bio-oil production from rice husk fast pyrolysis in a conical spouted bed reactor. **Fuel**, v. 128, p. 162-169, 2014.

ALVAREZ, J. *et al.* Fast co-pyrolysis of sewage sludge and lignocellulosic biomass in a conical spouted bed reactor. **Fuel**, v. 159, p. 810-818, 2015a.

ALVAREZ, J. *et al.* Sewage sludge valorization by flash pyrolysis in a conical spouted bed reactor. **Chemical Engineering Journal**, v. 273, p. 173-183, 2015b.

AMUTIO, M. *et al.* Fast pyrolysis of eucalyptus waste in a conical spouted bed reactor. **Bioresource Technology**, v. 194, p. 225-232, 2015.

AMUTIO, M. *et al.* Influence of temperature on biomass pyrolysis in a conical spouted bed reactor. **Resources, Conservation and Recycling**, v. 59, p. 23-31, 2012.

ANGIN, D. Effect of pyrolysis temperature and heating rate on biochar obtained from pyrolysis of safflower seed press cake. **Bioresource Technology**, v. 128, p. 593-597, 2013.

APAYDIN-VAROL, E.; PÜTÜN, E.; PÜTÜN, A. E. Slow pyrolysis of pistachio shell. **Fuel**, v. 86, n. 12-13, p. 1892-1899, 2007.

ARAZO, R. O. *et al.* Bio-oil production from dry sewage sludge by fast pyrolysis in an electrically-heated fluidized bed reactor. **Sustainable Environment Research**, v. 27, n. 1, p. 7-14, 2017.

ASADULLAH, M. *et al.* Production and detailed characterization of bio-oil from fast pyrolysis of palm kernel shell. **Biomass and Bioenergy**, v. 59, p. 316-324, 2013.

ASADULLAH, M. *et al.* Production of bio-oil from fixed bed pyrolysis of bagasse. **Fuel**, v. 86, p. 2514-2520, 2007.

ATEŞ, F.; İŞIKDAĞ, M. A. Evaluation of the role of the pyrolysis temperature in straw biomass samples and characterization of the oils by GC/MS. **Energy and Fuels**, v. 22, p. 1936-1943, 2008.

ATEŞ, F.; İŞIKDAĞ, M. A. Influence of temperature and alumina catalyst on pyrolysis of corncob. **Fuel**, v. 88, p. 1991-1997, 2009.

ATEŞ, F.; PÜTÜN, E.; PÜTÜN, A. E. Fast pyrolysis of sesame stalk: Yields and structural analysis of bio-oil. **Journal of Analytical and Applied Pyrolysis**, v. 71, n. 2, p. 779-790, 2004.

BAE, Y. J. *et al.* The characteristics of bio-oil produced from the pyrolysis of three marine macroalgae. **Bioresource Technology**, v. 102, p. 3512-3520, 2011.

BAKAR, M. S. A.; TITLOYE, J. O. Catalytic pyrolysis of rice husk for bio-oil production. **Journal of Analytical and Applied Pyrolysis**, v. 103, p. 362-368, 2013.

BALCI, S.; DOĞU, T.; YÜCEL, H. Pyrolysis Kinetics of Lignocellulosic Materials. **Industrial and Engineering Chemistry Research**, v. 32, p. 2573-2579, 1993.

BARTOLI, M. *et al.* Production of bio-oils and bio-char from *Arundo donax* through microwave assisted pyrolysis in a multimode batch reactor. **Journal of Analytical and Applied Pyrolysis**, v. 122, p. 479-489, 2016.

BEAUMONT, O.; SCHWOB, Y. Influence of physical and chemical parameters on wood

pyrolysis. **Industrial & Engineering Chemistry Process Design and Development**, v. 23, p. 637-641, 1984.

BEIS, S. H.; ONAY, O.; KOÇKAR, O. M. Fixed-bed pyrolysis of safflower seed: Influence of pyrolysis parameters on product yields and compositions. **Renewable Energy**, v. 26, n. 1, p. 21-32, 2002.

BIRADAR, C. H.; SUBRAMANIAN, K. A.; DASTIDAR, M. G. Production and fuel quality upgradation of pyrolytic bio-oil from *Jatropha Curcas* de-oiled seed cake. **Fuel**, v. 119, p. 81-89, 2014.

BISWAS, B. *et al.* Pyrolysis of agricultural biomass residues: Comparative study of corn cob, wheat straw, rice straw and rice husk. **Bioresource Technology**, v. 237, p. 57-63, 2017.

BOATENG, A. A. *et al.* Bench-scale fluidized-bed pyrolysis of switchgrass for bio-oil production. **Industrial and Engineering Chemistry Research**, v. 46, n. 7, p. 1891-1897, 2007.

BONELLI, P. R.; CERRELLA, E. G.; CUKIERMAN, A. L. Slow Pyrolysis of Nutshells: Characterization of Derived Chars and Process Kinetics. **Energy Sources**, v. 25, p. 767-778, 2003.

BORDOLOI, N. *et al.* Characterization of bio-oil and its sub-fractions from pyrolysis of *Scenedesmus dimorphus*. **Renewable Energy**, v. 98, p. 245-253, 2016.

BORDOLOI, N. *et al.* Pyrolysis of *Mesua ferrea* and *Pongamia glabra* seed cover: Characterization of bio-oil and its sub-fractions. **Bioresource Technology**, v. 178, p. 83-89, 2015.

BRIDGWATER, A. V.; MEIER, D.; RADLEIN, D. An overview of fast pyrolysis of biomass. **Organic Geochemistry**, v. 30, n. 12, p. 1479-1493, 1999.

ÇAĞLAR, A.; DEMİRBAŞ, A. Conversion of cotton cocoon shell to liquid products by pyrolysis. **Energy Conversion and Management**, v. 41, p. 1749-1756, 2000.

CARDOSO, C. A. L.; MACHADO, M. E.; CARAMÃO, E. B. Characterization of bio-oils obtained from pyrolysis of bocaiuva residues. **Renewable Energy**, v. 91, p. 21-31, 2016.

CASONI, A. I. *et al.* Pyrolysis of sunflower seed hulls for obtaining bio-oils. **Bioresource Technology**, v. 177, p. 406-409, 2015.

CHANG, G. *et al.* The lignin pyrolysis composition and pyrolysis products of palm kernel shell, wheat straw, and pine sawdust. **Energy Conversion and Management**, v. 124, p. 587-597, 2016.

CHOUDHURY, N. D. *et al.* Pyrolysis of jute dust: Effect of reaction parameters and analysis of products. **Journal of Material Cycles and Waste Management**, v. 16, p. 449-459, 2014.

CHUTIA, R. S.; KATAKI, R.; BHASKAR, T. Characterization of liquid and solid product from pyrolysis of *Pongamia glabra* deoiled cake. **Bioresource Technology**, v. 165, p. 336-

342, 2014.

DAS, P.; GANESH, A. Bio-oil from pyrolysis of cashew nut shell-a near fuel. **Biomass and Bioenergy**, v. 25, n. 1, p. 113-117, 2003.

DELLA ROCCA, P. A. *et al.* Pyrolysis of hardwoods residues: On kinetics and chars characterization. **Biomass and Bioenergy**, v. 16, n. 1, p. 79-88, 1999.

DEMIRAL, I.; AYAN, E. A. Pyrolysis of grape bagasse: Effect of pyrolysis conditions on the product yields and characterization of the liquid product. **Bioresource Technology**, v. 102, p. 3946-3951, 2011.

DEMIRBAS, A. Effect of temperature on pyrolysis products from four nut shells. **Journal of Analytical and Applied Pyrolysis**, v. 76, p. 285-289, 2006.

DEMIRBAS, A. Effects of temperature and particle size on bio-char yield from pyrolysis of agricultural residues. **Journal of Analytical and Applied Pyrolysis**, v. 72, n. 2, p. 243-248, 2004.

DEMIRBAS, A. The influence of temperature on the yields of compounds existing in bio-oils obtained from biomass samples via pyrolysis. **Fuel Processing Technology**, v. 88, p. 591-597, 2007.

DEMIRBAŞ, A. Analysis of Liquid Products from Biomass via Flash Pyrolysis. **Energy Sources**, v. 24, n. 4, p. 337-345, 2002a.

DEMIRBAŞ, A. Properties of charcoal derived from hazelnut shell and the production of briquettes using pyrolytic oil. **Energy**, v. 24, p. 141-150, 1999.

DEMIRBAŞ, A. Pyrolysis and steam gasification processes of black liquor. **Energy Conversion and Management**, v. 43, n. 7, p. 877-884, 2002b.

DHANAVATH, K. N. *et al.* Optimization of process parameters for slow pyrolysis of neem press seed cake for liquid and char production. **Journal of Environmental Chemical Engineering**, v. 7, 2019.

DI BLASI, C. *et al.* Product distribution from pyrolysis of wood and agricultural residues. **Industrial and Engineering Chemistry Research**, v. 38, p. 2216-2224, 1999.

DI BLASI, C.; BRANCA, C. Kinetics of Primary Product Formation from Wood Pyrolysis. **Industrial & Engineering Chemistry Research**, v. 40, p. 5547-5556, 2001.

DUFOUR, A. *et al.* Synthesis gas production by biomass pyrolysis: Effect of reactor temperature on product distribution. **International Journal of Hydrogen Energy**, v. 34, n. 4, p. 1726-1734, 2009.

DUMAN, G. *et al.* The slow and fast pyrolysis of cherry seed. **Bioresource Technology**, v. 102, p. 1869-1878, 2011.

ELLENS, C. J.; BROWN, R. C. Optimization of a free-fall reactor for the production of fast

pyrolysis bio-oil. **Bioresource Technology**, v. 103, p. 374-380, 2012.

ENCINAR, J. M. *et al.* Pyrolysis/gasification of agricultural residues by carbon dioxide in the presence of different additives: influence of variables. **Fuel Processing Technology**, v. 55, n. 3, p. 219-233, 1998.

ENCINAR, J. M.; GONZÁLEZ, J. F.; GONZÁLEZ, J. Fixed-bed pyrolysis of *Cynara cardunculus* L. Product yields and compositions. **Fuel processing technology**, v. 68, n. 3, p. 209-222, 2000.

FAGBEMI, L.; KHEZAMI, L.; CAPART, R. Pyrolysis products from different biomasses: application to the thermal cracking of tar. **Applied Energy**, v. 69, p. 293-306, 2001.

FAHMI, R. *et al.* The effect of lignin and inorganic species in biomass on pyrolysis oil yields, quality, and stability. **Fuel**, v. 87, p. 1230-1240, 2008.

FIGUEIREDO, J. L. *et al.* Pyrolysis of holm-oak wood: influence of temperature and particle size. **Fuel**, v. 68, p. 1012-1016, 1989.

FIGUEIREDO, M. K. *et al.* Pyrolysis Oil from the Fruit and Cake of *Jatropha curcas* Produced Using a Low Temperature Conversion (LTC) Process: Analysis of a Pyrolysis Oil-Diesel Blend. **Energy and Power Engineering**, v. 3, n. July, p. 332-338, 2011.

FONTS, I. *et al.* Study of the pyrolysis liquids obtained from different sewage sludge. **Journal of Analytical and Applied Pyrolysis**, v. 85, n. 1-2, p. 184-191, 2009.

FRENCH, R.; CZERNIK, S. Catalytic pyrolysis of biomass for biofuel production. **Fuel Processing Technology**, v. 91, n. 1, p. 25-32, 2010.

FU, P. *et al.* Effect of temperature on gas composition and char structural features of pyrolyzed agricultural residues. **Bioresource Technology**, v. 102, n. 17, p. 8211-8219, 2011.

FU, P. *et al.* FTIR study of pyrolysis products evolving from typical agricultural residues. **Journal of Analytical and Applied Pyrolysis**, v. 88, p. 117-123, 2010.

GARCIA-PEREZ, M. *et al.* Fast pyrolysis of oil mallee woody biomass: Effect of temperature on the yield and quality of pyrolysis products. **Industrial and Engineering Chemistry Research**, v. 47, p. 1846-1854, 2008.

GARG, R.; ANAND, N.; KUMAR, D. Pyrolysis of babool seeds (*Acacia nilotica*) in a fixed bed reactor and bio-oil characterization. **Renewable Energy**, v. 96, p. 167-171, 2016.

GERÇEL, H. F. The production and evaluation of bio-oils from the pyrolysis of sunflower-oil cake. **Biomass and Bioenergy**, v. 23, p. 307-314, 2002.

GRAY, M. R.; CORCORAN, W. H.; GAVALAS, G. R. Pyrolysis of a Wood-Derived Material. Effects of Moisture and Ash Content. **Industrial and Engineering Chemistry Process Design and Development**, v. 24, p. 646-651, 1985.

GÜLLÜ, D. Effect of catalyst on yield of liquid products from biomass via pyrolysis. **Energy**

Sources, v. 25, n. 8, p. 753-765, 2003.

GUO, X. *et al.* Properties of bio-oil from fast pyrolysis of rice husk. **Chinese Journal of Chemical Engineering**, v. 19, n. 1, p. 116-121, 2011.

HEIDARI, A. *et al.* Effect of process conditions on product yield and composition of fast pyrolysis of *Eucalyptus grandis* in fluidized bed reactor. **Journal of Industrial and Engineering Chemistry**, v. 20, p. 2594-2602, 2014.

HEO, H. S. *et al.* Bio-oil production from fast pyrolysis of waste furniture sawdust in a fluidized bed. **Bioresource Technology**, v. 101, p. S91-S96, 2010a.

HEO, H. S. *et al.* Fast pyrolysis of rice husk under different reaction conditions. **Journal of Industrial and Engineering Chemistry**, v. 16, n. 1, p. 27-31, 2010b.

HERNANDEZ-MENA, L. E.; PÉCORÀ, A. A. B.; BERÁLDO, A. L. Slow pyrolysis of bamboo biomass: Analysis of biochar properties. **Chemical Engineering Transactions**, v. 37, p. 115-120, 2014.

HORNE, P. A.; WILLIAMS, P.T. Influence of temperature on the products from the flash pyrolysis of biomass. **Fuel**, v. 75, n. 9, p. 1051-1059, 1996.

HSU, C. P.; HUANG, A. N.; KUO, H. P. Analysis of the Rice Husk Pyrolysis Products from a Fluidized Bed Reactor. **Procedia Engineering**, v. 102, p. 1183-1186, 2015.

HUANG, A. N. *et al.* Production and separation of rice husk pyrolysis bio-oils from a fractional distillation column connected to a fluidized bed reactor. **Powder Technology**, v. 323, p. 588-593, 2016.

HUANG, X. *et al.* Influences of pyrolysis conditions in the production and chemical composition of the bio-oils from fast pyrolysis of sewage sludge. **Journal of Analytical and Applied Pyrolysis**, v. 110, p. 353-362, 2014.

IMAM, T.; CAPAREDA, S. Characterization of bio-oil, syn-gas, and bio-char from switchgrass pyrolysis at various temperatures. **Journal of Analytical and Applied Pyrolysis**, v. 93, p. 170-177, 2012.

INGRAM, L. *et al.* Pyrolysis of wood and bark in an auger reactor: Physical properties and chemical analysis of the produced bio-oils. **Energy and Fuels**, v. 22, p. 614-625, 2008.

ISA, K. M. *et al.* Thermogravimetric analysis and the optimization of bio-oil yield from fixed-bed pyrolysis of rice husk using response surface methodology (RSM). **Industrial Crops and Products**, v. 33, n. 2, p. 481-487, 2011.

ISLAM, M. N. *et al.* Pyrolytic oil from fixed bed pyrolysis of municipal solid waste and its characterization. **Renewable Energy**, v. 30, n. 3, p. 413-420, 2005.

ISLAM, M. N.; ZAILANI, R.; NASIR ANI, F. Pyrolytic oil from fluidised bed pyrolysis of oil palm shell and its characterisation. **Renewable Energy**, v. 17, n. 1, p. 73-84, 1999.

ISLAM, M. R.; PARVEEN, M.; HANIU, H. Properties of sugarcane waste-derived bio-oils obtained by fixed-bed fire-tube heating pyrolysis. **Bioresource Technology**, v. 101, n. 11, p. 4162-4168, 2010.

JOURABCHI, S. A.; GAN, S.; NG, H. K. Pyrolysis of *Jatropha curcas* pressed cake for bio-oil production in a fixed-bed system—Energy **Conversion and Management**, v. 78, p. 518-526, 2014.

JUNG, S. H.; KANG, B. S.; KIM, J. S. Production of bio-oil from rice straw and bamboo sawdust under various reaction conditions in a fast pyrolysis plant equipped with a fluidized bed and a char separation system. **Journal of Analytical and Applied Pyrolysis**, v. 82, p. 240-247, 2008.

KANAUJIA, P. K. *et al.* Pyrolysis of *Jatropha Curcas* seed cake followed by optimization of liquid-liquid extraction procedure for the obtained bio-oil. **Journal of Analytical and Applied Pyrolysis**, v. 118, p. 202-224, 2016.

KANG, B. S. *et al.* Fast pyrolysis of radiata pine in a bench scale plant with a fluidized bed: Influence of a char separation system and reaction conditions on the production of bio-oil. **Journal of Analytical and Applied Pyrolysis**, v. 76, n. 1-2, p. 32-37, 2006.

KILIÇ, M.; PÜTÜN, E.; PÜTÜN, A. E. Optimization of *Euphorbia rigida* fast pyrolysis conditions by using response surface methodology. **Journal of Analytical and Applied Pyrolysis**, v. 110, p. 163-171, 2014.

KIM, K. H. *et al.* Influence of pyrolysis temperature on physicochemical properties of biochar obtained from the fast pyrolysis of pitch pine (*Pinus rigida*). **Bioresource Technology**, v. 118, p. 158-162, 2012.

KIM, K. H. *et al.* Investigation of physicochemical properties of biooils produced from yellow poplar wood (*Liriodendron tulipifera*) at various temperatures and residence times. **Journal of Analytical and Applied Pyrolysis**, v. 92, p. 2-9, 2011.

KIM, S. J.; JUNG, S. H.; KIM, J. S. Fast pyrolysis of palm kernel shells: Influence of operation parameters on the bio-oil yield and the yield of phenol and phenolic compounds. **Bioresource Technology**, v. 101, p. 9294-9300, 2010.

KIM, S. W. *et al.* Bio-oil from the pyrolysis of palm and *Jatropha* wastes in a fluidized bed. **Fuel Processing Technology**, v. 108, p. 18-124, 2013.

KIM, S. W. Pyrolysis conditions of biomass in fluidized beds for production of bio-oil compatible with petroleum refinery. **Journal of Analytical and Applied Pyrolysis**, v. 117, p. 220-227, 2016.

KRISHNA, B. B. *et al.* Pyrolysis of *Cedrus deodara* sawmill shavings in hydrogen and nitrogen atmosphere for the production of bio-oil. **Renewable Energy**, v. 98, p. 238-244, 2016.

LAZZARI, E. *et al.* Production and chromatographic characterization of bio-oil from the pyrolysis of mango seed waste. **Industrial Crops and Products**, v. 83, p. 529-536, 2016.

- LEE, Y. *et al.* Characteristics of biochar produced from slow pyrolysis of Geodae-Uksae 1. **Bioresource Technology**, v. 130, p. 345-350, 2013a.
- LEE, Y. *et al.* Comparison of biochar properties from biomass residues produced by slow pyrolysis at 500°C. **Bioresource Technology**, v. 148, p. 196-201, 2013b.
- LI, L.; ZHANG, H.; ZHUANG, X. Pyrolysis of waste paper: Characterization and composition of pyrolysis oil. **Energy Sources**, v. 27, n. 9, p. 867-873, 2005.
- LI, S. *et al.* Fast pyrolysis of biomass in free-fall reactor for hydrogen-rich gas. **Fuel Processing Technology**, v. 85, n. 8-10, p. 1201-1211, 2004.
- LIM, C. H. *et al.* Novel input-output prediction approach for biomass pyrolysis. **Journal of Cleaner Production**, v. 136, p. 51-61, 2016.
- LIU, Z. *et al.* Cellulose-lignin and xylan-lignin interactions on the formation of lignin-derived phenols in pyrolysis oil. **BioResources**, v. 12, n. 3, p. 4958-4971, 2017.
- LY, H. V. *et al.* Fast pyrolysis of macroalga *Saccharina japonica* in a bubbling fluidized-bed reactor for bio-oil production. **Energy**, v. 93, p. 1436-1446, 2015.
- LY, H. V. *et al.* Fast pyrolysis of *Saccharina japonica* alga in a fixed-bed reactor for bio-oil production. **Energy Conversion and Management**, v. 122, p. 526-534, 2016.
- MADDI, B.; VIAMAJALA, S.; VARANASI, S. Comparative study of pyrolysis of algal biomass from natural lake blooms with lignocellulosic biomass. **Bioresource Technology**, v. 102, p. 11018-11026, 2011.
- MAKIBAR, J. *et al.* Performance of a conical spouted bed pilot plant for bio-oil production by poplar flash pyrolysis. **Fuel Processing Technology**, v. 137, p. 283-289, 2015.
- MANDAL, S. *et al.* Optimization of process parameters for bio-oil synthesis from pine needles (*Pinus roxburghii*) using response surface methodology. **Chemical Papers**, v. 72, p. 603-616, 2018.
- MANURUNG, R. *et al.* Valorisation of *Jatropha curcas* L. plant parts: Nut shell conversion to fast pyrolysis oil. **Food and Bioproducts Processing**, v. 87, n. 3, p. 187-196, 2009.
- MENÉNDEZ, J. A.; INGUANZO, M.; PIS, J. J. Microwave-induced pyrolysis of sewage sludge. **Water Research**, v. 36, p. 3261-3264, 2002.
- MIAO, X.; WU, Q. High yield bio-oil production from fast pyrolysis by metabolic controlling of *Chlorella protothecoides*. **Journal of Biotechnology**, v. 110, p. 85-93, 2004.
- MIAO, X.; WU, Q.; YANG, C. Fast pyrolysis of microalgae to produce renewable fuels. **Journal of Analytical and Applied Pyrolysis**, v. 71, n. 2, p. 855-863, 2004.
- MISHRA, R. K.; MOHANTY, K. Thermocatalytic conversion of non-edible Neem seeds towards clean fuel and chemicals. **Journal of Analytical and Applied Pyrolysis**, v.134, p. 83-92, 2018.

MISHRA, R. K.; MURARAKA, A.; MOHANTY, K. Optimization of process parameters and catalytic pyrolysis of Cascabela thevetia seeds over low-cost catalysts towards renewable fuel production. **Journal of the Energy Institute**, v. 93, p. 2033-2043, 2020.

MORALI, U.; ŞENSÖZ, S. Pyrolysis of hornbeam shell (*Carpinus betulus* L.) in a fixed bed reactor: Characterization of bio-oil and bio-char. **Fuel**, v. 150, p. 672-678, 2015.

MORALI, U.; YAVUZEL, N.; ŞENSÖZ, S. Pyrolysis of hornbeam (*Carpinus betulus* L.) sawdust: Characterization of bio-oil and bio-char. **Bioresource Technology**, v. 221, p. 682-685, 2016.

MULLEN, C. A. *et al.* Bio-oil and bio-char production from corn cobs and stover by fast pyrolysis. **Biomass and Bioenergy**, v. 34, n. 1, p. 67-74, 2010.

MULLEN, C. A.; BOATENG, A. A. Chemical Composition of Bio-oils Produced by Fast Pyrolysis of Two Energy Crops. **Energy & Fuels**, v. 22, n. 7, p. 2104-2109, 2008.

MURADOV, N. *et al.* Pyrolysis of fast-growing aquatic biomass - *Lemna minor* (duckweed): Characterization of pyrolysis products. **Bioresource Technology**, v. 101, p. 8424-8428, 2010.

MURATA, K. *et al.* Catalytic fast pyrolysis of jatropha wastes. **Journal of Analytical and Applied Pyrolysis**, v. 94, p. 75-82, 2012.

NAYAN, N. K.; KUMAR, S.; SINGH, R. K. Characterization of the liquid product obtained by pyrolysis of Karanja seed. **Bioresource Technology**, v. 124, p. 186-189, 2012.

NAYAN, N. K.; KUMAR, S.; SINGH, R. K. Production of the liquid fuel by thermal pyrolysis of neem seed. **Fuel**, v. 103, p. 437-443, 2013.

NUNN, T. R. *et al.* Product Compositions and Kinetics in the Rapid Pyrolysis of Milled Wood Lignin. **Industrial and Engineering Chemistry Process Design and Development**, v. 24, p. 844-852, 1985a.

NUNN, T. R. *et al.* Product Compositions and Kinetics in the Rapid Pyrolysis of Sweet Gum Hardwood. **Industrial & Engineering Chemistry Process Design and Development**, v. 24, p. 836-844, 1985b.

OASMAA, A. *et al.* Fast pyrolysis bio-oils from wood and agricultural residues. **Energy and Fuels**, v. 24, p. 1380-1388, 2010.

ONAY, O. Influence of pyrolysis temperature and heating rate on the production of bio-oil and char from safflower seed by pyrolysis, using a well-swept fixed-bed reactor. **Fuel Processing Technology**, v. 88, p. 523-531, 2007.

ONAY, Ö. Production of bio-oil from biomass: Slow pyrolysis of rapeseed (*Brassica napus* L.) in a fixed-bed reactor. **Energy Sources**, v. 25, n. 9, p. 879-892, 2003.

ONAY, Ö.; BEIS, S. H.; KOÇKAR, Ö. M. Fast pyrolysis of rape seed in a well-swept fixed-bed reactor. **Journal of Analytical and Applied Pyrolysis**, v. 58-59, p. 995-1007, 2001.

ONAY, Ö.; BEIS, S. H.; KOÇKAR, Ö. M. Pyrolysis of walnut shell in a well-swept fixed-bed reactor. **Energy Sources**, v. 26, n. 8, p. 771-782, 2004.

ONAY, O.; KOÇKAR, O. M. Fixed-bed pyrolysis of rapeseed (*Brassica napus* L.). **Biomass and Bioenergy**, v. 26, n. 3, p. 289-299, 2004.

ONAY, O.; KOÇKAR, O. M. Slow, fast and flash pyrolysis of rapeseed. **Renewable Energy**, v. 28, n. 15, p. 2417-2433, 2003.

OUDENHOVEN, S. R. G. *et al.* Effect of temperature on the fast pyrolysis of organic-acid leached pinewood; the potential of low-temperature pyrolysis. **Biomass and Bioenergy**, v. 89, p. 78-90, 2016.

ÖZBAY, N. *et al.* Biocrude from biomass: Pyrolysis of cottonseed cake. **Renewable Energy**, v. 24, p. 615-625, 2001.

ÖZBAY, N. *et al.* Characterization of bio-oil obtained from fruit pulp pyrolysis. **Energy**, v. 33, n. 8, p. 1233-1240, 2008.

ÖZÇİMEN, D.; ERSOY-MERİÇBOYU, A. Characterization of biochar and bio-oil samples obtained from the carbonization of various biomass materials. **Renewable Energy**, v. 35, p. 1319-1324, 2010.

ÖZÇİMEN, D.; KARAOSMANOĞLU, F. Production and characterization of bio-oil and biochar from rapeseed cake. **Renewable Energy**, v. 29, p. 779-787, 2004.

PAENPONG, C.; PATTIYA, A. Effect of pyrolysis and moving-bed granular filter temperatures on the yield and properties of bio-oil from fast pyrolysis of biomass. **Journal of Analytical and Applied Pyrolysis**, v. 119, p. 40-51, 2016.

PARK, J. *et al.* Slow pyrolysis of rice straw: Analysis of products properties, carbon and energy yields. **Bioresource Technology**, v. 155, p. 63-70, 2014.

PĂRPĂRIȚĂ, E. *et al.* Pyrolysis behaviors of various biomasses. **Polymer Degradation and Stability**, v. 100, n. 1, p. 1-9, 2014.

PATTIYA, A. Bio-oil production via fast pyrolysis of biomass residues from cassava plants in a fluidized-bed reactor. **Bioresource Technology**, v. 102, p. 1959-1967, 2011.

PATTIYA, A.; SUKKASI, S.; GOODWIN, V. Fast pyrolysis of sugarcane and cassava residues in a free-fall reactor. **Energy**, v. 44, p. 1067-1077, 2012.

PATTIYA, A.; SUTTIBAK, S. Production of bio-oil via fast pyrolysis of agricultural residues from cassava plantations in a fluidised-bed reactor with a hot vapour filtration unit. **Journal of Analytical and Applied Pyrolysis**, v. 95, p. 227-235, 2012.

PÜTÜN, A. E. *et al.* Bio-oil from olive oil industry wastes: Pyrolysis of olive residue under different conditions. **Fuel Processing Technology**, v. 87, n. 1, p. 25-32, 2005.

PÜTÜN, A. E. Biomass to bio-oil via fast pyrolysis of cotton straw and stalk. **Energy**

Sources, v. 24, n. 3, p. 275-285, 2002.

PÜTÜN, A. E. *et al.* Comparison between the “slow” and “fast” pyrolysis of tobacco residue. **Industrial Crops and Products**, v. 26, p. 307-314, 2007a.

PÜTÜN, A. E. *et al.* Production of biocrudes from biomass in a fixed-bed tubular reactor: Product yields and compositions. **Fuel**, v. 80, p. 1371-1378, 2001.

PÜTÜN, A. E. *et al.* Rapid and slow pyrolysis of pistachio shell: effect of pyrolysis conditions on the product yields and characterization of the liquid product. **International Journal of Energy Research**, v. 31, n. 5, p. 506-514, 2007b.

PÜTÜN, A. E.; APAYDM, E.; PÜTÜN, E. Bio-oil production from pyrolysis and steam pyrolysis of soybean-cake: product yields and composition—**Energy**, v. 27, n. 7, p. 703-713, 2002.

PÜTÜN, A. E.; APAYDM, E.; PÜTÜN, E. Rice straw as a bio-oil source via pyrolysis and steam pyrolysis. **Energy**, v. 29, p. 2171-2180, 2004.

PÜTÜN, A. E.; ÖZEAN, A.; PÜTÜN, E. Pyrolysis of hazelnut shells in a fixed-bed tubular reactor: Yields and structural analysis of bio-oil. **Journal of Analytical and Applied Pyrolysis**, v. 52, p. 33-49, 1999.

QIANG, L.; XU LAI, Y.; XI FENG, Z. Analysis on chemical and physical properties of bio-oil pyrolyzed from rice husk. **Journal of Analytical and Applied Pyrolysis**, v. 82, p. 191-198, 2008.

QUAN, C.; GAO, N.; SONG, Q. Pyrolysis of biomass components in a TGA and a fixed-bed reactor: Thermochemical behaviors, kinetics, and product characterization. **Journal of Analytical and Applied Pyrolysis**, v. 121, p. 84-92, 2016.

RAJA, S. A. *et al.* Flash pyrolysis of jatropha oil cake in electrically heated fluidized bed reactor. **Energy**, v. 35, p. 2819-2823, 2010.

RAVEENDRAN, K; GANESH, A. Heating value of biomass and biomass pyrolysis products. v. 75, n. 15, p. 1715-1720, 1996.

RAVEENDRAN, K.; GANESH, A.; KHILAR, K. C. Influence of mineral matter on biomass pyrolysis characteristics. **Fuel**, v. 74, n. 12, p. 1812-1822, 1995.

RAVEENDRAN, K.; GANESH, A.; KHILAR, K. C. Pyrolysis characteristics of biomass and biomass components. **Fuel**, v. 75, n. 8, p. 987-998, 1996.

RAZUAN, R. *et al.* Pyrolysis and combustion of oil palm stone and palm kernel cake in fixed-bed reactors. **Bioresource Technology**, v. 101, n. 12, p. 4622-4629, 2010.

REZA, M. S. *et al.* Acacia holosericea: An invasive species for biochar, bio-oil, and biogas production. **Bioengineering**, v. 6, n. 2, 2019.

SÁNCHEZ, M. E. *et al.* Pyrolysis of agricultural residues from rape and sunflowers:

Production and characterization of bio-fuels and biochar soil management. **Journal of Analytical and Applied Pyrolysis**, v. 85, n. 1-2, p. 142-144, 2009.

SCHRÖDER, E. Experiments on the pyrolysis of large beechwood particles in fixed beds. **Journal of Analytical and Applied Pyrolysis**, v. 71, p. 669-694, 2004.

SCOTT, D. S. *et al.* The Role of Temperature in the Fast Pyrolysis of Cellulose and Wood. **Industrial and Engineering Chemistry Research**, v. 27, p. 8-15, 1988.

SCOTT, D. S.; PLSKORZ, J.; RADLELN, D. Liquid Products from the Continuous Flash Pyrolysis of Biomass. **Industrial and Engineering Chemistry Process Design and Development**, v. 24, p. 581-588, 1985.

ŞENSÖZ, S. Slow pyrolysis of wood barks from *Pinus brutia* Ten. and product compositions. **Bioresource Technology**, v. 89, p. 307-311, 2003.

ŞENSÖZ, S.; ANGIN, D. Pyrolysis of safflower (*Charthamus tinctorius* L.) seed press cake: Part 1. The effects of pyrolysis parameters on the product yields. **Bioresource Technology**, v. 99, n. 13, p. 5492-5497, 2008a.

ŞENSÖZ, S.; ANGIN, D. Pyrolysis of safflower (*Charthamus tinctorius* L.) seed press cake in a fixed-bed reactor: Part 2. Structural characterization of pyrolysis bio-oils. **Bioresource Technology**, v. 99, p. 5498-5504, 2008b.

ŞENSÖZ, S.; ANGIN, D.; YORGUN, S. Influence of particle size on the pyrolysis of rapeseed (*Brassica napus* L.): fuel properties of bio-oil. **Biomass and Bioenergy**, v. 19, n. 4, p. 271-279, 2000.

ŞENSÖZ, S.; CAN, M. Pyrolysis of Pine(*Pinus Brutia* Ten.) Chips: 1. Effect of Pyrolysis Temperature and Heating Rate on the Product Yields. **Energy Sources**, v. 24, n. 4, p. 347-355, 2002a.

ŞENSÖZ, S.; CAN, M. Pyrolysis of pine (*Pinus brutia* Ten.) chips: 2. Structural analysis of bio-oil. **Energy Sources**, v. 24, n. 4, p. 357-364, 2002b.

ŞENSÖZ, S.; DEMIRAL, I.; GERÇEL, H. F. Olive bagasse (*Olea europea* L.) pyrolysis. **Bioresource Technology**, v. 97, n. 3, p. 429-436, 2006.

ŞENSÖZ, S.; KAYNAR, I. Bio-oil production from soybean (*Glycine max* L.); Fuel properties of Bio-oil. **Industrial Crops and Products**, v. 23, n. 1, p. 99-105, 2006.

SHARMA, R. K. *et al.* Characterization of chars from pyrolysis of lignin. **Fuel**, v. 83, p. 1469-1482, 2004.

SHARMA, R.; SHETH, P. N. Thermo-Chemical Conversion of *Jatropha* Deoiled Cake: Pyrolysis vs. Gasification. **International Journal of Chemical Engineering and Applications**, v. 6, n. 5, p. 376-380, 2015.

SHEN, D. K. *et al.* The pyrolytic degradation of wood-derived lignin from pulping process. **Bioresource Technology**, v. 101, p. 6136-6146, 2010.

SHEN, J. *et al.* Effects of particle size on the fast pyrolysis of oil mallee woody biomass. **Fuel**, v. 88, n. 10, p. 1810-1817, 2009.

SINGH, R. K.; SHADANGI, K. P. Liquid fuel from castor seeds by pyrolysis. **Fuel**, v. 90, n. 7, p. 2538-2544, 2011.

SINGH, S.; CHAKRABORTY, J. P.; MONDAL, M. K. Pyrolysis of torrefied biomass: Optimization of process parameters using response surface methodology, characterization, and comparison of properties of pyrolysis oil from raw biomass. **Journal of Cleaner Production**, v. 272, 2020.

SONG, Y.; TAHMASEBI, A.; YU, J. Co-pyrolysis of pine sawdust and lignite in a thermogravimetric analyzer and a fixed-bed reactor. **Bioresource Technology**, v. 174, p. 204-211, 2014.

SRICHAROENCHAIKUL, V. *et al.* Preparation and characterization of activated carbon from the pyrolysis of physic nut (*Jatropha curcas* L.) waste. **Energy and Fuels**, v. 22, p. 31-37, 2008.

SRICHAROENCHAIKUL, V.; ATONG, D. Thermal decomposition study on *Jatropha curcas* L. waste using TGA and fixed bed reactor. **Journal of Analytical and Applied Pyrolysis**, v. 85, p. 155-162, 2009.

STEFANIDIS, S. D. *et al.* A study of lignocellulosic biomass pyrolysis via the pyrolysis of cellulose, hemicellulose and lignin. **Journal of Analytical and Applied Pyrolysis**, v. 105, p. 143-150, 2014.

SULAIMAN, F.; ABDULLAH, N. Optimum conditions for maximizing pyrolysis liquids of oil palm empty fruit bunches. **Energy**, v. 36, p. 2352-2359, 2011.

SUTTIBAK, S.; SRIPRATEEP, K.; PATTIYA, A. Production of bio-oil via fast pyrolysis of cassava rhizome in a fluidized-bed reactor. **Energy Procedia**, v. 14, p. 668-673, 2012.

THANGALAZHY-GOPAKUMAR, S. *et al.* Utilization of palm oil sludge through pyrolysis for bio-oil and bio-char production. **Bioresource Technology**, v. 178, p. 65-69, 2015.

THUNMAN, H. *et al.* Composition of volatile gases and thermochemical properties of wood for modeling of fixed or fluidized beds. **Energy and Fuels**, v. 15, p. 1488-1497, 2001.

TSAI, W. T.; LEE, M. K.; CHANG, Y. M. Fast pyrolysis of rice husk: Product yields and compositions. **Bioresource Technology**, v. 98, n. 1, p. 22-28, 2007.

TSAI, W. T.; LEE, M. K.; CHANG, Y. M. Fast pyrolysis of rice straw, sugarcane bagasse and coconut shell in an induction-heating reactor. **Journal of Analytical and Applied Pyrolysis**, v. 76, n. 1-2, p. 230-237, 2006.

TUNCEL, F.; GERÇEL, H. F. Production and characterization of pyrolysis oils from *Euphorbia macroclada*. **Energy Sources**, v. 26, n. 8, p. 761-770, 2004.

UÇAR, S.; KARAGÖZ, S. The slow pyrolysis of pomegranate seeds: The effect of

temperature on the product yields and bio-oil properties. **Journal of Analytical and Applied Pyrolysis**, v. 84, p. 151-156, 2009.

UCAR, S.; OZKAN, A. R. Characterization of products from the pyrolysis of rapeseed oil cake. **Bioresource Technology**, v. 99, n. 18, p. 8771-8776, 2008.

UZUN, B. B.; PÜTÜN, A. E.; PÜTÜN, E. Fast pyrolysis of soybean cake: Product yields and compositions. **Bioresource Technology**, v. 97, n. 4, p. 569-576, 2006.

VALENZUELA-CALAHORRO, C. *et al.* Influence of particle size and pyrolysis conditions on yield, density, and some textural parameters of chars prepared from holm-oak wood. **Journal of Analytical and Applied Pyrolysis**, v. 12, p. 61-70, 1987.

VARMA, A. K. *et al.* Pyrolysis of wood sawdust: Effects of process parameters on products yield and characterization of products. **Waste Management**, v. 89, p. 224-235, 2019.

VARMA, A. K.; MONDAL, P. Pyrolysis of sugarcane bagasse in semi-batch reactor: Effects of process parameters on product yields and characterization of products. **Industrial Crops and Products**, v. 95, p. 704-717, 2017.

VIEIRA, F. R. *et al.* Optimization of slow pyrolysis process parameters using a fixed bed reactor for biochar yield from rice husk. **Biomass and Bioenergy**, v. 132, 2020.

VOLLI, V.; SINGH, R. K. Production of bio-oil from de-oiled cakes by thermal pyrolysis. **Fuel**, v. 96, p. 579-585, 2012.

WANG, S. *et al.* Influence of the interaction of components on the pyrolysis behavior of biomass. **Journal of Analytical and Applied Pyrolysis**, v. 91, p. 183-189, 2011.

WANG, X. *et al.* Biomass Pyrolysis in a Fluidized Bed Reactor. Part 2: Experimental Validation of Model Results. **Ind. Eng. Chem. Res.**, v. 44, p. 8786-8795, 2005.

WANG, X. *et al.* Energy and exergy analysis of rice husk high-temperature pyrolysis. **International Journal of Hydrogen Energy**, v. 41, p. 21121-21130, 2016.

WILLIAMS, P. T.; BESLER, S. The influence of temperature and heating rate on the slow pyrolysis of biomass. **Renewable Energy**, v. 7, n. 3, p. 233-250, 1996.

WILLIAMS, P. T.; NUGRANAD, N. Comparison of products from the pyrolysis and catalytic pyrolysis of rice husks. **Energy**, v. 25, p. 493-513, 2000.

WORASUWANNARAK, N.; SONOBE, T.; TANTHAPANICHAKOON, W. Pyrolysis behaviors of rice straw, rice husk, and corncob by TG-MS technique. **Journal of Analytical and Applied Pyrolysis**, v. 78, n. 2, p. 265-271, 2007.

WU, S. R. *et al.* Comparison of oil-tea shell and Douglas-fir sawdust for the production of bio-oils and chars in a fluidized-bed fast pyrolysis system. **Fuel**, v. 175, p. 57-63, 2016.

XU, R. *et al.* Bio-oil production by flash pyrolysis of sugarcane residues and post treatments of the aqueous phase. **Journal of Analytical and Applied Pyrolysis**, v. 91, n. 1, p. 263-272,

2011.

XU, R. *et al.* Flash pyrolysis of grape residues into biofuel in a bubbling fluid bed. **Journal of Analytical and Applied Pyrolysis**, v. 86, n. 1, p. 58-65, 2009.

YANIK, J. *et al.* Fast pyrolysis of agricultural wastes: Characterization of pyrolysis products. **Fuel Processing Technology**, v. 88, n. 10, p. 942-947, 2007.

YIN, R. *et al.* Characterization of bio-oil and bio-char obtained from sweet sorghum bagasse fast pyrolysis with fractional condensers. **Fuel**, v. 112, p. 96-104, 2013.

YU, Q. *et al.* Temperature impact on the formation of tar from biomass pyrolysis in a free-fall reactor. **Journal of Analytical and Applied Pyrolysis**, v. 40-41, p. 481-489, 1997.

ZANZI, R.; SJÖSTRÖM, K.; BJÖRNBOM, E. Rapid high-temperature pyrolysis of biomass in a free-fall reactor. **Fuel**, v. 75, n. 5, p. 545-550, 1996.

ZANZI, R.; SJÖSTRÖM, K.; BJÖRNBOM, E. Rapid pyrolysis of agricultural residues at high temperature. **Biomass and Bioenergy**, v. 23, p. 357-366, 2002.

ZHAI, M. *et al.* Characteristics of rice husk tar pyrolysis by external flue gas. **International Journal of Hydrogen Energy**, v. 40, n. 34, p. 10780-10787, 2015.

ZHANG, H. *et al.* Comparison of non-catalytic and catalytic fast pyrolysis of corncob in a fluidized bed reactor. **Bioresource Technology**, v. 100, p. 1428-1434, 2009.

ZHANG, S. *et al.* High-quality syngas production from microwave pyrolysis of rice husk with char-supported metallic catalysts. **Bioresource Technology**, v. 191, p. 17-23, 2015.

ZHAO, N.; LI, B. The effect of sodium chloride on the pyrolysis of rice husk. **Applied Energy**, v. 178, p. 346-352, 2016.

ZHENG, J. Bio-oil from fast pyrolysis of rice husk: Yields and related properties and improvement of the pyrolysis system. **Journal of Analytical and Applied Pyrolysis**, v. 80, n. 1, p. 30-35, 2007.

ZHENG, J. *et al.* Thermal conversion of rice husks and sawdust to liquid fuel. **Waste Management**, v. 26, p. 1430-1435, 2006.

ZHENG, J.; YI, W.; WANG, N. Bio-oil production from cotton stalk. **Energy Conversion and Management**, v. 49, n. 6, p. 1724-1730, 2008.

ANEXO C – Valores de lambda da transformação Box-Cox para os casos de I a VIII

Variáveis	Caso I	Caso II	Caso III	Caso IV	Caso V	Caso VI	Caso VII	Caso VIII
Temperatura	0,9	0,9	0,3	0,4	0,7	1	-0,6	0,2
Tamanho de partícula	0,0	0	-1	-0,9	-0,3	-0,5	-0,5	-0,5
Residência do vapor	-0,3	-0,3	-0,2	-0,3	-0,1	-0,1	-0,6	-0,6
Taxa de aquecimento	-0,1	0	-0,4	-0,4	-	-	-	-
Tempo de reação	0,3	0,2	-0,2	-0,2	-	-	-	-
Cinza	-	0,4	-	0,3	-	0,3	-	0,1
Umidade	-	0,8	-	1,1	-	1,1	-	-
Voláteis	-	2	-	-	-	2	-	-0,9
Carbono fixo	-	0	-	0,1	-	0,9	-	0,7
Carbono	-	-1,6	-	-1	-	-1,7	-	-
Hidrogênio	-	-0,3	-	-1,5	-	-0,9	-	-
Oxigênio	-	-	-	-	-	-	-	-
Nitrogênio	-	0,2	-	0,6	-	0,1	-	0
Celulose	-	-	-	-1,1	-	-	0,4	0,9
Lignina	-	-	0,6	1	-	-	-0,7	-
Hemicelulose	-	-	0,3	-0,4	-	-	1,5	-

Fonte: O autor, 2023.

ANEXO D – Resultado do lazyprediction para os dados de teste dos casos de I a VIII

✓ **Estudo de Caso I**

Modelo	R ²	RMSE (%)	Tempo de corrida (s)
RandomForestRegressor	0,71	7,20	0,23
GradientBoostingRegressor	0,71	7,27	0,20
HistGradientBoostingRegressor	0,68	7,57	0,14
LGBMRegressor	0,67	7,71	0,13
XGBRegressor	0,65	7,91	0,14
BaggingRegressor	0,63	8,11	0,04
ExtraTreeRegressor	0,63	8,14	0,01
ExtraTreesRegressor	0,62	8,27	0,16
AdaBoostRegressor	0,60	8,53	0,11
DecisionTreeRegressor	0,58	8,66	0,01
KNeighborsRegressor	0,54	9,06	0,02
NuSVR	0,24	11,70	0,02
OrthogonalMatchingPursuitCV	0,23	11,74	0,03
LassoCV	0,20	12,00	0,06
LarsCV	0,20	12,01	0,05
LassoLarsCV	0,20	12,01	0,02
LassoLarsIC	0,20	12,01	0,02
SVR	0,20	12,02	0,01
SGDRegressor	0,19	12,07	0,01
TransformedTargetRegressor	0,19	12,07	0,01
LinearRegression	0,19	12,07	0,01
Lars	0,19	12,07	0,03
Ridge	0,19	12,07	0,01
RidgeCV	0,19	12,08	0,01
ElasticNetCV	0,19	12,08	0,07
OrthogonalMatchingPursuit	0,19	12,11	0,02
PoissonRegressor	0,18	12,12	0,03
BayesianRidge	0,18	12,13	0,04
LassoLars	0,18	12,18	0,01
Lasso	0,18	12,18	0,03
LinearSVR	0,16	12,33	0,01
ElasticNet	0,14	12,42	0,02
GammaRegressor	0,13	12,55	0,03
HuberRegressor	0,12	12,55	0,02
TweedieRegressor	0,12	12,56	0,01
GeneralizedLinearRegressor	0,12	12,56	0,01
DummyRegressor	-	13,67	0,01

Modelo	R ²	RMSE (%)	Tempo de corrida (s)
RANSACRegressor	-	13,97	0,14
PassiveAggressiveRegressor	-	16,13	0,01
MLPRegressor	-	20,86	0,38
KernelRidge	-	47,89	0,03
GaussianProcessRegressor	-	147,87	0,04

Fonte: O autor, 2023.

✓ Estudo de Caso II

Modelo	R ²	RMSE (%)	Tempo de corrida (s)
ExtraTreesRegressor	0,94	2,86	0,18
XGBRegressor	0,94	3,04	0,23
BaggingRegressor	0,93	3,11	0,04
LGBMRegressor	0,93	3,23	0,14
GradientBoostingRegressor	0,93	3,27	0,22
RandomForestRegressor	0,92	3,51	0,27
HistGradientBoostingRegressor	0,91	3,56	0,14
DecisionTreeRegressor	0,88	4,14	0,01
ExtraTreeRegressor	0,88	4,23	0,01
GaussianProcessRegressor	0,81	5,27	0,03
AdaBoostRegressor	0,78	5,71	0,12
KNeighborsRegressor	0,76	6,00	0,02
SVR	0,43	9,15	0,01
NuSVR	0,37	9,64	0,02
ElasticNetCV	0,36	9,73	0,08
Ridge	0,36	9,73	0,01
SGDRegressor	0,36	9,74	0,02
LassoCV	0,36	9,75	0,08
RidgeCV	0,36	9,76	0,01
LassoLarsCV	0,35	9,76	0,03
LassoLarsIC	0,35	9,76	0,02
TransformedTargetRegressor	0,35	9,76	0,01
LinearRegression	0,35	9,76	0,01
HuberRegressor	0,35	9,82	0,02
BayesianRidge	0,34	9,84	0,04
OrthogonalMatchingPursuitCV	0,34	9,84	0,02
PoissonRegressor	0,34	9,85	0,05
LarsCV	0,32	10,00	0,07
Lasso	0,30	10,16	0,02
LassoLars	0,30	10,16	0,01
LinearSVR	0,29	10,22	0,01
ElasticNet	0,28	10,29	0,01

Modelo	R ²	RMSE (%)	Tempo de corrida (s)
TweedieRegressor	0,28	10,34	0,02
GammaRegressor	0,27	10,39	0,02
OrthogonalMatchingPursuit	0,15	11,22	0,01
RANSACRegressor	-	12,23	0,17
DummyRegressor	-	12,77	0,01
MLPRegressor	-	13,35	0,37
PassiveAggressiveRegressor	-	14,08	0,01
KernelRidge	-	52,66	0,02
Lars	-	145,12	0,03

Fonte: O autor, 2023.

✓ Estudo de Caso III

Modelo	R ²	RMSE (%)	Tempo de corrida (s)
GradientBoostingRegressor	0,97	1,38	0,10
XGBRegressor	0,97	1,44	0,09
ExtraTreesRegressor	0,95	1,75	0,12
RandomForestRegressor	0,95	1,75	0,18
DecisionTreeRegressor	0,95	1,90	0,01
BaggingRegressor	0,95	1,91	0,03
GaussianProcessRegressor	0,94	2,09	0,02
KNeighborsRegressor	0,90	2,57	0,01
AdaBoostRegressor	0,90	2,62	0,09
HuberRegressor	0,88	2,88	0,02
RANSACRegressor	0,87	2,94	0,10
ExtraTreeRegressor	0,86	3,03	0,01
HistGradientBoostingRegressor	0,86	3,09	0,08
PassiveAggressiveRegressor	0,81	3,59	0,01
LGBMRegressor	0,78	3,90	0,13
LinearRegression	0,75	4,08	0,01
TransformedTargetRegressor	0,75	4,08	0,02
LassoLarsIC	0,75	4,08	0,02
Lars	0,75	4,08	0,01
SGDRegressor	0,75	4,09	0,01
PoissonRegressor	0,75	4,10	0,01
Ridge	0,75	4,14	0,01
RidgeCV	0,75	4,14	0,01
LarsCV	0,75	4,15	0,03
LassoLarsCV	0,75	4,15	0,02
LassoCV	0,75	4,15	0,08
ElasticNetCV	0,74	4,19	0,07
BayesianRidge	0,73	4,29	0,04

Modelo	R ²	RMSE (%)	Tempo de corrida (s)
OrthogonalMatchingPursuitCV	0,68	4,65	0,01
SVR	0,63	5,02	0,01
Lasso	0,50	5,81	0,01
LassoLars	0,50	5,81	0,01
NuSVR	0,43	6,21	0,01
ElasticNet	0,37	6,52	0,01
TweedieRegressor	0,32	6,79	0,02
GammaRegressor	0,28	7,00	0,02
OrthogonalMatchingPursuit	-	9,35	0,01
DummyRegressor	-	9,50	0,01
LinearSVR	-	9,76	0,01
MLPRegressor	-	37,33	0,11
KernelRidge	-	46,57	0,02

Fonte: O autor, 2023.

✓ Estudo de Caso IV

Modelo	R ²	RMSE (%)	Tempo de corrida (s)
GaussianProcessRegressor	0,98	1,21	0,02
ExtraTreesRegressor	0,97	1,26	0,12
GradientBoostingRegressor	0,95	1,72	0,10
DecisionTreeRegressor	0,95	1,82	0,02
ExtraTreeRegressor	0,95	1,82	0,01
BaggingRegressor	0,92	2,13	0,05
AdaBoostRegressor	0,90	2,43	0,14
RandomForestRegressor	0,88	2,74	0,16
XGBRegressor	0,85	2,98	0,09
HistGradientBoostingRegressor	0,80	3,46	0,07
TransformedTargetRegressor	0,80	3,46	0,01
LinearRegression	0,80	3,46	0,01
LassoLarsCV	0,80	3,50	0,03
HuberRegressor	0,80	3,51	0,04
LassoCV	0,79	3,52	0,08
SGDRegressor	0,76	3,78	0,01
ElasticNetCV	0,76	3,78	0,12
LassoLarsIC	0,76	3,79	0,02
Ridge	0,73	4,01	0,01
BayesianRidge	0,71	4,18	0,05
PoissonRegressor	0,68	4,36	0,02
LGBMRegressor	0,66	4,51	0,12
SVR	0,58	5,01	0,01
LarsCV	0,55	5,21	0,02

Modelo	R ²	RMSE (%)	Tempo de corrida (s)
Lasso	0,53	5,30	0,01
LassoLars	0,53	5,30	0,01
OrthogonalMatchingPursuitCV	0,53	5,30	0,02
RidgeCV	0,52	5,41	0,01
PassiveAggressiveRegressor	0,50	5,51	0,01
OrthogonalMatchingPursuit	0,45	5,75	0,01
KNeighborsRegressor	0,44	5,81	0,02
ElasticNet	0,37	6,18	0,01
TweedieRegressor	0,30	6,52	0,01
NuSVR	0,26	6,68	0,01
GammaRegressor	0,19	6,99	0,02
DummyRegressor	-	10,04	0,01
LinearSVR	-	11,00	0,01
Lars	-	15,05	0,01
MLPRegressor	-	29,35	0,10
KernelRidge	-	46,72	0,01
RANSACRegressor	-	154,21	0,06

Fonte: O autor, 2023.

✓ Estudo de Caso V

Modelo	R ²	RMSE (%)	Tempo de corrida (s)
BaggingRegressor	0,48	7,28	0,03
ExtraTreesRegressor	0,47	7,30	0,13
XGBRegressor	0,46	7,34	0,09
RandomForestRegressor	0,45	7,37	0,18
GradientBoostingRegressor	0,39	7,75	0,15
KNeighborsRegressor	0,34	8,08	0,01
HistGradientBoostingRegressor	0,32	8,20	0,13
DecisionTreeRegressor	0,30	8,34	0,01
LGBMRegressor	0,26	8,54	0,13
ExtraTreeRegressor	0,25	8,59	0,01
AdaBoostRegressor	0,15	9,15	0,05
NuSVR	0,13	9,26	0,03
SVR	0,13	9,27	0,01
SGDRegressor	0,06	9,64	0,01
PoissonRegressor	0,06	9,65	0,04
Lars	0,06	9,66	0,03
TransformedTargetRegressor	0,06	9,66	0,01
LinearRegression	0,06	9,66	0,01
Ridge	0,06	9,66	0,01
RidgeCV	0,06	9,66	0,01

Modelo	R ²	RMSE (%)	Tempo de corrida (s)
OrthogonalMatchingPursuitCV	0,05	9,68	0,02
LassoLarsIC	0,05	9,68	0,02
LarsCV	0,05	9,70	0,05
LassoLarsCV	0,05	9,70	0,02
LassoCV	0,05	9,72	0,06
BayesianRidge	0,05	9,72	0,04
ElasticNetCV	0,04	9,73	0,06
HuberRegressor	0,04	9,76	0,01
GammaRegressor	0,04	9,76	0,02
TweedieRegressor	0,04	9,77	0,01
ElasticNet	0,03	9,82	0,01
OrthogonalMatchingPursuit	0,02	9,84	0,01
LassoLars	0,02	9,87	0,01
Lasso	0,02	9,87	0,03
DummyRegressor	0,00	9,96	0,01
LinearSVR	-	10,03	0,01
PassiveAggressiveRegressor	-	11,26	0,02
RANSACRegressor	-	11,70	0,09
GaussianProcessRegressor	-	12,17	0,03
MLPRegressor	-	23,78	0,34
KernelRidge	-	53,90	0,03

Fonte: O autor, 2023.

✓ Estudo de Caso VI

Modelo	R ²	RMSE (%)	Tempo de corrida (s)
LGBMRegressor	0,81	4,10	0,13
HistGradientBoostingRegressor	0,78	4,38	0,11
KNeighborsRegressor	0,74	4,74	0,01
RandomForestRegressor	0,70	5,09	0,22
GradientBoostingRegressor	0,65	5,50	0,18
BaggingRegressor	0,64	5,60	0,05
AdaBoostRegressor	0,63	5,67	0,22
HuberRegressor	0,54	6,29	0,01
ExtraTreesRegressor	0,45	6,89	0,14
XGBRegressor	0,43	7,04	0,16
Lars	0,42	7,06	0,03
LinearRegression	0,42	7,06	0,01
TransformedTargetRegressor	0,42	7,06	0,01
RidgeCV	0,42	7,06	0,01
ElasticNetCV	0,42	7,08	0,15
LassoCV	0,42	7,08	0,07

Modelo	R ²	RMSE (%)	Tempo de corrida (s)
Ridge	0,42	7,09	0,01
LassoLarsCV	0,42	7,11	0,03
LassoLarsIC	0,41	7,16	0,02
SGDRegressor	0,41	7,16	0,02
PoissonRegressor	0,38	7,31	0,02
BayesianRidge	0,36	7,41	0,07
DecisionTreeRegressor	0,36	7,43	0,02
ExtraTreeRegressor	0,35	7,49	0,01
SVR	0,32	7,66	0,01
OrthogonalMatchingPursuitCV	0,27	7,96	0,02
NuSVR	0,24	8,10	0,03
LassoLars	0,18	8,40	0,01
Lasso	0,18	8,40	0,02
LarsCV	0,18	8,42	0,06
LinearSVR	0,14	8,59	0,01
ElasticNet	0,14	8,61	0,04
TweedieRegressor	0,12	8,72	0,01
GammaRegressor	0,10	8,84	0,02
PassiveAggressiveRegressor	-	10,10	0,01
DummyRegressor	-	10,30	0,02
OrthogonalMatchingPursuit	-	10,44	0,01
RANSACRegressor	-	17,69	0,14
MLPRegressor	-	35,97	0,22
KernelRidge	-	54,80	0,02
GaussianProcessRegressor	-	83,85	0,02

Fonte: O autor, 2023.

✓ Estudo de Caso VII

Modelo	R ²	RMSE (%)	Tempo de corrida (s)
RandomForestRegressor	0,76	5,45	0,15
GradientBoostingRegressor	0,75	5,60	0,09
ExtraTreesRegressor	0,72	5,89	0,12
BaggingRegressor	0,71	6,01	0,03
HistGradientBoostingRegressor	0,69	6,25	0,08
DecisionTreeRegressor	0,68	6,29	0,01
XGBRegressor	0,66	6,52	0,09
LGBMRegressor	0,65	6,58	0,11
AdaBoostRegressor	0,58	7,26	0,08
KNeighborsRegressor	0,53	7,68	0,01
ExtraTreeRegressor	0,52	7,70	0,01
SVR	0,20	9,97	0,01

Modelo	R ²	RMSE (%)	Tempo de corrida (s)
NuSVR	0,13	10,42	0,01
OrthogonalMatchingPursuit	0,04	10,92	0,01
TweedieRegressor	0,02	11,05	0,01
GammaRegressor	0,01	11,08	0,02
ElasticNet	0,01	11,11	0,01
DummyRegressor	0,00	11,15	0,01
BayesianRidge	0,00	11,16	0,03
LassoLars	-	11,19	0,01
Lasso	-	11,19	0,01
ElasticNetCV	-	11,26	0,06
RANSACRegressor	-	11,42	0,14
RidgeCV	-	11,44	0,01
LassoCV	-	11,45	0,07
LarsCV	-	11,45	0,02
LassoLarsCV	-	11,45	0,02
LassoLarsIC	-	11,53	0,02
SGDRegressor	-	11,56	0,01
OrthogonalMatchingPursuitCV	-	11,58	0,01
Ridge	-	11,62	0,01
LinearRegression	-	11,65	0,01
TransformedTargetRegressor	-	11,65	0,01
Lars	-	11,65	0,01
PoissonRegressor	-	11,67	0,01
HuberRegressor	-	11,74	0,02
PassiveAggressiveRegressor	-	11,97	0,01
LinearSVR	-	14,89	0,01
MLPRegressor	-	44,41	0,10
KernelRidge	-	57,85	0,01
GaussianProcessRegressor	-	90,47	0,01

Fonte: O autor, 2023.

✓ Estudo de Caso VIII

Modelo	R ²	RMSE (%)	Tempo de corrida (s)
GradientBoostingRegressor	0,96	2,24	0,18
ExtraTreeRegressor	0,94	2,56	0,01
DecisionTreeRegressor	0,93	2,87	0,01
BaggingRegressor	0,92	3,09	0,04
XGBRegressor	0,92	3,09	0,07
RandomForestRegressor	0,89	3,51	0,14
ExtraTreesRegressor	0,89	3,57	0,19
AdaBoostRegressor	0,84	4,38	0,14

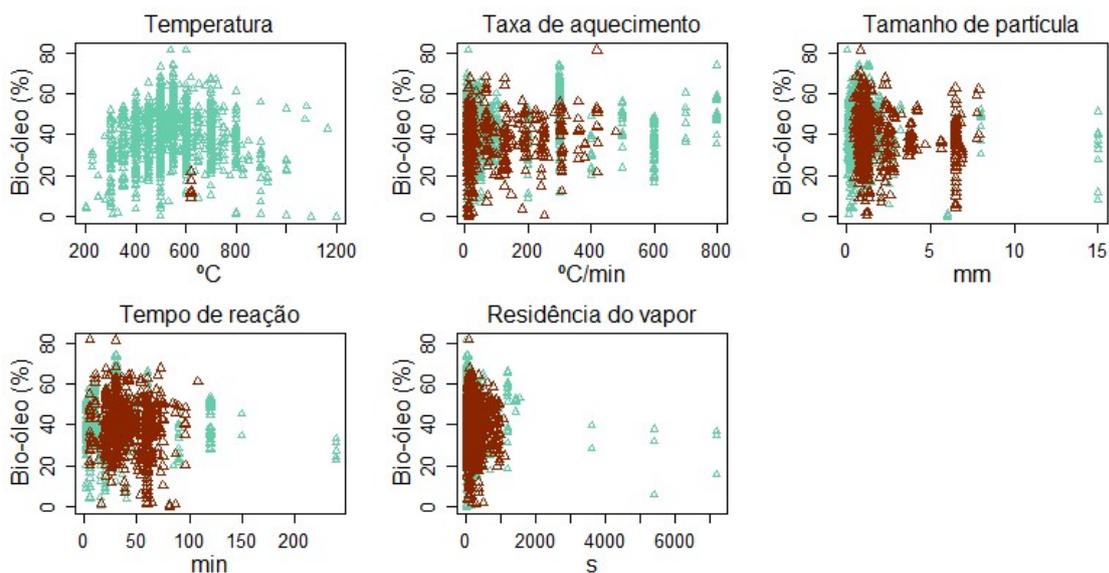
Modelo	R ²	RMSE (%)	Tempo de corrida (s)
HuberRegressor	0,83	4,45	0,03
TransformedTargetRegressor	0,83	4,49	0,01
LinearRegression	0,83	4,49	0,01
LassoCV	0,82	4,54	0,09
LassoLarsCV	0,82	4,54	0,02
SGDRegressor	0,82	4,59	0,01
RidgeCV	0,82	4,62	0,01
Ridge	0,82	4,62	0,01
PoissonRegressor	0,82	4,63	0,01
ElasticNetCV	0,80	4,78	0,11
LGBMRegressor	0,80	4,82	0,11
HistGradientBoostingRegressor	0,79	4,92	0,88
OrthogonalMatchingPursuitCV	0,79	4,96	0,01
KNeighborsRegressor	0,78	5,02	0,01
LassoLarsIC	0,78	5,03	0,02
BayesianRidge	0,78	5,09	0,03
LarsCV	0,77	5,22	0,02
LassoLars	0,60	6,81	0,01
Lasso	0,60	6,81	0,01
ElasticNet	0,50	7,63	0,01
TweedieRegressor	0,47	7,88	0,01
GammaRegressor	0,47	7,89	0,02
PassiveAggressiveRegressor	0,45	7,99	0,01
OrthogonalMatchingPursuit	0,31	8,95	0,01
NuSVR	0,26	9,29	0,01
SVR	0,22	9,53	0,01
DummyRegressor	-	10,84	0,01
LinearSVR	-	13,12	0,01
Lars	-	28,48	0,01
MLPRegressor	-	35,23	0,09
KernelRidge	-	55,79	0,01
GaussianProcessRegressor	-	106,78	0,02
RANSACRegressor	-	186,59	0,13

Fonte: O autor, 2023.

ANEXO E - Gráficos de distribuição e densidade dos valores imputados e observados para os estudos de caso I a VIII

✓ Estudo de Caso I

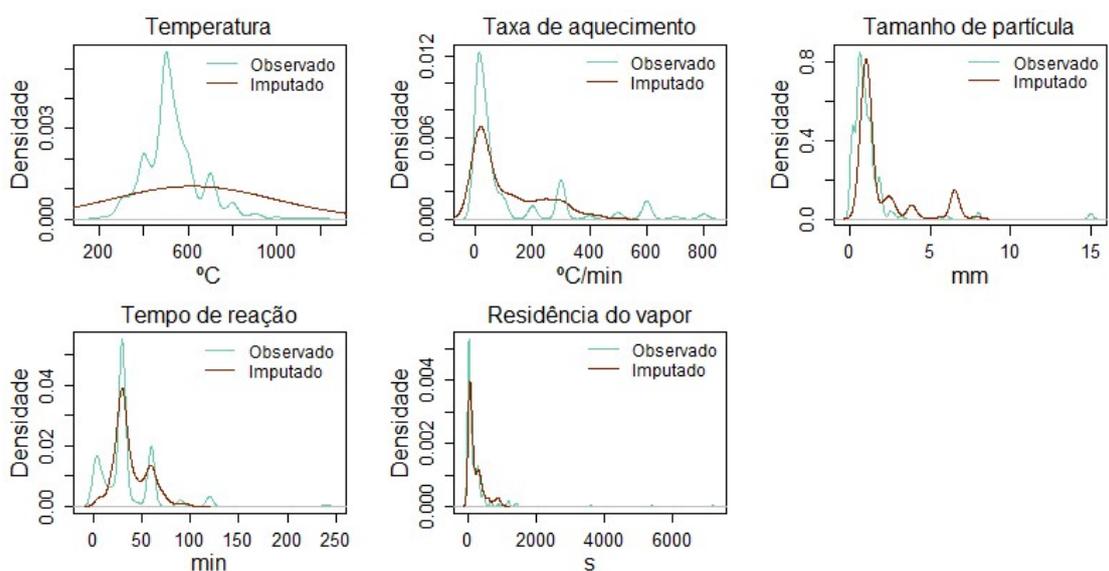
- Distribuição dos dados observados e imputados k-NN



• -Valores observados; • -Valores imputados

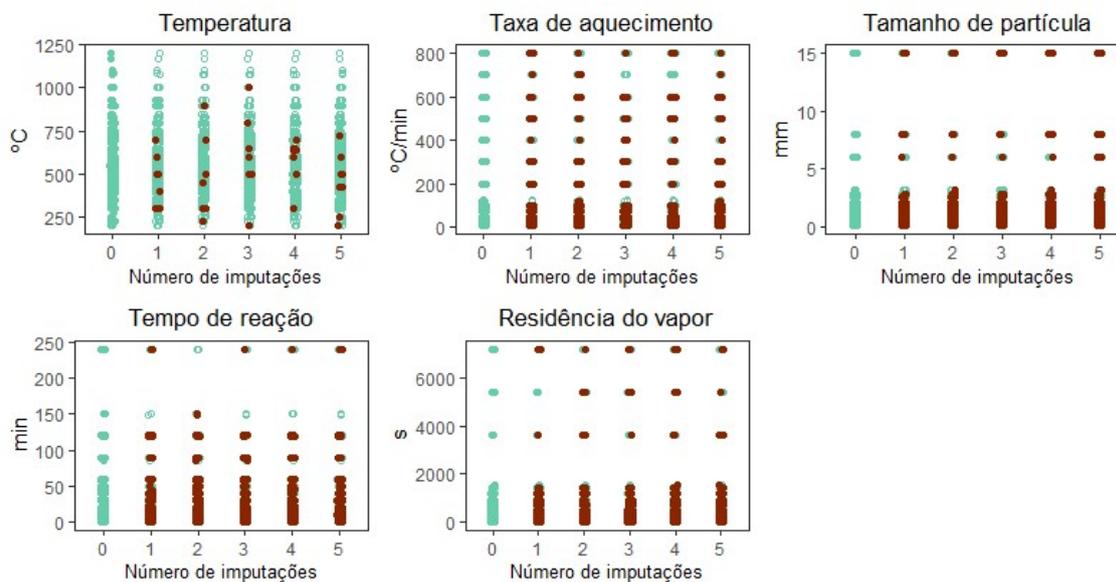
Fonte: O autor, 2023.

- Densidade dos dados observados e imputados por -kNN



Fonte: O autor, 2023.

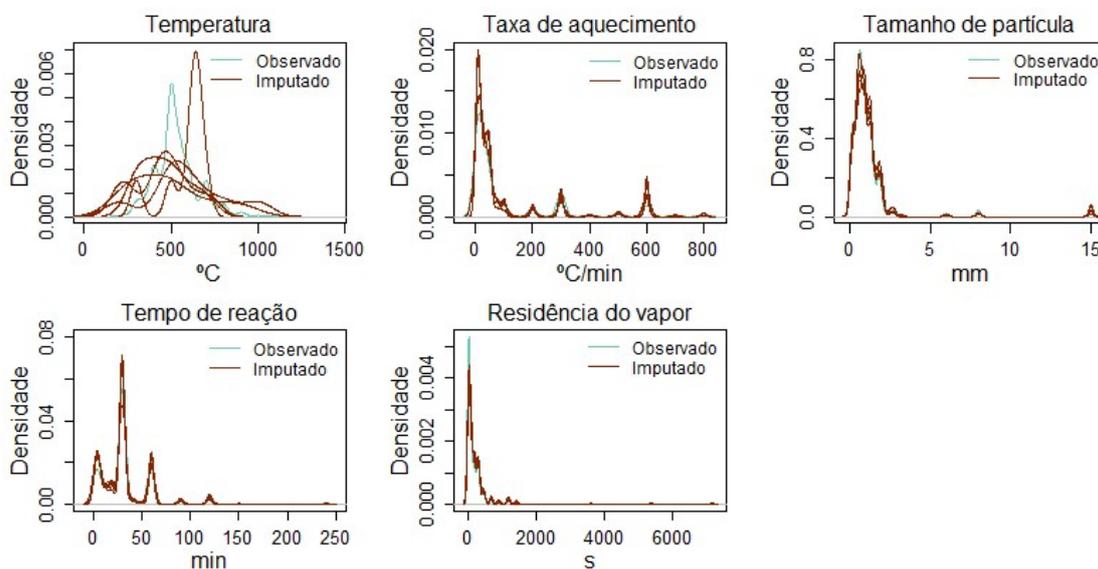
- Distribuição dos dados observados e imputados RF



● -Valores observados; ● -Valores imputados

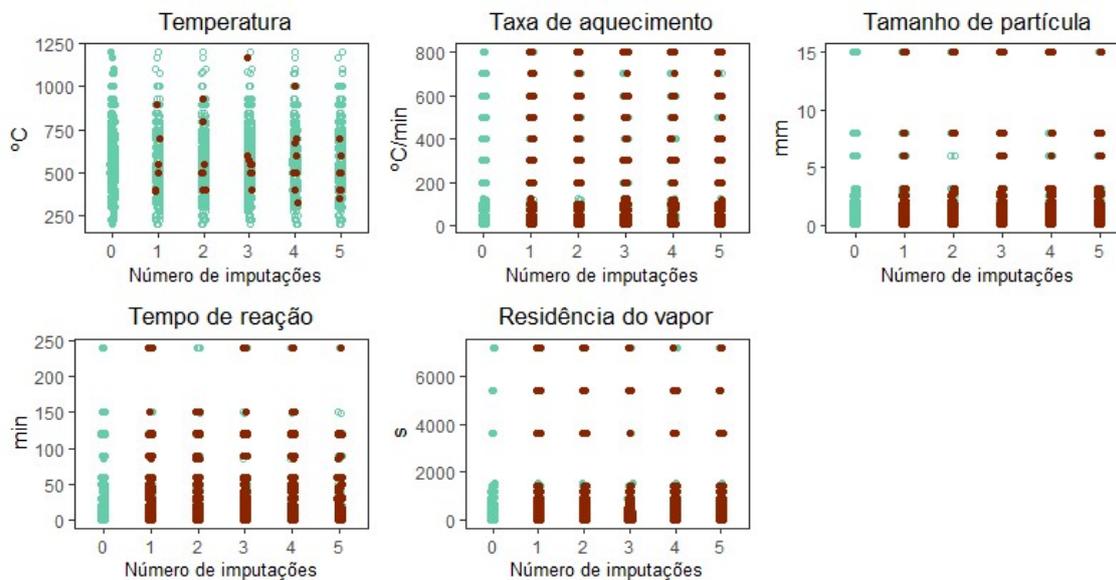
Fonte: O autor, 2023.

- Densidade dos dados observados e imputados por RF



Fonte: O autor, 2023.

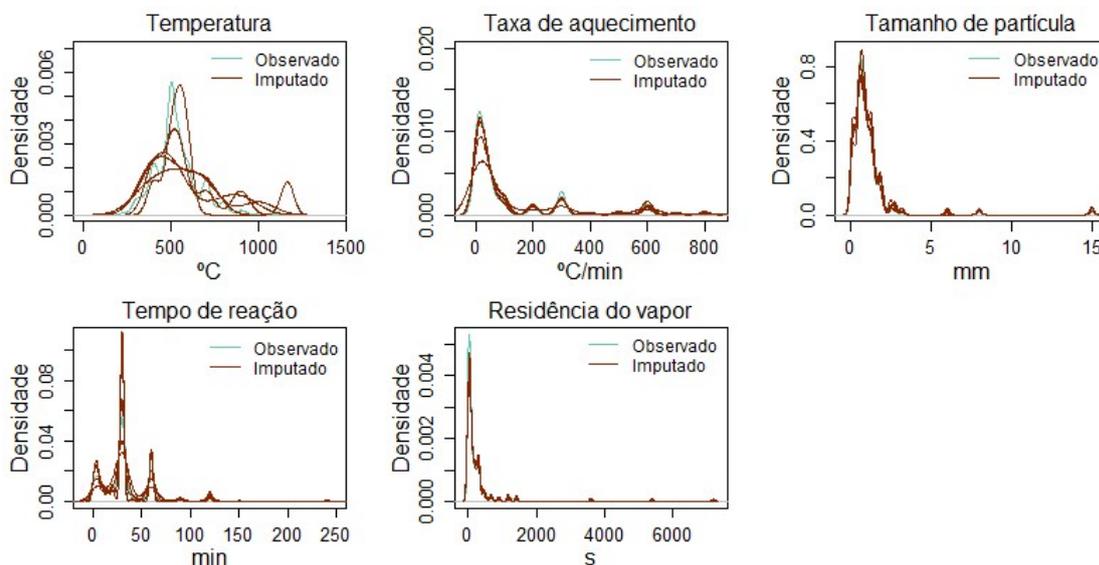
- Distribuição dos dados observados e imputados Midastouch



● -Valores observados; ● -Valores imputados

Fonte: O autor, 2023.

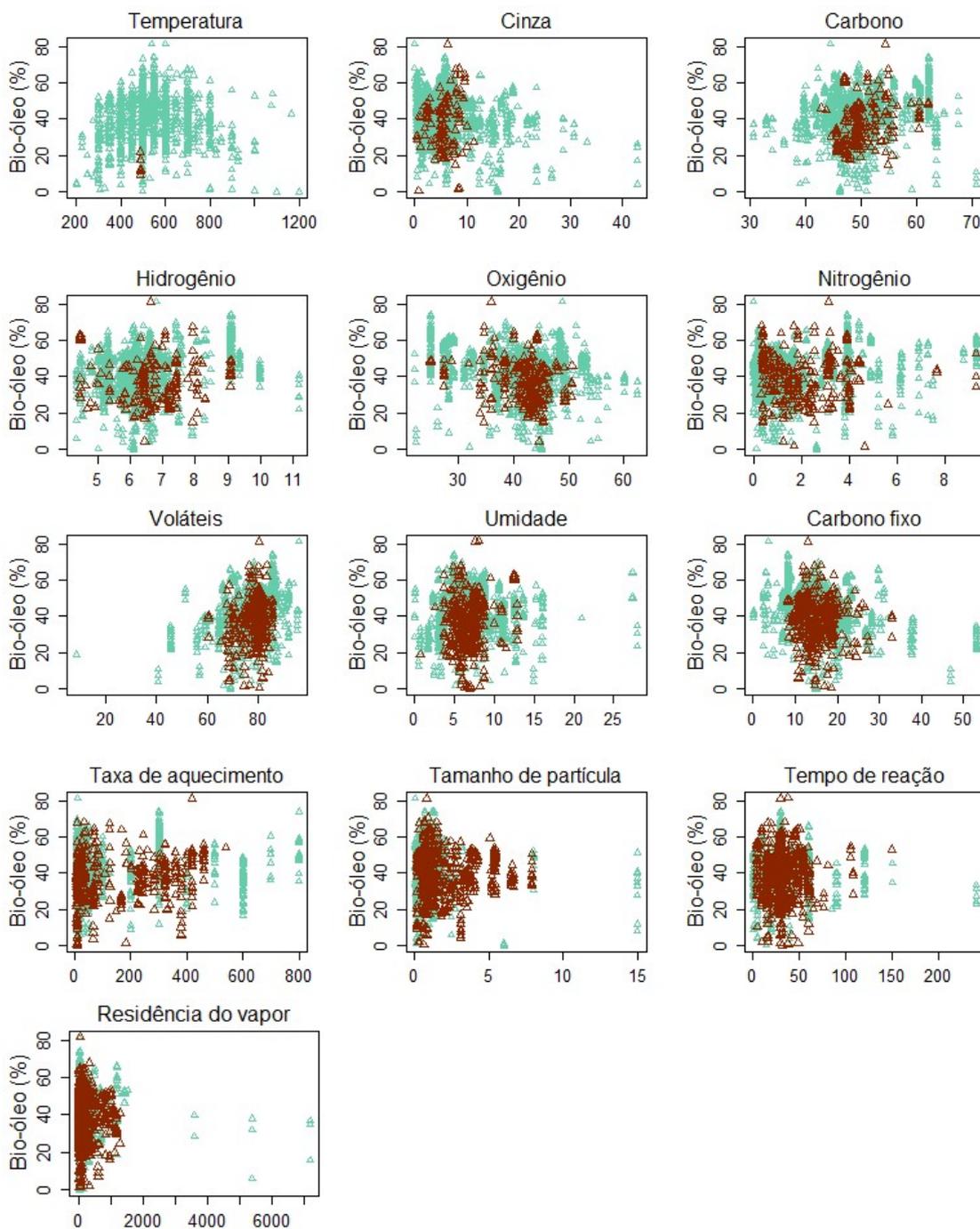
- Densidade dos dados observados e imputados por Midastouch



Fonte: O autor, 2023.

✓ Estudo de Caso II

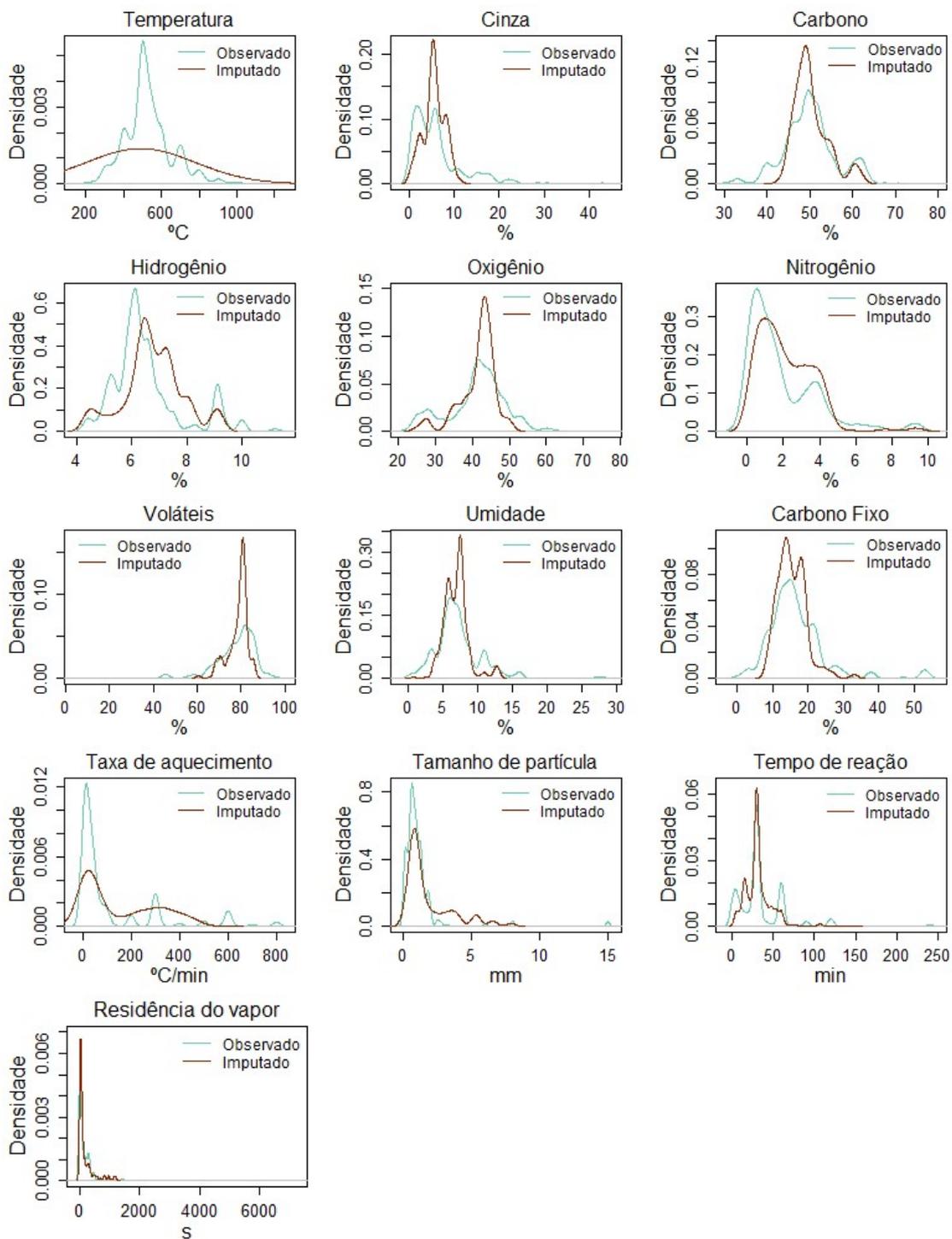
- Distribuição dos dados observados e imputados k-NN



● -Valores observados; ● -Valores imputados

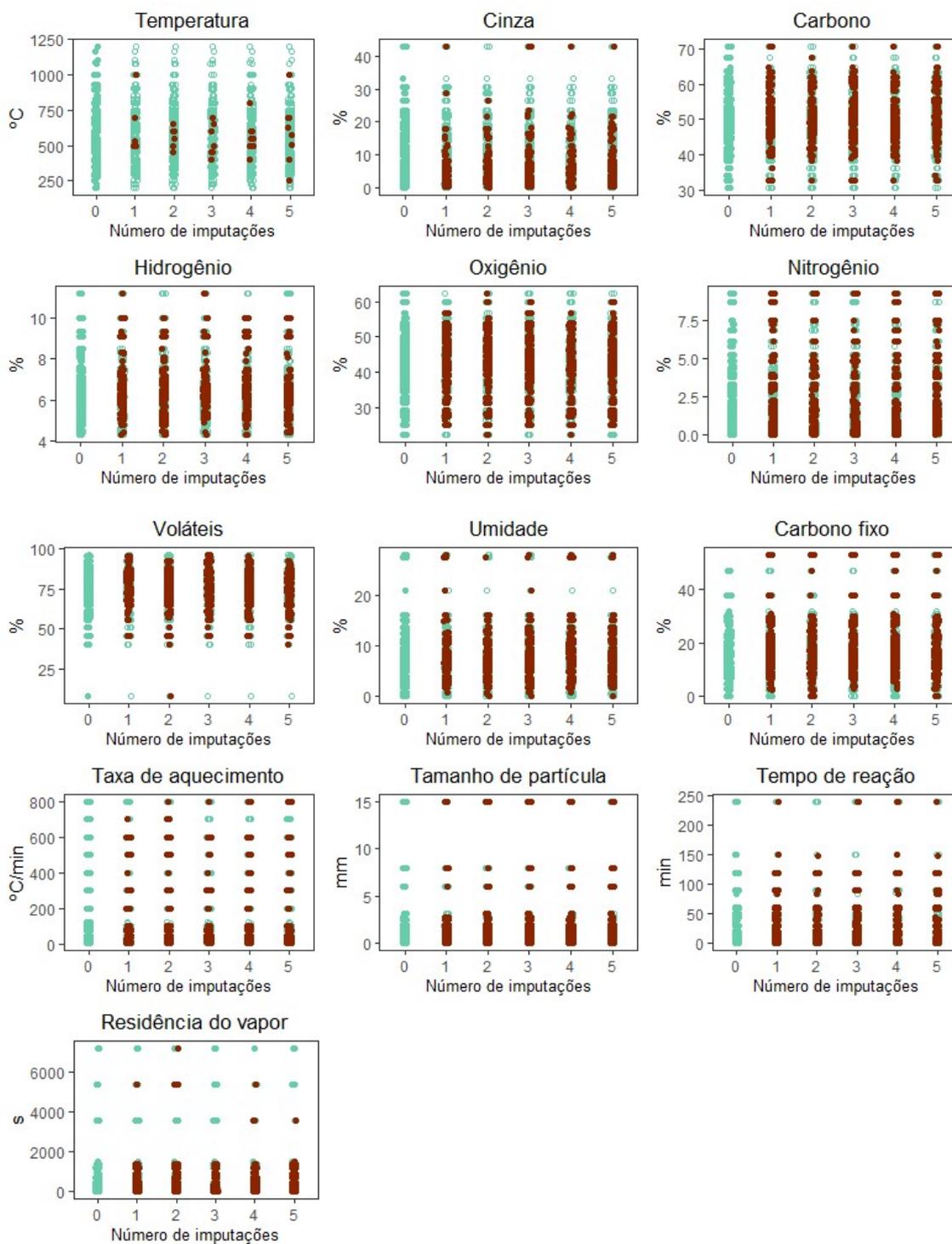
Fonte: O autor, 2023.

- Densidade dos dados observados e imputados por -kNN



Fonte: O autor, 2023.

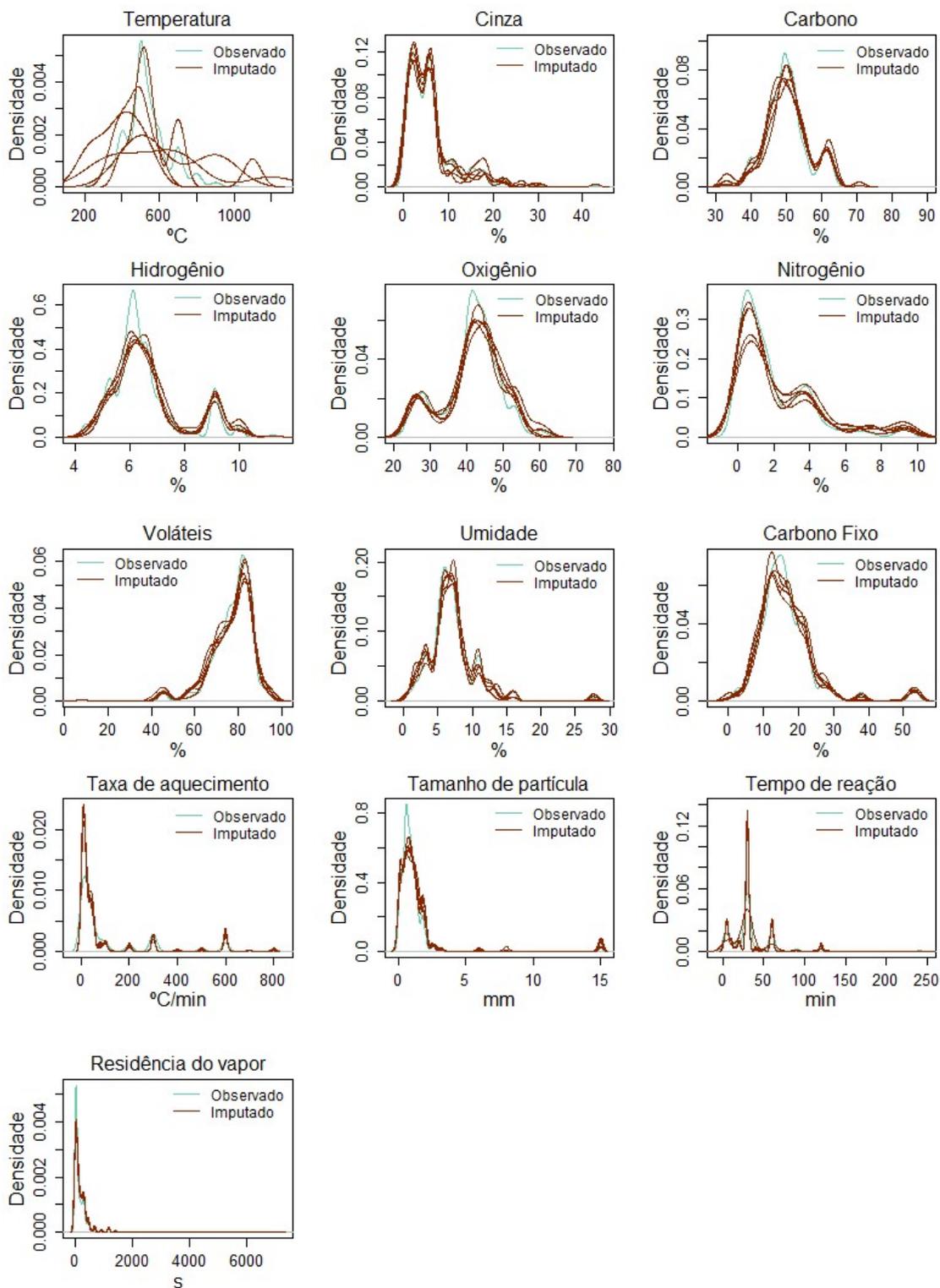
- Distribuição dos dados observados e imputados RF



● -Valores observados; ● -Valores imputados

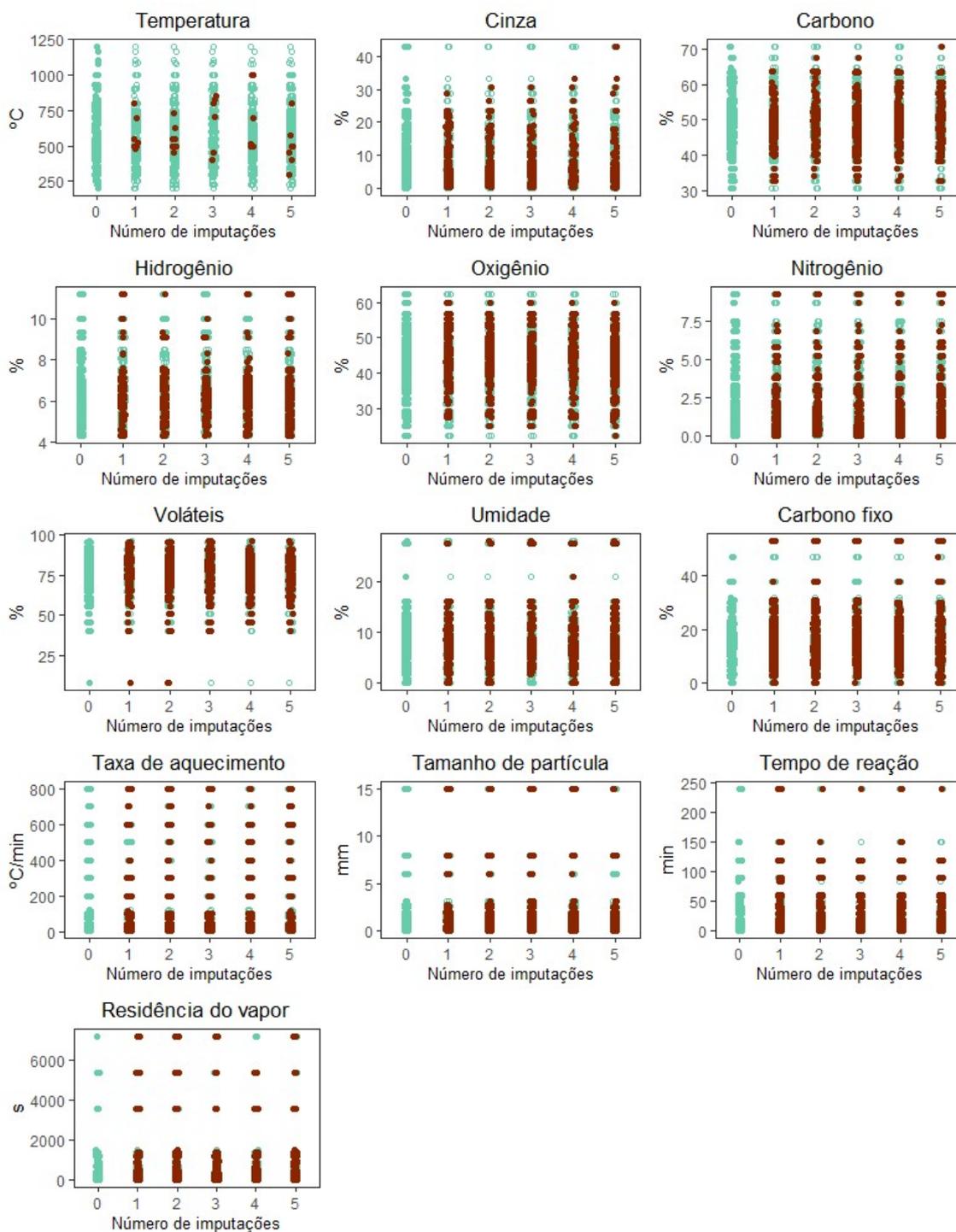
Fonte: O autor, 2023.

- Densidade dos dados observados e imputados por RF



Fonte: O autor, 2023.

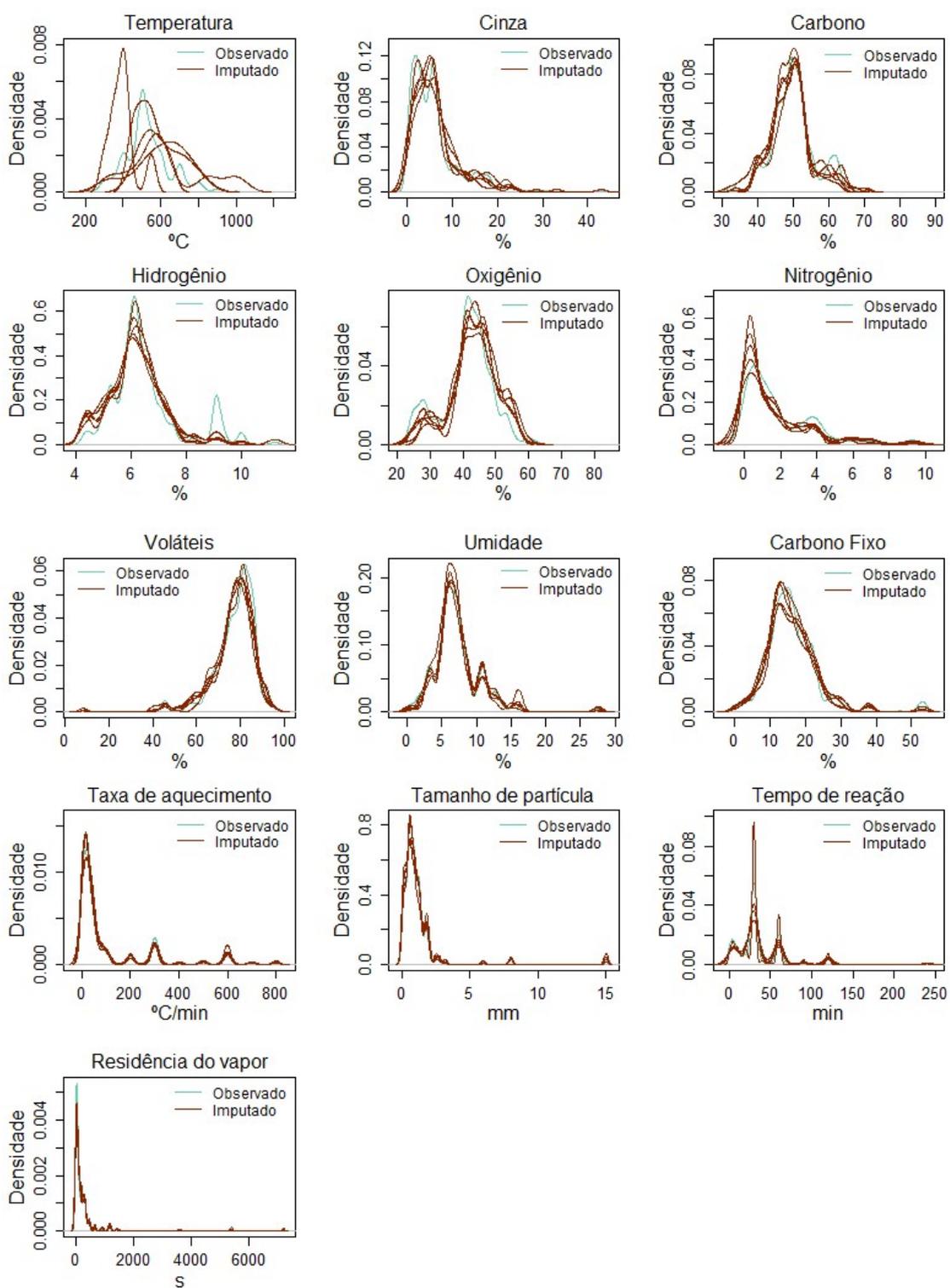
- Distribuição dos dados observados e imputados Midastouch



● -Valores observados; ● -Valores imputados

Fonte: O autor, 2023.

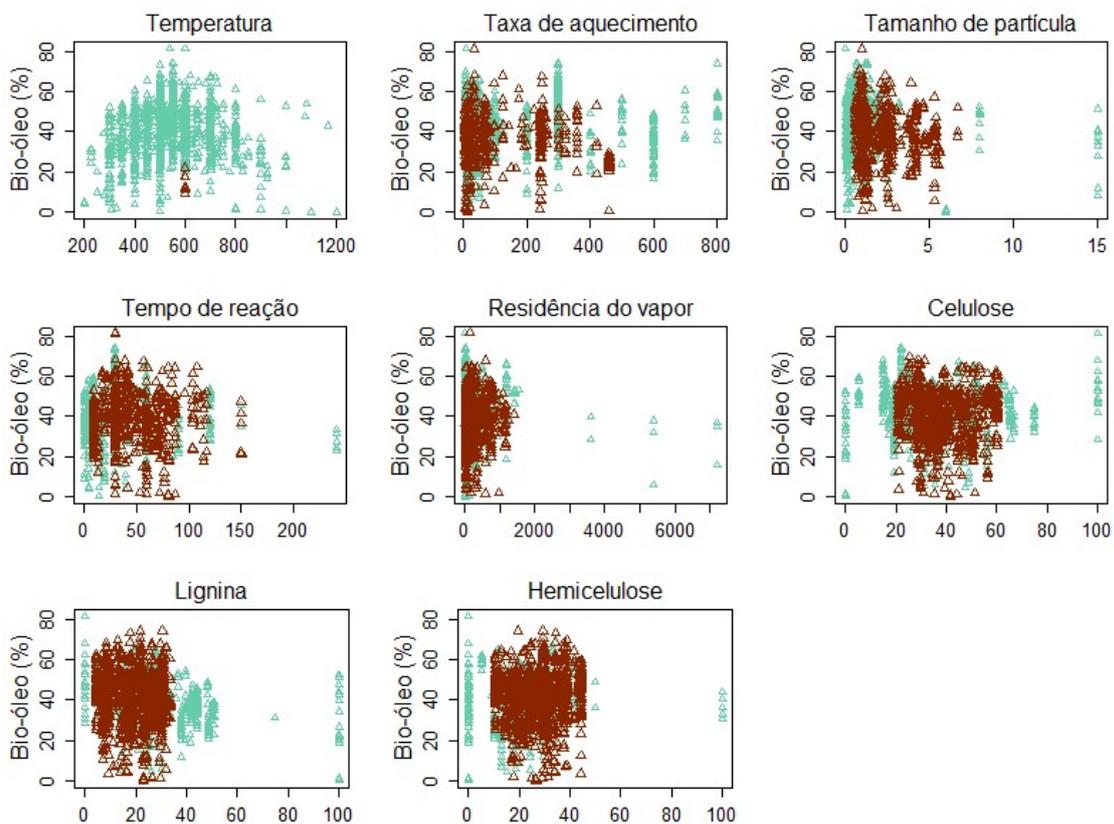
- Densidade dos dados observados e imputados por Midastouch



Fonte: O autor, 2023.

✓ Estudo de Caso III

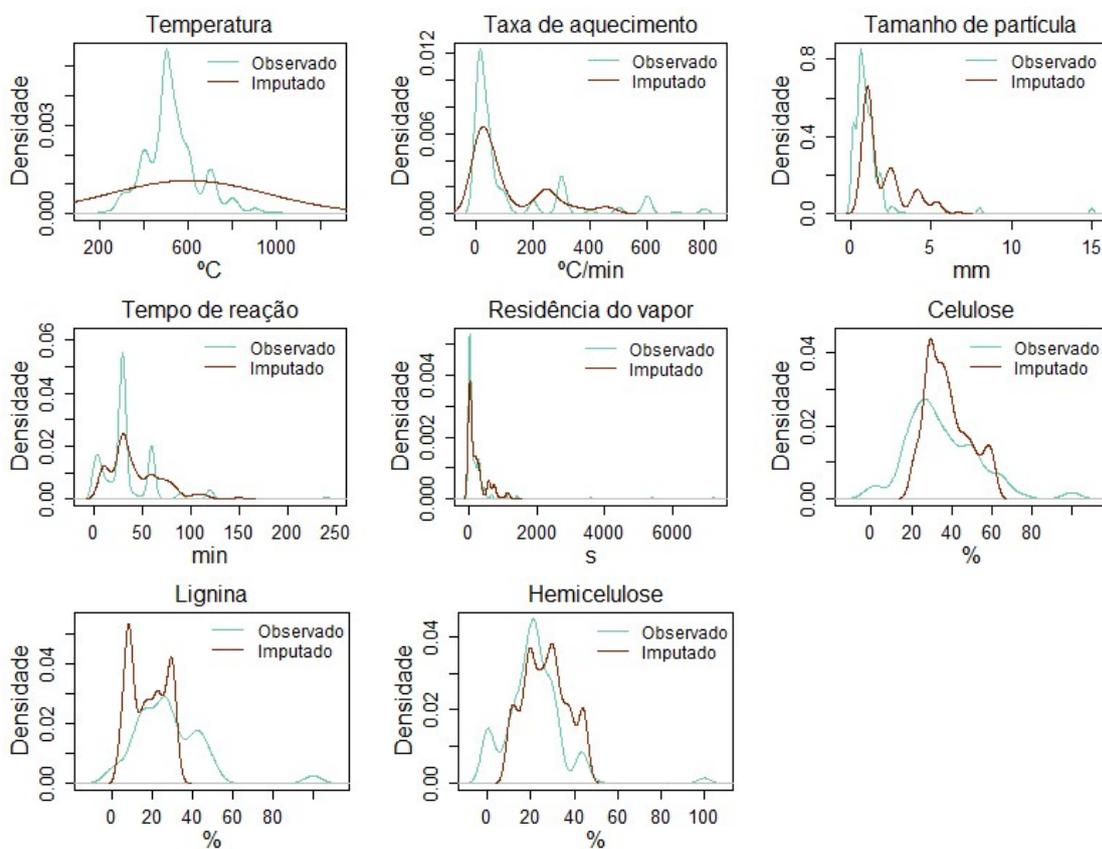
- Distribuição dos dados observados e imputados k-NN



● -Valores observados; ● -Valores imputados

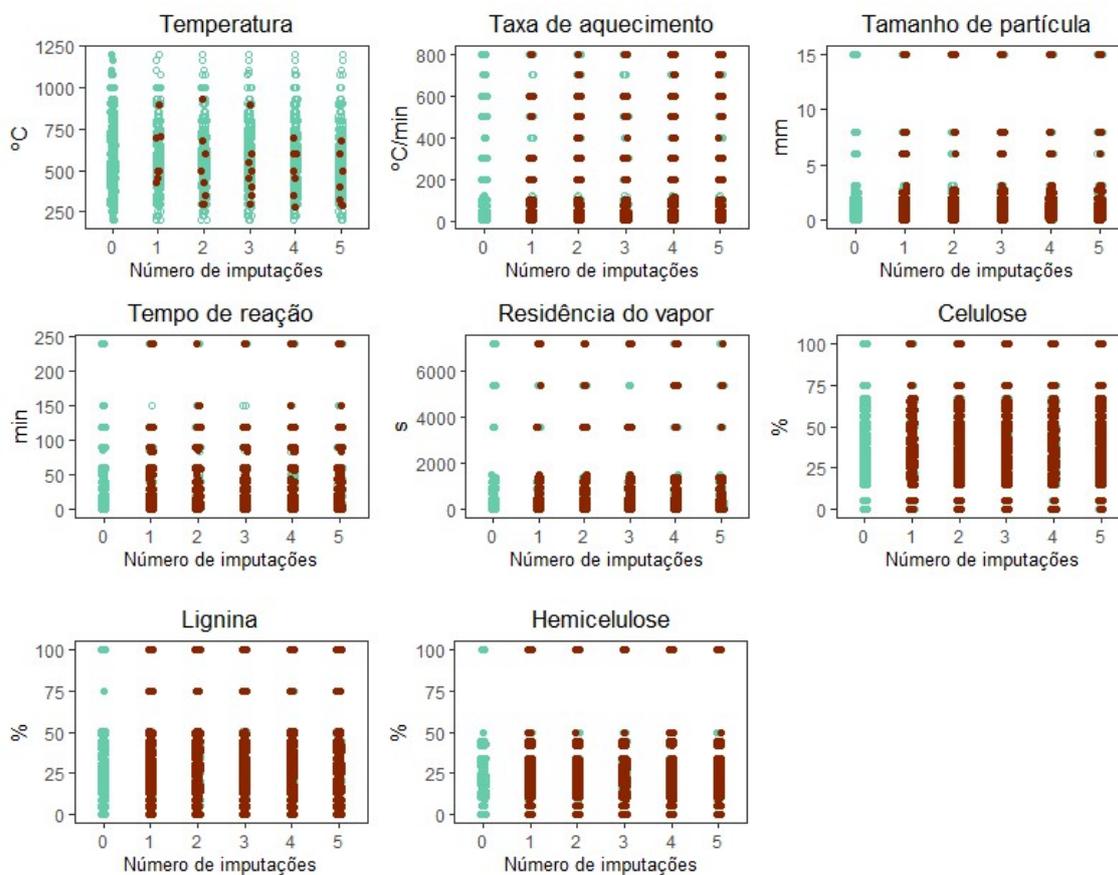
Fonte: O autor, 2023.

- Densidade dos dados observados e imputados por -kNN



Fonte: O autor, 2023.

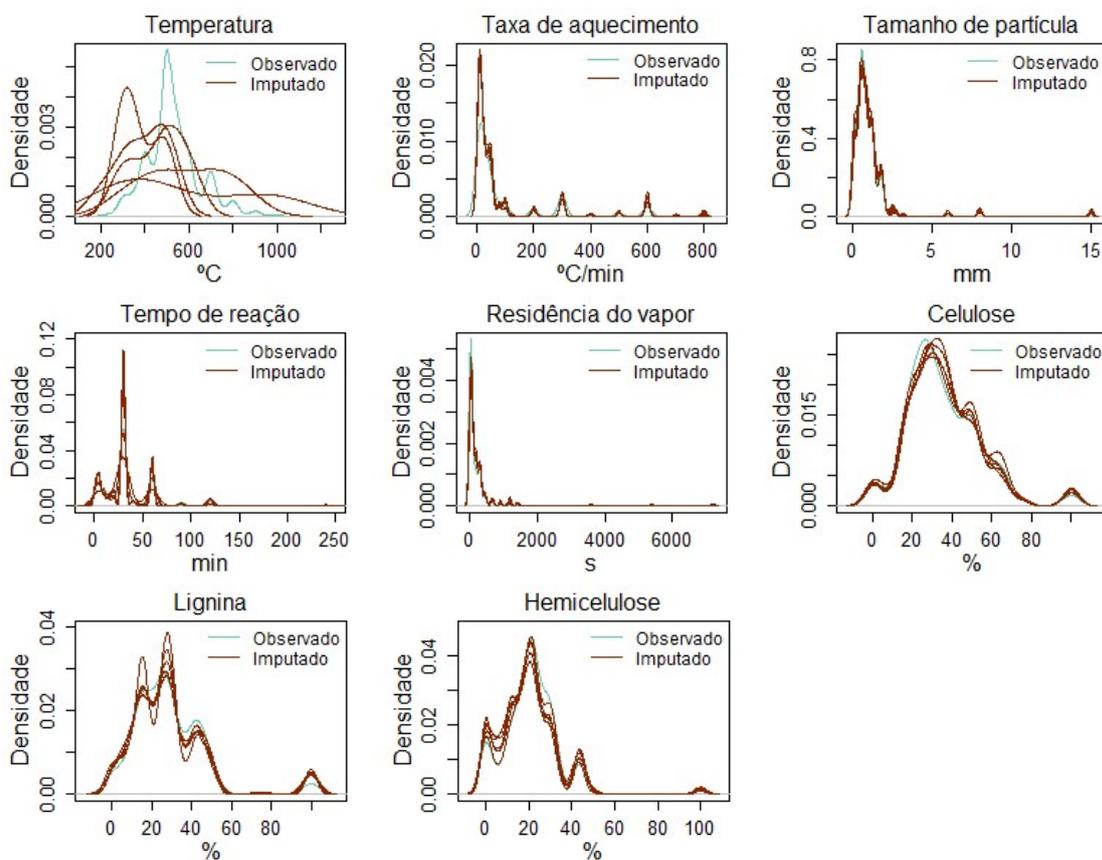
- Distribuição dos dados observados e imputados por RF



● -Valores observados; ● -Valores imputados

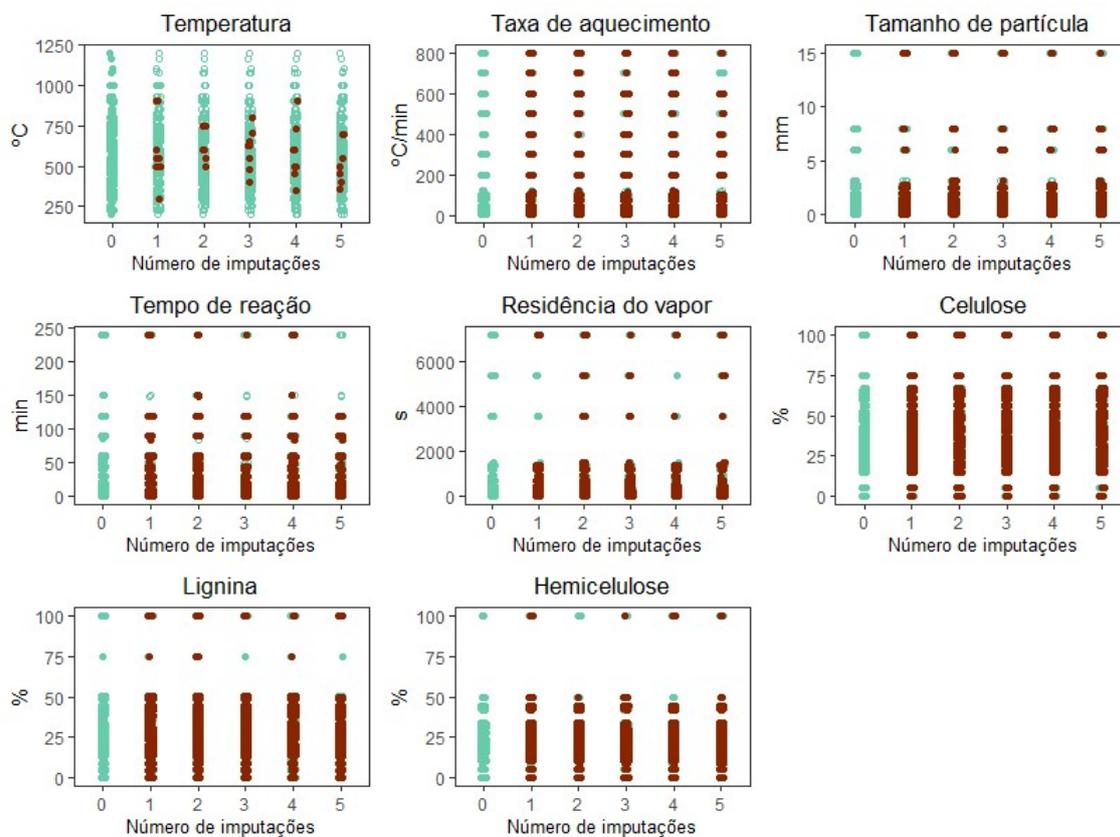
Fonte: O autor, 2023.

- Densidade dos dados observados e imputados por RF



Fonte: O autor, 2023.

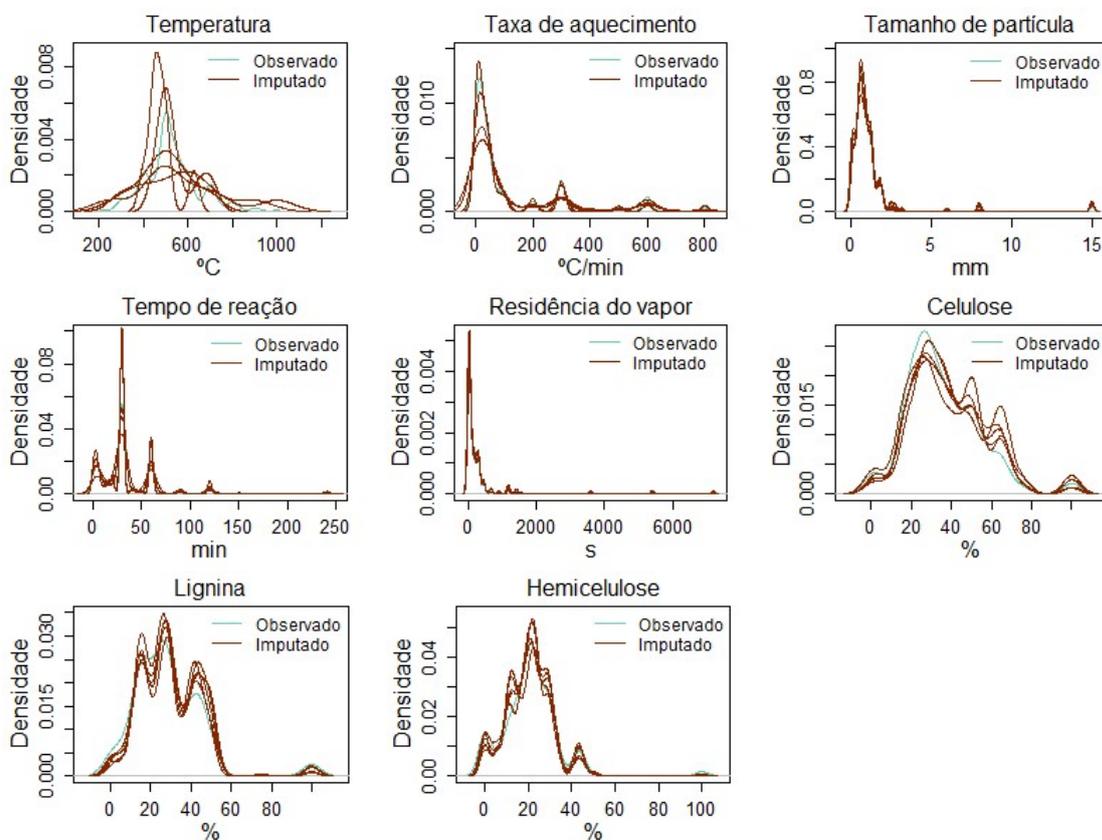
- Distribuição dos dados observados e imputados Midastouch



• -Valores observados; • -Valores imputados

Fonte: O autor, 2023.

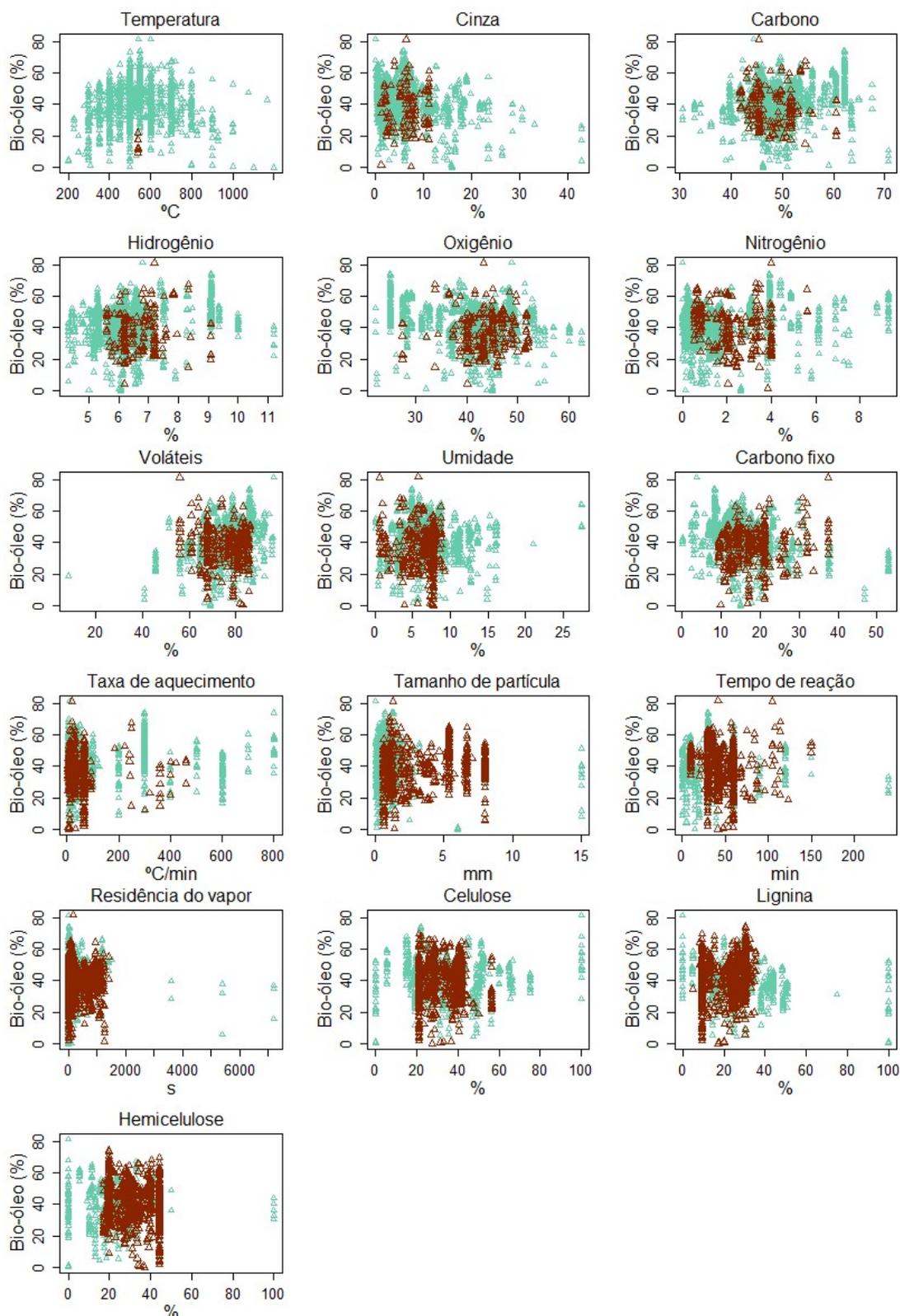
- Densidade dos dados observados e imputados por Midastouch



Fonte: O autor, 2023.

✓ Estudo de Caso IV

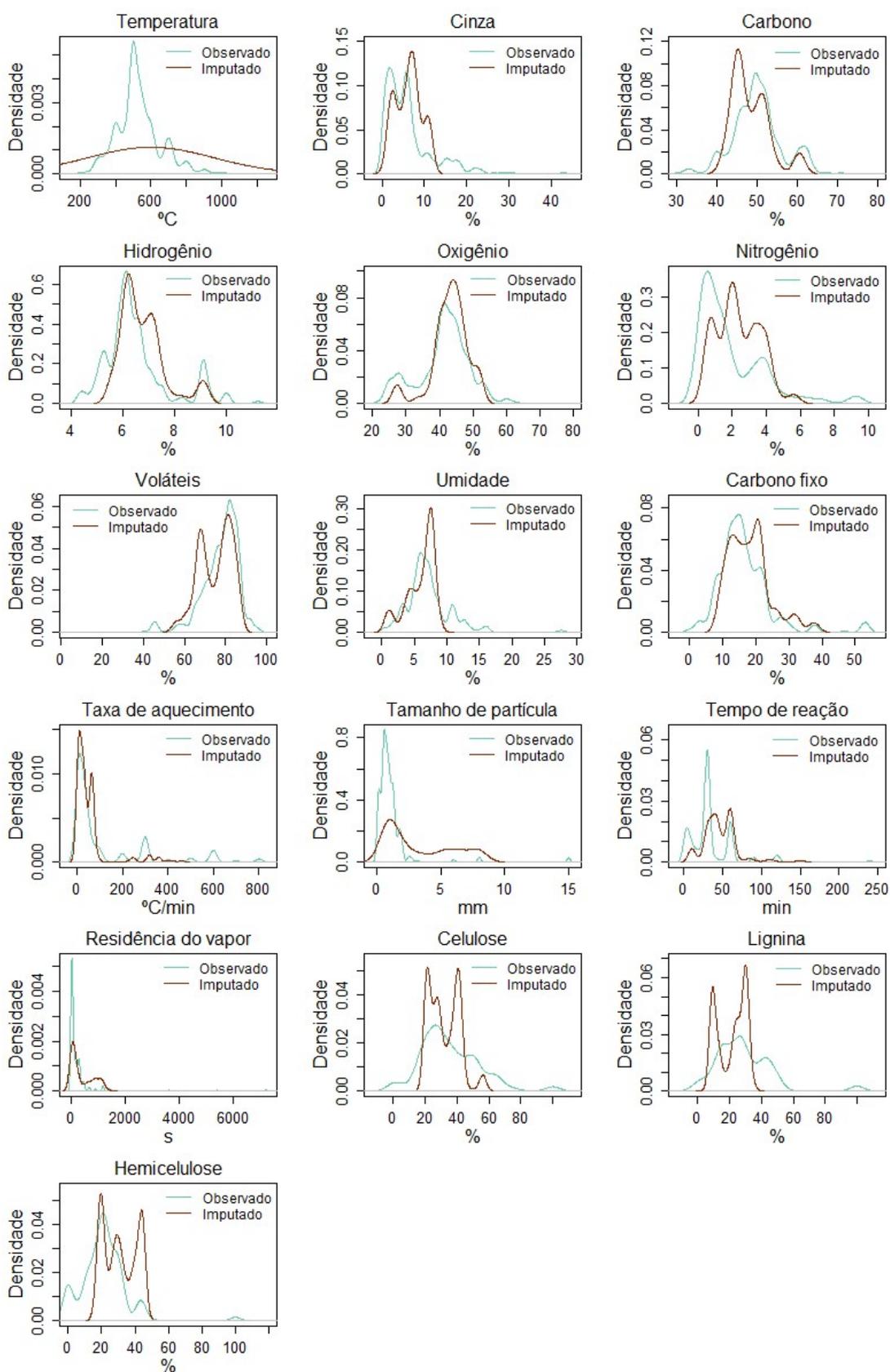
- Distribuição dos dados observados e imputados k-NN



● -Valores observados; ● -Valores imputados

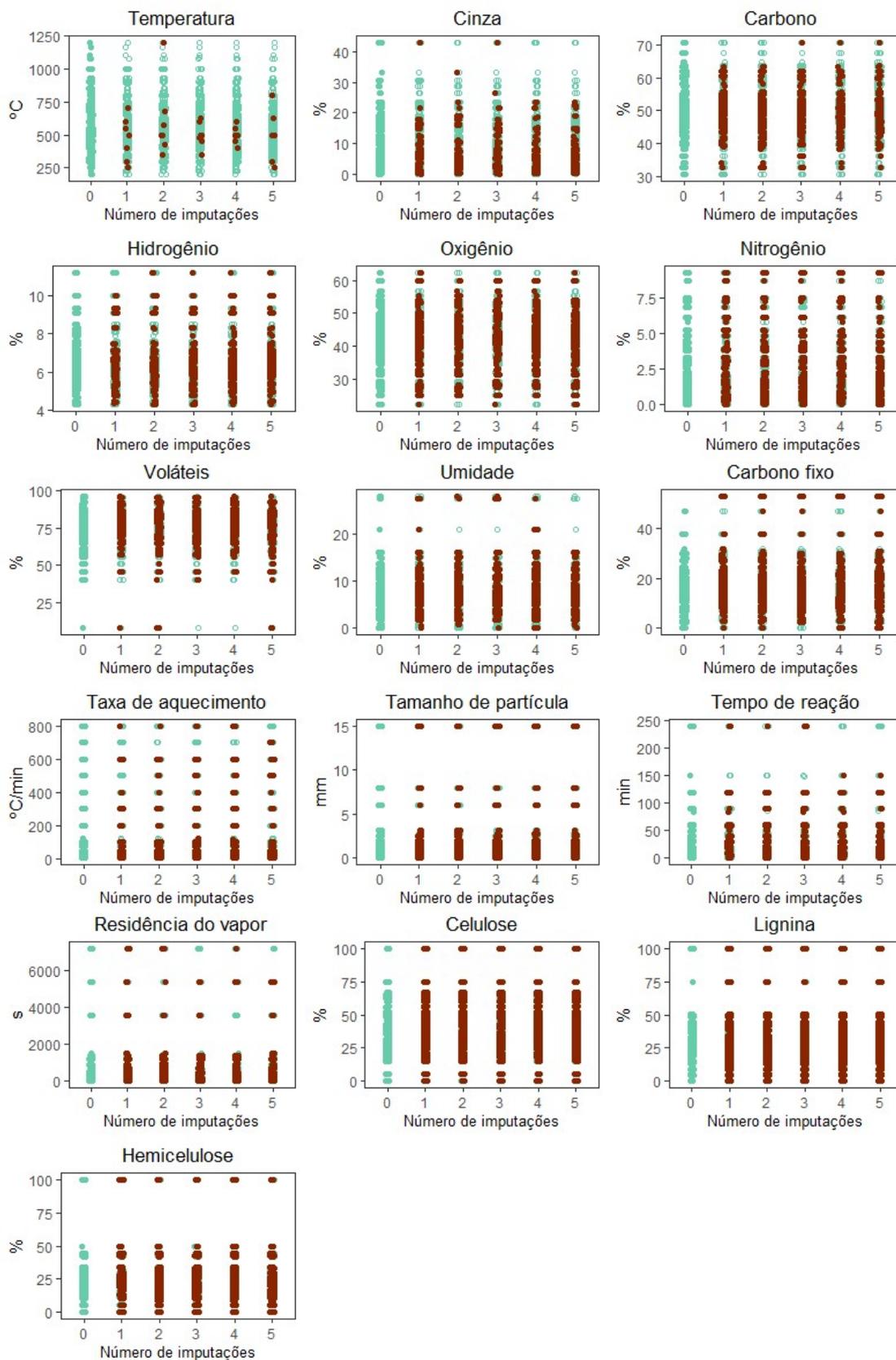
Fonte: O autor, 2023.

- Densidade dos dados observados e imputados por -kNN



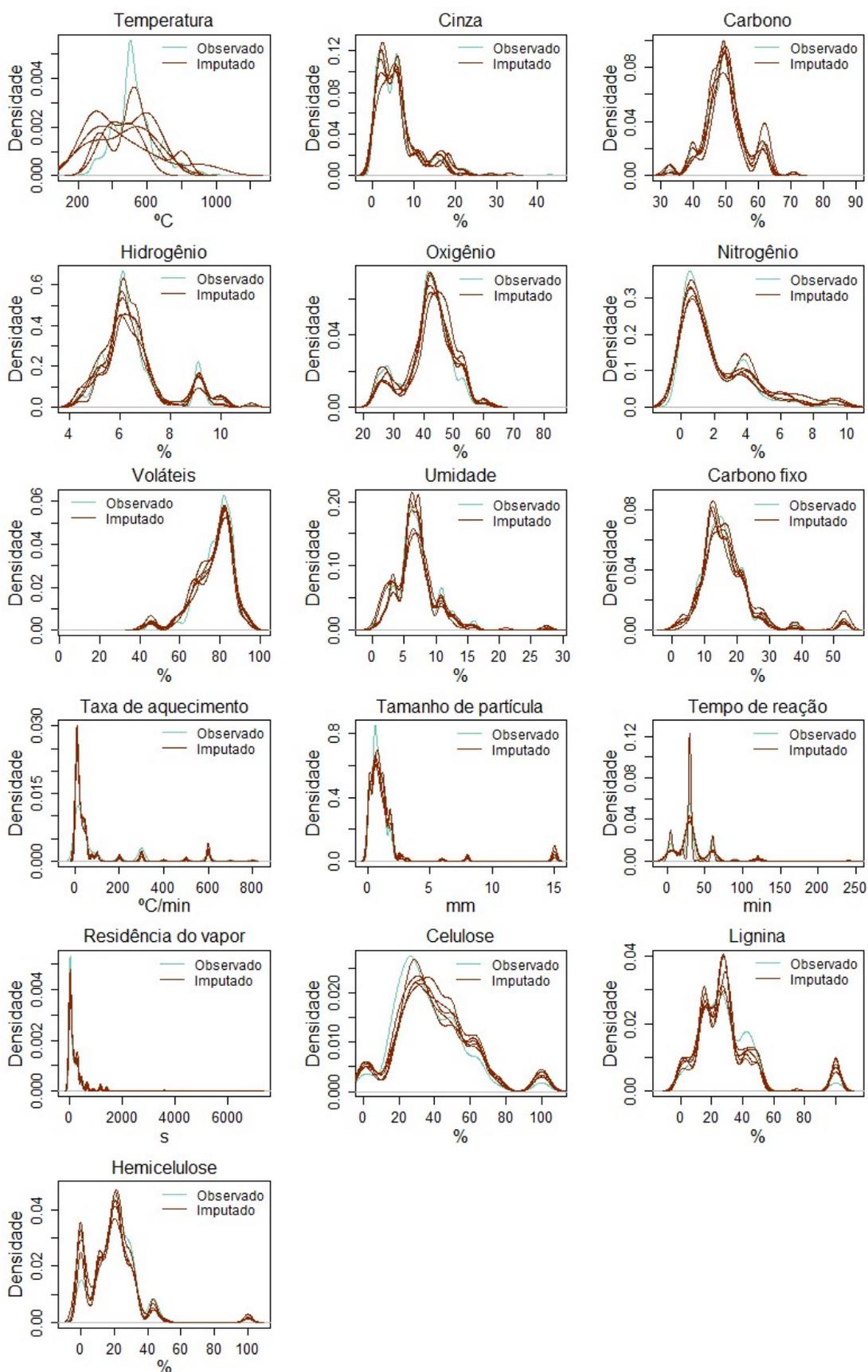
Fonte: O autor, 2023.

- Distribuição dos dados observados e imputados por RF



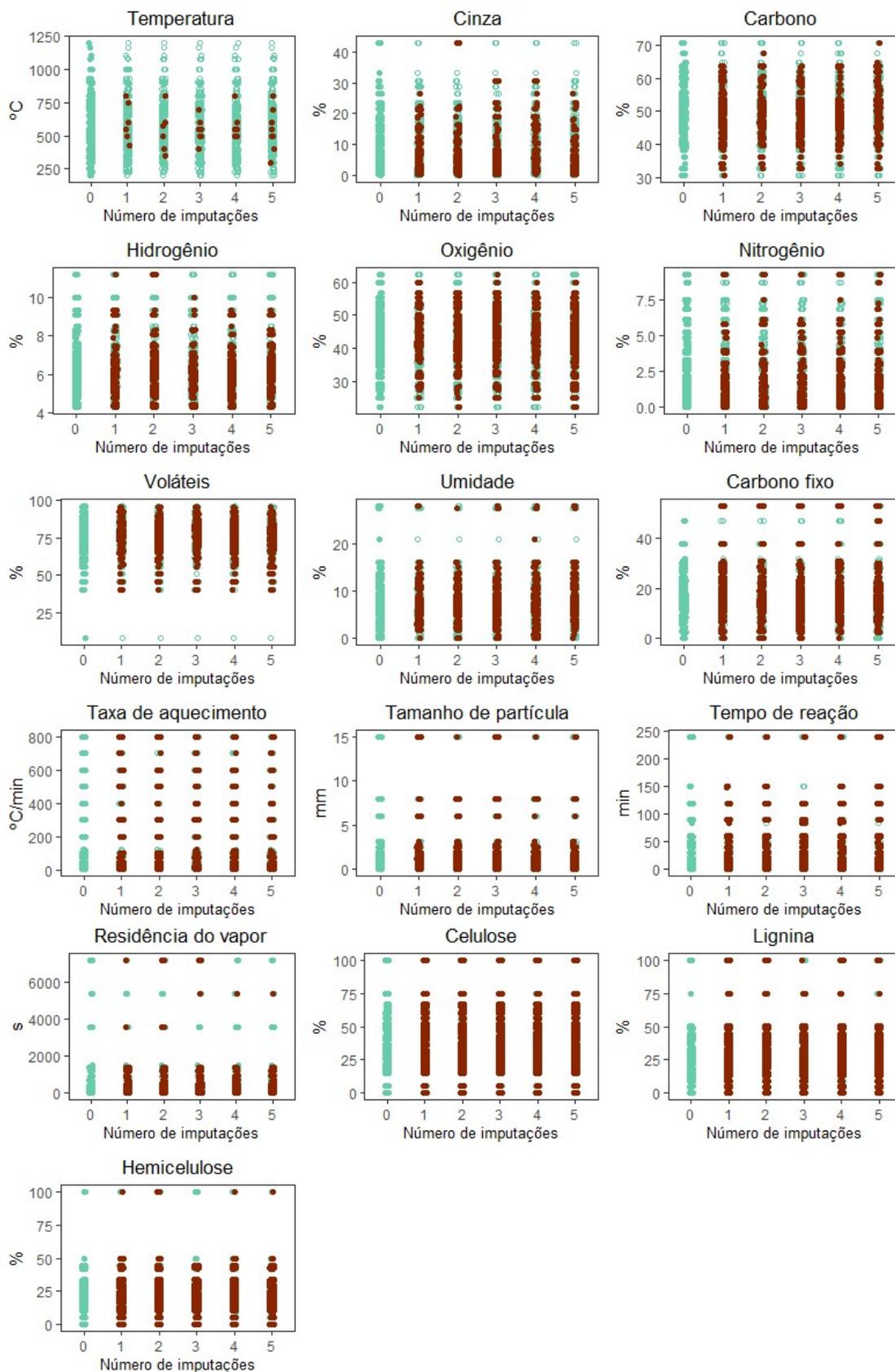
Fonte: O autor, 2023.

- Densidade dos dados observados e imputados por RF

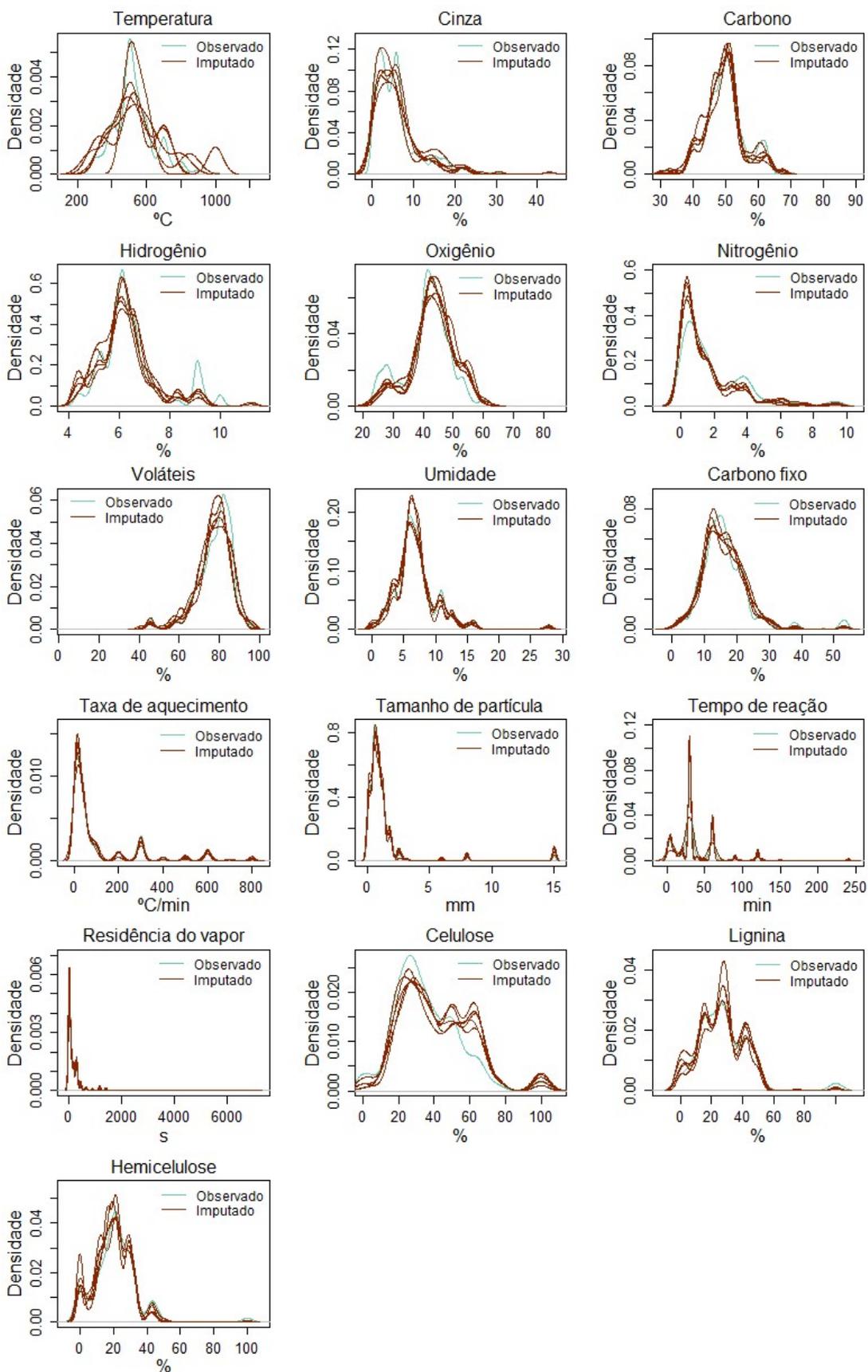


Fonte: O autor, 2023.

- Distribuição dos dados observados e imputados Midastouch



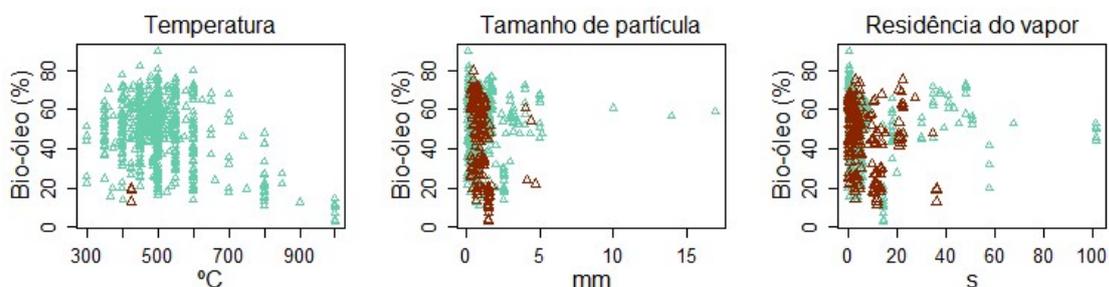
- Densidade dos dados observados e imputados por Midastouch



Fonte: O autor, 2023.

✓ Estudo de Caso V

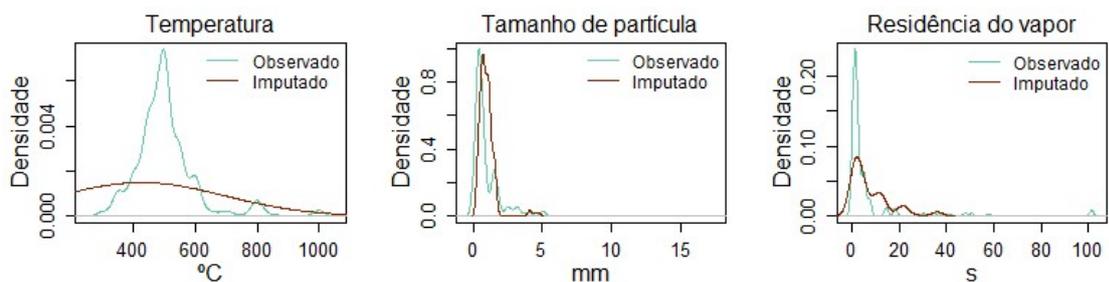
- Distribuição dos dados observados e imputados k-NN



● -Valores observados; ● -Valores imputados

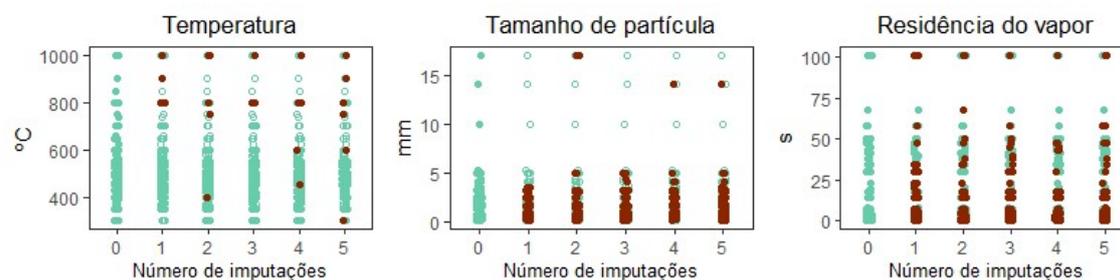
Fonte: O autor, 2023.

- Densidade dos dados observados e imputados por -kNN



Fonte: O autor, 2023.

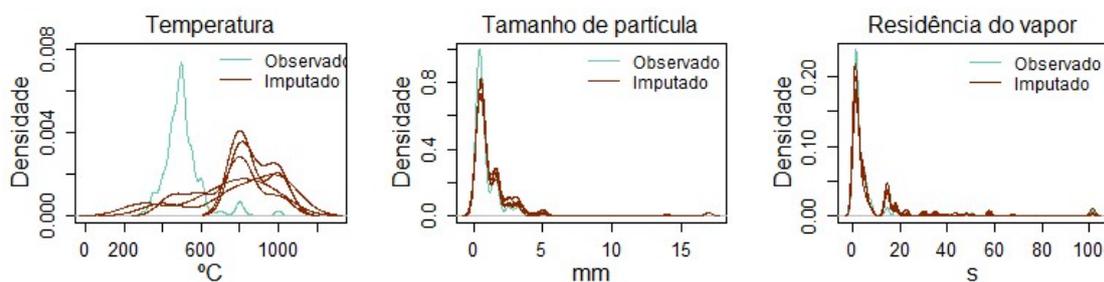
- Distribuição dos dados observados e imputados por RF



● -Valores observados; ● -Valores imputados

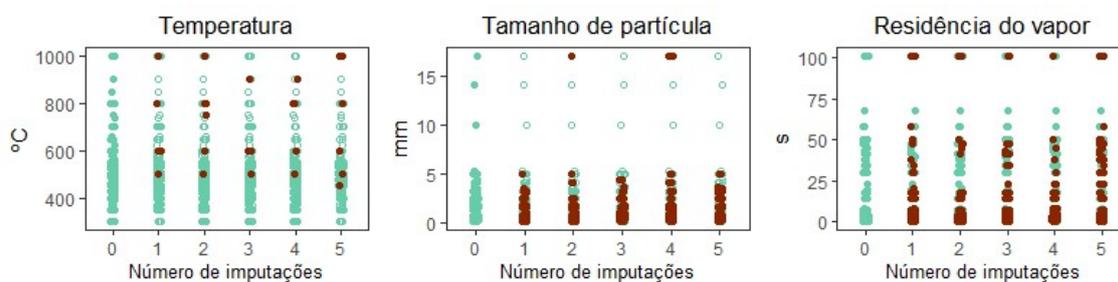
Fonte: O autor, 2023.

- Densidade dos dados observados e imputados por RF



Fonte: O autor, 2023.

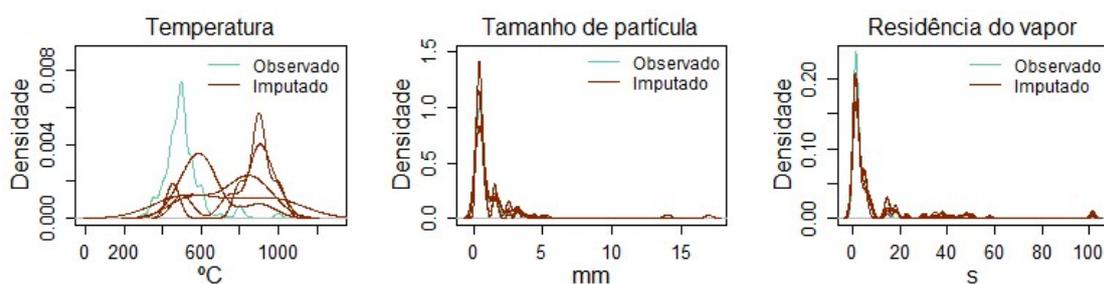
- Distribuição dos dados observados e imputados Midastouch



● -Valores observados; ● -Valores imputados

Fonte: O autor, 2023.

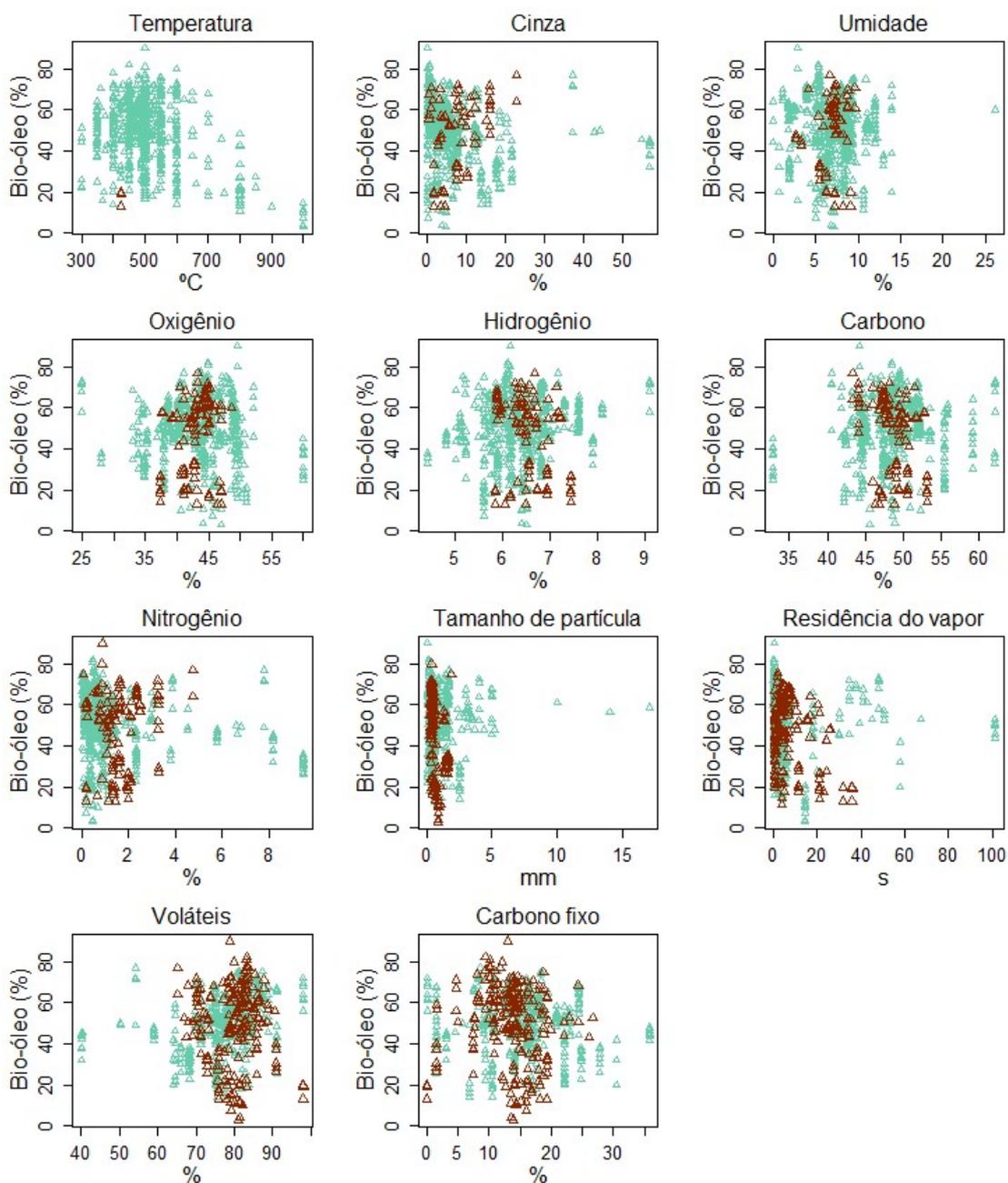
- Densidade dos dados observados e imputados por Midastouch



Fonte: O autor, 2023.

✓ Estudo de Caso VI

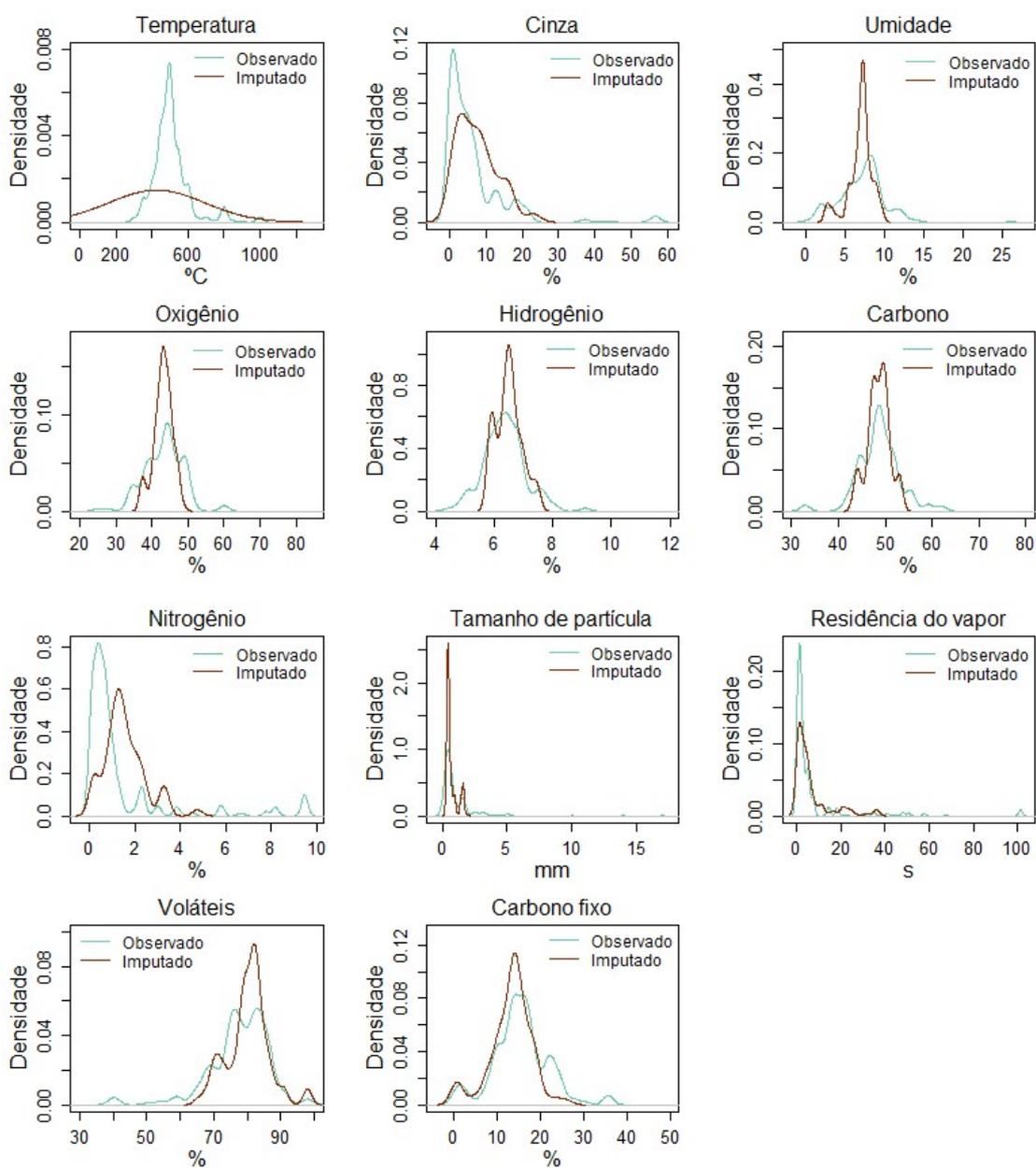
- Distribuição dos dados observados e imputados k-NN



● -Valores observados; ● -Valores imputados

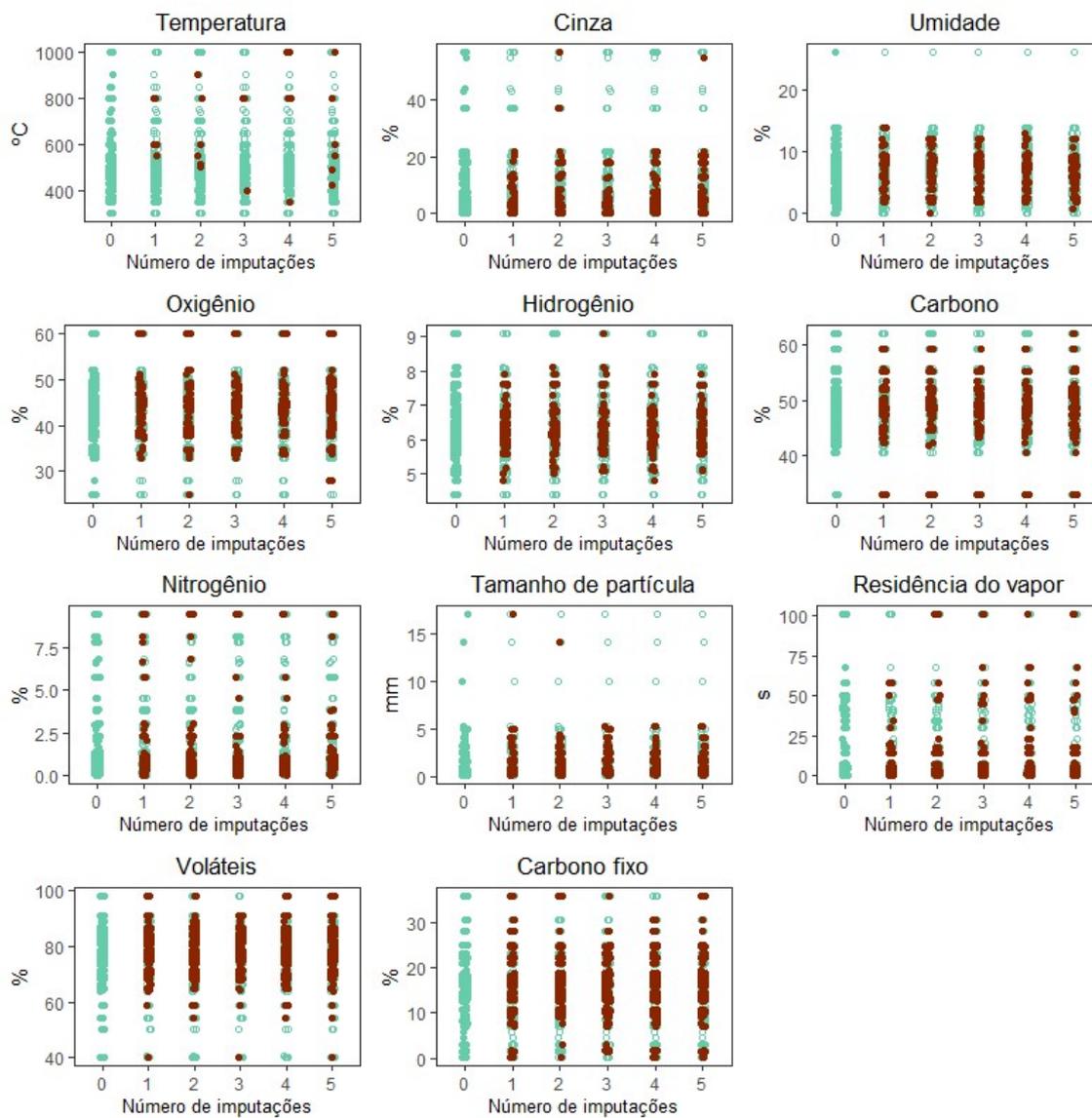
Fonte: O autor, 2023.

- Densidade dos dados observados e imputados por -k-NN



Fonte: O autor, 2023.

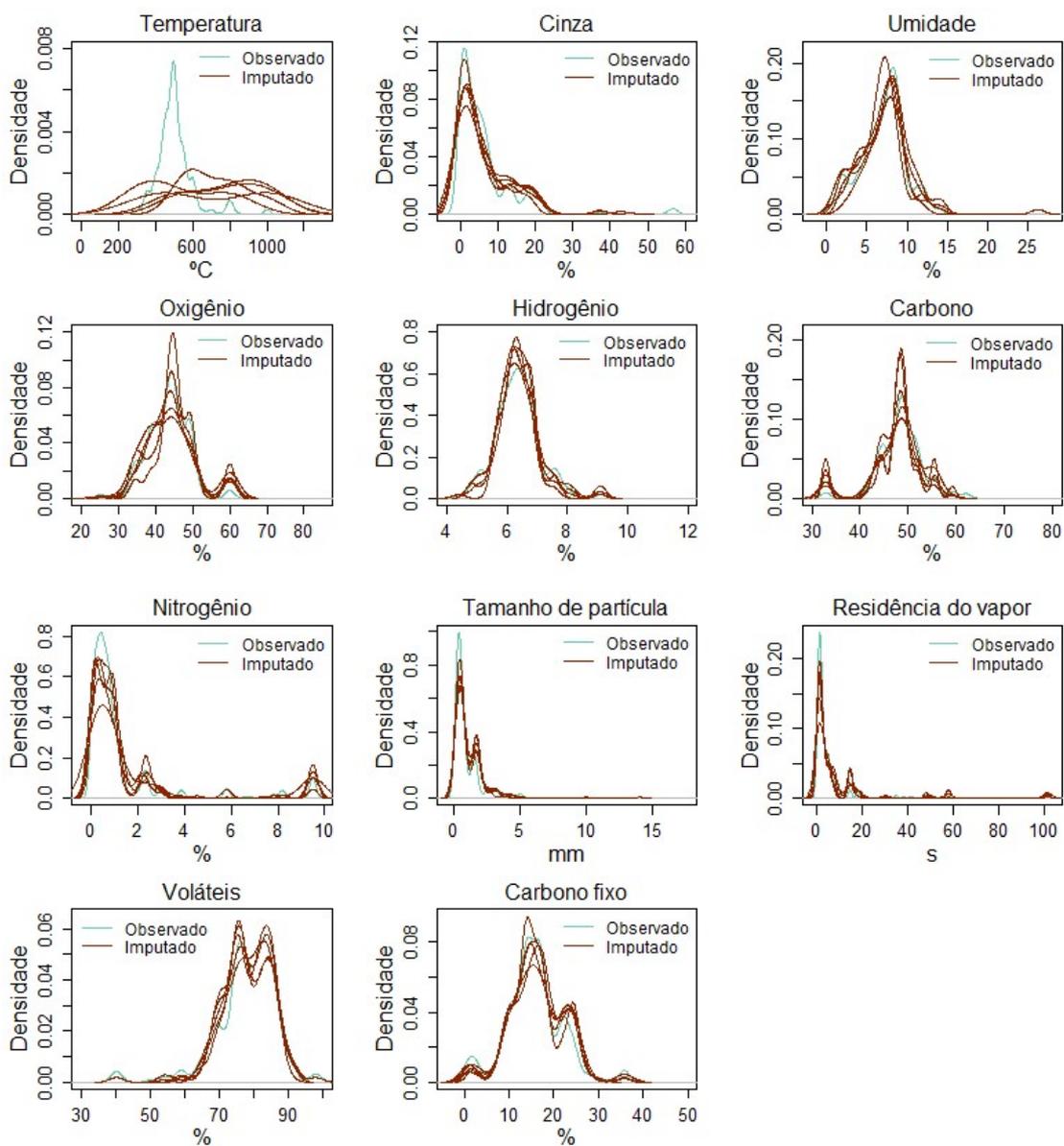
- Distribuição dos dados observados e imputados por RF



● -Valores observados; ● -Valores imputados

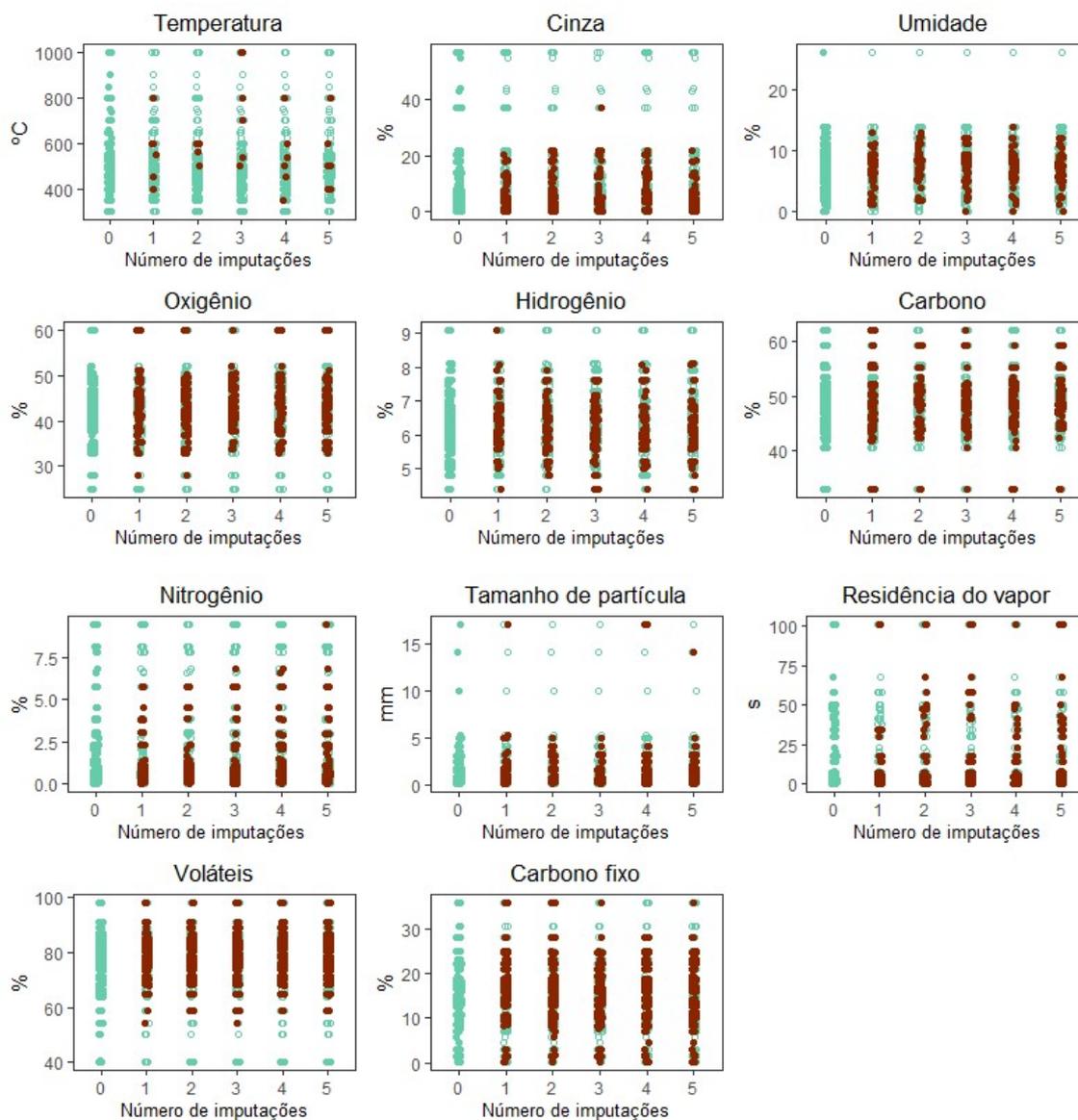
Fonte: O autor, 2023.

- Densidade dos dados observados e imputados por RF



Fonte: O autor, 2023.

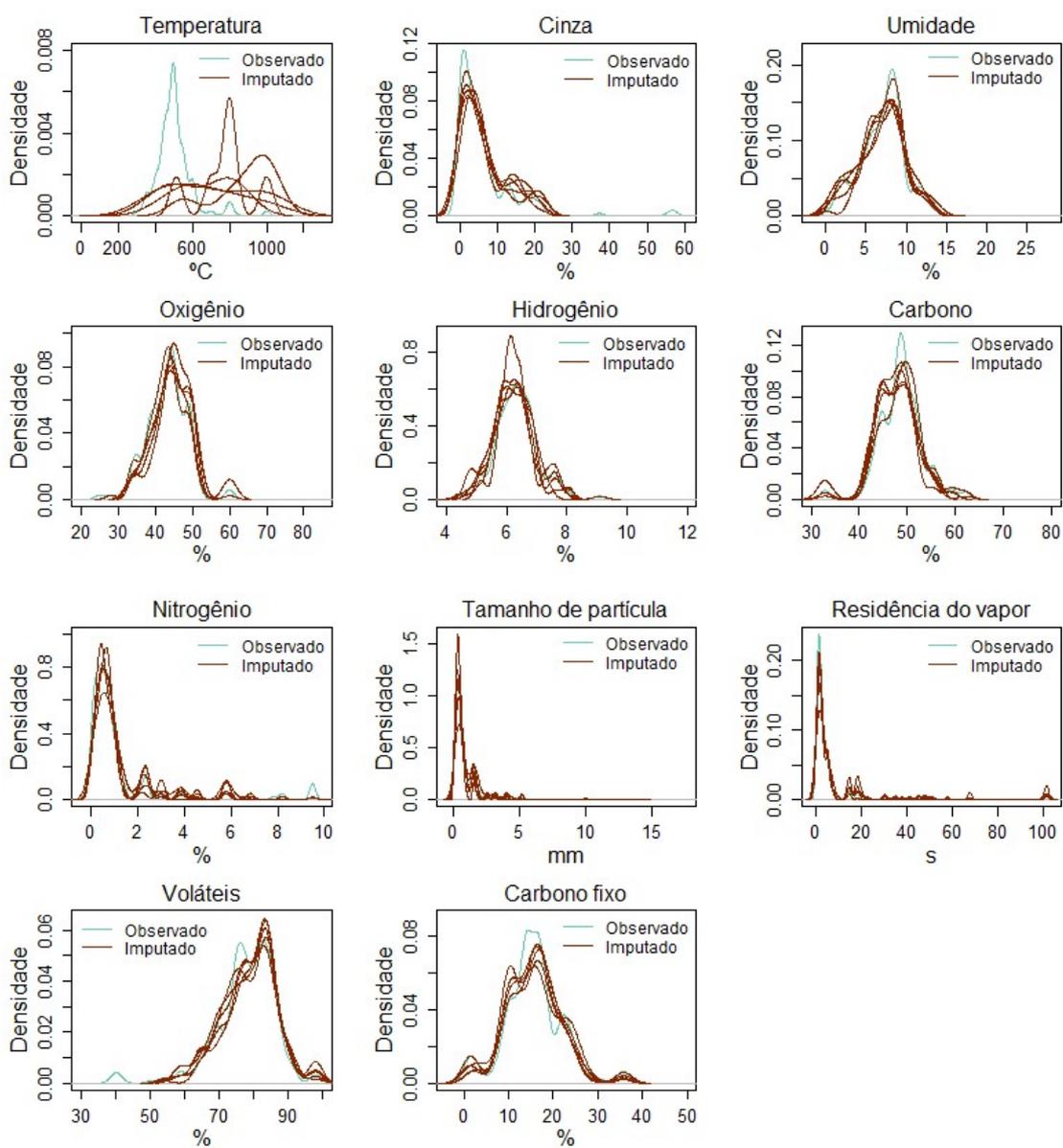
- Distribuição dos dados observados e imputados Midastouch



● -Valores observados; ● -Valores imputados

Fonte: O autor, 2023.

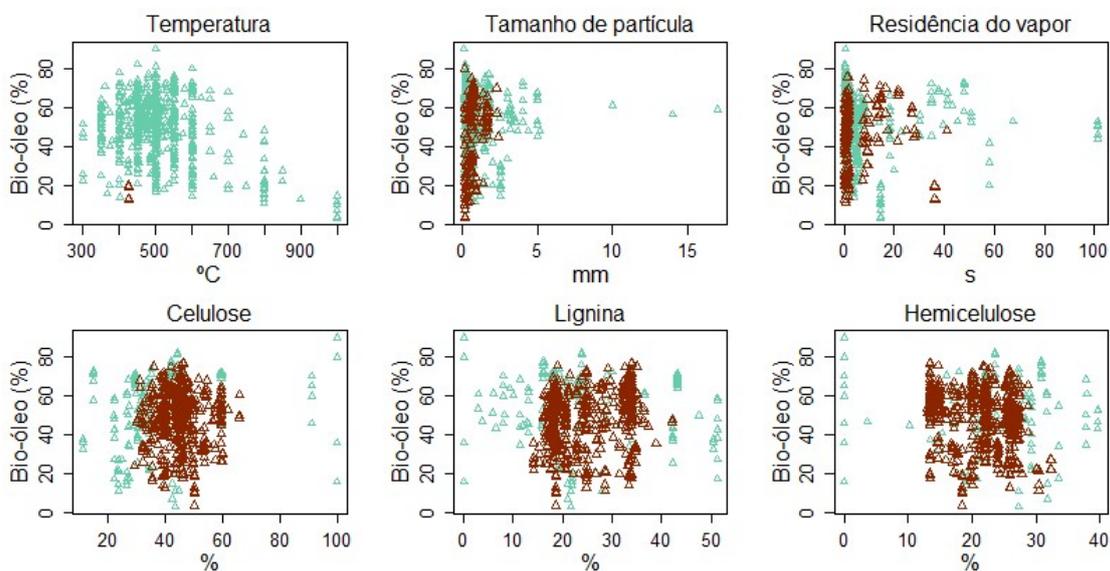
- Densidade dos dados observados e imputados por Midastouch



Fonte: O autor, 2023.

✓ Estudo de Caso VII

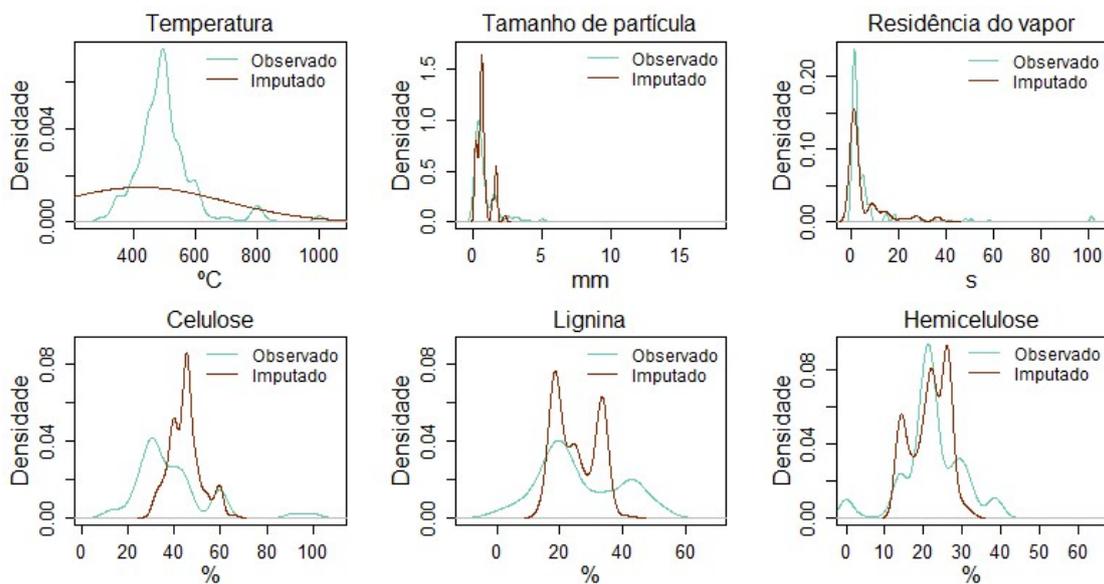
- Distribuição dos dados observados e imputados k-NN



● -Valores observados; ● -Valores imputados

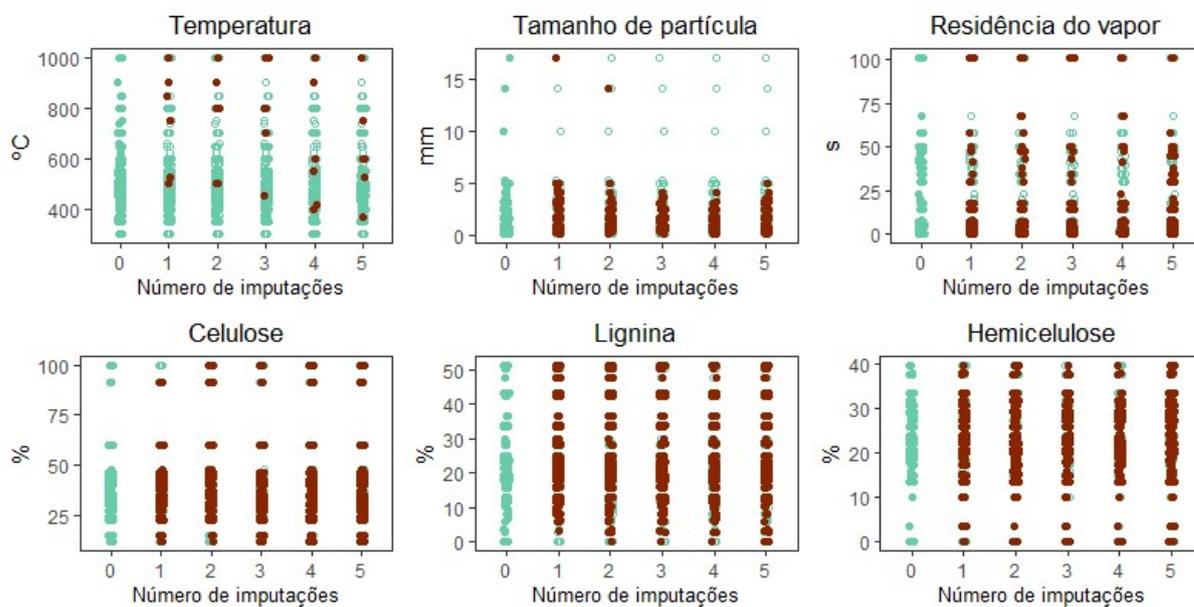
Fonte: O autor, 2023.

- Densidade dos dados observados e imputados por -kNN



Fonte: O autor, 2023.

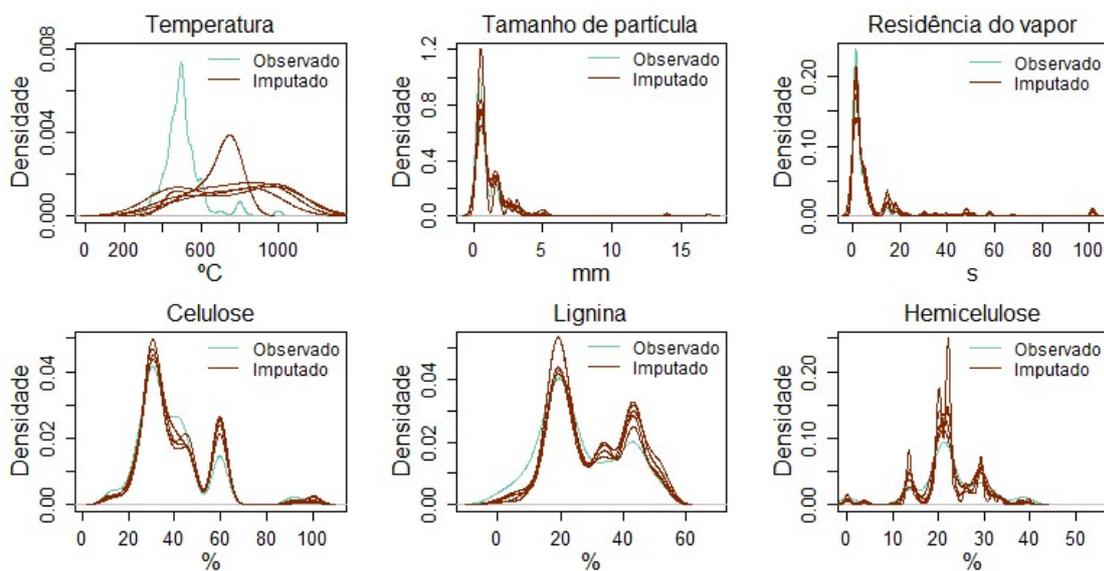
- Distribuição dos dados observados e imputados por RF



● -Valores observados; ● -Valores imputados

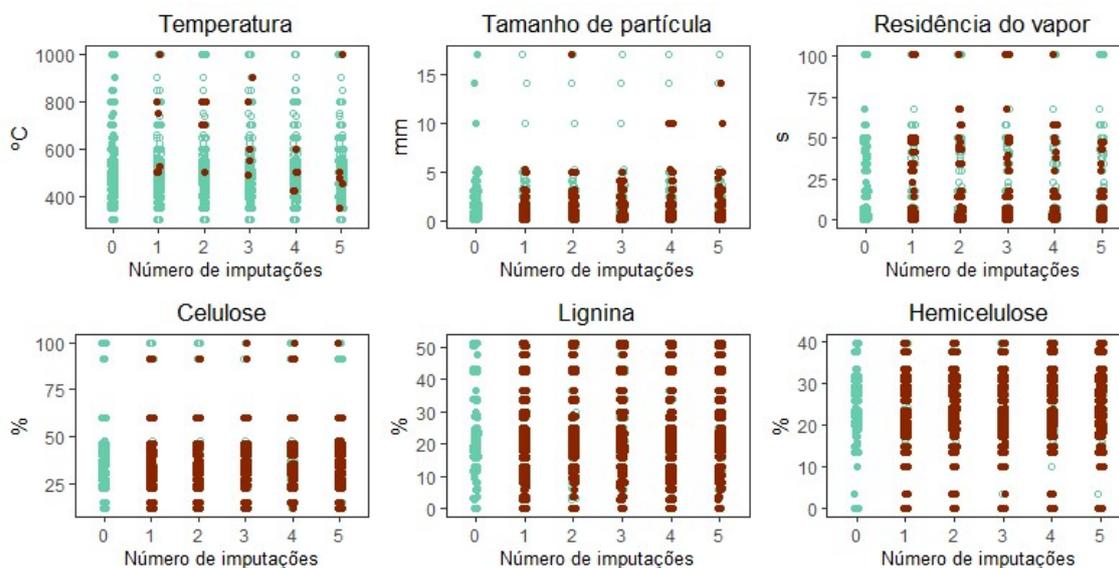
Fonte: O autor, 2023.

- Densidade dos dados observados e imputados por RF



Fonte: O autor, 2023.

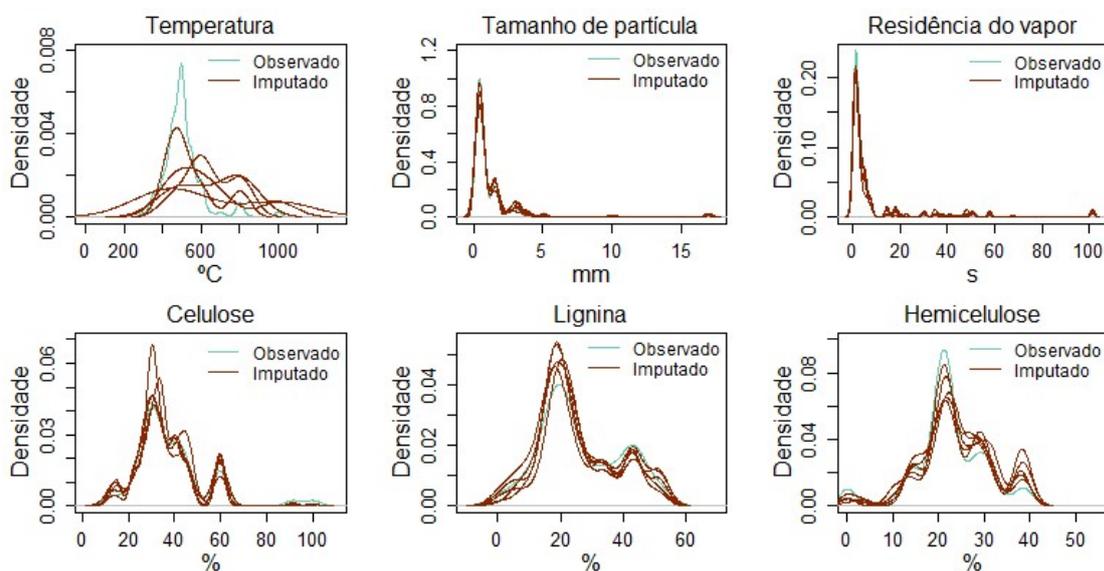
- Distribuição dos dados observados e imputados por Midastouch



● -Valores observados; ● -Valores imputados

Fonte: O autor, 2023.

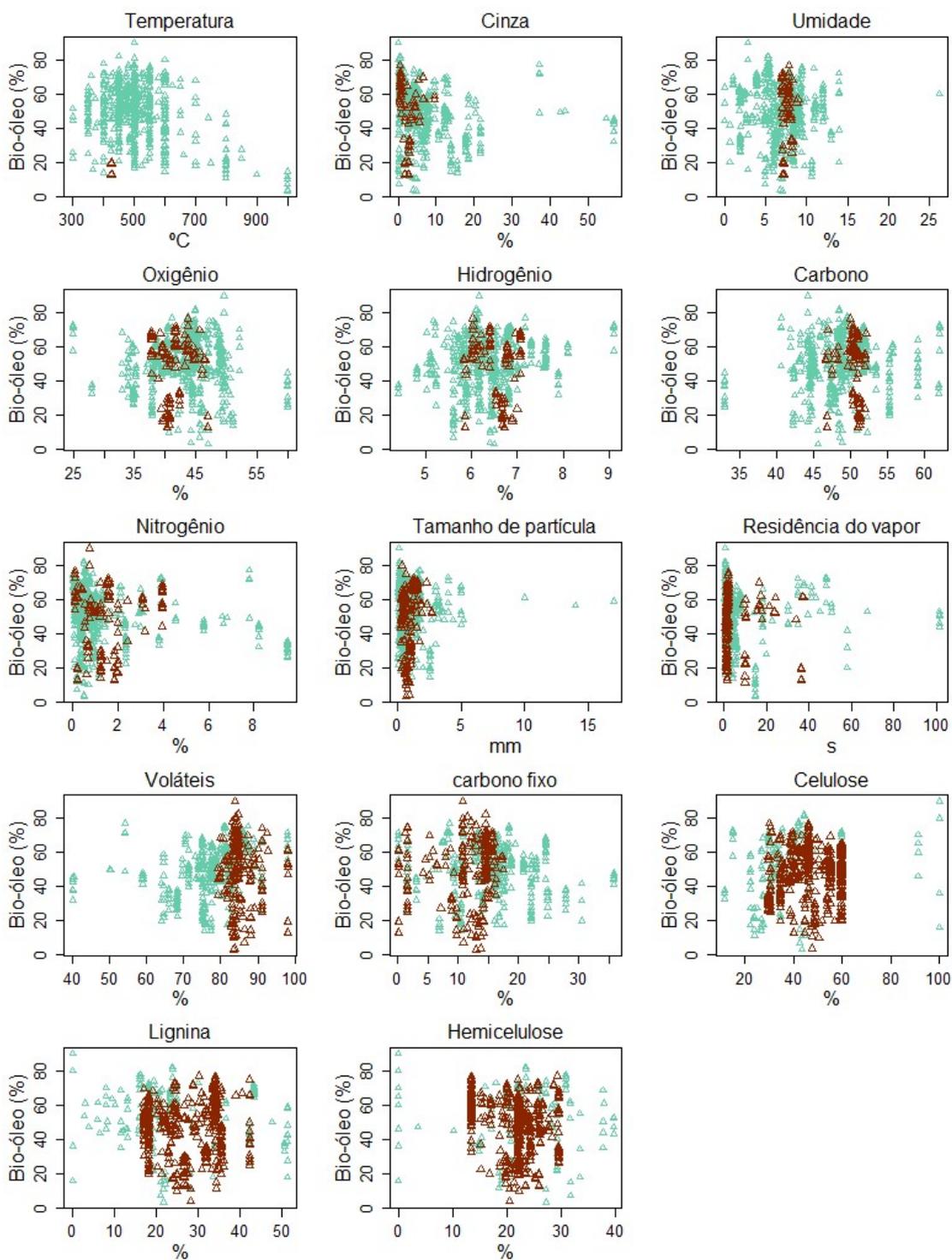
- Densidade dos dados observados e imputados por Midastouch



Fonte: O autor, 2023.

✓ Estudo de Caso VIII

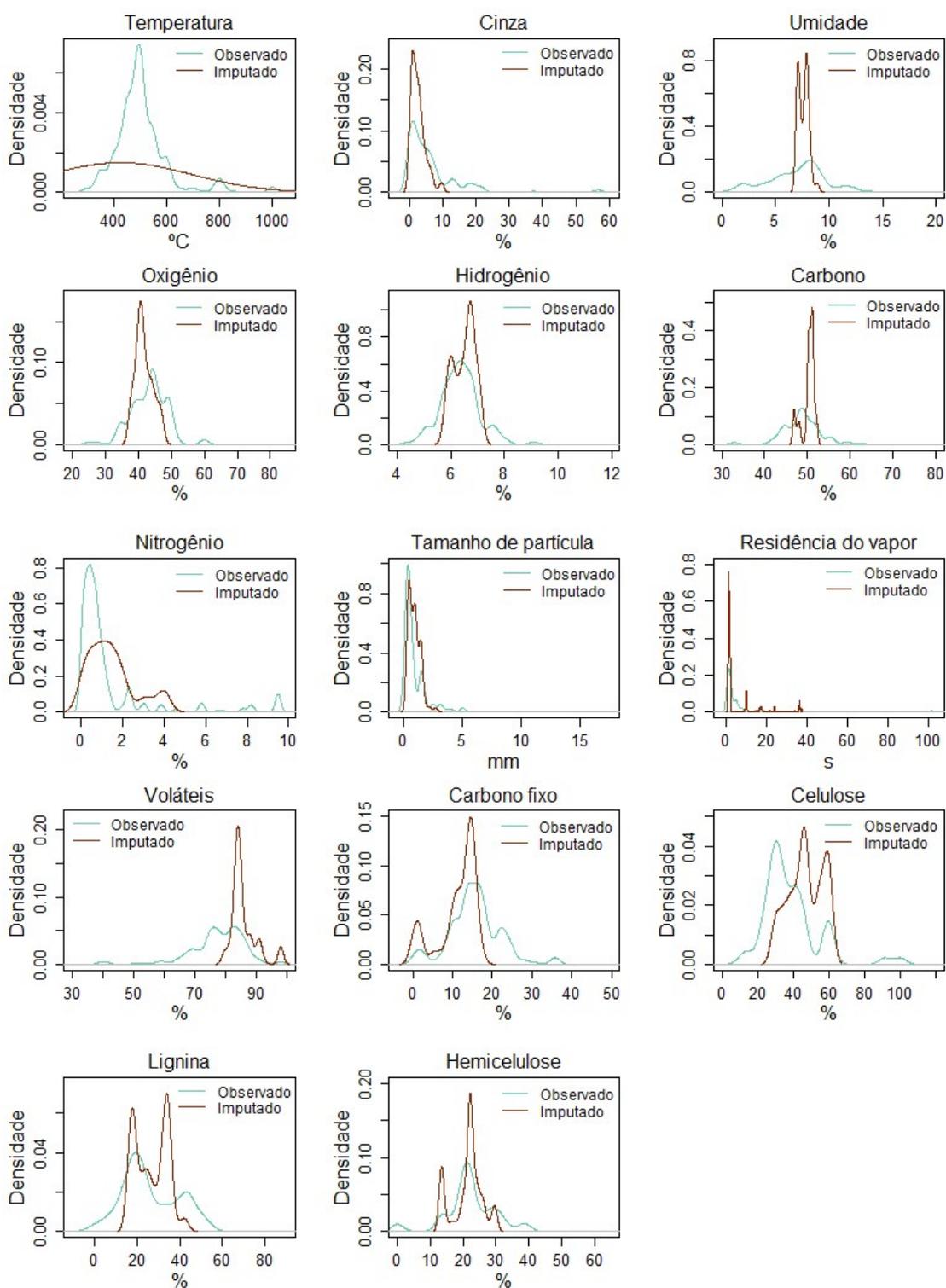
- Distribuição dos dados observados e imputados k-NN



● -Valores observados; ● -Valores imputados

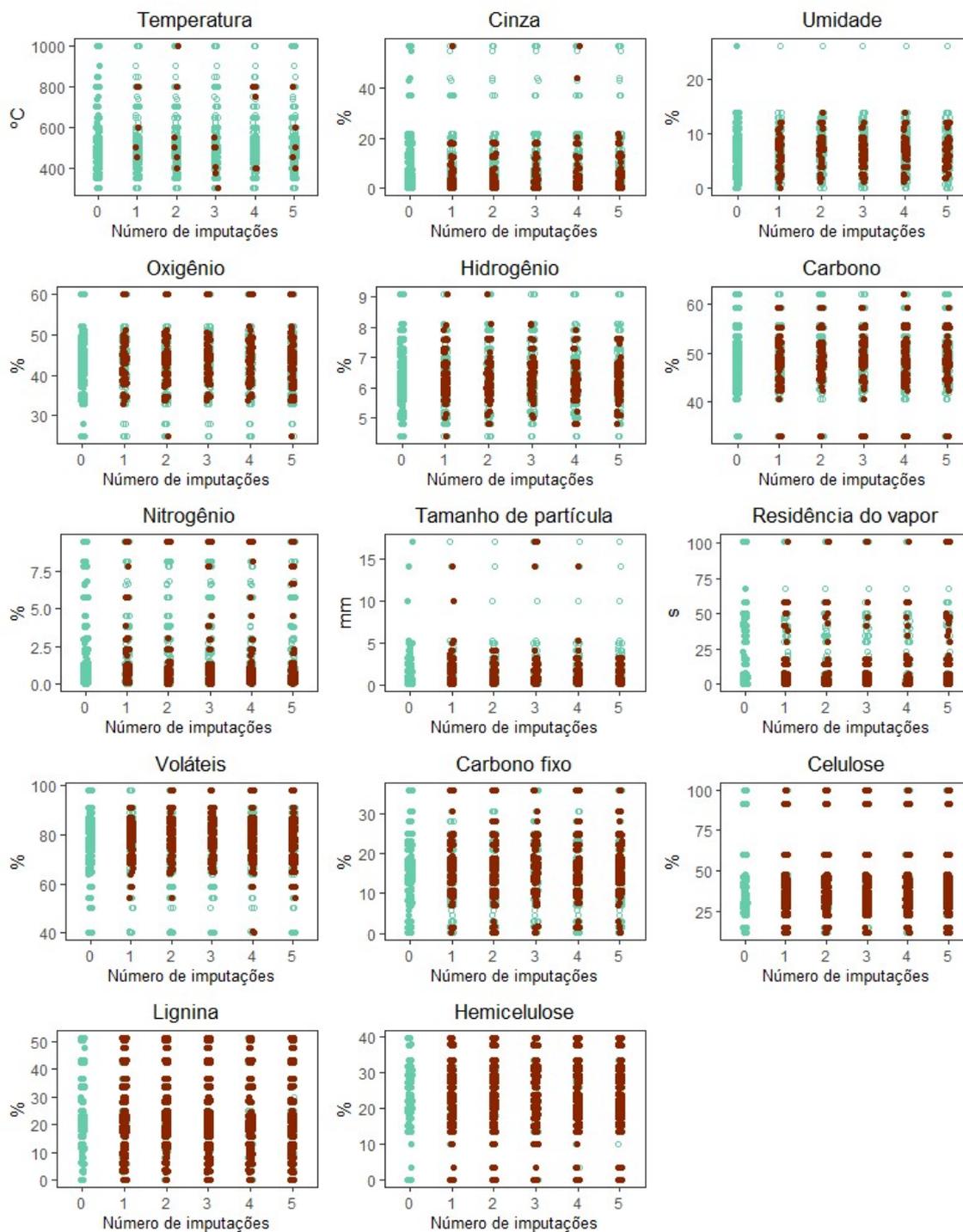
Fonte: O autor, 2023.

- Densidade dos dados observados e imputados por -kNN



Fonte: O autor, 2023.

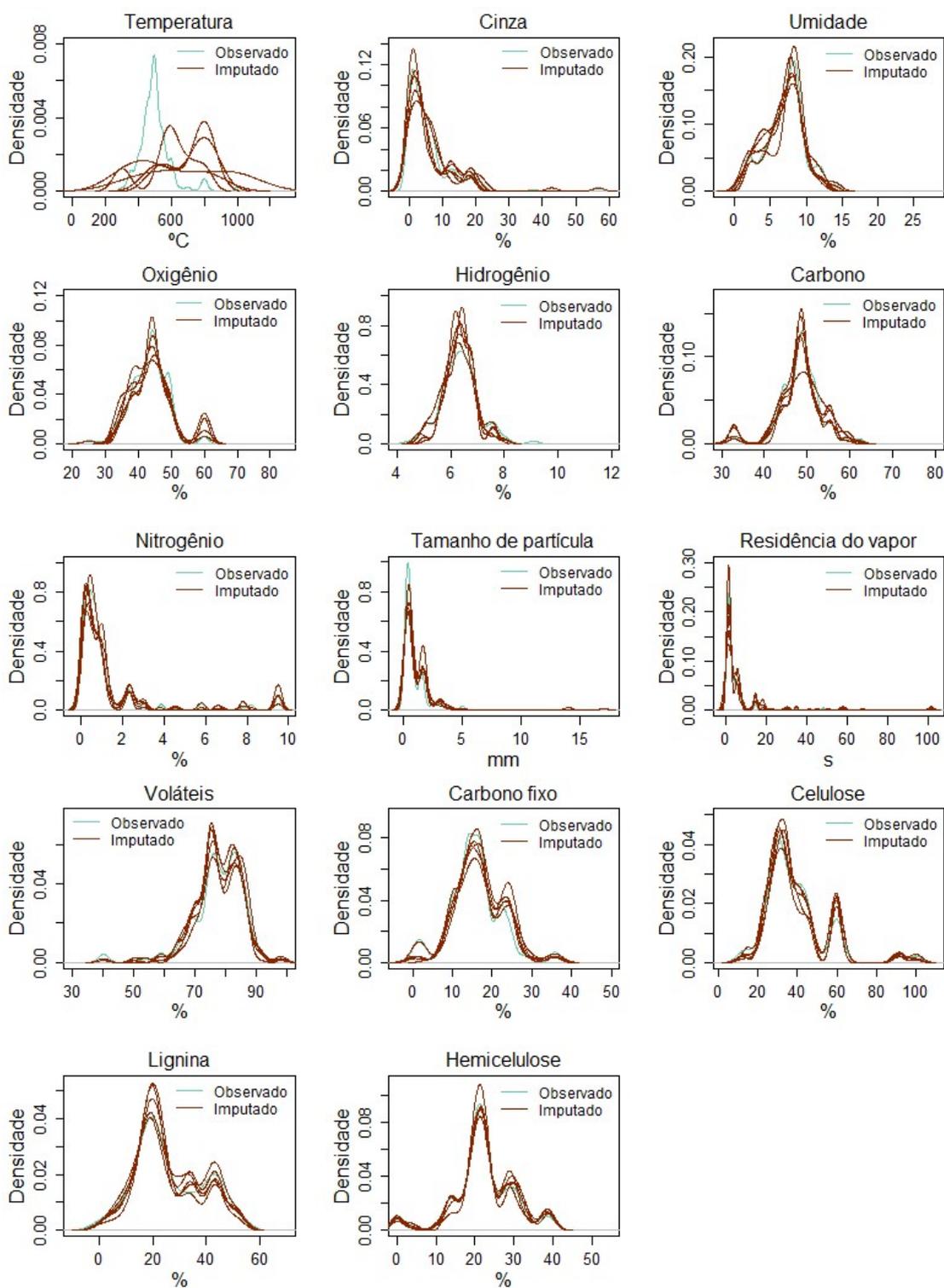
- Distribuição dos dados observados e imputados por RF



● -Valores observados; ● -Valores imputados

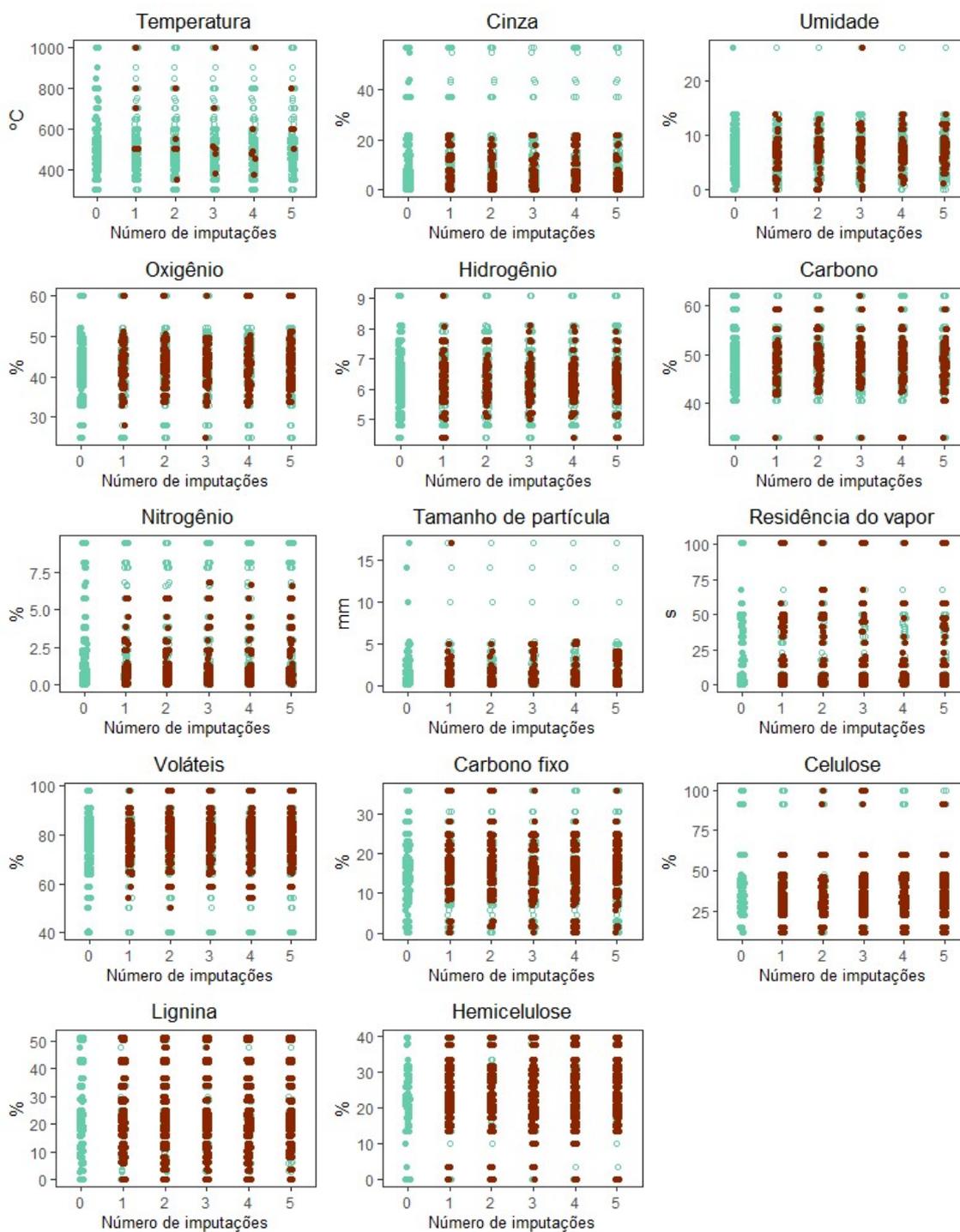
Fonte: O autor, 2023.

- Densidade dos dados observados e imputados por RF



Fonte: O autor, 2023.

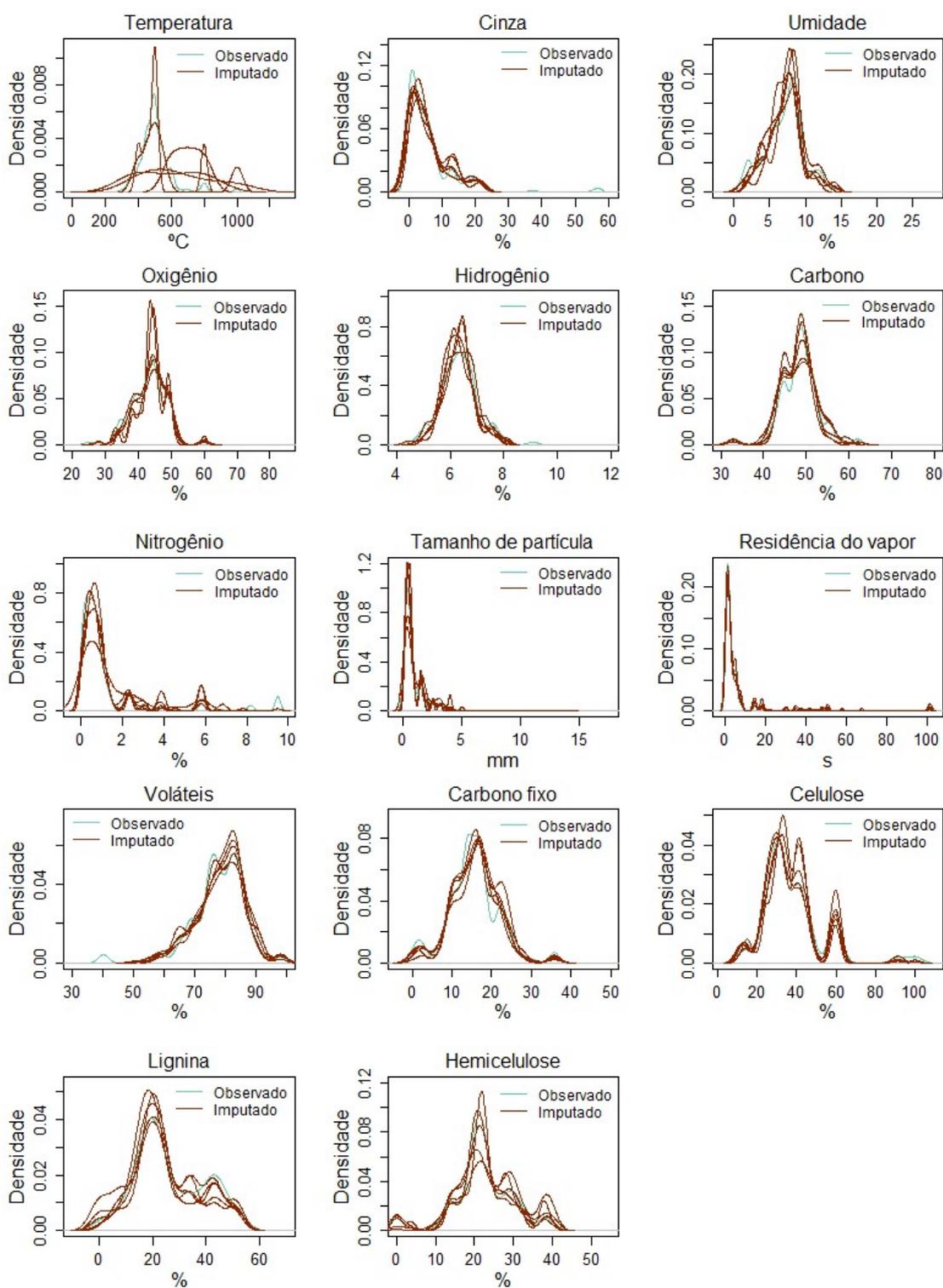
- Distribuição dos dados observados e imputados Midastouch



● -Valores observados; ● -Valores imputados

Fonte: O autor, 2023.

- Densidade dos dados observados e imputados por Midastouch



Fonte: O autor, 2023.