UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO TECNOLÓGICO
DEPARTAMENTO DE ENGENHARIA ELÉTRICA
PROJETO DE GRADUAÇÃO

BÁRBARA SUBTIL LIMA

APLICAÇÃO DE TÉCNICAS DE ANÁLISE DE DADOS EM UM PROCESSO DE SINTERIZAÇÃO

VITÓRIA
2019

BÁRBARA SUBTIL LIMA

APLICAÇÃO DE TÉCNICAS DE ANÁLISE DE DADOS EM UM PROCESSO DE SINTERIZAÇÃO

Parte manuscrita do Projeto de Graduação da aluna **Bárbara Subtil Lima**, apresentado ao Departamento de Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Engenheiro Eletricista.

Orientador: Prof. Dr. Klaus Fabian Côco

VITÓRIA
2019

BÁRBARA SUBTIL LIMA

APLICAÇÃO DE TÉCNICAS DE ANÁLISE DE DADOS EM UM PROCESSO DE SINTERIZAÇÃO

Parte manuscrita do Projeto de Graduação da aluna **Bárbara Subtil Lima**, apresentado ao Departamento de Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Engenheiro Eletricista.

Aprovada em 02, de dezembro de 2019.

COMISSÃO EXAMINADORA:

Prof. Dr. Klaus Fabian Côco
Universidade Federal do Espírito Santo
Orientador

Prof. Dr. Evandro Ottoni Teatini Salles
Universidade Federal do Espírito Santo
Examinador

MSc Ana Paula Miranda Diniz
Universidade Federal do Espírito Santo
Examinadora

Aos meus pais e a minha irmã, meus maiores amores.

Agradeço à Deus por ter sempre guiado os meus caminhos e me abençoado em cada uma das minhas escolhas e trabalhos. Agradeço aos meus pais por sempre me apoiarem, me incentivarem e me mostrarem a cada dia que todos os meus objetivos e sonhos podem ser realizados, independentemente do quão difíceis possam parecer. Agradeço à minha irmã por todo apoio e amparo e pelos abraços que me aconchegam todas as vezes. Agradeço à minha avó e a minha tia Aninha que sempre me colocaram nas suas orações e torceram para que eu concluísse essa caminhada da engenharia com sucesso. Agradeço ao meu namorado por toda a compreensão e apoio ao longo dessa caminhada e pela busca por livros e mais livros. Agradeço a minha família por todo amparo, amor e felicidade proporcionada.

Agradeço aos meus amigos da faculdade, não tenho dúvidas que sem vocês eu não chegaria até aqui. Agradeço aos meus colegas do trabalho que agregaram muitos novos conhecimentos e amizades. Por fim, agradeço ao meu orientador, Klaus, pelo grande amparo, auxílio e plena disponibilidade em me auxiliar na construção de novos conhecimentos e no desenvolvimento deste trabalho além da compreensão e apoio na mudança de temas e escolhas.

RESUMO

Neste trabalho é apresentado o projeto de graduação do Curso de Engenharia Elétrica da Universidade Federal do Espírito Santo (UFES), cujo desenvolvimento se baseia na realização da análise de dados do processo industrial de sinterização, em relação a um de seus principais parâmetros o BTP (do inglês *Burn Through Point*). O BTP apresenta grande importância para o desenvolvimento do processo, pois sua adequação tem grande influência na qualidade do produto, o sínter, e na velocidade do processo de produção, dois fatores relevantes para as empresas onde o processo está inserido. Dessa forma, o presente trabalho propõe a aplicação de técnicas de análise e tratamento de qualidade dos dados, análise descritiva e multivariada das variáveis envolvidas na determinação do BTP de forma a alcançar seu objetivo de propor um novo conjunto de variáveis de importância que podem ser inseridas em redes de inteligência para análises de previsão do parâmetro no processo.

Palavras-chave: Sinterização. Análise de dados. BTP.

LISTA DE FIGURAS

Figura 1 – Esboço do processo produtivo primário das usinas siderúrgicas 19
Figura 2 – Fluxograma esquemático típico da sequência de processamento do minério de ferro 20
Figura 3 – Entradas e saídas de um alto-forno 23
Figura 4 – Alimentação da esteira de sinterização 24
Figura 5 – Esquema do processo de sinterização por máquina contínua 25
Figura 6 – Zonas de sinterização em um processo de sinterização de máquina contínua 27
Figura 7 – Representação esquemática do sistema de exaustão de ar 28
Figura 8 – Posição do BTP em relação as temperaturas das CV 30
Figura 9 – Posição do BTP em relação as temperaturas das CV 31
Figura 10 – Processo de sinterização com adição de monitoramento por câmera infravermelha 32
Figura 11 – Processo de decomposição do sinal via DWT 36
Figura 12 – Processo de decomposição do sinal via MODWT 37
Figura 13 – Sobreposição das bandas de frequência de um processo MODWT 38
Figura 14 – Wavelet Haar (a) Função de escalonamento (b) Função base 40
Figura 15 – *Wavelets* Daubechies (a) db2 (b) db3 (c) db4 (d) db5 (e) db6 (f) db7 (g) db8 (h) db9 (i) db10 41
Figura 16 – *Wavelets* do tipo Symlet (a) sym2 (b) sym3 (c) sym4 (d) sym5 (e) sym6 (f) sym7 (g) sym8 42
Figura 17 – *Wavelets* do tipo Coiflets (a) coif1 (b) coif2 (c) coif3 (d) coif4 (e) coif5 42
Figura 18 – Modelo não linear de um neurônio 56
Figura 19 – Arquiteturas de redes neurais (a) Rede alimentada adiante com uma camada de neurônios (b) Rede alimentada adiante com duas camadas de neurônios (c) Rede recorrente 57
Figura 20 – Comportamento dos dados da máquina de sinterização 86
Figura 21 – Estrutura da topologia da rede NAR aplicada ao presente trabalho (15 estados regressivos) 107

LISTA DE GRÁFICOS

Gráfico 1 – Gráficos de linha (a) Consumo de Cal (b) Base de Coque 66
Gráfico 2 – Gráficos de linha (a) Total Pilha Homogeneizada (b) RDEG na Pilha Homogeneizada 67
Gráfico 3 – Gráficos de linha (a) SFBR no *sínter feed* (b) STFA no *sínter feed* 68
Gráfico 4 – Gráficos de linha (a) Porcentagem +1 mm (b) Relação >1 mm e <0,105 mm Pilha homogeneizada 68
Gráfico 5 – Gráficos de linha (a) Fração Maior que 1 mm (b) Fração Menor que 0,105 mm 69
Gráfico 6 – Gráficos de linha (a) Relação Maior que 1 mm e menor que 0,105 mm (b) Fração de RDEG 70
Gráfico 7 – Gráficos de linha (a) Densidade da Mistura (b) Altura da Camada 71
Gráfico 8 – Gráficos de linha (a) Velocidade de Sinterização (b) Velocidade da máquina de sínter 72
Gráfico 9 – Gráficos de linha temperaturas das Caixas de Vento (a) nº13 (b) nº15 (c) nº 18 (d) nº 19 (e) nº 20 (f) nº 21 73
Gráfico 10 – Gráficos de linha temperaturas das Caixas de Vento (a) nº 22 (b) nº23 74
Gráfico 11 – Gráficos de linha das pressões nas Caixas de Vento (a) nº22 (b) nº23 75
Gráfico 12 – Gráficos de linha das vazões dos exaustores (a) nº 1 (b) nº 2 75
Gráfico 13 – Gráfico de linha da Pressão do Exaustor 76
Gráfico 14 – Gráficos de linha do BTP (a) Temperatura (b) Posição 77
Gráfico 15 – Gráficos de linha (a) CO (b) Produtividade 78
Gráfico 16 – Gráfico de linha Geração de Finos de Retorno 78
Gráfico 17 – Exemplo de momento de parada da esteira de sinterização 86
Gráfico 18 – Formato final dos dados após tratamento dos momentos de parada 90
Gráfico 19 – Período de variação factível na planta analisada 92
Gráfico 20 – Período com ocorrência de *outliers* 92
Gráfico 21 – Exemplo de amostra interpolada (PCHIP) 94
Gráfico 22 – Temperatura e Posição do BTP da amostra (a) 20 (b) 4520 96
Gráfico 23 – Intervalo de variação crescente de temperatura (amostras 3444 a 3494) 97
Gráfico 24 – Intervalo de variação crescente de temperatura (amostras 4517 a 4644) 97
Gráfico 25 – Variações no valor fornecido da temperatura do BTP na amostra (a) 5077 (b) 6077. 98

Gráfico 26 – Variações dos valores do BTP nas amostras (a) 5666 (b) 2414 (c) 1291 (d) 3349 99
Gráfico 27 – Erro percentual associado a (a) Temperatura BTP (b) Posição BTP 100
Gráfico 28 – Comportamento da curva prevista e curva fornecida BTP (a) Temperatura (b) Posição 100
Gráfico 29 – Exemplo de amostras interpoladas a cada (a) 10 min (b) 1 hora..... 102
Gráfico 30 – Exemplo Formato de dados ao fim dos procedimentos..... 103
Gráfico 31 – Exemplo de deslocamento das amostras utilizando MODWT..... 105
Gráfico 32 – Exemplo tratamento de ruído com transformada *Wavelet* 106
Gráfico 33 – Regressão do treino da rede neural da variável "Velocidade da Máquina de Sínter" com dados de (a) Treino (b) Validação (c) Teste (d) Totais 108
Gráfico 34 – Autocorrelação do erro com 20 *bins* da "Velocidade da Máquina de Sínter". 109
Gráfico 35 – Histograma de erro de treino da Rede Neural para "Velocidade da Máquina de Sínter" 110
Gráfico 36 – Exemplo de completude dos dados da variável "Velocidade da Máquina de Sínter" 110
Gráfico 37 – Histograma das componentes principais e suas variáveis associadas 117
Gráfico 38 – Coeficientes ortogonais do componente principal 117
Gráfico 39 – Variáveis obtidas por meio da análise PCA (a) Primeira Componente (b) Segunda Componente 118
Gráfico 40 – Temperatura da CV nº 11 (a) Gráfico de linhas (b) Boxplot antes do tratamento (c) Boxplot após tratamento (d) Histograma antes do tratamento (e) Histograma após tratamento 120
Gráfico 41 – Temperatura da CV nº 13 (a) Gráfico de linhas (b) Boxplot antes do tratamento (c) Boxplot após tratamento (d) Histograma antes do tratamento (e) Histograma após tratamento 120
Gráfico 42 – Temperatura da CV nº 18 (a) Gráfico de linhas (b) Boxplot antes do tratamento (c) Boxplot após tratamento (d) Histograma antes do tratamento (e) Histograma após o tratamento 121
Gráfico 43 – Produtividade (a) Gráfico de linhas (b) Boxplot antes do tratamento (c) Boxplot após tratamento (d) Histograma antes do tratamento (e) Histograma após o tratamento..... 122
Gráfico 44 – Velocidade da Máquina de Sínter (a) Gráfico de linhas (b) Boxplot antes do tratamento (c) Boxplot após tratamento (d) Histograma antes do tratamento (e) Histograma após o tratamento 123

Gráfico 45 – Temperatura da CV nº 19 (a) Gráfico de linhas (b) Boxplot antes do tratamento (c) Boxplot após tratamento (d) Histograma antes do tratamento (e) Histograma após o tratamento 123
Gráfico 46 – Temperatura da CV nº 20 (a) Gráfico de linhas (b) Boxplot antes do tratamento c) Boxplot após tratamento (d) Histograma antes do tratamento (e) Histograma após o tratamento 124
Gráfico 47 – Temperatura da CV nº 21 (a) Gráfico de linhas (b) Boxplot antes do tratamento (c) Boxplot após tratamento (d) Histograma antes do tratamento (e) Histograma após o tratamento 125
Gráfico 48 – Temperatura da CV nº 22 (a) Gráfico de linhas (b) Boxplot antes do tratamento (c) Boxplot após tratamento (d) Histograma antes do tratamento (e) Histograma após o tratamento 125
Gráfico 49 – Temperatura da CV nº 23 (a) Gráfico de linhas (b) Boxplot antes do tratamento (c) Boxplot após tratamento (d) Histograma antes do tratamento (e) Histograma após o tratamento 126
Gráfico 50 – Pressão na CV nº22 (a) Gráfico de linhas (b) Boxplot antes do tratamento (c) Boxplot após tratamento (d) Histograma antes do tratamento (e) Histograma após o tratamento 127
Gráfico 51 – Pressão no exaustor (a) Gráfico de linhas (b) Boxplot antes do tratamento (c) Boxplot após tratamento (d) Histograma antes do tratamento (e) Histograma após o tratamento 127
Gráfico 52 – Emissão de CO (a) Gráfico de linhas (b) Boxplot antes do tratamento (c) Boxplot após tratamento (d) Histograma antes do tratamento (e) Histograma após o tratamento..... 128
Gráfico 53 – Geração de finos de retorno (a) Gráfico de linhas (b) Boxplot antes do tratamento (c) Boxplot após tratamento (d) Histograma antes do tratamento (e) Histograma após o tratamento 129
Gráfico 54 – Gráfico de linhas comparativo (a) Temperatura do BTP (b) Posição do BTP 130
Gráfico 55 – Amostras comparativas do BTP (a) Temperatura (b) Posição 130

LISTA DE QUADROS

Quadro 1 – Variáveis do processo 60
Quadro 2 – Intervalos de parada do processo de sinterização 88
Quadro 3 – Intervalos em que há registro de variação crescente de temperatura 96
Quadro 4 – Aplicação da MODWT para remoção de ruídos associados 104
Quadro 5 – Correlações de Pearson entre as variáveis de composição da mistura 113
Quadro 6 – Correlações de Pearson entre as variáveis de controle e indicação do processo 115
Quadro 7 – Correlações de Pearson entre as variáveis de indicação do processo 115

LISTA DE ABREVIATURAS E SIGLAS

ABM	Associação Brasileira de Metalurgia, Materiais e Mineração
BTP	*Burn Throught Point*
CV	Caixa de vento
DWT	*Discrete Wavelet Transform*
DWTP	*Discrete Wavelet Packet Transform*
FA	*Factor Analysis*
FT	*Fourrier Transform*
HPF	*High Pass Filter*
IEC	*International Electrotechnical Commission*
IEEE	*Institute of Electrical and Electronics Engineers*
IFES	Instituto Federal do Espírito Santo
ISO	*International Organization for Standardization*
KMO	*Kaiser-Meyer-Olkin*
LPF	*Low Pass Filter*
MODWT	*Maximal Overlap Discrete Wavelet Transform*
MODWPT	*Maximal Overlap Discrete Wavelet Packet Transform*
MRA	*MultiResolution Analysis*
MSA	*Measure of Sampling Adequacy*
MSE	*Mean Squared Error*
NAR	*Nonlinear AutoRegressive model*
PCA	*Principal Component Analysis*
PCHIP	*Piecewise Cubic Hermite Interpolating Polynomial*
RDEG	Quantidade de finos de minério do peneiramento utilizados no processo
RNA	Rede Neural Artificial
SFBR	Percentual de Minério de Ferro proveniente da mina de Brucutu
STFA	Percentual de Minério de Ferro proveniente da mina de Fábrica
STFT	*Short-Time Fourrier Transform*
UFES	Universidade Federal do Espírito Santo
WT	*Wavelet Transform*

LISTA DE SÍMBOLOS

$F[x(t)]$	Transformada de Fourrier de um sinal contínuo
$X(f_s)$	Sinal no domínio da frequência
$x(t)$	Sinal no domínio do tempo
f_s	Frequência do sinal
t	Tempo
τ	Instante de tempo onde é centrada a janela
$W(t-\tau)$	Janela centrada no instante de tempo τ
$\psi(t)$	*Wavelet* mãe
$\psi_{a,b}(t)$	Funções *wavelet* baseadas na *wavelet* mãe
a	Parâmetro de escala
b	Parâmetro de deslocamento
$CWT(a,b)$	Transformada de *wavelet* em tempo comtínuo
$\psi_{j,r}[m]$	Funções *wavelet* no tempo discreto
m	Variável independente discreta
X	Variável aleatória
$X_{a,i}$	Cada elemento registrado de uma determinada variável X
Y	Variável aleatória
$Y_{b,i}$	Cada elemento registrado de uma variável Y
$\overline{X_a}$	Média dos dados da variável X
$\overline{Y_b}$	Média dos dados da variável Y
n	Quantidade total dos dados a serem analisados
Z	Matriz de variáveis com dados associados
$\tilde{Z}$	Matriz de variáveis com dados associados normalizados
CP	Matriz de componentes principais
A	Matriz de autovetores
x_j	Sinais de entrada
w_{kj}	Pesos sinápticos
k	Índice do neurônio ao qual o peso se refere
b_k	Limiar correspondente ao neurônio k
φ	Função de ativação
y_k	Sinal de saída

$\hat{y}(l)$	Saída estimada do modelo no instante l
$y(l)$	Saída do modelo de uma rede NAR
n_u	Número de atrasos na memória
$R_{\varepsilon\varepsilon}$	Função de autocorrelação
E	Valor esperado

SUMÁRIO

1 INTRODUÇÃO 16
1.1 Objetivos 17
1.1.1 Objetivo geral 17
1.1.2 Objetivos específicos 17
2 USINAS SIDERÚRGICAS E O PROCESSO PRODUTIVO DO FERRO PRIMÁRIO 19
2.1 Processo de Sinterização 23
2.1.1 Processo de sinterização contínua por esteira rolante 24
2.1.2 Zonas de sinterização 26
2.1.3 O processo operacional de sinterização e o BTP 29
2.2 Revisão Bibliográfica 32
3 TRANSFORMADA *WAVELET* E SUA APLICAÇÃO NA FILTRAGEM DE SINAIS 34
3.1 Transformada *Wavelet* Discreta 34
3.2 Transformadas *Wavelet* Não-decimadas e Invariantes no Tempo 36
3.3 Procedimento de Aplicação de *Wavelets* como Filtros Lineares 38
3.4 Famílias de *Wavelets* 39
3.4.1 *Wavelet* de Haar 39
3.4.2 *Wavelet* de Daubechies 40
3.4.3 *Wavelets* Symlets e Coiflets 41
4 ANÁLISE DE DADOS 43
4.1 Análise de Qualidade dos Dados 43
4.2 Análise Descritiva 45
4.3 Análise Estatística Multivariada 46
4.3.1 Técnicas de interdependência 47
4.3.1.1 Adequação da Utilização da Análise Fatorial 48
4.3.1.2 Extração de Fatores 51
5 REDES NEURAIS ARTIFICIAIS 55
6 PLANTA EM ESTUDO E VARIÁVEIS EM ANÁLISE 59
7 EXAME DOS DADOS ORIGINAIS 65
7.1 Parâmetros Relativos à Composição da Mistura de Entrada do Processo 65
7.2 Parâmetros de Controle do Processo de Sinterização 70

7.3 Indicadores de Controle do Processo de Sinterização 72
7.4 Indicadores de Saída do Processo de Sinterização 77
8 ANÁLISE DA QUALIDADE DOS DADOS 79
8.1 Acurácia 79
8.2 Completude 82
8.3 Consistência 83
8.4 Credibilidade 83
8.5 Atualidade 83
8.6 Acessibilidade 83
8.7 Compreensibilidade 84
9 TRATAMENTO DOS DADOS 85
9.1 Tratamento por Eliminação de Conjunto de Dados 85
9.2 Remoção de *Outliers* 90
9.2.1 Tratamento por interpolação de amostras do entorno de *outliers* isolados 93
9.3 Cálculo da Posição e Temperatura do BTP 94
9.4 Reamostragem dos Dados para Suavização de Variações 101
9.5 Tratamento de Ruídos com Transformada *Wavelet* 103
9.6 Tratamento de Completude de Dados 106
10 ANÁLISE DE FATORES 112
10.1 Análise de Correlação Linear de Pearson 112
10.2 Estatística MSA 116
10.3 Análise de Componentes Principais 116
11 VARIÁVEIS TRATADAS 119
11.1 Variáveis Reamostradas a cada 1 hora 119
11.2 Variáveis Reamostradas a cada 10 minutos 122
11.3 Temperatura e Posição do BTP 129
12 CONCLUSÃO E SUGESTÕES PARA TRABALHOS FUTUROS 131
REFERÊNCIAS BIBLIOGRÁFICAS 133

1 INTRODUÇÃO

Em meio a um desenvolvimento industrial constante e um mercado competitivo, indústrias buscam continuamente sistemas cada vez mais precisos, eficientes, otimizados, dinâmicos, seguros e por um menor custo. Para que essa dinâmica ocorra de maneira eficaz, as tomadas de decisões que são realizadas durante a execução dos processos devem ser assertivas e efetuadas de forma a atingir sua melhor qualidade e capacidade. A assertividade dessas deliberações pode ser aperfeiçoada por meio da análise de medições de variáveis do processo e monitoramento de cada uma de suas etapas.

A análise de dados de um processo pode, por exemplo, tornar possível acompanhar o comportamento de variáveis e suas tendências, observar o comportamento de erros no processo e suas causas, além de avaliar as condições físicas dos equipamentos de trabalho e condições estruturais do processo. A geração de informação anteriormente descrita pode, então, ser tomada como embasamento para, por exemplo, prever o comportamento da qualidade do processo diante de alguma mudança de controle ou tomada de decisão.

A necessidade da precisão de tomada de decisões se faz presente em sistemas que precisam de acompanhamento periódico e que, por pequenas alterações de suas variáveis, pode prejudicar a qualidade do produto gerado e o seu funcionamento adequado como um todo. Esse é o caso da sinterização, um dos processos que ocorrem nas usinas siderúrgicas na área produtiva do ferro e aço, procedimento ao qual é submetido o minério de ferro de granulação fina.

Na planta operacional desse processo, há uma constante busca por assertividade da fusão parcial do sínter. Para isso, é necessário identificar o instante de tempo e a posição ao longo da esteira no qual se finaliza a formação do material alvo, ponto de queima (BTP, do inglês *Burn Throught Point*). A não assertividade desse parâmetro pode gerar consequências diretas a produtividade e qualidade da sinterização.

Caso esse ponto seja atingido, previamente a um ponto alvo determinado, não há um aproveitamento máximo da esteira de sinterização e, em decorrência disso, uma diminuição da capacidade produtiva da planta. Caso esse ponto seja atingido após o ponto alvo, é possível que o sínter não diminua sua temperatura de forma adequada para inserção na máquina de

resfriamento e gere danos estruturais. É possível, além disso, não formar o produto de forma adequada e qualidade, gerando a necessidade de reprocessamento (SAIZ; POSADA, 2013).

Os resultados e variações historicamente presentes na planta são registrados e armazenados temporalmente. Por meio deles, e de conhecimento adicional do processo, é possível entender a dinâmica operacional da planta, monitorar e verificar causas e influências processuais em momentos de falha operacional. Além disso, é possível analisar a aplicação de previsibilidade aos dados e objetivos chave da operação.

Entender a dinâmica dos processos e operações da planta, estudar o comportamento das variáveis, seus cálculos, constatar as correlações entre seus dados e a relação adequada para obtenção do BTP torna um meio apto a tomada de decisões e promove um alinhamento do processo em relação a suas demandas.

No presente trabalho será realizado um estudo conceitual, uma análise de dados históricos de variáveis, atentando-se ao seu comportamento e variação. Será aplicado às variáveis um tratamento, associação, análise de suas relações com o desempenho do processo e a garantia de uma adequada fusão parcial do sínter. Por fim, serão obtidos dados aptos à posteriores aplicações através de Modelos de Aprendizados de Máquina.

1.1 Objetivos

1.1.1 Objetivo geral

O objetivo geral do presente trabalho é realizar a análise dos dados de um banco, composto por variáveis de uma planta de sinterização, por meio de metodologias descritivas e multivariadas, com o propósito de observar o comportamento do sistema e as correlações entre suas variáveis em relação ao BTP do processo e, assim, obter, ao fim das análises, um novo conjunto de variáveis que servirão de entrada para a elaboração de Modelos de Aprendizagem de Máquina.

1.1.2 Objetivos específicos

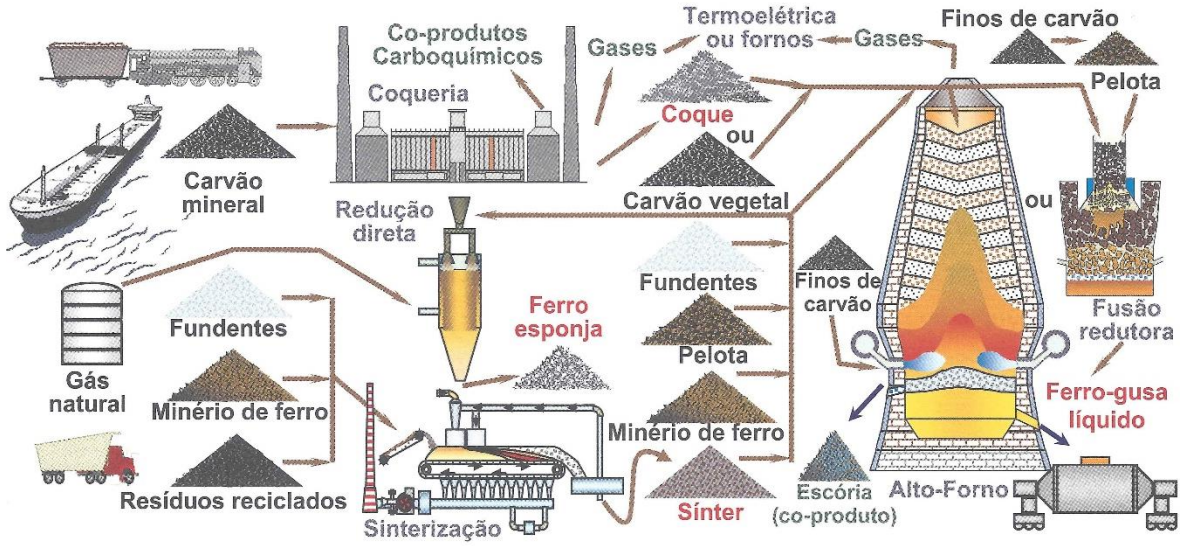
O presente trabalho apresenta como objetivos específicos realizar um estudo do processo de sinterização e os parâmetros monitorados de forma a compreender a dinâmica do processo

operacional observando, diretamente, suas relações com o ponto do final da fusão parcial do BTP. Serão aplicadas, em seguida, técnicas de tratamento e análises qualitativas, descritivas e multivariadas aos dados de 36 variáveis do processo de sinterização. Por fim selecionou-se e criou-se um novo conjunto de variáveis para ser utilizada na construção de aprendizados de máquina como auxílio para estimativa do BTP.

2 USINAS SIDERÚRGICAS E O PROCESSO PRODUTIVO DO FERRO PRIMÁRIO

Usinas siderúrgicas são compostas por variados processos com a finalidade de produção de ferro e aço. Suas principais matérias primas são, comumente, o minério de ferro, o carvão mineral e o calcário (ARCELORMITTAL, [20--?]). O esboço do processo produtivo primário, isto é, até a produção do ferro-gusa, pode ser observado por meio da Figura 1.

Figura 1 – Esboço do processo produtivo primário das usinas siderúrgicas

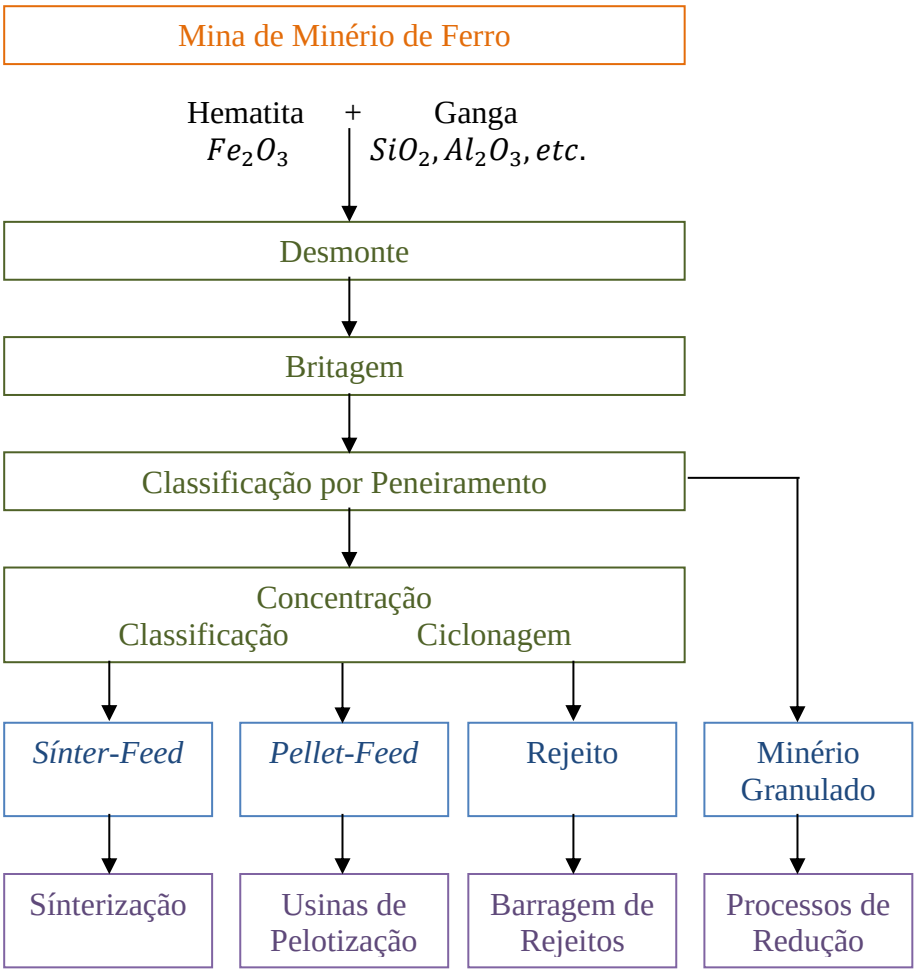


Fonte: Rizzo (2007).

O minério de ferro destinado a cada um dos processos apresentados no esboço possui diferentes granulometrias. O corpo mineral extraído da mina é inicialmente fragmentado por meio, ou não, de explosivos, processo denominado desmonte. Logo em seguida, é realizado o processo de britagem para efetuar a liberação de hematita da ganga, impurezas não desejáveis que se encontram associados ao minério (IFES, [20--?]).

O material é então classificado em relação a sua granulometria por meio de peneiras, com o objetivo de graduar as partículas em relação ao seu tamanho, e separado da ganga por meio de um processo de indução de rotação de gases dentro de um ciclone, denominado ciclonagem (IFES, [20--?]). Em seguida o material é classificado em relação à sua concentração e destinado a processos adequados à sua granulometria, conforme é exemplificado por meio do fluxograma apresentado na Figura 2.

Figura 2 – Fluxograma esquemático típico da sequência de processamento do minério de ferro



Fonte: Rizzo (2007).
Nota: Adaptado pelo autor.

O *sínter-feed* compõe minérios de granulometria entre 6,35 mm e 0,149 mm e 63% a 65% de teor de ferro. O *pellet-feed* é composto por minério de granulometria inferiores que 0,149 mm e 65% a 69% de teor de ferro. O minério granulado, por fim, apresenta granulometria maior que 6,35 mm e 66% a 68% de teor de ferro em sua composição. Os rejeitos do processo, como SiO_2 e Al_2O_3 são comumente depositados em barragens que os abrigam (RIZZO, 2007).

O minério granulado pode ser enviado diretamente aos altos-fornos, ou ainda podem sofrer algum tipo de redução de granulometrias. Ao *sínter-feed* e *pellet-feed*, são aplicados dois processos de remodelação das matérias-primas de forma que atendam às necessidades dos altos-fornos, a sinterização ou pelotização. Os processos citados compõem um grupo de processos de aglomeração a quente de minério de ferro de granulação fina, fundentes e adições, de forma

a viabilizar os processos metalúrgicos e facilitar o transporte adequado do produto (RIZZO, 2007).

Os processos de aglomeração surgiram a partir da necessidade de aproveitamento de minérios de granulação fina originados da extração, manuseio e transporte do minério ou por meio de processos de cominuição aplicados a minérios de baixo teor, de forma a permitir um aumento da concentração de ferro no material (IFES, [20--?]). O comportamento mais eficiente e adequado do sínter e das pelotas no processo de redução nos altos-fornos em relação ao minério granulado duro, fez com que os processos de aglomeração se mantivessem presentes nas usinas siderúrgicas. O minério de ferro de maior granulometria passou a ser, em parte, cominuído e adicionado como entrada aos processos de sinterização e pelotização. O minério fino, que é adquirido junto ao minério granulado, continuou a ser aproveitado, e não descartado.

Outro processo de remodelação das matérias-primas para atender as exigências dos altos-fornos é a coqueificação. A coqueificação é um processo de decomposição térmica do carvão mineral ao abrigo do ar. Uma matéria volátil é desprendida da matéria sólida e obtém-se, por fim, um resíduo sólido, poroso e carbonoso, o coque (FILHO, 2007). O produto gerado é a principal fonte de energia em todo o controle de processo nos altos-fornos e, ao mesmo tempo, elemento redutor do minério de ferro, permitindo a separação do minério de ferro e seu óxido.

Para uma adequada operação nos altos-fornos precisa-se de combustível sólido que contenha um alto teor de carbono, apresente baixa reatividade (carboreatividade) e baixo teor de enxofre, e que forneça a energia térmica necessária às reações químicas de redução do minério de ferro (MOURÃO et al., 2007). O coque produzido atende todas as necessidades citadas. A granulometria e resistência mecânica desse material devem ser adequados para permitir a ascensão dos gases gerados no processo dos altos-fornos (ARAÚJO, 1967).

Os materiais, após devido processamento, são encaminhados para os altos-fornos, onde é produzido o ferro-gusa a partir de quatro matérias primas: sínter, minério de ferro, pelotas e fundentes (ARECELORMITTAL, [20--?]).

O alto-forno é um reator metalúrgico destinado a produzir ferro-gusa em estado líquido, a uma temperatura em torno de 1500°C, para suprir os requisitos de quantidade e qualidade requeridas

para os processos produtivos subsequentes. Esse equipamento pode ser considerado um trocador de calor e de massa, de contracorrente, onde o calor é transferido do gás para a carga de matéria prima e o oxigênio passa da carga para o gás. O gás sobe pelo interior do forno enquanto a carga metálica e o combustível descem. A natureza das reações, em contracorrente, torna o processo como um todo extremamente eficiente (GEERDES; TOXOPEUS; VLIET, 2009).

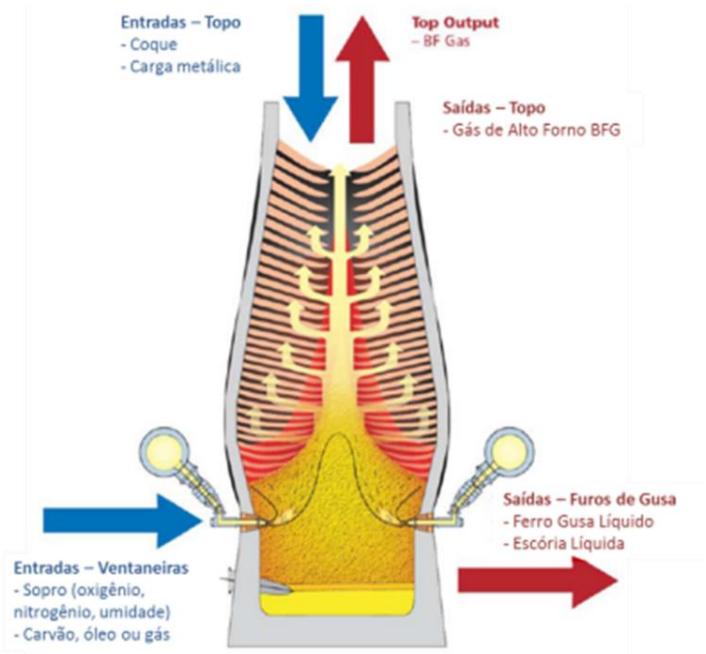
Conforme ilustrado na Figura 3, a carga metálica constituída por minério grosso, sínter e pelota, é carregada na parte superior do forno de forma intercalada com o combustível em estado sólido. Esse combustível é o coque siderúrgico ou carvão vegetal, responsável pelo processo de redução do ferro, que ocorre com o sopro de ar quente na parte inferior do forno. A maioria dos altos–fornos atuais utiliza injeção auxiliar de combustíveis tais como carvão pulverizado, gás natural, óleo e até rejeitos tais como plásticos. Qualquer material com alto percentual de hidrocarbonetos, desde que satisfeitas algumas premissas operacionais, pode ser injetado no forno como combustível. A queima do coque gera gases redutores em altas temperaturas. Esses gases ascendentes em contato com a carga de minérios descendentes reagem, causando redução e fusão do minério, dando origem ao ferro gusa, que é o principal produto do alto-forno, e a escória. Ambos são produzidos em estado líquido em temperaturas entre 1450ºC e 1520°C (MOURÃO et al., 2007).

Deve-se ressaltar o papel fundamental da seleção do material que será carregado no forno, assim como sua distribuição interna. A granulometria das partículas estará diretamente relacionada com a permeabilidade das camadas de material formadas dentro do forno, possibilitando a formação do perfil gasoso adequado para redução (GEEDERS; TOXOPEUS; VLIET, 2009). A distribuição do material realizada a partir do topo do forno também contribui para a redução adequada e é feita de acordo com o material recebido. Variações de materiais exigem uma mudança na distribuição, sempre visando o melhor escoamento do gás redutor, buscando a obtenção de todas as reações químicas necessárias para redução.

A remoção do produto é feita com a abertura e fechamento dos furos de gusa através de equipamentos adequados, permitindo o vazamento em canais que destinam a produção aos carros torpedo, que fazem a logística do ferro gusa. O aço líquido, então, pode sofrer adição de outros componentes e ser enviado para processos de lingotamento, que têm como produto aço

em placas. Essas placas podem ser enviadas diretamente para comercialização, ou podem ser, antes, transformadas em bobinas, a partir de processos de laminação, finalizando assim o processo de produção das usinas siderúrgicas (ARCELORMITTAL, [20--?]).

Figura 3 – Entradas e saídas de um alto-forno



Fonte: Geerdes; Toxopeu; Vliet (2009).

2.1 Processo de Sinterização

“O processo de sinterização é um processo de aglomeração de finos de minério de ferro, finos de fundentes (cal, calcário e quartzo), finos de combustíveis sólidos (carvão mineral do tipo antracito, coque, carvão vegetal e coque de petróleo) e finos de rejeitos ou resíduos industriais” (RIZZO, 2007, p. 42). Os materiais finos anteriormente citados comumente têm origem natural, ou são finos decorrentes dos processos de produção da própria usina siderúrgica.

O processo de sinterização se inicia com a mistura dos finos anteriormente citados em equipamentos próprios de forma que a sua saída seja homogênea. A mistura é então direcionada à máquina de sinterização, em modo de panela ou esteira contínua. Nesse equipamento a temperatura é elevada, entre 1200ºC e 1400°C, resultando uma fusão parcial dos componentes. O produto da fusão parcial é o sínter (RIZZO, 2007).

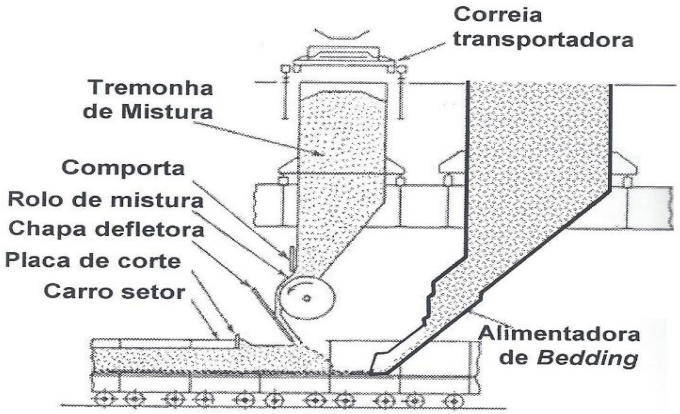
As propriedades do sínter e a dosagem dos seus componentes são definidas de acordo com as necessidades de processo dos altos-fornos. Dessa forma, o sínter deve apresentar características como boa resistência mecânica, elevado teor de ferro, alta porosidade para permeabilidade dos gases no alto-forno, alta redutibilidade dos óxidos contidos no minério, baixa umidade para aumentar o rendimento energético do material e baixa quantidade de impurezas, promovidas pelas reações dos fundentes adicionados na mistura. As características finais do material buscam evitar problemas de permeabilidade dos gases ascendentes no alto-forno e aumenta da redutibilidade dos óxidos contidos no minério (GALIAZZI et al., 2014; RIZZO, 2007).

Ao mesmo tempo que a qualidade de saída do sínter do processo é importante, deve-se levar em conta nas definições do processo um menor custo, segurança operacional e agilidade. Atualmente, o modelo de máquina de sinterização mais comum e que garante maior agilidade no processo de grandes empresas são modelos ininterruptos do tipo esteira rolante, a Máquina de Dwight-Lloyd. Porém, para pequenos processos de sinterização, que podem ocorrer de forma lenta e interrupta, existe o modelo de produção da Máquina de Greenawalt, processo de sinterização em panelas e setor de carga fixa, que não serão abordados no presente trabalho (RIZZO, 2007).

2.1.1 Processo de sinterização contínua por esteira rolante

Os finos já misturados, são adicionados à Máquina de Dwight-Lloyd. O material é adicionado à esteira por meio de dispositivos de gravidade que são controlados por meio de rolos alimentadores, conforme Figura 4.

Figura 4 – Alimentação da esteira de sinterização

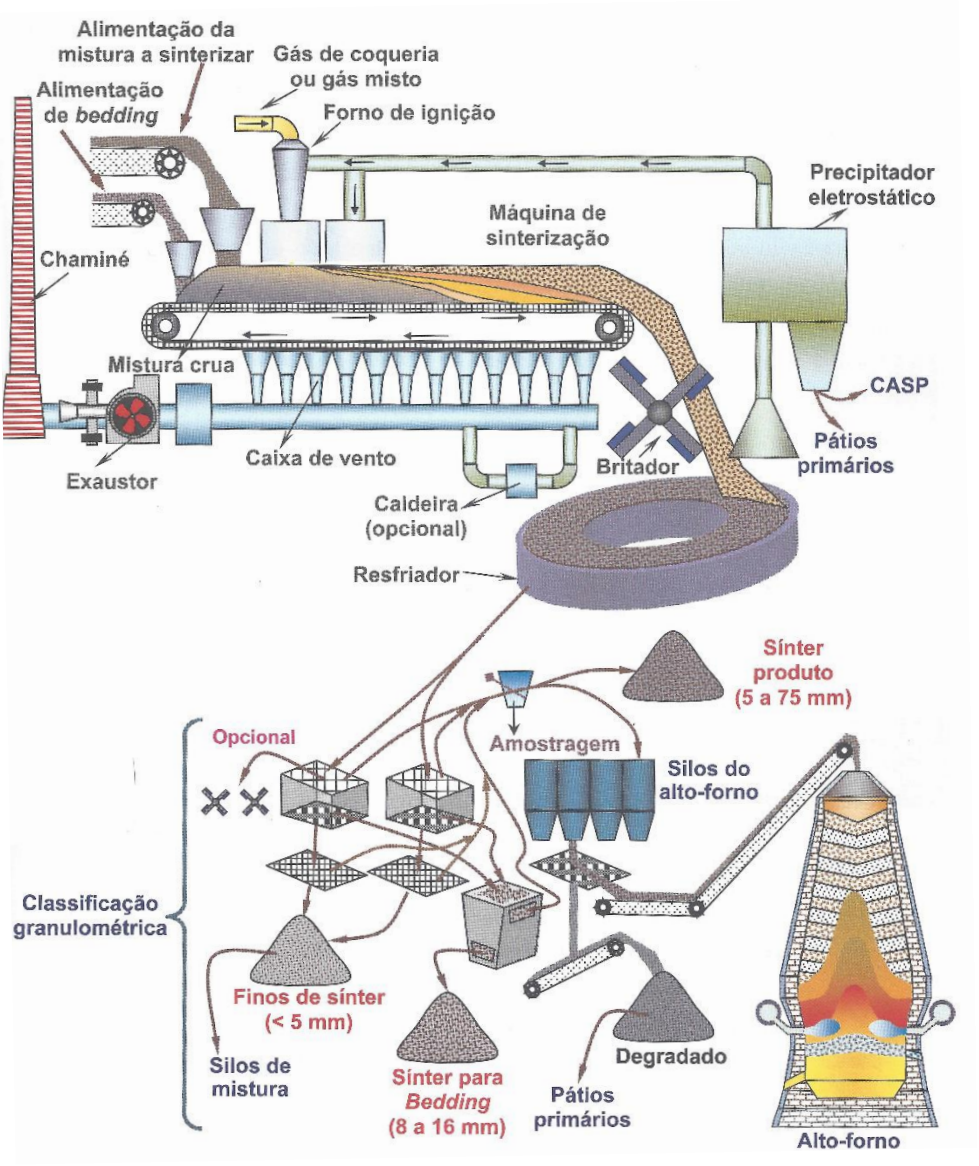


Fonte: ArcelorMittal ([20--?]).

Nessa máquina (Figura 5), a mistura é carregada em uma esteira de metal gradeada, de maneira uniforme, acima de uma camada de proteção chamada de *bedding*, composta por sínter de maior granulometria. A uniformidade da deposição do material e uma altura desejável é garantida por meio de equipamentos de vibração alocados ao longo da esteira (IFES, [20--?]; RIZZO, 2007).

A esteira é fabricada por meio de materiais de alta resistência térmica e ao desgaste, pois devem manter sua dureza mesmo com a elevação das temperaturas ao longo do processo. Ela apresenta espaços livres para permitir a passagem do ar aspirado e se desloca por meio de apoio sobre roletes que se deslocam sobre trilhos (RIZZO, 2007).

Figura 5 – Esquema do processo de sinterização por máquina contínua



Fonte: Rizzo (2007).

A máquina contém um forno de ignição e um sistema de exaustão, composto por exaustores e caixas de vento (CV), que promovem a combustão de todo o combustível presente na mistura. Inicialmente, o forno de ignição aquece a camada superficial da mistura e age na combustão do combustível presente na superfície dela.

Por conta da sucção do ar quente realizada pelos exaustores, há uma transmissão do calor presente na parte superior para a parte inferior que promove a combustão de todo o combustível presente na mistura até o final da esteira. Nesse processo de combustão são gerados gases quentes responsáveis, junto ao sistema de exaustão e o forno de ignição, pela fusão parcial da mistura de entrada.

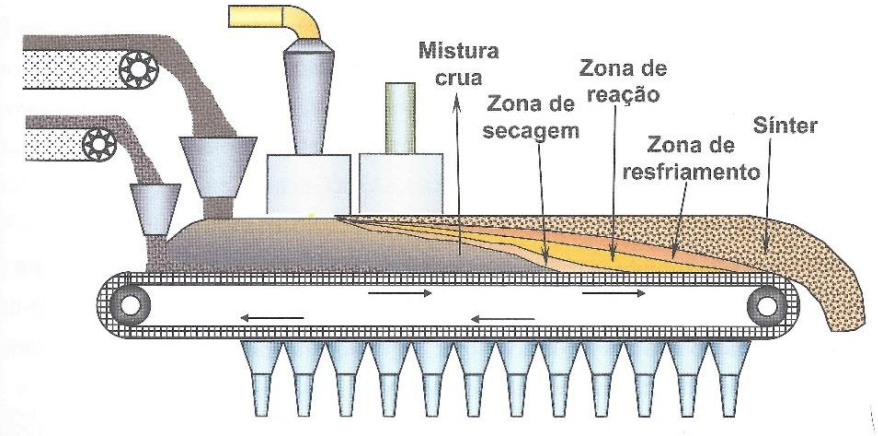
Com a continuidade da sucção do ar pelos exaustores, a mistura fundida parcialmente é resfriada, formando assim, o sínter. Após o resfriamento, o sínter é peneirado para ser feita a classificação granulométrica, separando o sínter produto, do sínter *bedding* e os finos de sínter. O sínter produto é então, enviado para os silos do alto-forno ou é estocado em pátios de armazenamento.

2.1.2 Zonas de sinterização

Durante o processo de produção do sínter, à medida que a mistura de finos passa ao longo da esteira, a aglomeração e fusão parcial da mistura é definida por dois fenômenos distintos, a saber, o fenômeno de transferência de calor da camada superior do processo de sinterização para a camada inferior, promovida pela sucção do ar pelos exaustores, além do fenômeno de combustão do combustível presente na mistura que gera uma gama de reações químicas (RIZZO, 2007).

Durante o processo de aquecimento, os fenômenos citados definem zonas de sinterização ao longo do processo que possuem características comuns, como pode ser observado por meio da Figura 6 (RIZZO, 2007). A temperatura vai aumentando ao longo do processo até atingir sua temperatura máxima, momento no qual é iniciado o processo de resfriamento.

Figura 6 – Zonas de sinterização em um processo de sinterização de máquina contínua



Fonte: Rizzo (2007).

Segundo Rizzo (2017), é possível identificar ao longo do processo, por meio da Figura 6, quatro zonas principais:

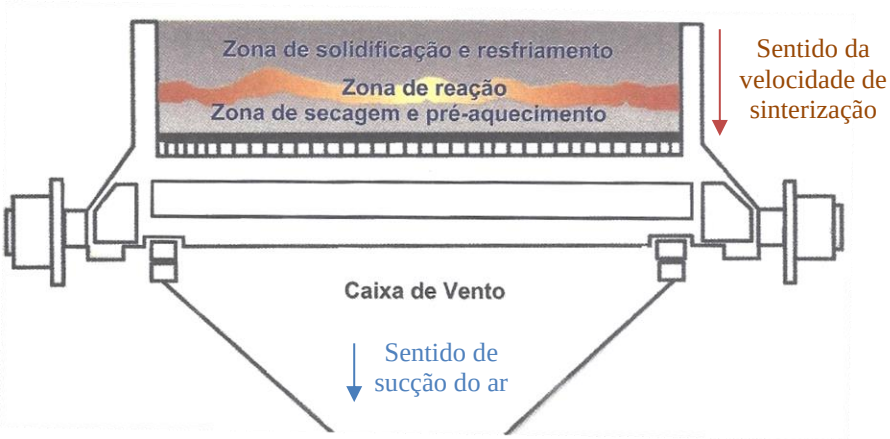
- Mistura Crua (Zona Úmida): zona onde aparecem temperaturas inferiores a 100ºC. Nesse local inicia-se uma evaporação da umidade presente na mistura. Nas regiões interiores da mistura, que apresentam temperatura inferior, a água irá condensar e gerar maior umidade na região;
- Zona de secagem: compreende a faixa de temperaturas entre 100ºC e 150ºC. Ocorre a vaporização total da umidade presente nas misturas e posterior desidratação dos hidróxidos;
- Zona de reação: compreende a faixa de temperaturas acima de 500ºC. Nessa zona se inicia a combustão do combustível presente na mistura até chegar à máxima temperatura e início do resfriamento. Os materiais, nesse momento, são fundidos parcialmente, formando o que é denominado magma do sínter;
- Zona de resfriamento: ocorre o resfriamento e solidificação do sínter, produto do processo.

Na zona de reação ou zona de sinterização forma o magma do sínter, que é composto por um "[...] conjunto de silicatos, ferritos de cálcio e óxidos primários ou secundários que, após cristalização, constituem o sínter" (RIZZO, 2007, p.127).

A reação de combustão do combustível presente na mistura gera uma quantidade significativa de calor que auxilia na fusão dos materiais e, ao mesmo tempo, gera um ambiente redutor que é suficiente para a transformação da hematita (Fe_2O_3) para magnetita (Fe_3O_4) (RIZZO, 2007).

Por meio da Figura 7, é possível observar a representação esquemática da esteira de sinterização e o sistema de exaustão de ar associado por meio de uma visão transversal dos equipamentos. A frente de queima, ou zona de reação, percorre toda a camada à medida que a esteira se desloca rumo ao equipamento de resfriamento, em uma velocidade denominada velocidade de sinterização.

Figura 7 – Representação esquemática do sistema de exaustão de ar



Fonte: Rizzo (2007).
Nota: Adaptado pelo autor.

Quando a zona de reação atinge a parte mais inferior da mistura ocorre o ponto de maior temperatura dos gases que estão sendo succionados ao longo da esteira por meio das caixas de vento que são conectadas aos exaustores, chamado de BTP (do inglês *Burn Through Point*). Esse parâmetro é importante para o processo de produção de aço, pois indica o ponto no qual o calor cedido pelo forno penetrou todo o material e realizou a fusão parcial do mesmo de forma adequada. Após o instante do BTP, inicia-se o resfriamento e solidificação do material produzido.

Dessa forma, o presente trabalho irá se atentar ao comportamento das variáveis do processo de sinterização, tanto de entradas, como leituras, com o objetivo de observar as relações entre elas e a posição de BTP, de forma a preparar as variáveis para desenvolvimento de uma lógica de previsão de comportamento da planta.

2.1.3 O processo operacional de sinterização e o BTP

O processo operacional de sinterização demanda de conhecimentos e habilidades dos operadores. Eles têm responsabilidades no acompanhamento das variáveis e dinâmicas do processo e operação dos equipamentos da planta desde o carreamento da esteira de sinterização até a obtenção do produto final, o sínter.

O operador possui responsabilidades relativas a controle e supervisão do funcionamento adequado da máquina e do processo. Dentre as responsabilidades do operador é possível citar a verificação de falhas na alimentação da esteira de sinterização, garantia da alimentação correta dos materiais de entrada do processo, verificação da composição irregular dos materiais, verificação de falhas no forno de ignição, garantia do funcionamento adequado da instrumentação, além da verificação constante do equilíbrio térmico (SAIZ; POSADA, 2013).

A planta é constantemente monitorada e cada um de seus processos deve ser acompanhado de uma análise de grande quantidade de variáveis, como por exemplo: velocidade da máquina, composição do material de entrada, umidade, fluxo de ar no forno de ignição, fluxo de gás no forno de ignição, velocidade do rolo de alimentação, velocidade do refrigerador, altura da camada além de temperaturas e pressões nas caixas de vento (SAIZ; POSADA, 2013). O operador também deve acompanhar fatores relativos ao cumprimento de normas ambientais. São eles, por exemplo, emissão de particulados e gases como CO_2, SO_2 e NO_x e temperatura de saída dos gases pelas chaminés.

Um dos fatores que o operador deve se atentar é na definição e garantia da finalização da fusão do sínter. É preciso observar onde ocorre o fim da sua queima, local onde o calor inserido inicialmente no material por meio do forno de ignição atinge o fundo da camada de sínter disposto na esteira. Para garantia da formação do sínter de forma adequada e conclusão do processo é definido o BTP.

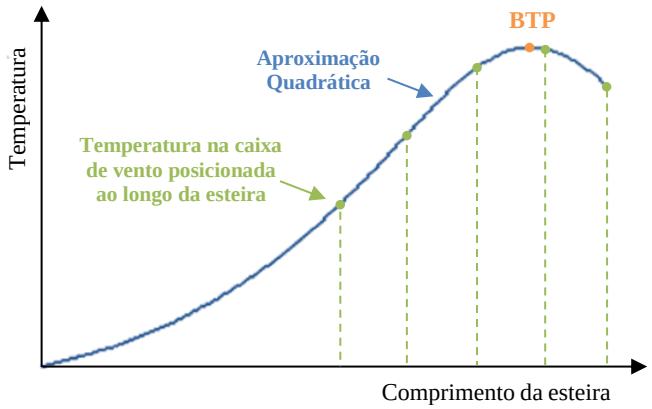
A fusão do sínter se dá quando o calor se espalha por toda a camada de material, esse fato pode ser identificado quando a temperatura do gás residual medido nas caixas de vento atinge seu valor mais elevado, por meio dessa estratégia é definido o BTP. Admite-se que ao manter esse ponto estável em um local da esteira é possível gerar uma estabilização do processo e a melhoria da qualidade e da produtividade (SAIZ; POSADA, 2013).

O BTP é registrado por meio do comportamento das temperaturas ao longo do tempo que são registradas nas caixas de vento que estão dispostas ao longo do processo de sinterização. Em geral, em processos de sinterização típicos, existem caixas de vento e monitoramento de temperatura ao longo de toda a esteira, igualmente espaçadas, porém atenta-se as medições de temperatura das últimas cinco a dez caixas de vento para a estimativa do local do BTP, pois a maior temperatura da esteira deve ocorrer no fim do processo confirmando a fusão parcial do material para a formação do sínter.

Logo após a máxima temperatura ao longo da esteira, quando a frente de queima atingiu a parte inferior do material, dá-se início ao resfriamento da camada mais inferior do sínter, esse processo de resfriamento continuará até o final da esteira, onde o material deve-se apresentar solidificado, conforme pôde ser compreendido por meio da Figura 4.

Dispondo das temperaturas e posições de cada uma das caixas de vento, ao final de cada intervalo de medição é então mapeada uma aproximação por meio de curvas quadráticas do comportamento da temperatura naquele instante, conforme é possível observar por meio da Figura 8 (SONG; WANG; ZHANG, 2013).

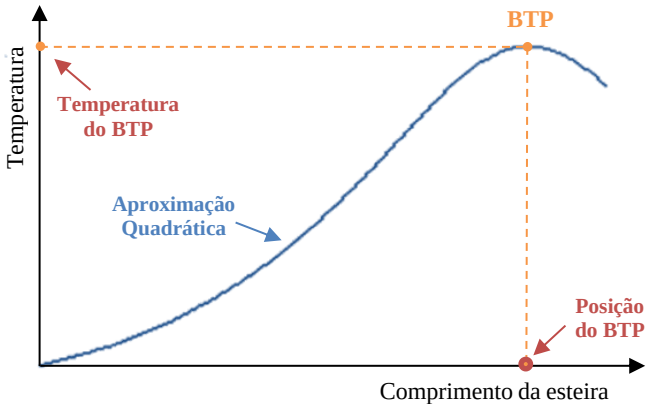
Figura 8 – Posição do BTP em relação as temperaturas das CV



Fonte: Produção do próprio autor.

A partir da aproximação quadrática, é então computado a temperatura máxima indicada pela curva aproximada, essa temperatura é a temperatura do BTP. A máxima temperatura é então associada a uma posição da esteira, de acordo com o equacionamento aproximado, que é a posição de ocorrência do BTP, conforme apresentado na Figura 9. As duas medidas são computadas a cada intervalo de periodicidade de registro das temperaturas.

Figura 9 – Posição do BTP em relação as temperaturas das CV



Fonte: Produção do próprio autor.

A obtenção do polinômio de interpolação de segundo grau pode ser realizada por meio de diferentes técnicas de regressão e pode variar entre as empresas que contém esse tipo de planta e realizam o cálculo dessa medida. Além disso, é possível que algumas plantas utilizem de dados de medições de instantes anteriores ao registro do BTP.

Uma alternativa de medição do BTP é por meio de aparelhos de câmera de termografia infravermelha conforme proposto e testado por Usamentiaga e outros (2013). O monitoramento apresentado se baseou na utilização de uma câmera infravermelha ao final da esteira de sinterização, conforme pode ser observada por meio da Figura 10.

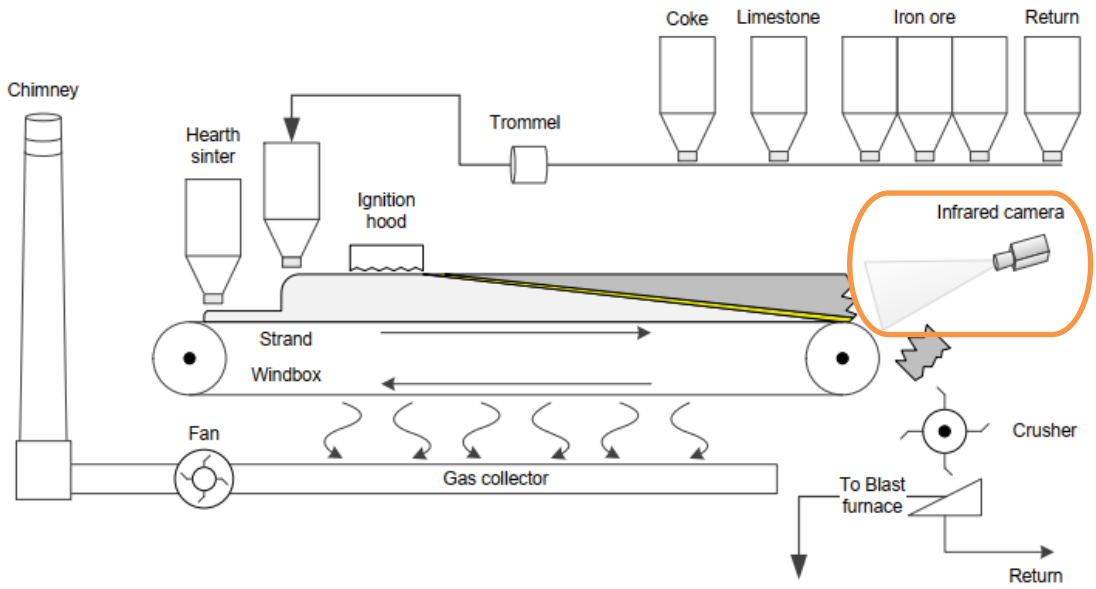
As imagens obtidas pela câmera alocada permitem observar a frente de queima e detectar sua posição. A posição da frente de queima e da máxima temperatura na frente de sinterização permite determinar a posição do BTP na esteira (USAMENTIAGA et al., 2013).

O operador é responsável por selecionar variáveis e gerenciar os indicadores do processo de forma a garantir uma a adequada execução do processo gerando um sínter com qualidade e fundição adequada. A responsabilidade pelo funcionamento adequado do sistema é baseada totalmente em definições e tomadas de decisão humanas do operador, fundamentadas em conhecimentos e vivências práticas do processo.

Nesse contexto, é possível observar que a automatização e a utilização de recursos para auxiliar na produção do sínter se torna válida de forma a garantir a produção, aumentar a eficiência e reduzir a necessidade de reprocesso de materiais com propriedades incorretas. Dessa forma,

uma exploração e reconhecimento dos dados do processo auxiliam na proposição de análises que contribuam para diminuição de possíveis erros cometidos por decisões humanas.

Figura 10 – Processo de sinterização com adição de monitoramento por câmera infravermelha



Fonte: Usamentiaga e outros (2013).
Nota: Adaptado pelo autor.

2.2 Revisão Bibliográfica

A dinâmica e as variáveis do processo de sinterização são frequentemente estudados de forma a garantir o adequado funcionamento da planta e obtenção do sínter de qualidade. A temperatura e posição do BTP constituem importantes parâmetros para indicação e garantia de um desempenho apropriado. Além disso, nota-se uma preocupação com o controle das emissões realizadas pelo processo de sinterização que pode ser controlada com o apropriado desempenho da planta.

Wang e outros (2009), aplicaram técnicas de mineração de dados para encontrar correlações entre dados do BTP com o intuito de empregar um modelo de previsão baseado em redes neurais regressivas a um modelo para controle do BTP na esteira de sinterização. Foram utilizados na pesquisa parâmetros como velocidade da esteira de sinterização, altura da camada e proporção de sínter de retorno, além de temperatura de ignição, permeabilidade do ar, proporção de combustível e relação de umidade final da mistura.

Song, Wang e Zhang (2013), realizaram um estudo acerca do BTP, sua forma de cálculo e influências das variáveis do processo em relação ao seu comportamento, para, em seguida, aplicar modelo de predição com base em uma estrutura Bayesiana. Em 2014, Wang e outros propuseram uma metodologia baseada no comportamento ao longo do tempo das curvas de temperatura para estimar e em seguida prever o comportamento do BTP. Para isso foram utilizadas para regressão máquinas de vetores de suporte de mínimos quadrados (LS-SVM, do inglês *Bayesian Least Squares Support Vector Machines*).

Galliazi e outros (2014), apresentam a influência sobre os parâmetros de controle do processo de sinterização para a qualidade física do sínter para aplicação nos altos-fornos. Foi analisada a influência da velocidade de sinterização, altura da camada e velocidade da máquina de sínter influenciam na granulometria e qualidade final do sínter produzido.

Kim e outros (2014), mostram diferentes modelos para calcular e estimar a curva de temperatura a partir das leituras de temperatura obtidas nas das caixas de vento em um determinado processo de sinterização. Foram propostas curvas de terceira e quinta ordem para maior precisão da estimativa do comportamento desse parâmetro que não era propriamente refletido por meio de aproximações de segundo grau.

Wang e outros (2018), já mudam o alvo da pesquisa e se atentam a prever as características do sínter e emissões do processo, tomando 220 variáveis, tendo a permeabilidade do leito, densidade de ignição, quantidade de dolomita e calcário na matéria-prima, analisadas ao longo do estudo proposto.

Nota-se, por meio das referências aqui tratadas que há uma preocupação com o desempenho do processo associado a suas variáveis. Os autores se atentaram a análise do conjunto de variáveis medidas, tratamentos de conjuntos de dados e correlação entre as variáveis com o objetivo de formar uma base de dados robusta que esteja associada a um desempenho adequado dos diferentes modelos de aprendizagem de máquina utilizados.

3 TRANSFORMADA *WAVELET* E SUA APLICAÇÃO NA FILTRAGEM DE SINAIS

A presente seção aborda uma revisão de conceitos e propriedades para o entendimento da aplicação da Transformada *Wavelet* Discreta em processamento de sinais com objetivo de empregar uma redução de ruídos.

3.1 Transformada *Wavelet* Discreta

A utilização da Transformada de *Wavelet* em forma contínua se torna restrita a análises e estudos teóricos. Para aplicações práticas, dessa forma, é utilizada a Transformada de *Wavelet* Discreta (DWT, do inglês Discrete *Wavelet Transform*). Nessas aplicações os sinais discretos a serem tratados são associados a *wavelets* com parâmetros de translação, dilatação e variável independente discretos. A operação tem como resultado um conjunto de dados discretos (DIAS, 2017).

O problema de resolução tempo-frequência é resultado de um fenômeno físico conhecido como Princípio de Gabor-Heisenberg, que estabelece que não é possível conhecer a frequência exata e o tempo exato de ocorrência dessa frequência em um sinal. Esse fenômeno está presente na DWT.

O sinal é analisado por uma forma alternativa denominada Análise Multi-resolução (MRA, do inglês *MultiResolution Analysis*) (OLIVEIRA, 2007). Essa técnica foi proposta por Mallat (1989) que descreve em sua publicação uma forma de extrair informações de um sinal a partir de sucessivas resoluções. Para isso, o sinal é decomposto em uma série de frequências independentes com ajuste espacial. São aplicados filtros baseados em *wavelets* ortogonais, por meio de janelamento variável. A técnica é utilizada para analisar sinais em diferentes componentes de frequências e resoluções temporais.

Segundo Oliveira (2007), a equação (1) mostra a DWT, onde há discretização apenas dos parâmetros de escala e translação.

$$\psi_{j,r}[m] = \frac{1}{\sqrt{|a_0^j|}}\psi\left(\frac{m - r b_0 a_0^j}{a_0^j}\right) \quad (1)$$

onde:

- j e r são números inteiros;
- $a_0 > 1$ é o parâmetro de escala fixo;
- b_0 é o parâmetro de translação fixo;
- m é a variável independente discreta.

Escolhendo $a_0 = 2$ e $b_0 = 1$, obtém se a equação (2).

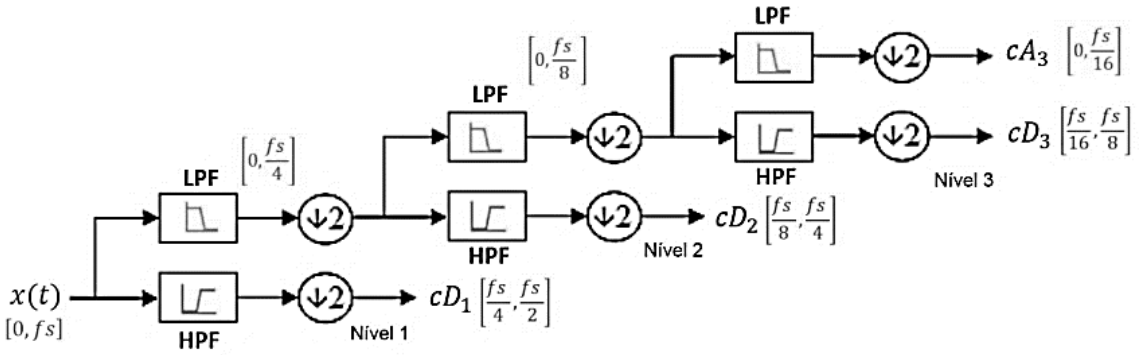
$$\psi_{j,r}[m] = 2^{-\frac{j}{2}}\ \psi(2^{-j}m - r) \quad (2)$$

A aplicação da análise por *wavelets* ortogonais por meio de janelas pequenas permite a extração de componentes de alta frequência já a aplicação de janelas maiores extrai componentes de baixa frequência. Esse formato de extração de sinais se mostra importante pois, segundo Rioul e Vetterli (1991), as componentes de baixa frequência carregam informações de aproximação do sinal, isto é, trazem informações genéricas sobre o seu comportamento, enquanto as componentes de alta frequência carregam o nível de detalhamento desses sinais, trazendo informações específicas.

Dessa forma, no primeiro passo da decomposição são, então, aplicados um filtro passa-altas (HPF, do inglês *High Pass Filter*) e outro passa-baixas (LPF, do inglês *Low Pass Filter*), determinados de acordo com a *wavelet* escolhida. A cada nova resolução, são aplicados novamente os filtros, dividindo o espectro de frequência em duas partes iguais por meio das filtragens. Em seguida os sinais são subamostrados, formando coeficientes de detalhe e de aproximação, conforme é possível observar por meio da Figura 11 (CUNHA, 2013).

O sinal decomposto por meio da MRA proposta por Mallat (1989), pode ser plenamente recuperado. Dessa forma, é possível selecionar informações suficientes para a reconstrução do sinal por meio da filtragem e eliminar frequências não necessárias a análise do sinal em estudo.

Figura 11 – Processo de decomposição do sinal via DWT



Fonte: Diniz (2018).
Nota: Adaptado pelo autor.

É possível realizar, além disso, a simplificação de análises de sinais, permitir a identificação de ruídos e compressão de dados, ao identificar frequências passíveis de supressão que não influenciam significativamente no seu comportamento.

A abordagem DWT, entretanto, apresenta deficiências. O comprimento do sinal a ser inserido em uma dessas transformadas é limitado a um múltiplo inteiro de uma potência de dois e elas apresentam sensitividade em relação a escolha de um ponto de partida da série temporal. Isso se deve a necessidade de reamostragem das saídas da *wavelet* e também a utilização de filtros de escala em cada estágio do algoritmo (PERCIVAL; WALDEN, 2000).

Os dados de sinais coletados de aplicações com sensoriamento constante são atualizados com grande frequência e suas análises necessitam de atualização continuada. Ao aplicar a abordagem DWT a esse tipo sinal, a cada nova adição de dados e mudança da estrutura do conjunto, os coeficientes resultantes das reamostragens da transformada seriam, em sua grande parte, atualizados. Dessa forma, essa transformada não se mostra adequada para essa finalidade.

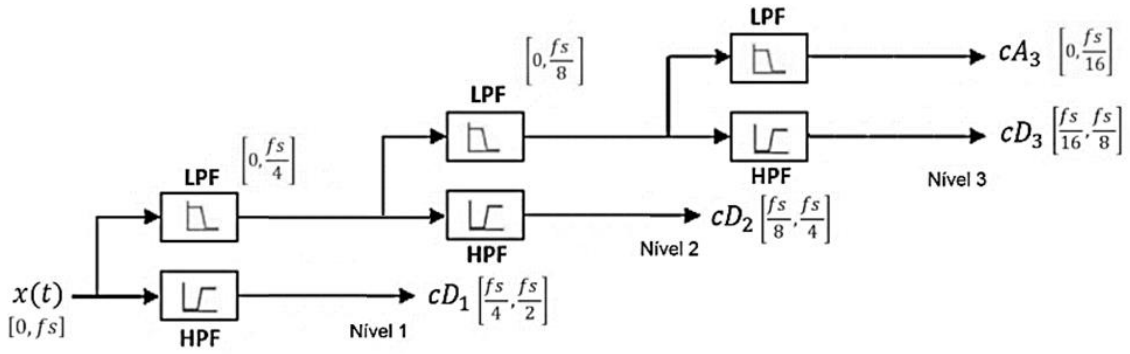
3.2 Transformadas *Wavelet* Não-decimadas e Invariantes no Tempo

As Transformadas *Wavelets* não decimadas e invariantes no tempo, por não apresentarem decimação, podem ser aplicadas a qualquer série temporal de qualquer comprimento. Esse tipo de transformada apresenta formato semelhante a DWT, sendo utilizadas por meio de MRA para decompor um sinal em amostras de altas e baixas frequências. Entretanto, ela se torna interessante para análises de sinais pois é capaz de superar algumas das principais deficiências

apresentadas da DWT como o comprimento do sinal e a definição do ponto de partida para a análise (PERCIVAL; WALDEN, 2000).

A transformada *wavelet* discreta não-decimada e invariante no tempo utilizada nesse trabalho foi a Trasnformada de *Wavelet* de sobreposição máxima (MODWT, do inglês *Maximal Overlaped Discrete Wavelet Transform*). O processo de decomposição em três níveis de uma MODWT pode ser observado por meio da Figura 12.

Figura 12 – Processo de decomposição do sinal via MODWT



Fonte: Diniz (2018).
Nota: Adaptado pelo autor.

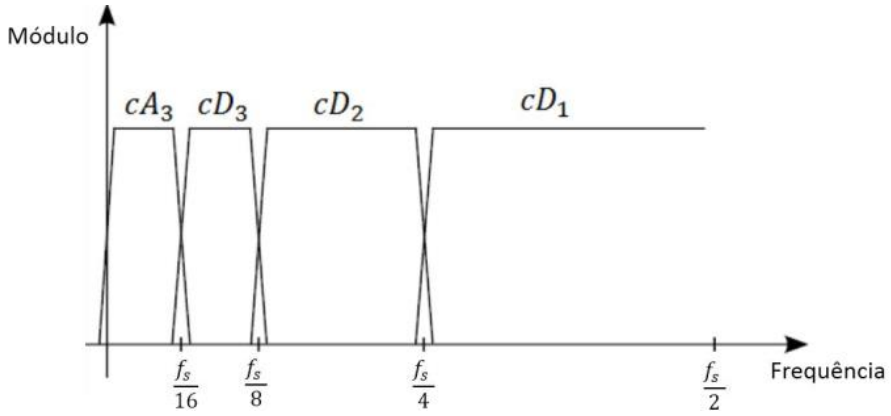
A MODWT, por exemplo, pode ser definida para sinais que apresentem qualquer comprimento. A DWT, entretanto, apresenta limitação em relação aos comprimentos de sinais. Nessa abordagem, há a necessidade de que o comprimento do sinal seja múltiplos de potências de 2, dependendo do número de níveis definidos previamente para a MRA. A MODWT não apresenta dependência do ponto de partida, como na DWT e nem da quantidade de amostras que são utilizadas para análise (PERCIVAL; WALDEN, 2000).

Deve-se salientar que, por conta da ausência da necessidade de decimação, a MODWT mantém o número de coeficientes *wavelets* associados independentemente da quantidade de decomposições a serem realizadas. Enquanto isso, a DWT apresenta diminuição de coeficientes pela metade por conta da decimação associada. Esses componentes são descartados após reamostragem do sinal (PERCIVAL; WALDEN, 2000).

A MODWT não apresenta largura de banda uniforme e são invariantes no tempo. Ela não apresenta efeito de borda nas extremidades da série temporal porque, nesse caso, as componentes não são ortogonais e há sobreposição das bandas de frequências. Esse fato ocorre

como efeito colateral para resolver o problema de borda e permitir o seu uso em séries temporais (PERCIVAL; WALDEN, 2000). É possível observar a sobreposição das bandas (do inglês *overlaped*) por meio da Figura 13.

Figura 13 – Sobreposição das bandas de frequência de um processo MODWT



Fonte: Diniz (2018).

3.3 Procedimento de Aplicação de *Wavelets* como Filtros Lineares

O procedimento de filtragem de sinais por meio de Transformada *Wavelet* se inicia por meio da aplicação de uma MRA, seguida de separação e seleção de faixas de frequência adequadas para compor o sinal filtrado e por fim, a realização da síntese por meio de bancos de filtro de síntese.

O primeiro passo para desenvolvimento da aplicação de inicia por meio da definição dos parâmetros a serem aplicados na MRA. Deve-se definir, inicialmente, a quantidade de decomposições em frequência que devem ser utilizadas na análise para realização da adequada separação do ruído do sinal em estudo (CUNHA, 2013). Além disso, deve-se selecionar a família de *wavelets* a serem utilizada no procedimento. As ponderações necessárias para essa definição são apresentadas na seção 3.4 do presente trabalho.

Em seguida, ao final da decomposição, deve-se analisar o conjunto de frequências discriminadas e identificar os coeficientes que concentram a maior parte da energia do sinal baseado em um limite pré-determinado. Nesse processo, é possível eliminar ruído branco e ruídos de espectro discreto, desde que as frequências desses ruídos não coincidam com as componentes de frequência do sinal. Dessa forma, remove-se boa parte do ruído permitindo

uma boa estimativa do sinal desejado. Após a seleção dessas frequências, o sinal pode ser recuperado por meio de um banco de filtros de síntese (CUNHA, 2013).

3.4 Famílias de *Wavelets*

Existem, atualmente, uma grande quantidade de funções *wavelets* destinadas a análises contínuas e discretas. A escolha da *wavelet* que será utilizada para processamento do sinal de interesse deve ser tomada com base nas características da forma da *wavelet* comparadas as características da forma do sinal.

Quanto maiores forem as semelhanças entre o sinal e a *wavelet*, melhor será a representação do sinal pela *wavelet*. Deve ser levado em conta, ao realizar essa comparação, nas diferentes variações entre propriedades das *wavelets* como, por exemplo, a sua simetria, regularidade, suavidade e escala (MISITI et al., 2017).

No presente trabalho, será utilizada MODWT com as bases de Haar, Daubechies e suas famílias de funções derivadas, Symlets e Coiflet.

3.4.1 *Wavelet* de Haar

A *Wavelet* de Haar é a mais simples. Ela é baseada em uma *wavelet* de escalamento cuja função é dada por meio da equação (3).

$$\phi(t) = \begin{cases} 1, se\ 0 \leq t \leq 1 \\ 0, caso\ contrário \end{cases} \tag{3}$$

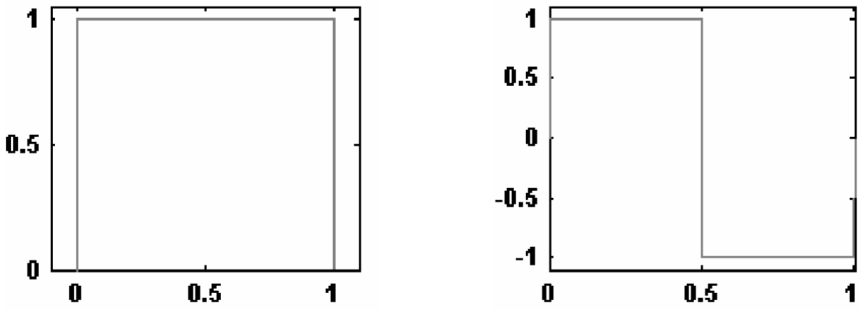
A *wavelet* de Haar é construída a partir das operações de mudança de escala e translação aplicada a *wavelet* mãe[1] $\phi(t)$. Nesse caso, a função $\psi(t)$ é definida como $\psi(t) = \phi(2t) - \phi(2t-1)$, e pode ser representada por meio da equação (4).

[1] A Wavelet Mãe é também conhecida como Wavelet Base (MISSITI et al., 2013).

$$\psi(t) = \begin{cases} 1, se\ 0 \leq t \leq \frac{1}{2} \\ -1, se\ \frac{1}{2} \leq t \leq 1 \\ 0, \quad caso\ contrário \end{cases} \tag{4}$$

As duas funções podem ser observadas na Figura 14.

Figura 14 – Wavelet Haar (a) Função de escalonamento (b) Função base



(a) (b)

Fonte: Missiti e outros (2013).

A *wavelet* de Haar é descontínua e não diferenciável, não possuindo uma transição suave escala/tempo (SANTOS, 2015). Sua utilização é adequada para sinais que possuem mudanças abruptas de valores ou descontinuidades no decorrer do tempo.

3.4.2 *Wavelet* de Daubechies

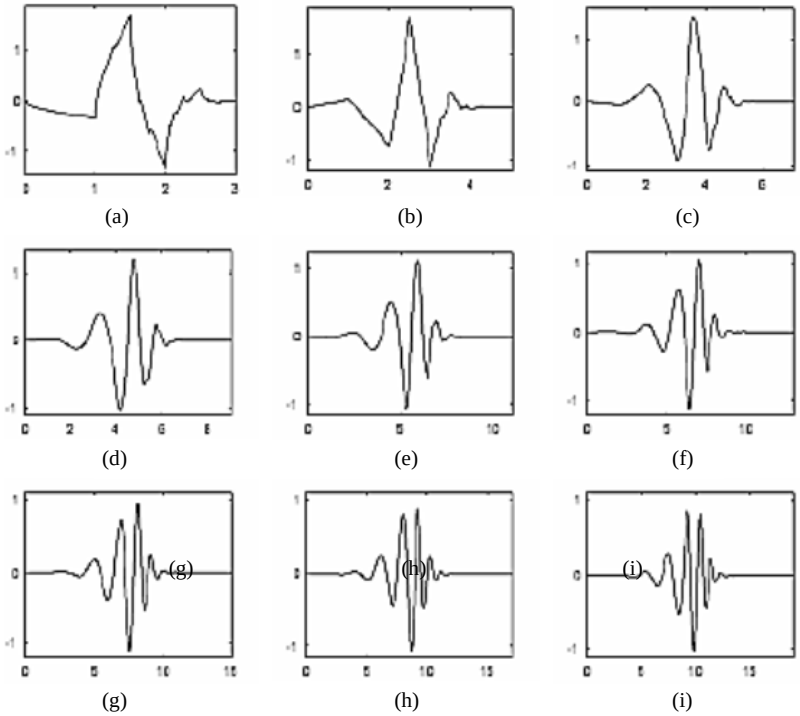
As *wavelets* de Daubechies, são perfeitamente compactas no tempo, mas no domínio da frequência, tem um alto grau de superposição espectral entre as escalas (GUTIERREZ, 2002). Sua maior vantagem é serem ortogonais, isto é, quando ocorre um erro no sinal de entrada, ele não cresce com a transformação. Dessa forma, é mantida a estabilidade numérica computacional garantindo a fidelidade na transformada (SANTOS, 2015).

Trata-se como N a ordem da *wavelet* e dbN é a *wavelet* de ordem N. Essa família de *wavelets* contém, como db1, a *wavelet* de Haar, a mais simples dentre as demais (MISSITI et al., 2013).

A ordem N também expressa o número de vezes no qual as *wavelets* passam pelo eixo zero. As *wavelets* são assimétricas, especialmente quando a sua ordem apresenta valores baixos. Quanto

maior a ordem da *wavelet*, maior a sua regularidade e simetria, conforme é possível observar por meio da Figura 15 (MISSITI et al., 2013).

Figura 15 – *Wavelets* Daubechies (a) db2 (b) db3 (c) db4 (d) db5 (e) db6 (f) db7 (g) db8 (h) db9 (i) db10



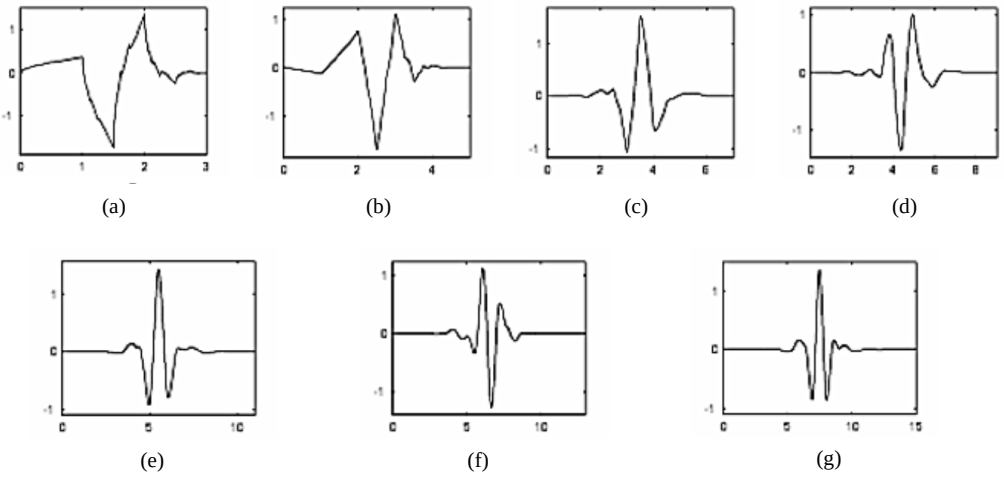
Fonte: Missiti e outros (2013).

3.4.3 *Wavelets* Symlets e Coiflets

A família de *wavelets* Symlets foi proposta por Daubechies e foi obtida de modificações realizadas na família de *wavelets* Daubechies. As Symlets apresentam características semelhantes as Daubechies, porém foram alteradas de forma a se tornarem mais simétricas sem diminuição de sua simplicidade, conforme pode ser observado por meio da Figura 16 (BIANCHI, 2006; MISSITI et al., 2013).

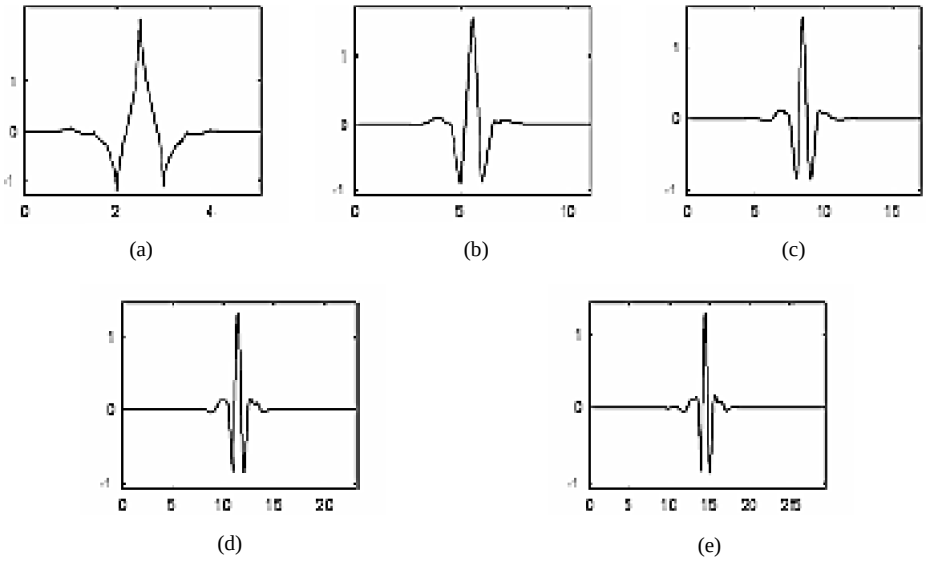
As Coiflets também foram construídas de forma que se tornassem mais simétricas, e nesse caso, são ainda mais simétricas que as Symlets, conforme é possível observar por meio da Figura 17. Com relação ao número de zeros, pode-se observar que ela apresenta 2N zeros em sua composição (BIANCHI, 2006; MISSITI et al., 2013).

Figura 16 – *Wavelets* do tipo Symlet (a) sym2 (b) sym3 (c) sym4 (d) sym5 (e) sym6 (f) sym7 (g) sym8



Fonte: Missiti e outros (2013).

Figura 17 – *Wavelets* do tipo Coiflets (a) coif1 (b) coif2 (c) coif3 (d) coif4 (e) coif5



Fonte: Missiti e outros (2013).

4 ANÁLISE DE DADOS

A análise de dados compreende desde a verificação de sua qualidade e correspondência a realidade até aplicações estatísticas para inferências de informações. É abordado na presente seção metodologias de análise e suas definições, englobando análise da qualidade dos dados, análise descritiva e análise estatística multivariada.

4.1 Análise de Qualidade dos Dados

A Qualidade de Dados corresponde às características dos dados para satisfazer um sistema e seus requerimentos técnicos. É uma visão ou uma avaliação de adequação de dados para servir a um propósito em determinado contexto (MAHANTI, 2019). Ela deve ser garantida por quaisquer instituições que utilizem de dados para análise de tomada de decisão, comportamento de variáveis, análises de dados brutos ou análises de previsão de processos.

As análises da qualidade de dados podem ser realizadas por meio de duas diferentes perspectivas: contextual e intrínseca. A qualidade de dados contextual se refere a percepções individuais a respeito da qualidade de dados. Ela é influenciada por visões que dependem do cenário da análise, como as características da informação e como o usuário desses dados as utilizam (HAIR et al., 2009).

A qualidade de dados intrínseca, diferentemente da contextual, é baseada apenas nos elementos de dados presentes em um conjunto, independente do seu contexto. Esse formato de análise é direcionado a verificar, por exemplo, a adequação das dimensões e a completude e a precisão dos dados (BATINI; SCANNAPIECO, 2016).

Atualmente existe uma norma da ISO (do inglês *International Organization for Standardization*) e IEC (do inglês *International Electrotechnical Commission*), a ISSO/IEC 25012:2008. Essa norma define o modelo de qualidade de dados e pode ser utilizada para estabelecer requisitos, definir medidas e realizar avaliações de qualidade de dados, além de identificar critérios de garantia de análise de dados (ISO, 2019a).

A ISO (2019b) apresenta 15 características presentes no padrão ISO/IEC 20512 para qualidade dos dados:

- Acurácia (accuracy): o grau em que os dados apresentam atributos que representam corretamente o valor verdadeiro do atributo pretendido de um conceito ou evento em um contexto específico de uso;
- Completude (*completeness*): o grau em que os dados de assunto associados a uma entidade têm valores para todos os atributos esperados e instâncias de entidades relacionadas em um contexto específico de uso;
- Consistência (*consistency*): o grau no qual os dados de um contexto específico não se contradizem, entre si, e apresentam coerência uns com os outros;
- Credibilidade (*credibility*): o grau no qual os dados têm atributos considerados verdadeiros e verossímeis pelos usuários em um contexto específico de uso;
- Atualidade (*currentness*): o grau no qual os dados se ajustam temporalmente em um contexto de uso;
- Acessibilidade (*accessibility*): o grau em que os dados podem ser acessados em um contexto específico de uso, particularmente por pessoas que precisam de suporte tecnológico ou configuração especial devido a alguma deficiência;
- Conformidade (*compliance*): o grau em que os dados possuem atributos que aderem a normas, convenções ou regulamentos em vigor e regras semelhantes relacionadas à qualidade de dados em um contexto específico de uso;
- Confidencialidade (*confidentiality*): o grau em que os dados possuem atributos que garantem que sejam acessíveis e interpretáveis apenas por usuários autorizados em um contexto específico de uso;
- Eficiência (*efficiency*): o grau em que os dados possuem atributos que podem ser processados e fornecem os níveis esperados de desempenho usando as quantidades e os tipos apropriados de recursos em um contexto específico de uso;
- Precisão (*precision*): o grau segundo o qual os dados apresentaram atributos exatos e discrimináveis no contexto de uso;
- Rastreabilidade (*traceability*): o grau em que os dados possuem atributos que fornecem uma trilha de auditoria de acesso aos dados e de quaisquer alterações feitas nos dados em um contexto específico de uso;

- Compreensibilidade (*understandability*): o grau em que os dados têm atributos que permitem que sejam lidos e interpretados pelos usuários e expressos em idiomas, símbolos e unidades apropriados em um contexto de uso específico;
- Disponibilidade (*availability*): o grau em que os dados possuem atributos que permitem recuperá-los por usuários ou aplicativos autorizados em um contexto específico de uso;
- Portabilidade (*portability*): o grau em que os dados possuem atributos que permitem que sejam instalados, substituídos ou movidos de um sistema para outro, preservando a qualidade existente em um contexto específico de uso;
- Recuperabilidade (*recoverability*): o grau em que os dados possuem atributos que permitem manter e preservar um nível especificado de operações e qualidade, mesmo em caso de falha, em um contexto específico de uso.

Utilizando as informações contidas na instrução normativa anteriormente citada, é possível realizar uma adequação na leitura, no armazenamento e no tratamento dos dados antes que sejam realizadas análises acerca do conteúdo de seu conjunto de dados. Esse fato garante uma análise mais precisa e fiel à realidade do conteúdo dos dados.

4.2 Análise Descritiva

A análise descritiva dos dados utiliza de meios como tabelas, gráficos e medidas básicas de resumo em *dashboards*, por exemplo, de forma a permitir ao pesquisador entender melhor como estão se comportando os dados ao longo do tempo. As técnicas permitem identificar tendências, variabilidade e valores atípicos no conjunto de dados (FÁVERO et al., 2009).

Na estatística descritiva, as técnicas são aplicadas a cada variável, de forma a observar o seu comportamento individual. São realizadas, para cada variável, análises em relação aos valores centrais e não centrais por meio das medidas de posição, medidas de dispersões e assimetrias.

Por meio das medidas de posição é possível realizar uma síntese dos dados por meio de valores representativos. As medidas de posição são divididas em medidas de tendência central, medidas de tendência não central e verificação de *outliers*. As tendências de medida central são as medidas estatísticas de maior uso popular, podendo-se elencar, por exemplo, a média, a mediana e a moda (FÁVERO et al., 2009). Dentre as tendências não centrais, pode-se citar os

quartis, decis e percentis, que dividem o conjunto de dados em quatro, dez e cem partes em ordem crescente dos dados.

As medidas de dispersão englobam parâmetros como: amplitude, que representa a diferença entre o maior e menor valor do conjunto; desvio médio, média dos desvios em relação à média; variância, dispersão dos dados em relação à média; desvio padrão, além do coeficiente de variação, que mede a homogeneidade dos dados em relação à média.

As medidas de assimetria e curtose têm objetivo de caracterizar a forma de distribuição dos dados em relação à média, sendo a assimetria o grau de desvio da simetria de uma distribuição e a curtose, o achatamento ou alongamento da distribuição em relação à altura do pronto máximo da distribuição.

4.3 Análise Estatística Multivariada

Segundo Hair e outros (2009, p. 23), “Análise multivariada se refere a todas as técnicas estatísticas que simultaneamente analisam múltiplas medidas sobre indivíduos ou objetos sob investigação”. Diferentes autores e pesquisadores podem conceituar essa forma de análise com pequenas distinções. Alguns definem como uma análise sob um conjunto de variáveis reais enquanto outros afirmam ser uma análise sob um conjunto de variáveis estatísticas (agrupamento de variáveis).

“A realização de análise multivariada se insere em um contexto no qual as variáveis são aleatórias e inter-relacionadas, de forma que os seus efeitos não podem ser significativamente interpretados separadamente [...]” (FÁVERO et al., 2009, p. 5). As variáveis de análise, por exemplo, podem pertencer a um mesmo processo ou pesquisa, ou a uma mesma dinâmica operacional.

As técnicas de análise multivariada se difundiram rapidamente e se tornaram amplamente utilizadas por diferentes instituições, pois permitiram o fornecimento de embasamentos mais sólidos perante tomadas de decisão. Essas técnicas são comumente divididas em dois grupos, as técnicas de interdependência, que estão focadas na identificação da estrutura dos dados, e as

técnicas de dependência, que estão focadas na utilização de variáveis para prever ou explicar alguma variável dependente.

Serão apresentadas nas seguintes seções os conceitos de técnicas de interdependência e dependência e as análises a serem realizadas no presente trabalho.

4.3.1 Técnicas de interdependência

As técnicas de interdependência têm como base realizar a identificação de estruturas e correlações entre um conjunto definido de variáveis, observações ou objetos. Essas técnicas podem gerar maior simplicidade durante a análise e também proporcionar descobertas iniciais acerca do conjunto de dados, porém não definirão estruturas para otimização e maximização de dependência entre as variáveis (HAIR et al., 2009). Os pesquisadores, comumente, utilizam dessas técnicas para avaliar a estrutura de seus dados e, em seguida, as empregam em novas análises. Será tratada, então, nessa seção, uma das técnicas de interdependência de análises estatísticas multivariada, a análise de fatores.

A análise de fatores ou análise fatorial é um conjunto de técnicas de interdependência utilizadas, regularmente, para realizar análises de inter-relações entre variáveis e sintetizar as informações presentes em grupos de variáveis (fatores) sem que haja perda de informação significativa (HAIR et al., 2009). Os grupos de variáveis devem ser formados de tal maneira que as variáveis que estão inseridas em um mesmo fator devem ser altamente correlacionadas e, ao mesmo tempo, devem ter uma baixa correlação com variáveis presentes em outros fatores.

A análise fatorial pode ser realizada a partir de duas propostas, de forma exploratória ou de forma confirmatória, de acordo com o conhecimento que o pesquisador possui acerca dos dados que estão sendo tratados. Em uma análise confirmatória, o pesquisador tem conhecimento, por meio da execução do processo, da dinâmica de correlação e influências entre as variáveis, enquanto, em uma análise exploratória, o pesquisador tem pouco ou nenhum conhecimento das estruturas dos fatores (FÁVERO et al., 2009).

Deve-se realizar a identificação de *outliers* da distribuição e verificar se a distribuição seja tendenciosa antes de realizar a análise fatorial, pois esses dois aspectos podem distorcer os

resultados das correlações, uma vez que alteram as estimativas das médias e dos desvios padrão (FÁVERO et al., 2009).

O presente trabalho terá interesse no desenvolvimento de uma análise fatorial exploratória. Para a realização dela deve-se, primeiramente, efetuar análises de correlações e estatísticas associadas com o objetivo de verificar se, de fato, esse tipo de análise pode ser corretamente aplicado as variáveis que estão inseridas no conjunto de análise. Em seguida deve-se realizar a extração de fatores iniciais, definir o número de fatores e, por fim, analisar a rotação deles. A descrição de cada etapa do processo de análise fatorial e a metodologia utilizada para o desenvolvimento do processo são definidas por meio das seções seguintes.

4.3.1.1 Adequação da Utilização da Análise Fatorial

Previamente ao início da aplicação da Análise Fatorial, devem ser realizadas suposições críticas de forma a analisar o cumprimento de pressupostos conceituais, acerca da validade e adequação do grupo de dados, e pré-requisitos estatísticos, relacionados à métricas de correlações entre as variáveis.

Em relação aos princípios conceituais, deve-se estudar o caso em análise de forma a entender, a princípio, se há uma interação e integração conceitual entre as variáveis escolhidas para a realização da Análise Fatorial, observando a dinâmica de processos, ambientes ou pesquisas em observação, atentando-se a relação entre variáveis dependentes e independentes. Deve-se, também, garantir que os padrões a serem observados são, de fato, válidos e adequados a planta em que se está atuando.

Deve-se, então, efetuar análises estatísticas de forma a garantir que as variáveis selecionadas apresentem correlações suficientes que justifiquem a aplicação da Análise Fatorial (HAIR et al., 2009, p. 109). Para alcançar esse propósito é interessante que se utilize diferentes abordagens de identificação de correlações, são elas: análise visual da matriz de correlações, análise da matriz de correlação anti-imagem, teste de esfericidade de Bartlett e Medida de Adequação de Amostra (MSA, do inglês *Measure of Sampling Adequacy*) (FÁVERO et al., 2009, p. 240).

A análise fatorial é baseada nas correlações entre variáveis mostradas, comumente, por uma matriz de correlações. A matriz de correlações mede a associação linear entre as variáveis por um determinado coeficiente correlação sendo o de utilização mais comum, o coeficiente de Pearson.

Sendo $X_{a,i}$ cada elemento registrado de uma determinada variável X, $Y_{b,i}$ cada dado registrado de uma determinada variável Y, $\overline{X_a}$ e $\overline{Y_b}$ a média dos dados de cada uma das variáveis, o coeficiente de correlação de Pearson pode ser descrito por meio da equação (5).

$$rho(a,b) = \frac{\sum_{i=1}^{n}(X_{a,i} - \overline{X_a})(Y_{b,i} - \overline{Y_b})}{\left\{\sum_{i=1}^{n}(X_{a,i} - \overline{X_a})^2 \sum_{i=1}^{n}\left(Y_{b,i} - \overline{Y_b}\right)^2\right\}^{1/2}} \quad (5)$$

onde:

- n é quantidade total dos dados a serem analisados;
- $\overline{X_a} = \sum_{i=1}^{n}(X_{a,i})/n$
- $\overline{Y_b} = \sum_{i=1}^{n}(Y_{b,i})/n$

O coeficiente de correlação de Pearson varia de -1 a 1. O sinal apresentado indica a direção positiva ou negativa da correlação, que é perfeita nos extremos do intervalo e não existe quando o valor da correlação for igual a 0, em outras palavras, quando a correlação foi igual a 1 ou -1 o valor de uma variável pode ser determinado exatamente ao conhecer o valor da outra, enquanto que, se o valor for igual a 0, não há nenhuma correlação de primeiro grau entre as variáveis. Por fim, o coeficiente é calculado a cada duas variáveis que apresentem coerência conceitual ao serem correlacionadas, formando a matriz de correlações.

O estudo anterior à Análise Fatorial deve-se iniciar, então, examinando a matriz de correlações e realizando uma análise exploratória de forma a verificar se existem valores significativos de correlações que tornem a análise adequada. É possível inferir que, caso seja observada alta correlação entre determinadas variáveis, elas tendam a compartilhar o mesmo fator, e, caso apresentem baixa correlação, a realização da Análise Fatorial pode não seja adequada (FÁVERO et al., 2009). Se a inspeção visual da matriz de correlação não revelar valor superior a 0,30, considerando o coeficiente de Pearson, o formato de análise é, provavelmente, inadequado (HAIR et al., 2009).

As correlações entre as variáveis também podem ser analisadas por meio das correlações parciais entre as variáveis. Uma correlação parcial é aquela que não é explicada quando os efeitos de outras variáveis são levados em conta. Se as correlações parciais não são altas, então a análise fatorial se mostra inadequada. Uma elevada correlação parcial é aquela com significância prática e estatística, uma regra prática seria considerar acima de 0,7 como elevadas (HAIR et al., 2009). Uma matriz de correlação anti-imagem é construída de forma a mostrar o valor negativo das correlações parciais.

Uma estatística usual para verificação da adequação das variáveis a Análise Fatorial é a estatística KMO (Kaiser-Meyer-Olkin), que compara as correlações simples com as correlações parciais. O cálculo da estatística é dado pela equação (6).

$$KMO = \frac{\sum_{i \neq j} \sum {r_{ij}}^2}{\sum_{i \neq j} \sum {r_{ij}}^2 + \sum_{i \neq j} \sum {a_{ij}}^2} \tag{6}$$

onde:

- r_{ij} é o coeficiente de correlação entre as variáveis;
- a_{ij} é o coeficiente de correlação parcial.

Segundo Fávero e Belfiore (2017), a Análise Fatorial pode ser inadequada caso os valores da estatística KMO sejam iguais ou inferiores a 0,6.

Outra medida para quantificar o grau de intercorrelações entre as variáveis e a adequação da análise fatorial é a MSA (HAIR et al., 2009). A MSA é uma medida semelhante a medida KMO, ou seja, ela relaciona os valores de correlação comum com as correlações parciais, porém, nesse caso são incluídos apenas os coeficientes que se deseja analisar (FÁVERO et al., 2009). A equação (7) descreve a medida MSA.

$$MSA = \frac{\sum_{i \neq j} {r_{ij}}^2}{\sum_{i \neq j} {r_{ij}}^2 + \sum_{i \neq j} {a_{ij}}^2} \tag{7}$$

onde:

- r_{ij} é o coeficiente de correlação entre as variáveis;

- a_{ij} é o coeficiente de correlação parcial.

O índice varia de 0 a 1, alcançando 1 quando cada variável é perfeitamente prevista sem erro pelas outras variáveis (HAIR et al., 2009). Quanto maior o valor dessa variável, melhor será a realização da Análise Fatorial. Segundo Hair e outros (2009), o pesquisador deve ter um valor MSA geral acima de 0,50 antes de proceder com a análise fatorial. Caso não tenha, o pesquisador deve analisar os valores de MSA individuais para cada variável, eliminar as variáveis que tenham MSA menores que 0,5 e depois efetuar o cálculo do MSA geral para verificar o objetivo atendido.

Ao realizar as análises apresentadas na presente seção, é possível concluir se o conjunto de variáveis é apto ou não a realização da Análise Fatorial e selecionar as variáveis para execução dela, para que, em seguida, seja aplicada a extração de fatores iniciais.

4.3.1.2 Extração de Fatores

Após a validação da aplicação da Análise Fatorial, deve-se examinar as variáveis de forma a extrair os fatores iniciais e, logo em seguida, definir a quantidade de fatores que irão descrever adequadamente os dados. Segundo Hair e outros (2009), as decisões devem ser tomadas com relação ao método de extração dos fatores e ao número de fatores selecionados para explicar a estrutura latente dos dados.

Os dois principais métodos de extração dos fatores são o de Análise dos Componentes Principais (PCA, do inglês *Principal Component Analysis*) e Análise dos Fatores Comuns (FA, do inglês *Factor Analysis*). A escolha do método deve ser baseada no objetivo da análise fatorial e no conhecimento prévio acerca da variância das variáveis. Ao se tratar do objetivo da análise, utiliza-se PCA quando o objetivo é reduzir os dados a um número mínimo de fatores necessários para explicar o máximo de variância representada pelas variáveis originais para fins de previsão. Quando o objetivo é analisar fatores ou dimensões latentes que reflitam o que as variáveis têm em comum, deve-se utilizar FA (HAIR et al., 2009).

Em relação ao conhecimento acerca da variância das variáveis, deve-se levar em conta a metodologia aplicada em cada um dos procedimentos. A PCA considera a variância total em seus cálculos, dessa forma, para a realização dessa análise, deve-se ter conhecimento acerca dos

dados de forma que a variância presente nos dados que corresponde aos erros que são resultados da não confiabilidade da aquisição de dados do processo tenham uma proporção pequena em relação a variância total dos dados das variáveis (HAIR et al., 2009).

A Análise de Fatores comuns, diferentemente, considera que as variâncias correspondentes a erros são específicas e únicas para apenas uma variável e não são interessantes para a definição da estrutura das variáveis. Dessa forma, considera-se apenas as variâncias comuns (HAIR et al., 2009).

No presente trabalho, como o objetivo é reduzir a quantidade de variáveis relacionadas para aplicação em uma máquina de inteligência e observar as correlações entre as variáveis, será utilizada a técnica de PCA. Essa nova apresentação dos dados pode, também, auxiliar na interpretação dos mesmos, evidenciar tendências, tornar clara a relevância das variâncias das variáveis para a variância original dos dados e mostrar redundâncias (ROSSI, 2017).

A PCA converte um conjunto original de variáveis em um novo conjunto de variáveis que carregue maior parte da informação do sinal original de forma que os fatores possam ser obtidos por meio de uma combinação linear das variáveis originais (HAIR et al., 2009).

Segundo Rossi (2017), o algoritmo para aplicação de PCA se resume aos seguintes passos. Sejam as variáveis Z_1, Z_2, ... Z_p cada uma com n dados associados que compõe uma matriz de dados n x p, conforme equação (8).

$$Z = \left[Z_1\ Z_2\ \dots\ Z_p\right] = \begin{bmatrix} z_{11} & \cdots & z_{1p} \\ \vdots & \ddots & \vdots \\ z_{n1} & \cdots & z_{np} \end{bmatrix} \tag{8}$$

Para evitar a sensibilidade da PCA em relação a diferença de escala entre medições pode-se optar por realizar a normalização dos dados. É então calculada a matriz de covariância da matriz resultante dos passos anteriores. Caso os dados estejam normalizados, essa matriz corresponderá a matriz de correlação (ROSSI, 2017).

São calculados, então, os autovetores e autovalores a partir da matriz de covariância. Os autovalores devem ser ordenados de acordo com os autovalores associados. O primeiro será o

componente principal, o segundo será a segunda componente principal e assim por diante (ROSSI, 2017).

O pesquisador, deve então, definir o número de fatores adequados a sustentação e representação de um determinado grupo de variáveis. Para isso, combina-se o entendimento conceitual do processo e estrutura das variáveis além de se basear em diferentes critérios como: critério da raiz latente, critério *a priori*, critério de percentagem de variância, critério do teste Scree e heterogeneidade dos respondentes (HAIR et al., 2009).

O Critério *a Priori* é realizado de forma que o pesquisador utiliza de uma quantidade de fatores pré-definida. O pesquisador comumente aplica esse critério quando ele possui conhecimento prévio acerca das variáveis e objetivos, se deseja aferir uma teoria pré-definida ou ainda replicar a análise de outro pesquisador.

"O critério de percentual de variância é uma abordagem baseada na conquista de um percentual cumulativo especificado da variância total extraída por meio de fatores sucessivos" (HAIR et. al., 2009, p. 114). Objetiva-se, por meio dessa metodologia, garantir que os fatores escolhidos explicitem ao menos um valor percentual da variância.

O pesquisador pode utilizar os critérios anteriormente citados de forma a orientá-lo a interpretação dos fatores. Dessa forma, deve-se avaliar diferentes estruturas de fatores previamente a definição final da solução. A seleção dos fatores é realizada a partir de avaliações das estruturas montadas de forma a não selecionar fatores em excesso ou em falta, garantindo uma estrutura correta sem omissão de resultados e facilidade no momento de análise. No presente trabalho, será aplicado, diante da quantidade pequena de conjuntos possíveis e facilidade na definição da quantidade de valores associados, o critério percentual da variância.

Para obter a matriz de componentes principais, é efetuado o produto matricial entre os dados de entrada normalizados e os autovetores selecionados, conforme equação (9).

$$CP = \tilde{Z}A \tag{9}$$

Sendo CP os componentes principais. A a matriz de autovetores e $\tilde{Z}$ a matriz de dados normalizados. Cada coluna da nova matriz apresenta um novo componente que reúne informações das variáveis que a PCA foi aplicada. Os novos componentes podem contribuir de maneira mais eficazes a treinamentos de máquinas de aprendizado.

5 REDES NEURAIS ARTIFICIAIS

Uma Rede Neural Artificial (RNA) é uma máquina projetada para modelar a maneira como o cérebro realiza uma tarefa de interesse. A rede é normalmente implementada utilizando-se de programação em um ambiente digital. As redes neurais empregam uma interligação de células computacionais simples unidas por um conjunto de conexões, normalmente unidirecionais, denominadas neurônios, que lembram a estrutura do cérebro humano (HAYKIN, 2001; BRAGA; LUDEMIR; CARVALHO, 2007).

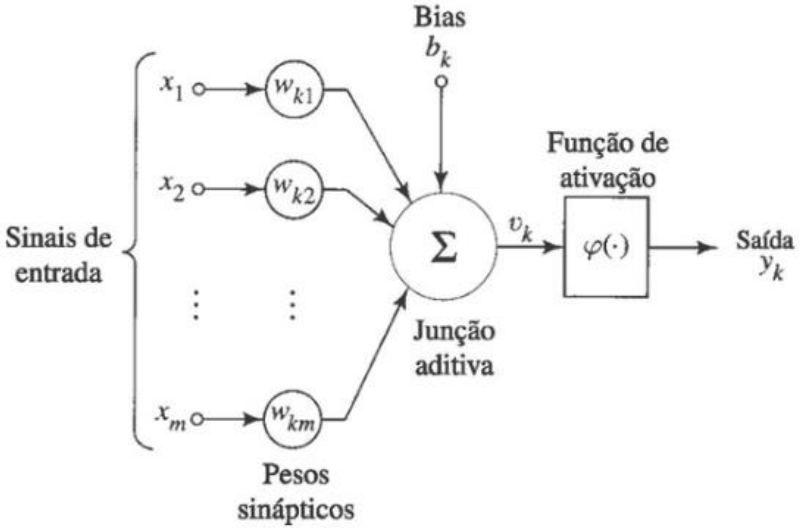
"Uma rede neural é um processador maciçamente paralelamente distribuído constituído de unidades de processamento simples, que têm a propensão natural para armazenar conhecimento experimental e torná-lo disponível para uso" (HAYKIN, 2001, p. 28). O neurônio é uma unidade de processamento de informação que é fundamental para a operação de uma rede neural. Um neurônio, é formado por um conjunto de elos de conexão de entrada, cada um associado a um peso particular denominados pesos sinápticos, um somador que associa os sinais de entrada ponderados pelos pesos e uma função de ativação para restringir a amplitude da saída de um neurônio (HAYKIN, 2001).

O modelo de neurônio artificial proposto por McCulloch e Pitts (MCCULLOCH; PITTS, 1943) é formado por uma simplificação do que se tinha conhecimento acerca de um neurônio no momento que foi criado. Conforme pode ser observado na Figura 18, o neurônio proposto é composto pelos sinais de entrada (x_1, x_2, x_3, ..., x_n), seus pesos sinápticos w_{kj}, onde k representa o índice do neurônio ao qual o peso se refere, um limiar correspondente ao neurônio k, b_k, uma função de ativação φ e uma saída y_k.

A função de ativação é geralmente utilizada com dois propósitos: limitar a saída do neurônio e introduzir não-linearidade no modelo. O *bias*, b_k, tem o papel de aumentar ou diminuir a influência do valor da entrada líquida para a ativação do neurônio k. É possível descrever a equação matemática que representa um neurônio por meio da equação (10).

$$y_k = \varphi\left(\sum_{j=1}^{n} w_{kj}.x_j + b_k\right) \tag{10}$$

Figura 18 – Modelo não linear de um neurônio



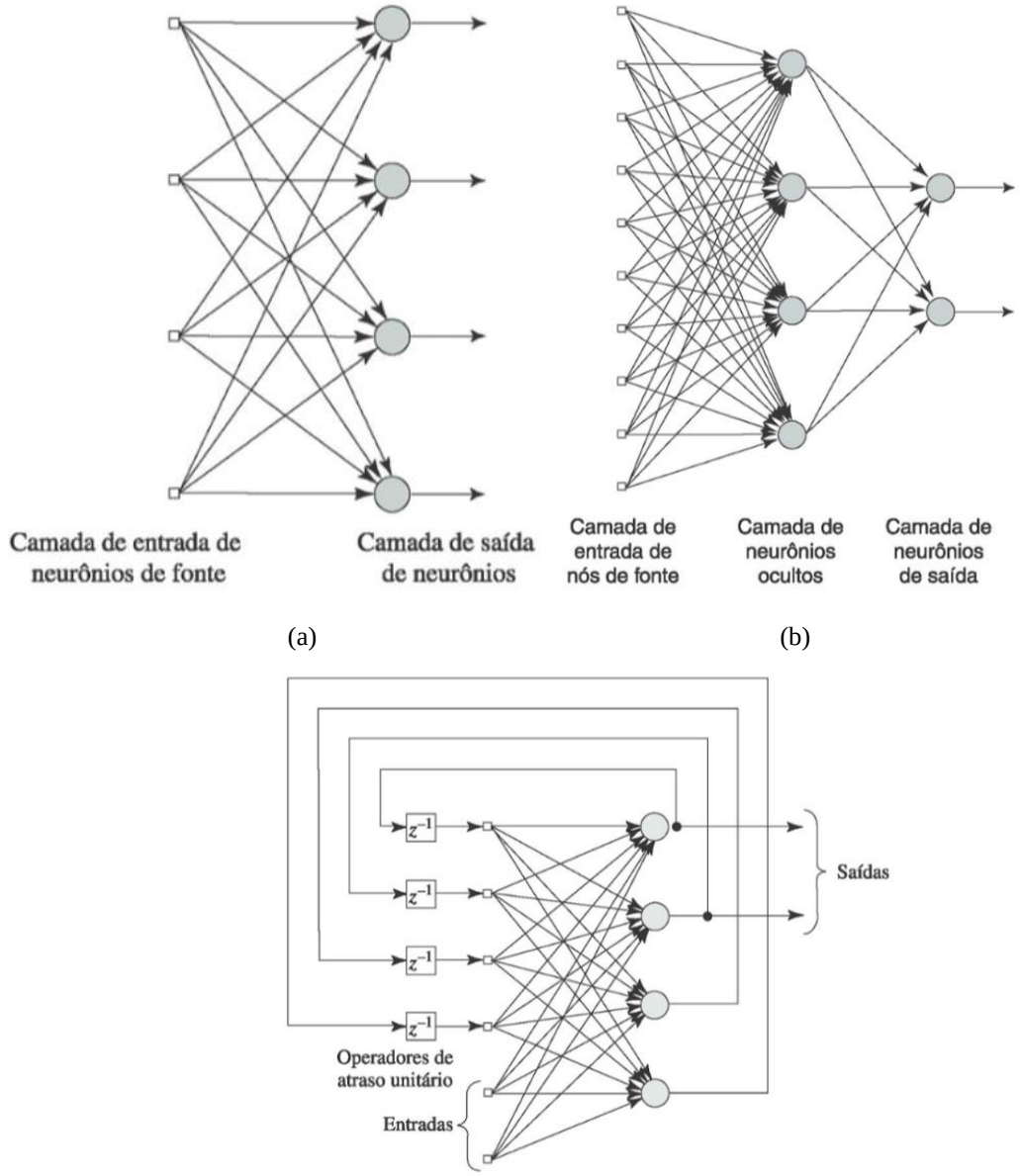
Fonte: Haykin (2001).

Os neurônios, quando estruturados em uma rede, se tornam capazes de resolver problemas de alta complexidade (BRAGA; LUDEMIR; CARVALHO, 2007). A estrutura de uma rede neural pode assumir três classes principais de arquitetura: redes alimentadas adiante com camada única (Figura 19 (a)), redes alimentadas diretamente com múltiplas camadas (Figura 19 (b)) e redes recorrentes (Figura 19 (c)).

A rede mais simples é a de camada única de neurônios. A camada única é referente a quantidade de saída dos neurônios. Uma rede *feedfoward* de uma única camada é apresentada pela Figura 19 (a). Nessa rede a propagação do sinal ocorre em apenas um único sentido, da entrada para a saída. Ao adicionar uma ou mais camadas ocultas, conforme é possível observar pela Figura 19 (b), torna-se possível discriminar estatísticas de ordens mais elevadas. As saídas de uma camada servem, em geral, como alimentação para a próxima camada (HAYKIN, 2001; BRAGA; LUDEMIR; CARVALHO, 2007).

Como é possível observar por meio da Figura 19 (c), um laço de realimentação pode ser adicionado a rede tornando uma rede neural recorrente. Essa adição impacta positivamente na capacidade de aprendizado da rede e no desempenho dela, pois as saídas da rede neural passam a se basear, além das entradas, de informações passadas da própria variável (HAYKIN, 2001). Ela pode, além disso, se basear apenas na própria variável em intervalos de tempo anteriores para seu próprio aprendizado.

Figura 19 – Arquiteturas de redes neurais (a) Rede alimentada adiante com uma camada de neurônios (b) Rede alimentada adiante com duas camadas de neurônios (c) Rede recorrente



Fonte: Haykin (2001).

Essa alternativa é solucionada por um modelo não linear autorregressivo (NAR, do inglês *Nonlinear AutoRegressive model*). Nesse modelo autorregressivo a rede neural utiliza apenas valores passados da saída para previsão de seu comportamento futuro. A rede NAR pode ser representado pela equação (11).

$$\hat{y}(l) \;=\; \varphi\big[\, y(l-1), y(l-2), \ldots, y\big(l-n_y\big)\big]^T \tag{11}$$

onde $\hat{y}(l)$ é a saída estimada do modelo no instante l; $y(l)$ é a saída do modelo; n_y é o número de atrasos da memória de saída e $\varphi(\cdot)$ é a função de ativação.

A capacidade de aprender por exemplos é uma das características mais importantes de uma RNA. O aprendizado de uma RNA está presente em um processo iterativo de ajuste dos pesos das conexões. Ao final do processo, esses pesos armazenam o aprendizado externo dos dados da rede (MENDEL, 1970 apud SCHIAVONI, 2010).

"O conceito de aprendizado está relacionado a melhora do desempenho da rede segundo um critério preestabelecido" (SCHIAVONI, 2010, p. 23). O erro quadrático médio é uma das alternativas a serem utilizadas como critério de desempenho pelos algoritmos de correlação de erros. O que diferencia os algoritmos de aprendizado está centrado principalmente na forma de atualização dos pesos. Inicialmente, os pesos sinápticos são configurados aleatoriamente e conforme novas informações vão sendo inseridas, seus valores vão sendo alterados (BRAGA; LUDEMIR; CARVALHO, 2007; SCHIAVONI, 2010).

Em geral, algoritmos de aprendizagem podem ser divididos em duas classes: aprendizado supervisionado e aprendizado não-supervisionado. Aprendizado supervisionado implica, necessariamente, a existência de um supervisor, ou professor externo, o qual é responsável por estimular as entradas da rede por meio de padrões de entrada e observar a saída calculada, comparando-a com a saída desejada. Como a resposta da rede é função dos valores atuais do seu conjunto de pesos, estes são ajustados de forma a aproximar a saída da rede da saída desejada (BRAGA; LUDEMIR; CARVALHO, 2007). Em problemas de modelagem de sistemas e previsão, nos quais a resposta desejada é conhecida, o algoritmo é considerado supervisionado (HAYKIN, 2001).

No presente trabalho as redes neurais artificiais serão utilizadas para regressão de forma a realizar a interpolação suave dos dados em análise, especialmente em momentos nos quais há ausência contínua de registro das variáveis. Para essa aplicação, é suficiente utilizar a estrutura NAR, considerando, apenas os registros passados da própria variável.

6 PLANTA EM ESTUDO E VARIÁVEIS EM ANÁLISE

O processo de sinterização da planta industrial é monitorado em diferentes aspectos para garantia de um processo de dinâmico e de qualidade. O operador da planta toma decisões em relação a essas medições para realizar um adequado controle do processo de produção.

O presente trabalho tem o foco de analisar a qualidade e o comportamento das variáveis do processo ao longo do tempo e estudar a relação entre suas variações e o local em que se encontra BTP. Nesse ponto ocorre o maior valor de temperatura, local onde o calor fornecido pelo forno de ignição, por meio da contribuição dos exaustores e combustão do material interno adicionado, promove a fusão parcial de todo o material para a formação do sínter.

Após atingir o valor de máxima temperatura na zona de reação (ver Figura 3), o sínter é resfriado e solidificado por maquinário específico. Para que o resfriamento inicie em um ponto apropriado em relação a esteira, deve-se garantir uma posição adequada do BTP. Caso o início do resfriamento ocorra em local inadequado a produtividade pode ser diminuída e serem gerados danos aos equipamentos responsáveis pelo resfriamento.

A esteira de sinterização da planta a ser estudada apresenta área útil de 484 m², largura de 5,5 m e comprimento de 88 m (TELLES, 2010). A planta contém 23 caixas de vento ao longo da esteira dispostas entre distâncias equivalentes a 4m uma da outra, cada uma delas apresenta monitoramento de temperatura por meio de termopares. Para o cálculo do BTP, a cada instante, são utilizadas as medições das últimas cinco caixas de vento para obtenção da curva quadrática associada e registro da posição e temperatura do BTP, conforme explicado na seção 2.1.3.

Foram disponibilizados, pela empresa detentora da planta, os dados de trinta e seis variáveis do processo, selecionadas por engenheiros especialistas, as quais incluem as temperaturas das caixas de vento, os valores de posição e temperatura do BTP e variáveis consideradas potencialmente significativas para observação do comportamento do BTP.

Para a realização de uma análise conceitualmente adequada destinada a observação de influências entre as variáveis do processo, não se pode levar em conta uma possível correlação entre as temperaturas das caixas de vento e o valor de temperatura e posição do BTP, já que

esses valores são obtidos diretamente por meio de aproximação quadrática da curva de temperatura formada por essas medições. Considera-se, então, uma prudência conceitual observar e analisar as possíveis influências das demais variáveis em relação a alteração e disposição das temperaturas das caixas de vento quantificadas ao final do processo.

Para todas as variáveis do processo foram fornecidos dados a cada uma hora. Entretanto existiram variáveis que apresentavam medições equivalentes a cada dia ou a cada seis horas. Isso ocorreu, pois, a periodicidade de medições dessas variáveis eram, respectivamente, 24 horas e 6 horas e a empresa responsável pelo processo atribuiu uma única medição a todo o intervalo. Não é informado o horário no qual é realizado o registro dessas variáveis para estimativa de seu comportamento ao longo do intervalo tratado.

As variáveis fornecidas, junto de sua unidade de medida e respectiva taxa de amostragem, são apresentadas no Quadro 1.

Quadro 1 – Variáveis do processo

(continua)

Nº	Variáveis do processo	Unidade de medida	Amostragem
1	Consumo de Cal	%	24 horas
2	Base de Coque	%	24 horas
3	Pilha Homogeneizada A	ton	1 hora
4	Pilha Homogeneizada B	ton	1 hora
5	Total Pilha Homogeneizada	ton	1 hora
6	RDEG na Pilha Homogeneizada	%	24 horas
7	SFBR no sínter *feed*	%	24 horas
8	STFA no sínter *feed*	%	24 horas
9	+1mm na Pilha Homogeneizada	%	24 horas
10	Relação +1/-0,105 na Pilha Homogeneizada	-	24 horas
11	Fração +1 mm na Mistura Total	%	24 horas
12	Fração -0,105 mm na Mistura Total	%	24 horas
13	Relação +1/-0,105 mm na Mistura Total	-	24 horas
14	RDEG na Mistura Total	%	24 horas
15	Densidade da Mistura	t/m³	1 hora
16	Altura da Camada	mm	6 horas
17	Velocidade de Sinterização	mm/min	1 hora
18	Velocidade da Máquina de Sínter	m/min	1 hora
19	Temperatura da Caixa de Vento nº11	°C	6 horas
20	Temperatura da Caixa de Vento nº13	°C	6 horas
21	Temperatura da Caixa de Vento nº18	°C	6 horas
22	Temperatura da Caixa de Vento nº19	°C	1 hora
23	Temperatura da Caixa de Vento nº20	°C	1 hora
24	Temperatura da Caixa de Vento nº21	°C	1 hora
25	Temperatura da Caixa de Vento nº22	°C	1 hora

Quadro 1 – Variáveis do processo

(conclusão)

Nº	Variáveis do processo	Unidade de medida	Amostragem
26	Temperatura da Caixa de Vento nº23	°C	1 hora
27	Pressão na Caixa de Vento nº22	mmH2O	1 hora
28	Pressão na Caixa de Vento nº23	mmH2O	1 hora
29	Vazão do Exaustor Nº 1	Nm^3/min	1 hora
30	Vazão do Exaustor Nº 2	Nm^3/min	1 hora
31	Pressão do Exaustor	mmH2O	1 hora
32	Temperatura BTP	°C	1 hora
33	Posição do BTP	%	1 hora
34	CO	mg/Nm^3	1 hora
35	Produtividade	$ton/dia*m^2$	6 horas
36	Geração de Finos de Retorno	ton/dia	1 hora

Fonte: Reis (2018).
Nota: Adaptado pelo autor.

As variáveis de número um a quatorze, indicadas no Quadro 1, correspondem a parâmetros relativos à composição da mistura de entrada do processo. O produto homogeneizado utilizado como entrada é alocado em duas pilhas, denominadas A e B. Em relação a essas pilhas, as medidas da quantidade do produto homogeneizado pelas variáveis, são denominadas “Pilha Homogeneizada A” e “Pilha Homogeneizada B” e a soma dos pesos das duas pilhas na variável “Total Pilha Homogeneizada”. Como se é interessado em manter os pesos das pilhas de entrada na medida do possível constante, no presente trabalho será utilizada apenas efeito de informação de medida de massa de entrada a variável “Total Pilha Homogeneizada”.

É identificado no sínter *feed*, mistura constituída essencialmente por minérios e minerais de ferro, o percentual SFBR e STFA, isto é, o percentual de minério de ferro provenientes das minas de Brucutu e de Fábrica, respectivamente, que correspondem a minérios base, de menor qualidade.

Na pilha homogeneizada, formada, por exemplo, de minério de ferro, sínter, pelotas e calcário são medidas as variáveis “RDEG na Pilha Homogeneizada”, “+1 mm na Pilha Homogeneizada” e “Relação +1/-0,105 mm na Pilha Homogeneizada”. A variável “RDEG na Pilha Homogeneizada” corresponde a quantidade de finos de minério do peneiramento da mistura que são reutilizados no processo. A variável “+1 mm na Pilha Homogeneizada”, corresponde ao percentual de componentes com granulometria maior que 1mm. A variável “+1/-0,105 mm na Pilha Homogeneizada”, corresponde a relação do percentual de componentes com

granulometria maior que 1 mm e o percentual de componentes de granulometria menor que 0,105 mm.

Ao final, é formada a mistura total composta pela pilha homogeneizada e o combustível sólido adicionado a mistura. Sob ela são medidos os percentuais da composição de cal, coque, percentual de componentes de granulometria maior que 1 mm, percentual de componentes de granulometria inferior a 0,105 mm, relação entre os percentuais de granulometria maior que 1 mm e de granulometria menor que 0,105 mm, e o percentual de RDEG, finos de reutilização em relação a mistura total.

As variáveis de número quinze a dezoito, indicadas no Quadro 1, correspondem a parâmetros de controle do processo de sinterização (RIZZO, 2007). A "Densidade da Mistura" é definida de acordo com a composição do material fino inserido para produção do sínter. A composição do material é definida de acordo com a necessidade de composição química exigida pelo alto-forno no momento da produção.

A "Altura da Camada" é um parâmetro da mistura que deve ser definido e operado de forma a se obter uma máxima permeabilidade e eficiência térmica do processo. A máxima permeabilidade ocorre em pequena altura da camada, enquanto a máxima eficiência térmica ocorre para maiores alturas (RIZZO, 2007). O operador deve, então, definir uma altura adequada da camada de modo que os dois parâmetros estejam balanceados.

A "Velocidade de Sinterização" e a "Velocidade da Máquina de Sínter" são os parâmetros a serem controlados que são mais sensíveis, pois, qualquer alteração na permeabilidade da camada é acompanhada de uma alteração da velocidade da esteira e na de sinterização (RIZZO, 2007). A velocidade de sinterização corresponde a velocidade com que a combustão se propaga ao longo da camada. A velocidade da máquina de sínter, corresponde a velocidade com que é comandada a esteira. As duas velocidades devem estar em sincronia, pois, dessa forma obtém-se sua máxima capacidade de produção e condução adequada da queima durante o processo.

Vale ressaltar que uma variável do processo, a "Intensidade de Ignição", promovida pelo forno de ignição, não foi fornecida pela empresa que conduz o processo de sinterização a ser

analisado. Ela influencia diretamente no início do processo de fusão do material fino, durante aquecimento e formação da camada superficial do sínter (RIZZO, 2007).

As variáveis de número dezenove a trinta e um, indicadas no Quadro 1, são indicadores de controle do processo de sinterização. A "Temperatura das Caixas de Vento" normalmente é medida por termopares próximos as caixas de vento. Uma das informações importantes, advinda das medições de temperatura, é o local onde ocorre a máxima temperatura do processo. Normalmente, controla-se a posição do ponto final de queima para que se situe próximo à penúltima ou antepenúltima caixas de vento (RIZZO, 2007).

A vazão de ar nos exaustores e a pressão nas caixas de vento e nos exaustores são variáveis que apresentam relação direta com a permeabilidade da combustão através das camadas do sínter. A pressão na caixa de vento e nos exaustores no início do processo de sinterização é normalmente mais alta, indicando uma menor permeabilidade do material, e menor ao final do processo. As medidas de pressão, entretanto, podem variar de acordo com o material adicionado e pode ser um indicativo da qualidade e porosidade do material, sendo utilizada, por exemplo, como medida para alteração de outros parâmetros, como altura da camada, e velocidade da máquina de sinterização (RIZZO, 2007).

As variáveis trinta e dois e trinta e três são referentes a temperatura e a posição do BTP que corresponde, respectivamente, a maior temperatura durante o processo e o local da esteira no qual ela ocorre. Essas informações são obtidas por cálculo com base nas últimas medições nas caixas de vento presentes ao longo da esteira.

As variáveis de número trinta e quatro a trinta e seis, no Quadro 1, são indicadores de saída do processo, englobando emissões, produtividade e produtos a serem reutilizados. A variável "CO" indica a quantidade de monóxido de carbono que foi emitido durante o processo de sinterização. Essa variável é medida por meio da relação, em miligramas, da quantidade de CO emitido em relação a uma unidade de volume utilizada apenas pela empresa responsável pela planta, sem significado revelado, Nm^3 [2].

[2] Leia-se, metro cúbico normalizado.

A cada dia são medidas a “Produtividade” em relação a produção de sínter, a “Geração de Finos de Retorno”, que corresponde a quantidade, em toneladas por dia, de material fino que não foi aglomerado ao longo do processo de sinterização e deve ser utilizado novamente como matéria-prima.

7 EXAME DOS DADOS ORIGINAIS

Serão realizados, na presente seção, o exame inicial dos dados do processo de sinterização. As etapas inicialmente desenvolvidas são consideradas essenciais para o início da realização de análises multivariadas. É possível, por meio dela, compreender, de maneira mais aprofundada, o comportamento dos dados ao longo do período em estudo (HAIR et al., 2009). Será levado em conta o conhecimento conceitual e operacional do processo para avaliar cada um dos parâmetros.

Esses exames são essenciais para identificar observações faltantes e atípicas. Eles auxiliam a evitar uma distorção e invalidação da análise multivariada dos resultados, pois, por meio deles, juntamente ao conhecimento conceitual prévio do pesquisador, que se constrói um domínio dos fundamentos associados aos dados. Nesse sentido, serão realizadas investigações em relação ao comportamento, *outliers* e *missing data*.

Será avaliada, inicialmente, a dinâmica de cada uma das variáveis do processo por meio dos dados crus. Será plotado, para cada uma delas, o gráfico de linhas com o objetivo de observar o intervalo de variação dos dados associados em relação ao seu papel no processo, a ausência de registro de dados e a presença de *outliers*.

A apresentação dos gráficos de linhas será dada a cada hora, tomando como hora e data inicial, 00h00 do dia 24 de agosto de 2017, e hora e data final 23h00 do dia 23 de agosto de 2018, totalizando 8760 registros horários por variável.

7.1 Parâmetros Relativos à Composição da Mistura de Entrada do Processo

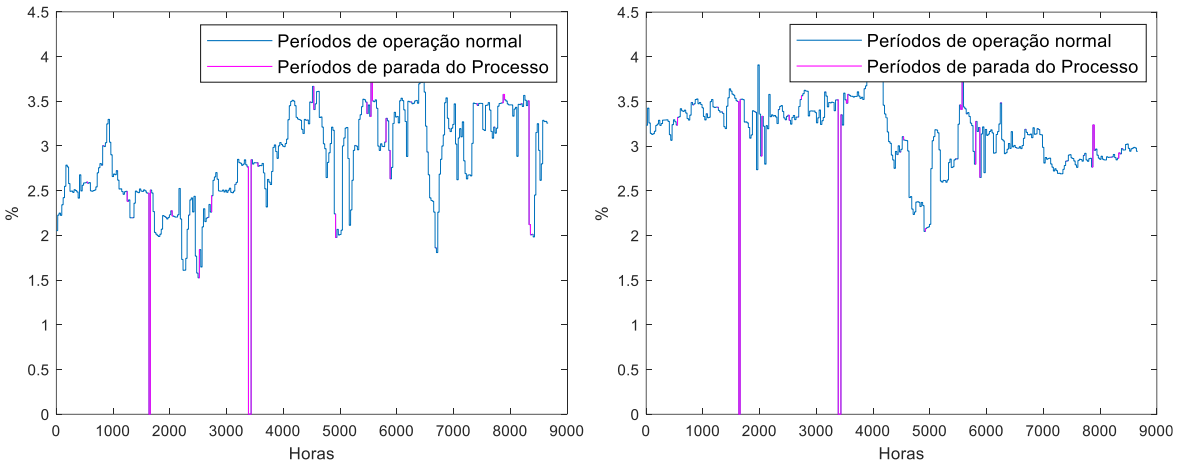
Os parâmetros aqui tratados são relativos à composição da mistura de entrada do processo. Com exceção dos parâmetros de medida de peso das pilhas homogeneizadas, os registros das variáveis são realizados a cada 24 horas.

As duas primeiras variáveis, “Consumo de Cal” e “Base de Coque”, são relativas à composição percentual em relação ao total da massa da composição da pilha homogeneizada. São mostrados por meio do Gráfico 1, o gráfico de linhas de cada uma das variáveis. É possível que a variação

do percentual de cada componente varia ao longo do tempo, refletindo a alteração da demanda desses componentes para a composição do sínter a ser utilizado nos altos-fornos da empresa.

Por meio do Gráfico 1, é possível observar com clareza dois intervalos nos quais as varáveis se igualam a zero, entre as amostras 1615 a 1656 e 3373 a 3432. Nesses instantes ocorreram medições iguais a zero ou ausência consecutiva de um número considerável de dados em grande parte das variáveis devido a paradas do processo. Os instantes de parada do processo são indicados em todos os gráficos de forma a observar sua influência em relação a mudanças bruscas nos valores das variáveis. Os momentos de parada serão devidamente explicitados e desconsiderados na seção 9.1.

Gráfico 1 – Gráficos de linha (a) Consumo de Cal (b) Base de Coque

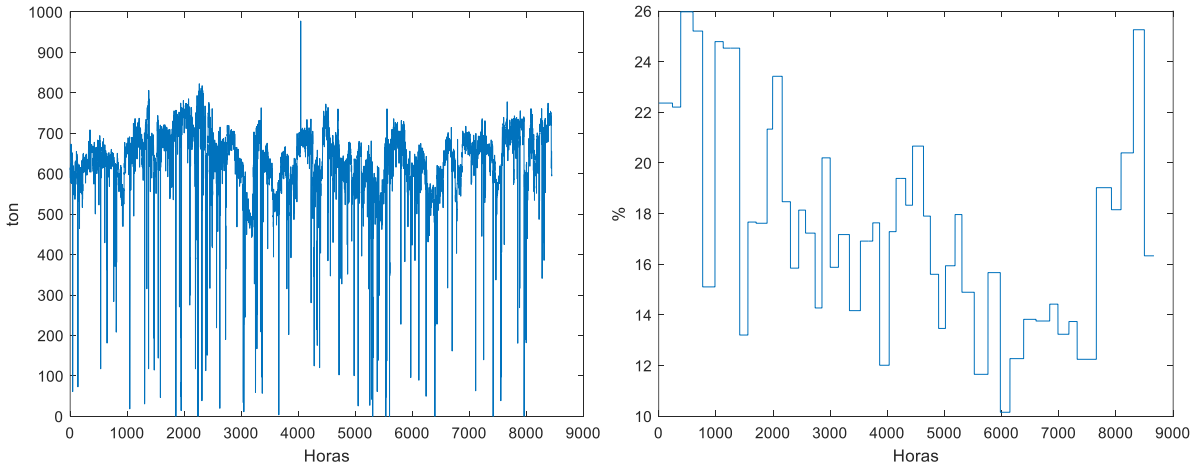


(a) (b)

Fonte: Produção do próprio autor.

A variável “Total Pilha Homogeneizada” se configura como a soma das duas medições de massa (em toneladas) das pilhas homogeneizadas A e B a cada hora. Ela, então, apresenta uma leitura a cada hora do peso total das pilhas de entrada do processo que devem ser mantidos mais próximo da constância ao longo do tempo. Observa-se, por meio do Gráfico 2 (a), uma quantidade significativa de valores destoantes que representam baixa carga nos silos. Entretanto, sua variação gera impactos na dinâmica da esteira, exceto em casos nos quais o carregamento seja descontinuado, o que não ocorre no período tratado.

Gráfico 2 – Gráficos de linha (a) Total Pilha Homogeneizada (b) RDEG na Pilha Homogeneizada



(a) (b)

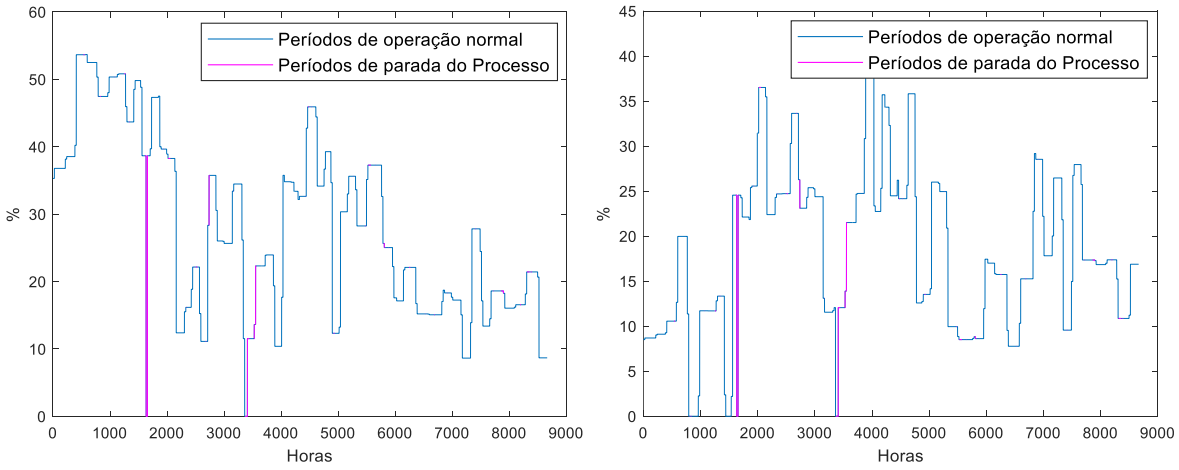
Fonte: Produção do próprio autor.

A variável “RDEG na Pilha Homogeneizada” representa uma composição percentual de finos de minério em relação a massa total da Pilha Homogeneizada. A variável é medida a cada 6 horas e está dentro dos valores típicos esperados. Conforme mostrado pelo Gráfico 2 (b), os seus valores se apresentam constantes dentro de um intervalo de 5 a 7 dias, e, portanto, essa variável não influencia no andamento do processo de forma dinâmica.

As variáveis “SFBR no sínter *feed*” e “STFA no sínter *feed*” correspondem a percentuais de minério de ferro em relação a massa de sínter *feed* de duas origens diferentes, como já mencionado. A escolha do percentual das origens poderá depender da disponibilidade do material e sua qualidade, por essas razões é conceitualmente aceitável que o percentual de cada um dos componentes variar de zero a composição total do sínter *feed*, conforme é possível observar por meio do Gráfico 3.

Deve-se atentar que a soma das duas porcentagens pode não alcançar 100% da composição por conta da adição de outros compostos, entretanto, não pode ultrapassar esse valor. Em momentos de parada da esteira de sinterização, como pode ser visto no Gráfico 3, em dois momentos, os registros das variáveis foram iguais a zero.

Gráfico 3 – Gráficos de linha (a) SFBR no *sínter feed* (b) STFA no *sínter feed*

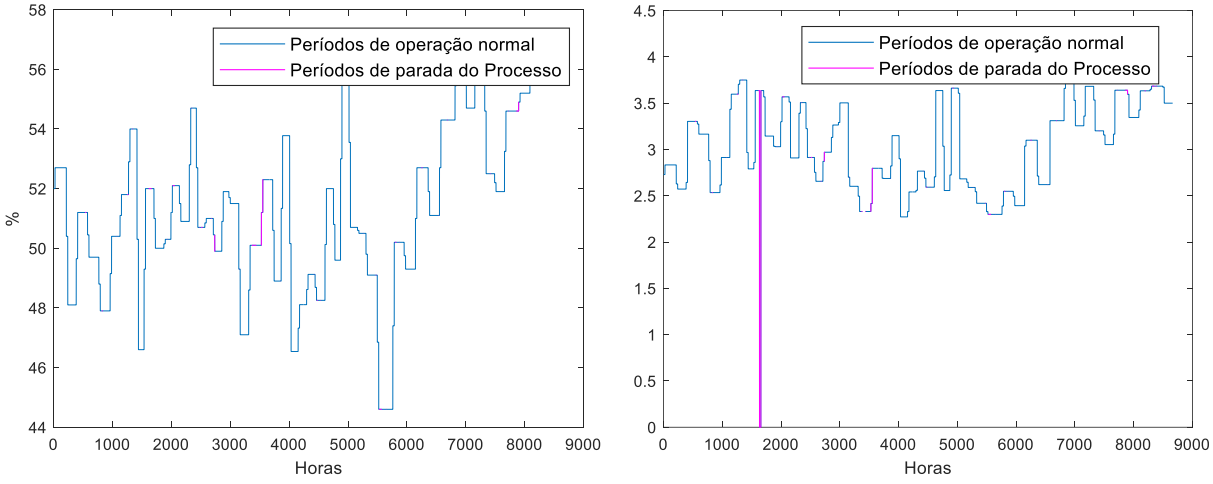


(a) (b)

Fonte: Produção do próprio autor.

Por meio do Gráfico 4 é possível observar variáveis referentes a composição granulométrica da Pilha Homogeneizada. Tanto a variável "+1 mm na Pilha Homogeneizada" como "Relação +1/-0,105 na Pilha Homogeneizada", apresentam valores possíveis em relação ao processo considerando suas unidades de medida. Nota-se que apenas a primeira não apresentou variações em relação a parada da esteira. A segunda, por se tratar de uma relação granulométrica, é esperada conceitualmente a necessidade de maior concentração compostos com granulometria mais alta em relação aos de granulometria mais baixa.

Gráfico 4 – Gráficos de linha (a) Porcentagem +1 mm (b) Relação >1 mm e <0,105 mm Pilha homogeneizada



(a) (b)

Fonte: Produção do próprio autor.

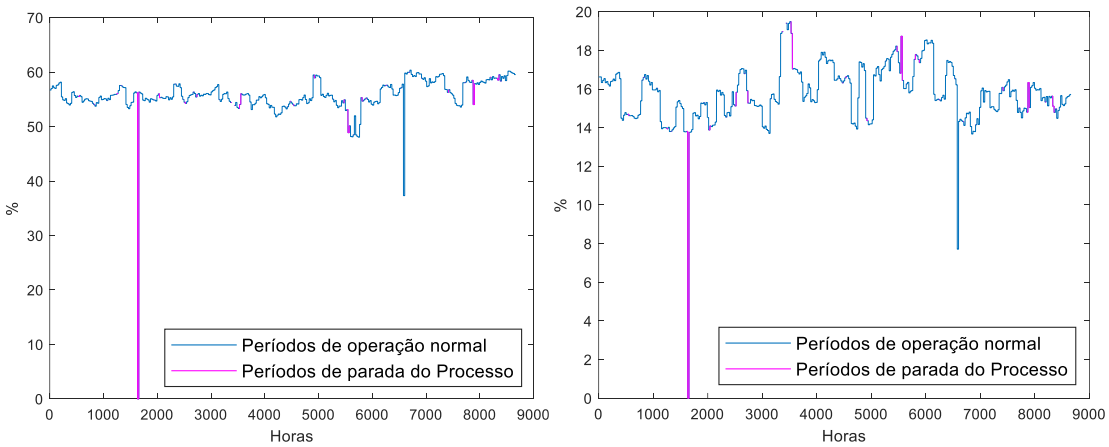
As variáveis “Fração +1 mm na Mistura Total”, “Fração -0,105 mm na Mistura Total”, “Relação +1/-0,105 mm na Mistura Total” e RDEG na Mistura Total identificam a composição final, após *blending* das matérias primas utilizadas como entrada da esteira de sinterização. A representação de cada uma delas ao longo do período analisado pode ser observado por meio do Gráfico 5 e Gráfico 6.

Em cada um dos gráficos, é possível inferir características semelhantes. A mistura de entrada é composta por uma receita, e, assim, os valores se apresentam próximos a um alvo específico. É possível observar variações bruscas associadas a paradas no processo, contudo, no Gráfico 5 (a), (b) e Gráfico 6 (a) e (b) é possível observar uma variação de um dia desassociada a parada do processo, uma falha de medição entre as amostras 6577 e 6600. Nota-se, além disso, pelo Gráfico 5b, falta de completude da medida de Fração de RDEG pelo período de um dia.

Os quatro últimos dias das leituras desse grupo não foram registrados, e, por isso, não aparecem no gráfico. A ausência dos registros referentes aos quatro dias mencionados também ocorre com as demais variáveis que se incluem no grupo de parâmetros relativos à composição da mistura do processo.

Conceitualmente, a composição da matéria-prima influencia, mesmo que de forma sutil, o comportamento das variáveis ao longo do tempo. Os materiais são diferenciados por granulometria e composição e essas características influenciam na absorção e propagação da temperatura ao longo da camada de matéria prima na esteira.

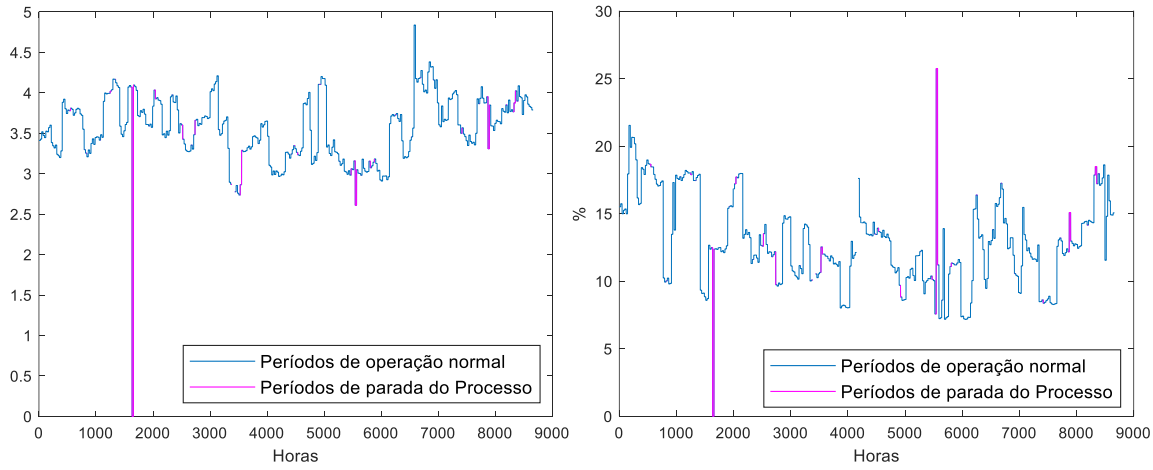
Gráfico 5 – Gráficos de linha (a) Fração Maior que 1 mm (b) Fração Menor que 0,105 mm



(a) (b)

Fonte: Produção do próprio autor.

Gráfico 6 – Gráficos de linha (a) Relação Maior que 1 mm e menor que 0,105 mm (b) Fração de RDEG



(a) (b)

Fonte: Produção do próprio autor.

7.2 Parâmetros de Controle do Processo de Sinterização

Os parâmetros de controle do processo de sinterização incluem variáveis que determinam e alteram a qualidade do sínter e propriedades do sínter ao final do processo, e portanto, essas variáveis são medidas com maior frequência. Dentro do grupo de variáveis estão inclusas “Densidade da Mistura”, “Altura da Camada”, “Velocidade de Sinterização” e “Velocidade da Máquina de Sínter”.

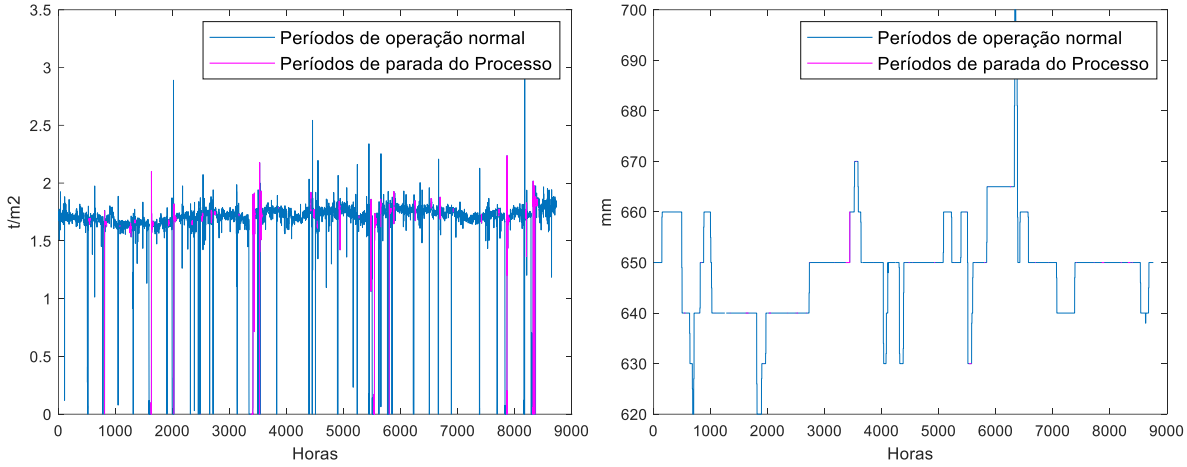
A variável “Densidade da Mistura” apresentou medições iguais a zero em determinados instantes de tempo, conforme o Gráfico 7a. Esses registros indicam a ausência de medição da variável naqueles instantes já que o valor não está compreendido entre os valores possíveis de medição de densidade. Foram identificadas medições abaixo do intervalo possível também em paradas do processo.

Os últimos 29 amostras da variável “Densidade da Mistura”, correspondentes aos dias 22 a partir das 18h00 e o dia 23 de agosto de 2018, não foram registrados no banco de dados em análise e, por isso, não aparecem no Gráfico 7 (a).

A variável “Altura da Camada” é um parâmetro, avaliado por RIZZO (2007), como um parâmetro de controle significativo para mudanças e controle do processo produtivo. Observou-se, entretanto, que as medições, mesmo que realizadas a cada 6 horas, permaneciam constantes,

durante a maior parte do tempo, por seis ou mais dias. Deve-se ressaltar que a variação entre o valor mínimo e o valor máximo dessa amostra é de 5 mm. Dessa forma, entende-se que a variável deve se manter constante, próximo a 650 mm.

Gráfico 7 – Gráficos de linha (a) Densidade da Mistura (b) Altura da Camada



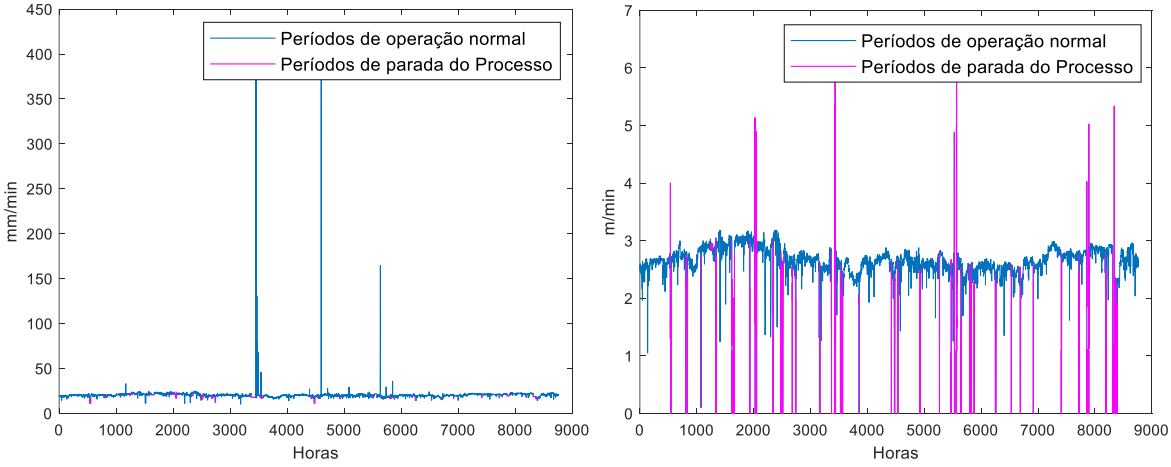
(a) (b)

Fonte: Produção do próprio autor.

A variável “Velocidade de Sinterização”, conforme pode ser observada no Gráfico 8 (a), apresenta-se com valores abruptos em alguns instantes, em instantes próximos as variações abruptas presentes na variável “Temperatura do BTP”, conforme pode ser observado no Gráfico 14 (a). Não se têm informações acerca de como é medido ou registrada essa variável. Assume-se, então, que ela é calculada por ausência de tecnologias associadas a essa medição na planta e por apresentar falhas nos mesmos que a outra, que é calculada.

Em relação a variável “Velocidade da Máquina de Sínter” é diretamente controlada de forma a atingir o funcionamento adequado do processo. Observa-se por meio do Gráfico 8 (b), as paradas do processo, no qual a velocidade é zerada. É possível notar que, durante o momento de desaceleração ou aceleração da esteira durante momentos de parada, há uma falha na leitura da velocidade da esteira, pois apresenta altas medições de velocidade.

Gráfico 8 – Gráficos de linha (a) Velocidade de Sinterização (b) Velocidade da máquina de sínter



(a) (b)

Fonte: Produção do próprio autor.

7.3 Indicadores de Controle do Processo de Sinterização

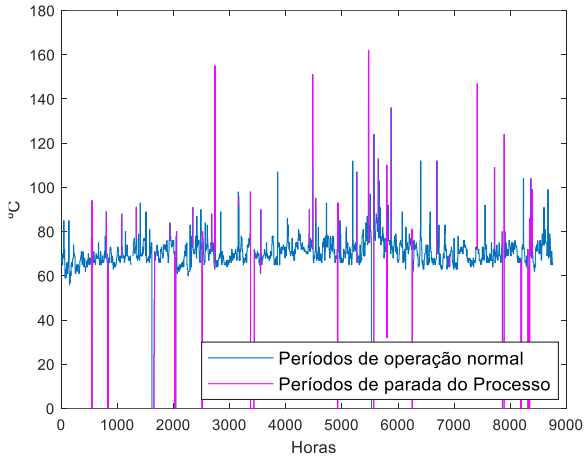
Os indicadores de controle do processo de sinterização correspondem a variáveis que validam a atuação do sistema de controle. Nesse contexto, estão incluídas nesse grupo medições de temperatura, pressão e vazão dos equipamentos do processo.

Como pode ser observado por meio do Gráfico 9 e do Gráfico 10, as temperaturas aumentam à medida que se aproxima das últimas caixas de vento presentes na esteira de sinterização e, nas duas últimas, apresentam diminuição. É possível notar que, apenas para as últimas 5 caixas de vento, a amostragem de registro fornecida foi a cada 1 hora. O registro dessas últimas caixas de vento é de suma importância pois elas indicam a posição de fusão do material que garante a regularização do processo e sua execução de forma adequada.

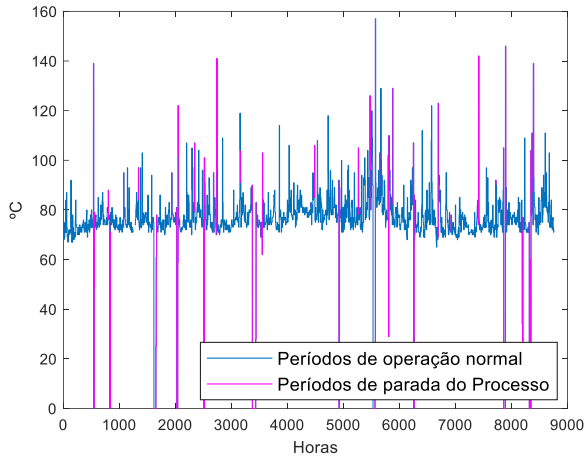
Observa-se que grande percentual de variações abruptas anômalas nos valores das temperaturas se dá em momentos de parada do processo, conforme refletido no Gráfico 9 na cor magenta. Esses dados não correspondem ao funcionamento normal do processo, portanto, devem ser desconsiderados das análises.

Os dados aparecem com grande variação a cada amostra/hora por conta da ausência de medição dos dados de forma mais recorrente. Não é possível ser observada a variação gradual de cada uma das temperaturas ao longo do processo. Essa análise será pormenorizada na seção 8.1.

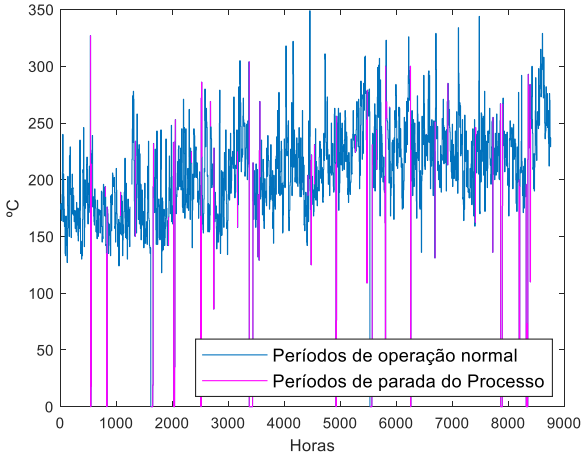
Gráfico 9 – Gráficos de linha temperaturas das Caixas de Vento (a) nº13 (b) nº15 (c) nº 18 (d) nº 19 (e) nº 20 (f) nº 21



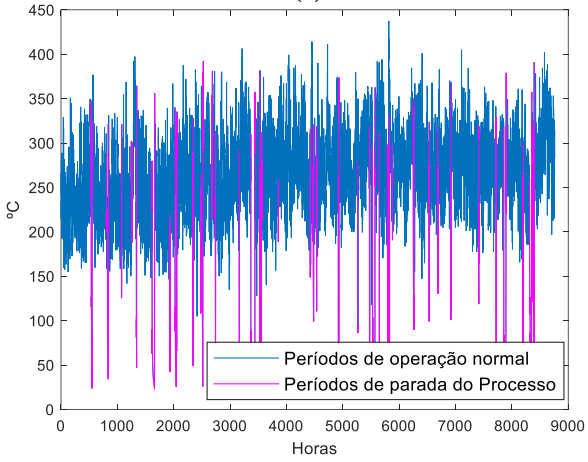
(a)



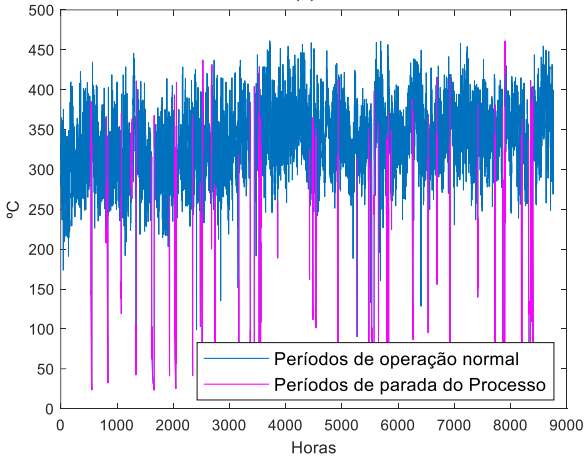
(b)



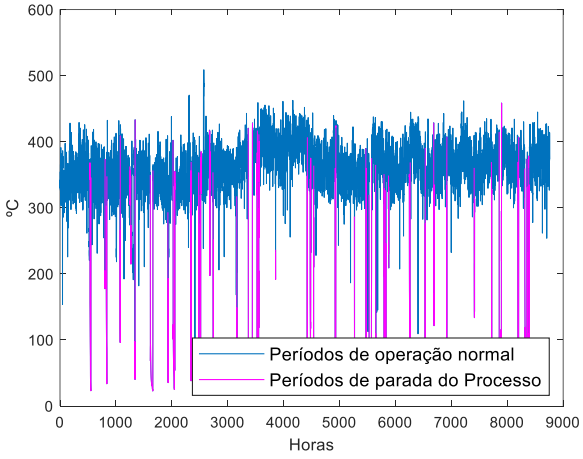
(c)



(d)



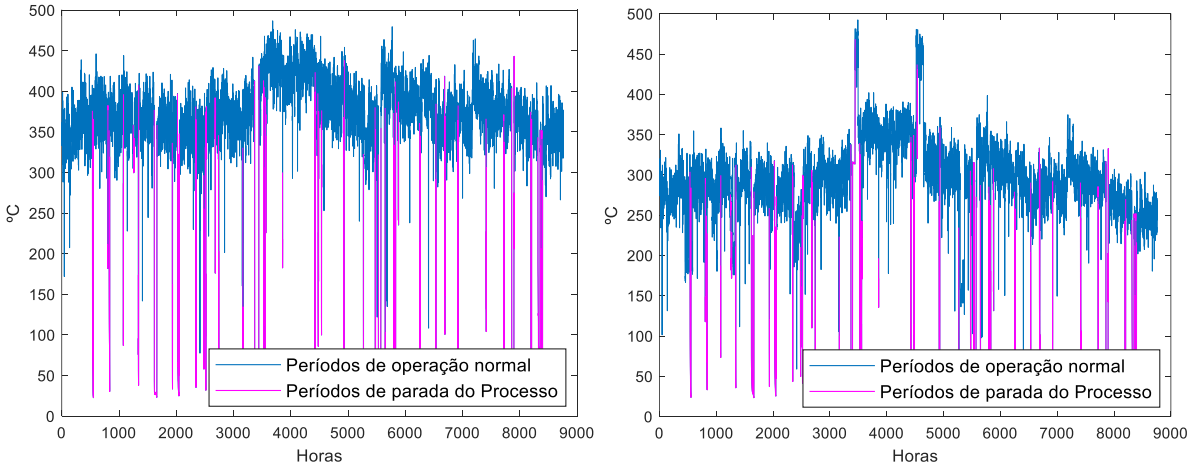
(e)



(f)

Fonte: Produção do próprio autor.

Gráfico 10 – Gráficos de linha temperaturas das Caixas de Vento (a) nº 22 (b) nº23



Fonte: Produção do próprio autor.

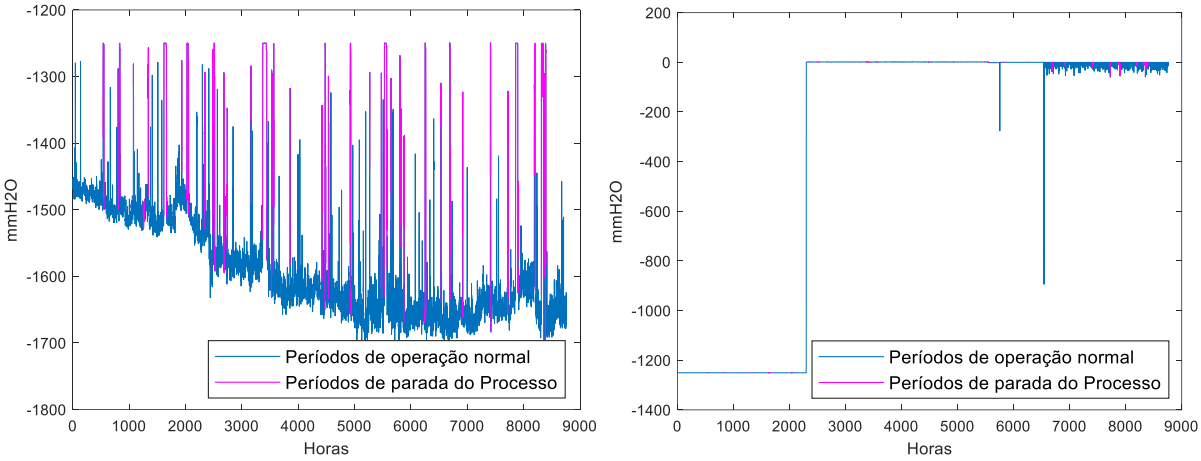
Nota-se que existem dois intervalos nos quais a temperatura da caixa de vento nº 23 se apresenta mais alta do que a média dos outros momentos. Pelo fato de as variações estarem presentes, também, na medição da temperatura da caixa de vento nº 22 e ter se estendido por muitas horas, compreende-se que as amostras não constituem falhas. Esses instantes serão detalhados na seção 9.3 durante o cálculo do BTP.

As pressões nas caixas de vento nº 22 e nº 23 apresentam comportamentos diferentes ao longo do tempo. Enquanto na nº 22, a pressão tende a sofrer um aumento de forma suave ao longo do período em análise e apresenta variações significativas durante as medições, a caixa de vento nº 23 apresenta valores estáticos ao longo do tempo, conforme é possível observar por meio do Gráfico 11 (a).

Verifica-se, também, que, para a caixa de vento nº 22, os valores mínimos de medição, em módulo, indicados como *outliers* são limitados em -1250 mmH_2O, esse fato pode indicar que o medidor apresenta limite mínimo de sua medida nessa faixa de pressão.

Observa-se, por meio do Gráfico 11 (b), que, no instante de tempo 2297, há uma variação de valores próximos a -1250 mmH_2O para valores próximos de 0 mmH_2O na pressão da caixa de vento nº 23. Apesar da variação aparentar brusca, ela não se mostra significativa ao se atentar a unidade de medida que está sendo analisada. A variável, todavia, mostra inconsistência ao apresentar continuidade por um longo período e valores variáveis logo em seguida.

Gráfico 11 – Gráficos de linha das pressões nas Caixas de Vento (a) nº22 (b) nº23

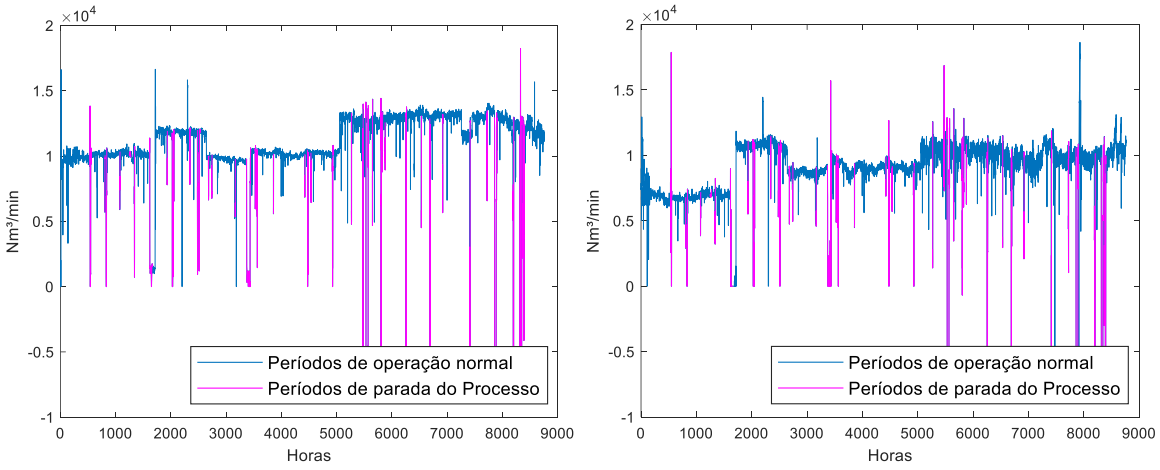


(a) (b)

Fonte: Produção do próprio autor.

As variáveis relacionadas a vazão nos exautores, apresentam tendência a constância por intervalos, conforme é possível observar pelo Gráfico 12. Como sua medição é dada em uma variável particular (Nm³/min), calculada e utilizada apenas pelos responsáveis pela planta, não é possível inferir sobre a significância da variação da vazão em relação ao processo.

Gráfico 12 – Gráficos de linha das vazões dos exaustores (a) nº 1 (b) nº 2



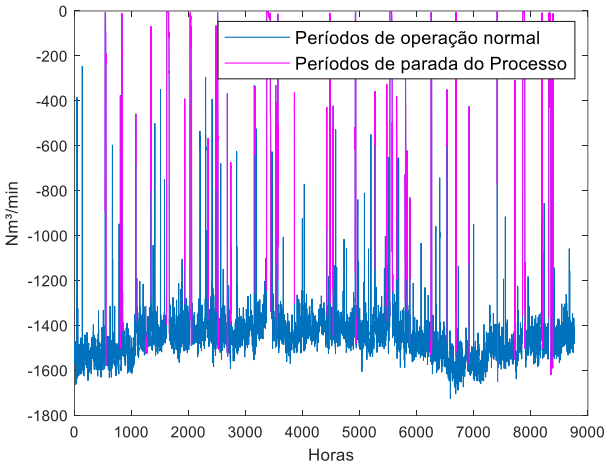
(a) (b)

Fonte: Produção do próprio autor.

Observa-se nos gráficos das variáveis de vazão dos exaustores a presença muitos *outliers*, inclusive negativos, durante as paradas de processo, que conferem medição fora do domínio possível do parâmetro (inversão de sentido) e serão desconsiderados.

A Pressão do Exaustor apresenta tendência próxima a constância considerando sua unidade de medida. Apresenta uma medição densa, ruidosa e com grande quantidade de *outliers* conforme pode ser observado no Gráfico 13. Diante das observações deve ser aplicado um filtro para observar melhor o seu comportamento e verificar sua influência em relação as demais variáveis.

Gráfico 13 – Gráfico de linha da Pressão do Exaustor

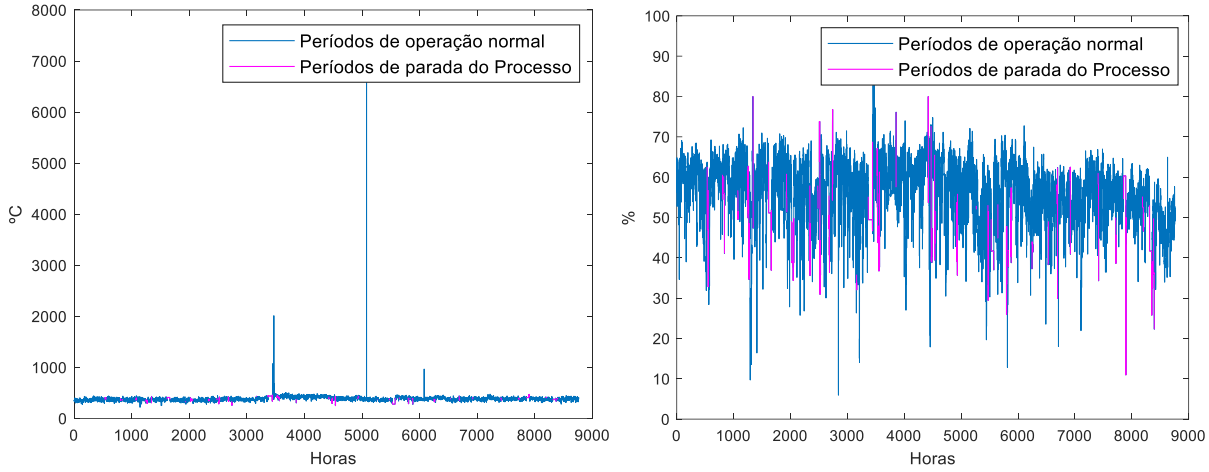


Fonte: Produção do próprio autor.

A “Temperatura do BTP” e “Posição do BTP” constituem variáveis calculadas tendo como base as temperaturas e posições das cinco caixas de vento. Ao analisar o Gráfico 14 (a), infere-se que existem variações abruptas na temperatura do BTP (*outliers*), fora do seu intervalo possível de 200ºC a 500ºC. Nesses casos, é possível inferir que os algoritmos para seu cálculo não convergiram a uma equação quadrática que representasse adequadamente o comportamento do BTP nesses instantes, ou que o BTP não foi atingido ao longo da esteira. Esse caso será discutido na seção 0.

O BTP é um parâmetro alvo, pois se deseja que seu valor seja de 60% das distâncias entre as últimas cinco caixas de vento, apesar disso, observa-se que essa posição apresenta variações significativas, conforme pode ser observada por meio do Gráfico 14 (b). Pode-se inferir que o controle desse parâmetro e o alcance do alvo desejado ainda não é precisamente atingido na planta em análise.

Gráfico 14 – Gráficos de linha do BTP (a) Temperatura (b) Posição



(a) (b)

Fonte: Produção do próprio autor.

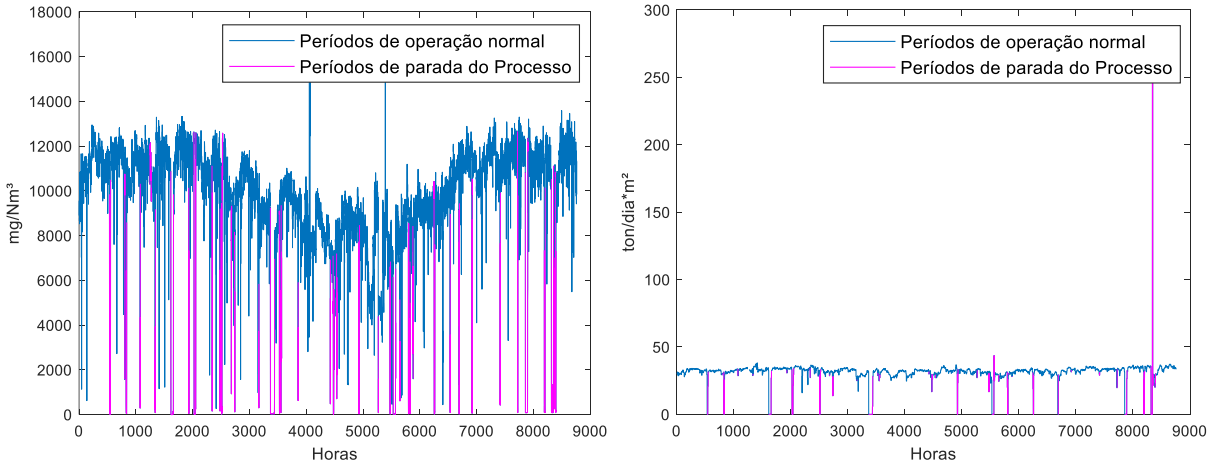
7.4 Indicadores de Saída do Processo de Sinterização

Os indicadores de saída do processo, como já citado, correspondem aos fatores resultantes do processo, isto é, sua produtividade, geração de produtos que devem ser retornados e a emissão de gases, como o CO.

Em relação a emissão do gás “CO” durante o período em análise, nota-se, a partir do Gráfico 15 (a) que até aproximadamente metade do período selecionado, a emissão desse gás seguia uma tendência de redução. Em seguida, a emissão passou a ter uma tendência de aumento. Valores baixos de emissão foram, em sua maioria, justificadas pelo processo, apesar disso, existe ainda a presença de *outliers*.

Por meio do Gráfico 15 (b), em relação a variável “Produtividade” observa-se que há variações abruptas durante as paradas do processo, que devem ser desconsideradas, e acompanha uma tendência como alvo próximo a 30 ton/dia.m².

Gráfico 15 – Gráficos de linha (a) CO (b) Produtividade

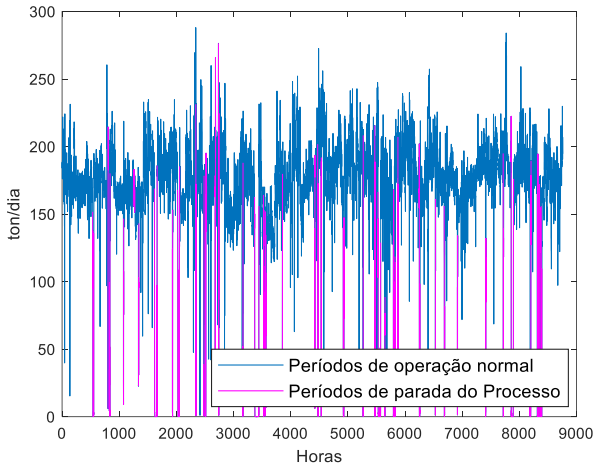


(a) (b)

Fonte: Produção do próprio autor.

As variáveis de “Geração de Finos de Retorno” representam a quantidade, em ton/dia, de material a ser reutilizado para retorno, que não originou sínter no processo, e o peso em toneladas do material de retorno de pequena granulação. Como é possível observar por meio do Gráfico 16, reduções bruscas no valor da variável esteve diretamente relacionada a parada do processo ou mudanças de comportamento dele.

Gráfico 16 – Gráfico de linha Geração de Finos de Retorno



Fonte: Produção do próprio autor.

8 ANÁLISE DA QUALIDADE DOS DADOS

Na presente seção serão analisadas de forma intrínseca e extrínseca, isto é, será realizada uma análise baseada apenas nos elementos do conjunto e também considerando o seu contexto. As características a serem analisadas são sete das quinze características presentes no padrão ISO/IEC 20512:2008, a citar, acurácia, completude, consistência, credibilidade, atualidade, acessibilidade e compressibilidade.

Deve-se ressaltar que não foi possível obter conclusões efetivas em relação as características de conformidade, confidencialidade, eficiência, precisão, rastreabilidade, disponibilidade, portabilidade e recuperabilidade, pois essas são dependentes de informações de segurança, disponibilidade e planejamento da própria empresa. Infere-se, entretanto, que por serem dados já utilizados e regularizados em uma planta convencional de sinterização, aderem as normas de qualidade de dados em sua coleta, apresentam precisão suficiente para seu funcionamento adequado, apresentam conformidade necessária e são confidencializados a pessoas que sejam responsáveis pelo processo.

8.1 Acurácia

A acurácia, pode ser analisada em relação a precisão estrutural dos dados e a precisão temporal dos dados. A precisão estrutural é responsável por caracterizar os dados em consideração a um tempo específico enquanto a precisão temporal analisa como a mudança temporal dos dados na realidade está refletida nos dados analisados (BATINI; SCANNAPIECO, 2016).

A precisão estrutural pode ser dividida em duas formas de análise: *Syntactic Accuracy* e *Semantic Accuracy*. A *Syntactic Acuracy* tem como objetivo analisar se os elementos do banco de dados estão de acordo com seu domínio possível e a *Semantic Accuracy* busca analisar o quão perto o valor medido está do valor real (BATINI; SCANNAPIECO, 2016).

Em relação a *Syntactic Accuracy*, pode-se inferir, a partir da observação dos dados e suas unidades de medida, os valores de medição como peso, temperatura, pressão, vazão e porcentagens estão dentro de um domínio possível durante a maior parte dos registros disponibilizados para o presente estudo. Entretanto, é possível verificar que, quando a variável

“Velocidade da Máquina de Sínter” é reduzida ou se torna zero, indicando uma parada no processo, há uma variação brusca dos parâmetros, especialmente relativos às variáveis do grupo de indicadores do controle do processo de sinterização. Foram computadas, por exemplo, medições de temperatura e de velocidade da esteira de sinterização que seriam inalcançáveis por meio dos equipamentos disponibilizados ao longo da esteira de sinterização. Nesses instantes de tempo é recomendado que os dados do processo sejam invalidados.

Segundo Batini e Sacannapieco (2016), deve-se analisar as dimensões de acurácia relacionadas ao tempo (do inglês *Time-Related Accuracy Dimensions*). As variáveis devem ser analisadas com base à rapidez na qual os dados são atualizados à medida que ocorrem alterações na realidade, a volatidade, ou seja, a frequência da mudança dos dados de forma e em relação a atualização dos dados.

O processo de sinterização é lento e de longa duração. Nesse sentido, a mudança temporal dos dados é lenta e isso reflete a necessidade de medições de menor periodicidade para refletir nos dados a dinâmica real dos parâmetros do processo e garantir uma precisão temporal, na ordem de minutos. Assim sendo, foram realizados diagnósticos que garantem a necessidade de alteração em relação a periodicidade fornecida.

A variação da composição da pilha homogeneizada e a mistura total de entrada dependem de um novo empilhamento e posterior homogeneização. A alteração das medições desse parâmetro ocorre somente após uma nova mistura, novo empilhamento ou alteração da pilha a ser utilizada como matéria-prima do processo. Atualmente, a medição desses parâmetros é realizada a cada 24 horas. Recomenda-se, entretanto, que, uma forma mais precisa e adequada à realidade do processo seria o alinhamento do registro dos dados referentes a esses parâmetros logo após ser realizada alguma alteração na pilha de entrada.

Conforme mencionado na seção 6, a esteira de sinterização apresenta 88 m de comprimento. Como já explicado, a matéria prima do sínter é depositada no início da esteira e ao final dela conclui-se o processo de fundição parcial desse material, que forma o sínter. Considerando uma velocidade média do maquinário de 2,5 m/min a 3 m/min, constata-se que o mesmo material se mantém disposto no processo de sinterização por cerca de a 29,3 min a 35,2 min. Considera-se, assim, necessária a utilização de uma amostragem de medição das variáveis de controle e

indicadores do processo de sinterização precisam de registros e amostragem menor que 5 min para que seja possível observar o comportamento desse material ao longo do processo e sua influência nas variações.

Os parâmetros de controle do processo de sinterização, nos dados disponibilizados, com exceção da altura da camada, apresentam periodização apresentada de 1 hora, enquanto a altura da camada, tem periodicidade de medição a cada 6 horas. É possível observar Gráfico 7 (a) e Gráfico 8 (a) e (b), que os parâmetros de Densidade da Mistura, Velocidade de Sinterização e Velocidade da máquina de sínter apresentam variações significativas ao longo do tempo. Entretanto, é possível constatar por meio do Gráfico 7 (b) que a Altura da camada não apresenta variações constantes e, dessa forma, é possível considerar uma periodicidade inferior para os seus registros.

Sabe-se, também, que a velocidade da esteira influencia o tempo de permanência do material sobre ela ao longo do processo e deve ser levado em conta para uma análise fundamentada acerca do BTP. Como esses são parâmetros de controle que afetam, conceitualmente, as temperaturas ao longo do processo de sinterização, dessa forma, como sugerido anteriormente, esses parâmetros devem apresentar maior periodicidade de medição ou registro para uma análise adequada.

Temperaturas, pressões e vazões são variáveis que se alteram de forma gradual. Ao observar os dados apresentados do grupo de indicadores do processo de sinterização, com periodicidade de uma hora, não é possível observar a variação gradativa desses parâmetros ao longo do tempo e, sim, variações bruscas desses parâmetros. Esse fato indica que a precisão temporal dos dados não está de acordo com o necessário. Levando em conta a velocidade da esteira de sinterização e o tempo de permanência do material nela, constata-se que para que exista uma precisão temporal, a periodicidade de medição deve ser ajustada.

Justifica-se, entretanto, a necessidade de menor periodicidade de registros das temperaturas das caixas de vento que não estão localizadas ao final da esteira pois as variações de temperatura nessas regiões são pequenas ao longo do tempo. Contudo, para um monitoramento eficaz ainda deve-se levar em conta a velocidade da esteira e o tempo de permanência do material sob ela.

Conclui-se, por fim, que a precisão temporal dos dados não se encontra de acordo com a realidade e dificulta a realização de análises adequadas para a estimação da variação e comportamento das variáveis ao longo do processo. Além disso, as variáveis apresentam valores dentro do seu domínio possível, com exceção dos períodos de parada da esteira de sinterização.

8.2 Completude

Ao se realizar uma análise de completude dos dados do processo observa-se que existem três instantes de tempo durante o ano que não apresentam registros de variáveis do processo, esses instantes correspondem as amostras 1995 (02h00 do dia 15 de novembro de 2017), 3664 (15h00 do dia 23 de janeiro de 2018) e 4238 (13h00 do dia 16 de fevereiro de 2018).

Foi possível identificar que a ausência de completude das variáveis esteve diretamente atrelada as paradas de processo, identificadas por meio da variável “Velocidade da Máquina de Sínter”. Nesses instantes, mais de uma variável apresentava valor não registrado ou igual a 0, que representavam ausência de registro, já que não estavam dentro do intervalo possível de valores para essas variáveis.

Detectou-se, também, um intervalo de medições, 1248 a 1272, referentes a medição de um dia completo do mês de outubro de 2018 e uma hora, onde não foram registradas seis variáveis do processo. São elas: Altura da camada, Temperaturas das Caixas de Vento nº 11, 13 e 18, Produtividade e Retorno. Nesse intervalo não ocorreu redução ou alteração significativa de velocidade da esteira de sinterização.

As variáveis que apresentavam ausência de informações fora dos intervalos já tratados foram a “Densidade da Mistura”, “Retorno”, “Pilha Homogeneizada A”, “Pilha Homogeneizada B” e “Total Pilha Homogeneizada”. Essas variáveis necessitarão de tratamento de completude para visualização apropriada do seu comportamento ao longo do tempo.

Ao final do intervalo, entre as medições 8665 a 8760, foi detectado um intervalo com ausência de registro de medição de todas as quatorze variáveis relativas à composição da mistura do processo. Acredita-se que, neste caso, a ausência de registro no banco de dados ocorreu pois

eram relativos aos últimos quatro dias antes da consulta e os dados não estavam completamente carregados no banco. Esse intervalo foi desconsiderado por conta da ausência de um número significativo de variáveis.

8.3 Consistência

Os dados a serem analisados não apresentam contradição aparente entre si, e apresentam coerência. Medidas que são muito destoantes, denominadas de *outliers*, serão estudadas e tratadas ou removidas para que não comprometam a análise dos dados do processo.

8.4 Credibilidade

Em relação a credibilidade dos dados a serem analisados, é possível constatar que, os valores atribuídos as variáveis podem ser consideradas verossímeis pelos usuários para análise e compreensão do processo da planta de sinterização. Os valores são medidos por meio de equipamentos robustos e confiáveis ou registrados por operadores. Atualmente, são utilizados de forma efetiva para tomada de decisões durante o processo.

8.5 Atualidade

Os dados disponibilizados para análise datam de vinte e quatro de agosto de 2017 a vinte e três de agosto de 2018. A variação temporal entre o fim de sua coleta e o início de sua análise, realizada no presente trabalho, é de, aproximadamente, um ano. O nível de atualização dos dados pode ser considerado de médio a alto grau, pois, o processo de sinterização não apresentou atualização ou variação considerável ao longo do intervalo de tempo disponibilizado para análise e não passou por novas atualizações até o presente momento do estudo.

8.6 Acessibilidade

O nível de acessibilidade dos dados aos operadores da planta e responsáveis por analisar e acompanhar os dados do processo é alto. Os dados estão disponíveis aos operadores da planta de sinterização de forma interativa por meio de seu sistema supervisório e aos responsáveis por

acompanhamento dos dados também em bancos de dados acessíveis por meio de software de leitura de dados do tipo .xlsx.

8.7 Compreensibilidade

Os dados têm atributos que, suficientemente, permitem que sejam lidos e interpretados pelos usuários que entendem os conceitos fundamentais do processo de sinterização. É possível identificar cada variável e observar seu comportamento de forma simples e clara.

9 TRATAMENTO DOS DADOS

Na presente seção, será realizado o tratamento dos dados que foram caracterizados e examinados nas seções 7 e 9. Serão, inicialmente, identificados e removidos do conjunto, os dados que são referentes a paradas do processo de sinterização. Os *outliers* do processo serão analisados, para remover, individualmente, medições incorretas das variáveis. Será realizada uma reamostragem e filtragem dos dados por meio da MODWT. Por fim, serão tratados os dados faltantes por meio de interpolação de dados e redes neurais regressivas.

9.1 Tratamento por Eliminação de Conjunto de Dados

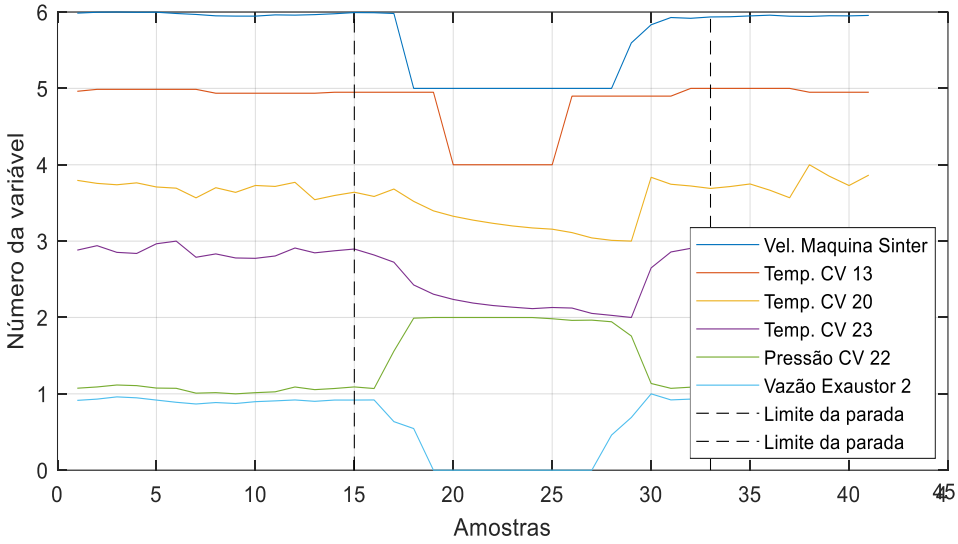
Por meio do exame preliminar dos dados e da análise de sua qualidade, foi possível identificar momentos nos quais os registros não coincidiram com o comportamento regular da máquina de sinterização baseado em seu funcionamento padrão ao analisar o conjunto de dados disponibilizados. Nesses momentos, a variável de velocidade da máquina de sínter se apresenta igual a zero ou próxima desse valor, isto é, ocorreu uma parada do processo de sinterização.

Nos instantes em que a máquina se apresenta parada, assim como nos momentos em que ela desacelera e retorna a acelerar, as variáveis do grupo de indicadores de controle do processo de sinterização e indicadores de saída, conforme descritas na seção 6, apresentam seus registros como valores iguais a zero ou destoantes de outras medições próximas. É valido ressaltar que em alguns momentos, durante a desaceleração e aceleração da esteira, a velocidade lida pode apresentar variações bruscas em relação a medições próximas, erros de medição.

O Gráfico 17 exemplifica um momento de parada do processo em meio uma operação normal. Nela são ilustradas uma parcela das variáveis que são diretamente influenciadas pela parada do processo e os limites dos dados de parada que deverão ser desconsiderados.

Por meio do Gráfico 17, observa-se que a variável de velocidade da máquina de sínter reduz, permanece em zero e retorna a aumentar novamente. Nesses instantes, nota-se um comportamento anômalo nas demais variáveis. As temperaturas e vazões são reduzidas consideravelmente e a pressão da caixa de vento apresenta aumento dentro do intervalo delimitado.

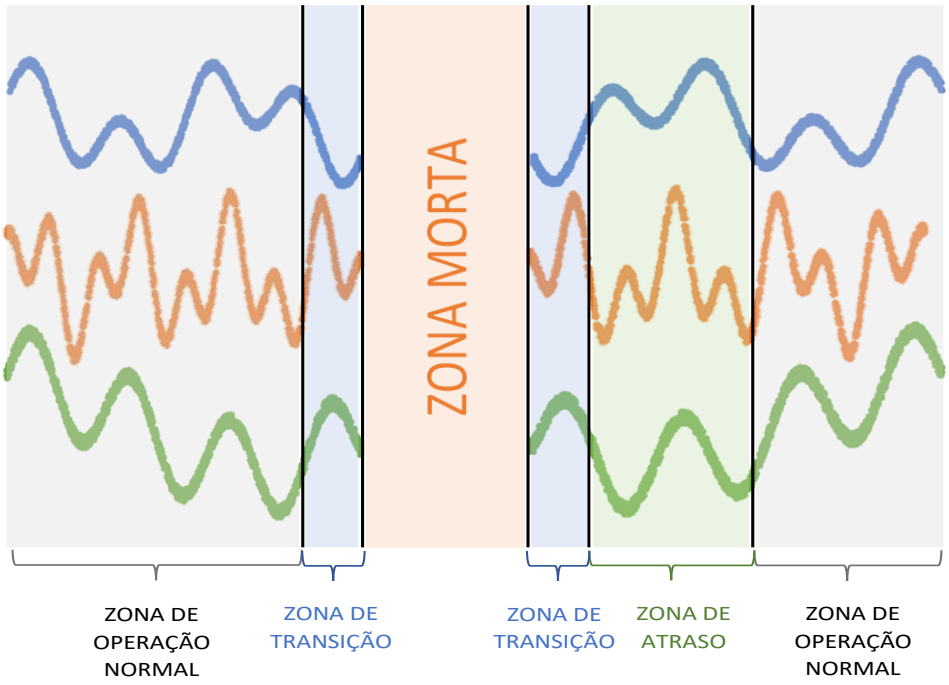
Gráfico 17 – Exemplo de momento de parada da esteira de sinterização



Fonte: Produção do próprio autor.

Para considerar as paradas do processo de sinterização, optou-se, no presente trabalho, definir os períodos de parada da esteira, além de sua desaceleração e retorno a aceleração. Para realizar essa identificação, os dados foram segmentados em zonas, conforme pode ser observado na Figura 20.

Figura 20 – Comportamento dos dados da máquina de sinterização



Fonte: Produção do próprio autor.

É possível identificar, por meio da Figura 20, quatro zonas distintas dos dados associadas ao comportamento da máquina de sinterização, são elas:

- Zona Morta: períodos nos quais existe parada completa da esteira de sinterização para realização de procedimentos operacionais de manutenção ou reparo. Nesses instantes a Velocidade da Máquina de Sínter equivale ou é muito próxima de zero. Os valores de temperatura e posição do BTP permanecem constantes, há variação abrupta ou ausência de medição de variáveis relativas a temperaturas e pressões;

- Zona de Transição: períodos nos quais está sendo realizado um procedimento de desaceleração da esteira para início da parada operacional ou aceleração da esteira após o fim da parada do processo. Nesses momentos, na maior parte das vezes a Velocidade da Máquina de Sínter varia entre zero e a velocidade operacional seguinte a parada. Em alguns casos, aparecem velocidades destoantes aos valores próximos, representando erros de leitura. Os valores de temperatura e posição do BTP ainda permanecem constantes ou destoantes em mais de 50% da média operacional, há um retorno de aumento ou ausência de medição de variáveis relativas a temperaturas e pressões;

- Zona de Atraso: período no qual há retorno de operação normal, porém, nesses momentos, por se tratar de uma série temporal, os dados não podem ser utilizados diretamente para análise. Para aplicação a uma rede de treinamento, deve-se retomar a análise dos dados após essa zona;

- Zona de Operação normal: período no qual há funcionamento pleno da operação da Máquina de Sinterização, sem anomalias associadas a parada da esteira e que garante dados para análise temporal defasada no tempo.

As paradas do processo não apresentam uma duração padrão, elas variaram de intervalos de seis a sessenta e seis horas de parada. Essa variação ocorre, pois, a duração é planejada de acordo com as demandas de reparo dos equipamentos ou solução de problemas associados ao processo de sinterização. O comportamento dos dados não reflete a realidade do sistema, portanto é necessário remover esses intervalos de tempo do conjunto de dados para que não exista inferências errôneas sobre o comportamento real da planta.

Os intervalos de parada podem ser observados por meio do Quadro 2. Nela são mostrados os dados que serão desconsiderados por se tratar de dados de parada do processo. São indicados a data e horário de seu acontecimento e o intervalo em relação ao total de amostras. Também é delimitado o intervalo de parada completa da esteira, instantes horários nos quais ela aparece com velocidade zero.

Quadro 2 – Intervalos de parada do processo de sinterização

(continua)

Dados desconsiderados da análise				Período de parada da esteira			
Início		Término		Início	Término	Início	Término
Data	Hora	Data	Hora	Amostras	Amostras	Amostras	Amostras
15/09/2017	06h00	16/09/2017	02h00	535	555	541	549
26/09/2017	09h00	26/09/2017	15h00	804	808	804	804
27/09/2017	09h00	27/09/2017	22h00	826	839	827	837
07/10/2017	17h00	07/10/2017	21h00	1074	1078	1074	1074
18/10/2017	15h00	18/10/2017	23h00	1336	1344	1339	1341
30/10/2017	06h00	01/11/2017	06h00	1615	1663	1621	1651
12/11/2017	09h00	12/11/2017	14h00	1930	1935	1931	1933
16/11/2017	04h00	17/11/2017	16h00	2021	2057	2025	2040
29/11/2017	11h00	29/11/2017	17h00	2340	2346	2342	2343
05/12/2017	06h00	05/12/2017	11h00	2479	2484	2480	2482
06/12/2017	09h00	07/12/2017	08h00	2506	2529	2507	2515
13/12/2017	12h00	13/12/2017	21h00	2677	2686	2680	2680
15/12/2017	23h00	16/12/2017	07h00	2736	2744	2740	2741
02/01/2018	17h00	02/01/2018	21h00	3162	3166	3163	3164
11/01/2018	06h00	14/01/2018	10h00	3367	3443	3370	3432
17/01/2018	18h00	18/01/2018	02h00	3523	3531	3526	3528
18/01/2018	20h00	19/01/2018	05h00	3549	3558	3550	3552
19/01/2018	08h00	19/01/2018	13h00	3561	3566	3564	3564
31/01/2018	14h00	31/01/2018	16h00	3855	3857	3855	3855
24/02/2018	04h00	24/02/2018	08h00	4421	4425	4422	4422
26/02/2018	06h00	26/02/2018	21h00	4471	4486	4475	4482
28/02/2018	20h00	01/03/2018	02h00	4533	4539	4535	4535
16/03/2018	22h00	17/03/2018	15h00	4919	4936	4920	4931
31/03/2018	09h00	31/03/2018	12h00	5266	5269	5267	5267
08/04/2018	22h00	09/04/2018	07h00	5471	5480	5473	5477
11/04/2018	06h00	13/04/2018	05h00	5527	5574	5530	5561
16/04/2018	05h00	16/04/2018	09h00	5646	5650	5647	5647
22/04/2018	11h00	23/04/2018	02h00	5796	5811	5797	5808
23/04/2018	16h00	23/04/2018	22h00	5825	5831	5826	5828
25/04/2018	19h00	26/04/2018	00h00	5876	5881	5877	5877
11/05/2018	05h00	11/05/2018	23h00	6246	6264	6250	6261
23/05/2018	01h00	23/05/2018	5h00	6530	6534	6531	6531
29/05/2018	12h00	29/05/2018	23h00	6685	6696	6687	6692
08/06/2018	03h00	08/06/2018	8h00	6916	6921	6917	6918
28/06/2018	15h00	29/06/2018	0h00	7408	7417	7409	7413
11/07/2018	12h00	11/07/2018	20h00	7717	7725	7719	7721
17/07/2018	06h00	19/07/2018	03h00	7855	7900	7862	7888

Quadro 2 – Intervalos de parada do processo de sinterização

(conclusão)

Dados desconsiderados da análise				Período de parada da esteira			
Início		Término		Início	Término	Início	Término
Data	Hora	Data	Hora	Amostras	Amostras	Amostras	Amostras
31/07/2018	03h00	31/07/2018	20h00	8188	8205	8189	8202
05/08/2018	08h00	07/08/2018	01h00	8313	8354	8314	8347
07/08/2018	11h00	07/08/2018	15h00	8364	8368	8366	8366
08/08/2018	06h00	08/08/2018	19h00	8383	8396	8385	8393

Fonte: Produção do próprio autor.

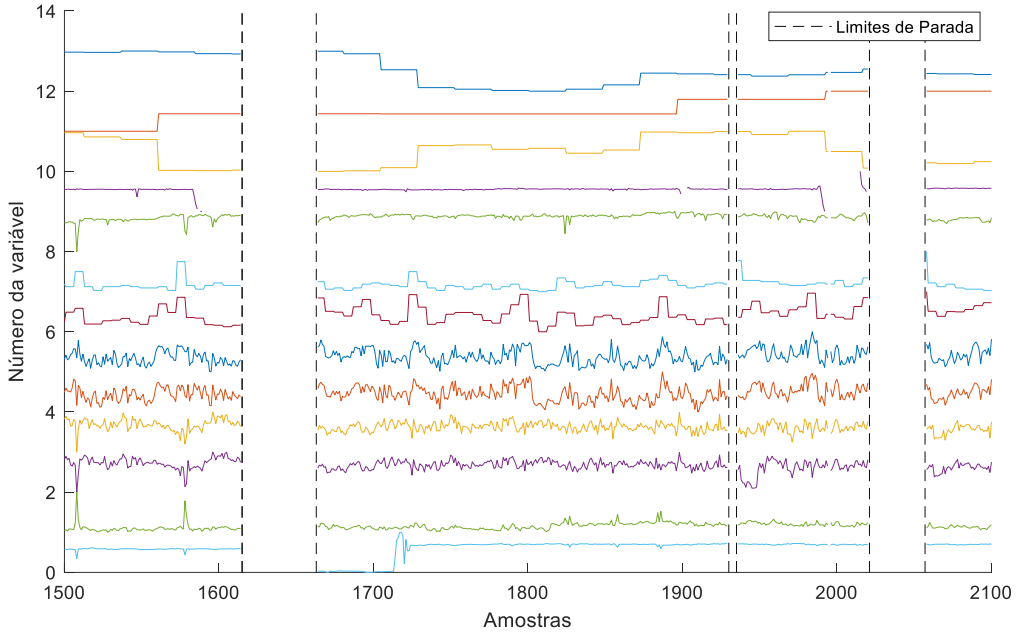
Para as variáveis medidas a cada 06h00 e 24h00 foi apenas informado o valor medido para um determinado conjunto de horas, porém não é possível identificar o instante exato da medição que representará o grupo. Esse registro pode ter sido efetuado dentro de uma zona de operação normal ou de parada quando ele engloba um conjunto de horas nos quais há amostras das duas zonas. Foram, então, para essas variáveis, desconsideradas todas as amostras que incluíam momentos de parada.

Como exemplo, é possível citar a parada que ocorreu entre a data 16/11/2017 às 04h00 e a data 17/11/2017 às 16h00. Para as variáveis amostradas a cada 06h00, primeiro dado, correspondente ao conjunto de 00h00 às 06h00 do dia 16/11/2017 foi desconsiderado, mesmo que apenas metade das horas compreendidas nessa amostragem se incluam em momentos de parada. O mesmo ocorre para variáveis medidas a cada 24h00, no qual os dados dos dois dias completos foram desconsiderados.

Também foi removido do conjunto de dados as medições 1248 a 1272, referentes a medição de um dia completo do mês de outubro de 2018. Nesses momentos ocorreram ausências de registros de seis variáveis durante todo o período. As amostras 8665 a 8760, foram desconsideradas do conjunto por conta da ausência do registro de quatorze variáveis. Essas ausências foram desassociadas a redução de velocidade da esteira de sinterização, conforme tratado na seção 8.2.

Ao final do tratamento realizado na presente seção foram removidos os dados que correspondiam a intervalos de parada e foi mantida a representação temporal deles de forma a não perder a informação cronológica. Os dados não foram agregados, sendo mantidos em grupos de dados válidos. Seu formato final pode ser observado por meio do Gráfico 18.

Gráfico 18 – Formato final dos dados após tratamento dos momentos de parada



Fonte: Produção do próprio autor.

9.2 Remoção de *Outliers*

Após a remoção de dados inválidos, foi realizado um estudo de verificação de *outliers*. Considerou-se, inicialmente, que, como constatado e relatado na seção 8.1, os dados não estão adequadamente amostrados no tempo de forma que seja possível observar a variação gradativa dos parâmetros.

Ao realizar uma análise, constata-se que grande parte das oscilações presentes a cada instante estão dentro de uma variação possível de valores, levando em conta a amostragem fornecida. Foi confirmado por responsáveis pelo processo, além disso, que variações relacionadas a sua dinâmica são passíveis e comuns de ocorrer ao longo do intervalo fornecido.

Tendo isso em vista, será aplicada uma detecção e remoção de *outliers* baseada em variações não fundamentadas conceitualmente considerando a sua unidade de medida, isto é, variações de ocorrência intangíveis se baseando nos equipamentos presentes na planta e na dinâmica do processo. Também serão removidos valores disruptivos em relação à média local e os que estão fora do intervalo factível das variáveis.

Variações significativas comuns às variáveis dos grupos de parâmetros e indicadores de controle, não relacionadas a parada da esteira, serão mantidas, pois se tratam de alterações da dinâmica operacional inerente a planta analisada. Esse comportamento foi confirmado por operadores e, por isso, eles compõem um conjunto de valores válidos.

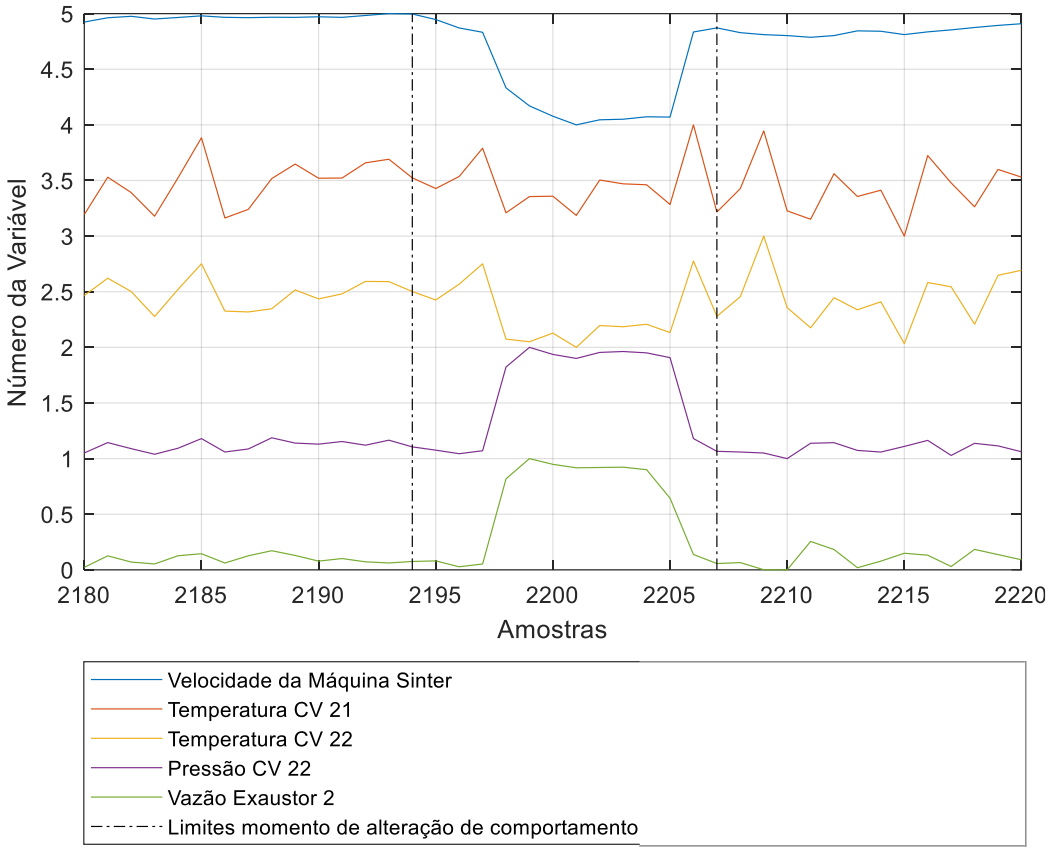
Esses eventos, ainda que reais e factíveis na planta analisada, ocorrem com menor frequência, considerando a quantidade total dos eventos que acontecem naquela faixa. Para aplicações nos quais se deseja observar o comportamento quando o sistema opera por mais tempo e apresenta mais eventos, esses dados podem passar a ser considerados como pontos fora da curva, *outliers*, e podem não contribuir para um treinamento baseado em máquina de inteligência.

Optou-se por manter esses pontos para sinalizar a quem vai usar esses dados em sequência que os valores mantidos são reais e que acontecem no processo dentro do esperado. A partir desse conhecimento, o pesquisador pode optar pelo tratamento desses eventos separadamente por meio de modelos probabilísticos. Um exemplo desses momentos pode ser observado por meio do Gráfico 19. A Velocidade da máquina de sínter apresenta diminuição em sua velocidade, influenciando no comportamento das demais variáveis do processo, não relacionada a paradas, e retorna ao seu modo de operação mais comum logo em seguida.

A técnica utilizada para identificação, reconhecimento e remoção de *outliers* se baseou no desvio absoluto em relação à média. Foram mantidos apenas os dados selecionados dentro do intervalo da média e variação de dois desvios padrão.

Levou-se em consideração a existência de uma variação frequente da média e mediana dos parâmetros ao longo do tempo, considerando janelas de dados. Diante disso, o cálculo e definição de *outliers* não foi baseado no conjunto total dos dados disponíveis, pois considera-se, nesse caso, mudança de *setpoints* de operação como *outliers*. Optou-se, então, por aplicar o cálculo dos desvios em relação à média por meio de janelas deslizantes. Foi definida, para cada variável, levando em conta sua variação ao longo do tempo, o intervalo das janelas variando de 500 a 1000 amostras.

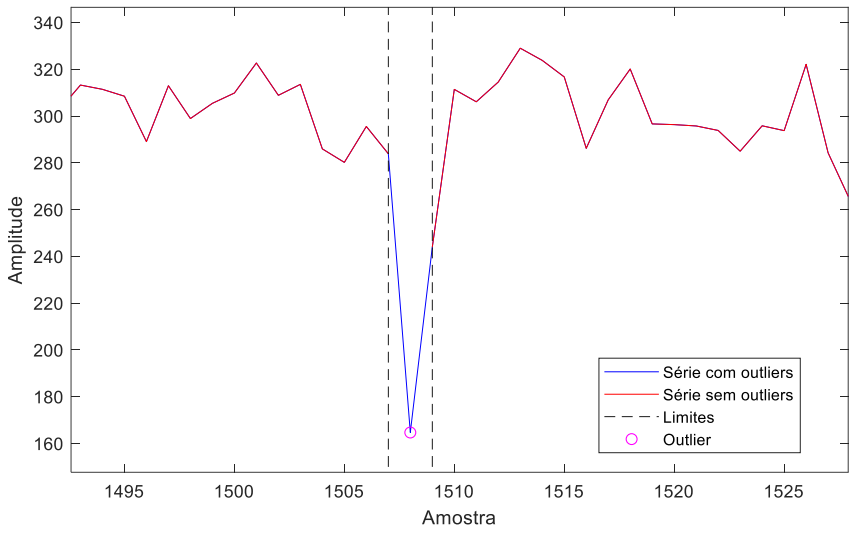
Gráfico 19 – Período de variação factível na planta analisada



Fonte: Produção do próprio autor.

O Gráfico 20 mostra em exemplo de identificação de *outliers* ao longo dos dados amostrados.

Gráfico 20 – Período com ocorrência de *outliers*



Fonte: Produção do próprio autor.

Nas variáveis apresentavam periodicidade de 24 horas, todas pertencentes ao grupo de parâmetros relativos a composição de entrada do processo, a maior parte variações identificadas na seção 7.1, estavam diretamente relacionadas a paradas do processo. Como nesses casos, os dados foram desconsiderados da análise e não se tem informações suficientes para garantia de que as variações diárias apresentados são discrepantes da realidade, não foi possível efetuar remoção de *outliers* desses dados da forma convencionada.

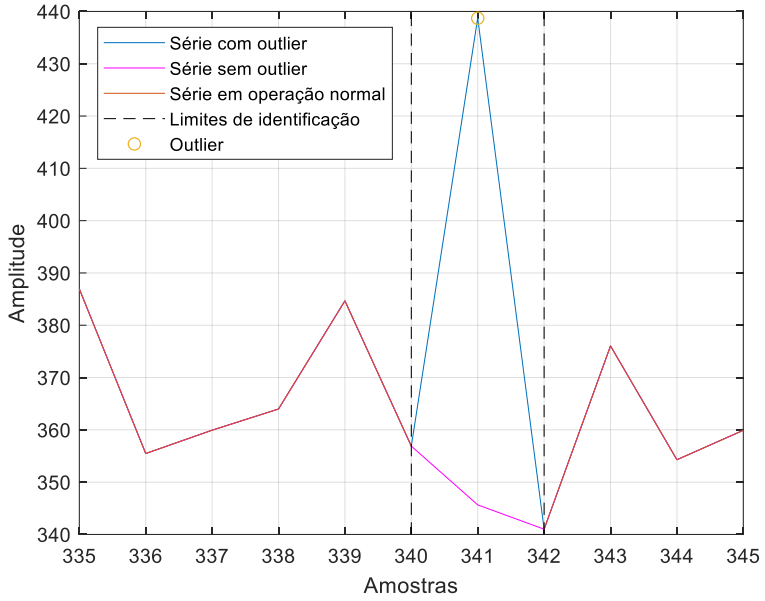
As variáveis “Fração +1 mm na mistura Total”, “Fração -0,105 mm na Mistura Total” e “Relação +1/-0,105 mm na Mistura Total”, diferentemente, apresentaram uma medição destoante das demais desassociada de variações em outras variáveis. Essas medições foram tratadas como *outliers* do processo.

Um estudo mais apurado e preciso acerca de *outliers* associados a erros de medição seria realizado com o fornecimento dos dados de manutenção da esteira e equipamentos de sinterização, e relatórios de operação do processo, pois, dessa forma, se torna possível discernir variações relacionadas às dinâmicas operacionais, bem como necessidades de mudanças de *setpoints* para determinadas demandas e possíveis leituras de sensores inválidas e quais mudanças foram aplicadas. Não se tem acesso a esses dados para comparação e garantia de compatibilidade operacional.

9.2.1 Tratamento por interpolação de amostras do entorno de *outliers* isolados

A completude de dados referentes a falhas isoladas, de uma amostra, foi realizada por meio de interpolação por amostras do entorno utilizando o polinômio interpolador por partes de Hermite (PCHIP, do inglês *Piecewise Cubic Hermite Interpolating Polynomial*). A maior parte dos dados incompletos serão tratados por meio dessa metodologia. É possível observar, por meio do Gráfico 21, um exemplo de interpolação realizada através desse método.

Gráfico 21 – Exemplo de amostra interpolada (PCHIP)



Fonte: Produção do próprio autor.

9.3 Cálculo da Posição e Temperatura do BTP

Segundo responsáveis pela planta de sinterização tratada nesse trabalho, os parâmetros de posição e temperatura do BTP são obtidos por meio de uma aproximação quadrática baseada na medição de temperatura das últimas cinco caixas de vento presentes no processo, conforme elucidado na seção 2.1.3. Não foi fornecida pela empresa, por ser uma informação sigilosa, o método de obtenção da aproximação quadrática utilizada para extrair informações acerca do BTP.

Li e Wang (2006), sugerem em seu trabalho uma metodologia de cálculo da curva quadrática de temperatura aproximada, baseada em temperaturas das caixas de vento presentes BTP na planta de sinterização. A equação de segundo grau é estimada de forma simples utilizando três amostras de medição de temperatura no entorno da temperatura máxima. Eles mostram que ela é suficiente e confiável para utilização para predição do BTP em redes neurais artificiais e controles Fuzzy. Ainda assim, a forma de aplicação utilizada pelos pesquisadores não era a que mais coincidia com os valores obtidos pela planta.

Foi possível obter aproximações e variações semelhantes aos dados originais fornecidos ao se utilizar da técnica simples de aproximação da curva fornecida pelo MATLAB®. A técnica

associa os pontos de temperatura fornecidos a forma de Vandermonde, na qual os termos de cada linha estão em progressão geométrica, por meio de uma matriz de pesos conforme pode ser observada por meio das equações (12) e (13).

$$\begin{pmatrix} v_1^n & v_1^{n-1} & \dots & 1 \\ v_2^n & v_2^{n-1} & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ v_m^n & v_m^{n-1} & \dots & 1 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_{n+1} \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} \tag{12}$$

$$V.P = Y \tag{13}$$

Onde:

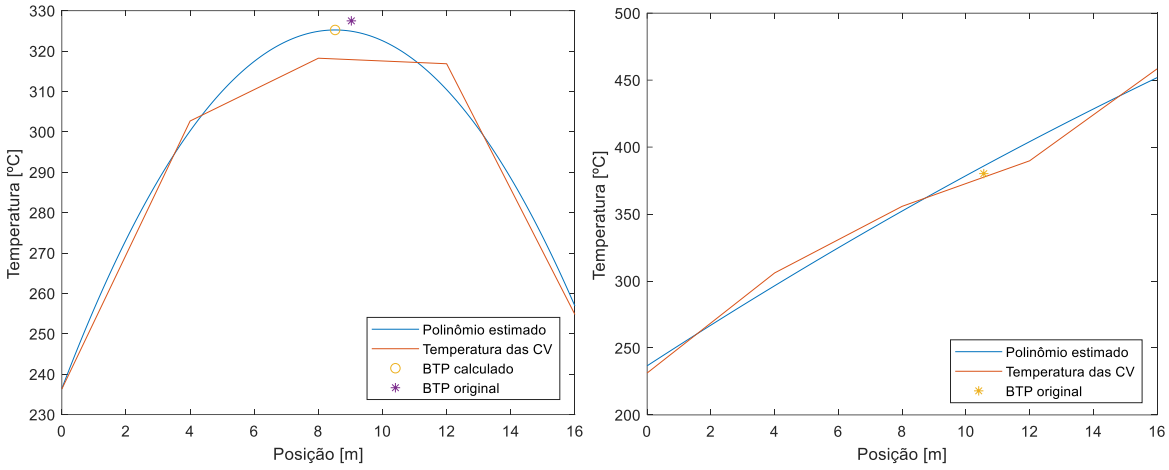
- V é a matriz de Vandermonde;
- P é a matriz de pesos ($p_1, p_2, p_3 \dots p_n$);
- Z é a matriz que contém os pontos originais fornecidos ($z_1, z_2, z_3 \dots z_n$).

Para resolver o problema, busca-se obter P por meio da igualdade apresentada. Como a solução requerida é uma equação quadrática, essa forma de solução é satisfatória. Para solução, a cada medição, utilizou-se as 5 medições de temperatura das caixas de vento para estimar a curva. Extraiu-se o ponto máximo e a posição a ele associada e obteve-se a temperatura e posição do BTP, respectivamente.

Foram comparados os dados obtidos em relação aos dados fornecidos pela empresa de ambas as variáveis. Nota-se que o resultado apresenta uma boa aproximação do valor calculado pela empresa detentora da planta. É possível observar um exemplo em que a temperatura e a posição do BTP se mostraram próximos por meio do Gráfico 22 (a). Nos gráficos seguintes as caixas de vento nº 19, 20, 21, 22 e 23 estão localizadas, respectivamente, nas posições 0, 4, 8, 12 e 16.

O método de cálculo utilizado no presente trabalho diverge dos dados reais em momentos nos quais a temperatura continua crescente à medida que se avança para as últimas caixas de vento presentes na planta. Observa-se, por meio do Gráfico 22 (b), que registro da posição e temperatura do BTP, por meio dos dados originais fornecidos, foram coletados fora do ponto máximo de temperatura e em posição adequada ao longo do processo. Não se tem conhecimento como o cálculo foi realizado nesse caso particular.

Gráfico 22 – Temperatura e Posição do BTP da amostra (a) 20 (b) 4520



(a) (b)

Fonte: Produção do próprio autor.

O comportamento crescente de temperatura da Gráfico 22 (b) denota que o BTP não foi atingido em local adequado ou não foi possível chegar ao seu resfriamento até o final da esteira de sinterização. Esse registro de variação crescente de temperatura se deu nos intervalos que podem ser observados por meio do Quadro 3.

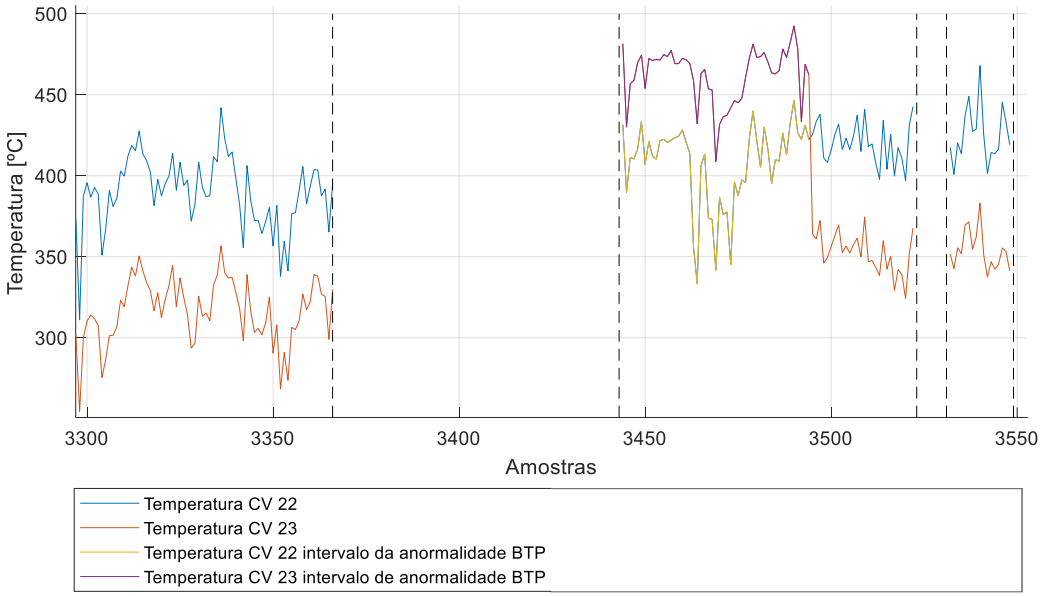
Quadro 3 – Intervalos em que há registro de variação crescente de temperatura

Data e Hora		Amostras	
Início	Término	Início	Término
14/01/2018 11h00	16/01/2018 13h00	3444	3494
28/02/2018 04h00	05/03/2018 11h00	4517	4644

Fonte: Produção do próprio autor.

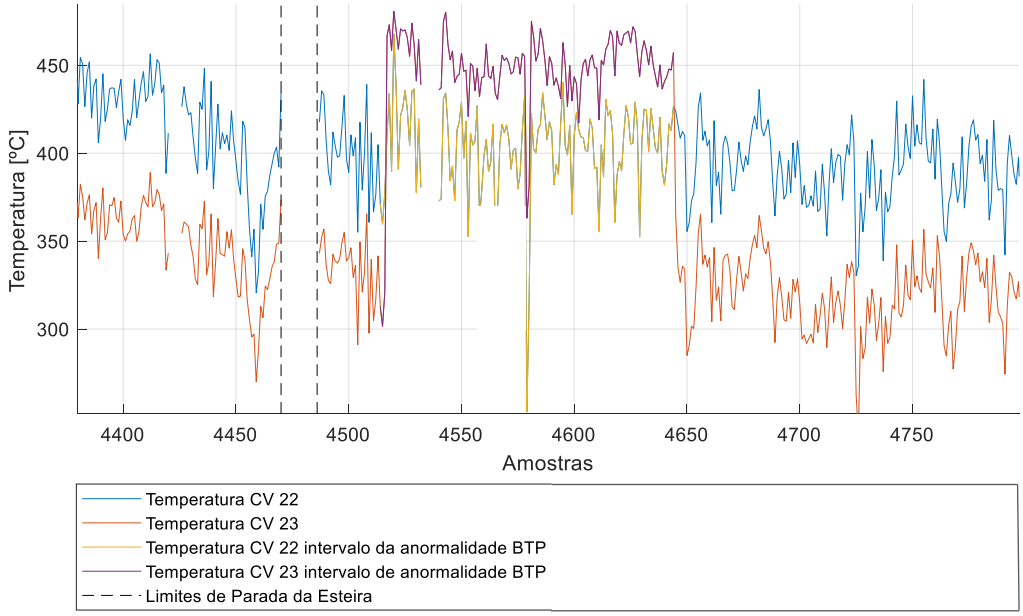
É possível observar as medições de cada uma das amostras por meio do Gráfico 23 e do Gráfico 24. Constata-se que apenas o primeiro intervalo se iniciou após uma parada de processo (Gráfico 23) e o segundo intervalo, 31 horas após uma parada (Gráfico 24). Não é possível identificar então uma relação direta entre os dois momentos e paradas no processo. Constata-se, por fim, que a mudança de temperatura e posição do BTP não ocorreram por conta de falhas no último sensor da esteira, o da caixa de vento número 23, pois as alterações de temperatura são refletidas na medição dos últimos dois sensores presentes ao longo da esteira.

Gráfico 23 – Intervalo de variação crescente de temperatura (amostras 3444 a 3494)



Fonte: Produção do próprio autor.

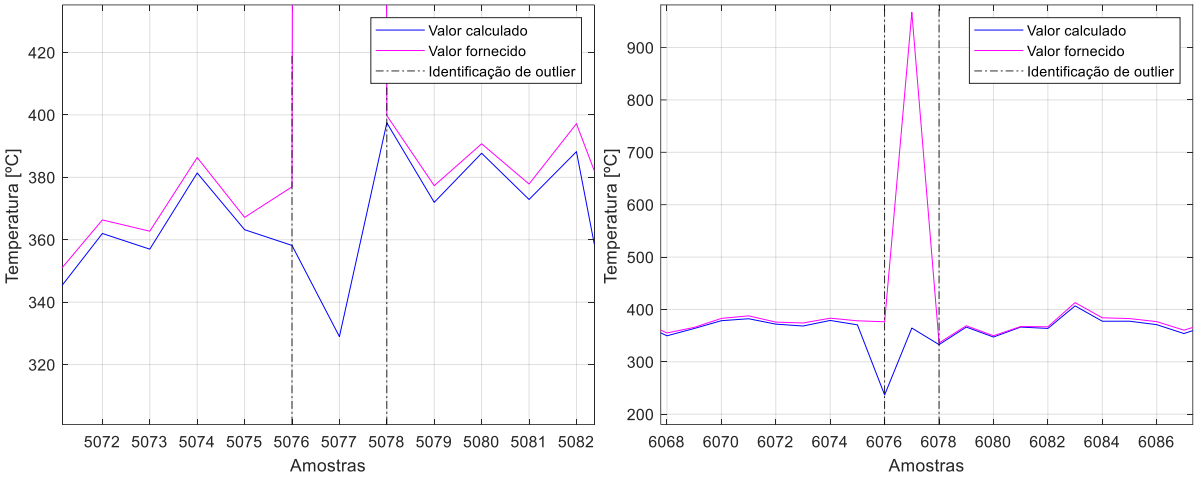
Gráfico 24 – Intervalo de variação crescente de temperatura (amostras 4517 a 4644)



Fonte: Produção do próprio autor.

Além desses casos, em dois momentos 5077 e 6077, foram apresentadas variações bruscas no cálculo de temperatura do BTP, que foram considerados *outliers*, pois a estimativa proposta, alcançou valores inferiores e adequados.

Gráfico 25 – Variações no valor fornecido da temperatura do BTP na amostra (a) 5077 (b) 6077.



(a) (b)

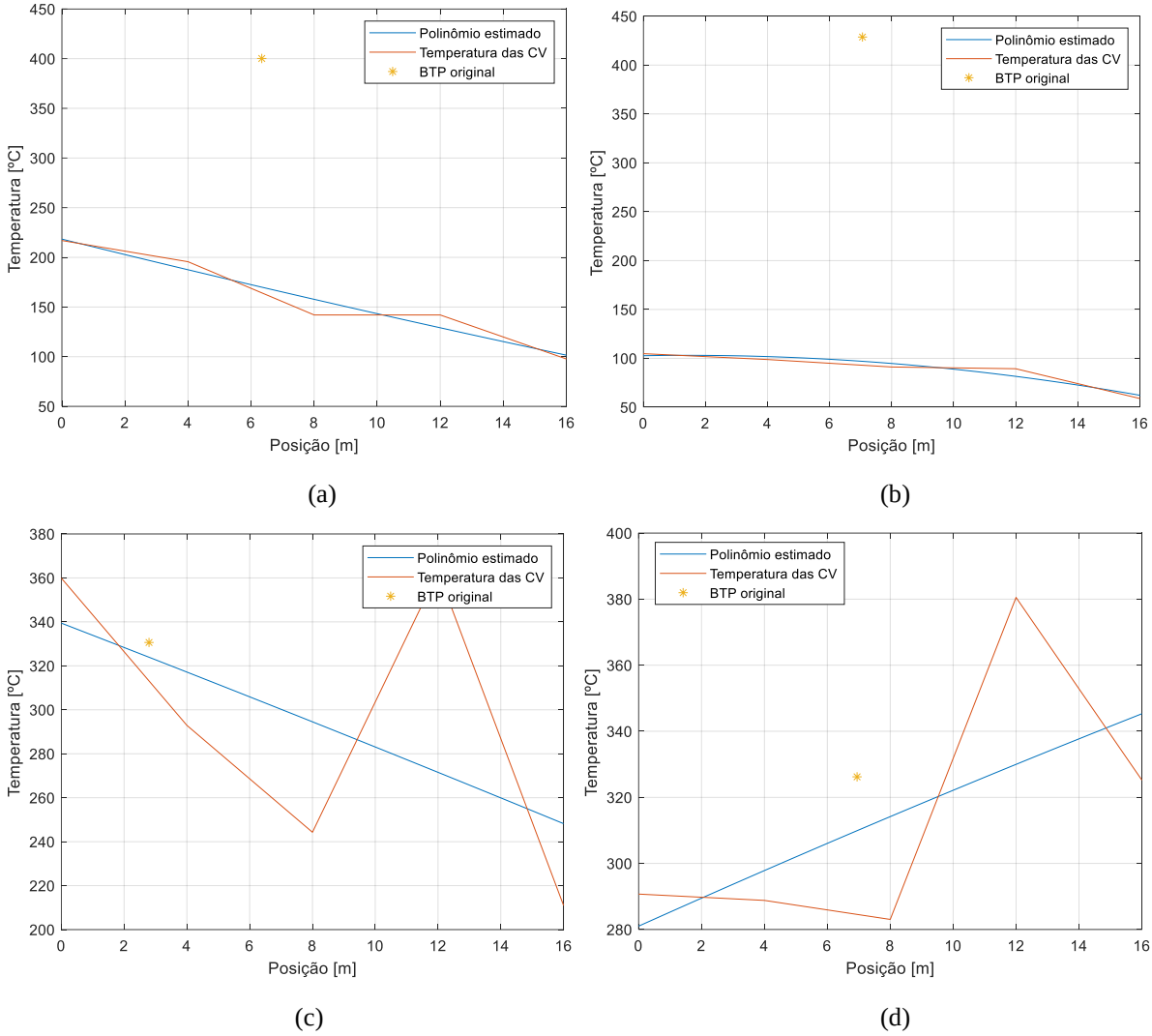
Fonte: Produção do próprio autor.

Durante os momentos em que há variações da velocidade da esteira, que correspondiam a valores reais e factíveis, a curva de temperatura formada apresentava-se decrescente, conforme é possível observar no Gráfico 26 (a) e (b). Infere-se que, nesses casos, o valor registrado do BTP fornecido pela empresa não corresponde a uma aproximação quadrática da curva. E uma aproximação quadrática ajustada obteve uma posição e temperatura do BTP em uma posição anterior ao local analizado. Ou seja, o BTP foi obtido antes do previsto na esteira por conta de uma velocidade mais baixa da esteira.

Em casos diferentes, como na medição 1291 (Gráfico 24 (c)), também não foi possível prever o BTP de forma adequada por meio do algoritmo proposto. Observou-se que entre aos instantes 1291 e 1335, desassociado de quaisquer outras alterações na planta, a curva de temperatura apresentou mudança de comportamento.

Na maior parte das vezes, nessas medições, as maiores temperaturas ocorriam nas CV 19 e 20 (conforme Gráfico 24 (a) e (b)), ou um polinômio de segundo grau não era suficiente para realização de uma interpolação de segundo grau das amostras (conforme Gráfico 24 (c) e (d)). O comportamento descrito não foi considerado falha de sensoriamento e comportamento possível já que a temperatura medida por todos os sensores apresenta a mesma tendência.

Gráfico 26 – Variações dos valores do BTP nas amostras (a) 5666 (b) 2414 (c) 1291 (d) 3349



Fonte: Produção do próprio autor.

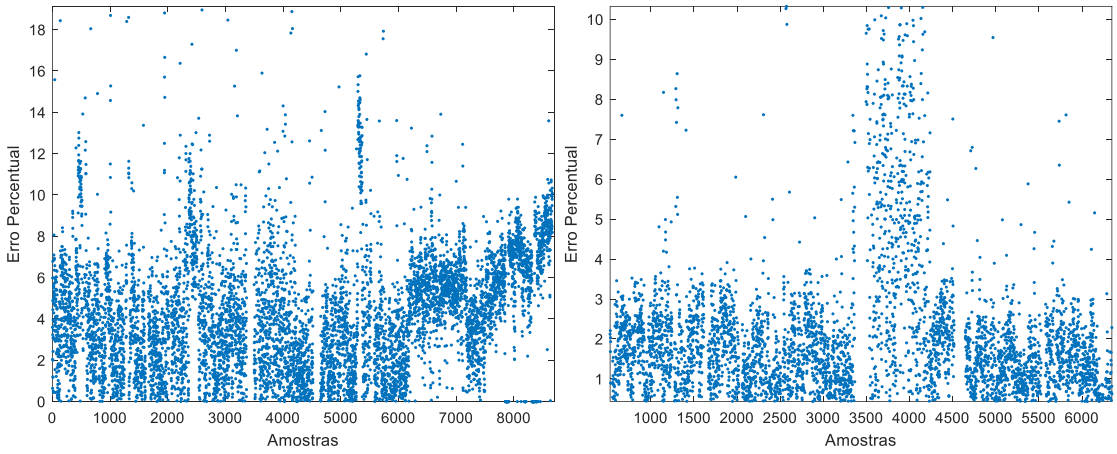
O instante 3349 (Gráfico 26 (d)), diferentemente, apareceu com diferença de medição isolada, desacompanhada de quaisquer outras alterações. É possível inferir que nessa situação ocorreu uma falha de medição em algum dos sensores, como por exemplo no sensor da caixa de vento 20, posicionada em 8 m no gráfico.

Desconsiderando os registros discriminados anteriormente para a estimativa do BTP foi calculado o erro percentual para cada uma das amostras, tanto da temperatura do BTP como de sua posição, conforme pode ser observado por meio do Gráfico 27 (a) e (b).

Nota-se que o erro associado a temperatura variou em média de 0 a 8% e a temperatura de 0 a 4%. Em um grupo de valores, o erro associado a posição do BTP aumentou para até cerca de

10%. Nota-se, ao comparar os gráficos do BTP calculado e do BTP fornecido pela empresa, que os valores estipulados pela empresa apresentam os dados deslocados linearmente em relação ao calculado. Essa alteração pode ser um caso particular dos cálculos por eles realizados.

Gráfico 27 – Erro percentual associado a (a) Temperatura BTP (b) Posição BTP

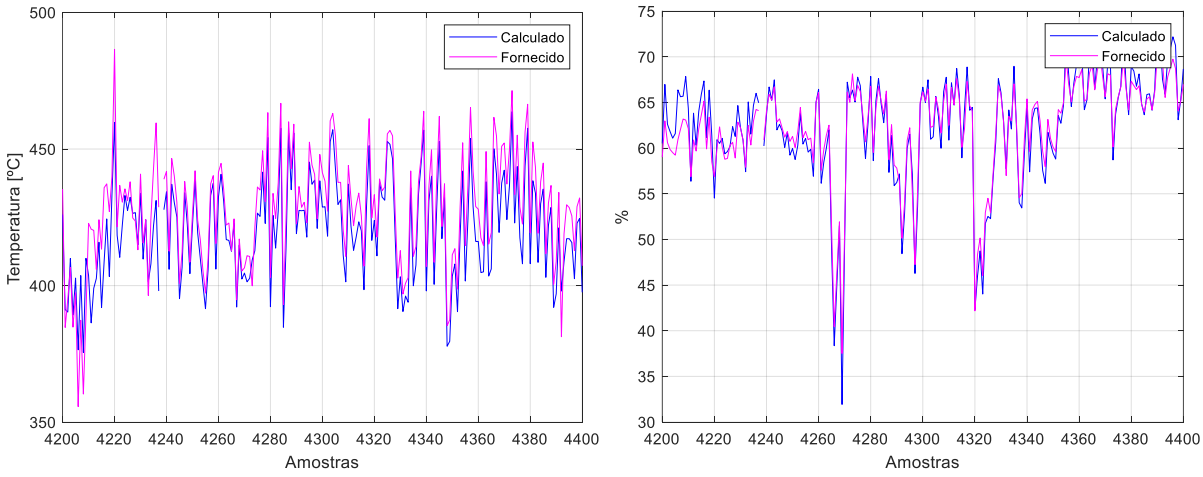


(a) (b)

Fonte: Produção do próprio autor.

Por meio do Gráfico 28 (a) e (b) é possível observar que, as curvas apresentam formatos muito próximos. Devido a metodologia de cálculo não ser coincidente, em alguns momentos existe uma defasagem em sua amplitude acompanhando sua variação.

Gráfico 28 – Comportamento da curva prevista e curva fornecida BTP (a) Temperatura (b) Posição



(a) (b)

Fonte: Produção do próprio autor.

É válido destacar que os dados fornecidos para a análise consistem em médias de um conjunto medições realizadas à cada segundo. Assim sendo, não é possível identificar, de forma precisa, se a disparidade entre os cálculos considerados e os dados originais se dá por conta do fato de que os fornecidos sejam formados por uma média de medições ou se ela aparece por conta do uso de metodologia de cálculo diferente para obtenção da curva quadrática.

Optou-se, no presente trabalho, para fins acadêmicos, utilizar os dados calculados por meio da nova métrica apresentada já que as variações propostas e a medição de temperatura do BTP se mostraram próximas as registradas e garantem um comportamento semelhante, mesmo que defasado. Para análises pormenorizadas como, por exemplo, previsão de variáveis, é considerado fundamental o fornecimento da metodologia utilizada na planta para aumentar a confiabilidade e credibilidade dos dados.

Essa metodologia de cálculo será utilizada para realizar novos cálculos dos valores de posição e temperatura do BTP após a realização de reamostragem, filtragem e completude dos dados medidos por meio de redes neurais para observação de seu comportamento e análise de correlação com as demais variáveis do processo. O resultado do cálculo será mostrado por meio da seção 11.3.

9.4 Reamostragem dos Dados para Suavização de Variações

Conforme abordado na seção 8.1, identificou-se a necessidade de uma maior recorrência e registro de medições em relação a cada variável do processo, já que o material que entra na esteira é encaminhado para o equipamento de resfriamento após cerca de 30 min. Verificou-se a necessidade de efetuar a reamostragem de alguns conjuntos de dados, já que as medições disponibilizadas foram fornecidas em diferentes variáveis em intervalos: 24 horas, 6 horas ou a cada uma hora.

As variáveis que foram medidas a cada 24 horas ou 6 horas eram, pela própria empresa detentora da planta, replicadas como medições para as demais horas do intervalo tratado. As variáveis amostradas a cada 24 horas não possuem variação de valores recorrente que justifique a reamostragem por intervalos menores. Nas variáveis originalmente amostradas a cada 6 horas,

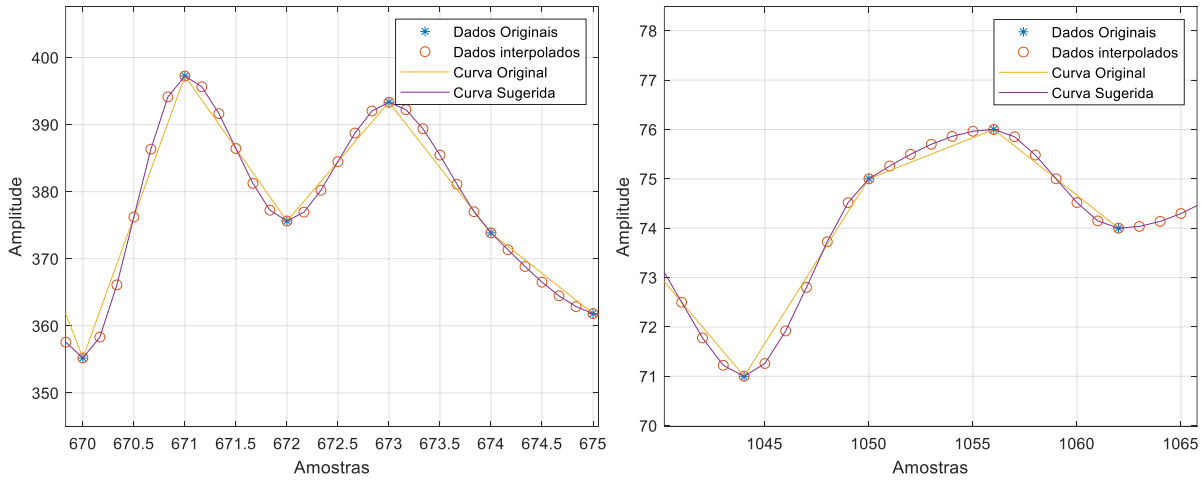
por apresentarem variações significativas ao longo do tempo, optou-se por realizar uma reamostragem a cada 1 hora.

As variáveis originalmente amostradas a cada 1 hora, devido refletirem a dinâmica temporal do processo, foram reamostradas a cada 10 minutos, perfazendo uma melhor apresentação dos dados para estudos associados. Foi possível, então, obter um sinal mais suavizado, com variações menores ao longo dos intervalos, retratando os sinais de temperatura e pressão, por exemplo, de forma mais próxima ao comportamento real.

A reamostragem foi realizada por meio da inserção de 5 amostras entre cada medição realizada, elas foram obtidas por meio de interpolação preservando a forma do polinômio PCHIP.

É possível observar, por meio do Gráfico 29 (a) e (b), a forma de interpolação aplicadas aos dados. Foram adicionados cinco novos valores a cada intervalo de 1 hora para a reamostragem a cada 10 min, e, da mesma forma, cinco novos valores a cada intervalo de 6 horas para reamostragem dos sinais a cada 1 hora.

Gráfico 29 – Exemplo de amostras interpoladas a cada (a) 10 min (b) 1 hora

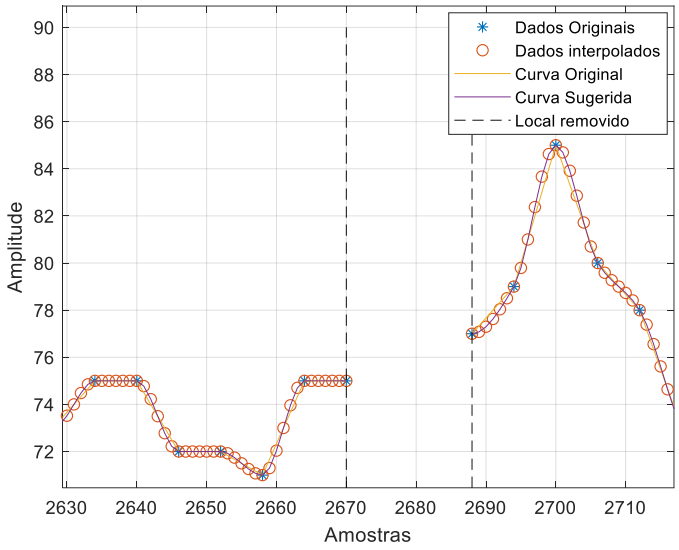


(a) (b)

Fonte: Produção do próprio autor.

É importante destacar que essa técnica foi aplicada apenas aos dados completos originais. Os dados que correspondiam a falta de completude dos dados, *outliers*, ou paradas de esteira, que foram removidos, conforme explicado nas seções 9.1 e 9.2, continuaram sem completude de seus registros, como pode ser observado por meio do Gráfico 30.

Gráfico 30 – Exemplo Formato de dados ao fim dos procedimentos



Fonte: Produção do próprio autor.

As variáveis de temperatura e posição do BTP, por serem variáveis calculadas, não foram reamostradas conforme descrito na presente seção. Nesse caso, elas serão recalculadas ao final do tratamento dos dados medidos por meio da metodologia apresentada na seção 0.

9.5 Tratamento de Ruídos com Transformada *Wavelet*

Foi possível observar, após os tratamentos até então realizados, que ainda permaneceram ruídos indesejáveis em certas variáveis que são introduzidos durante a captação dos dados. Esse fato dificulta a análise desses sinais e a aplicação dos dados em aprendizados de máquina para previsões, por exemplo. Para realizar a remoção desses ruídos é efetuado um tratamento por meio de MODWT, de forma a remover as componentes de altas frequências e selecionar aquelas que carregam a maior parte de informações de cada sinal.

As *wavelets* de base escolhidas para cada um dos sinais foram determinadas de forma que apresentassem maior similaridade com o comportamento de cada uma das variáveis por meio de análise visual. Foram verificadas, nessa análise, as *wavelets* de Daubechies e Symlets de diferentes ordens. Para todas as variáveis foi suficiente decompor o sinal em 8 subséries distintas, definidos de forma empírica para propiciar um bom resultado na atenuação do ruído. Esse procedimento foi aplicado às variáveis que, originalmente, tinham periodicidade de medição a cada 1 hora e às que, originalmente, tinham periodicidade a cada 6 horas e ainda

apresentam variações significativas em um pequeno espaço de tempo, que caracterizam ruídos após a reamostragem aplicada. A quantidade de frequências removidas foi baseada na obtenção de uma suavização da curva sem que ela perca suas características de variação. É apresentada, no Quadro 4, a *wavelet* de base aplicada a cada uma das variáveis tratadas, a quantidade de subséries em que foi decomposto o sinal e níveis de frequência removidos.

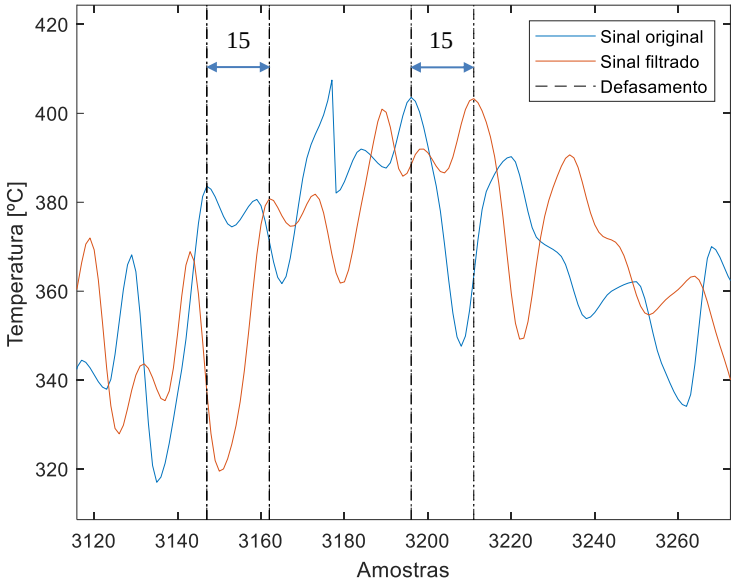
Quadro 4 – Aplicação da MODWT para remoção de ruídos associados

Nº	**Amostragem**	***Wavelet* mãe definida**	**Altas frequências removidas**	**Amostras Deslocadas**
5	10 min	sym4	4/8	30
15	10 min	sym4	4/8	29
17	10 min	sym4	4/8	29
18	10 min	sym4	4/8	29
19	1 hora	sym4	4/8	30
20	1 hora	sym4	4/8	30
21	1 hora	sym4	4/8	30
22	10 min	sym5	4/8	35
23	10 min	sym5	4/8	35
24	10 min	sym5	4/8	34
25	10 min	sym5	3/8	15
26	10 min	sym5	3/8	15
27	10 min	sym4	4/8	32
29	10 min	sym4	4/8	32
30	10 min	sym4	4/8	32
31	10 min	sym4	4/8	31
34	10 min	sym5	3/8	15
35	1 hora	sym4	3/8	14
36	10 min	sym4	4/8	31

Fonte: Produção do próprio autor.

Após ser efetuada a seleção do conjunto de frequências que concentravam a maior parte da energia do sinal e compará-la com o sinal sem o devido filtro associado, detectou-se um atraso associado em relação ao sinal original, conforme pode ser observado no Gráfico 31. Esse atraso foi corrigido para cada um dos sinais, separadamente, pois ele variava em relação a quantidade de frequências removidas e ao sinal analisado. O efeito de atraso ocorre devido à sobreposição das bandas (*overlaped*) dos filtros da transformada em cada nível da decomposição, entretanto a verificação e análise desse efeito não será escopo do presente trabalho.

Gráfico 31 – Exemplo de deslocamento das amostras utilizando MODWT

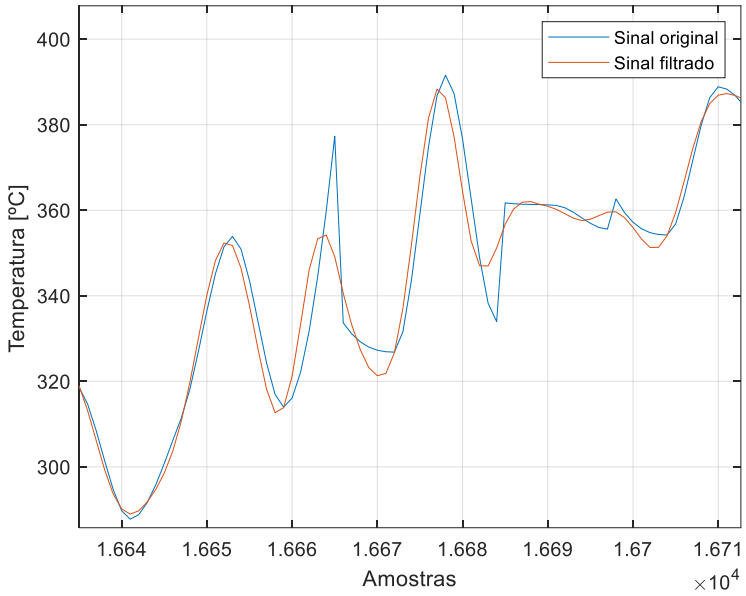


Fonte: Produção do próprio autor.

É possível observar, por meio do Gráfico 32, exemplos de sinais de temperatura das caixas de vento antes e após a aplicação do filtro *wavelet* associado. É possível observar que, em alguns momentos, sejam removidas frequências que fazem com que o módulo do sinal se distoe do sinal original conforme as amostras próximas a 16670 e 16680 (Gráfico 30). Essas diferenças, na maioria dos casos, representam menos que 4% do valor da temperatura do instante de medição associado. Considerando que o cálculo do BTP já é baseado em uma aproximação de segundo grau, essa variação nas temperaturas não implicará em grande impacto na aferição do BTP real do processo.

O tratamento de ruídos por meio de Transformada *Wavelet* foi significativo para o desempenho das redes neurais recorrentes utilizadas para a completude dos dados, conforme será tratado na seção 9.6. Sem o tratamento, o desempenho era inferior ao necessário e o erro associado se mostrava alto.

Gráfico 32 – Exemplo tratamento de ruído com transformada *Wavelet*



Fonte: Produção do próprio autor.

9.6 Tratamento de Completude de Dados

Na presente seção será apresentado o tratamento de completude dos dados após serem realizadas as adequações apresentadas nas seções 9.1 a 9.5. Serão tratados os dados que não foram registrados no banco e *outliers* removidos. É válido ressaltar que os instantes de parada não são considerados dados incompletos, eles são desconsiderados das análises.

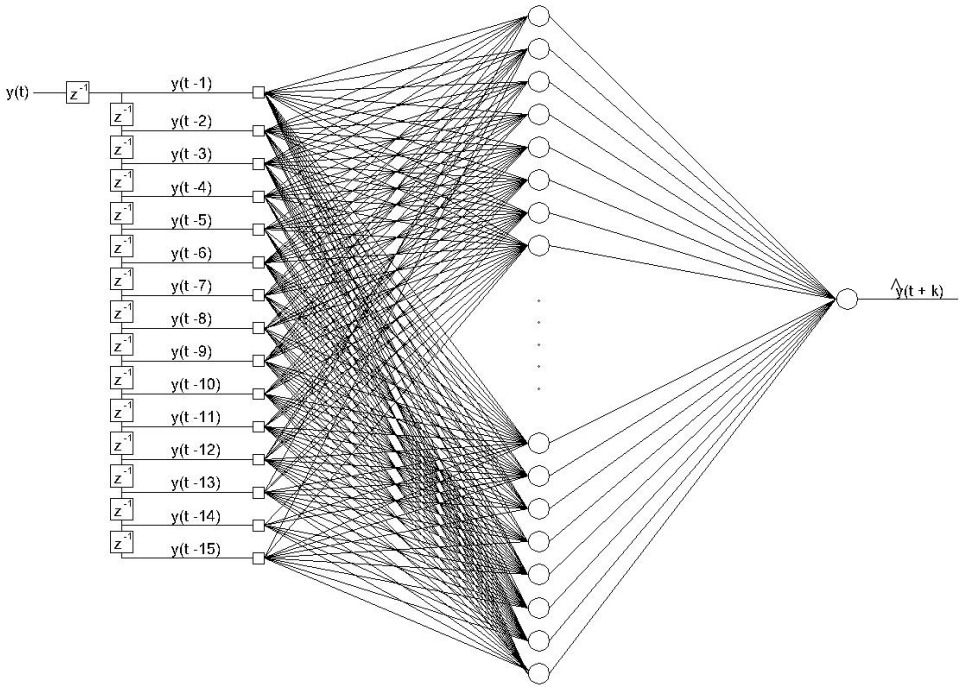
Foram utilizadas redes neurais recorrentes para a completude dos dados da variável de Velocidade da Máquina de Sinterização e Pressão no exaustor, que apresentaram ausências consecutivas de registros. Dessa forma, foi necessário completar os dados por meio de um método com maior precisão que reflita o comportamento da variável ao longo do tempo.

Para isso, para cada uma das variáveis, foi utilizada uma rede neural NAR com 100 neurônios na camada oculta. Para definição dessa quantidade de neurônios foram testados, considerando 15 estados regressivos, 25, 50, 75, 100 e 125 neurônios, e escolheu-se o que atribuía menor erro quadrático médio e correlação do erro, e regressão de teste mais próxima de 1.

Quanto a quantidade de estados regressivos, foram testados de 15 em 15 estados regressivos para se obter um menor erro quadrático médio e correlação além de regressão próxima de 1.

Foram utilizados 15 estados regressivos para a “Velocidade de Sinterização” e 15 estados regressivos para a “Pressão no Exaustor”. O algoritmo de treinamento utilizado foi o *Levenberg-Marquardt backpropagation*. A estrutura da rede neural utilizada para a variável “Velocidade de Sinterização” pode ser observada por meio da Figura 21.

Figura 21 – Estrutura da topologia da rede NAR aplicada ao presente trabalho (15 estados regressivos)



Fonte: Produção do próprio autor.

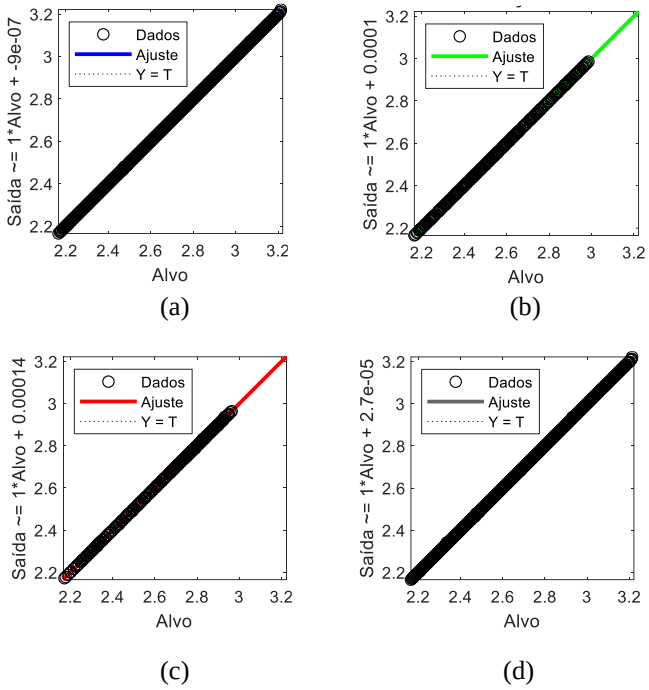
Para treinamento das redes neurais foram utilizados os dados tratados, sem *outliers* e já interpolados. Nesse caso, são também removidos os instantes que apresentam muita variação e que, da mesma forma, são dados factíveis do processo, para treinamento da rede neural.

Os grupos de dados em completude dessas variáveis foram divididos em três segmentos destinados a treino (70%), teste (15%) e validação (15%). A validação e verificação da adequação do modelo da rede neural foi observado por meio das técnicas de regressão, da autocorrelação do erro e do histograma do erro aplicados a esses conjuntos de dados.

A metodologia de regressão é utilizada para delinear a resposta da rede neural em comparação aos valores esperados. Nesse gráfico, é plotado o alvo (abcissa) em função da saída obtida pela rede neural (ordenada) junto a uma linha de ajuste, conforme é possível observar por meio do

Gráfico 33. A aproximação dos dados sobre uma linha de ajuste, e, consequentemente, a regressão da rede neural mais próxima a um, compreende um treinamento adequado e ajustado da rede.

Gráfico 33 – Regressão do treino da rede neural da variável "Velocidade da Máquina de Sínter" com dados de (a) Treino (b) Validação (c) Teste (d) Totais



(a) (b)

(c) (d)

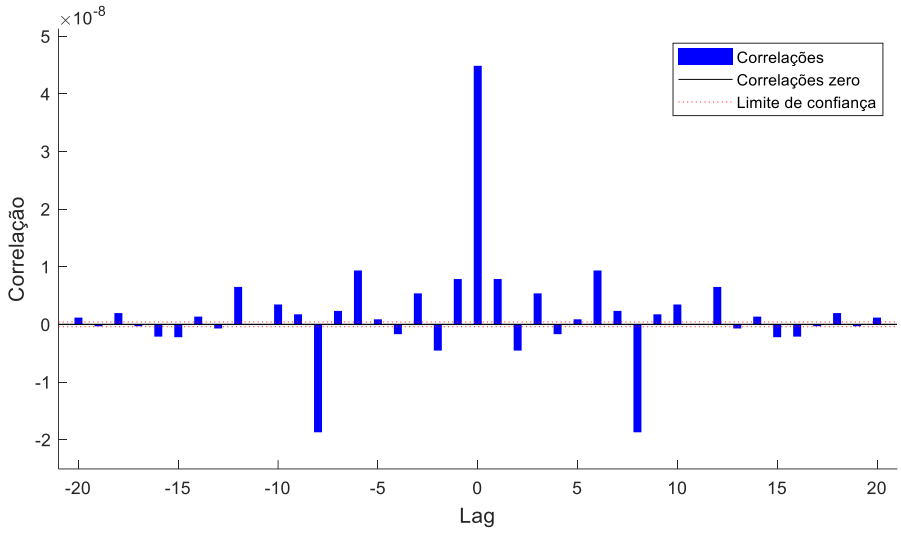
Fonte: Produção do próprio autor.

A verificação por meio dos coeficientes de autocorrelação do erro foi baseada na função de autocorrelação do erro $R_{\varepsilon\varepsilon}$ entre o instante t_k e seu defasamento t_{k-1}. Essa variável pode ser calculada por meio da equação (14):

$$R_{\varepsilon\varepsilon}(t_{k-1}, t_k) \triangleq E[\varepsilon(t_{k-1})\varepsilon^*(t_k)] \tag{14}$$

em que E é o valor esperado entre o MSE (ε) e seu conjugado. Constata-se que o modelo da rede neural está adequado, quando não há correlação entre os erros nos instantes de tempo tratados, isto é, quando os coeficientes resultantes são nulos (TÁPIA, 2000). Para aplicações concretas, busca-se o menor erro associado considerando a precisão necessária a ser aplicada a essa variável. É possível observar, por meio do Gráfico 34, os erros apresentados e a correlação do erro em relação a diferentes defasagens da variável de Vazão no exaustor 2.

Gráfico 34 – Autocorrelação do erro com 20 *bins* da “Velocidade da Máquina de Sínter”



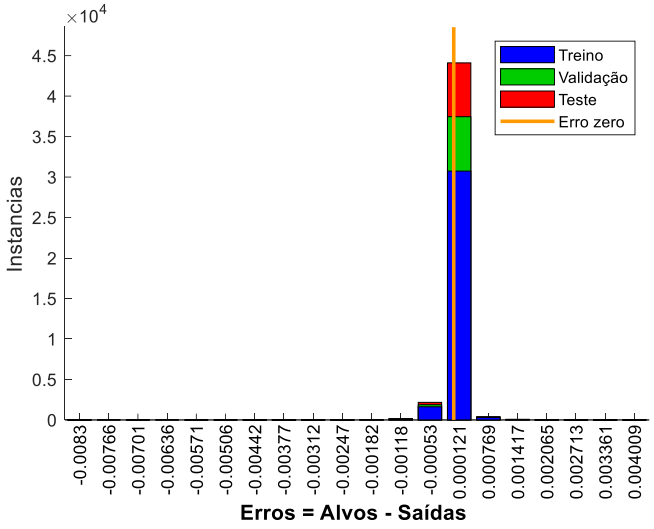
Fonte: Produção do próprio autor.

Por meio do Gráfico 34, compreende-se que a autocorrelação do erro associada a variável “Velocidade da Máquina de Sínter” se apresenta pequena, na ordem de 10^{-8}. Esse valor não é considerável em relação a ordem de grandeza associada e suficiente para a aplicação em questão. Para a variável “Pressão no Exaustor” a autocorrelação do erro foi da ordem de 10^{-3}.

O histograma do erro, conforme mostrado no Gráfico 35, exibe de forma gráfica, a diferença entre o alvo e as saídas obtidas como resultado da aplicação da rede neural ao conjunto de treino, teste ou validação de forma a avaliar o desempenho do treinamento. É quantificada o número de instâncias (ordenada), associada aos erros (abcissa).

A variável “Velocidade de Sinterização” apresenta registros que variam, a cada duas amostras, em ao menos na unidade de centena e o erro associado varia na quarta casa decimal. Além disso, busca-se no presente trabalho, a relação das variáveis com o objetivo de observar a sua influência sobre o BTP. Como o BTP é calculado por meio de uma aproximação de segundo grau, o erro associado a medição dessa amostra se torna desprezível. Para o caso da variável “Pressão no Exaustor” o erro varia na segunda casa decimal, todavia, a cada duas amostras, seu valor é variado em mais de uma unidade.

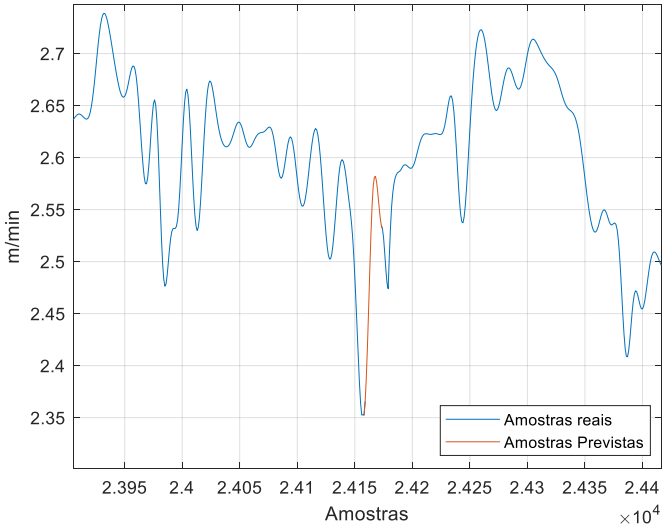
Gráfico 35 – Histograma de erro de treino da Rede Neural para “Velocidade da Máquina de Sínter”



Fonte: Produção do próprio autor.

É possível concluir por meio do Gráfico 35 e análises realizadas, que a maior parte das instâncias tiveram erros associados desprezíveis, no que tange a necessidade de precisão da variável a ser medida. Em vista disso, constata-se que o treinamento da rede foi satisfatório para a aplicação de interesse. Observa-se, por meio do Gráfico 36, um exemplo da aplicação da rede neural às variáveis.

Gráfico 36 – Exemplo de completude dos dados da variável “Velocidade da Máquina de Sínter”



Fonte: Produção do próprio autor

Para realizar a completude, as variáveis foram previstas por meio da rede neural, em seguida foram removidas duas amostras previstas e duas amostras reais imediatamente após ao fim da previsão e os valores ausentes foram interpolados para garantir uma continuidade do sinal.

As variáveis “Densidade da Mistura”, “Vazão no Exaustor 1”, “Vazão no Exaustor 2” apresentaram dados sequencialmente incompletos, chegando à ausência de 9 horas seguidas de medições e 48 horas de ausência de medições ao final do banco de dados. As variáveis relacionadas a vazão, apresentaram ausências de registros, além disso, nas primeiras amostras e após paradas de processo.

Por meio das variáveis amostrada a cada 1 hora, não foi possível treinar a rede neural, pois as variações bruscas dificultavam sua convergência. Enquanto isso, após amostrado e filtrado a cada 10 minutos, como a quantidade de dados para completude era muito grande, mesmo que o modelo de rede neural tenha sido validado, ao ser aplicado a uma grande quantidade de amostras ele era invalidado. Com exceção desses grandes intervalos, as variáveis foram completadas. Considera-se, nesse caso, a possibilidade de descarte dessas variáveis, não utilizando-as para previsões associadas a curva de temperatura da planta.

10 ANÁLISE DE FATORES

Será utilizada, no presente trabalho, a técnica de análise de fatores para estudar a interdependência linear entre as variáveis do processo. Examina-se, primeiramente, a integração conceitual entre as variáveis, seu intervalo de amostragem e a separação entre variáveis dependentes e independentes do processo, por meio de análise de correlação linear garantindo a validade da aplicação.

São aplicadas análises estatísticas para garantir, matematicamente, que as variáveis apresentam correlação suficiente para aplicação da análise fatorial e, em seguida, é aplicada a PCA, com o objetivo de efetuar a redução da quantidade de variáveis que reflitam o comportamento do processo e observar o nível de correlação entre as variáveis.

10.1 Análise de Correlação Linear de Pearson

Primeiramente, deve-se verificar, de acordo com pressupostos conceituais e análises prévias, a adequação da aplicação da análise aos dados da planta de sinterização. São selecionadas variáveis do processo que podem ser linearmente associadas, baseada em critérios conceituais e de amostragem para a formação da matriz de correlações, calculadas com base no coeficiente de Pearson. Segundo Hair e outros (2009), deve-se validar a análise de fatores, partindo de uma correlação de 0,3. Para ele, são consideradas correlações acima de 0,5.

O Quadro 5 apresenta a matriz aplicada ao grupo de dados composto por informações, conceitualmente correlatadas, que abordam a composição da matéria-prima de entrada. As cores nos quadros que serão apresentados na presente seção representam o nível da variação detectado. Quanto mais próximo de 1 ou -1, é evidenciada a correlação entre as variáveis do processo, considerando o coeficiente de Pearson.

É possível observar, por meio do Quadro 5, uma alta correlação entre as variáveis 6 e 14, que correspondem a presença de compostos RDEG em dois conjuntos preparação da matéria-prima, a mistura homogênea e a mistura total de entrada. Essa correlação já era esperada, uma vez que a massa total de compostos RDEG não se altera quando se adiciona os fundentes e combustível

sólido que forma a mistura total. Logo, a porcentagem em relação ao total do RDEG em cada uma das misturas é proporcional, linearmente correlatado.

Quadro 5 – Correlações de Pearson entre as variáveis de composição da mistura

Variáveis	1	2	6	7	8	9	10	11	12	13	14
1	1,000	-0,259	-0,365	-0,226	-0,047	-0,001	-0,174	-0,092	0,415	-0,350	-0,321
2	-0,259	1,000	0,135	0,275	0,027	-0,417	-0,274	-0,307	0,024	-0,141	0,088
6	-0,365	0,135	1,000	0,533	-0,050	0,036	0,224	0,033	-0,267	0,209	0,824
7	-0,226	0,275	0,533	1,000	-0,281	-0,575	-0,211	-0,478	-0,126	-0,097	0,369
8	-0,047	0,027	-0,050	-0,281	1,000	0,160	0,225	-0,006	-0,161	0,126	-0,121
9	-0,001	-0,417	0,036	-0,575	0,160	1,000	0,816	0,882	-0,438	0,680	0,167
10	-0,174	-0,274	0,224	-0,211	0,225	0,816	1,000	0,725	-0,828	0,920	0,275
11	-0,092	-0,307	0,033	-0,478	-0,006	0,882	0,725	1,000	-0,445	0,733	0,251
12	0,415	0,024	-0,267	-0,126	-0,161	-0,438	-0,828	-0,445	1,000	-0,931	-0,310
13	-0,350	-0,141	0,209	-0,097	0,126	0,680	0,920	0,733	-0,931	1,000	0,327
14	-0,321	0,088	0,824	0,369	-0,121	0,167	0,275	0,251	-0,310	0,327	1,000

Fonte: Produção do próprio autor.

Verifica-se, além disso, uma forte correlação entre as variáveis correspondentes a granulometria da matéria-prima do processo. As variáveis 9 a 13 apresentam alto grau de correlação linear de Pearson.

A variável 13 é fornecida diretamente como uma relação entre duas variáveis também registradas, as variáveis 11 e 12. Portanto, a variável 13 não adiciona informação extra para a análise de dados e, por ela ser resultado de uma divisão simples entre as variáveis 12 e 11, ela apresenta alta correlação linear com a variável 12. Esse dado é registrado no banco, pois é por meio dele que, visualmente, o operador pode ter consciência da proporção entre as duas granulometrias, mas não é necessário manter as três informações pois não acrescentaria informações para treinamento de máquinas de inteligência.

As variáveis 9 e 11 se referem a fração da granulometria do material acima de 1mm em dois momentos, na mistura que forma e pilha homogeneizada e a outra que formará a mistura total que será utilizada como entrada do processo de sinterização. A segunda amostra apresenta informações acerca da mudança da granulometria da mistura após a adição do combustível sólido que forma a mistura final. É possível concluir que a fração se mantém proporcional nas duas amostras. Por isso deve-se avaliar, ao treinar uma máquina de inteligência, se as duas informações contribuem ou não para o desempenho do treinamento.

Embora as variáveis 9 e 10 estejam diretamente relacionadas, não se pode suprimir uma delas por conta da perda de informações acerca de granulometrias baixas presentes na estrutura. Como não foram fornecidas a informação acerca do percentual de baixa granulometria no conjunto estudado e não foi informada a metodologia de cálculo dele, não é possível eliminar, por vias conceituais uma dessas variáveis. Contudo é possível realizar uma análise de fatores que inclua as três variáveis, 9, 10 e 11, já que elas se apresentam altamente correlatadas.

As variáveis 6 e 7 apresentam uma correlação linear acima da média, entretanto não é possível considerar uma correlação entre elas, pois a primeira é referente a um percentual de finos de retorno que é reutilizado na planta de sinterização enquanto a segunda refere-se a um percentual do minério proveniente da mina localizada em Brucutu. Essas variáveis não apresentam relação conceitual, diante disso a correlação linear apresentada não é passível de interpretação.

Há também uma correlação significativa entre as variáveis 7 e 9, que representam, respectivamente, o percentual de minério proveniente da mina de Brucutu e a granulometria acima de 1 mm na pilha homogeneizada. A granulometria do composto de minério pode estar associada à sua origem, mas, por não se ter informações acerca da qualidade do minério dessa mina, não é possível validar conceitualmente uma relação de influência direta entre esses dois parâmetros.

O Quadro 6 apresenta a correlação de Pearson aplicada aos grupos de dados que são compostos por variáveis de controle e indicação do processo de sinterização, amostrados a cada 10 minutos.

Identifica-se pelo Quadro 6 que as variáveis 17 e 18 apresentam alto coeficiente de correlação. Conforme informado anteriormente presume-se que a variável “Velocidade de Sinterização” seja um parâmetro calculado a partir da variável “Velocidade da Máquina de Sínter”. Por meio do resultado da correlação de Pearson constata-se que esse cálculo seja baseado em uma equação linear. Por esse motivo, essa correlação não será considerada para a formação de um novo fator.

As variáveis 27 e 30 também apresentaram coeficiente de correlação acima do esperado. O alto valor apresentado mostra que existe uma correlação linear entre a pressão nas últimas caixas de

vento e vazão nos exaustores. Esse fato confere o processo real da planta e elucida a influência da vazão dos exaustores nas pressões das caixas de vento.

Quadro 6 – Correlações de Pearson entre as variáveis de controle e indicação do processo

Variáveis	17	18	27	31
17	1,000	0,911	0,183	-0,124
18	0,911	1,000	0,269	-0,121
27	0,183	0,269	1,000	0,293
31	-0,010	-0,010	0,021	-0,027

Fonte: Produção do próprio autor.

O Quadro 7 apresenta a correlação de Pearson aplicada ao grupo de indicação do processo de sinterização, referente as temperaturas das caixas de vento número 11, 13 e 18. Nota-se que ocorreu uma maior correlação linear entre as três temperaturas, especialmente entre as variáveis 19 e 20.

Quadro 7 – Correlações de Pearson entre as variáveis de indicação do processo

Variáveis	19	20	21	35
19	1,000	0,744	0,403	-0,246
20	0,744	1,000	0,488	-0,392
21	0,403	0,488	1,000	-0,244
22	-0,246	-0,392	-0,244	1,000

Fonte: Produção do próprio autor.

Foi avaliada, além disso, a correlação dessas temperaturas e produtividade em relação a variáveis de controle e indicação do processo, considerando a amostragem a cada 1 hora. Conclui-se que por meio dessa análise não foi possível obter correlações lineares significativas.

Compreende-se que, mesmo que as variáveis de temperatura das últimas 5 caixas de vento apresentem uma alta correlação linear entre si, a análise fatorial, baseada nos componentes principais não é aplicável sobre elas, já que as mesmas são utilizadas diretamente para o cálculo estimativo do BTP, conforme apresentado nas seções 2.1.3 e 0. Além disso, já é conhecido que as relações entre as temperaturas do processo e as variáveis relativas à temperatura e posição do BTP são quadráticas e, desse modo, não é válida a realização de análises lineares sobre elas.

As demais correlações não foram apresentadas por serem insignificantes e não aplicáveis, considerando a métrica da correlação de Pearson, apresentada na seção 4.3.1.1.

10.2 Estatística MSA

Além da métrica de Pearson, foi aplicada uma técnica de verificação da adequação das variáveis a Análise Fatorial, a estatística MSA. Optou-se por não aplicar a estatística KMO, que apresenta mesma base conceitual que a MSA, por ela levar em conta o conjunto completo de variáveis do processo. Nesse caso, como a maior parte das variáveis não se correlacionam linearmente, o seu valor seria baixo, mesmo com a comprovação da aplicabilidade da técnica por meio das estatísticas de Pearson e MSA.

Segundo Hair e outros (2009), o pesquisador deve ter um valor de MSA, em geral, maior ou igual a 0,5. Os valores calculados para o grupo de variáveis 9, 10 e 11, extraídos da matriz de correlação de Pearson foi equivalente a 0,58, valor acima do estipulado.

10.3 Análise de Componentes Principais

Aplicou-se, a partir das análises anteriores, a técnica PCA com o objetivo de exibir possibilidades de combinações entre variáveis que possam resultar em um melhor desempenho ao serem introduzidas em uma máquina de aprendizagem. É possível que a aplicação dessas combinações possa contribuir para uma melhoria associada ao treinamento de uma máquina de inteligência para a predição mais adequada de uma variável alvo do processo.

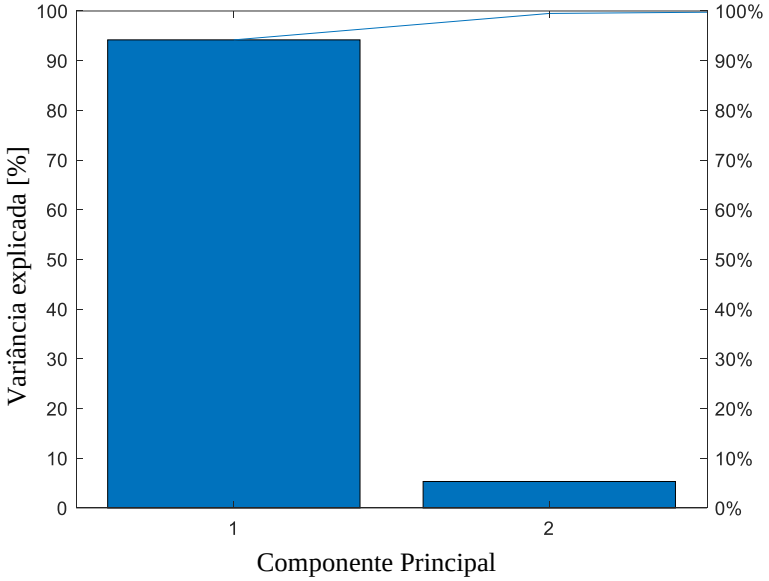
Analisou-se a possibilidade de aplicar a análise de fatores às variáveis 9, 10 e 11. Ao aplicar essa técnica às três variáveis, foi possível chegar aos seguintes resultados. Por meio do Gráfico 37, é possível observar o histograma das componentes principais e suas variâncias associadas. É possível compreender, que as duas primeiras componentes principais associadas carregam consigo informação de mais que 90% da variância dos dados.

E possível reduzir a quantidade de variáveis mantendo uma ou duas componentes principais mantendo, ainda assim, a variância associada a elas, garantindo uma perda de informação muito pequena.

Por meio do Gráfico 38, pode ser visualizado os coeficientes ortogonais do componente principal para cada variável e as pontuações do componente principal para cada observação

normalizadas. Cada uma das variáveis, 9, 10 e 11, são representadas por meio de um vetor. Esses vetores indicam como cada variável contribui para os dois componentes principais no gráfico.

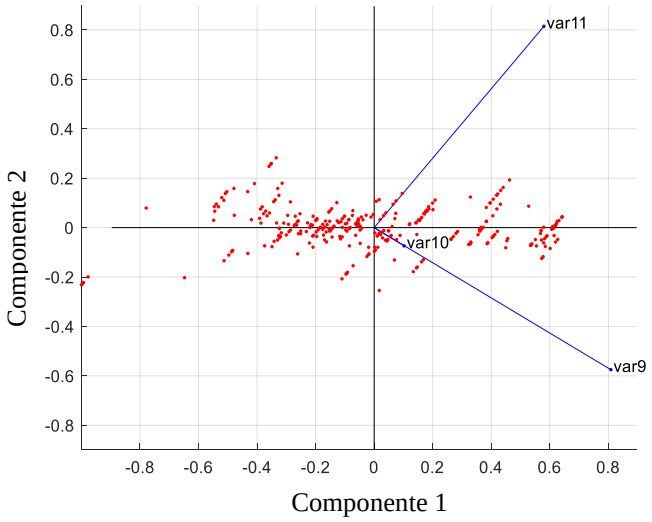
Gráfico 37 – Histograma das componentes principais e suas variáveis associadas



Fonte: Produção do Próprio autor.

Observa-se que as três variáveis apresentam um coeficiente positivo significativo para a formação do primeiro componente principal. Enquanto, ao tratar da segunda componente principal, a variável 11, é a única a apresentar coeficientes positivos, as variáveis 9 e 10, acrescentam coeficientes negativos para a formação do segundo fator.

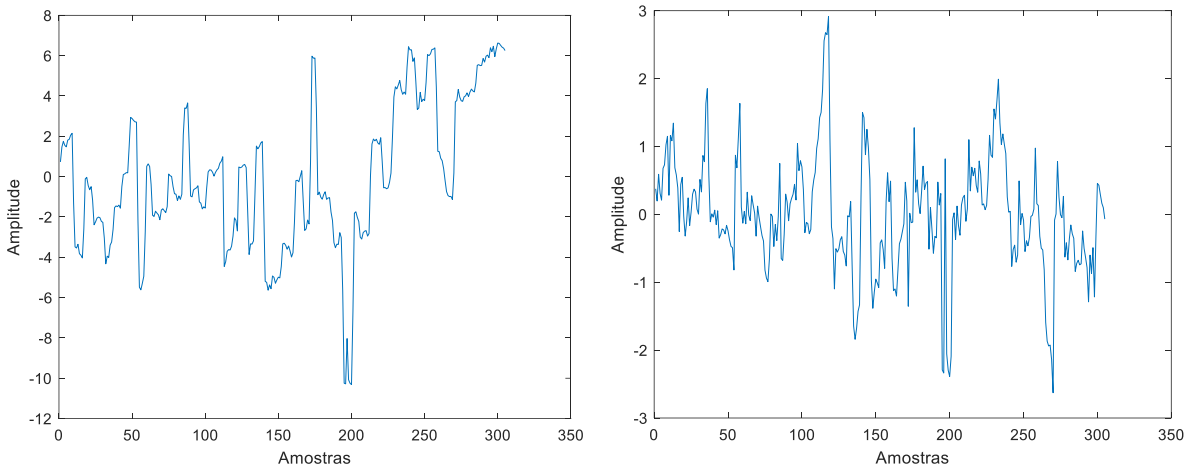
Gráfico 38 – Coeficientes ortogonais do componente principal



Fonte: Produção do Próprio autor.

Foram selecionados, então, dois componentes principais para representar as três variáveis apresentadas. Essas componentes são apresentadas por meio do Gráfico 39 e podem ser utilizadas como uma possibilidade para aplicação de um algoritmo de aprendizado de máquina. A amostragem presente no gráfico é realizada a cada 24 horas.

Gráfico 39 – Variáveis obtidas por meio da análise PCA (a) Primeira Componente (b) Segunda Componente



(a) (b)

Fonte: Produção do Próprio autor.

11 VARIÁVEIS TRATADAS

Foram obtidas, ao final das análises e tratamentos aplicados, as variáveis tratadas. Elas serão apresentadas a seguir, segregadas em relação a sua amostragem final. Poucas mudanças foram aplicadas as variáveis de amostragem a cada 24 horas, foram removidos os dados de parada e variações anômalas exemplificadas pela seção 9.2. Essas variáveis, por isso, não serão exibidas na presente seção.

É válido ressaltar que, nesse caso, as variáveis originais podem apresentar amostragens inferiores a amostragem final das variáveis após tratamento. Deve-se atentar ao se realizar uma comparação entre os dois gráficos *boxplot*, a quantidade de dados fora do intervalo entre os quartis 1 e 3. Os gráficos de linhas serão plotados com base horária e não amostral para que seja possível realizar a comparação da variável nos dois momentos. Vale lembrar que amostras que apresentaram variações, mesmo que significativas, mas que eram reais e faziam parte do processo, foram mantidas.

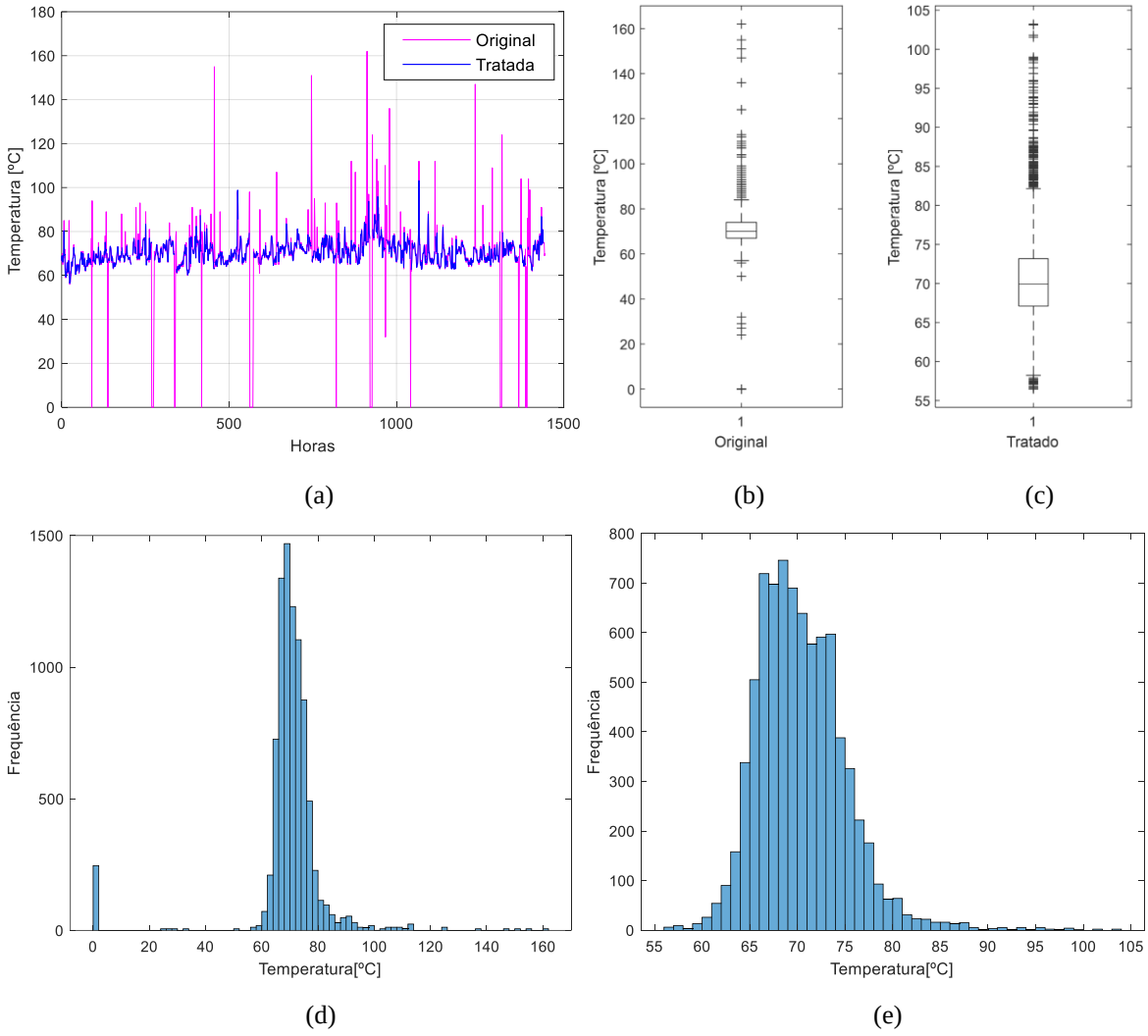
Os Histogramas dos dados antes e após o tratamento também foram plotados de forma a observar a distribuição de frequência de ocorrências em relação aos valores de cada uma das variáveis.

Por meio das seções 11.1 e 11.2 são apresentadas as variáveis que foram reamostradas a cada 1 hora e a cada 10 minutos. Por meio dos gráficos, é possível observar o tratamento aplicado a composição final de cada uma das variáveis.

11.1 Variáveis Reamostradas a cada 1 hora

Na presente seção serão apresentados gráficos finais das variáveis que foram tratadas e reamostradas a cada uma hora. Originalmente, essas variáveis apresentavam amostragem a cada 6 h. São apresentados do Gráfico 40 ao 43, os gráficos de linha, *boxplots* e histogramas de cada uma delas antes e após o tratamento apresentado no presente trabalho.

Gráfico 40 – Temperatura da CV nº 11 (a) Gráfico de linhas (b) Boxplot antes do tratamento (c) Boxplot após tratamento (d) Histograma antes do tratamento (e) Histograma após tratamento

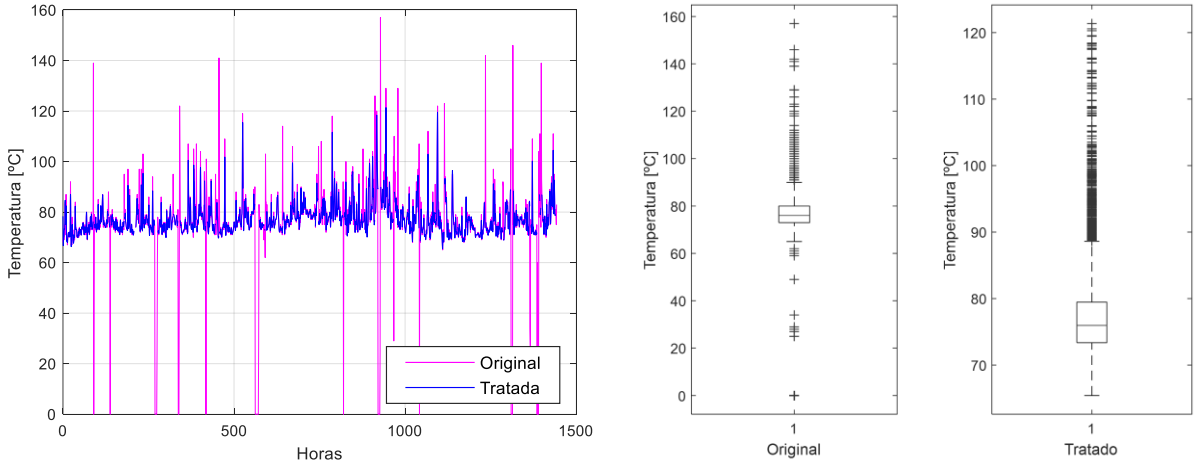


(a) (b) (c)

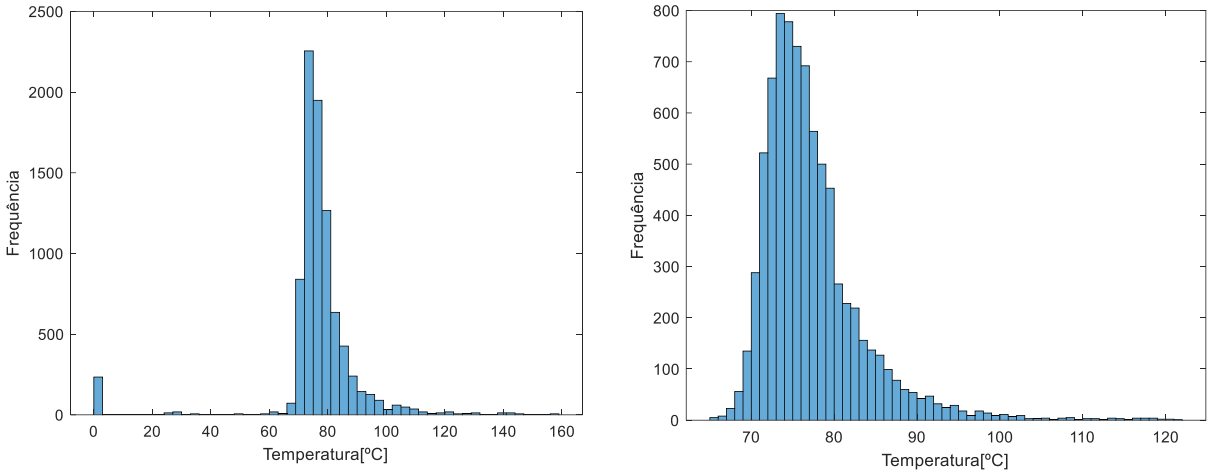
(d) (e)

Fonte: Produção do Próprio autor.

Gráfico 41 – Temperatura da CV nº 13 (a) Gráfico de linhas (b) Boxplot antes do tratamento (c) Boxplot após tratamento (d) Histograma antes do tratamento (e) Histograma após tratamento



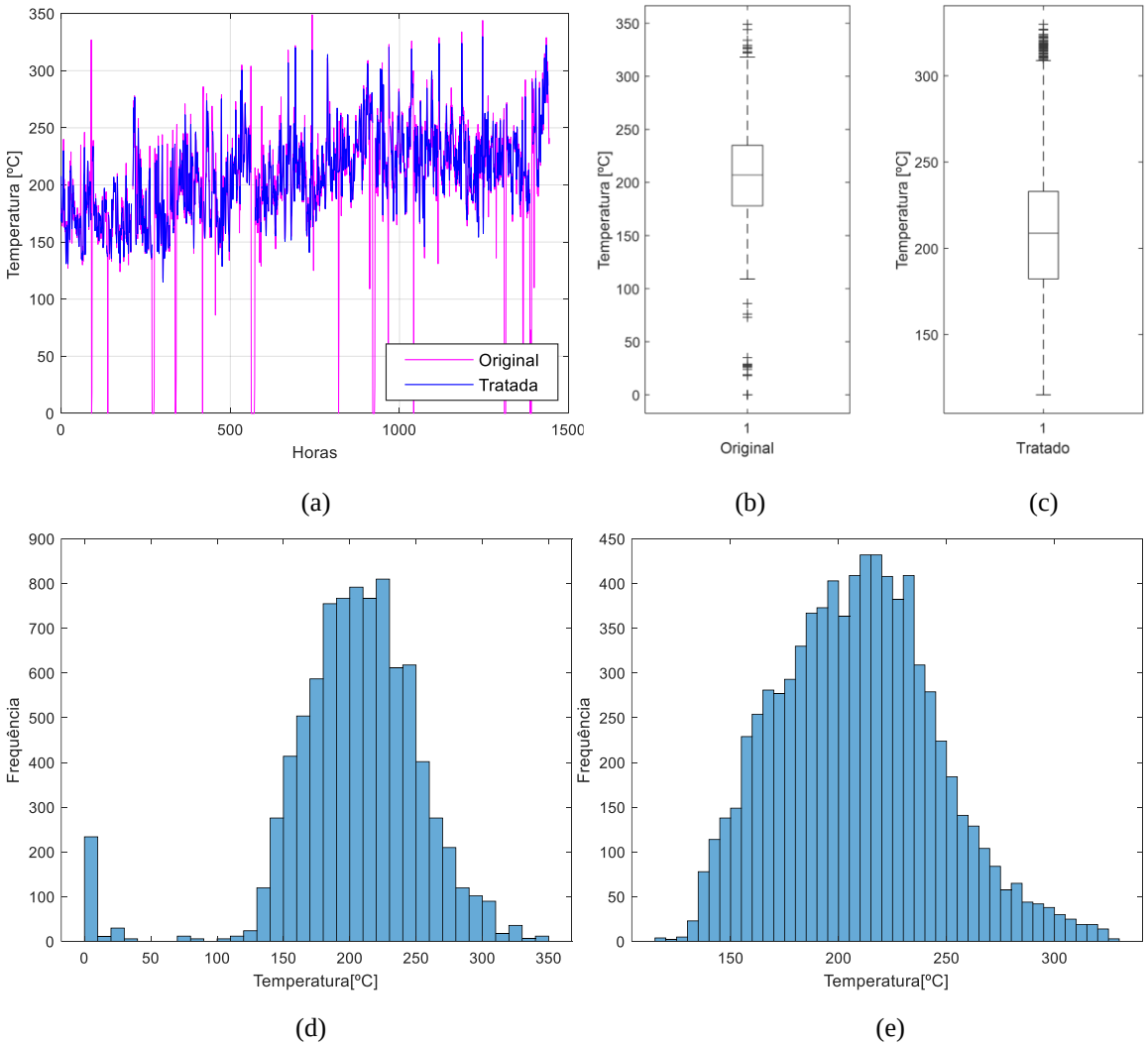
(a) (b) (c)



(d) (e)

Fonte: Produção do Próprio autor.

Gráfico 42 – Temperatura da CV nº 18 (a) Gráfico de linhas (b) Boxplot antes do tratamento (c) Boxplot após tratamento (d) Histograma antes do tratamento (e) Histograma após o tratamento

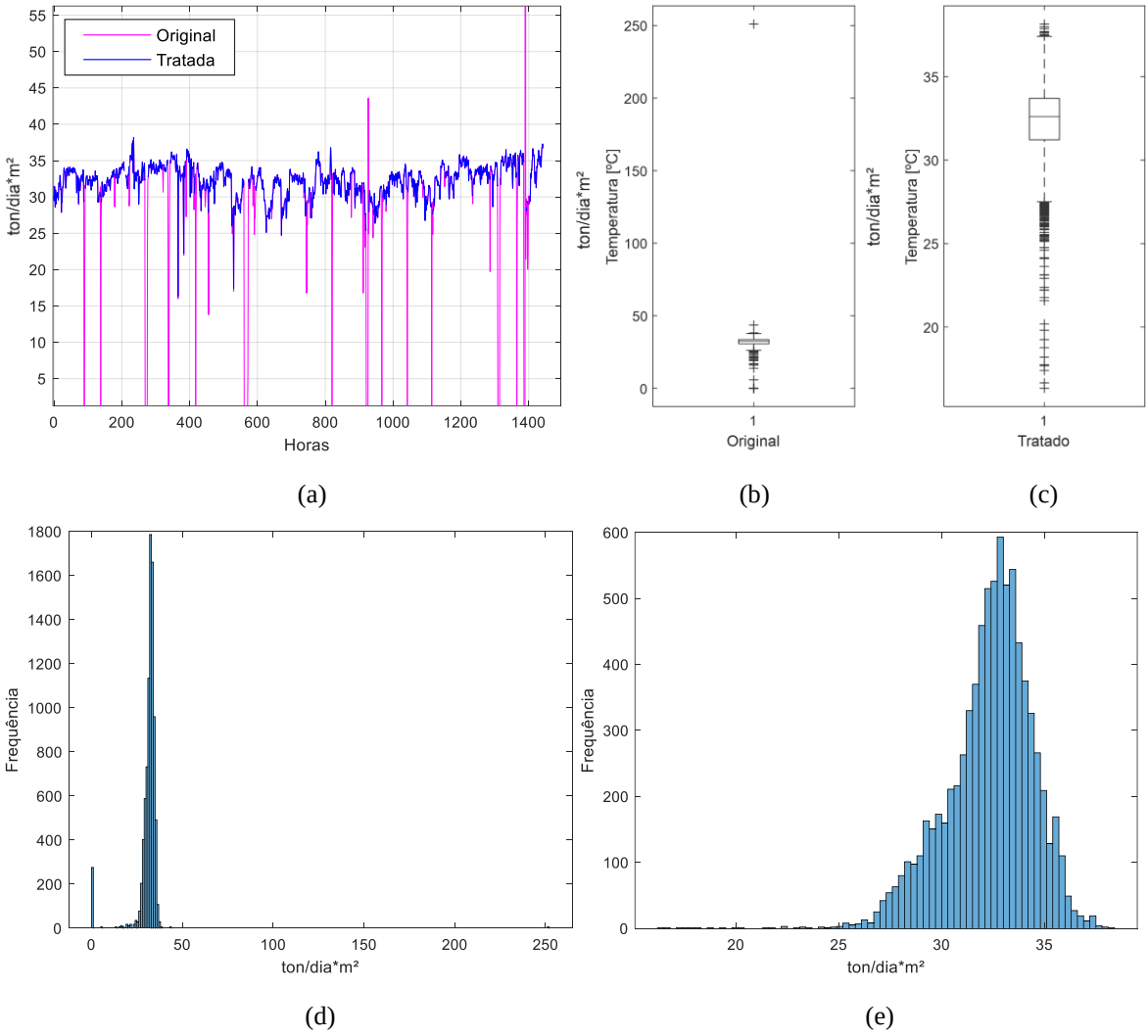


(a) (b) (c)

(d) (e)

Fonte: Produção do Próprio autor.

Gráfico 43 – Produtividade (a) Gráfico de linhas (b) Boxplot antes do tratamento (c) Boxplot após tratamento (d) Histograma antes do tratamento (e) Histograma após o tratamento



(a) (b) (c)

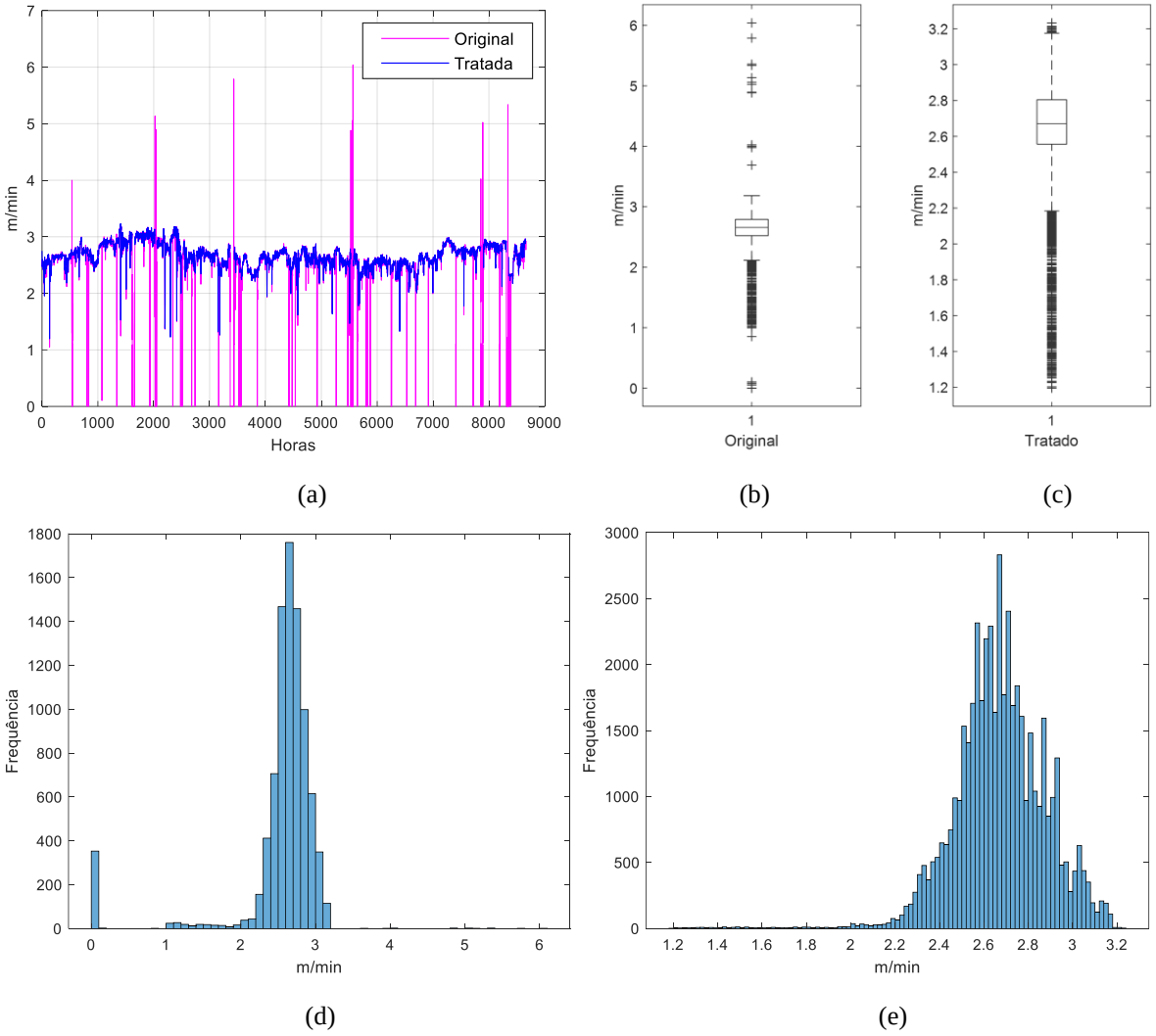
(d) (e)

Fonte: Produção do Próprio autor.

11.2 Variáveis Reamostradas a cada 10 minutos

Na presente seção serão apresentados gráficos finais das variáveis que foram tratadas e reamostradas a cada 10 min. Originalmente, essas variáveis apresentavam amostragem a cada 1 h. São apresentados do Gráfico 44 ao 53, os gráficos de linha, *boxplots* e histogramas de cada uma delas antes e após o tratamento apresentado no presente trabalho.

Gráfico 44 – Velocidade da Máquina de Sínter (a) Gráfico de linhas (b) Boxplot antes do tratamento (c) Boxplot após tratamento (d) Histograma antes do tratamento (e) Histograma após o tratamento

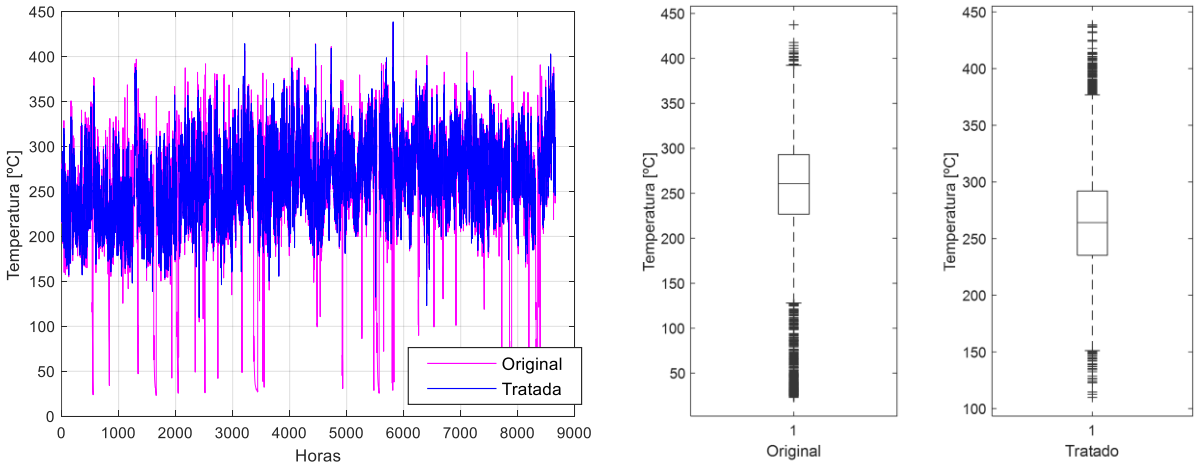


(a) (b) (c)

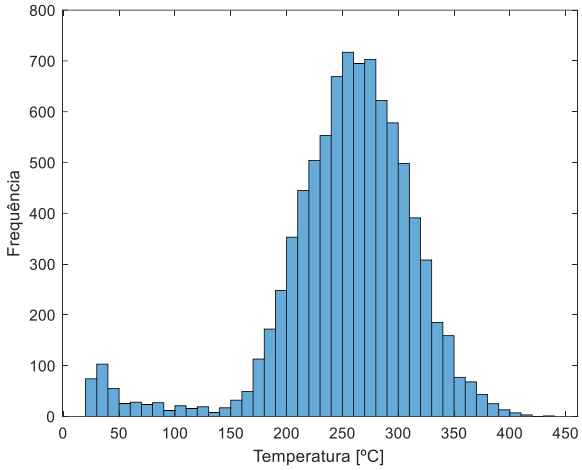
(d) (e)

Fonte: Produção do Próprio autor.

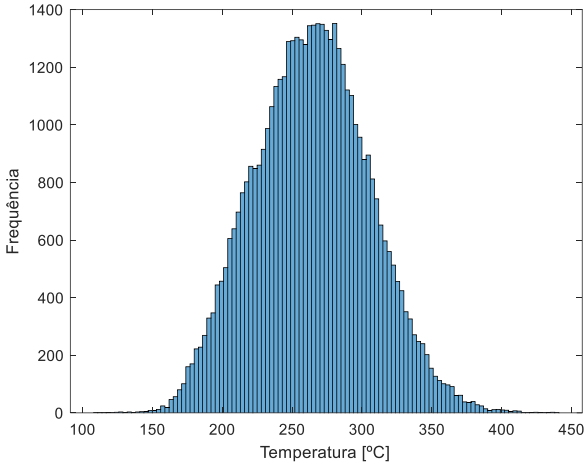
Gráfico 45 – Temperatura da CV nº 19 (a) Gráfico de linhas (b) Boxplot antes do tratamento (c) Boxplot após tratamento (d) Histograma antes do tratamento (e) Histograma após o tratamento



(a) (b) (c)



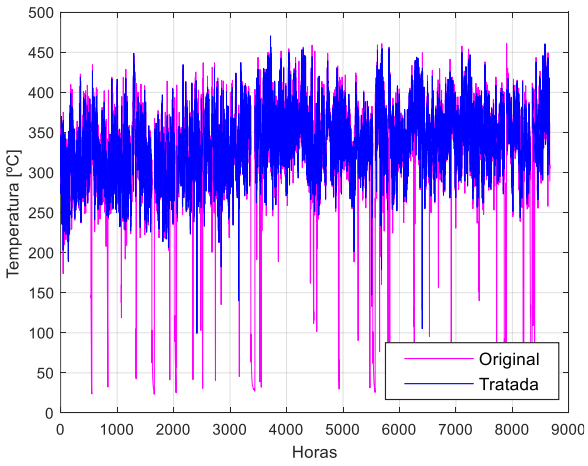
(d)



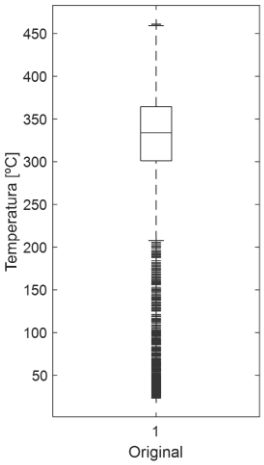
(e)

Fonte: Produção do Próprio autor.

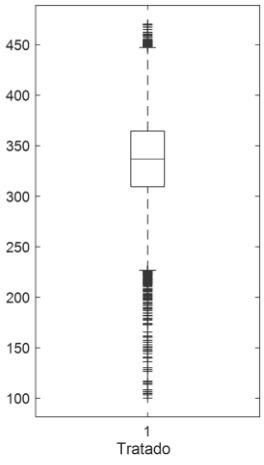
Gráfico 46 – Temperatura da CV nº 20 (a) Gráfico de linhas (b) Boxplot antes do tratamento c) Boxplot após tratamento (d) Histograma antes do tratamento (e) Histograma após o tratamento



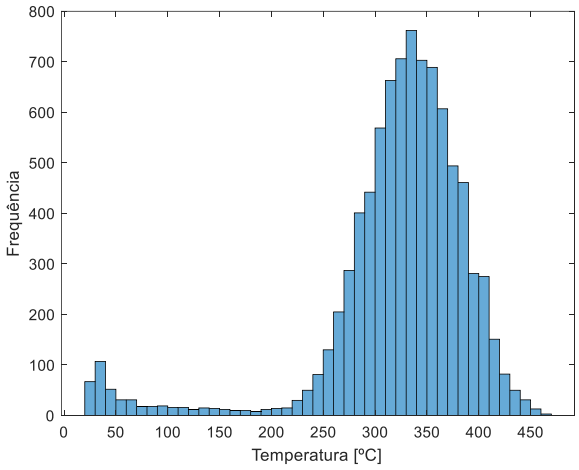
(a)



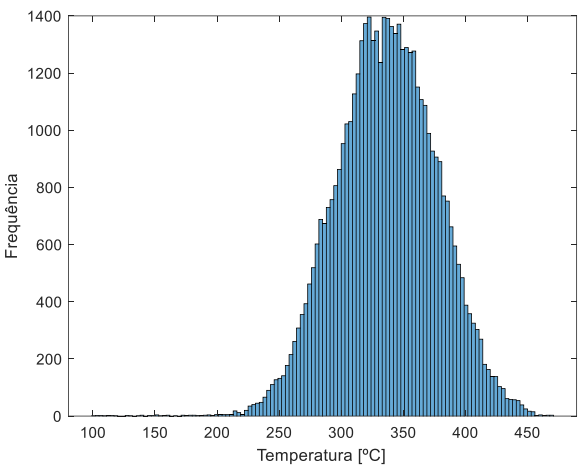
(b)



(c)



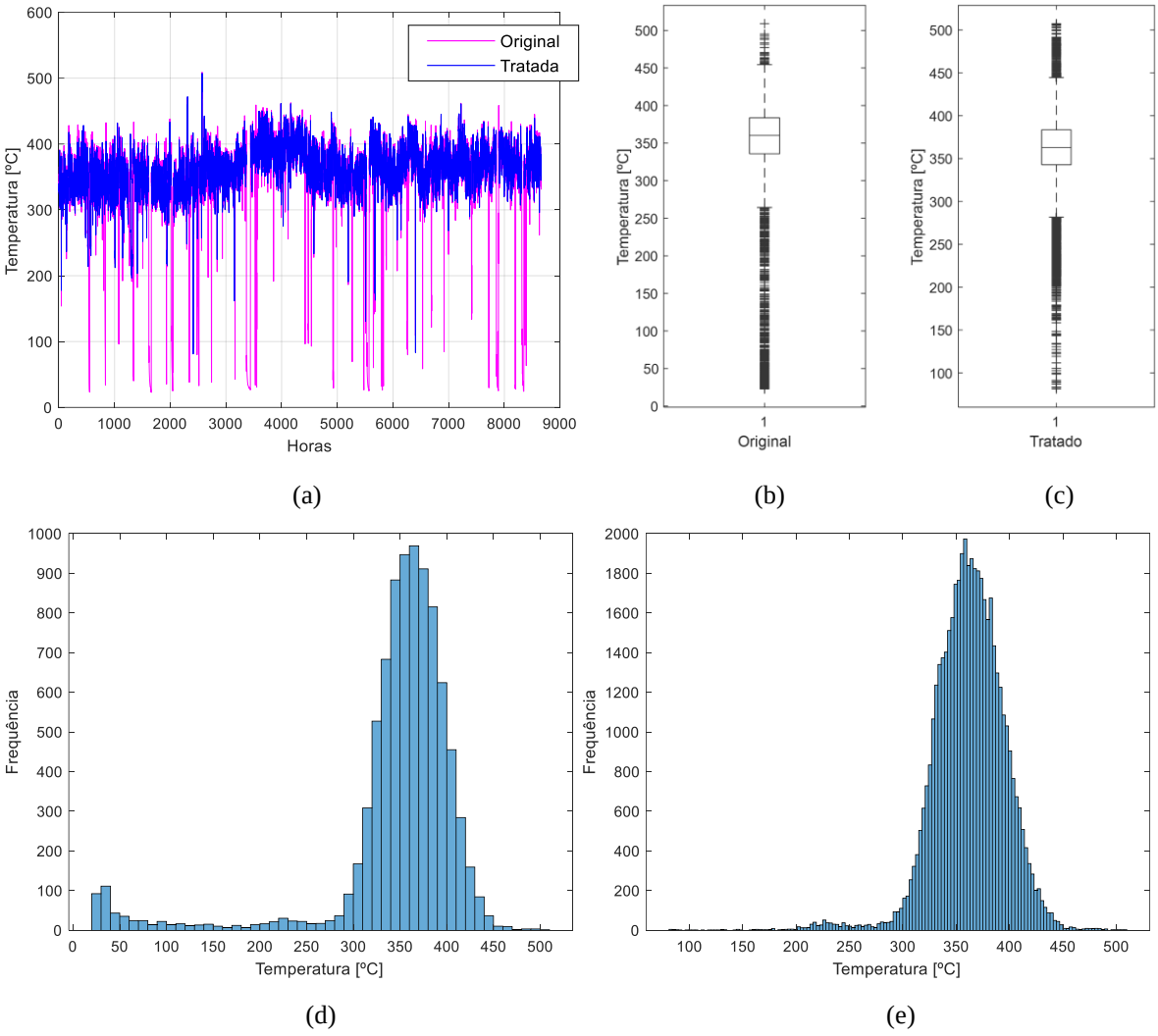
(d)



(e)

Fonte: Produção do Próprio autor.

Gráfico 47 – Temperatura da CV nº 21 (a) Gráfico de linhas (b) Boxplot antes do tratamento (c) Boxplot após tratamento (d) Histograma antes do tratamento (e) Histograma após o tratamento

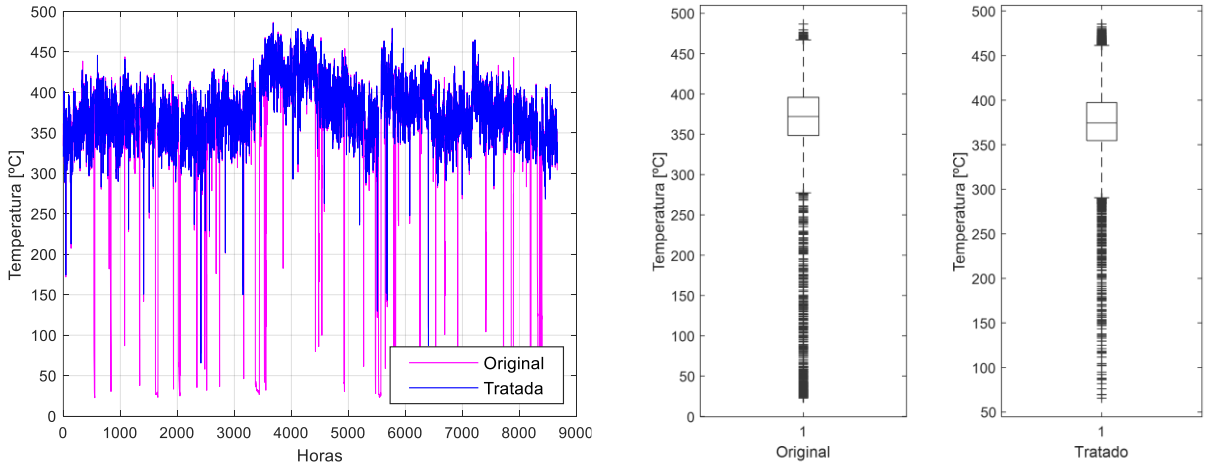


(a) (b) (c)

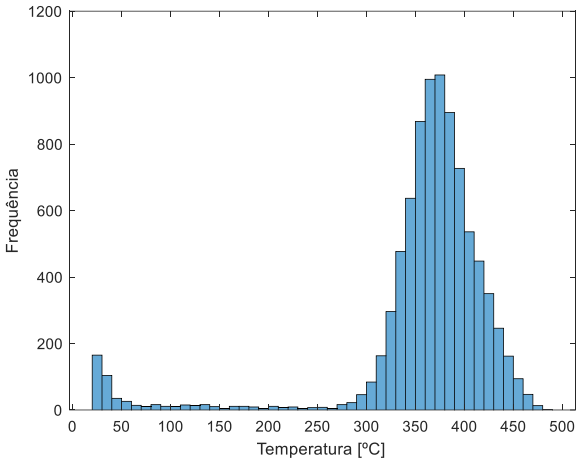
(d) (e)

Fonte: Produção do Próprio autor.

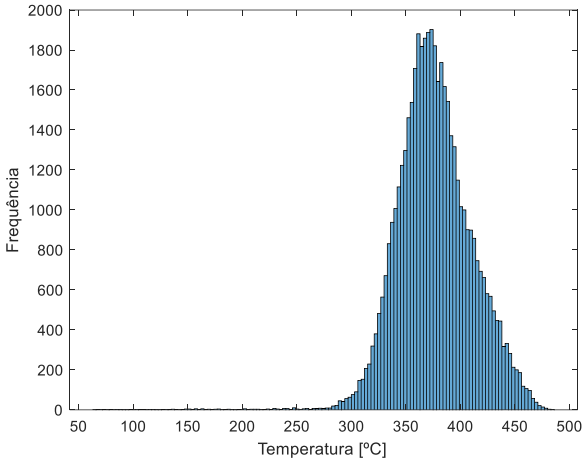
Gráfico 48 – Temperatura da CV nº 22 (a) Gráfico de linhas (b) Boxplot antes do tratamento (c) Boxplot após tratamento (d) Histograma antes do tratamento (e) Histograma após o tratamento



(a) (b) (c)



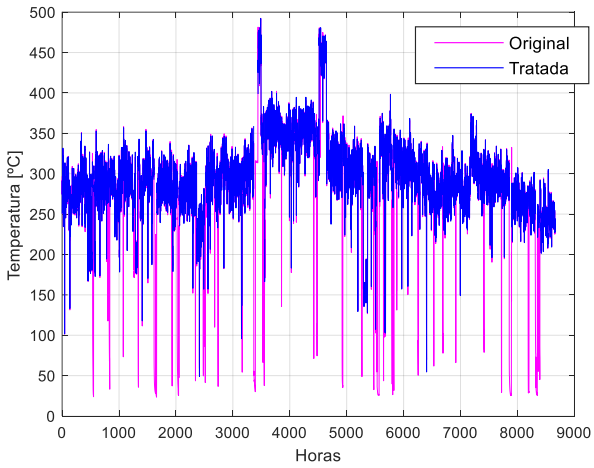
(d)



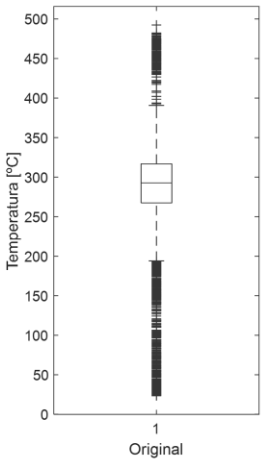
(e)

Fonte: Produção do Próprio autor.

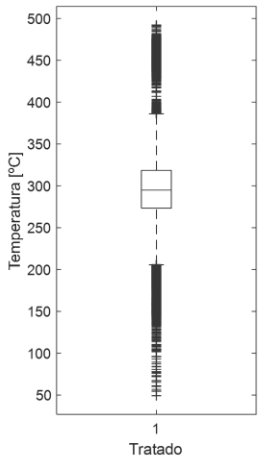
Gráfico 49 – Temperatura da CV nº 23 (a) Gráfico de linhas (b) Boxplot antes do tratamento (c) Boxplot após tratamento (d) Histograma antes do tratamento (e) Histograma após o tratamento



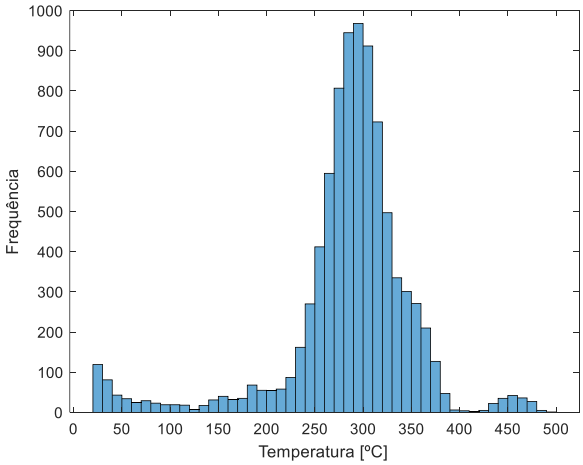
(a)



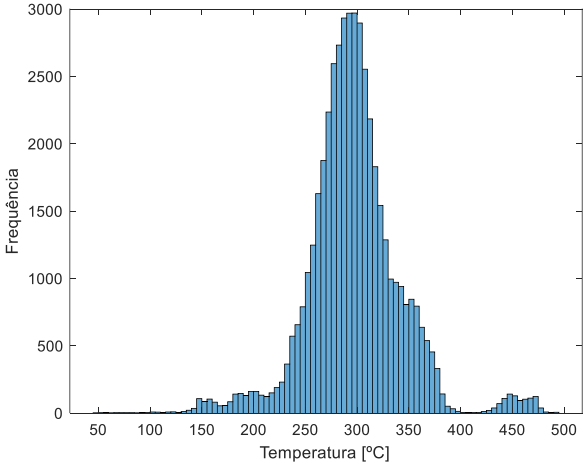
(b)



(c)



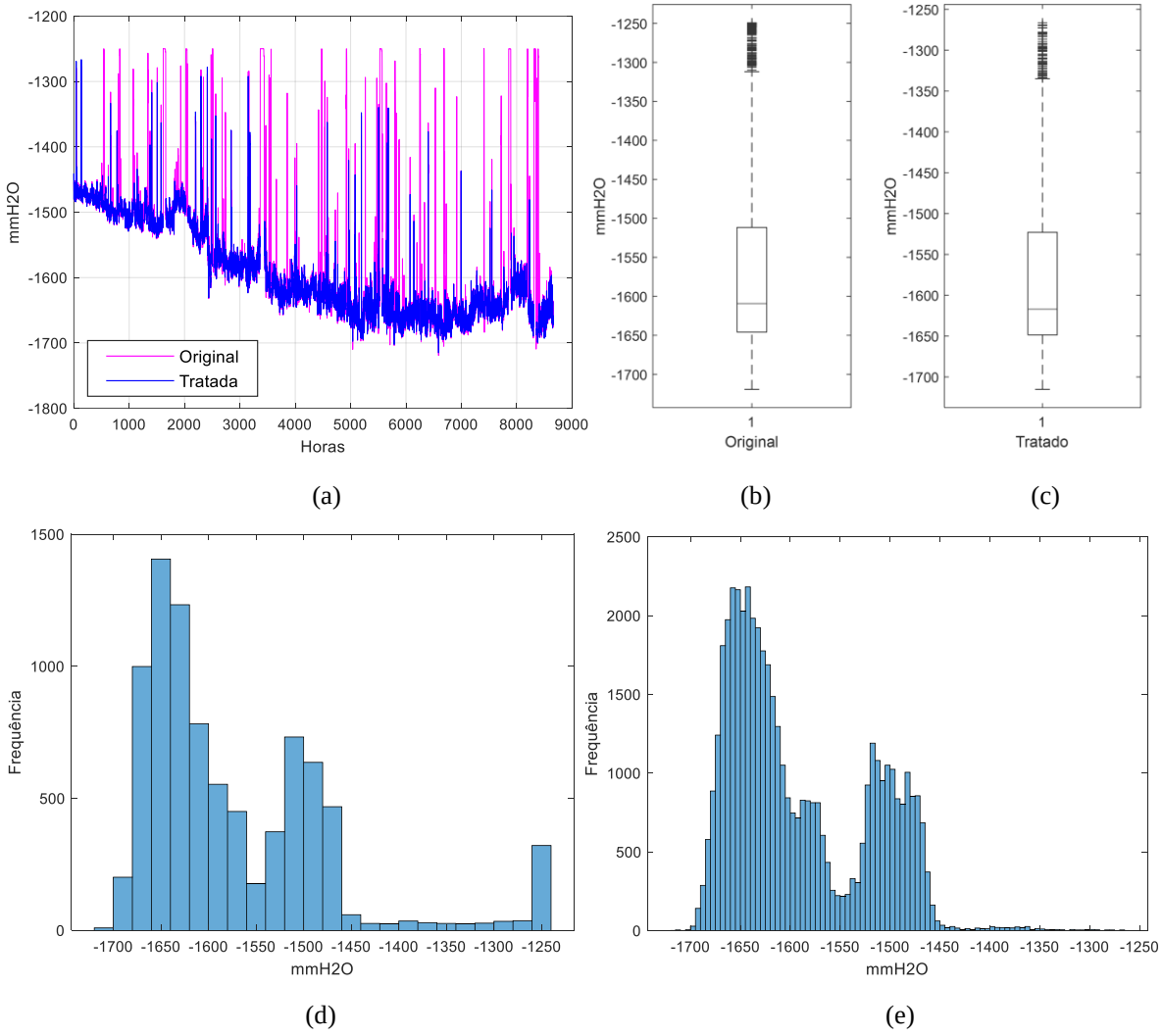
(d)



(e)

Fonte: Produção do Próprio autor.

Gráfico 50 – Pressão na CV nº22 (a) Gráfico de linhas (b) Boxplot antes do tratamento (c) Boxplot após tratamento (d) Histograma antes do tratamento (e) Histograma após o tratamento

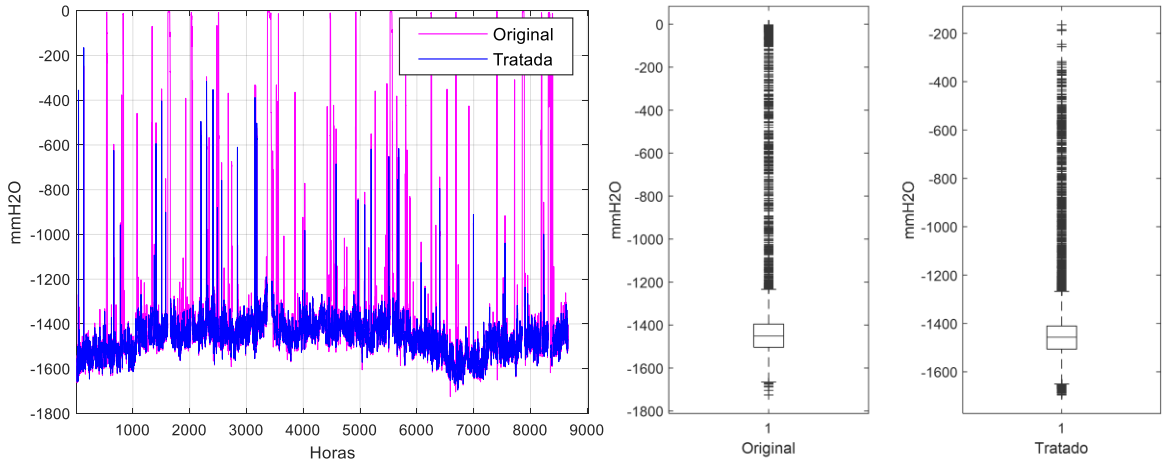


(a) (b) (c)

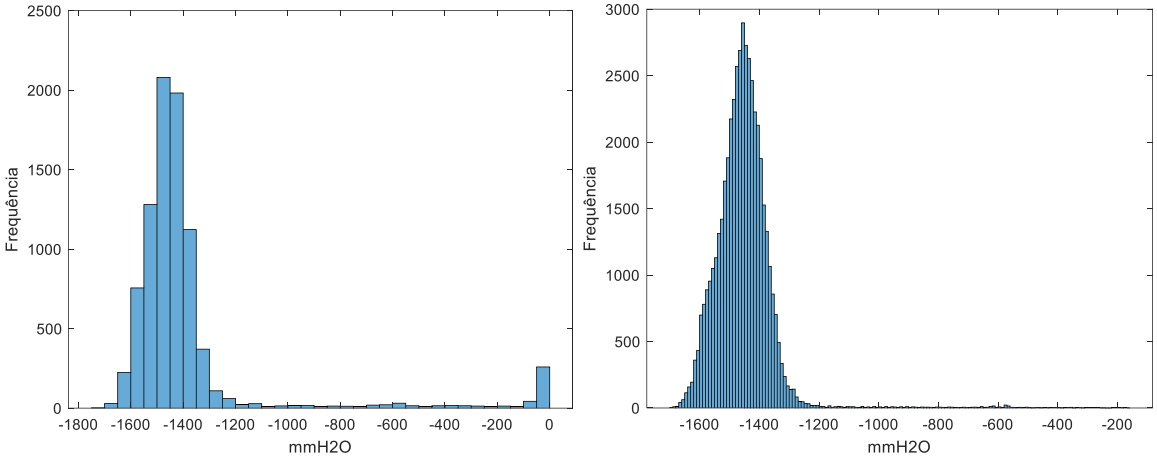
(d) (e)

Fonte: Produção do Próprio autor.

Gráfico 51 – Pressão no exaustor (a) Gráfico de linhas (b) Boxplot antes do tratamento (c) Boxplot após tratamento (d) Histograma antes do tratamento (e) Histograma após o tratamento



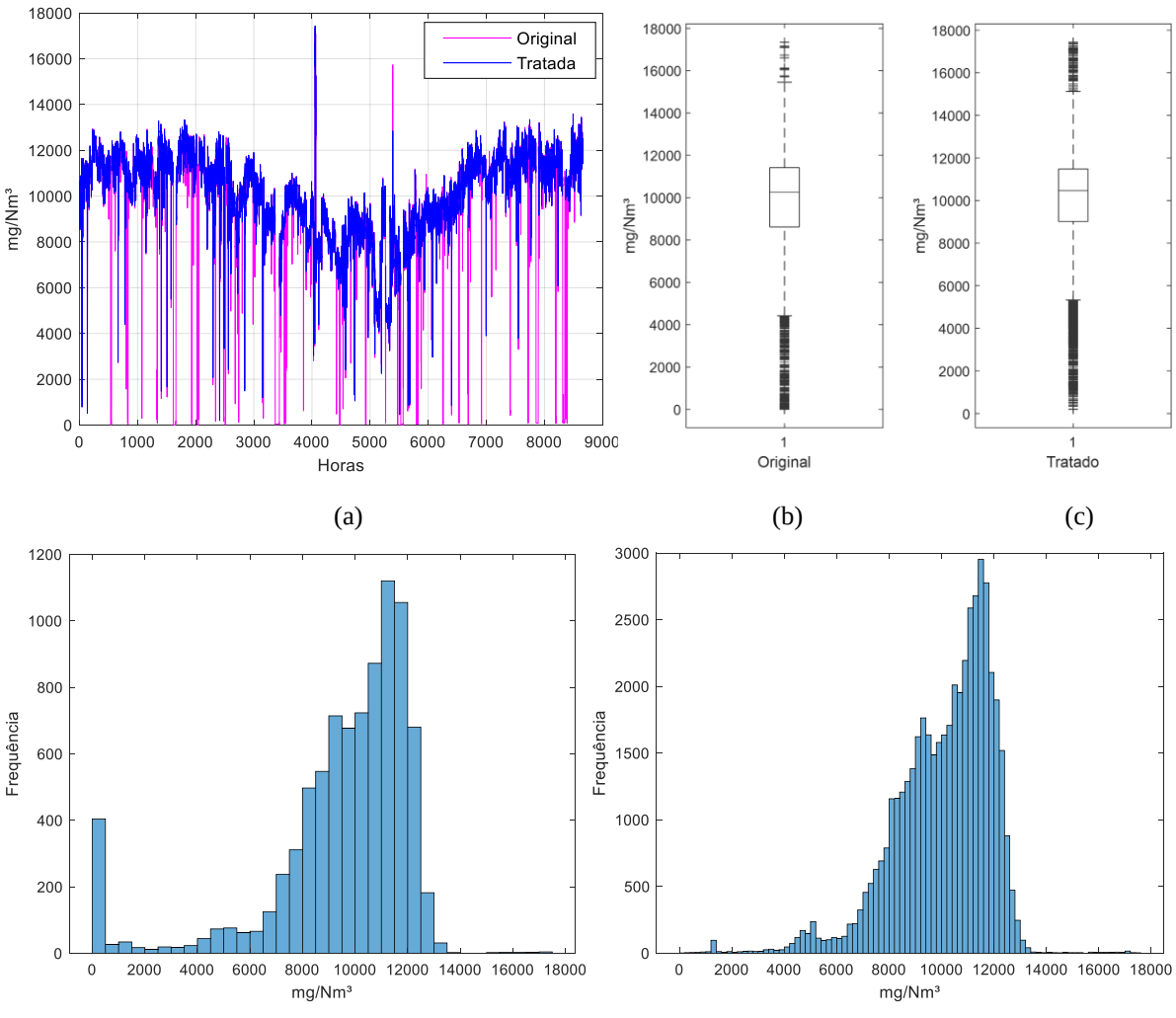
(a) (b) (c)



(d) (e)

Fonte: Produção do Próprio autor.

Gráfico 52 – Emissão de CO (a) Gráfico de linhas (b) Boxplot antes do tratamento (c) Boxplot após tratamento (d) Histograma antes do tratamento (e) Histograma após o tratamento

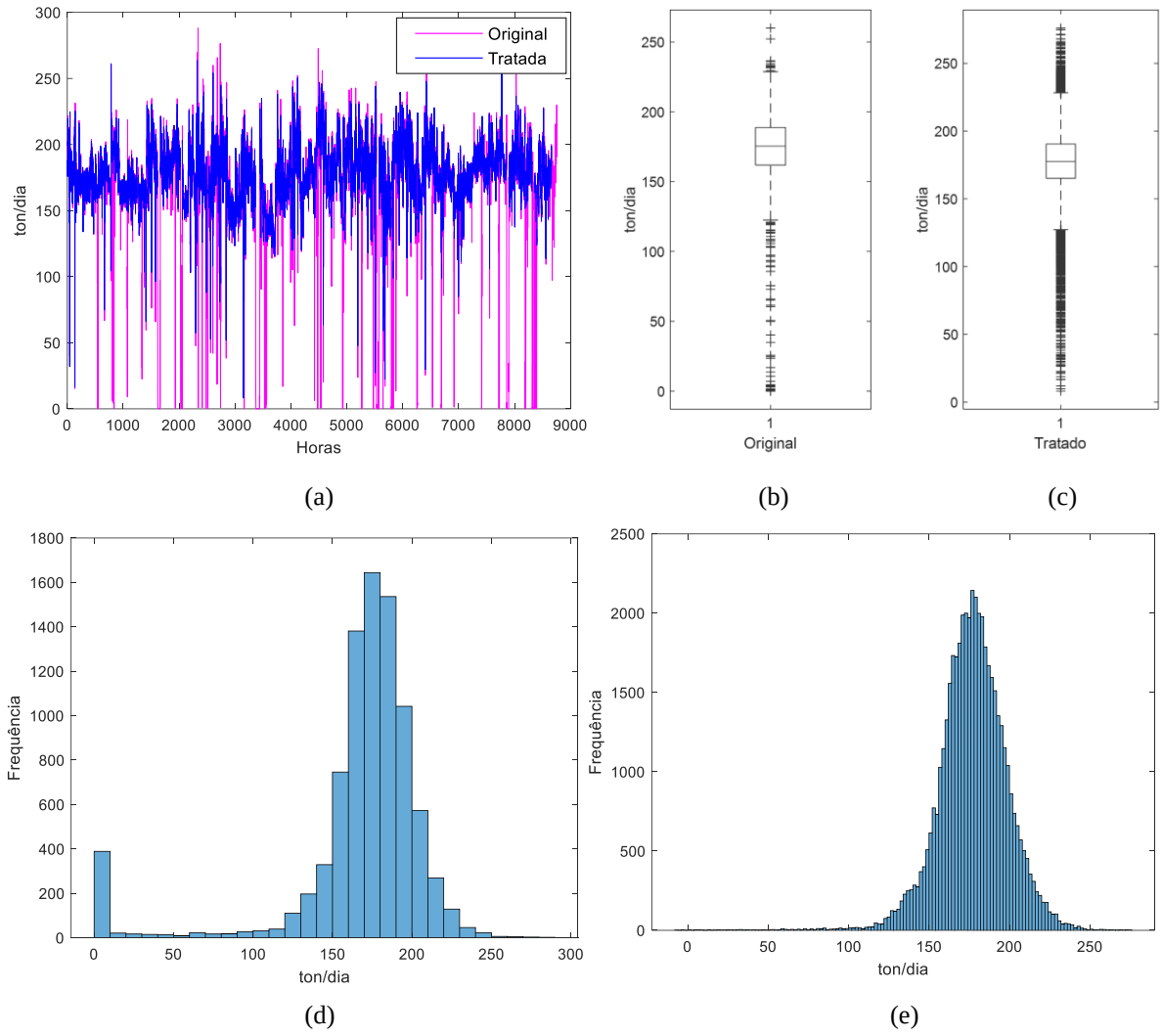


(a) (b) (c)

(d) (e)

Fonte: Produção do Próprio autor.

Gráfico 53 – Geração de finos de retorno (a) Gráfico de linhas (b) Boxplot antes do tratamento (c) Boxplot após tratamento (d) Histograma antes do tratamento (e) Histograma após o tratamento

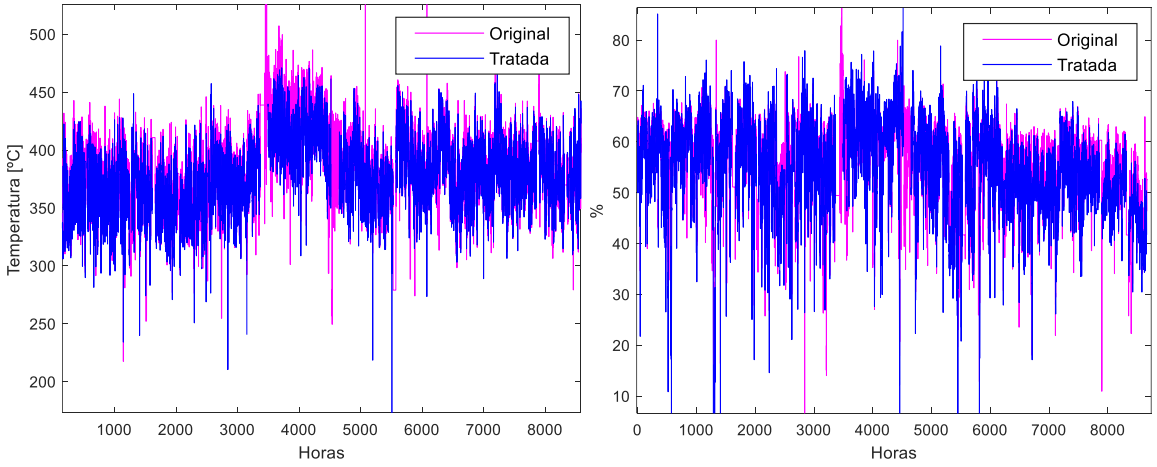


Fonte: Produção do Próprio autor.

11.3 Temperatura e Posição do BTP

Por meio do Gráfico 54 (a) e (b) são plotados os valores originais da temperatura e posição do BTP e os valores dessas mesmas variáveis baseadas no cálculo apresentado na seção 9.3 com base nos dados de temperatura já tratados.

Gráfico 54 – Gráfico de linhas comparativo (a) Temperatura do BTP (b) Posição do BTP

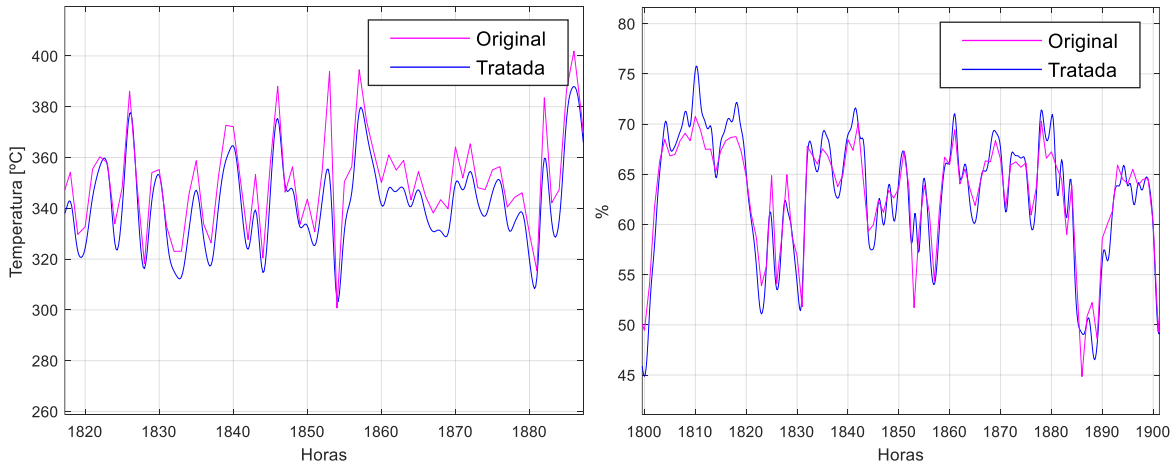


(a) (b)

Fonte: Produção do Próprio autor.

Para se observar como se comporta as variáveis de posição e temperatura em relação aos dados originais, são plotados, por meio do Gráfico 55 (a) e (b), algumas amostras de cada uma das variáveis.

Gráfico 55 – Amostras comparativas do BTP (a) Temperatura (b) Posição



(a) (b)

Fonte: Produção do Próprio autor.

12 CONCLUSÃO E SUGESTÕES PARA TRABALHOS FUTUROS

O presente trabalho apresentou um estudo conceitual acerca do processo de sinterização. Abordando a composição da mistura de entrada, os equipamentos e comportamento do material ao longo da esteira de sinterização até o início do processo de resfriamento e emissões associadas.

Utilizou-se de variáveis fornecidas por engenheiros responsáveis pelo processo, julgadas como potencialmente influenciáveis a dinâmica do BTP. Entretanto, análises semelhantes, dispostas na Seção 0, utilizaram de variáveis conceitualmente influenciáveis no BTP, que não foram disponibilizadas pelo detentor da planta.

Para uma análise mais aprofundada, considera-se necessário o fornecimento da técnica aplicada pelo detentor da planta destinado ao cálculo do BTP, pois esse fato permitiria uma análise mais apropriada para se chegar a uma consolidação final da medição do parâmetro. É necessário também ser fornecida a forma de validação do cálculo desse parâmetro e precisão associada, pois, ao que se constatou ao longo do trabalho, a estimativa dessa variável é realizada por meio de uma aproximação quadrática da curva, metodologia que carrega consigo uma adição de erro considerável e diminui a precisão de medição desse parâmetro de alta significância ao funcionamento adequado da planta. Além disso, em alguns momentos, observou-se que o parâmetro não foi registrado de acordo com a realidade da planta. A mesma necessidade se aplica ao cálculo da Velocidade de Sinterização, que não é fornecido pelo detentor da planta.

Foi analisada a qualidade dos dados fornecidos para realização de análises do comportamento do BTP ao longo do tempo. Concluiu-se, por meio de análises do processo e das variações bruscas observadas nas variáveis fornecidas, que a amostragem dos dados era insuficiente para realização de uma análise que necessita de precisão associada. De forma a contornar essa problemática, para fins acadêmicos, os dados foram reamostrados, entretanto, sugere-se, para posteriores análises da mesma planta, a utilização de medições de amostras com maior recorrência para refletir uma proximidade a realidade do processo.

Identificou-se instantes de tempo que devem ser desconsiderados de análises dos dados por constituírem momentos de parada do processo de sinterização para manutenções de rotina.

Foram removidos *outliers* que correspondiam a falhas nos sensores presentes na planta e foram mantidos dados que apresentam variações significativas, que eram dados válidos e retratavam ocorrências reais na planta. Foram aplicadas técnicas de interpolação dos dados por meio do polinômio interpolador PCHIP e por meio de redes neurais autorregressivas.

A MODWT foi aplicada como alternativa de filtro aos ruídos de alta frequência associados aos sinais lidos durante o processo. Foram evidenciadas as relações lineares entre as variáveis do processo, durante a realização de análise fatorial, e foi proposta uma nova variável de entrada das redes neurais que possam contribuir para o seu desempenho.

Por fim, sugere-se que, para obter a estimativa do BTP ao longo do processo, o objetivo alvo seja a análise da variação da curva de temperatura ao longo do tempo. Deve-se sugerir, além disso, alternativas para a metodologia utilizada para o cálculo desse parâmetro, já que ele é associado a uma carga significativa de imprecisão e valores registrados erroneamente, conforme foi possível observar na seção 9.3.

REFERÊNCIAS BIBLIOGRÁFICAS

ARAÚJO, L. A. **Siderurgia**. 1. ed. São Paulo: FTA, 1967.

ARCELORMITTAL. **Quem somos**. [20--?]. Disponível em: https://tubarao.arcelormittal.com/quem-somos/usina/fluxo-producao/index.asp. Acesso em: 20 ago. 2019.

BATINI, C.; SCANNAPIECO, M. **Data and Information Quality**: Dimensions, Principles and Techniques. 1. ed. Suíça: Springer Nature, 2016.

BIANCHI, M. F. **Extração de características de imagens de faces humanas através de wavelets, PCA e IMPCA**. 2006. Dissertação (Mestrado em Engenharia Elétrica) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2006. Disponível em: https://teses.usp.br/teses/disponiveis/18/18133/tde10072006002119/publico/Capitulo_4_mestrado.pdf. Acesso em: 29 out. 2019.

BRAGA, A. P.; LUDEMIR, T. B.; CARVALHO, A. C. P. L. F. **Redes Neurais Artificiais**: Teoria e Aplicações. Rio de Janeiro: LTC, 2007.

CUNHA, C. F. F. C. **Aplicação da Transformada Wavelet na Redução de Ruídos em Medições de Descargas Parciais**. 2013. Dissertação (Mestrado em Engenharia Elétrica) – Programa de Pós-Graduação em Engenharia Elétrica (COPPE), Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2013.

DIAS, V. O. C. **Wavelets e uma aplicação na Espectometria**. 2017. Dissertação (Mestrado em Matemática) – Departamento de Matemática, Universidade Federal do Espírito Santo, Vitória, 2017. Disponível em: http://portais4.ufes.br/posgrad/teses/tese_10994_DissertacaoViniciusOliari.pdf. Acesso em: 10 out. 2019.

DINIZ, A. P. M. **Modelos de previsão do conteúdo de silício no ferro gusa usando redes neurais artificiais**. 2018. Dissertação (Mestrado em Engenharia Elétrica) – Programa de Pós-Graduação em Engenharia Elétrica (PPGEE), Centro Tecnológico da Universidade Federal do Espírito Santo, Vitória, 2018. Disponível em: http://portais4.ufes.br/posgrad/teses/tese_10911_DISSERTA%C7%C3O%20-%20ANA%20PAULA%20MIRANDA%20DINIZ.pdf. Acesso em: 3 out. 2019.

FÁVERO, L. P.; BELFIORE, P.; SILVA F. L.; CHAN, B. L. **Análise de dados**: modelagem multivariada para tomada de decisões. 1. ed. Rio de Janeiro: Elsevier, 2009.

FÁVERO, L. P.; BELFIORE, P. **Análise de dados**: Técnicas Multivariadas Exploratórias com SPSS® e STATA®. 1. ed. Rio de Janeiro: Elsevier, 2017.

FILHO; J. C. F. **Sistema para determinação de penalidades para desvios de especificações em contratos de compra de carvões siderúrgicos**. 2007. Dissertação (Mestrado em Engenharia Industrial) – Programa de Pós-Graduação em Engenharia Industrial, Departamento de Engenharia Industrial, Pontífica Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2007. Disponível em: www2.dbd.puc-rio.br/pergamum/tesesabertas/0511101_07_pretextual.pdf. Acesso em: 20 ago. 2019.

GALIAZZI C. V. R.; SANTOS, E. T. B.; SILVEIRA, W. L. G.; RIBEIRO, L. H. C.; SILVA, R. O.; SILVA, B. P. Controle da qualidade física do sínter para Alto-Forno. *In*: SEMINÁRIO DE REDUÇÃO DE MINÉRIO DE FERRO E MATÉRIAS-PRIMAS, 44., Simpósio Brasileiro de Minério de Ferro, 15. e Simpósio Brasileiro de Aglomeração de Minério de Ferro, 2., 2014, Belo-Horizonte. **Anais [...]**. Belo-Horizonte: 2014. p. 234-240. Disponível em: https://abmproceedings.com.br/ptbr/article/controle-da-qualidade-fsica-do-sinter-para-alto-forno. Acesso em: 25 ago. 2019.

GEERDES, M.; TOXOPEUS, H.; VLIET, C.V. D. **Modern Blast Furnace Ironmaking an Introduction**. 2. ed. Amsterdam: IOS Press Ebooks, 2009.

GUTIERREZ, C. E. C. **Eliminação do Ruído por Encolhimento de Wavelets**. 2002. Dissertação (Mestrado em Engenharia Elétrica) – Programa de Pós-Graduação em Sistemas de Energia, Departamento de Engenharia Elétrica, Pontífica Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2002. Disponível em: www2.dbd.pucrio.br/pergamum/tesesabertas/ 5000045931_02_pretexto.pdf. Acesso em: 29 out. 2019.

HAIR, J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E; TATHAM, R. L. **Análise multivariada de dados.** 6. ed. Porto Alegre: Bookman, 2009

HAYKIN, S. **Redes neurais**: princípios e prática. 2. ed. Porto Alegre: Bookman, 2001.

INSTITUTO FEDERAL DO ESPÍRITO SANTO. **Preparação de matérias-primas**. [20--?]. Material do curso de Pós-Graduação Lato Sensu em Engenharia Metalúrgica com Ênfase em Siderurgia – Instituto Federal do Espírito Santo, Vitória, [20--?].

ISO. **ISO/IEC 25012:2008 Software engineering, Software product Quality Requirements and Evaluation (SQUARE)**: Data quality model. Disponível em: https://www.iso.org/standard/35736.html. Acesso em: 22 ago. 2019a.

_____. **ISO/IEC 25012:2008: System-Dependent Data Quality.** Disponível em: https://iso25000.com/index.php/en/iso-25000-standards/iso-25012?limit=5&start=15. Acesso em: 23 ago. 2019b.

KIM, B; JEONG, J. W.; HWANG, K.; LEE, J.; DONCHANG, H.; PARK, K. Estimation of burn-through point in the sinter process. *In*: INTERNATIONAL CONFERENCE ON CONTROL, AUTOMATION AND SYSTEMS (ICCAS 2014), 14., 2014, Seoul. **Anais […]**. Seoul: 2014. p. 1531-1533.

LI, M. H.; WANG, J. The Research for Soft Measuring Technique of Sintering Burning Through Point. *In:* IEEE CONFERENCE ON INDUSTRIAL ELECTRONICS AND APPLICATIONS, 1., 2006, Singapore. **Anais […]**. Singapore: 2006. p. 1-4. Disponível em: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4025857&isnumber=4025694. Acesso em: 20 out. 2019.

MAHANTI, R. **Data Quality**: Dimensions, Measurement, Strategy, Management, and Governance. [eBook Kindle]. Milwaukee, Wisconsin: ASQ Quality Press, 2019.

MALLAT, S. G. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 11, n. 7, p. 674-693, jul. 1989. Disponível em: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=192463&isnumber=4963. Acesso em: 1 nov. 2019.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, v. 5, n. 4, p. 115-133, 1943.

MISITI, M.; MISITI, Y.; OPPENHEIM, G.; POGGI, J. M. **Wavelet Toolbox User's Guide**. 1. Ed. Natick: The MathWorks, 2017.

MOURÃO, M. B.; YOKOJI, A.; MALYNOWSKYJ, A.; LEANDRO, C. A. S.; TAKANO, C.; QUITES, E. E. C.; GENTILE, F. E.; SILVA, G. F. B. L.; BOLOTA, J. R., GONÇALVES, M.; FACO, R. J. **Introdução à Siderurgia**. São Paulo: Associação Brasileira de Metalurgia e Materiais, 2007.

OLIVEIRA, H. M. D. **Análise de Fourier e Wavelets**: sinais estacionários e não estacionários. 1. ed. Recife: Editora Universitária da UFPE, 2007.

PERCIVAL, D. B.; WALDEN, A. T. **Wavelet Methods for Time Series Analysis**. 1. ed. Cambridge: Cambridge University Press, 2000.

REIS, I. A. **Uso de redes neurais recorrentes para Estimação do BTP em um processo de Sinterização**. 2018. Trabalho de Conclusão de Curso (Graduação em Engenharia Elétrica) – Centro Tecnológico, Universidade Federal do Espírito Santo, Vitória, 2018.

RIZZO, E. M. S. **Operador de sinterização**. Vitória: [s.n], 2007.

ROSSI, R. G. **Análise de Componentes Principais em Data Warehouses**. 2017. Dissertação (Mestrado em Ciência da Computação) – Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2017.

SAIZ, J.; POSADA, M. J. Non-linear State Estimator for the On-line Control of a Sinter Plant. **ISIJ International**, v. 53, n. 9, p. 1658-1664, 2013. Disponível em: ttps://www.jstage.jst.go.jp/article/isijinternational/53/9/53_1658/_html/-char/en. Acesso em: 28 out. 2019.

SANTOS, J. C. M. **Redução de ruído em sinais de voz combinando filtro de Kalman e Transformada Wavelet**. 2015. Dissertação (Mestrado em Engenharia Elétrica) – Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal de Uberlândia, Uberlândia, 2015. Disponível em: https://repositorio.ufu.br/bitstream/123456789/14611/1/ReducaoRuidoSinais.pdf. Acesso em: 20 nov. 2019.

SCHIAVONI, A. S. **Um Estudo Comparativo de Métodos para Balanceamento do Conjunto de Treinamento em Aprendizado de Redes Neurais Artificiais**. 2010. Trabalho de Conclusão de Curso (Graduação em Ciência da Computação) – Departamento de Ciência da Computação, Universidade Federal de Lavras, Lavras, 2010.

SONG, Q.; WANG A.; ZHANG Y. The Combination Prediction of BTP in Sintering Process based on Bayesian Framework and LS-SVM. **Telkomnika Indonesian Journal of Electrical Engineering**, v. 11, n. 8, p. 4616-4626, ago. 2013.

TÁPIA, M. **Redes Neurais Artificiais**: Uma Aplicação na Previsão de Preços de Ovos. 2000. Dissertação (Mestrado em Ciências da Computação) – Programa de Pós-Graduação em Ciências da Computação, Universidade Federal de Santa Catarina, Florianópolis, 2000.

TELLES, V. B. **Reciclagem da poeira de aciaria elétrica na sinterização de minério de ferro visando a eliminação de zinco**. 2010. Dissertação (Mestrado em Engenharia Metalúrgica e de Materiais) – Departamento de Engenharia Metalúrgica e de Materiais, Escola Politécnica de São Paulo, São Paulo, 2010.

USAMENTIAGA, R.; MOLLEDA, J.; GARCIA, D. F.; BULNES F. G. Monitoring Sintering Burn-Through Point Using Infrared Thermography. **Sensors**, Basel, v. 13, n. 8, p. 10287-10305, ago. 2013. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3812604/. Acesso em: 30 out. 2019.

WANG, B.; FANG, Y; SHENG, J; GUI, W.; SUN, Y. BTP Prediction Model Based on ANN and Regression Analysis. *In:* INTERNATIONAL WORKSHOP ON KNOWLEDGE DISCOVERY AND DATA MINING, 2., 2009, Moscow. **Anais [...]**. Moscow: 2009. p. 108-111. Disponível em: https://ieeexplore.ieee.org/document/4771890/authors#authors. Acesso em: 30 out. 2019.

WANG, J; LI, X; LI, Y; WANG, K. BTP prediction of sintering process by using multiple models. *In*: THE CHINESE CONTROL AND DECISION CONFERENCE, 26., 2014, Changsha. **Anais [...].** Changsha: 2014. p. 4008-4012. Disponível em: https://ieeexplore.ieee.org/document/6852882/authors#authors. Acesso em: 1 nov. 2019.