



Técnicas de *Machine Learning* aplicadas ao monitoramento de partículas em caldeira de recuperação Kraft

Elisa Carneiro Werneck do Carmo

Projeto de Final de Curso

Orientadores

Prof. Maurício Bezerra de Souza Jr., D.Sc.

Prof. Andréa Pereira Parente, M.Sc.

Agosto de 2017

TÉCNICAS DE *MACHINE LEARNING* APLICADAS AO MONITORAMENTO DE PARTÍCULAS EM CALDEIRA DE RECUPERAÇÃO KRAFT

Elisa Carneiro Werneck do Carmo

Projeto de Final de Curso submetido ao Corpo Docente da Escola de Química, como parte dos requisitos necessários à obtenção do grau de Engenheiro Químico.

Aprovado por:

Flavio Waltz, M.Sc.

Heloísa Lajas Sanches, D.Sc.

Elene de Souza Freitas, M.Sc.

Orientador por:

Mauricio Bezerra de Souza Jr., D.Sc.

Andréa Pereira Parente, M.Sc.

Rio de Janeiro, RJ – Brasil

Agosto de 2017

Carmo, Elisa Carneiro Werneck do.

Técnicas de *Machine Learning* aplicadas ao monitoramento de partículas em caldeira de recuperação Kraft / Elisa Carneiro Werneck do Carmo – Rio de Janeiro: UFRJ/EQ. 2017.

XII, 124 p. ; il.

(Projeto de Final de Curso) – Universidade Federal do Rio de Janeiro, Escola de Química, 2017.

Orientadores: Mauricio Bezerra de Souza Jr. e Andréa Pereira Parente

1. Indústria de Papel e Celulose. 2. Redes Neurais Artificiais. 3. Caldeira de Recuperação Kraft. 4. Projeto Final (Graduação – UFRJ/EQ). 5. Mauricio Bezerra de Souza Jr. e Andréa Pereira Parente. I. Técnicas de *Machine Learning* aplicadas ao monitoramento de partículas em caldeira de recuperação Kraft.

Dedicatória

À minha mãe, minha melhor amiga e ao meu querido avô Mariano, que me acompanha e me protege em todos os momentos de minha vida.

“I don’t pretend we have all the answers. But the questions are certainly worth thinking about.”

Arthur C. Clarke

AGRADECIMENTOS

Agradeço imensamente ao meu orientador, professor Maurício, e à minha co-orientadora, professora Andréa, pela paciência, disponibilidade e confiança durante todo o projeto. Foi um prazer trabalhar e, principalmente, aprender com vocês. Agradeço, também, à Andréa Costa pela disponibilização dos dados utilizados neste estudo, à Elene e ao Cauê pela importante ajuda e contribuição para o projeto.

Obrigada, também, à Escola de Química e à Universidade Federal do Rio de Janeiro por todo o aprendizado e pela formação de excelência oferecida. Agradeço, também, à École Centrale Marseille, pela incrível oportunidade da Dupla Diplomação.

Agradeço aos meus pais, Ana Lucia e Roberto, meus primeiros professores e futuros colegas de profissão. Vocês que me ensinaram o prazer da descoberta e compartilharam comigo o amor pela Engenharia Química. Devo minhas conquistas à dedicação e ao carinho de vocês. Mãe, seu companheirismo e apoio foram essenciais para minha caminhada até aqui, me dando forças para superar todos os desafios. Meu muito obrigada à minha amada família, pela dedicação e torcida, em especial às minhas “segundas mães”, Avany e Maria de Jesus, e ao nosso pequeno Vicente.

Agradeço imensamente a todos os professores que contribuíram de forma única para meu crescimento profissional e, muitas vezes, pessoal, em especial aos professores Armando, Carlos André, Charlls, Elaine, Heloísa, Marcelo, Nei e Ofélia, da UFRJ; M. Ibaseta, Mme Loubet, Mme Duprat e M. Denis, da École Centrale Marseille.

Agradeço aos queridos colegas com quem tive o prazer de trabalhar durante meus estágios e a todos das equipes YUKON Automação Industrial, Sanofi Vertolay e PPGN. Todo o aprendizado foi essencial para meu crescimento pessoal e profissional.

Gostaria, também, de agradecer aos meus amigos e às muitas “famílias” que a vida me deu: à “família” inaciana, que terá sempre um lugar especial em meu coração; à “família” EQ. Julia, Thais, Sabrina, Lucchesi, Diego, Luyne, Raísa e Mariana, obrigada por todos os ensinamentos, risadas e carinho; às “famílias” *marseillaise* e *vichyssoise*, sem as quais a experiência do Duplo Diploma não teria sido tão surpreendente. Agradeço, especialmente, à Ana Beatriz, minha grande companheira, ao Felipe, meu eterno *coloc* e ao Daniel, Pedro, Raísa e Roberto, pelo carinho, aprendizado, paciência e amizade.

E claro, agradeço ao meu companheiro Adrien, que foi essencial para a minha jornada até aqui. Obrigada por todo o apoio, por todas as palavras de carinho que me reconfortaram e pelos sermões, que me fizeram encarar de frente todos os desafios.

Resumo do Projeto de Final de Curso apresentado à Escola de Química como parte dos requisitos necessários para obtenção do grau de Engenheiro Químico.

TÉCNICAS DE *MACHINE LEARNING* APLICADAS AO MONITORAMENTO DE PARTÍCULAS EM CALDEIRA DE RECUPERAÇÃO KRAFT

Elisa Carneiro Werneck do Carmo

Agosto, 2017

Orientadores: Prof. Mauricio Bezerra de Souza Jr., D.Sc.
Prof. Andréa Pereira Parente, M.Sc.

A caldeira de recuperação é um equipamento chave nas plantas de produção de papel e celulose pelo processo Kraft. A deposição de material particulado formado na caldeira representa um dos maiores problemas operacionais, devendo ser monitorada e controlada para prevenção de paradas operacionais não programadas. O processo de formação de partículas neste equipamento é complexo e modelos fenomenológicos são de difícil desenvolvimento. Uma alternativa promissora, que se insere no contexto da Indústria 4.0, é a aplicação de técnicas de aprendizado de máquinas, ou “*Machine Learning*”, para modelagem de tal fenômeno. O objetivo do presente trabalho foi, então, propor uma metodologia para o emprego de redes neurais artificiais na modelagem da formação de material particulado em uma caldeira de recuperação Kraft. Foram utilizadas redes neurais dos tipos Multicamadas Perceptron (MLP), Base Radial (RBF) e de mapeamento auto-organizável para o desenvolvimento de modelos preditivos e classificatórios a partir de dados históricos de um ano de operação em uma planta industrial. Foi possível obter uma modelagem satisfatória do número de partículas formadas na caldeira de recuperação Kraft a partir de um modelo “híbrido” de predição e classificação, que realiza a classificação dos valores preditos por uma rede preditiva em duas classes. A classe 1 engloba dados de operação normal, caracterizada por um número de partículas inferior a 200 partículas por minuto, enquanto a classe 2 agrupa os dados de operação anormal e potencialmente insegura, com emissão acima de 200 partículas por minuto. O melhor modelo encontrado se baseia em uma rede MLP de 3 camadas, com 11 neurônios na camada oculta, utilizando conjuntos adicionais de dados obtidos por adição de ruído gaussiano aos dados originais correspondentes à operação anormal. Esta adição foi realizada para equilibrar a distribuição dos dados operacionais disponíveis. O desempenho da rede de predição, dado pelo coeficiente de correlação entre os valores calculados pela rede e os valores reais, foi superior a 0,87 e o desempenho total de classificação foi de cerca de 87%. Este método é capaz de classificar corretamente aproximadamente 94% dos dados de classe 1 e 77% dos dados de classe 2. Foi obtido, também, um mapa auto-organizável a partir uma rede de mapeamento auto-organizável de topologia 20x5 que permite a identificação de regiões de operação normal, regiões de transição e regiões de maior probabilidade de operação anormal. Este método pode ser usado na geração de recomendações referentes à segurança operacional, sinalizando condições operacionais com maior potencial de risco de elevada formação de partículas.

ÍNDICE

ÍNDICE DE FIGURAS	ix
ÍNDICE DE TABELAS	xi
NOMENCLATURA	xii
1. INTRODUÇÃO	1
2. OBJETIVOS	5
3. A INDÚSTRIA DE PAPEL E CELULOSE	6
3.1 Matéria-Prima – Fontes Celulósicas	7
3.2 Processos e Etapas da Produção de Papel	10
3.2.1 Principais Processos de Deslignificação da Madeira	11
3.2.2 Produção da Polpa de Celulose pelo Processo Kraft	13
3.2.3 Produção de Papel a Partir da Polpa Celulósica	18
3.3 A Caldeira de Recuperação Kraft	20
3.3.1 Descrição Física da Caldeira de Recuperação	22
3.3.2 Processo de Recuperação: Transformações físico-químicas do licor negro nas diferentes zonas da fornalha	28
3.3.3 Formação de Material Particulado na Caldeira de Recuperação Kraft	33
3.4 Parâmetros de Interesse para Planta de Produção de Papel e Celulose com Ênfase no Processo Kraft	38
3.4.1 Modelagem dos Parâmetros	39
4. REDES NEURONAIIS ARTIFICIAIS	42
4.1 As Redes Neurais Artificiais	42
4.2 Histórico	43
4.3 Arquitetura das Redes Neurais Artificiais	45
4.3.1 Principais Redes Neurais Artificiais	48
4.4 Função de Ativação	51
4.5 Treinamento	54
4.5.1 Método de retropropagação (<i>backpropagation</i>)	55
4.5.2 Método de Treinamento Kohonen	59
4.6 Teste e Validação	61
4.7 Aplicações	62
5. DADOS E MÉTODOS	64
5.1 Dados Operacionais da Caldeira de Recuperação Kraft	64
5.2 Software STATISTICA®	70

5.2.1	STATISTICA® Automated Neural Networks (SANN)	70
5.2.2	Escolha das Entradas Relevantes a Partir da Análise de Sensibilidade 71	
5.2.3	Determinação do Erro e da Performance das Redes.....	71
5.3	Predição do Número de Partículas a Partir de Redes de Regressão	73
5.4	Classificação do Número de Partículas	74
5.5	Análise de <i>Cluster</i>	78
6.	RESULTADOS.....	82
6.1	Predição	82
6.1.1	Redes MLP	82
6.1.2	Redes RBF	86
6.2	Redes de Classificação	88
6.3	Abordagem “híbrida”: Predição e Classificação	91
6.4	Análise Cluster – Mapas Auto-Organizáveis	101
7.	CONSIDERAÇÕES FINAIS	107
8.	REFERÊNCIAS BIBLIOGRÁFICAS	111
	APÊNDICES	119

ÍNDICE DE FIGURAS

Figura 1: Estrutura da celulose. FONTE: PANDEY, 1999.....	8
Figura 2: Estrutura da hemicelulose. FONTE: PANDEY, 1999.....	8
Figura 3: Estrutura da lignina originária de eucaliptos. FONTE: CORREIA, 2016.	9
Figura 4: Estrutura química da lignina presente em madeiras de fibra longa. FONTE: PANDEY, 1999.....	10
Figura 5: Reações envolvidas na degradação da lignina no Processo Kraft. FONTE: CORREIA, 2016.....	13
Figura 6: Diagrama de blocos da produção de polpa de celulose pelo processo Kraft. FONTE: CORREIA, 2016.....	14
Figura 7: Funcionamento de um digestor contínuo do tipo Kamyr. CORREIA, 2016. .	15
Figura 8: Esquema e seções de uma máquina de papel. FONTE: RENOVA, 2000.	19
Figura 9: Representação da natureza cíclica do processo Kraft, indicando a recuperação de reagentes. FONTE: COSTA, 2004.	22
Figura 10: Caldeira de Recuperação Kraft na planta Celulosa Arauco Y Constitución, Nueva Aldea Mill in Chile (Cortesia de Metso Power). Fonte: PRASZKIER, 2011. ...	23
Figura 11: Caldeira de Recuperação Kraft dividida em duas seções. Fonte: PAOLIELLO et al., 2002.	23
Figura 12: Interior da fornalha de uma caldeira de recuperação da Andritz com os tubos de aço-carbono. Fonte: VAKKILAINEN, 2005.....	26
Figura 13: Corte transversal da parede da caldeira. Fonte: PRASZKIER, 2011.....	27
Figura 14: Representação esquemática das três zonas da fornalha de uma caldeira de recuperação Kraft. FONTE: COSTA, 2004.	29
Figura 15: Variação do tamanho da gotícula/partícula do licor negro após entrada na caldeira de recuperação. FONTE: HOLMBLAD, 2015.....	31
Figura 16: Formação de depósitos de material particulado nos superaquecedores de uma caldeira de recuperação Kraft. FONTE: VAKKILAINEN, 2005.	37
Figura 17: Composição dos depósitos de material particulado. FONTE: TRAN, 2007.	38
Figura 18: Esquema de um neurônio biológico. FONTE: CASTRO (2014).	45
Figura 19: Modelo do j-ésimo neurônio artificial de uma rede multicamadas. Adaptado de DE SOUZA Jr., 2005.....	46
Figura 20: Esquema de fluxo de informações em uma rede <i>feedforward</i> . FONTE: BAUGHMAN e LIU, 1995.	48
Figura 21: Esquema de uma rede RBF. FONTE: STATISTICA, 2004.....	50
Figura 22: Estrutura de uma rede de mapeamento auto-organizável, em que os neurônios de saída estão organizados em uma grade bidimensional. FONTE: ALVAREZ e LUQUE, 2003.	51
Figura 23: Função logística com fator p igual a 3, 1,5 e 0,8. FONTE: CARDON e MÜLLER, 1994.....	52
Figura 24: Função tangente hiperbólica com fator p de 2,2, 1, e 0,6. FONTE: CARDON e MÜLLER, 1994.	53
Figura 25: Representação gráfica ilustrativa do momento de interrupção do treinamento que corresponde ao início do aumento do erro de teste.....	62
Figura 26: Evolução de Epart durante o período de aquisição de dados.	66

Figura 27: Definição das classes estabelecidas e estudadas na classificação do número de partículas.	75
Figura 28: Representação das topologias das redes testadas.	78
Figura 29: Mapa auto-organizável com 100 neurônios dispostos em uma matriz 20x5, indicando regiões de grande incidência de casos de classe 2.	80
Figura 30: Comparação entre os valores preditos pela melhor rede neuronal e os valores reais para cada conjunto de dados.	85
Figura 31: Comparação entre os valores de Epart preditos e os reais para todos os dados.	86
Figura 32: Relação entre o número de alarmes falsos e os índices de desempenho local para uma razão n_2/n_1 igual a 0,062 (curva azul) e para valores iguais de n_1 e n_2 (curva verde).	95
Figura 33: Influência do número de repetições no desempenho da rede.	97
Figura 34: Desempenhos geral e individual da classe 2 com o aumento do número de repetições.	100
Figura 35: Efeito do equilíbrio da distribuição dos dados em duas classes no desempenho de classificação da classe 2.	101
Figura 36: Rede 5x8.	102
Figura 37: Rede 10x10.	103
Figura 38: Resumo dos resultados da primeira etapa da análise de clusters.	103
Figura 39: Mapa referente à rede 20x5.	104
Figura 40: Comparação da rede de topologia 20x5 para análise em 2, 3, 4, 5 e 7 classes.	105

ÍNDICE DE TABELAS

Tabela 1: Estágios de branqueamento com seus respectivos símbolos e produto químico utilizado. FONTE: CASTRO, 2009.	18
Tabela 2: Compostos presentes no licor negro.	21
Tabela 3: Composição química de um licor negro. FONTE: FREDERICK, 1997.	21
Tabela 4: Principais etapas da queima do licor negro para uma gotícula de 2 mm. Fonte: VAKKILAINEN, 2006.	31
Tabela 5: Principais reações e mudanças de estado na fornalha. FONTES: GRACE et al., 1998 e COSTA, 2004, JARVINEN et al., 2002.	32
Tabela 6: Variáveis de processo empregadas. Adaptado de COSTA, 2004.	65
Tabela 7: Estatísticas das variáveis de entrada.	67
Tabela 8: Variação das variáveis operacionais normalizadas.	68
Tabela 9: Variação das variáveis de entrada para os casos em que E_{part} é superior a 200.	69
Tabela 10: Parâmetros iniciais de geração de redes.	74
Tabela 11: Parâmetros iniciais de geração de redes de classificação.	76
Tabela 12: Parâmetros utilizados para a geração de redes de mapeamento auto-organizável.	79
Tabela 13: Resultados do teste inicial com redes MLP para predição.	82
Tabela 14: Resultados do segundo teste com redes MLP para predição.	83
Tabela 15: Resultados do terceiro teste com redes MLP para predição.	84
Tabela 16: Resultados do quarto teste com redes MLP para predição.	84
Tabela 17: Resultados para melhor rede MLP para predição.	85
Tabela 18: Resultados do primeiro teste para a melhor rede RBF para predição.	87
Tabela 19: Resultados para melhor rede RBF para predição obtida.	88
Tabela 20: Melhores redes MLP para classificações em 7,5,4,3 e 2 classes.	89
Tabela 21: Melhores redes RBF para classificações em 7,5,4,3 e 2 classes.	89
Tabela 22: Melhores redes MLP para classificações em 7,5,4,3 e 2 classes.	90
Tabela 23: Melhores redes RBF para classificações em 7,5,4,3 e 2 classes.	90
Tabela 24: Resultados classificação após predição do número de partículas.	92
Tabela 25: Comparação entre os desempenhos e porcentagens de alarmes falsos dos métodos de classificação tratados neste estudo.	94
Tabela 26: Desempenhos das redes de predição de acordo com o número de repetições.	97
Tabela 27: Comparação do desempenho de classificação de cada classe para divisão dos dados em 7 classes de acordo com o número de repetições.	98
Tabela 28: Comparação do desempenho de classificação de cada classe para divisão dos dados em 5 classes de acordo com o número de repetições.	98
Tabela 29: Comparação do desempenho de classificação de cada classe para divisão dos dados em 4 classes de acordo com o número de repetições.	99
Tabela 30: Comparação do desempenho de classificação de cada classe para divisão dos dados em 3 classes de acordo com o número de repetições.	99
Tabela 31: Comparação do desempenho de classificação de cada classe para divisão dos dados em 2 classes de acordo com o número de repetições.	100

NOMENCLATURA

LETRAS LATINAS

Símbolo	Descrição	Dimensão
E	Função erro para treinamento	-
Epart	Número de partículas por minuto formadas na caldeira de recuperação	partículas por minuto
f(.)	Função de ativação de um neurônio	-
L	Camada de saída da rede	-
n _h	Número de neurônios na camada oculta	-
n _i	Número de neurônios na camada de entrada	-
n' _i	Número de conjuntos de dados da classe i	-
n _k	Número de neurônios na camada k	-
n _L	Número de neurônios na camada de saída	-
N _p	Número de padrões para treinamento	-
p _{af}	Porcentagem de alarmes falsos	
p _i	Percentual de acerto da classe i	
Par1	Pressão de alimentação do ar primário.	mmca
Par2	Pressão de alimentação do ar secundário.	mmca
Pl	Pressão de alimentação do licor negro à caldeira.	kgf/cm ²
p _r	Desempenho geral da rede	
s _{pl,k}	Saída do neurônio i da camada k, para o padrão p	-
Tar1	Temperatura de alimentação do ar primário.	°C
Tar2	Temperatura de alimentação do ar secundário.	°C
Tlu	Temperatura de alimentação do licor negro.	°C
t _{pl}	Saída desejada do neurônio l da camada de saída, para o padrão p	-
Var1	Vazão de ar primário alimentado à caldeira.	t/h
Var2	Vazão de ar secundário alimentado à caldeira.	t/h
Var3	Vazão de ar terciário alimentado à caldeira.	t/h
Vvlu	Vazão volumétrica de licor negro alimentado à caldeira.	m ³ /h
w _{jik}	Peso entre o neurônio j da camada (k+1) e o neurônio i da camada anterior	-
x	Concentração de sólidos na alimentação do licor negro à caldeira.	%

LETRAS GREGAS

Símbolo	Descrição	Dimensão
α	Constante real positiva	-
ε	Constante de momento	-
η	Taxa de aprendizagem	-
θ	Bias do neurônio j da camada (k+1)	-
λ _{pj,k+1}	Soma ponderada das entradas de um neurônio j da camada (k+1), adicionada ao bias, para um padrão p	-
μ _{ji}	Centros da função de ativação Gaussiana para um neurônio j	-
σ	Desvio-padrão da função de ativação Gaussiana	-

1. INTRODUÇÃO

Os processos industriais estão em constante evolução, sobretudo com o desenvolvimento tecnológico cada vez mais intenso. A Humanidade já presenciou três revoluções industriais e, atualmente, a necessidade das indústrias de aumentar a sua produtividade, dinamizar a sua produção, adaptando-as às exigências dos clientes e ao aumento de sua competitividade está levando à quarta revolução, chamada de Indústria 4.0.

Em retrospectiva, a primeira revolução industrial, iniciada na Inglaterra em 1782, foi marcada pela invenção de máquinas a vapor e teve como marco a mecanização do processo de produção (DOPICO et al., 2016). A segunda revolução, vivenciada no início do século XX, baseou-se nos conceitos de racionalização, na padronização, na divisão de tarefas e na linha de montagem com uso de energia elétrica e com produção em massa. Já a terceira revolução industrial se caracterizou pelo desenvolvimento de computadores e da tecnologia da informação (TI), que permitiu o uso de máquinas controladas e automatizadas, flexibilizando os sistemas e aumentando ainda mais a produtividade. No entanto, a automação dos processos na terceira revolução industrial se limita a situações conhecidas e de certa forma repetitivas e mensuráveis. Deste modo, ainda há necessidade de intervenção humana, uma vez que os sistemas, em geral, não são capazes de aprender e agir de forma autônoma em situações inusitadas (DOPICO et al., 2016).

A necessidade de encontrar soluções ótimas para os problemas operacionais ainda existentes levou a uma nova revolução. A quarta revolução industrial, ou Indústria 4.0, como será tratada no presente estudo, se baseia no conceito de fábricas inteligentes (do inglês, *smart factories*) com elevada automação, eficiência e integração.

A Indústria 4.0 engloba tecnologias dotadas da capacidade de interagir com o ambiente em seu entorno e de resolver problemas, buscando soluções e tomando decisões de forma autônoma. Este avanço tecnológico se expande por toda a cadeia produtiva e inclui os sensores e atuadores de campo, as máquinas e os sistemas de TI, de automação e administrativos (RÜSSMANN et al., 2015), formando um conjunto de sistemas automatizados, otimizados e integrados.

A Indústria 4.0 está fundamentada em diversos conceitos que são essenciais para compreender a sua complexidade e os potenciais desta nova revolução industrial, destacando-se os seguintes: “*Big Data*”, Internet das coisas (IoT, em inglês, *Internet of Things*), armazenamento em nuvem, segurança cibernética, sistemas físico-cibernéticos e

técnicas de aprendizado de máquina (do inglês, “*Machine Learning*”) (RÜSSMANN et al., 2015; WANG, 2016; DOPICO et al., 2016).

O conceito de “*Big Data*” não se limita a um grande conjunto de dados gerados durante a produção. Ele inclui a ideia de que esses dados podem ser analisados de forma a descobrir padrões, regras e causas de falhas (WANG, 2016) para otimizar a produção, reduzir gastos energéticos e melhorar os serviços e a qualidade do produto final, ajudando nos processos de tomada de decisão.

Outro conceito importante e complementar é o de “*Machine Learning*” que envolve técnicas de inteligência artificial baseadas no aprendizado a partir de exemplos, que são usadas no reconhecimento de padrões. Desta forma, máquinas são capazes de aprender, de se adaptar e de agir de forma autônoma a partir dos dados de processo aos quais elas têm acesso. O “*Machine Learning*” é uma solução interessante para tarefas complexas demais para programas computacionais ou técnicas convencionais. Um exemplo é a modelagem de sistemas pouco conhecidos e com dinâmicas complexas ou processos produtivos que apresentam grandes variações (NILSSON et al., 2005).

O presente estudo se insere, justamente, no contexto da Indústria 4.0 e propõe o uso de técnicas de “*Machine Learning*” na busca de soluções inovadoras para um dos desafios enfrentados pela indústria de papel e celulose: o monitoramento da formação de partículas em uma caldeira de recuperação.

A indústria de papel e celulose apresenta uma grande importância no cenário econômico brasileiro. O Brasil é o quarto maior produtor mundial de celulose e o nono maior produtor de papel (IBÁ, 2016), sendo o processo Kraft o mais empregado na produção de polpa celulósica (COSTA, 2004). Neste processo, a caldeira de recuperação é um equipamento fundamental por ser responsável pela geração de vapor e energia elétrica para a planta a partir da queima do licor negro, assim como pela recuperação de reagentes não consumidos na digestão, o que permite a redução de rejeitos e, conseqüentemente, de impactos ambientais. O monitoramento e o controle de sua operação são essenciais para a indústria de papel e celulose e representam um desafio com grande potencial para estudos de modelagem dos processos físicos e químicos que nela ocorrem.

Durante o processo de queima de licor negro nas caldeiras de recuperação ocorre a formação de material particulado, que pode ser arrastado pelos gases de exaustão e depositado nas superfícies de troca térmica. Esta deposição de partículas provoca perda de eficiência de troca térmica e até mesmo bloqueio do escoamento dos gases, tornando

necessária a realização de paradas operacionais não programadas: um dos maiores problemas operacionais para a indústria de papel e celulose. Desta forma, a quantidade de partículas formadas na caldeira é um parâmetro que auxilia na avaliação da operação deste equipamento e o seu controle tem o potencial de reduzir a frequência de paradas para manutenção.

A natureza complexa da formação de material particulado na caldeira dificulta o desenvolvimento de modelos matemáticos convencionais a partir de métodos analíticos e fenomenológicos, tornando interessante, como alternativa, o uso de técnicas de Inteligência Artificial e de “*Machine Learning*” para a abordagem deste problema.

As redes neurais artificiais (RNA), ou em inglês *Artificial Neuronal Networks* (ANN), são técnicas empíricas de processamento de informação que possibilitam a geração de modelos de sistemas complexos a partir da identificação e generalização de padrões encontrados em um determinado conjunto de dados. Elas são dotadas da capacidade de aprendizagem através de exemplos, fazendo parte das técnicas de Inteligência Artificial e de “*Machine Learning*”.

O objetivo do presente trabalho é propor uma metodologia para a aplicação de redes neurais na modelagem da formação de partículas em uma caldeira de recuperação Kraft a partir de dados históricos operacionais de uma planta de produção de polpa de celulose. Foram adotadas três abordagens para a modelagem do problema:

- A predição do número de partículas formadas na caldeira empregando redes neurais preditivas;
- A classificação das condições operacionais em diferentes números de classes pré-estabelecidas em função da emissão de partículas;
- A análise por meio de mapas auto-organizáveis.

Testou-se o uso de redes do tipo Multicamadas Perceptron (MLP), Base Radial (RBF) e rede de mapeamento auto-organizável.

Este trabalho está organizado em sete capítulos, além da bibliografia e dos apêndices. A motivação e a contextualização da problemática, assim como os principais objetivos do estudo, são apresentados nos capítulos 1 e 2.

A importância da Indústria de Papel e Celulose no cenário brasileiro e suas principais etapas e processos são expostos no capítulo 3, no qual também são descritos os principais aspectos da caldeira de recuperação Kraft, detalhando o seu funcionamento e os processos físico-químicos que nela ocorrem, com ênfase na formação de material particulado.

No capítulo 4 são descritos os fundamentos teóricos sobre Redes Neurais Artificiais necessários para o desenvolvimento e a boa compreensão do presente estudo. São abordados os principais tipos de redes, suas características e suas aplicações, bem como os principais métodos de aprendizagem das redes.

A metodologia empregada, detalhando as diferentes abordagens utilizadas para a resolução do problema proposto, assim como as principais funcionalidades e aspectos do software usado, é apresentada no capítulo 5.

No capítulo 6 são discutidos os resultados obtidos e no capítulo 7 são apresentadas as considerações finais e as sugestões decorrentes do projeto.

2. OBJETIVOS

O objetivo geral do presente estudo é propor uma metodologia com base em técnicas de “*Machine Learning*”, mais especificamente a aplicação de redes neurais, para modelagem da formação de material particulado em uma caldeira industrial de recuperação Kraft. A modelagem da formação de tais partículas para fins de monitoramento e controle da caldeira permite que o processo seja operado em faixas seguras, reduzindo a frequência das paradas decorrentes de incrustações e bloqueios, levando assim a uma melhora do desempenho global do equipamento.

Os objetivos específicos do presente estudo são listados a seguir:

- Compreensão dos principais processos e etapas envolvidas na produção de papel e celulose, com ênfase na caldeira de recuperação Kraft e na formação de material particulado neste equipamento;
- Compreensão e estudo das principais técnicas de predição e classificação utilizando redes neurais artificiais e entendimento do seu potencial para aplicações à Engenharia de Processos;
- Elaboração de modelos de predição e classificação para o número de partículas produzidas na fornalha;
- Desenvolvimento de modelos para predição de condições anormais e indesejadas de operação da caldeira;
- Avaliação e comparação dos modelos propostos e análise da representatividade dos dados;
- Comparação dos resultados obtidos com estudos similares, na tentativa de aprimorar modelagens propostas anteriormente.

3. A INDÚSTRIA DE PAPEL E CELULOSE

A necessidade do Homem de transcrever suas ideias, feitos e descobertas foi a força motriz para o desenvolvimento do papel. Sua história remonta ao Egito Antigo, quando o papiro, uma planta abundante no delta do Nilo, foi utilizado como matéria-prima para produção das folhas de papiro, precursoras do papel.

O papel surgiu na China no ano 105 D.C. Inicialmente, ele era produzido a partir de uma pasta de bambu e amoreira obtida com instrumentos rudimentares como martelos de madeira e almofariz de pedras (COSTA, 2004). Em 751 D.C., o papel começou a ser produzido também na Índia e foi, então, introduzido no ocidente pelos árabes (LIMA et al., 1988).

Em 1144, surgiram as primeiras fábricas europeias em Valência e o processo de fabricação de papel se tornou mais barato pela utilização de trapos de tecidos como matéria-prima. Devido ao crescimento da demanda, houve a necessidade da exploração de outras fontes mais abundantes e economicamente interessantes, sendo atualmente a madeira a principal matéria-prima utilizada. A fabricação de papel permaneceu durante muitos anos um processo artesanal (RENOVA, 2000).

No Brasil, o primeiro papel foi produzido em 1809, no Rio de Janeiro, e, posteriormente, outras três fábricas foram instaladas em 1837, 1841 e 1852 (ROBERT, 2007).

A celulose, intermediário da fabricação do papel, possui diversas aplicações além da fabricação de papel, tais como a fabricação de nitrato de celulose, de plásticos e produção de acetato de celulose para a indústria têxtil (MORRISON e BOYD, 1992). A indústria, hoje chamada de papel e celulose, ganhou importância no cenário brasileiro. Atualmente, o setor conta com 222 empresas em 18 estados brasileiros, sendo as regiões Sul e Sudeste as principais produtoras, de acordo com o Centro de Gestão e Estudos Estratégicos (CGEE) (2013). Dentre as principais empresas do setor, podem ser destacadas: Fibria, Suzano Papel e Celulose, Klabin SA, Celulose Nipo-Brasileira SA (Cenibra), International Paper do Brasil Ltda., Grupo Orsa, Rigesa Celulose e Papel e Embalagens Ltda. (CENTRO DE GESTÃO E ESTUDOS ESTRATÉGICOS, 2013).

O Brasil é o quarto maior produtor mundial de celulose, tendo produzido 17,4 milhões de toneladas em 2015 e o nono maior produtor de papel, com uma produção de 10,4 milhões de toneladas em 2015 (IBÁ, 2016). Outros países líderes no mercado de papel e celulose são a Finlândia, a Suécia, o Canadá e os Estados Unidos da América.

Segundo o Centro de Gestão e Estudos Estratégicos (2013), o Brasil exportou em 2012 mais da metade (61,4%) de sua produção de celulose e 18,2% de sua produção total de papel. Os principais destinos das exportações brasileiras são, respectivamente a Europa, a China e a América do Norte (CGEE, 2013). O mercado mundial de celulose vem se mostrando favorável e em crescimento e, em 2016, as produções brasileiras de celulose aumentaram, em relação ao ano anterior, 8,1%. Já a produção de papel está estável e teve um pequeno decréscimo de 0,2% no ano de 2016 (IBÁ, 2016).

3.1 Matéria-Prima – Fontes Celulósicas

A pasta de celulose, necessária para a fabricação de papel, pode ser obtida de diferentes fontes celulósicas, classificadas de acordo com o tamanho de suas fibras. A primeira categoria, utilizada em pequena escala para produção industrial de papel, é formada por fibras muito longas chamadas de fibras têxteis e são provenientes de vegetais, tais como o algodão, o linho e até mesmo o bambu e o bagaço de cana.

A segunda categoria compreende as madeiras, que são as principais fontes para fabricação de papel. No Brasil, as principais matérias-primas utilizadas na fabricação de papel são o eucalipto e o pinus (COSTA, 2004).

As madeiras se subdividem em duas outras categorias: conífera e folhosa. As madeiras coníferas, também chamadas de madeira de fibra longa, madeira macia ou de baixa densidade (*softwood*, em inglês), apresentam fibras um pouco menores que as fibras têxteis, em geral, de 2,0 a 3,5 milímetros de comprimento, e são majoritariamente produzidas nos Estados Unidos da América, no Canadá, no Chile e em países escandinavos. Exemplos típicos são o pinho, a araucária e o cipreste. Já as madeiras folhosas, ou de fibra curta (*hardwood*, em inglês), apresentam fibras com 0,5 a 2,0 milímetros de comprimento e agrupam árvores como o eucalipto, o álamo, a bétula e o carvalho, entre outras. O Brasil possui uma importante participação na produção mundial deste tipo de madeira junto com a Indonésia, o Chile, Portugal, a Espanha e o Uruguai.

A madeira é uma fonte lignocelulósica constituída, basicamente, de celulose, lignina, hemicelulose e uma fração extrativa compreendendo resinas, pectinas, entre outros. Há ainda a presença de água, compostos inorgânicos, tais como fosfatos, sulfatos e compostos de potássio e de magnésio.

A Celulose

A celulose, principal componente da parede celular de um vegetal, é um biopolímero (polissacarídeo) linear de glicose, cristalino e insolúvel em água. A celulose, cuja estrutura química é mostrada na Figura 1, compreende longas cadeias poliméricas que se dispõem umas ao lado das outras em feixes e são mantidas posicionadas através de pontes de hidrogênio entre os numerosos grupos -OH adjacentes. Estes feixes formam estruturas entrelaçadas, que se agrupam e formam as fibrilas e as fibras que podem ser vistas a olho nu. Na madeira, estas fibrilas de celulose estão aglutinadas por lignina (MORRISON e BOYD, 1972).

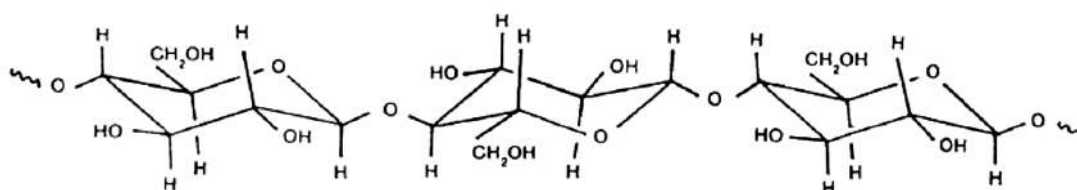


Figura 1: Estrutura da celulose. FONTE: PANDEY, 1999.

A Hemicelulose

A hemicelulose é um polímero formado de diferentes unidades de açúcares, tais como D-xilose, D-manose, D-glicose, D-galactose e substâncias pécticas. Assim como a celulose, a hemicelulose mostrada na Figura 2, é um polissacarídeo, porém de cadeias ramificadas e de tamanhos menores. Ao contrário da celulose, a hemicelulose é amorfa e não forma arranjos fibrosos (CASTRO, 2009).

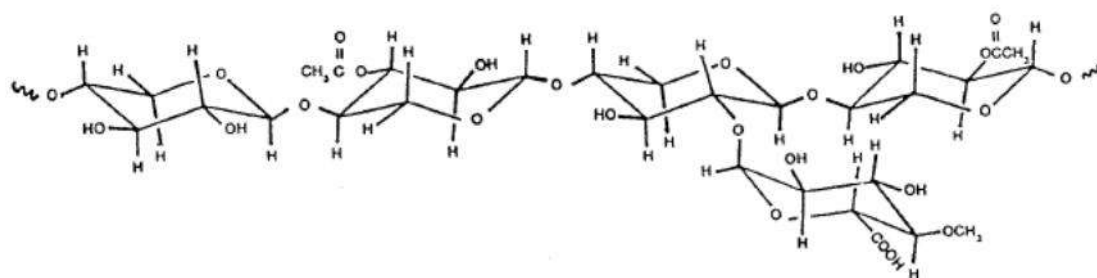


Figura 2: Estrutura da hemicelulose. FONTE: PANDEY, 1999.

A Lignina

A lignina, presente, principalmente, na parede celular de células vegetais, é constituída de moléculas de elevado grau de aromaticidade. A lignina é um conjunto

amorfo de polifenóis e tem diversas funções na estrutura da planta tais como a rigidez da parede celular, a união das células umas com as outras, a redução da permeabilidade à água e a proteção da madeira contra micro-organismos, uma vez que sua estrutura essencialmente fenólica age como um fungicida (MORRISON e BOYD, 1972).

Como a lignina pode sofrer oxidação, ela pode causar o escurecimento do produto final caso esteja presente na polpa de celulose usada na fabricação de papel. Segundo Correia (2016), madeiras folhosas apresentam menor teor de lignina que as coníferas e a quantidade dos grupos funcionais presentes na estrutura química da lignina varia de acordo com a espécie da planta.

Estruturas moleculares da lignina foram propostas por diversos pesquisadores como Freudenberg (1968), Adler (1977), Sakakibara (1980) e Nimz (1974), conforme indicado por Davin et al. (2008). A Figura 3 apresenta a estrutura da lignina encontrada em madeiras de eucalipto e a Figura 4 mostra a estrutura da lignina, segundo Adler (1977), para madeiras de fibra longa.

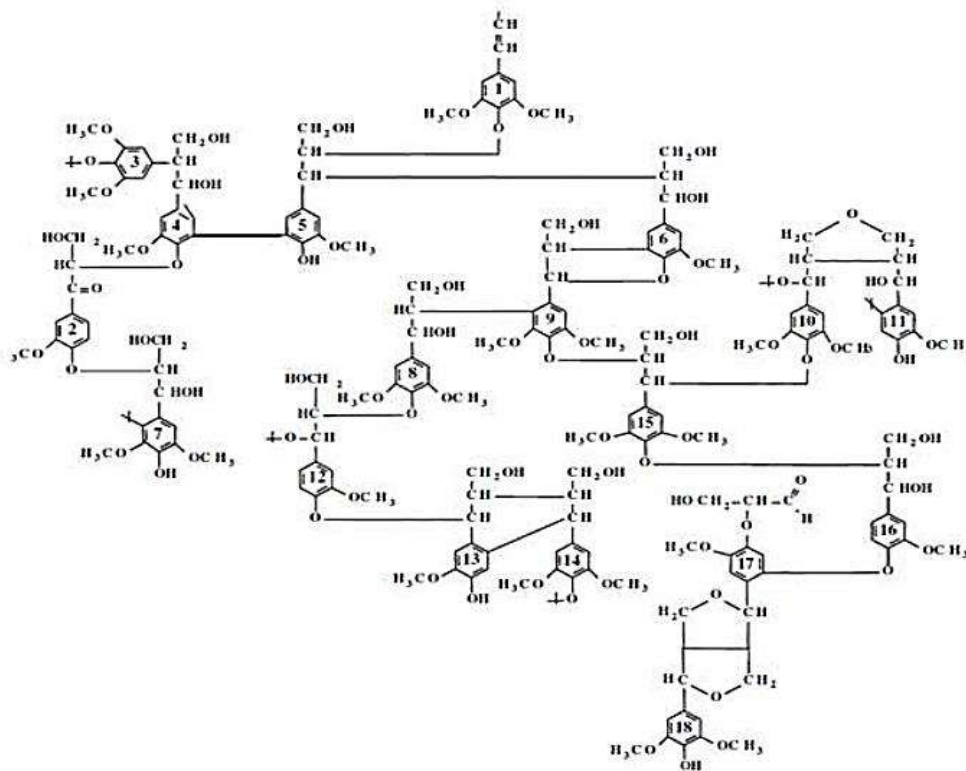


Figura 3: Estrutura da lignina originária de eucaliptos. FONTE: CORREIA, 2016.

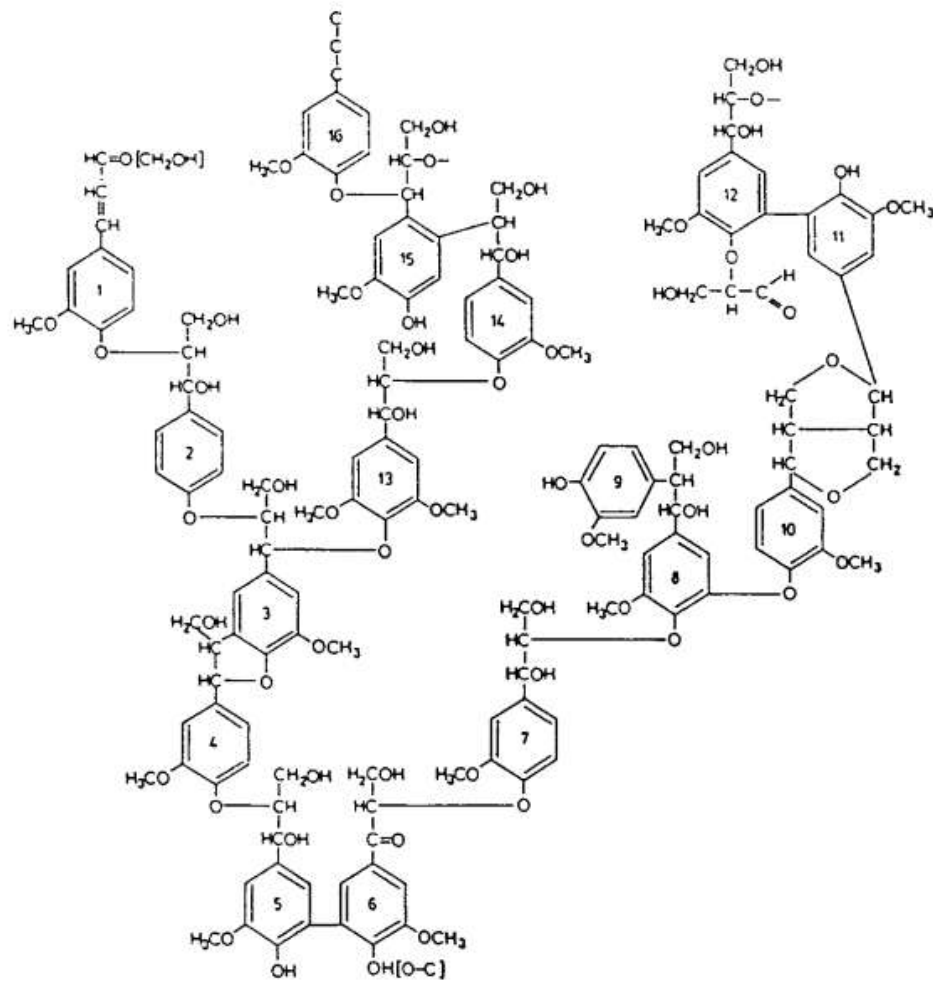


Figura 4: Estrutura química da lignina presente em madeiras de fibra longa.
 FONTE: PANDEY, 1999.

3.2 Processos e Etapas da Produção de Papel

O processo de obtenção de papel compreende duas etapas essenciais: a produção da polpa de celulose e a fabricação do papel propriamente dito.

Na fabricação da polpa de celulose, a lignina e a fração extrativa devem ser retiradas da matéria-prima celulósica, quebrando-se a estrutura da parede celular. Este processo é chamado de deslignificação e pode ser realizado de forma mecânica, física, biotecnológica ou química.

A produção do papel em si envolve a mistura da polpa de fibras de celulose com aditivos que conferem propriedades aos diferentes tipos de papel, tais como opacidade, resistência mecânica e à umidade, cor, colagem, impermeabilização, brilho, etc.

3.2.1 Principais Processos de Deslignificação da Madeira

O processo de fabricação de papel envolve uma série de etapas de processos químicos e físicos, sendo que a etapa principal consiste na formação de uma pasta de celulose, obtida a partir da quebra da estrutura celular da madeira e posterior dissolução e separação da lignina. A lignina é responsável pela cor da pasta e sua remoção é muito importante para se alcançar a pureza e a alvura desejadas para o papel final.

O Processo Mecânico

Nos processos mecânicos de deslignificação, não há tratamento químico da madeira (COSTA, 2004). A madeira é umedecida e, em seguida, prensada em rolos giratórios de material abrasivo ou processada em moinhos de bola. Apesar da obtenção de altos rendimentos, que podem chegar a 98% (RENOVA, 2000), a lignina e os demais componentes não-celulósicos da madeira não são completamente separados, possibilitando, assim, a posterior oxidação da lignina, assim como a degradação dos componentes não-celulósicos. A presença de tais componentes provoca o escurecimento da pasta celulósica e, conseqüentemente, do papel assim produzido. Esse processo é mais utilizado para papéis que exigem um menor grau de alvura e uma menor vida útil, tais como jornais e revistas (RENOVA, 2000).

Processos Físicos e Biotecnológicos

Os processos físicos e biotecnológicos não são ainda usados industrialmente, seja por inviabilidade operacional ou econômica. O processo físico se baseia, por exemplo, no uso de irradiação por raios gama, e apresenta como principal inconveniente a degradação oxidativa da própria celulose. Já o processo biotecnológico ainda está em fase de desenvolvimento, mas tem o potencial de reduzir os impactos ambientais dos rejeitos da fabricação de papel. O princípio explorado é a produção de enzimas ligninases por micro-organismos para degradação biológica da lignina (CASTRO, 2009).

Processos Termomecânicos

Há ainda os processos termomecânicos que se dividem em dois grupos: o processo por tratamento a vapor e o processo por explosão com vapor, ambos com elevada demanda energética. No processo por tratamento a vapor, o objetivo é aquecer a madeira com uso de vapor até temperaturas entre 150°C e 250°C (CASTRO, 2009) para o rompimento por hidrólise da estrutura da parede celular, provocando uma mudança do

estado rígido para o plástico (RENOVA, 2000) e provocar o aumento dos poros. No entanto, a celulose pode ser removida parcialmente pela ocorrência de reações indesejáveis. Já no processo por explosão a vapor, é realizado um aquecimento a temperaturas entre 180°C e 250°C, durante aproximadamente 20 minutos. Em seguida, o material sofre uma brusca queda de pressão para evaporação da água intracelular, causando uma quebra da estrutura da parede celular e permitindo a retirada por solubilização da lignina em soluções de hidróxido de sódio (CASTRO, 2009).

Processos Químicos

Dentre os processos químicos mais empregados industrialmente, destacam-se os chamados Processo Kraft (ou Sulfato), Processo Soda e Processo Sulfito. O princípio básico desses processos é a digestão, ou seja, o cozimento da matéria-prima celulósica sob pressão com auxílio de reagentes químicos.

No processo Soda, emprega-se o hidróxido de sódio como reagente digestor, aproveitando a elevada solubilidade da lignina (pelo seu caráter fenólico). Já no processo Sulfito, a digestão é realizada com sulfitos alcalinos como agentes químicos e pode apresentar rendimentos entre 40 e 60% (RENOVA, 2000). Apesar de o branqueamento da polpa de celulose obtida ser considerado fácil, este processo vem sendo substituído pelo processo Kraft, no qual a digestão é feita com sulfeto de sódio e hidróxido de sódio, por causa da complexidade do tratamento de resíduos e da regeneração dos reagentes.

Embora a maior parte da literatura defina o processo químico de deslignificação que emprega a mistura de hidróxido de sódio e sulfeto de sódio como sendo o processo Sulfato ou Kraft, alguns autores preferem denominar de forma diferente e estabelecem que, dependendo das concentrações desses reagentes, têm-se o processo Sulfato e o processo Kraft, conforme, por exemplo, em RENNOVA (2000). No entanto, neste estudo será adotada a abordagem de CASTRO (2009) e os dois processos serão tratados conjuntamente.

Segundo Costa (2004), o processo Kraft é o mais utilizado desde a década de 1930. No Brasil, em 2009, já correspondia a, aproximadamente, 81% dos processos industriais de deslignificação da madeira CASTRO (2009).

3.2.2 Produção da Polpa de Celulose pelo Processo Kraft

Nessa seção será descrita a fabricação de polpa de celulose a partir de madeiras pelo processo Kraft, devido à sua importância no cenário brasileiro.

A digestão sob pressão é realizada com sulfeto de sódio e hidróxido de sódio e permite uma boa dissolução da lignina e a preservação da resistência das fibras de celulose, produzindo uma pasta celulósica concentrada e branqueável. O rendimento pode ser de 50% a 60% segundo RENOVA (2000).

Nesse processo, os reagentes penetram e se difundem na estrutura celular da madeira, quebrando as pontes de hidrogênio e inchando a madeira. A lignina é degradada pelos reagentes químicos, formando fragmentos solúveis, pela hidrólise de ligações do tipo carbono-oxigênio das funções éter. Além da digestão da lignina, outras reações podem ocorrer, tais como a dissolução indesejável dos carboidratos (hemicelulose e, em menor grau, celulose), o que reduz o rendimento do processo pela degradação da própria celulose, e as reações envolvendo resinas e ácidos orgânicos presentes na madeira (CORREIA, 2016).

A Figura 5 abaixo mostra as reações que ocorrem na degradação da lignina no processo Kraft. Em meio básico e na presença de sulfeto (SH^-), um grupamento $-\text{OR}_2$ (com R_2 sendo H ou um grupamento alquil) é retirado da estrutura da lignina e, em seguida, há um rearranjo formando uma ponte de sulfeto e liberando um grupo fenólico.

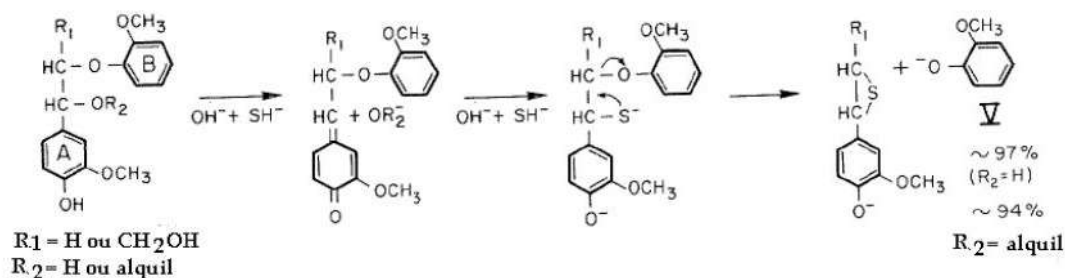


Figura 5: Reações envolvidas na degradação da lignina no Processo Kraft.
FONTE: CORREIA, 2016.

Podem ser destacadas como principais vantagens do processo Kraft, o menor tempo de residência na etapa de cozimento, a flexibilidade no processamento de madeiras de diferentes tipos, a manutenção da resistência da fibra de celulose, a possibilidade de obtenção de grande alvura final da pasta após branqueamento e a possibilidade de recuperação dos reagentes. Esse último fator é interessante, uma vez que

aproximadamente 50% dos resíduos sólidos podem ser recuperados (COSTA, 2004). Por outro lado, o elevado custo de branqueamento e de investimento inicial para construção das unidades fabris, além do desprendimento de gases com forte odor, configuram inconvenientes importantes desse processo.

A produção da polpa de celulose pelo processo Kraft pode ser dividida em dois ciclos principais: a deslignificação da madeira e a recuperação de reagentes. A Figura 6 mostra as principais etapas da produção de polpa de celulose, que serão detalhadas a seguir.

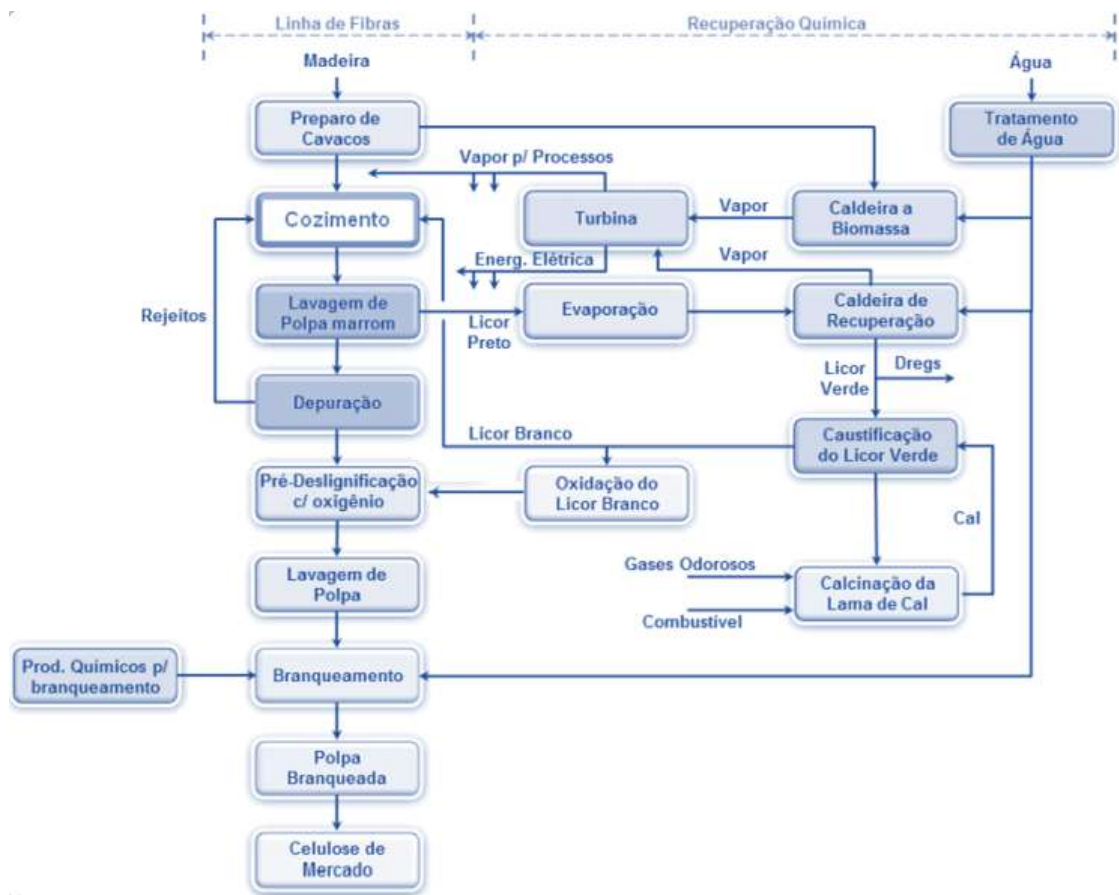


Figura 6: Diagrama de blocos da produção de polpa de celulose pelo processo Kraft. FONTE: CORREIA, 2016.

Primeiramente, a casca da madeira é retirada em descascadores rotativos, já que sua remoção facilita os processos subsequentes de peneiração e lavagem. A casca, que apresenta baixo teor de fibras, é queimada em caldeiras para geração de vapor. A madeira segue para picadores de disco ou tambor (CASTRO, 2009), onde é fragmentada em cavacos para que haja maior superfície de contato com os reagentes durante a digestão. Antes da estocagem do material ocorre uma etapa de peneiramento, que é necessária para seleção dos cavacos com tamanhos adequados. Aqueles que se encontram fora da

especificação podem ser reprocessados na unidade de picagem ou, quando muito finos, podem ser queimados na caldeira de geração de vapor (caldeira de biomassa).

Os cavacos de tamanho adequado são submetidos a um pré-tratamento a vapor para aquecimento e expulsão do ar presente nos poros, com o objetivo de facilitar a difusão dos reagentes. Em seguida, são alimentados em vasos de pressão ou digestores de cozimento mantidos sob pressão, onde são adicionados os reagentes. O processo de digestão pode ser operado em regime contínuo ou em batelada e pode atingir temperaturas de até 160°C (CORREIA, 2016), pois a reação é exotérmica.

As principais vantagens dos digestores contínuos consistem em seu tamanho e gasto energético reduzidos e o baixo custo de instalação. O digestor pode ser dividido em três zonas. Na primeira, chamada de zona de impregnação, ocorre a penetração e difusão dos reagentes. Na segunda zona, ocorre o cozimento propriamente dito da madeira, ou seja, a deslignificação (CORREIA, 2016). Por fim, há uma zona de lavagem onde a lignina residual é retirada. A Figura 7 ilustra o funcionamento de um digestor contínuo do tipo Kamyr.

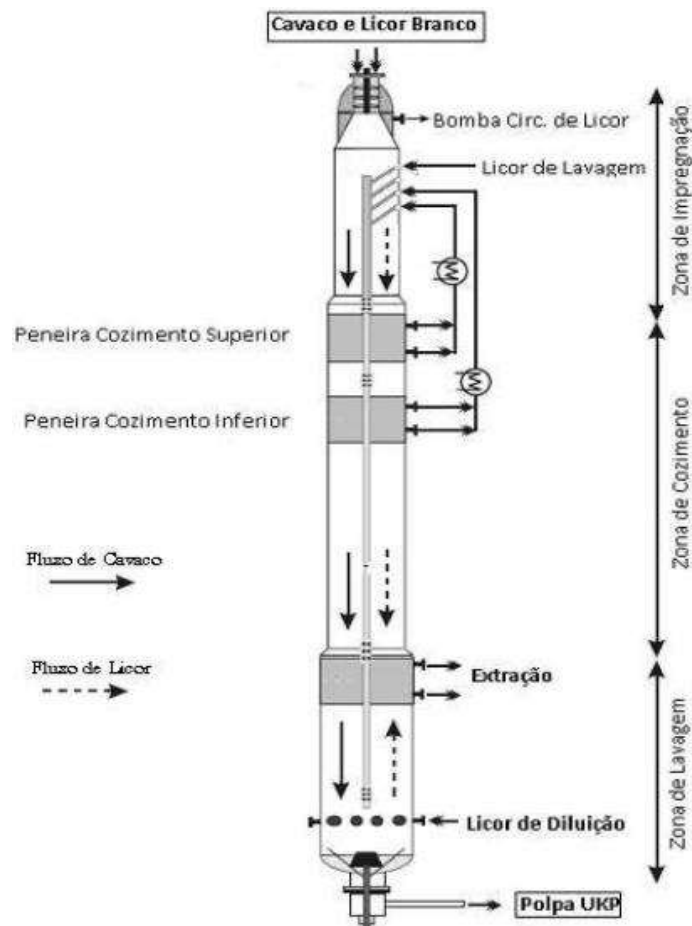


Figura 7: Funcionamento de um digestor contínuo do tipo Kamyr. CORREIA, 2016.

A mistura de reagentes alimentada no digestor recebe o nome de licor branco que, no caso do processo Kraft, é composto por sulfeto de sódio, hidróxido de sódio e outros sais de sódio em menores quantidades. Ao final do cozimento, tem-se uma polpa marrom que passará por diversas etapas de processamento. As principais correntes obtidas serão a polpa celulósica lavada que seguirá para a etapa de branqueamento, e o licor negro, que será concentrado por evaporação e alimentado na caldeira de regeneração Kraft. O licor negro é constituído por reagentes não consumidos, lignina retirada da madeira, matéria orgânica derivada da degradação da lignina e outros elementos residuais do cozimento. O alto poder calorífico da lignina justifica o seu uso como combustível.

As principais variáveis envolvidas no processo de cozimento são a espécie, a densidade e o tempo de estocagem da madeira, o tamanho dos cavacos e as condições operacionais, tais como pressão, temperatura, tempo de residência e concentração dos reagentes. O controle dessa etapa é fundamentalmente baseado no teor de lignina. Para tal controle, o procedimento mais empregado atualmente é a retirada de uma amostra da massa reacional do digestor para análise química em laboratório para a determinação do número Kappa, que é diretamente proporcional à quantidade de lignina presente na mistura reacional e que representa uma quantificação do grau de deslignificação da matéria-prima usada. Este parâmetro será tratado em maiores detalhes na seção 3.4.

Após o cozimento da madeira no digestor, a corrente de saída, constituída pelo licor negro e pela pasta de celulose, passa por uma etapa de separação mecânica dos cavacos residuais e por uma etapa de lavagem em filtros lavadores, formando o licor negro fraco ou diluído, ou seja, um licor contendo materiais solúveis presentes nas fibras e reagentes não consumidos.

O licor negro fraco, recuperado na lavagem, segue para uma etapa de recuperação dos reagentes. Inicialmente, o licor negro fraco, cuja composição é de aproximadamente 16% (mássico) em sólidos, é concentrado em uma série de evaporadores até atingir 70% a 85% em sólidos (CORREIA, 2016). Nesta etapa remove-se o “*tall oil*” (ácidos graxos e ácidos resinosos). A etapa de concentração é importante, uma vez que a alimentação de licores negros com elevados teores de água na caldeira não é recomendada, pois pode provocar explosões devido ao contato da água com materiais fundidos (em inglês, *smelt*¹), conforme mencionado por COSTA (2004).

¹ Smelt: Sais fundidos resultantes da queima do licor negro na caldeira de recuperação. Formam uma fase líquida contínua e não costumam apresentar material carbonáceo.

O licor negro concentrado, chamado de licor forte, é pré-aquecido a 115°C e alimentado em uma caldeira de recuperação. O pré-aquecimento do licor negro concentrado antes de sua alimentação na caldeira de recuperação é importante para que se possa diminuir a sua viscosidade (COSTA, 2004).

Na caldeira, a matéria orgânica, como, por exemplo, a lignina, é queimada para geração de vapor para uso na planta enquanto os sais inorgânicos são recuperados para reciclo. A operação da caldeira de recuperação é complexa e será tratada em detalhes na seção 3.3. Os reagentes químicos são fundidos na caldeira e recuperados, sendo conduzidos a um tanque de dissolução para obtenção do licor verde. O licor verde, rico em carbonato de sódio e sulfeto de sódio (CASTRO, 2009), é levado para caustificadores. Nos caustificadores, há adição de cal, que reage com a água formando hidróxido de cálcio, que, por sua vez, reage com o carbonato de sódio presente no meio e forma hidróxido de sódio. Posteriormente, em um sedimentador, o carbonato de cálcio e a cal residuais são recuperados. O carbonato é alimentado ao forno, onde é transformado em cal, que retorna ao caustificador, de acordo com a reação 1.



O líquido resultante da etapa de caustificação é o próprio licor branco, composto de hidróxido de sódio e sulfeto de sódio, que é bombeado para a etapa de digestão. Evidentemente, há necessidade de se adicionar uma corrente de *make-up* de sulfeto de sódio no processo de digestão para repor o que foi consumido na reação.

A polpa celulósica lavada, que ainda contém até 5% de lignina (CASTRO, 2009), segue para uma etapa de branqueamento, ou seja, um tratamento físico-químico para a continuação da retirada de lignina usando agentes alvejantes, para obtenção de uma polpa estável. Usualmente, esse processo é feito em etapas consecutivas intercaladas por lavagens para evitar o desgaste das fibras de celulose e para utilizar menores quantidades de reagentes (CASTRO, 2009). Costuma-se usar como primeira etapa a cloração seguida de uma extração alcalina e, por fim, etapas de branqueamento propriamente dito da polpa celulósica. A Tabela 1 resume as principais etapas envolvidas no branqueamento e a simbologia utilizada.

Diferentes conjuntos de estágios podem ser usados. A título ilustrativo, a função de cada etapa de um tratamento chamado CEHD será discutida. A cloração inicial tem por objetivo a retirada da lignina da pasta pela formação da cloro-lignina que é facilmente removida por extração alcalina no estágio subsequente. As etapas posteriores têm a

função de aumentar a alvura da pasta. No estágio final de dióxido de cloro, a lignina residual é oxidada, mas a celulose é preservada. Dessa forma, atinge-se o nível desejado de alvura.

Tabela 1: Estágios de branqueamento com seus respectivos símbolos e produto químico utilizado. FONTE: CASTRO, 2009.

Estágio	Código	Produto Químico
Cloração	C	Cloro gasoso
Extração alcalina	E	Soda cáustica
Hipocloração	H	Hipoclorito de sódio ou cálcio
Dióxido de cloro	D	Dióxido de cloro
Peróxido	P	Peróxido de hidrogênio
Oxigênio	O	Oxigênio
Ozônio	Z	Ozônio
Extração oxidativa	E _o	Soda cáustica e Oxigênio
Extração alcalina com peróxido	E _p	Soda cáustica e peróxido de hidrogênio

O controle do processo de branqueamento é feito basicamente a partir do monitoramento da viscosidade, proporcional à quantidade de celulose, e do número Kappa. Deve-se, portanto, evitar uma diminuição da viscosidade e buscar uma redução do número Kappa para atender às especificações de alvura desejadas.

Após o branqueamento, pode ser necessária a remoção de impurezas em uma etapa suplementar de depuração em centrífugas. Caso a pasta seja o produto final da unidade industrial em questão, é realizada uma etapa de secagem compreendendo uma drenagem ou uma prensagem, seguida da passagem por secadores a vapor ou a ar quente.

3.2.3 Produção de Papel a Partir da Polpa Celulósica

O papel é basicamente uma mistura da polpa de fibras de celulose com aditivos que conferem propriedades aos diferentes tipos de papel, tais como opacidade, resistência mecânica e à umidade, cor, colagem, impermeabilização, brilho, etc. Existem diversos tipos de aditivos, como a cola (formada por breu saponificado), cargas minerais, sulfato de alumínio (que permite a precipitação da cola nas fibras), corantes, plastificantes, dentre outros.

Após a fabricação da polpa de celulose, ela é desagregada em equipamentos similares a liquidificadores e as partículas maiores são separadas em uma centrífuga. Segue-se, então, uma etapa de refino para hidratação e desfibramento das fibras em

refinadores cônicos ou de disco, nos quais a superfície de contato das fibras com a água é aumentada. Como a celulose é higroscópica, a água fica retida. Posteriormente, há o preparo da receita final do papel na mesa plana, ou tanque de formação com adição da cola. A mistura segue para a caixa de nível, na qual a cola é precipitada pela ação do sulfato de alumínio, podendo haver inclusão de outros aditivos. A consistência da polpa, ou seja, a porcentagem de polpa seca por polpa úmida, é reduzida a 1 ou 2% (base mássica) em uma bomba de diluição.

Em seguida, as partículas menores e os contaminantes são retirados em ciclones, depuradores verticais (peneiramento com retenção de fibras enroladas ou impurezas) ou centrífugas. A mistura resultante segue para a máquina de papel, mostrada na Figura 8, constituída de diferentes partes: formação, prensagem, secagem, calandra e enrolamento.

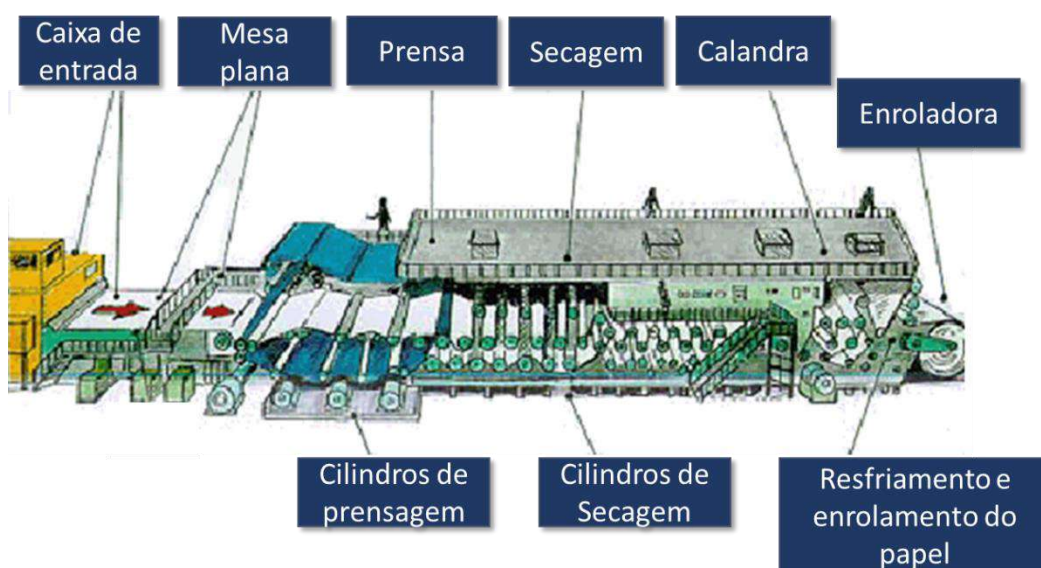


Figura 8: Esquema e seções de uma máquina de papel. FONTE: RENOVA, 2000.

Na máquina de papel, a polpa celulósica é espalhada uniformemente na mesa plana sobre uma tela com pequenos poros que permitem a passagem de água por gravidade ou com a ajuda de caixas de sucção ou demais técnicas de drenagem. Essa água, que ainda contém celulose, é recuperada em canaletas e recirculada.

As fibras retidas seguem para a seção de prensagem para a retirada da água residual por meio de rolos cilíndricos. O papel resultante possui ainda de 60 a 65% de água e segue para uma etapa de secagem que também tem como função curar as resinas aditivas. A quantidade de água é reduzida para 3 a 8%. Outros processos de secagem podem ser realizados em prensas ou cilindros secadores, que também contribuem para

outras características finais do papel tal como brilho (RENOVA, 2000). Por fim, o papel tem sua lisura e espessura definidas nas calandras e é enrolado em bobinas.

3.3 A Caldeira de Recuperação Kraft

O licor negro, subproduto do digestor de deslignificação, é uma solução altamente alcalina. Ele é obtido, como visto anteriormente, após a o processamento da polpa marrom que separa o licor negro da polpa de celulose.

O licor negro tem composição complexa: além de conter a lignina e os açúcares extraídos da madeira, material orgânico combustível, apresenta os sais inorgânicos, que resultam da reação de deslignificação. Em sua composição são encontradas várias classes de compostos inorgânicos e orgânicos conforme indicado na Tabela 2 (BOUCARD, 2014).

A Tabela 3 exemplifica as quantidades de compostos químicos presentes no licor negro de uma planta de papel e celulose na América do Norte.

As propriedades e a composição do licor negro variam em cada planta de acordo com as condições operacionais estratégicas de cada empresa. Essas diferenças são atribuídas, principalmente, aos seguintes fatores: condições operacionais do digestor, espécie da madeira utilizada, razão entre as quantidades de reagentes químicos alimentados ao digestor e a quantidade de madeira, propriedades do licor branco, rendimento da polpa e eficiência da lavagem do material descarregado do digestor.

A recuperação dos sais inorgânicos contidos no licor negro é crítica para reduzir os custos e o impacto ambiental do processo (GALDIANO, 2006).

Para esta finalidade, foram desenvolvidas caldeiras de recuperação que permitem recuperar sulfeto de sódio e o carbonato de sódio. Com isso, o descarte dos subprodutos é reduzido e há um ganho importante em termos ambientais.

Além da recuperação dos sais, a caldeira permite aproveitar o poder calorífico da lignina para produzir vapor, que pode ser utilizado para geração de energia elétrica e para suprir o consumo de vapor da própria planta.

Dessa forma, é possível analisar o processo Kraft como um processo cíclico, uma vez que há recuperação e reciclo de reagentes. A Figura 9 mostra esquematicamente a natureza cíclica de uma planta de produção de pasta de celulose com a recuperação dos sais inorgânicos e geração de vapor em uma caldeira de recuperação Kraft.

Tabela 2: Compostos presentes no licor negro.

Compostos Inorgânicos	
Hidróxido de sódio	NaOH.
Sais de sódio	Tais como Na ₂ S, Na ₂ CO ₃ , Na ₂ SO ₄ , NaCl, Na ₂ S ₂ O ₃ e Na ₂ SO ₃ .
Sais de cálcio	Tal como o CaCO ₃ .
Sais de potássio	Provenientes da madeira.
Compostos Orgânicos	
Lignina	Constituinte da madeira, cuja estrutura é dada na Figura 3 e na Figura 4.
Ácidos orgânicos	Formados principalmente devido à degradação da celulose e da hemicelulose durante a digestão da madeira, destacando-se o ácido acético e o ácido fórmico.
Carboidratos e hidróxi-ácidos	Formados pela degradação da celulose e da hemicelulose.
Ácidos graxos e ácidos resinosos	Também chamados de compostos “ <i>tall oil</i> ”, sendo que a maior parte dos ácidos graxos consiste em ácido oléico e ácido linoleico.
Metanol	Produzido pela desmetilação de grupos metoxila na estrutura da lignina.

Tabela 3: Composição química de um licor negro. FONTE: FREDERICK, 1997.

Composto Químico ou Espécie	Concentração (% peso/peso total)
Lignina alcalina	30 - 45%
Hidróxi-ácidos	25 - 35 %
Sódio	15 - 20 %
Ácido acético	5 %
Enxofre	3 - 5 %
Extrativos*	3 - 5 %
Ácido fórmico	3 %
Metanol	1 %

*Extrativos: materiais solúveis em solventes orgânicos, tais como terpenos e seus derivados, compostos graxos, ceras e seus componentes, e fenóis (CASTRO, 2009).

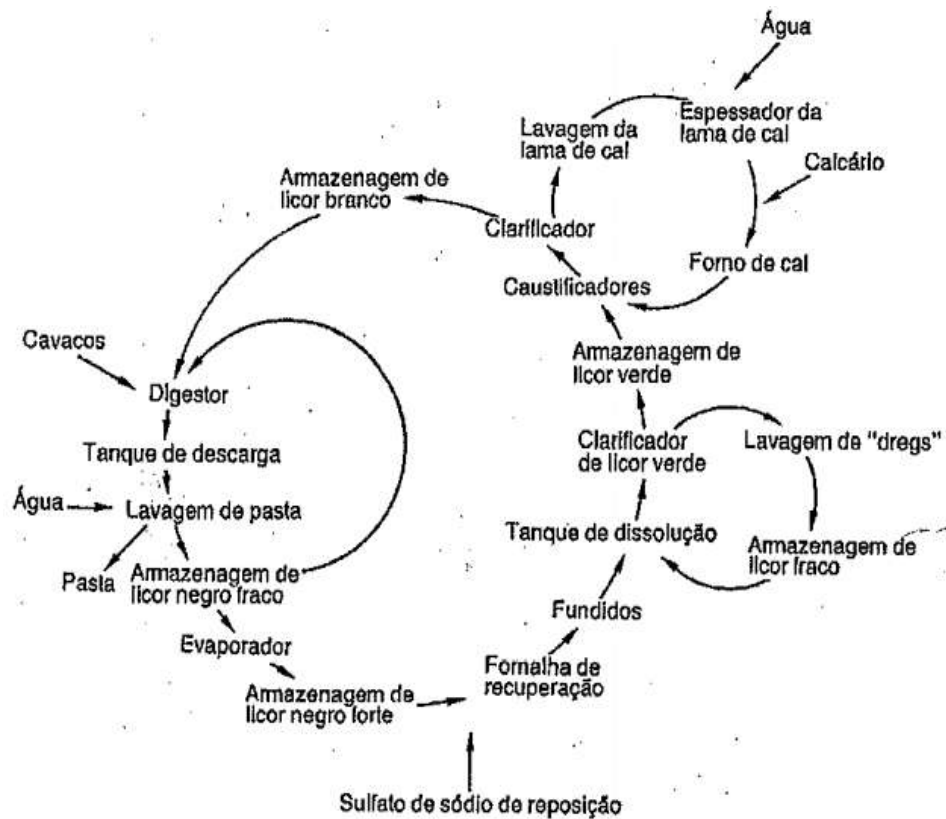


Figura 9: Representação da natureza cíclica do processo Kraft, indicando a recuperação de reagentes. FONTE: COSTA, 2004.

3.3.1 Descrição Física da Caldeira de Recuperação

As caldeiras de recuperação são essencialmente caldeiras aquatubulares de grande porte, com até 40 metros de altura (COSTA, 2004), em que a seção de combustão é bem diferente das caldeiras convencionais usadas para queima de combustíveis líquidos ou gasosos. A Figura 10 mostra, esquematicamente, uma caldeira de recuperação Kraft com detalhes de seus principais componentes, que serão explicados a seguir.

Uma caldeira de recuperação Kraft típica se divide basicamente em duas seções, conforme representado na Figura 11. Na primeira seção, denominada fornalha, ocorre a queima do licor negro. A segunda seção é uma região de transferência de calor por convecção, onde fica localizado um sistema de troca térmica para aproveitamento de calor, que compreende os seguintes componentes: superaquecedores, feixe convectivo, tubulões e economizadores. Entre as duas seções, é formado um estreitamento chamado “nariz” (Figura 10).

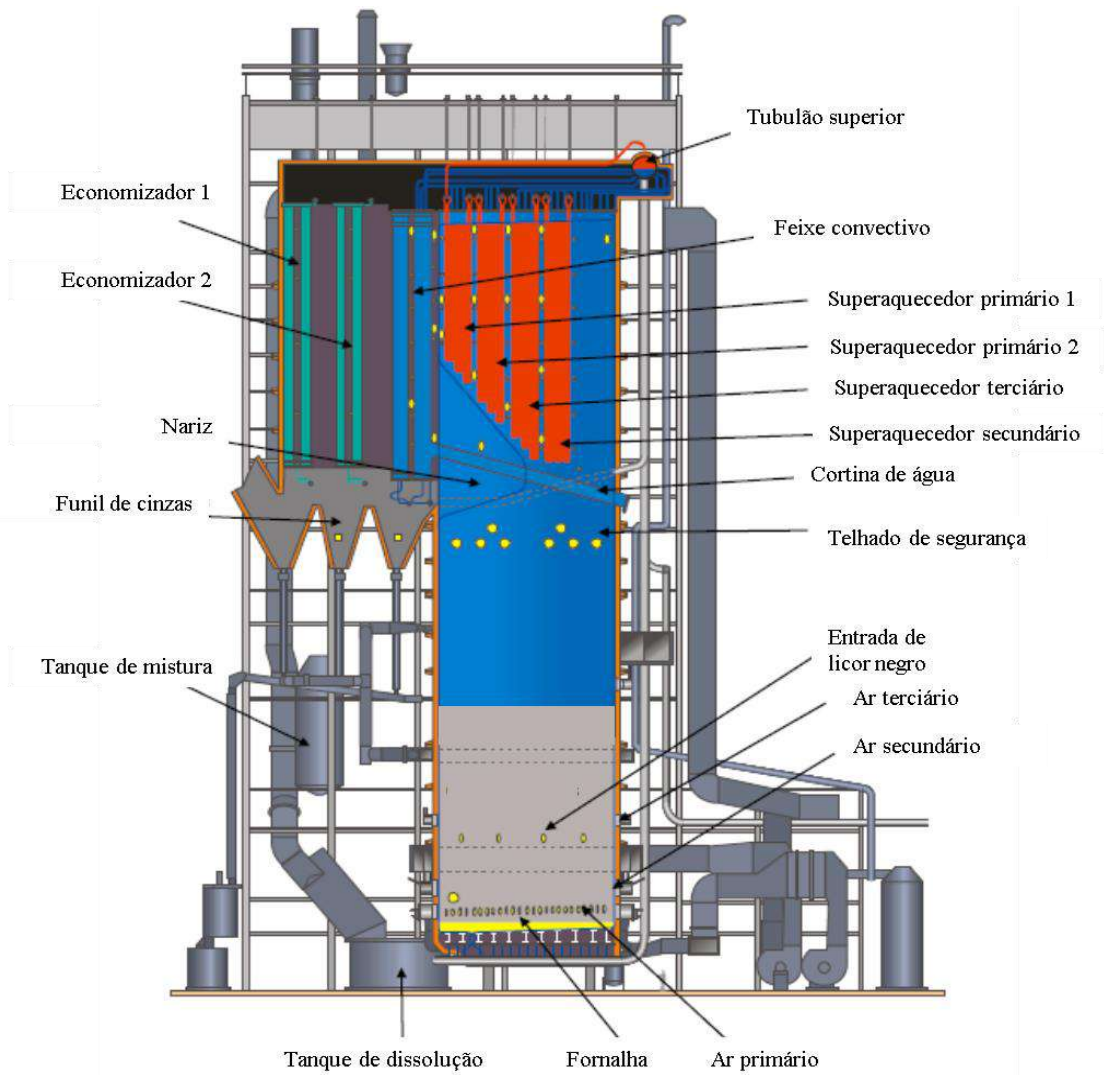


Figura 10: Caldeira de Recuperação Kraft na planta Celulosa Arauco Y Constitución, Nueva Aldea Mill in Chile (Cortesia de Metso Power). Fonte: PRASZKIER, 2011.

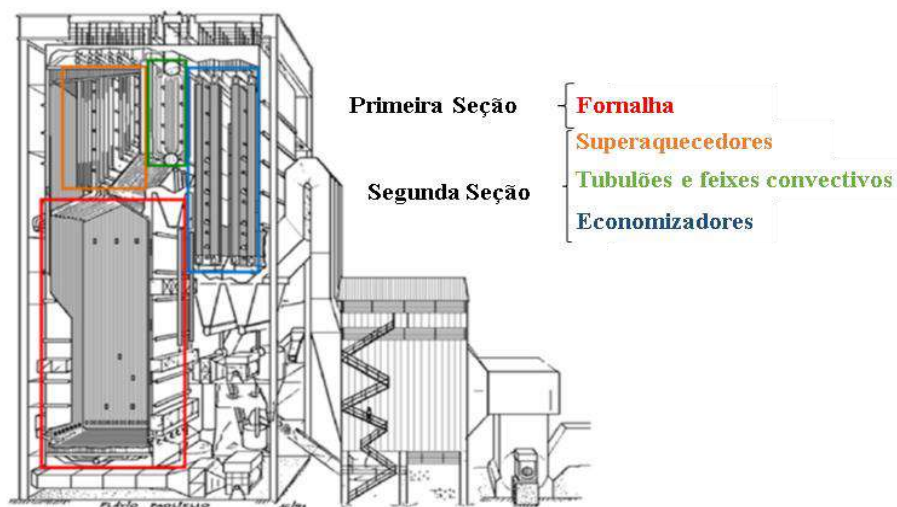


Figura 11: Caldeira de Recuperação Kraft dividida em duas seções. Fonte: PAOLIELLO et al., 2002.

A Fornalha

A queima do licor negro, mais precisamente a etapa de pirólise, se dá na fornalha da caldeira e acarreta na formação do chamado “*char*”, que consiste nos produtos sólidos residuais da pirólise do licor negro. O “*char*” é formado por uma mistura de material carbonáceo e de sais inorgânicos, que compreende, predominantemente, Na_2CO_3 , Na_2S e Na_2SO_4 , e que se depositam no fundo da fornalha (GRACE et al., 1985). Uma mistura de sais inorgânicos fundidos, chamada de “*smelt*”, também se encontra no fundo da fornalha e, em conjunto com o “*char*”, forma um leito ou “*char bed*”.

A fornalha é a região da caldeira com seção transversal retangular. Ao invés dos queimadores presentes em caldeiras convencionais onde é feita a mistura ar-combustível, a fornalha de uma caldeira de recuperação Kraft possui entradas para alimentação de licor negro concentrado, que se distribuem em suas quatro paredes, e injetores de ar de combustão localizados em diferentes alturas. O número de níveis de entradas de ar de combustão pode variar de três a cinco, mas a maioria das caldeiras de recuperação apresentam entradas de ar de combustão em três níveis diferentes, denominados alimentadores de ar primário, ar secundário e ar terciário.

No nível primário, o ar é alimentado na parte inferior da fornalha, pouco acima da superfície do leito (ou “*char bed*”). As entradas de ar primário são localizadas nas quatro paredes e fornecem oxigênio para a queima de material proveniente das partes superiores da fornalha. Nessa região, tem-se predominantemente um ambiente redutor.

O ar primário tem por função queimar o “*char*”, da superfície do leito formado no fundo da fornalha. O “*char*” é formado em etapas, conforme as partículas aspergidas do licor negro vão sofrendo queima. O fluxo de ar primário é responsável, também, pela topologia do leito de “*char*”, que favorece a recuperação dos sais fundidos. O piso da fornalha em algumas configurações é inclinado para facilitar a remoção dos sais fundidos do “*smelt*” (FORNETTI, 1998).

No nível secundário, acima do nível primário, o ar é, em geral, alimentado através das paredes frontal e traseira da fornalha, para definir adequadamente os perfis de fluxos dos gases que circulam e a altura do leito de “*char*”. Nesta região, tem-se também, predominantemente, um ambiente redutor. A principal função do ar secundário é a queima dos gases que são formados no leito de “*char*”. Ele influencia o padrão de escoamento dos gases na fornalha e, conseqüentemente, o desempenho da combustão.

No nível terciário, acima do nível secundário, o ar é alimentado de forma semelhante ao ar secundário, em um ponto acima da entrada do licor negro. Nessa região,

tem-se predominantemente um ambiente oxidante. A função do ar terciário é fornecer oxigênio suplementar para queima de gases e materiais arrastados pela corrente gasosa.

Uma quarta entrada de ar pode estar presente, em geral, em caldeiras de recuperação de grandes dimensões. A entrada de ar quaternário está localizada próximo ao “nariz” da caldeira. O fornecimento de ar quaternário regula o perfil de temperatura do gás de exaustão, resultando em uma redução da emissão de poluentes e inibindo a formação NO_x (FERREIRA, 2010).

A fornalha opera com um excesso de ar para garantir uma combustão completa. No entanto, quantidades muito acima das estequiométricas podem diminuir a eficiência da caldeira e aumentar os custos. Segundo Saviharju et al. (2007), tipicamente, a distribuição de ar total da fornalha pode variar nas seguintes faixas: ar primário de 20 a 35% do ar total, ar secundário de 35% a 60% do ar total, e ar terciário de 10 a 40% do ar total.

Fluxo de água na caldeira

O sistema de troca térmica consiste de diversos elementos tubulares com a função de gerar vapor a partir da água de alimentação de caldeira. A água é proveniente de uma unidade de desmineralização para que o teor de sais, que podem causar incrustações no equipamento, seja minimizado. No sistema de troca térmica da caldeira, a água segue um fluxo contrário em relação aos gases de exaustão.

A água de alimentação deve ser pré-aquecida em trocadores de calor e é introduzida nos economizadores, onde ocorre o aproveitamento final da energia contida nos gases de exaustão. A água segue, então, para o tubulão de vapor, ou tubulão superior.

O tubulão de vapor é um vaso horizontal que se localiza em um ponto alto da caldeira entre o superaquecedor e o economizador. Ele é o principal componente do sistema de água, por funcionar simultaneamente como reservatório da água e de vapor saturado. O seu nível, crítico para a segurança operacional, é mantido por meio de estratégias de controle e condições de intertravamento rigorosas. No Brasil, o controle automático de nível é exigido pela NR-13.

Abaixo do tubulão de vapor e paralelo a ele, encontra-se o tubulão de água, ou tubulão inferior. Os tubulões se comunicam por um feixe de tubos chamado de feixe convectivo.

Do tubulão de água saem diversos tubos que vão formar os diversos componentes do sistema na área da fornalha: as paredes de água, o nariz e as cortinas, que serão

descritos a seguir. Todos os tubos de água (feixe convectivo, paredes de água, nariz e cortinas) recebem calor e geram um fluxo de água parcialmente vaporizada que retorna para o tubulão de vapor, onde a água líquida e o vapor se separam.

No topo do tubulão de vapor existem saídas de vapor que alimentam os superaquecedores, nos quais o vapor é superaquecido e enviado a seu destino na planta. No caso de caldeiras modernas, que operam a pressões elevadas, em geral, o vapor é enviado a turbinas para geração de energia elétrica. Já o vapor de média ou baixa pressão pode ser utilizado nos demais processos da planta.

O sistema de água será descrito a seguir na ordem inversa, do ponto de vista da combustão: começando pelos componentes na área de radiação da fornalha e seguindo para a área de convecção.

Paredes de água

As paredes da fornalha são formadas por um conjunto de tubos contendo água para resfriamento, chamados de paredes de água. A água que circula nos tubos da parede de água recebe o calor da queima do licor negro, especialmente o calor emitido por radiação. Com a troca térmica, a água é transformada em vapor saturado, que é alimentado na seção dos superaquecedores. Isso possibilita, também o resfriamento da caldeira. A Figura 12 mostra o interior da fornalha de uma caldeira de recuperação evidenciando os tubos de aço-carbono que compõem a parede de água.



Figura 12: Interior da fornalha de uma caldeira de recuperação da Andritz com os tubos de aço-carbono. Fonte: VAKKILAINEN, 2005.

Existem três tipos clássicos de configurações de paredes de água dependendo da maneira como os tubos se unem. Os tubos podem ser unidos por barras, chamadas de membranas ou podem ser unidos tangencialmente, ou seja, diretamente uns com os outros. A Figura 13 mostra as seções transversais de dois tipos de arranjos da parede de água de uma caldeira de recuperação.

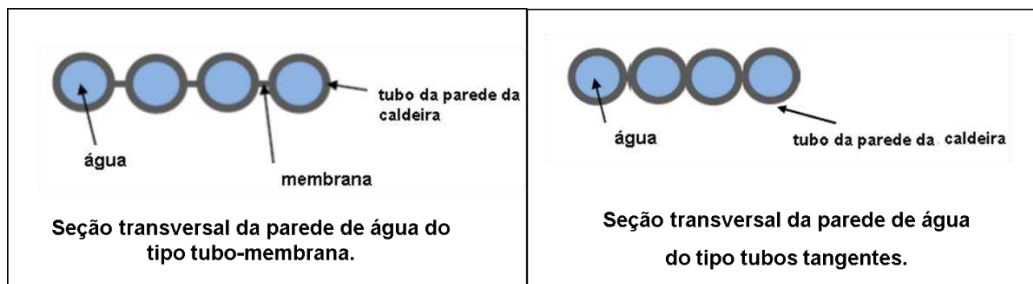


Figura 13: Corte transversal da parede da caldeira. Fonte: PRASZKIER, 2011.

Superaquecedores

Os superaquecedores são tubos tangenciais, cuja função é aquecer o vapor saturado proveniente do tubulão de vapor e aumentar a sua temperatura acima do ponto de saturação, transformando-o em vapor superaquecido. O vapor superaquecido segue para outras unidades da planta. O superaquecedor é projetado para fornecer uma saída de vapor em temperaturas em torno 400 - 550°C (COSTA, 2004), dependendo da pressão de operação da caldeira. O interesse de gerar vapor superaquecido é justificado pelas perdas térmicas ao longo das tubulações da planta que ocorrem até que o vapor chegue a seu destino.

Removedores de Fuligem

Os removedores de fuligem, também chamados de ramonadores, são elementos móveis e rotativos que, em sua extremidade, dirigem jatos de vapor sob pressão para a superfície externa dos tubos para retirar partículas depositadas, evitando assim o seu acúmulo. A deposição de material particulado de baixa condutividade causa perda de desempenho de troca térmica e consequente formação de pontos quentes.

Outros componentes na área de radiação

O nariz, como já mencionado acima, está situado na parte superior da fornalha e delimita a região da fornalha e a região do sistema de trocadores de calor. Ele protege o superaquecedor da radiação incidente na parte inferior da fornalha, tal que a principal

forma de troca de calor com os gases de exaustão seja a convecção. Além disso, o nariz permite reduzir a corrosão do superaquecedor.

A cortina ou “*screen*” é uma superfície formada por tubos, situada na área dos superaquecedores, e recebe calor do gás de exaustão, gerando vapor saturado. A circulação de água da cortina opera da mesma maneira que a água das paredes de água. Uma das funções da cortina é reduzir a temperatura do gás de exaustão antes que ele atinja o superaquecedor. Além disso, a cortina retém parte do material particulado de arraste, captura partículas de licor negro não queimadas e bloqueia a radiação, reduzindo as temperaturas superficiais do superaquecedor.

Economizadores

Os economizadores, localizados na parte superior da caldeira, formam um sistema de troca de calor, constituído de tubos verticais, que recupera o calor residual dos gases de exaustão para pré-aquecer a água que entra no sistema.

Desta forma, as principais finalidades dos economizadores são pré-aquecer a água que entra no sistema e resfriar o gás de exaustão que deixa a caldeira e segue para precipitadores eletrostáticos, onde são removidos os particulados arrastados pelo gás. Usualmente, segundo Costa (2004), os gases saem da caldeira com temperaturas em torno de 177°C. Além disso, os economizadores reduzem o potencial de choque térmico e flutuações de temperatura da água que alimenta os demais componentes da caldeira.

Tubulões e feixe convectivo

Os dois tubulões (de vapor e de água) são as seções que permitem dar capacitância ao sistema. Junto com o feixe convectivo, ao redor do qual escoam o gás de exaustão entre os superaquecedores e os economizadores, os tubulões formam um sistema que tem a função de receber a água de alimentação pré-aquecida, de fornecer vapor saturado aos superaquecedores, e de integrar os tubos de água das diversas seções – estes tubos saem do tubulão inferior e retornam para o tubulão superior. Além disso, o tubulão inferior permite a instalação de um tubo para a purga contínua da caldeira, o que permite manter uma concentração de sais que evite incrustação na parte interna dos tubos.

3.3.2 Processo de Recuperação: Transformações físico-químicas do licor negro nas diferentes zonas da fornalha

Enquanto o circuito de água de uma caldeira de recuperação é muito semelhante ao das caldeiras aquatubulares utilizadas em geração de vapor em termoelétricas e na

indústria de petróleo e gás, a fornalha é uma seção muito específica da indústria de papel e celulose.

A fornalha tem diversas funções: é nela que se remove a umidade remanescente do licor negro e que se faz a combustão dos compostos orgânicos e a redução dos sais de sódio de forma a recuperar o sulfeto de sódio em forma fundida.

Os perfis de temperatura e de composição dos gases de exaustão, tanto no eixo horizontal, quanto no eixo vertical, da caldeira variam muito. A partir do nível de entrada do licor negro, injetado por aspersão de forma a criar um movimento descendente do licor, grandes quantidades de vapor d'água e de materiais voláteis orgânicos são liberadas. Abaixo desta região, onde a maior parte da combustão ocorre, a temperatura do gás chega ao seu valor máximo, de cerca de 1300°C a 1400°C, no caso de caldeiras modernas. No nível de entrada do ar secundário, a temperatura dos gases diminui a valores próximos de 1100°C, onde as reações de combustão ocorrem preponderantemente, (MIKKANEN, 2000). A temperatura de saída dos gases e de material particulado residual a montante dos superaquecedores é de cerca de 900°C. Outros autores mencionam um perfil de temperatura na fornalha, em que a parte inferior está a 900°C, junto ao leito de “char”, e a parte superior se encontra a 1.000°C (COSTA, 2004).

Pode-se dividir a fornalha em três zonas para melhor compreensão dos processos físico-químicos que nela ocorrem, conforme a Figura 14.

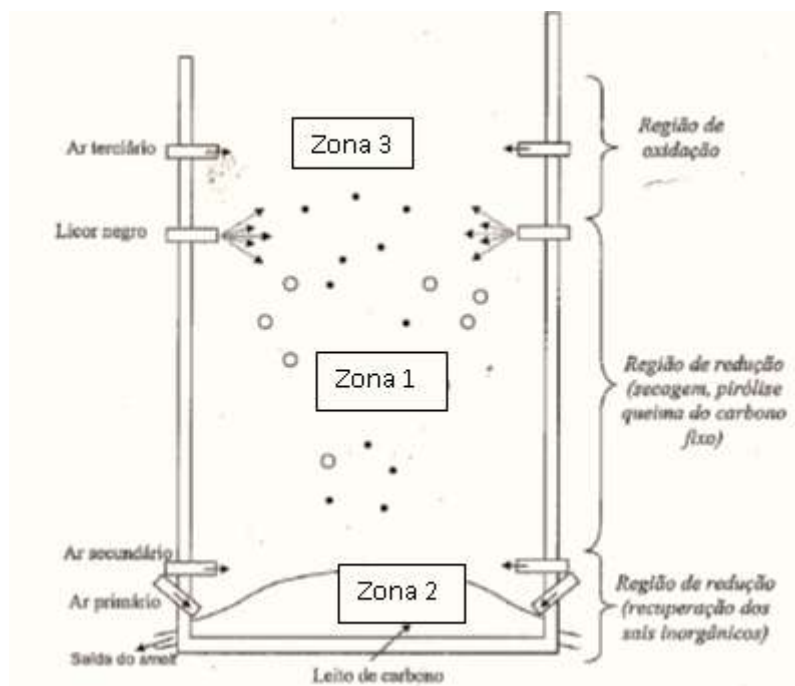


Figura 14: Representação esquemática das três zonas da fornalha de uma caldeira de recuperação Kraft. FONTE: COSTA, 2004.

O licor negro, contendo sólidos em suspensão, é alimentado na caldeira, na primeira zona, através de uma série de bocais de aspersão na forma de gotículas com tamanhos controlados. As gotas descem em contrafluxo em relação ao ar ascendente, perdendo inicialmente a umidade e em seguida entrando em combustão. Dessa forma, a primeira zona da fornalha é uma região de redução, onde ocorrem, predominantemente, as etapas de secagem, pirólise e queima do carbono fixo do licor negro. A queima do licor negro é realizada em um ambiente deficiente de oxigênio de tal forma a promover reações de redução do enxofre de compostos de enxofre com a formação de sulfeto de sódio (Na₂S).

O controle do tamanho das gotículas de licor negro na alimentação é muito importante para otimizar a etapa de secagem (perda de umidade da partícula), favorecer sua queima e para garantir que elas atinjam a parte inferior da fornalha, formando uma camada onde a parte orgânica ainda está em combustão. Por um lado, gotículas excessivamente pequenas seriam arrastadas para a parte superior da caldeira, por outro, gotículas muito grandes podem atingir o fundo da caldeira antes de serem adequadamente queimadas (VAKKILAINEN, 2005). Vakkilainen (2005) estabelece que o licor deve ser aspergido ou pulverizado como gotículas “grosseiras”, com diâmetro médio entre 2 e 3 mm.

Segundo Grace et al. (1985), a queima das gotículas de licor negro na fornalha compreende três estágios: secagem, queima de voláteis e queima do produto sólido residual da pirólise, ou seja, do “char”. A queima de voláteis consiste de dois processos distintos: a pirólise e a combustão gasosa. Estes estágios tendem a ocorrer sequencialmente, embora devido aos tamanhos diferenciados das gotículas e das partículas, e dos gradientes de temperatura, possam ocorrer simultaneamente.

Após a secagem das gotículas por evaporação da água, quando as partículas de licor negro atingem uma temperatura de aproximadamente 200°C, as reações de pirólise se iniciam. A pirólise é a decomposição, ou quebra térmica dos sólidos do licor negro e envolve reações de fragmentação e polimerização competitivas. A pirólise produz gases combustíveis e o “char” (GRACE et al., 1985). Quando a pirólise é rápida, devido a altas temperaturas, a formação de gases é favorecida, ao passo que a pirólise mais lenta tende a aumentar a formação de “char”.

Durante a pirólise, as partículas aumentam de volume devido à grande liberação de gases de combustão, conforme mostrado na Figura 15, e se encontram em chamas. Os gases formados durante a pirólise são SO₂, CO₂, CO, CH₄, H₂O e TRS (enxofre total

reduzido). Conforme Silva (2008), os gases TRS são H_2S , CH_3SH , CH_3SCH_3 e $CH_3S_2CH_3$.

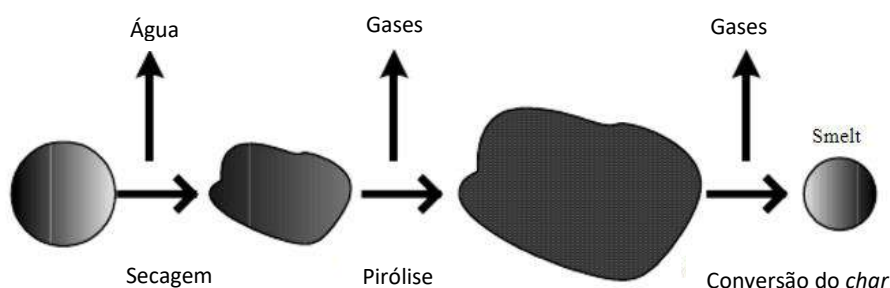


Figura 15: Variação do tamanho da gotícula/partícula do licor negro após entrada na caldeira de recuperação. FONTE: HOLMBLAD, 2015.

Vakkilainen (2006) apresenta uma “escala de tempo” para a queima de uma gotícula de 2 mm de diâmetro, dividindo em quatro etapas o processo global de queima do licor negro desde a primeira zona até o material atingir a terceira zona e formar o “*smelt*” (mistura de sais fundidos), conforme a Tabela 4.

Tabela 4: Principais etapas da queima do licor negro para uma gotícula de 2 mm. Fonte: VAKKILAINEN, 2006.

Etapa	Características	Escala de tempo na fornalha
Secagem	Evaporação da água, diâmetro constante, seguindo-se o inchamento	0,1 - 0,2 s
Desvolatilização	Material em chamas, ignição, inchamento e expansão da gotícula, com liberação de voláteis	0,2 - 0,3 s
Queima do “ <i>char</i> ”	Desaparecimento da chama, reações de redução, diâmetro aumenta	0,5 - 1 s
Fusão (“ <i>smelt</i> ”)	Diâmetro aumentando ou constante, re-oxidação	Longo

As diversas reações de redução e oxidação que podem ocorrer na caldeira devido à queima do licor negro e à combustão da fase gasosa, conforme sugeridas na literatura

por Grace et al. (1998) e Costa (2004), estão agrupadas na Tabela 5. Essas reações podem ocorrer simultaneamente em zonas diferentes da caldeira.

Tabela 5: Principais reações e mudanças de estado na fornalha. FONTES: GRACE et al., 1998 e COSTA, 2004, JARVINEN et al., 2002.

REAÇÃO QUÍMICA OU MUDANÇA DE ESTADO	COMENTÁRIOS
$\text{Na}_2\text{S} + \text{H}_2\text{O} + \text{CO}_2 \rightarrow \text{H}_2\text{S} + \text{Na}_2\text{CO}_3$ $\text{Na}_2\text{S} + 2 \text{CO}_2 \rightarrow \text{Na}_2\text{CO}_3 + \text{COS}$	Formação de gases com enxofre.
$\text{Na}_2\text{CO}_3 + \text{SO}_2 + \frac{1}{2} \text{O}_2 \rightarrow \text{Na}_2\text{SO}_4 + \text{CO}_2$ $2\text{NaCl} + \text{SO}_2 + \frac{1}{2}\text{O}_2 + \text{H}_2\text{O} \rightarrow 2\text{HCl} + \text{Na}_2\text{SO}_4$ $\text{Na}_2\text{S}_{(s)} + 2 \text{O}_2_{(g)} \rightarrow \text{Na}_2\text{SO}_4_{(g)}$	Formação de sulfato de sódio.
$\text{Na}_2\text{SO}_4_{(s)} + 2 \text{C}_{(s)} \rightarrow \text{Na}_2\text{S}_{(s)} + 2 \text{CO}_2_{(g)}$ $\text{Na}_2\text{SO}_4_{(s)} + 4 \text{C}_{(s)} \rightarrow \text{Na}_2\text{S}_{(s)} + 4 \text{CO}_{(g)}$	Redução do sulfato.
$\text{Na}_2\text{CO}_3 + 2 \text{C} \rightarrow 2 \text{Na} + 3 \text{CO}$ $\text{NaCl}_{(s,l)} \rightarrow \text{NaCl}_{(v)}$	Segundo Grace et al. (1998), acredita-se que o sódio se libera apenas durante a queima do “char” através da redução de carbonato de sódio pelo carbono, produzindo vapor de sódio elementar e pela vaporização direta de NaCl.
$\text{H}_2\text{O}_{(l)} \rightarrow \text{H}_2\text{O}_{(g)}$	Controlada por transferência de calor.
Material pirolisável $\rightarrow \text{CH}_n\text{O}_o_{(s)}$, $\text{H}_2\text{O}_{(g)}$, $\text{CO}_{(g)}$, $\text{CH}_4_{(g)}$, $\text{Na}_{(g)}$, $\text{M}_2\text{S}_{(s)}$, $\text{M}_2\text{SO}_4_{(s)}$, $\text{M}_2\text{CO}_3_{(s)}$ M = Na ou K	Abordagem estabelecida por Jarvinen (2002) para definir o resultado da queima da partícula do material pirolisável presente no licor negro.
$\text{C}_{(s)} + \text{H}_2\text{O}_{(g)} \rightarrow \text{CO}_{(g)} + \text{H}_2_{(g)}$ $\text{C}_{(s)} + \text{CO}_2_{(g)} \rightarrow 2 \text{CO}_{(g)}$ $\text{C}_{(s)} + \frac{1}{2} \text{O}_2_{(g)} \rightarrow \text{CO}_{(g)}$ $\text{C}_{(s)} + \text{O}_2_{(g)} \rightarrow \text{CO}_2_{(g)}$	Reações do carbono elementar.
$\text{H}_2_{(g)} + \frac{1}{2} \text{O}_2_{(g)} \rightarrow \text{H}_2\text{O}_{(g)}$	
$\text{CO}_{(g)} + \text{H}_2\text{O}_{(g)} \rightarrow \text{CO}_2_{(g)} + \text{H}_2_{(g)}$	
$\text{CH}_4_{(g)} + \text{H}_2\text{O}_{(g)} \rightarrow \text{CO}_{(g)} + 3 \text{H}_2_{(g)}$ $\text{CH}_4_{(g)} + \text{CO}_2_{(g)} \rightarrow 2 \text{CO}_{(g)} + 2 \text{H}_2_{(g)}$	Reações do metano produzido.
As reações indicadas para compostos de sódio ocorrem de forma correspondente para os sais de potássio.	

A zona 2, localizada abaixo da primeira, é uma região de redução, compreendendo entradas de ar primário e ar secundário e saídas dos sais inorgânicos recuperados na forma de uma massa fundida, chamada “*smelt*”, que se encontra misturada ao “*char*”, formando o leito de “*char*”. O ar injetado nessa região não penetra no leito, e, por isso, ocorrem reações de redução dos sais inorgânicos, propiciando a formação de sulfeto de sódio.

A massa de sais inorgânicos fundida, formada, predominantemente, de sulfeto de sódio (Na_2S) e carbonato de sódio (Na_2CO_3), é vazada pelo fundo da caldeira em um tanque de dissolução contendo água, obtendo-se assim o chamado licor verde, que compreende basicamente Na_2CO_3 , Na_2S e que, após a etapa de caustificação, é transformado em licor branco retornando ao processo de deslignificação.

A zona 3, localizada acima da primeira zona, é uma região de oxidação, com injeção de ar terciário. Nessa zona, o objetivo da injeção de ar é completar a combustão dos gases e promover sua mistura. Os seguintes gases podem estar presentes na terceira zona da fornalha: CO_2 , H_2O , SO_2 , O_2 , Na_2SO_4 , Na_2CO_3 , NaCl , HCl , CO , H_2S , N_2 e NO_x .

É inevitável que exista o arraste de pequenas partículas sólidas pelos gases de exaustão, além da formação de sais na parte superior da fornalha, que seguem o fluxo de gases de exaustão. Evidentemente, este arraste é indesejável, pois pode causar diversos problemas, tais como deposição na superfície externa dos tubos de água, chegando a causar bloqueio da passagem dos gases e redução da eficiência da troca térmica, além do custo de remoção destas partículas no separador eletrostático na saída da caldeira. Esta questão é de grande interesse para o presente estudo e será abordada em mais detalhes na seção seguinte.

3.3.3 Formação de Material Particulado na Caldeira de Recuperação Kraft

O acúmulo de partículas pode ser causado tanto pela vaporização e posterior condensação de compostos alcalinos quanto por fragmentos de partículas de “*char*” que são arrastados pelo gás de exaustão. O progresso das incrustações é acompanhado, principalmente, pelo monitoramento das seguintes variáveis operacionais: queda de pressão, temperatura do gás de exaustão ao longo da caldeira e temperatura do vapor produzido (VAKKILAINEN, 2005). Como dito anteriormente, o material particulado depositado nas superfícies de troca térmica é retirado por ramonadores, porém, algumas incrustações podem se tornar resistentes a esse mecanismo de remoção. Quando há parada

de operação da caldeira, os ramonadores podem ejetar água líquida ao invés de vapor para lavagem.

Para que o acúmulo de material particulado ocorra sobre as superfícies de troca térmica, as partículas devem ser capazes de aderir às superfícies. Para isso, é necessário que haja uma certa quantidade de fase líquida. Assim, destacam-se quatro temperaturas importantes na deposição de materiais particulados: a dita “primeira” temperatura de fusão, abaixo da qual todo o material está na fase sólida; a “segunda” temperatura de fusão, acima da qual todo o material inorgânico se encontra na fase líquida e duas temperaturas intermediárias, a saber: a chamada “*sticky temperature*”, que é a temperatura na qual o material apresenta características de aderência aos tubos e paredes e contém entre 15% e 20% de líquido; e a temperatura de deformação radical, na qual 70% do material inorgânico está líquido e apresenta uma grande fluidez e pouca capacidade de aderência.

Além do processo de deposição de partículas sobre as superfícies do sistema de troca térmica, há ainda outro fenômeno importante chamado de sinterização. A sinterização se caracteriza pela aglomeração e pela união de partículas vizinhas com densificação e aumento da rigidez. O processo de sinterização dificulta a remoção das partículas pelos ramonadores e ocorre de forma mais rápida para partículas menores.

Os depósitos de material particulado consistem majoritariamente em compostos alcalinos solúveis em água (99%), principalmente Na_2SO_4 e Na_2CO_3 , de uma pequena quantidade de NaCl e de compostos de enxofre reduzidos, tal como Na_2S . Há ainda a presença de compostos de potássio.

Os depósitos de partículas nas regiões superiores da caldeira são compostos basicamente de três tipos de material: partículas de arraste (em inglês, *carryover*), fumo (em inglês, *fume*) e partículas de tamanho intermediário ou ISP do inglês “*intermediate sized particle*”.

Partículas de Arraste

Os depósitos de partículas de arraste são formados por impactação inercial² de partículas maiores do “*smelt*” ou partículas do licor negro parcialmente queimadas, que

² Impactação inercial: A impactação inercial é a separação de partículas presentes em um fluxo de gases que possuem uma inércia suficiente para colidirem em um obstáculo e se desprenderem do fluxo gasoso.

foram arrastadas pelo gás de exaustão. O tamanho das partículas de arraste varia de 0,01 a 3 mm e, quando depositadas, elas apresentam uma coloração rosada (TRAN, 2007).

Estas partículas são carregadas pelo fluxo de gás e ascendem na fornalha até atingirem a zona predominantemente oxidativa, onde o sulfeto de sódio (Na_2S) é majoritariamente oxidado, formando sulfato de sódio (Na_2SO_4). Como esperado, por causa de sua origem, as partículas de arraste possuem uma composição próxima à do “*smelt*” oxidado, entretanto, uma vez que as partículas são depositadas sobre as superfícies dos tubos, uma parte do carbonato presente pode reagir com o dióxido de enxofre (SO_2) e formar mais sulfato de sódio e os cloretos podem ser vaporizados ou convertidos em sulfatos (TRAN, 2007).

Dessa forma, o depósito de partículas de arraste apresenta uma quantidade de sulfato mais alta que a do “*smelt*” oxidado, e quantidade menores de cloreto e de carbonato.

Fumos

Os depósitos de fumos são originários de compostos alcalinos de sódio e potássio do licor negro que vaporizam e seguem o fluxo de gás de exaustão. Eles reagem com os gases produzidos na caldeira e os compostos formados se condensam quando a temperatura é reduzida, formando pequenas partículas.

As partículas de fumos, com aparência branca, costumam apresentar uma distribuição uniforme de tamanho, que varia de 0,1 a 1 μm . O tamanho reduzido dessas partículas as torna muito reativas.

A composição do fumo varia com a temperatura do leito de “*char*”, a sulfidez³, a concentração de voláteis no licor e com a posição na caldeira. Quando a temperatura é elevada no leito de “*char*”, há maior produção de fumo.

Na região inferior da fornalha, o enxofre presente na corrente de gás se encontra predominantemente na sua forma reduzida (H_2S), por causa das temperaturas elevadas e do ambiente redutor. Segundo Tran (2007), os depósitos de fumos são formados majoritariamente de hidróxidos alcalinos (KOH e NaOH) e carbonatos nesta parte da fornalha. Já na região superior da fornalha, onde a atmosfera é oxidativa, há conversão de H_2S em SO_2 , que reage, por sua vez, com os hidróxidos e forma sulfatos de sódio e potássio. Além disso, os cloretos e os carbonatos também reagem e são sulfatados, porém

³ Sulfidez: É a razão entre as concentrações de sulfeto de sódio e de álcali ativo, ou seja, de uma solução de sulfeto de sódio e hidróxido de sódio usada na deslignificação.

com uma cinética mais lenta. No nariz, os fumos possuem mais sulfatos, uma quantidade menor de carbonatos e cloretos e não apresentam hidróxidos nem compostos de enxofre reduzido.

Quando comparados às partículas de arraste, os fumos são mais ricos em sulfatos, cloretos e compostos de potássio, mas apresentam menor quantidade de carbonatos. Isso pode ser explicado tanto pelas reações de sulfatação quanto pela alta volatilidade dos cloretos de sódio e potássio.

Partículas de tamanho intermediário (ISP)

As partículas ISP apresentam coloração branca ou acinzentada e tamanhos variando de 1 a 100 μm e são provenientes do material fundido arrastado pelo gás. Elas podem ser formadas pela fragmentação de gotículas de licor negro durante a combustão, pelo escape de partículas do material fundido durante a queima do “*char*”, ou pela aglomeração de fumos que foram depositados, sofreram sinterização e retornaram ao fluxo de gás. As partículas ISP podem ainda ser resultantes da formação de partículas irregulares nos evaporadores devido à presença de impurezas minerais, compreendendo compostos de silício, magnésio ou cálcio (VAKKILAINEN, 2005). Segundo Tran (2007), a composição das ISP é de difícil determinação, uma vez que sua separação das partículas de arraste e dos fumos é complicada. No entanto, sugere-se que a composição seja similar à das partículas de arraste, porém com mais sulfatos e com uma quantidade menos significativa de potássio, cloreto e carbonato (TRAN, 2007).

Tipo de Depósito de Acordo com a Localização na Caldeira

A composição dos depósitos de material particulado varia de acordo com a sua localização no sistema de troca térmica da caldeira, principalmente, por causa da variação de temperatura.

Nos superaquecedores são observados dois fenômenos: a deposição de fumos e a impactação inercial de partículas maiores, uma vez que a velocidade do gás de exaustão é relativamente baixa (3 a 5 m/s), permitindo a separação de partículas grandes (VAKKILAINEN, 2005). Na região inferior dos superaquecedores, as temperaturas ainda são elevadas e o tempo de residência não é suficiente para que a vaporização e a formação de fumos ocorram. Assim, a maior parte do depósito de partículas corresponde às partículas de arraste e ISP. Já na parte superior, a redução da temperatura permite a formação de fumos.

É interessante ressaltar que, na parte inferior dos superaquecedores, a temperatura pode atingir a temperatura de deformação radical, tornando o depósito fluido. Parte do depósito pode então escorrer e evitar o aumento da espessura das incrustações, em um fenômeno que regula o acúmulo e impede a obstrução. No entanto, na parte superior, pode haver um acúmulo significativo de particulados, uma vez que a temperatura é inferior àquela de deformação e a espessura do depósito não é limitada pela fluidez do material. A Figura 16 mostra o depósito de material particulado nos tubos dos superaquecedores, dando uma dimensão do problema aqui tratado.



Figura 16: Formação de depósitos de material particulado nos superaquecedores de uma caldeira de recuperação Kraft. FONTE: VAKKILAINEN, 2005.

O feixe convectivo é a parte da caldeira mais problemática em relação à deposição de particulados, principalmente por ser o local mais comum de ocorrência de obstruções. Como a maior parte das partículas de arraste e ISP já foi depositada anteriormente e a temperatura do gás de exaustão é menor nessa região, os fumos são o tipo de partícula mais presente. A entrada da região e o centro do feixe são os locais onde mais se observa a obstrução da passagem do gás de exaustão. Isso porque o espaçamento entre os tubos é reduzido e a ação dos ramonadores, localizados nas extremidades do feixe convectivos, pode empurrar as partículas e provocar seu deslocamento para a região do centro do feixe convectivo.

Nos economizadores, a ocorrência de obstruções pode ser evitada pela presença de ramonadores em sua entrada e pode ser explicada pela impactação de partículas causada pela mudança na direção do gás e pela formação de bissulfato de sódio (NaHSO_4) a partir da reação do sulfato de sódio com SO_x e vapor de água presentes na corrente de gás (TRAN, 2007). Segundo Vakkilainen (2005), pH baixos podem tornar as partículas

mais pegajosas, aumentando a sua aderência aos tubos e provocando incrustações de difícil remoção.

Nos precipitadores eletrostáticos, são encontradas partículas mais finas que não foram depositadas ao longo da caldeira. Segundo Tran (2007), 95% do material particulado é formado por fumos e os outros 5% são partículas de arraste e ISP, apresentando uma quantidade significativa de Na_2SO_4 e Na_2CO_3 .

A Figura 17 apresenta de forma resumida a composição dos depósitos encontrados em diferentes regiões da caldeira e no precipitador eletrostático.

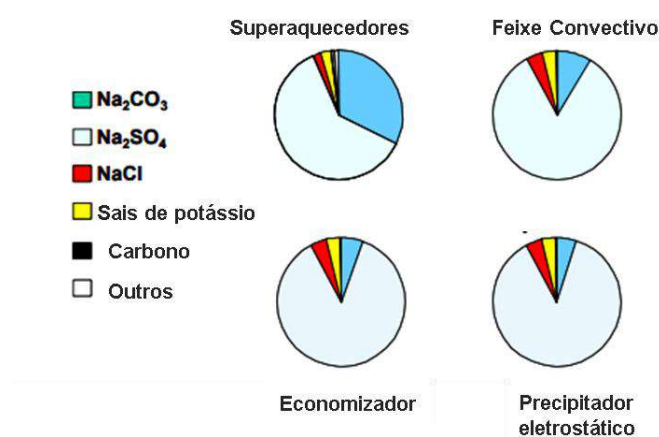


Figura 17: Composição dos depósitos de material particulado. FONTE: TRAN, 2007.

A quantidade de partículas que sai da fornalha pode ser medida por instrumentos localizados na região de entrada dos superaquecedores. Em algumas instalações industriais são utilizadas câmeras e processamento de imagens para contar o número de partículas que deixam a fornalha por unidade de tempo. É comum haver redundância nos sensores, uma vez que a quantidade de partículas pode não ser uniforme e variar de acordo com a posição do sensor. Os dados usados no presente estudo são referentes a uma caldeira dotada de tal sistema de medição, em que existem dois sensores localizados no lado direito e no lado esquerdo da fornalha pouco abaixo dos superaquecedores.

3.4 Parâmetros de Interesse para Planta de Produção de Papel e Celulose com Ênfase no Processo Kraft

Serão destacados a seguir três parâmetros operacionais fundamentais para o Controle e Automação da Indústria de Papel, relacionados a três diferentes unidades de uma planta típica de processo Kraft.

O número Kappa, já citado, indica a quantidade de lignina presente na polpa e é essencial para avaliação do desempenho dos processos de deslignificação e de branqueamento.

O parâmetro *Freeness* representa uma medida da drenabilidade da polpa na máquina de papel e é usado, entre outros, para monitoramento e controle do consumo energético, da temperatura e da consistência da polpa na etapa de desfibramento.

Por fim, destaca-se o parâmetro de formação de material particulado na caldeira de recuperação do processo Kraft. As partículas podem ser arrastadas e depositadas na caldeira, reduzindo sua eficiência. Desta forma, este parâmetro auxilia na avaliação da operação da caldeira e o seu controle tem o potencial de reduzir a frequência de paradas para manutenção.

3.4.1 Modelagem dos Parâmetros

Um problema comum aos três parâmetros mencionados é o seu monitoramento. O procedimento mais usual para a determinação dos números Kappa e *Freeness* consiste na coleta de uma amostra e na sua análise em laboratório. Todavia, o uso de análises laboratoriais introduz um atraso, que dificulta o controle do processo em tempo real. No entanto, esforços vêm sendo empregados para encontrar soluções que diminuam o intervalo de tempo entre a obtenção do resultado da análise e o momento de coleta, ou mesmo que seja possível a determinação deste parâmetro de forma *online* e instantânea. Por isso, métodos *online* vêm sendo desenvolvidos e aplicados na indústria para possibilitar uma determinação em tempo real. Entretanto, os métodos *online* podem oferecer desvantagens, tais como a necessidade de se fazer manutenções frequentes nos instrumentos de medição e de se ter condições operacionais estáveis.

Outras soluções buscadas estão ligadas ao desenvolvimento de modelos matemáticos, fenomenológicos, empíricos e estatísticos para a predição dos números Kappa e *Freeness*.

Vroom (1957) foi um dos pioneiros na modelagem da deslignificação com a criação do fator H para digestores em batelada. Este fator agrupa a temperatura e o tempo de cozimento e foi muito utilizado em outros trabalhos para o desenvolvimento de correlações entre o fator H e o número Kappa, como o realizado por Hatton (1973).

Outros pesquisadores desenvolveram modelos para a deslignificação nos anos seguintes, que se basearam em métodos estatísticos e em modelos fenomenológicos, tal

como mostrados por Rantanen em 2006 e, até mesmo, a partir de redes neuronais artificiais, como divulgados por Dufour et al. em 2005. Correia (2016) estudou a predição do número Kappa para madeiras folhosas a partir de métodos estatísticos, tais como média móvel simples (MMS), alisamento exponencial simples (AES), Regressão Linear Múltipla (RLM), entre outras, e de redes neuronais artificiais e concluiu que este último método apresentava uma boa predição para o valor do parâmetro.

Para a determinação do *Freeness*, foram desenvolvidos métodos de simulação. Segundo Rocha et. al. (2006), o *Freeness* pode ser calculado matematicamente com base em correlações, utilizando os resultados de testes laboratoriais, e de curvas de drenagem. Além disso, o uso de redes neuronais também vem sendo explorado para a determinação do *Freeness* (WALTZ et al., 2017).

O terceiro parâmetro importante é o número de partículas formadas na fornalha da caldeira de recuperação, que será chamado simplesmente de “número de partículas” no texto. A incrustação das superfícies de troca térmica do sistema de recuperação de calor da caldeira por material particulado representa um dos maiores problemas operacionais do equipamento, tornando necessárias paradas não programadas (*shutdown*). Os depósitos de sólidos aumentam a resistência à troca térmica, reduzindo a eficiência térmica da caldeira; contribuem para a corrosão e, em casos extremos, dificultam ou até mesmo bloqueiam o escoamento do gás de exaustão.

O presente trabalho se insere justamente no contexto de formação de material particulado nas caldeiras de recuperação Kraft e busca uma alternativa para a medição do número de partículas produzidas na caldeira a partir da sua predição ou classificação utilizando redes neuronais artificiais.

Diversos estudos foram realizados para a detecção de obstruções e a modelagem da formação de depósitos no sistema de troca térmica de caldeiras de recuperação como evidenciado nos trabalhos de VAKKILAINEN (2005 e 2006), porém poucos estudos foram feitos na modelagem da formação de material particulado, principalmente pela variabilidade e complexidade do processo de formação conforme visto anteriormente.

Entretanto, Costa et al. (2005) se empenharam nesta problemática e modelaram a formação de material particulado de uma caldeira industrial a partir de redes neuronais. O trabalho realizado pelo grupo foi de grande importância para a indústria de papel e celulose, tendo sido, inclusive premiado pela sua originalidade. Primeiramente, testou-se a predição do número de partículas produzidas na fornalha e que seguem para os demais componentes da caldeira usando redes MLP (explicadas posteriormente) com as seguintes

características: três camadas; função de ativação tangente sigmoideal tanto para a camada oculta quanto para a camada de saída; método de aprendizado retropropagação com 1.000 épocas; e seleção de 80% dos dados para o treinamento das redes e os 20% restantes para teste.

A melhor rede encontrada apresentava 12 neurônios ocultos e recebia como variáveis de entrada a vazão, a pressão e a temperatura de alimentação do licor negro, as vazões de entrada do ar primário, do ar secundário e do ar terciário, e a pressão e a temperatura do ar secundário. No entanto, a predição não se mostrou uniforme e apresentou erros significativos para grandes valores de número de partículas, superiores a 300 partículas/min, o que pode ser justificado pela pouca quantidade de dados correspondentes à geração elevada de partículas.

Costa et al. (2005) adotaram também uma outra abordagem para o problema tentando classificar as partículas com redes do tipo RBF. A quantidade de partículas foi dividida em classes de acordo com intervalos pré-estabelecidos da quantidade de partículas formadas. Foram testadas entre duas e sete classes, mas apenas a classificação das partículas em duas classes apresentou resultados satisfatórios. A autora sugere que a divisão em duas classes, especificando um corte nos casos de uma quantidade de partículas superior a 200 partículas/min, é suficiente para o monitoramento do bom funcionamento da caldeira. Dessa forma, a melhor rede neuronal testada apresentou um desempenho global de 80,8% e 80,1% para o treinamento e para o teste, respectivamente, e classificou corretamente mais de 80% dos dados. A rede apresentou 32 neurônios ocultos e 5 variáveis de entrada relevantes: pressões das entradas de ar primário e secundário, vazões do ar primário e terciário e temperatura de entrada do licor negro.

4. REDES NEURONAIS ARTIFICIAIS

4.1 As Redes Neurais Artificiais

Redes Neurais Artificiais (RNA), ou em inglês *Artificial Neuronal Networks* (ANN), são técnicas não algorítmicas (DE SOUZA Jr., 2005) e empíricas de processamento de informações (OTERO, 1998) usadas para geração de modelos. As RNA são um exemplo de inteligência artificial inspirado no sistema neuronal biológico e que, assim como o cérebro humano, são dotadas da capacidade de aprendizagem através de exemplos.

As redes neuronais são treinadas a partir de um conjunto de dados, que apresentam alguma relação entre si, e identificam padrões para generalizá-los (CORREIA, 2016), sendo, ainda, capazes de isolar ruídos (DE SOUZA Jr., 2005).

As RNA permitem, assim, a modelagem de sistemas complexos para os quais o uso de métodos matemáticos, analíticos e fenomenológicos convencionais não são viáveis. Essa é uma das principais vantagens das redes neuronais, que permite sua aplicação a processos e plantas industriais que apresentam dinâmicas complexas não lineares e com incertezas (DE SOUZA Jr., 2005), gerando problemas de modelagem e de controle cada vez mais desafiadores.

Além da possibilidade de resolução de problemas complexos, as RNA apresentam inúmeras outras vantagens:

- Capacidade de aprendizado a partir de exemplos mesmo na presença de ruídos, como já mencionado. Isso implica que a rede pode ser treinada apenas a partir dos dados históricos da própria planta;
- Processamento em paralelo e não sequencial, diferente de outros métodos algorítmicos (ANTSAKLIS, 1990);
- Robustez, facilidade de uso (CORREIA, 2016) e maior flexibilidade pela facilidade de generalização para outras plantas (DE SOUZA Jr., 2005);
- Tolerância a falhas, pois o comprometimento de alguns neurônios não implica, necessariamente, em queda expressiva de performance (ANTSAKLIS, 1990). Isto é possível pela distribuição dos dados na rede;
- Aplicação como sensor virtual, prevendo o valor de uma variável cuja medida é inviável na prática.

Infelizmente, as RNA apresentam algumas limitações como o longo tempo de treinamento e a necessidade da escolha correta das variáveis de processo que serão usadas

como entradas para alimentação da rede, geralmente baseada em tentativa e erro (BAUGHMAN e LIU, 1995). É importante ressaltar que as RNA podem ser vulneráveis à utilização de dados que se encontram fora da faixa utilizada para treinamento e teste (DE SOUZA Jr., 2005).

O desenvolvimento de uma determinada rede para resolução de problemas, de forma resumida, inicia-se com a escolha de sua topologia e sua arquitetura, incluindo a definição da quantidade de neurônios e camadas. Em seguida, é realizado o treinamento da rede escolhida a partir de um conjunto de dados de entrada (variáveis disponíveis) e saídas (variáveis que devem ser preditas). Por fim, há uma etapa de validação a partir de outro conjunto de dados que verificará a eficiência da rede a partir do erro entre o valor predito e o real.

Uma das características das RNA é que a etapa de maior complexidade do desenvolvimento se refere a esta fase – seleção da topologia e treinamento. Ao final deste processo obtém-se uma RNA parametrizada que faz cálculos em alta velocidade a partir de equações e parâmetros já definidos.

4.2 Histórico

O primeiro estudo de grande importância sobre as Redes Neurais Artificiais foi realizado em 1943, por McCulloch e Pitts, que propuseram o primeiro modelo matemático de um neurônio artificial. Eles concluíram, por um lado, que um neurônio só está ativado, ou seja, só é capaz de transmitir a outros neurônios sua saída quando esta excede um determinado valor. Por outro lado, eles introduziram o conceito de sinapses inibitórias a partir da ideia dos pesos sinápticos que ponderam as entradas de cada neurônio. O neurônio é excitado quando o peso é superior a zero e inibido quando ele é negativo (CARDON et al., 1994).

Em 1949, Donald O. Hebb contribuiu para a evolução das RNA a partir do conceito não formalizado matematicamente do ajuste e atualização dos pesos sinápticos durante a etapa de aprendizado (ou treinamento). Ele também afirmou que a informação em uma rede é armazenada nos próprios pesos.

O primeiro modelo de RNA, chamado *perceptron*, foi implementado em 1958 por Frank Rosenblatt, revolucionando os estudos na área. O *perceptron* possui uma arquitetura simples com uma camada de saída e outra de entrada. O algoritmo de

aprendizagem supervisionado se baseia na atualização dos pesos a partir do erro entre a saída desejada e a saída calculada pela rede.

O modelo *perceptron* foi criticado por Minsky e Papert no livro *Perceptrons* em 1969 pela sua simplicidade excessiva e pelo fato de um modelo com apenas as camadas de entrada e saída não ser capaz de simular determinadas funções simples como a operação lógica “ou exclusivo” (CARDON et al., 1994) ou de classificar padrões não linearmente separáveis (DE SOUZA Jr., 2005).

Embora os estudos sobre as RNA tenham perdido força após a publicação de *Perceptrons*, alguns trabalhos de grande importância foram realizados nesse período, como o novo modelo de mapa auto-organizável proposto por Kohonen em 1972. A principal contribuição deste trabalho foi o estudo de uma rede com um grande conjunto de nós que interagem entre si para gerar uma representação de padrões de entrada. Esta rede utiliza um algoritmo de aprendizagem não-supervisionado, isto é, em que não é necessário fornecer à rede as saídas desejadas.

No início da década de 80, o interesse nas RNA voltou a crescer e novos tipos de treinamentos foram desenvolvidos motivados pelo avanço e aperfeiçoamento dos computadores digitais. Em 1982, o modelo de Hopfield foi proposto com conexões do tipo recorrente e, em 1986, Rumelhart e McClelland desenvolveram o importante método de aprendizagem chamado retropropagação (em inglês *backpropagation*), em que o erro entre o valor da saída calculado pela rede e o desejado é retroalimentado à rede para ajuste dos pesos sinápticos.

Em 2006, o estudo das RNA ganhou um novo impulso com o desenvolvimento da rede “*deep belief*” por Hinton. Esta rede utiliza o algoritmo de treinamento não supervisionado “*greedy layer-wise pre-training*” que se destaca por não necessitar de arquiteturas diferenciadas com conexões recorrentes, por exemplo. A partir de então as pesquisas se voltaram para a busca de técnicas de treinamento não supervisionado e para melhora do desempenho de redes treinadas com pequenos conjuntos de dados. Entretanto, a digitalização das indústrias e a maior facilidade de aquisição de dados operacionais, assim como o avanço no desempenho de computadores com unidades de processamento cada vez mais rápidas, provocaram uma mudança no foco dos estudos. A principal frente de pesquisa passou a ser o aperfeiçoamento dos métodos de treinamento supervisionado já existentes e o tratamento de grandes conjuntos de dados por redes mais profundas, com mais conexões entre neurônios (GOODFELLOW et al., 2016).

Desde então, o aprimoramento das técnicas de redes neuronais permitiu a melhora do desempenho das redes, capazes de gerar modelos de maior complexidade com maior precisão. Isto possibilitou a aplicação das RNA com sucesso em diversas áreas industriais e comerciais (MACUKOW, 2016), sendo empregas em grandes empresas, tais como Google, Microsoft e IBM. Atualmente, as redes neuronais contribuem para diversos setores tecnológicos e para domínios do conhecimento científico, tais como neurociência, química, farmacologia e processamento de imagens (GOODFELLOW et al., 2016).

4.3 Arquitetura das Redes Neuronais Artificiais

Os neurônios biológicos são formados pelo corpo celular, pelos dendritos e pelo axônio conforme mostra a Figura 18. Os sinais nervosos são enviados do axônio de uma célula aos dendritos de outra, através de conexões chamadas sinapses. Em alguns órgãos, como o cérebro, diversos neurônios se ligam formando redes complexas.

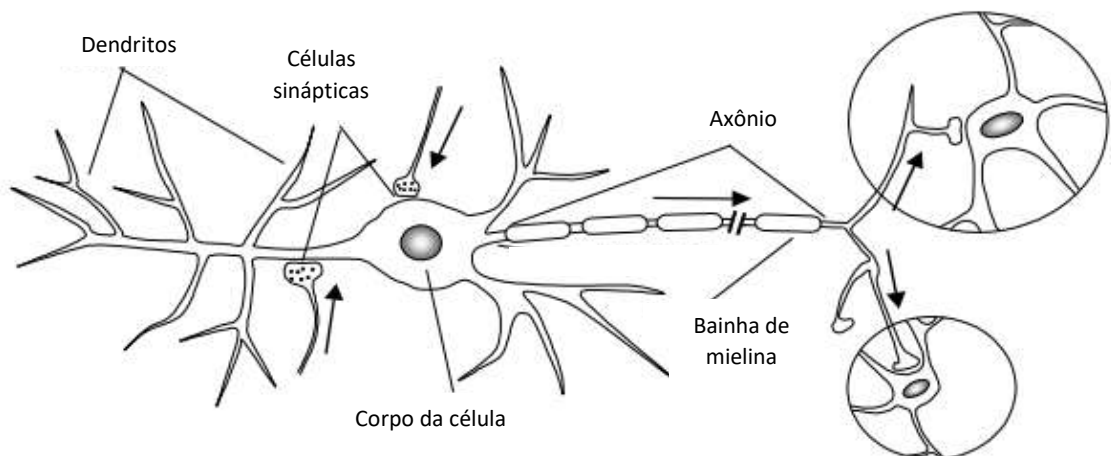


Figura 18: Esquema de um neurônio biológico. FONTE: CASTRO (2014).

As redes neuronais se inspiram nesse modelo e são formadas por unidades de processamento não lineares, chamadas neurônios ou nós, que são organizados em diferentes camadas e são interconectados (ANTSAKLIS, 1990).

A Figura 19 mostra um modelo de neurônio artificial em uma rede multicamadas e sumariza a nomenclatura utilizada para os sinais e para os pesos sinápticos. Considerando o j -ésimo neurônio presente na camada $(k+1)$, os sinais de entrada desse neurônio correspondem às saídas dos neurônios da camada anterior (k) representadas por $s_{pi,k}$ ($i = 1, \dots, n_k$), em que n_k é o número de neurônios na camada anterior. Esses sinais são ligados ao neurônio j e multiplicados pelos pesos sinápticos w_{jik} . É feita, então, uma

soma dos sinais de entrada ponderados, à qual é adicionado, ainda, um termo bias (θ_k), que garante um maior grau de liberdade às redes (OTERO, 1998) e modifica o valor de entrada no neurônio influenciando na sua excitação. Essa soma será aqui representada por $\lambda_{pj,k+1}$. Por fim, o resultado da soma é transformado em sinal de saída ($s_{pj, k+1}$) pela chamada função de ativação $f(\cdot)$ (DE SOUZA Jr., 2005). O índice p corresponde ao padrão (vetor de entrada) que está sendo tratado pela rede.

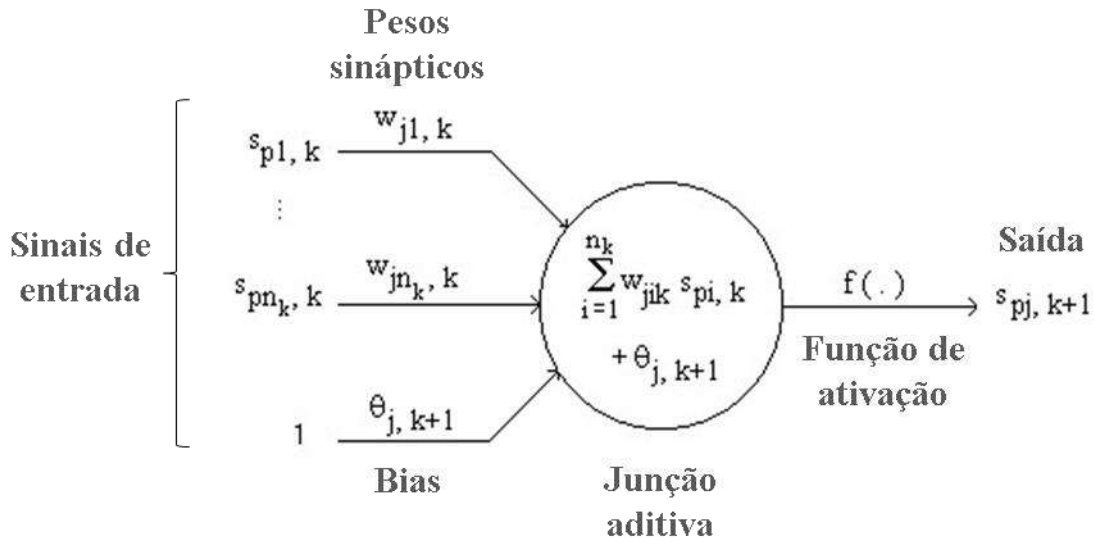


Figura 19: Modelo do j -ésimo neurônio artificial de uma rede multicamadas. Adaptado de DE SOUZA Jr., 2005.

Dessa forma, pode-se escrever matematicamente a saída em função das entradas de acordo com a Equação (1), que mostra que o cálculo da rede neuronal artificial é rápido e não requer iterações:

$$s_{pj,k+1} = f(\lambda_{pj,k+1}) = f \left[\left(\sum_{i=1}^{n_k} w_{jik} \times s_{pi,k} \right) + \theta_{j,k+1} \right] \quad (1)$$

Neste estudo, os sinais de entrada dos neurônios da primeira camada serão tratados pelo vetor $\underline{x}(x_1, x_2, \dots, x_n)$.

É possível dividir as redes em componentes físicos, compostos pelas conexões e unidades de processamento, e não físicos, que são os padrões e as funções (CARDON et al., 1994). Os padrões são os dados de entrada que alimentam a rede e que correspondem a uma determinada saída. Eles podem ser compostos de variáveis de diferentes tipos: contínuas, binárias ou categóricas. As funções são modelos matemáticos usados para transferência de sinais ou para o aprendizado. No primeiro caso, elas são chamadas de

funções de ativação e determinam a intensidade dos sinais enviados de um neurônio para o outro. Já no segundo, elas têm por objetivo o ajuste dos pesos durante o processo de treinamento.

A arquitetura de uma RNA consiste da estrutura de comunicação dos nós, ou seja, da forma como eles estão interconectados. Ela engloba a definição da quantidade de camadas e dos tipos de conexão, incluindo o fluxo de informações.

As conexões entre neurônios determinam o fluxo e a modulação da informação (CARDON et al., 1994). Comumente, a ligação pode ser classificada como “inter-camadas”, “intra-camadas” ou recorrente. As conexões “inter-camadas” são realizadas entre neurônios de diferentes camadas, as “intra-camadas” entre neurônios de uma mesma camada e as conexões recorrentes são aquelas em que a saída de um neurônio é enviada ao próprio neurônio.

É possível, ainda, classificar as conexões em função do sentido do fluxo de informação. Dessa forma, existem dois tipos: alimentação à frente (*feedforward*) ou retro-alimentação (*feedback*). Essas classificações correspondem, respectivamente, à transmissão de sinais de entrada para os neurônios de saída (sentido positivo) e ao fluxo contrário de informações, ou seja, da saída para a entrada (sentido negativo).

As redes neuronais costumam ser completamente conectadas, ou seja, todas as saídas dos neurônios de uma camada estão conectadas a todos os neurônios da camada seguinte. Redes nas quais isso não ocorre são classificadas como parcialmente conectadas.

Na arquitetura de rede mais comumente usada, chamada *Feedforward* (CORREIA, 2016), os nós são dispostos em camadas, com conexões do tipo inter-camadas, e o fluxo de informações é do tipo alimentação à frente, sendo assim, unidirecional, seguindo dos neurônios da camada de entrada para os da camada de saída, passando pelas demais camadas intermediárias, chamadas de camadas ocultas ou escondidas (em inglês, *hidden layers*). Sendo uma rede completamente conectada, os nós de cada camada recebem os sinais de todos os neurônios da camada anterior conforme mostrado na Figura 20.

A quantidade de nós presentes nas camadas de entrada e saída são estabelecidos de acordo com o problema analisado e são iguais, respectivamente, ao número de variáveis relevantes de entrada e ao número de variáveis que se deseja prever. Normalmente, a função de ativação dos neurônios da camada de entrada é a função

identidade e, portanto, a informação transmitida para os neurônios da camada seguinte não é alterada.

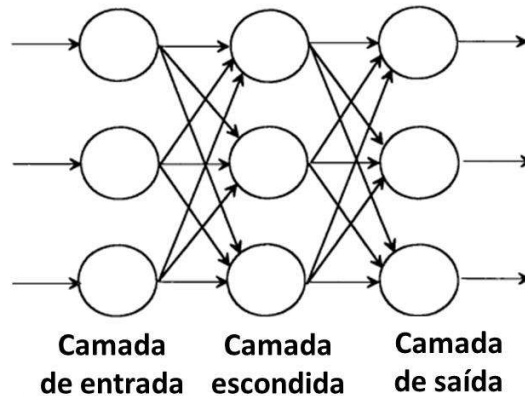


Figura 20: Esquema de fluxo de informações em uma rede *feedforward*. FONTE: BAUGHMAN e LIU, 1995.

O número de neurônios é um fator determinante para a capacidade de generalização da rede, ou seja, para a eficiência de sua predição a partir de entradas diferentes daquelas usadas para o treinamento. Caso a rede possua um número excessivo de neurônios, ocorre o fenômeno chamado de sobre-ajuste, *overfitting* ou *overtraining* no qual o erro encontrado para os dados de treinamento é pequeno, mas para demais dados de entrada o erro se torna significativo. Já para uma quantidade insuficiente de neurônios, o treinamento é comprometido e a convergência pode não ser atingida (CORREIA, 2016). A complexidade e a flexibilidade da rede é função da quantidade de neurônios na camada oculta.

4.3.1 Principais Redes Neurais Artificiais

Dentre as arquiteturas de redes neurais artificiais com maior aplicação e importância no contexto atual, destacam-se as redes MLP (*Perceptrons* Multicamadas), as redes RBF (*Radial Basis Function*) e rede de mapeamento auto-organizável. Essas três arquiteturas serão discutidas a seguir.

A arquitetura mais utilizada para a resolução de problemas de predição é a chamada *Perceptrons* Multicamadas (MLP, em inglês *Multi-Layer Perceptron*), na qual as redes podem apresentar uma ou mais camadas ocultas. Nessa arquitetura, a rede é caracterizada pela quantidade de nós em cada camada assim como pela quantidade de camadas e o tipo de conexão existente entre os neurônios. Ela apresenta um fluxo de

informação do tipo *feedforward* e seu treinamento é supervisionado, sendo normalmente utilizado o algoritmo de retropropagação (*backpropagation*). O treinamento de redes será tratado em detalhes adiante, na seção 4.5. Uma rede MLP com uma única camada oculta tem a topologia mostrada na Figura 20.

Outra arquitetura importante, mais utilizada para problemas de classificação de dados e predição de séries temporais, é a RBF (do inglês *Radial Basis Function*) que também é fundamentada na topologia *feedforward* e, assim como as redes MLP, apresenta um treinamento do tipo supervisionado. As redes RBF apresentam apenas três camadas, ou seja, apenas uma camada oculta.

A principal diferença em relação às redes MLP é a modelagem da relação entre as entradas e as saídas. Ao contrário das redes MLP, o processo de treinamento consiste em duas etapas. Na primeira etapa, a camada oculta realiza a modelagem da distribuição de probabilidade dos dados de entrada e na segunda ocorre a aprendizagem propriamente dita da relação entre entradas e saídas, com o ajuste dos pesos sinápticos. Outra diferença importante é que os neurônios da camada oculta das redes RBF não recebem um termo bias, somente os neurônios da camada de saída.

O processo de treinamento é fundamentado na formação de agrupamentos (*clusters*) dos padrões de entrada, formado no espaço multidimensional das entradas. Cada agrupamento possui um centro e é possível, portanto, conhecer a distância de um determinado conjunto de dados de entrada do centro de cada “*cluster*”. Dessa forma, os neurônios calculam a distância euclidiana entre o vetor de entrada (de um conjunto de dados) e um vetor referência, que representa a localização do centro da função de base radial no espaço das entradas. Essa distância é, então, usada para classificar o vetor de entrada, apontando a qual agrupamento ou “*cluster*” ele pertence. Assim, além dos pesos sinápticos, as redes RBF admitem outros dois parâmetros: o centro da função radial e sua dispersão. A função de base radial mais usada é a função gaussiana que será tratada na seção 4.4.

As funções radiais apresentam uma natureza local, ou seja, elas geram saídas ativas apenas na vizinhança de seu centro de ativação. Dessa forma, as redes RBF costumam apresentar um maior número de neurônios na camada oculta que as redes MLP, uma vez que mais neurônios são necessários para atender ao espaço multidimensional das entradas.

A Figura 21 apresenta um esquema de uma rede RBF.

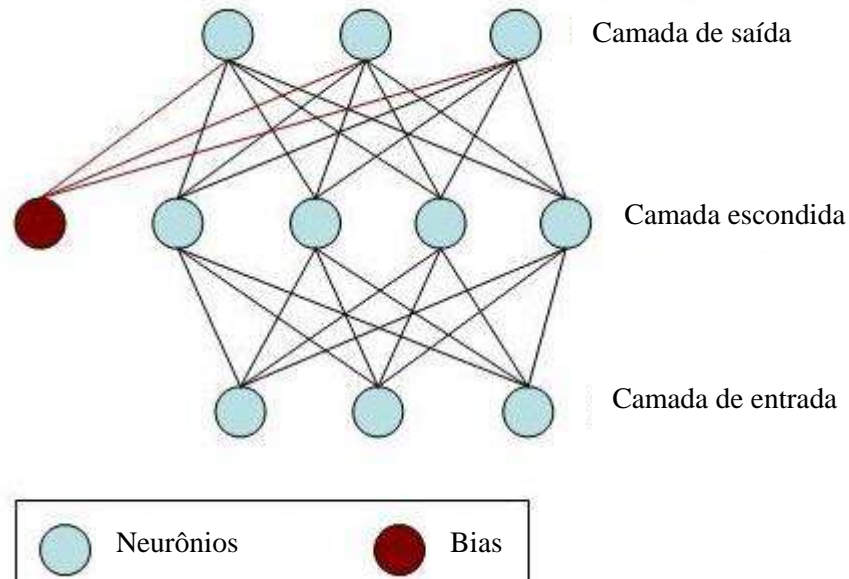


Figura 21: Esquema de uma rede RBF. FONTE: STATISTICA, 2004.

Por fim, outra arquitetura de grande importância é a de mapeamento auto-organizável ou SOM (do inglês, *Self-Organized Map*) que se destaca por ser uma rede do tipo não-supervisionada, ou seja, que não recebe os valores de saída desejada, mas apenas os dados de entradas durante o treinamento.

O princípio de funcionamento dessa rede é baseado na organização do córtex humano, em que cada região é excitada por sensações diferentes e específicas (ROSA e VALLE, 2012). Dessa forma, em uma rede de mapeamento auto-organizável, entradas diferentes excitam neurônios diferentes de tal modo que neurônios que se encontram próximos uns aos outros são ativados por padrões (entradas) apresentando similaridades.

Em outras palavras, a rede de mapeamento auto-organizável realiza a modelagem da distribuição de probabilidade dos vetores de entrada e agrupa, em um mapa, vetores similares. Assim, vetores próximos geometricamente estão localizados em pontos próximos no mapa e são separados em uma mesma classe.

Ao contrário das redes MLP e RBF que são organizadas em camadas, a SOM é uma rede de camada única em forma de grade uni ou bidimensional, na qual os neurônios se posicionam lado a lado, formando vizinhanças. Existem conexões do tipo intra-camada, ou laterais, entre os neurônios, que têm a capacidade de excitar neurônios na sua vizinhança topológica. A camada única é, portanto, uma representação da saída de rede de mapeamento auto-organizável. A Figura 22 mostra a estrutura característica das redes SOM.

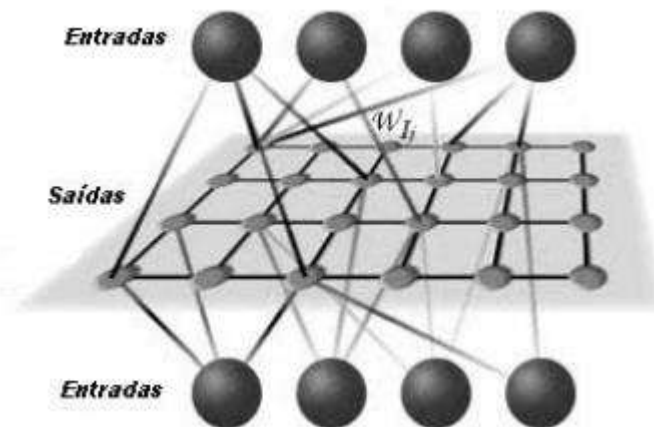


Figura 22: Estrutura de uma rede de mapeamento auto-organizável, em que os neurônios de saída estão organizados em uma grade bidimensional. FONTE: ALVAREZ e LUQUE, 2003.

O treinamento da rede é dividido basicamente em três etapas. Primeiramente, na etapa de competição, os neurônios são estimulados pelos dados de entrada e competem entre si de tal forma que o neurônio vencedor (seguindo uma abordagem do tipo “o vencedor leva tudo” ou do inglês “*winner-takes-all*”) seja aquele mais similar e que melhor represente a informação de entrada. Para determinação da similaridade entre neurônio e entrada, são empregadas funções discriminantes, sendo a mais usada a distância euclidiana entre os pontos que representam, respectivamente, a entrada e o neurônio. Em seguida, ocorre o chamado processo de cooperação, no qual o neurônio vencedor ativa somente os neurônios vizinhos. Na última etapa, no processo de adaptação sináptica, os pesos sinápticos desses neurônios excitados são ajustados (ROSA e VALLE, 2012).

Em resumo, as entradas, que podem apresentar n dimensões, são projetadas em um mapa de dimensão inferior (uni ou bidimensional) e a rede pode ser interpretada como uma transformação não-linear permitindo essa organização. Quando uma nova entrada é apresentada à rede após o treinamento, ela é representada no mapa auto-organizável e incorporada ao *cluster* (agrupamento) mais próximo.

4.4 Função de Ativação

A função de ativação atua como uma função de transferência que transforma a soma ponderada dos sinais de entrada em sinal de saída do neurônio, limitando sua amplitude (CORREIA, 2016) e introduzindo não-linearidades.

As funções de ativação podem ser consideradas globais ou locais (DE SOUZA Jr., 2005) de acordo com o intervalo de valores de entrada que elas admitem. As funções locais geram saídas ativas somente para valores próximos aos dados de entrada na vizinhança do seu centro de ativação, enquanto as funções de ativação globais admitem um intervalo grande de valores das entradas.

As funções de ativação mais largamente empregadas e que estão disponíveis no software de geração de redes neuronais usado no presente estudo, o STATISTICA®, são detalhadas abaixo:

- **Identidade:** Muito utilizada para neurônios das camadas de entrada, a função identidade, como o próprio nome sugere, realiza uma multiplicação do sinal de entrada pela unidade (número um), fazendo com que a saída possua o mesmo valor da entrada. É considerada uma função global.
- **Função Logística ou Sigmoidal:** É uma função global cujo gráfico apresenta uma curva em forma de “S” com valores entre 0 e 1 como mostrado na Figura 23. A equação matemática dessa função é (Equação (2)):

$$s_{pj,k+1} = f(\alpha \cdot \lambda_{pj,k+1}) = \frac{e^{\alpha \cdot \lambda_{pj,k+1}}}{e^{\alpha \cdot \lambda_{pj,k+1}} + 1} = \frac{1}{1 + e^{-\alpha \cdot \lambda_{pj,k+1}}} \quad (2)$$

Em que α é um número real positivo.

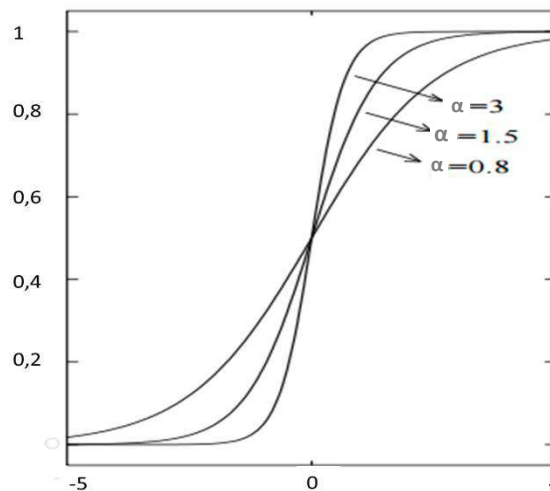


Figura 23: Função logística com fator α igual a 3, 1,5 e 0,8. FONTE: CARDON e MÜLLER, 1994.

- **Tangente Hiperbólica:** Essa função global é muito utilizada para camadas ocultas de redes MLP por apresentar, normalmente, melhores desempenho devido

à sua simetria (STATISTICA, 2004). A forma da curva da função tangente hiperbólica se assemelha àquela da logística, como mostrado na Figura 24. No entanto, sua saída possui valores entre -1 e 1, sendo a “média” igual a zero. A equação matemática é fornecida na Equação (3):

$$s_{pj,k+1} = f(\alpha \cdot \lambda_{pj,k+1}) = \tanh(\alpha \cdot \lambda_{pj,k+1}) = \frac{e^{\alpha \cdot \lambda_{pj,k+1}} - e^{-\alpha \cdot \lambda_{pj,k+1}}}{e^{\alpha \cdot \lambda_{pj,k+1}} + e^{-\alpha \cdot \lambda_{pj,k+1}}} \quad (3)$$

Em que α é um número real positivo.

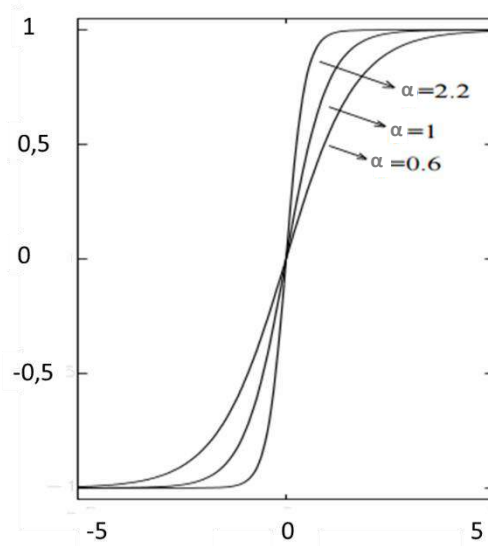


Figura 24: Função tangente hiperbólica com fator p de 2,2, 1, e 0,6. FONTE: CARDON e MÜLLER, 1994.

- **Exponencial:** Outra função de ativação comum é a exponencial negativa segundo a equação matemática abaixo (Equação (4)):

$$s_{pj,k+1} = f(\alpha \cdot \lambda_{pj,k+1}) = \exp(-\alpha \cdot \lambda_{pj,k+1}) = e^{-\alpha \cdot \lambda_{pj,k+1}} \quad (4)$$

Em que α é um número real positivo.

- **Seno:** O software STATISTICA® não utiliza a função seno como parâmetro padrão como faz com as demais funções, mas permite que o usuário a selecione caso haja um conhecimento prévio de sua eficiência para o problema em questão. Caso contrário, seu uso não é altamente recomendado pelos desenvolvedores do software (STATISTICA, 2004).
- **Gaussiana:** A função gaussiana é uma função do tipo local, ou seja, ela produz uma resposta ativa somente na vizinhança do seu centro de ativação (DE SOUZA Jr., 2005). Ela é usada somente para camadas ocultas de redes RBF e pode ser

escrita matematicamente da seguinte forma (Equação (5)) para o neurônio j da camada $(k+1)$:

$$\prod_{i=1}^{n_k} e^{-\frac{(s_{pi,k} - \mu_{ji})^2}{2\sigma_{j1}^2}} \quad (5)$$

Em que μ_{ji} são os centros da função gaussiana e σ_j , o desvio-padrão (dispersão) (DE SOUZA Jr, 1993).

- **Softmax:** A função Softmax é muito usada para problemas de classificação com mais de uma saída e permite que a rede interprete as saídas como probabilidades, uma vez que os valores de saída são normalizados para que sua soma seja unitária (HOCHULI, 2016). A função Softmax é dada pela Equação (6) seguir.

$$f(\lambda_{pj,k+1}) = \frac{e^{\lambda_{pj,k+1}}}{\sum_{m=1}^{N_k} e^{\lambda_{pm,k+1}}} \quad (6)$$

4.5 Treinamento

O treinamento de uma rede neuronal é o processo de aprendizagem a partir de padrões contidos em um conjunto de dados de entrada. Esse processo confere à rede a sua capacidade de aprendizado a partir de exemplos e a capacidade de generalização para valores de entrada “inéditos” (que ela não conhece previamente). O processo consiste no ajuste dos parâmetros da rede, tais como pesos sinápticos e bias de tal forma a otimizar os valores de saída da rede.

Existem dois tipos de treinamento, o supervisionado e o não supervisionado, que diferem entre si pela presença ou ausência de dados de saída nos dados apresentados à rede durante o treinamento.

Nos métodos de treinamento supervisionados, as saídas desejadas são conhecidas e a modificação dos pesos é realizada para atingir valores próximos aos dessas saídas. Para tal, o ajuste é feito com base na minimização de uma função erro entre a saída calculada pela rede e aquela desejada. As funções erro utilizadas no presente trabalho serão abordadas na seção 5.2.3.

Já nos métodos não supervisionados, a adaptação dos pesos se baseia na representação da distribuição de probabilidade ou na correlação dos padrões de entrada, já que somente dados de entrada são fornecidos à rede durante o treinamento. A rede é autônoma e busca propriedades similares entre as entradas para separá-las em agrupamentos. Estes métodos são utilizados para problemas de classificação e agrupamento dos dados (CORREIA, 2016).

O treinamento supervisionado, na prática, se baseia em algoritmos iterativos que modificam gradualmente os pesos sinápticos até otimizar seus valores e permitir que a rede seja capaz de gerar saídas o mais próximo possível do esperado.

O treinamento começa com a inicialização dos pesos, que devem assumir valores pequenos para que eles aumentem gradualmente durante o treinamento e possam ajustar os dados. O software de geração de redes neuronais STATISTICA® usado no presente estudo fornece duas opções de inicialização de pesos de forma aleatória. Na primeira, os valores dos pesos assumem valores aleatórios com distribuição normal (média igual a zero e desvio padrão de um). A segunda opção, a distribuição dos pesos é uniforme variando de zero a um (STATISTICA, 2004). Rumelhart e McClelland (1986) recomendam pesos aleatoriamente distribuídos entre -1 e 1.

O processo iterativo possui 5 etapas para cada iteração, ciclo ou época. Ele é iniciado com a apresentação de um conjunto de dados de entrada e a saída correspondente (1). A rede, a partir do seu estado atual (pesos e bias), realiza a predição para a saída (2). A função erro selecionada calcula a diferença entre o valor desejado e aquele predito (3). Essas etapas são repetidas até que todos os dados de treinamento sejam fornecidos à rede (4). Um algoritmo de treinamento é utilizado para que o ajuste dos pesos sinápticos seja realizado de acordo com o erro calculado (5). São realizadas tantas iterações quanto necessárias para que a eficiência da rede seja satisfatória.

Os principais algoritmos de treinamento usados para as redes tratadas neste estudo são apresentados a seguir.

4.5.1 Método de retropropagação (*backpropagation*)

O método de retropropagação foi desenvolvido de forma independente por diferentes especialistas, porém ele foi popularizado após a publicação dos estudos de Rumelhart, Hinton e Williams, em 1986. Desde então, ele se tornou o principal algoritmo de treinamento supervisionado usado em redes neuronais (DE SOUZA Jr, 2005).

Este método é aplicado para redes *feedforward*, com conexões do tipo intercamadas e com múltiplas camadas e ele se baseia no cálculo do gradiente do erro em relação a um determinado peso ou bias, o que requer que as funções de ativação sejam continuamente diferenciáveis (DE SOUZA Jr., 2005).

O método de retropropagação se dá em duas fases. Primeiramente, o conjunto de dados de entradas é propagado pela rede e uma função erro, que será apresentada posteriormente, é calculada. A segunda etapa explica o nome do método, uma vez que o erro é propagado no sentido contrário (*feedback*) para que a atualização dos pesos seja efetuada. É importante lembrar que uma etapa de inicialização dos pesos, anterior às duas fases referidas acima, é necessária.

Conforme dito anteriormente, o treinamento consiste em minimizar o erro entre a saída desejada (t , de *target*, em inglês) e a calculada pela rede (s) a partir da modificação dos pesos e bias. Matematicamente, isso implica na minimização da função erro descrita pela Equação (7), em que N_p é o número de padrões (conjunto de dados), L a camada de saída e o índice l , o neurônio correspondente da camada de saída.

$$E(W) = \sum_{p=1}^{N_p} E_p = \sum_{p=1}^{N_p} \frac{1}{2} \sum_{l=1}^{N_L} (s_{pl,L} - t_{pl})^2 \quad (7)$$

Para que o erro seja minimizado, tem-se a chamada função objetivo cujo gradiente deve ser igual ao vetor nulo de acordo com a Equação (8), em que W é o vetor dos pesos sinápticos.

$$\underline{\nabla} E(W) = \underline{0} \quad (8)$$

Ajuste dos pesos

Uma das estratégias de ajuste dos pesos mais conhecida é a chamada gradiente descendente, utilizada por Rumerhart e McClelland (1986), na qual a cada iteração os pesos são modificados pelo negativo do gradiente do erro multiplicado por um passo η , positivo e normalmente inferior a 1. Esse passo é chamado de taxa de aprendizagem. Matematicamente, o ajuste dos pesos é descrito pela Equação (9), em que \underline{W}_k é o vetor dos pesos sinápticos na k -ésima iteração e E é a função erro.

$$\Delta \underline{W}_{k+1} = -\eta \underline{\nabla} E(\underline{W}_k) \quad (9)$$

Entretanto, a estratégia de gradiente descendente apresenta limitações. A desvantagem do passo fixo é a dificuldade da adoção de um valor para a taxa de

aprendizagem. Para valores elevados, ocorrem oscilações nos pesos, enquanto para valores pequenos, a convergência do método torna-se excessivamente lenta.

Para reduzir os problemas desta estratégia devido a oscilações, propôs-se a inclusão de um termo de momento de acordo com a Equação (10), em que ε é fixo.

$$\Delta \underline{W}_{k+1} = -\eta \underline{\nabla} E(\underline{W}_k) + \varepsilon \Delta \underline{W}_k \quad (10)$$

No entanto, as oscilações não foram superadas por esse método e outros foram desenvolvidos ao longo dos anos. Um exemplo é o método do gradiente com direções conjugadas de Leonard e Kramer (1990), que propõe o uso de η e ε como fatores dinâmicos e não fixos. Apesar da maior demanda computacional, o método permite ultrapassar os problemas com oscilações. As Equações (11) e (12) abaixo mostram como o método calcula recursivamente a saída s a cada iteração (indicada pelos índices).

$$\underline{s}_0 = -\underline{\nabla} E(\underline{W}_0) \quad (11)$$

$$\underline{s}_{k+1} = -\underline{\nabla} E(\underline{W}_{k+1}) + \varepsilon \underline{s}_k \quad (12)$$

Os vetores dos pesos sinápticos e o fator ε são calculados, respectivamente, de acordo com as Equações (13) e (14).

$$\underline{W}_{k+1} = \underline{W}_k + \eta \underline{s}_k \quad (13)$$

$$\varepsilon = \frac{\underline{\nabla} E(\underline{W}_{k+1})^T \underline{\nabla} E(\underline{W}_{k+1})}{\underline{\nabla} E(\underline{W}_k)^T \underline{\nabla} E(\underline{W}_k)} \quad (14)$$

Há ainda, os métodos de Newton e os quase-Newton, que trabalham, respectivamente, com a matriz Hessiana⁴ analítica e com aproximações desta matriz. Um exemplo de método quase Newton é o BFGS (Broyden-Fletcher-Goldfarb-Shanno) (DE SOUZA Jr., 1993). O software STATISTICA[®], utilizado no presente estudo, emprega e testa diferentes algoritmos de treinamento, tais como o gradiente descendente, o gradiente com direções conjugadas e o BFGS.

O método de retropropagação

Será utilizada como exemplo ilustrativo para explicar o método de retropropagação neste estudo uma rede com três camadas e que usa a função sigmoideal (logística) como função de ativação. A escolha desse exemplo se apoia no fato de que,

⁴ Matriz Hessiana: Matriz das derivadas segundas parciais de uma função. No contexto, a matriz Hessiana contém as derivadas segundas da função objetivo em relação aos pesos sinápticos.

conforme provado por Hecht-Nielsen (1989), uma rede com três camadas, treinada por retropropagação e com número suficiente de neurônios é capaz de aproximar qualquer função contínua quando as funções de ativação dos neurônios ocultos são funções sigmoidais (DE SOUZA Jr.,1993). Além disso, dentre as funções de ativação empregadas em redes neuronais, a função sigmoial é a mais utilizada. Adotou-se a nomenclatura usada por De Souza Jr. (1993).

Como o método se baseia no gradiente do erro, é necessário conhecer a derivada da função sigmoial, que apresenta uma propriedade interessante: a sua derivada é igual à própria função subtraída de seu quadrado, conforme mostrado matematicamente pelas Equações (15) e (16):

$$f(x) = \frac{1}{(1 + e^{-x})} \quad (15)$$

$$f'(x) = \frac{e^{-x}}{(1 + e^{-x})^2} = f(x) - f^2(x) = f(x)(1 - f(x)) \quad (16)$$

Na primeira fase do treinamento, os sinais propagam para frente e é possível, com base no modelo de neurônio explicado anteriormente, calcular a saída dos neurônios da camada oculta, fornecida pela Equação (17).

$$s_{pj,2} = f(\lambda_{pj,2}) = f \left[\left(\sum_{i=1}^{n_i} w_{ji1} s_{pi,1} \right) + \theta_{j,2} \right] \quad (17)$$

Em que n_i é o número de neurônios na camada de entrada, n_h é o número de neurônios na camada oculta e $j=1,2,\dots,n_h$.

A saída da rede é, então, dada pela Equação (18):

$$s_{pj,L} = f(\lambda_{pj,L}) = f \left[\left(\sum_{i=1}^{n_h} w_{ji,L-1} \times s_{pi,L-1} \right) + \theta_{j,L} \right] \quad (18)$$

Em que n_h é o número de neurônios na camada oculta e $j=1,2,\dots,n_L$.

O erro é calculado pela Equação (7) e a segunda fase é iniciada com a sua propagação no sentido contrário (retropropagação). Os pesos e bias das conexões entre a camada oculta e a de saída são ajustados antes dos pesos e bias correspondentes às conexões entre a camada de entrada e a oculta.

A atualização dos pesos e bias pode seguir um dos métodos citados anteriormente nesta seção e é ilustrada a seguir com base no método do gradiente descendente.

As derivadas da função erro em relação aos pesos devem ser calculadas para a determinação do gradiente e da função objetivo utilizando a regra da cadeia. De acordo com DE SOUZA Jr., (1993) a derivada para as conexões entre a camada oculta e a de saída é dada pela Equação (19) e para as conexões entre a camada de entrada e a oculta pela Equação (20).

$$\frac{\partial E}{\partial w_{ji2}} = \sum_{p=1}^{N_p} \frac{\partial E_p}{\partial s_{pj,L}} \frac{\partial s_{pj,L}}{\partial \lambda_{pj,L}} \frac{\partial \lambda_{pj,L}}{\partial w_{ji2}} = \sum_{p=1}^{N_p} (s_{pj,L} - t_{pj}) f'(\lambda_{pj,L}) s_{pi,2} \quad (19)$$

$$\begin{aligned} \frac{\partial E}{\partial w_{ji1}} &= \sum_{p=1}^{N_p} \sum_{l=1}^{n_L} \frac{\partial E_p}{\partial s_{pl,L}} \frac{\partial s_{pl,L}}{\partial \lambda_{pl,L}} \frac{\partial \lambda_{pl,L}}{\partial s_{pj,2}} \frac{\partial s_{pj,2}}{\partial \lambda_{pj,2}} \frac{\partial \lambda_{pj,2}}{\partial w_{ji1}} = \\ &= \sum_{p=1}^{N_p} \left(\sum_{l=1}^{n_L} (s_{pl,L} - t_{pl}) f'(\lambda_{pl,L}) w_{lj2} \right) f'(\lambda_{pj,2}) s_{pi1} \end{aligned} \quad (20)$$

As derivadas para os bias da camada de saída ($\theta_{j,L}$) e da camada oculta ($\theta_{j,2}$) são fornecidas, respectivamente, pelas Equações (21) e (22).

$$\frac{\partial E}{\partial \theta_{j,L}} = \sum_{p=1}^{N_p} \frac{\partial E_p}{\partial s_{pj,L}} \frac{\partial s_{pj,L}}{\partial \lambda_{pj,L}} \frac{\partial \lambda_{pj,L}}{\partial \theta_{j,L}} = \sum_{p=1}^{N_p} (s_{pj,L} - t_{pj}) f'(\lambda_{pj,L}) \quad (21)$$

$$\begin{aligned} \frac{\partial E}{\partial \theta_{j,2}} &= \sum_{p=1}^{N_p} \sum_{l=1}^{n_L} \frac{\partial E_p}{\partial s_{pl,L}} \frac{\partial s_{pl,L}}{\partial \lambda_{pl,L}} \frac{\partial \lambda_{pl,L}}{\partial s_{pj,2}} \frac{\partial s_{pj,2}}{\partial \lambda_{pj,2}} \frac{\partial \lambda_{pj,2}}{\partial \theta_{j,2}} = \\ &= \sum_{p=1}^{N_p} \left(\sum_{l=1}^{n_L} (s_{pl,L} - t_{pl}) f'(\lambda_{pl,L}) w_{lj2} \right) f'(\lambda_{pj,2}) \end{aligned} \quad (22)$$

De forma iterativa, as duas fases (propagação e retropropagação) são repetidas até que o erro seja considerado aceitável, com base em uma tolerância pré-estabelecida, para que o treinamento seja interrompido.

4.5.2 Método de Treinamento Kohonen

Conforme dito anteriormente, o treinamento das redes de mapeamento auto-organizável compreende três etapas: competição, cooperação e adaptação sináptica. Entretanto, uma etapa preliminar é necessária para a inicialização dos pesos e do raio de vizinhança, que decresce com o tempo nas redes de mapeamento auto-organizável.

Cardon e Müller (1994) sugerem a metade do diâmetro da rede como raio de vizinhança de tal forma que ele diminua ao longo do treinamento.

Após inicialização dos pesos, é iniciado o processo competitivo que tem por objetivo a definição do neurônio vencedor. Para tal, um padrão de entrada é alimentado à rede e é realizada a busca pelo neurônio que minimiza a distância em relação ao vetor de entrada. Assim, as distâncias são calculadas entre o vetor de entrada (\underline{x}) e o vetor de pesos (\underline{w}) correspondente a cada neurônio. Geralmente, utiliza-se a distância euclidiana. A Equação (23) mostra o cálculo da distância (D_j) do neurônio j em relação à entrada \underline{x} de dimensão ($N \times 1$), sendo w_{ij} o peso entre a entrada i e o neurônio j , e n o número da iteração.

$$D_j(n) = \left\| \underline{x} - \underline{w}_j \right\| = \sqrt{\sum_{i=1}^N (x_i(n) - w_{ij}(n))^2} \quad (23)$$

Para encontrar o neurônio vencedor, deseja-se encontrar a distância mínima, ou seja, o neurônio vencedor $v(x)$ é dado pela Equação (24):

$$v(x) = \underset{j}{\operatorname{argmin}} \left\| \underline{x} - \underline{w}_j \right\| \quad (24)$$

Uma vez determinado o neurônio vencedor, inicia-se o processo cooperativo. Uma função $h_{j,v}$ define a vizinhança do neurônio vencedor, centrada nele próprio. Ela deve ser simétrica e atingir seu valor máximo no neurônio vencedor. Sua amplitude deve diminuir monotonicamente até ser nula conforme a distância lateral entre o vencedor e demais neurônios aumenta. Um exemplo de função que atende a essas exigências é a Gaussiana. Neste caso, o centro da gaussiana é nulo ($\mu=0$) e corresponde ao neurônio vencedor. Para que o raio da vizinhança diminua com o tempo de treinamento, conforme dito anteriormente, pode-se utilizar um desvio-padrão σ dinâmico, que diminui conforme o tempo aumenta. O decaimento pode ser linear ou exponencial.

Tem-se, então, o processo adaptativo para o ajuste dos pesos dos neurônios contidos na vizinhança topológica do neurônio vencedor. O ajuste é dado pela Equação (25) (CASTRO e CASTRO, 2014).

$$\underline{w}_j(n+1) = \underline{w}_j(n) + \eta(n) \times h_{j,v(\underline{x})}(n) \times (\underline{x}(n) - \underline{w}_j(n)) \quad (25)$$

Uma vez que os pesos tenham sido ajustados, retorna-se ao início do processo iterativo com a apresentação de um novo padrão à rede e as etapas de competição, cooperação e adaptação são repetidas até que não haja mais modificações significativas no mapa auto-organizável.

4.6 Teste e Validação

Conforme visto anteriormente, os dados usados na geração de uma rede neuronal são divididos em três grupos e cada um deles é destinado a um dos três processos a seguir: treinamento, teste e validação. Recomenda-se que a maioria dos conjuntos de dados seja destinada ao treinamento para garantir que a quantidade de exemplos seja suficiente para uma aprendizagem eficiente.

A etapa de teste é usada para evitar o sobre-ajuste (*overfitting*) e verificar a real capacidade de generalização da rede, ou seja, de modelar dados desconhecidos que nunca foram usados no treinamento. Dessa forma, o teste consiste em fornecer à rede, durante a evolução do treinamento, novos dados e verificar sua eficiência na predição ou classificação da variável de saída.

O teste contribui para o impedimento de sobre-ajuste com a técnica chamada de *early stopping* que permite a interrupção do treinamento da rede quando o erro do teste começa a aumentar mesmo enquanto o erro do treinamento está em declínio.

Assim, ao final de cada iteração do processo de treinamento, após o ajuste dos pesos, a rede é apresentada a todos os padrões do grupo de dados de teste para que a predição seja realizada e o erro de teste calculado. O erro é então comparado com aquele encontrado no teste da iteração anterior. Caso ele seja inferior, uma nova iteração de treinamento é realizada. Caso contrário, o treinamento é interrompido.

A Figura 25 mostra a evolução do erro de teste e de treinamento durante os ciclos (iterações) do processo de treinamento. Enquanto o erro de teste está sendo reduzido, o treinamento prossegue. Quando ele começa a aumentar o treinamento é encerrado (técnica de *early stopping*).

O processo de validação, por sua vez, consiste no uso de padrões inéditos para a comprovação da eficiência da rede. A validação é realizada somente no final do treinamento para uma verificação suplementar àquela feita no processo de teste. Por isso, quando o número de dados é limitado, é possível não executar a validação, visto que o uso de dados na validação implica na diminuição da quantidade de dados disponíveis para

os outros dois processos que estão, de fato, envolvidos na otimização da rede e de suas capacidades de aprendizagem e de generalização.

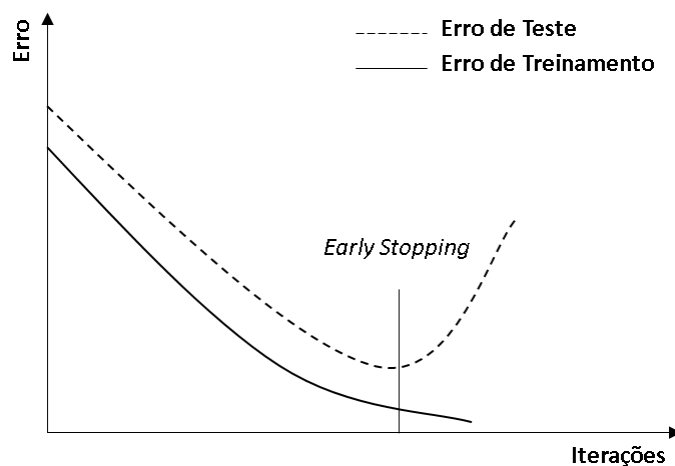


Figura 25: Representação gráfica ilustrativa do momento de interrupção do treinamento que corresponde ao início do aumento do erro de teste.

4.7 Aplicações

As Redes Neurais Artificiais estão ocupando cada vez mais um lugar de destaque nas pesquisas em diversas áreas da Tecnologia e da Indústria principalmente pela sua capacidade de modelar processos complexos e com não-linearidades.

São exemplos de aplicações bem-sucedidas e reconhecidas das RNA o seu uso em processamento de sinais, controle de processos, robótica, pré-processamento de dados, reconhecimento de padrões, análise de imagens, diagnóstico e monitoramento médico, reconhecimento de notas musicais para composições automáticas, previsões para o mercado de ações e na avaliação de empréstimos e créditos.

O uso das RNA se torna especialmente interessante para a Indústria Química, que envolve processos com dinâmicas complexas, cuja modelagem é extremamente difícil ou mesmo inviável. Isto se deve à presença inerente aos processos químicos de tempos mortos, ruídos e não-linearidades (DE SOUZA Jr., 2005).

Uma das aplicações mais interessantes e importantes das redes neuronais na Indústria Química é no controle de processos, principalmente para estratégias de controle preditivo. A maior parte das indústrias químicas dispõe de sistemas de controle que permitem armazenar conjuntos históricos de dados de processo, gerando milhares de registros (conjuntos de dados associados a um instante de tempo) por dia. A possibilidade de modelar processos com dinâmicas complexas a partir destes dados históricos e de gerar

predições satisfatórias para variáveis de difícil medição permite a criação de controladores eficientes baseados em redes neuronais. Segundo De Souza Jr. (2005), atualmente, diversas indústrias utilizam técnicas de redes neuronais. Um dos fatores de sucesso das redes neuronais é que a complexidade do algoritmo está na fase de treinamento, feita antes da configuração do controlador. A execução do modelo representado pela rede neuronal é rápida, atendendo aos requisitos de rápida resposta dos processos a serem controlados.

Outra aplicação das RNA com grande potencial para uso industrial é a sua utilização para geração dos chamados sensores virtuais. Os sensores virtuais são unidades de processamento que não são fisicamente ligadas à planta, mas que são capazes de estimar e calcular parâmetros de processo em tempo real apenas pelo processamento de outros dados operacionais fornecidos por transmissores no campo, tais como temperatura, vazão, pressão, etc. Esta aplicação se torna interessante para a predição de variáveis que não podem ser medidas no campo e em tempo real de maneira eficaz.

No caso da indústria de papel, destacam-se os parâmetros Kappa e *Freeness* que são, usualmente, determinados em análises laboratoriais, que, além de laboriosas, introduzem um tempo morto elevado, dificultando o seu controle em tempo real.

Modelos para o número Kappa utilizando redes neuronais foram estudados por diferentes pesquisadores como Dufour et al. em 2005 e Correia em 2016. Este último propôs a predição do número Kappa para madeiras folhosas a partir de redes neuronais artificiais. A melhor rede encontrada no estudo foi uma MLP (*Multi-Layer Perceptron*) que apresentou um erro percentual médio absoluto de 3,7%. A rede possui três camadas (de entrada, oculta e de saída), com três neurônios na camada oculta. As funções de ativação para as camadas oculta e de saída foram, respectivamente, a função tangente hiperbólica e a identidade. Foram usadas 11 variáveis de entrada, sendo elas: tempo de retenção dos cavacos no silo, velocidade de dosagem dos cavacos, carga alcalina total, temperatura do topo do digestor, álcali efetivo de circulação superior (no digestor), temperatura de circulação superior (no digestor), álcali efetivo de circulação inferior (no digestor), temperatura de circulação inferior (no digestor), fluxo do licor diluído no fundo e Kappa anterior. Já para a determinação do parâmetro *Freeness*, que pode ser utilizado no controle da etapa de refino da polpa celulósica, o uso de redes neuronais também vem sendo explorado (RADIX, 2017).

5. DADOS E MÉTODOS

O objetivo principal do presente trabalho é o desenvolvimento de um modelo para a estimativa do número de partículas formadas na caldeira de recuperação de um processo Kraft de produção de polpa de celulose, usando técnicas de “*Machine Learning*” por meio de redes neurais a partir de dados de operação. Para o treinamento das redes, foram utilizados dados históricos de processo de uma planta industrial conforme será explicado a seguir.

Foram adotadas três abordagens diferentes para a resolução do problema proposto de modelagem, tendo sido estudados três tipos de rede no total (MLP, RBF e de mapeamento auto-organizável). Em um primeiro momento, testou-se a eficiência de redes neurais MLP e RBF para predição do número de partículas formadas. Em seguida, estudou-se o emprego de redes do tipo RBF e MLP para classificação do número de partículas formadas em classes pré-estabelecidas de operação normal e de falha. Por fim, realizou-se uma análise de agrupamentos, ou análise *cluster*, com mapas auto-organizáveis.

Desta forma, foram testadas redes neurais com diferentes arquiteturas para a predição e classificação da variável de interesse, ou seja, a quantidade de partículas formadas na caldeira de recuperação. A metodologia usada para cada abordagem será descrita em detalhes nas seções 5.3 e 5.4.

Os resultados finais das melhores redes encontradas pelas três abordagens usadas serão comparados entre si e, também, aos resultados obtidos por Costa (2004), que realizou uma modelagem da caldeira Kraft a partir de redes neurais, conforme apresentado anteriormente na seção 3.4.1. Espera-se refinar o trabalho já realizado, explorando diferentes abordagens, e encontrar redes com melhores desempenhos que possam modelar a formação de partículas na caldeira de forma a permitir o seu uso na indústria.

5.1 Dados Operacionais da Caldeira de Recuperação Kraft

Neste estudo foram usados, para o treinamento e para o teste das redes neurais, os dados operacionais tratados por Costa (2004) e Costa et al. (2005), que tiveram por objetivo a modelagem da caldeira de recuperação Kraft de uma importante fábrica de papel e celulose brasileira e da formação de material particulado neste equipamento.

Os dados correspondem a doze meses de operação da caldeira de recuperação de uma planta para produção de polpa de celulose pelo processo Kraft e incluem 12 variáveis de processos e 3.381 pontos operacionais (ou conjuntos de dados). O nome da empresa será omitido por razões de confidencialidade, assim como os valores absolutos das variáveis de processo.

A obtenção dos dados foi realizada através do histórico do sistema de supervisão e aquisição de dados da indústria e as 12 variáveis operacionais disponibilizadas para este estudo serão detalhadas na Tabela 6.

Tabela 6: Variáveis de processo empregadas. Adaptado de COSTA, 2004.

TAG	Descrição	Local de medição	Unidades
Vvlu	Vazão volumétrica de licor negro alimentado à caldeira.	Alimentação de licor negro.	m ³ /h
x	Concentração de sólidos na alimentação do licor negro.	Alimentação de licor negro.	%
Tlu	Temperatura de alimentação do licor negro.	Alimentação de licor negro.	°C
Var1	Vazão de ar primário.	Alimentação de ar primário.	t/h
Var2	Vazão de ar secundário.	Alimentação de ar secundário.	t/h
Var3	Vazão de ar terciário.	Alimentação de ar terciário.	t/h
Par1	Pressão de alimentação do ar primário.	Alimentação de ar primário.	mmca
Par2	Pressão de alimentação do ar secundário.	Alimentação de ar secundário.	mmca
Tar1	Temperatura de alimentação do ar primário.	Alimentação de ar primário.	°C
Tar2	Temperatura de alimentação do ar secundário.	Sistema de alimentação de ar secundária.	°C
Pl	Pressão de alimentação do licor negro.	Sistema de alimentação de licor negro.	Kgf/cm ²
Epart	Número médio de partículas por minuto.	Média entre as quantidades medidas na entrada da região dos superaquecedores no sentido do fluxo dos gases do lado direito e esquerdo.	partículas/ min

Outras variáveis de processo tais como a vazão, a temperatura e a pressão do vapor superaquecido formado na caldeira foram utilizadas no trabalho de Costa (2004). Porém, de acordo com Costa (2004), as variáveis mais relevantes e que apresentam maior influência sobre a formação de material particulado são, justamente, as 11 utilizadas neste trabalho.

O número de partículas formadas no interior da caldeira é fornecido pela variável chamada Epart. Sua predição corresponderá à saída das redes que utilizam treinamento supervisionado testadas ao longo deste trabalho. O valor de Epart é obtido pela média entre os valores de número de partículas que são arrastadas com o fluxo dos gases, na fornalha, na entrada da região dos superaquecedores, do lado direito e esquerdo. A medição é feita por câmeras e posterior processamento de imagem.

Os dados operacionais coletados foram previamente tratados por Costa (2004) para retirada de dados não confiáveis, assim como de dados cujo valor da quantidade de partículas medida pelas duas câmeras apresentou diferenças significativas, superiores a 100 partículas/min. A quantidade de partículas (variável Epart), de acordo com os dados disponíveis, variou entre 9,62 e 806,12 (valores mínimo e máximo, respectivamente). O valor médio de Epart é igual a 96,99 e o desvio padrão a 61,01.

A Figura 26 mostra a evolução de Epart durante o período de aquisição de dados.

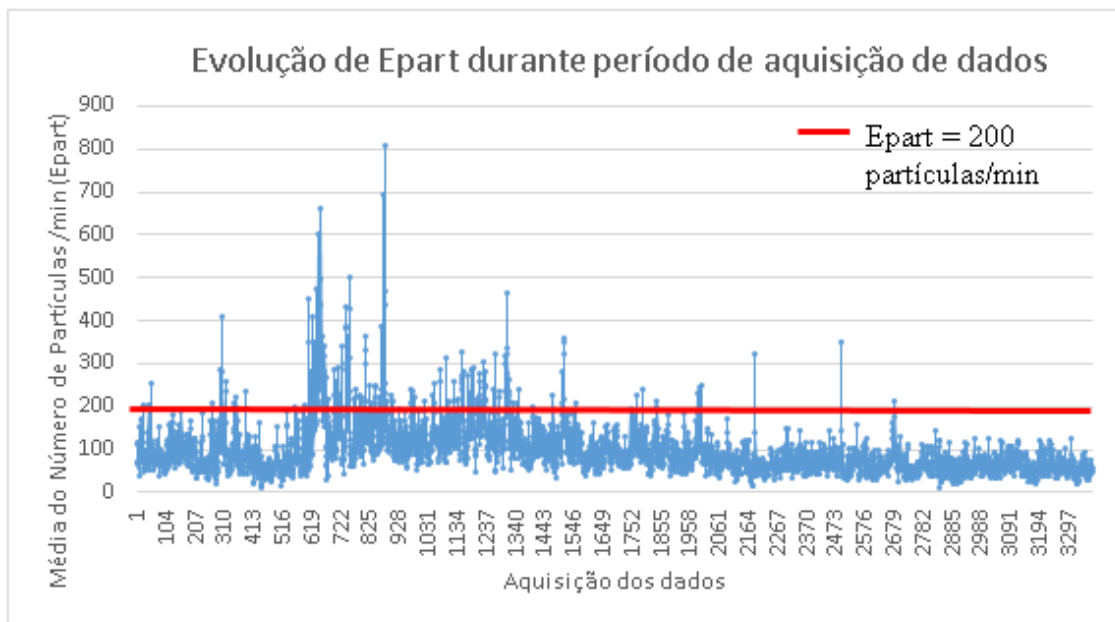


Figura 26: Evolução de Epart durante o período de aquisição de dados.

Segundo Costa (2004), valores de Epart de até 200 partículas/min correspondem a condições operacionais normais. Desta forma, a modelagem proposta deve ser capaz de

identificar condições anormais com números de partículas/min superiores a 200 partículas/min. Percebe-se pelo gráfico que este limite, representado pela linha vermelha horizontal, é ultrapassado em alguns momentos da operação da caldeira: existem 197 pontos operacionais (5,8% do total) com valores de Epart superiores a 200 partículas/min.

É notável, também, que a operação apresenta grandes variações do valor de Epart, assim como das demais variáveis, como será abordado a seguir, e que os valores de Epart não são distribuídos uniformemente. A distribuição não uniforme é esperada, já que os dados foram obtidos durante operação rotineira da caldeira com objetivo de acompanhar o processo e não de realizar uma varredura experimental de todos os valores possíveis de Epart. A baixa quantidade de pontos com Epart elevado (acima de 200) também é esperada e evidencia que a caldeira funciona, na maior parte do tempo, dentro das condições desejadas em relação à formação de material particulado. Entretanto, quando há uma má distribuição dos dados, o processo de treinamento da rede pode ser prejudicado, conforme será discutido posteriormente.

A Tabela 7 fornece os valores máximos e mínimos, a média e o desvio padrão de cada variável. Os gráficos da evolução de cada variável durante o período de aquisição de dados são fornecidos no apêndice 1.

Tabela 7: Estatísticas das variáveis de entrada.

	Vvlu	x	Tlu	Var1	Var2	Var3	Par1	Par2	Tar1	TAr2	PI	Epart
Valor Mínimo	25,09	70,71	127,15	82,35	99,64	18,02	12,72	123,31	120,37	109,34	1,76	9,62
Valor Máximo	74,73	87,39	143,46	156,1	143,61	93,07	247,78	510,71	165,23	205,94	4,31	806,12
Média	63,81	81,46	140,23	122,09	129,14	65,82	112,91	324,84	153,36	141,24	3,91	96,99
Desvio padrão	3,96	2,24	1,07	14,49	8,13	7,37	20,53	103,08	4,61	13,73	0,23	61,01

É interessante ressaltar que as variáveis que apresentaram maiores variações e maiores valores de desvio padrão são a pressão de alimentação do ar secundário, Par2, e a quantidade de partículas, Epart. Ademais, de acordo com os gráficos do apêndice 1, é possível observar que mudanças significativas ocorreram em relação às variáveis Var1 e Var2 a partir, aproximadamente, do conjunto de dados de número 884 e em relação à variável Par2, que sofreu um claro aumento a partir dos dados de número 1380, caracterizando uma mudança na forma de operação da caldeira.

Foi realizada uma análise a partir dos dados normalizados para verificar a influência de variações em cada uma das variáveis de entrada no valor de Epart. Para isso, foram identificadas as situações em que pelo menos uma das variáveis operacionais (incluindo Epart) apresentou um grande desvio da média.

Verificou-se que nos casos em que as variáveis de entrada apresentam grandes variações, a variável Epart apresenta desvios pequenos. A Tabela 8 mostra um exemplo em que isso ocorre. O desvio em relação à média de cada variável é apresentado e uma escala de cores é usada para evidenciar os maiores desvios (em vermelho escuro). Percebe-se que em apenas três casos (destacados na figura) variações significativas das variáveis de entrada correspondem a desvios significativos no Epart. Estes casos correspondem a valores de Epart iguais a 234,09, 187,33 e 350,6 partículas por minuto.

Tabela 8: Variação das variáveis operacionais normalizadas.

\ Wlu	x	Tlu	Var1	Var2	Var3	Par1	Par2	Tar1	TAr2	PI	Epart
5.6689	0.5428	6.1702	1.2545	2.1188	1.6490	0.1661	1.7176	0.5787	0.9441	1.2452	0.3094
0.2382	0.5218	0.5513	1.9282	1.6417	0.5948	5.8192	1.3108	1.4981	1.7400	0.4078	0.1757
0.1756	0.4684	0.2436	1.9682	1.5654	0.5474	6.1446	1.1826	1.4959	1.7808	0.3637	0.1842
8.2767	3.1977	13.1037	1.7732	3.6279	6.4827	1.9207	1.0255	2.8878	0.8253	3.2065	0.1248
2.1472	1.8347	1.3225	0.1786	0.1650	5.0081	0.5202	0.3362	0.0021	0.5617	1.2893	0.3628
5.9264	3.1710	5.8346	1.5854	1.0218	6.1666	3.3134	0.6324	1.9903	1.8659	0.1874	0.8209
2.1371	1.8302	1.2505	0.8502	1.0587	5.0813	2.2938	0.4916	0.5420	1.7909	0.2315	0.7629
3.2176	0.3557	0.1852	0.3909	0.4659	3.8400	5.2536	1.2794	1.0493	0.3550	0.3637	2.2471
3.5635	0.3201	0.1945	0.5752	0.3417	3.9214	5.4743	1.2846	1.3788	0.1781	0.3637	0.2996
3.8740	0.2132	0.8657	0.8810	0.1761	4.0584	6.5698	0.8862	1.6910	0.7482	0.4078	1.4807
5.3912	0.7356	1.1479	1.5081	0.9730	4.1941	0.8490	0.3129	0.6071	0.1656	0.0552	0.3286
7.2442	1.4293	6.2075	1.4722	0.9263	4.5644	1.3766	0.5094	1.9816	2.4856	2.2368	0.3464
0.3798	0.4137	0.2598	0.6570	0.6951	0.3697	0.9128	0.2995	7.1523	2.3232	0.4078	0.0540
0.1175	0.8069	0.6072	0.5451	0.5315	0.5963	0.5002	0.5065	0.3057	0.2625	5.0137	0.0984
0.7983	0.2132	1.1547	0.5672	0.2695	0.5434	0.8106	1.1491	1.0407	0.0025	5.3222	0.7616
1.0134	0.4047	1.8446	0.6004	0.2708	0.4756	0.9743	1.2310	1.0537	0.0200	6.7326	0.2219
1.0109	1.7589	2.1336	0.6300	0.1920	0.6845	0.6513	0.5263	0.8694	0.6622	5.6307	0.4957
0.4858	0.5161	1.6115	0.5652	0.1822	0.6465	0.2874	0.4241	0.8065	0.4889	6.3360	0.4233
0.5944	0.4226	1.4251	0.5976	0.2203	0.6669	0.2128	0.4844	0.8174	0.4860	7.5260	0.4141
0.8822	0.5983	1.3971	0.5955	0.2683	0.6791	0.2055	0.4104	0.8325	0.4977	9.4854	0.3752
0.7509	0.7745	1.4251	0.5527	0.3064	0.7564	0.2124	0.3513	0.8542	0.5108	8.8924	0.8275
1.1043	1.1887	1.4531	0.4913	0.4638	0.9124	0.1792	0.3302	0.8737	0.5239	7.1734	0.5483
2.3363	3.3403	1.4531	0.5389	0.1933	1.0142	0.1656	0.3364	0.8954	0.0863	5.5426	0.9053
1.5310	1.3268	0.6047	0.5451	0.4823	0.6872	0.3366	0.0738	0.9453	0.4256	5.4103	0.0787
1.9904	1.7188	0.1411	0.6349	0.7824	1.1634	0.6235	0.2492	0.7068	0.4438	6.8208	0.2369
2.0005	1.7723	0.0665	0.6473	0.7565	1.2272	0.5957	0.1839	0.7241	0.4285	5.7630	0.4630
0.9402	1.8124	0.2132	0.6722	0.4712	0.7483	0.5733	0.0527	0.8130	0.3804	5.2781	1.1385
1.0614	1.0462	0.2063	0.7902	0.5438	1.0128	0.5324	0.1468	0.7935	0.2799	6.8208	0.0437
1.4754	1.7589	0.0547	0.7543	0.5524	1.0711	0.5115	0.2228	0.8651	0.2981	5.9833	0.7454
9.7737	3.0730	4.7625	1.8394	2.2590	4.5414	2.2631	1.0107	1.8667	0.6324	4.0880	0.0543
6.7923	1.0150	2.6184	1.6696	2.3501	3.7179	2.1599	0.7027	2.1008	0.3295	3.5591	0.8733
9.4935	1.2422	1.7141	1.3114	0.4081	6.4855	1.9816	0.0795	0.3100	0.1300	0.5178	4.1569
5.5830	2.2635	0.7843	1.8739	0.4503	3.5944	1.6664	0.7075	0.0109	0.3746	0.4078	0.1359
6.4489	2.4105	0.0267	2.7196	1.0886	4.0258	0.7847	0.6653	0.4422	0.4059	0.4518	0.6240
6.8125	1.9384	0.8844	2.7430	1.2706	4.1289	1.0346	0.6210	0.1517	0.4212	0.4078	1.1558
5.9718	1.3459	1.9005	2.6119	0.2142	5.2183	0.6172	0.2478	0.0303	0.4103	0.1433	0.7258
6.0223	1.3548	2.7862	2.3806	0.2040	4.9714	0.4111	0.1368	0.2124	0.4241	0.0771	0.8360
7.2618	1.1098	3.7650	1.7173	2.2234	4.3894	1.7658	0.7925	1.1989	0.6943	3.3387	0.4093
0.0996	0.3557	0.7725	0.7778	0.9201	0.0631	0.9148	0.8671	5.4483	0.1998	0.0111	0.4365

Outra situação avaliada foi a análise dos casos em que a variável Epart apresentou valores superiores ao limite sugerido por Costa (2004). Observa-se que valores de Epart

acima do limite não correspondem diretamente a situações em que pelo menos uma das variáveis de entrada apresenta grandes desvios de seus valores médios. A Tabela 9 ilustra o comportamento das variáveis de entrada em alguns casos em que o valor de Epart foi superior a 200. Em apenas 8 dos 197 casos foram verificadas variações significativas em alguma das variáveis de entrada, sem que um padrão tenha sido identificado. Isto parece indicar que a influência das variáveis sobre a formação de partículas é de natureza complexa, possivelmente integrada e não linear, de modo que as redes neuronais se tornam uma abordagem interessante para resolução deste problema de modelagem.

Tabela 9: Variação das variáveis de entrada para os casos em que Epart é superior a 200.

Vvlu	x	Tlu	Var1	Var2	Var3	Par1	Par2	Tar1	TAr2	PI	Epart
1.04907	1.3459	0.63517	1.52436	1.48916	0.71965	1.44677	1.6034	0.19515	1.00308	1.24523	1.72049
1.13238	0.4149	0.25046	1.43945	1.65027	0.89331	0.88658	1.66607	0.30572	1.03876	1.2893	1.71902
0.86226	0.53072	0.25979	1.43876	1.56541	0.80105	1.16083	1.60292	0.26236	0.97978	1.2893	1.74786
0.46591	0.33471	0.25046	1.58786	1.43505	0.759	1.41559	1.55946	0.21033	1.13124	1.2893	2.59756
0.09229	0.26344	0.96146	2.13043	1.57279	0.29637	1.09604	1.23603	0.50518	1.15235	1.33338	1.83211
0.51135	1.4439	0.50466	1.96614	1.5236	0.07795	1.12771	1.45837	0.49645	1.32857	1.2893	3.09551
0.29677	1.1499	0.12927	1.98754	1.59001	0.0169	1.18178	1.45779	0.52246	1.33366	1.20115	3.00012
0.31192	1.10535	0.10379	1.94751	1.56787	0.00062	1.2037	1.44722	0.6352	1.36425	1.15707	5.15862
0.64263	0.77126	0.19702	1.95993	1.76096	0.41983	1.06487	1.43296	1.7062	1.3191	0.40777	2.26187
0.67544	0.8069	0.19702	1.95165	1.69332	0.352	1.14183	1.56858	1.71487	1.31546	0.40777	2.60854
0.48863	0.77572	1.64199	1.85225	1.88149	0.41033	1.75122	1.45294	1.7127	1.46328	0.40777	1.79195
0.14501	0.3659	1.28774	2.04484	1.71054	0.333	2.04349	1.39144	1.5436	1.41085	0.40777	1.69295
0.06171	0.33917	1.29706	2.00894	1.74374	0.30451	1.80529	1.39755	1.53059	1.43779	0.40777	2.07076
0.06704	0.74008	0.97078	1.97029	1.83475	0.32351	1.25484	1.49989	1.67802	1.40284	0.31962	2.28482
0.72341	1.35926	0.90552	1.81428	1.83475	1.24061	3.50923	1.40172	2.24387	1.65915	0.45185	1.75344
0.62748	0.5129	0.84026	1.48294	1.13005	1.23654	0.02243	1.95507	1.59997	1.42832	1.37746	5.80179
0.44572	0.50844	0.73772	1.51331	1.15587	1.1877	0.01849	1.92451	1.59563	1.48002	1.33338	4.12616
0.67039	0.47726	0.1504	1.32417	1.07347	1.37221	0.01074	1.78947	1.84929	1.32056	1.20115	2.99389
0.6603	0.34808	0.01057	1.38837	1.05749	1.40341	0.01557	1.72641	1.84712	1.3293	1.20115	2.55216
0.56689	0.66435	0.54878	1.38837	0.95664	1.4821	0.05794	1.73544	1.59997	1.48585	1.24523	2.39104
0.1352	1.13208	0.05718	1.23513	1.77572	1.36542	0.35119	1.67801	1.74956	1.39701	1.2893	5.15534
0.06957	1.31472	0.30888	1.21925	1.56787	1.15243	0.40916	1.86979	1.56311	1.44507	1.2893	2.1455
0.05442	0.84253	0.38346	1.24755	1.57402	1.05203	0.02048	1.7237	1.35715	1.56085	1.20115	2.03782
0.05161	0.77572	0.19702	1.24824	1.4744	1.05882	0.07646	1.77967	1.48723	1.42541	1.24523	3.79982
0.01627	0.74453	0.29956	1.19371	1.48547	1.05203	0.34973	1.68664	1.55227	1.46837	1.2893	4.13977
0.09229	0.67772	0.82162	1.20544	1.46457	1.00998	0.05989	1.62969	1.50024	1.43488	1.33338	3.66739
0.54669	0.67201	0.11995	1.28483	1.45965	0.84582	0.47399	1.62193	1.52842	1.39847	1.37746	6.14846
0.23113	1.16326	1.30639	1.31658	1.56787	0.84854	0.04825	1.58003	1.51758	1.37444	1.33338	3.11256
0.55427	1.34144	1.12926	1.29311	1.38217	0.84989	0.02535	1.58458	1.50891	1.35041	1.2893	2.85571
0.77137	1.36372	1.08265	1.34971	1.42275	1.08459	0.30345	1.597	1.57612	1.55721	1.20115	8.28155
0.44824	1.07863	1.12926	1.38768	1.46703	1.09545	0.14563	1.73291	1.56962	1.53391	1.20115	4.57855
0.51893	0.97617	1.18519	1.4077	1.3711	1.01947	0.1203	1.74252	1.56528	1.48876	1.20115	3.1242
0.04403	0.05283	1.11062	1.38561	1.44981	0.90687	0.28202	1.60641	1.67368	1.47347	1.20115	2.42267
0.00645	0.0229	1.064	1.39113	1.43997	0.90823	0.4077	1.6749	1.71054	1.48803	1.20115	1.8762
0.09957	0.65544	1.02671	1.37042	1.53467	0.91772	0.41793	1.60932	1.62598	1.46182	1.24523	1.72819
0.17307	0.86926	0.92417	1.39113	1.3834	0.98284	0.08084	1.69285	1.50674	1.35988	1.33338	5.56576
0.21599	0.6109	1.14542	1.22546	1.28624	1.11444	0.1559	1.66947	1.71921	1.41813	1.33338	2.38087
0.11753	1.01626	0.91484	1.39113	1.43259	0.84447	0.02092	1.78404	1.68669	1.42832	1.2893	2.40284
0.15006	0.41044	0.21317	1.49813	1.24811	1.50245	0.30102	1.67791	1.76691	1.50551	1.50969	3.64804
0.21094	0.72672	0.23182	1.34833	1.34773	1.43326	0.00582	1.66801	1.57395	1.51134	1.2893	6.57888
0.27405	0.63317	0.25046	1.4008	1.29116	1.36	0.08283	1.63474	1.56528	1.49459	1.20115	3.65034
0.9077	0.19092	0.33436	1.25652	1.52852	1.53636	0.00489	1.62649	1.59346	1.53318	0.80446	9.21648
0.39018	0.8069	0.34369	1.36697	1.54697	1.53365	0.29424	1.5549	1.61514	1.44289	0.45185	2.304
0.04432	0.61981	0.06401	1.36145	1.713	1.71137	0.04873	1.69062	1.50674	1.33002	0.45185	2.51692
0.10491	0.63762	0.15724	1.40908	1.64781	1.64489	0.03948	1.62281	1.48506	1.31764	0.49592	3.55413
0.03899	0.43272	0.18521	1.41322	1.58755	1.5445	0.01849	1.78588	1.49157	1.43197	0.45185	2.83146

5.2 Software STATISTICA®

5.2.1 STATISTICA® Automated Neural Networks (SANN)

A geração de redes e análise dos resultados neste projeto foi realizada no software STATISTICA® versão 8.0, que permite a análise e o gerenciamento de dados a partir de métodos estatísticos e pacotes com técnicas avançadas de reconhecimento e tratamento de dados, como as redes neurais artificiais.

Utilizou-se a ferramenta *Automated Neural Networks* (SANN) do pacote *Data Mining* para a geração semiautomática de redes neurais, que permite a resolução de problemas de classificação, de predição (regressão), séries temporais e análises de agrupamento de dados (do tipo *cluster*).

Os dados fornecidos devem ser tratados antes de serem usados para a geração de redes. Fica a cargo do usuário realizar um pré-tratamento por conta própria, ou permitir que o software normalize todos os dados de forma automática para valores entre 0 e 1 (STATISTICA, 2004).

A ferramenta permite, entre outros, a seleção dos seguintes parâmetros:

- Tipo de problema: classificação, regressão, séries temporais ou análise *cluster*;
- Entradas e saídas de acordo com o tipo de variável (contínua ou categórica);
- Distribuição dos dados a serem usados nas etapas de geração das redes, ou seja, treinamento, testes e validação;
- Tipo de rede que deve ser gerada: MLP, RBF ou mapeamento auto-organizável;
- Intervalo de números de neurônios ocultos que deve ser testado;
- Funções de ativação a serem testadas;
- Número de redes a serem testadas;

Uma vez geradas as redes com os parâmetros definidos, o software fornece como resultado as melhores redes, ou seja, aquelas que apresentam menores erros de treinamento e de teste e, portanto, melhor desempenho.

Em seguida, é possível analisar os resultados gerados a partir da sensibilidade das variáveis de entrada, dos erros e do desempenho, dos algoritmos de treinamento e funções usados nas redes selecionadas. Pode-se, ainda, saber os pesos de cada conexão das redes e as predições realizadas ao longo do treinamento e dos testes. Por fim, os modelos representados pelas redes de maior interesse podem ser salvos em linguagem C/C++.

5.2.2 Escolha das Entradas Relevantes a Partir da Análise de Sensibilidade

O software STATISTICA, a partir da ferramenta de Análise Global de Sensibilidade (em inglês, *Global Sensitivity Analysis*), possibilita a geração de uma análise da sensibilidade das entradas, permitindo o estudo da importância de cada variável para o desempenho da rede. Para tal, analisa-se a influência de alterações em cada entrada no erro calculado.

A rede treinada é executada com diversos conjuntos de dados. Para cada variável, prepara-se um conjunto de dados em que esta variável tem sempre o mesmo valor (seu valor médio) e as demais variáveis são exatamente as mesmas usadas no treinamento. Um novo erro correspondente a essa alteração é calculado (STATISTICA, 2004). Desta forma, caso uma variável apresente uma grande influência sobre o resultado da rede, o novo erro será superior ao antigo. Caso contrário, será possível observar um decréscimo do erro, o que significa que essa variável pode ser removida para que o desempenho da rede neuronal melhore.

Para a visualização do decréscimo ou aumento do erro, o software disponibiliza os resultados na forma da razão entre o novo erro (com alterações em uma variável) e o erro original. Assim, caso o erro aumente, a razão será superior a 1 e é aconselhável manter a variável, uma vez que ela possui uma influência importante para a rede. Se o erro diminuiu, apresentando razão inferior a 1, significa que a rede responde melhor sem a presença daquela entrada. Por fim, se a razão é igual a 1, sabe-se que o erro não foi modificado pela alteração na entrada e que esta, portanto, não é relevante para a rede neuronal. Nestes dois casos, é recomendável a retirada da variável.

É importante ressaltar que as entradas podem apresentar correlações e redundâncias entre si e que, por isso, a análise de sensibilidade possui limitações. Isto porque ela analisa alterações em uma variável por vez e desconsidera tais relações entre as entradas, além de realizar análises independentes para cada rede, ou seja, a análise é específica para cada rede gerada (STATISTICA, 2004).

5.2.3 Determinação do Erro e da Performance das Redes

A escolha das melhores redes se apoia na análise da qualidade de seu funcionamento traduzida pela sua performance e seu erro de predição ou classificação.

Os erros calculados pelo software têm por objetivo determinar a proximidade entre a saída desejada e aquela calculada pela rede, determinando a distância entre esses

dois valores usando uma função erro específica. Os erros calculados são usados no treinamento para o ajuste dos pesos sinápticos a cada iteração conforme explicado anteriormente.

Neste estudo, foram usados dois tipos de erro: a soma do erro quadrático (*Sum of squared error – SOS*) e o chamado erro de entropia cruzada ou, em inglês, *Cross Entropy Function* (CE). O primeiro tipo de erro calcula a soma dos quadrados dos erros entre os valores de saída desejados e aqueles calculados pela rede. Ele é o único erro empregado para problemas de predição, mas pode ser usado, também, para problemas de classificação. A Equação (26) abaixo, em que s_i representa o valor predito pela rede, t_i a saída desejada e N o número total de dados de treinamento, mostra como o erro SOS é determinado.

$$E_{SOS} = \sum_i^N (s_i - t_i)^2 \quad (26)$$

Já o *Cross Entropy Function* (CE) é o mais utilizado em problemas de classificação e é dado pela Equação (27), em que s_i é o valor predito pela rede, t_i a saída desejada e N o número total de dados de treinamento.

$$E_{CE} = - \sum_i^N t_i \times \ln \left(\frac{s_i}{t_i} \right) \quad (27)$$

Deve-se analisar tanto a performance do treinamento, que verifica a capacidade de aprendizagem da rede, quanto a do teste ou validação, que avalia a capacidade de generalização da rede a partir do uso de novos valores para as entradas não usadas no treinamento. Esta última é fundamental para verificar se houve fenômeno de sobre-ajuste (*overfitting*) e se é necessário, então, alterar a arquitetura da rede.

Em problemas de predição, o software STATISTICA® calcula a performance da rede a partir do coeficiente de correlação entre as saídas desejadas (fornecidas previamente para treinamento e teste) e as saídas calculadas pela rede. O coeficiente de correlação varia de -1 a 1 e em caso de performance ou predição perfeita ele assume seu valor máximo.

Deve-se ressaltar que um coeficiente igual a 1 pode indicar a presença de *overfitting* na rede, que ocorre quando a rede apresenta uma boa performance somente para os dados de treinamento. Normalmente, os dados de entrada apresentam ruídos por serem provenientes de processos reais e isso dificulta a obtenção de redes com

performances próximas a 1. Por causa do risco de generalizações ruins e de *overfitting*, recomenda-se sempre basear a escolha da melhor rede na análise do coeficiente de correlação ligado ao teste (e à validação, quando houver). É importante, também, analisar o número de neurônios da camada oculta.

Em problemas de classificação, as saídas são previamente classificadas e separadas em classes pré-definidas. A eficiência da rede é, então, dada pela porcentagem ou taxa de classificações corretas. Entretanto, é necessário realizar um estudo mais aprofundado além da análise das performances do treinamento e do teste. Isso porque a quantidade de dados correspondentes a cada classe pode variar muito. Esta questão é particularmente relevante nos problemas de detecção e diagnóstico de falhas em que o número de exemplos de operação normal é muito maior do que os de classes de falhas, o que pode prejudicar o processo de treinamento da rede neuronal.

Dessa forma, a eficácia da classificação de uma rede pode variar de acordo com a classe. Se determinadas classes apresentarem dados suficientes de treinamento, a rede irá classificar com alta taxa de acertos os dados pertencentes a estas classes. No entanto, se alguma classe não for bem treinada, a rede será incapaz de realizar a generalização e de classificar futuras entradas pertencentes à mesma. Assim, é possível que a rede apresente uma performance global elevada, uma vez que ela consegue classificar bem a maioria das classes. Porém, ela apresenta um desempenho deteriorado para uma das classes, o que não é, obviamente, desejado. Este aspecto também é especialmente relevante no controle de processos: uma rede que representa bem as situações normais de operação, mas que não consegue modelar uma situação anormal decorrente de um distúrbio operacional pode gerar uma situação de controle insegura. Para garantir que a performance da rede de classificação seja boa para todas as classes desejadas, é recomendável analisar o desempenho individual de cada classe.

5.3 Predição do Número de Partículas a Partir de Redes de Regressão

A primeira abordagem adotada foi a predição do número de partículas usando redes MLP e RBF com 3 camadas (entrada, oculta e saída) para a modelagem do problema. Os primeiros parâmetros de análise fixados foram o número de redes geradas a cada análise, a distribuição dos 3.381 pontos operacionais em amostras de treinamento e testes, a faixa de número de neurônios ocultos, o decaimento dos pesos, as funções de

ativação e o número de redes a serem mostradas após cada análise. Estes parâmetros e suas definições são mostrados na Tabela 10.

Tabela 10: Parâmetros iniciais de geração de redes.

Parâmetros	Valores definidos
Número de redes geradas a cada análise	1.000
Amostra de treinamento	80% do conjunto de dados
Amostra de teste	20% do conjunto de dados
Amostra de validação	0% do conjunto de dados
Número de neurônios ocultos para redes MLP	De 4 a 14 (faixa sugerida pelo software)
Número de neurônios ocultos para redes RBF	De 10 a 100 e de 144 a 252 (faixa sugerida pelo software)
Decaimento dos pesos	0,0001 (mínimo) e 0,001 (máximo)
Funções de ativação	Inicialmente todas as disponíveis conforme seção 4.4
Entradas	Vvlu, x, Tlu, Var1, Var2, Var3, Par1, Par2, Tar1, Tar2, PI
Número de redes a serem mostradas após cada análise (seleção automática do software)	10

Primeiramente, foram testadas redes com diferentes funções de ativação para as camadas oculta e de saída, variando o número de neurônios na camada oculta. Em seguida, as arquiteturas e funções de ativação que apresentaram os melhores desempenhos foram selecionadas para sucessivos refinamentos da busca. Dessa forma, novas redes foram geradas e este processo foi repetido até que não houvesse uma melhora significativa do desempenho das redes. Foi realizada também a análise de sensibilidade das redes geradas para verificar a influência de cada entrada no desempenho da rede. Caso alguma das variáveis não apresentasse influência no desempenho da rede, ela era retirada para os testes subsequentes.

5.4 Classificação do Número de Partículas

A segunda abordagem proposta foi o uso de redes neuronais para a classificação dos dados em função do número de partículas. Esta estratégia também foi estudada por Costa (2004) e se justifica pelo fato de que o número exato de partículas não é essencial para o controle do processo e acompanhamento da operação. Segundo Costa (2004), o

monitoramento pode ser feito de forma satisfatória a partir do conhecimento da faixa de valor do número de partículas ao invés de um valor numérico. Como dito anteriormente, a operação pode ser considerada normal para valores de Epart inferiores a 200 partículas/min. Assim, uma rede capaz de prever e classificar o número de partículas em dois intervalos (Epart superior ou inferior a 200 partículas/min) já seria satisfatória para solucionar o problema de modelagem proposto.

O estudo do uso de redes para classificação se baseou na separação dos dados em 7, 5, 4, 3 e 2 classes. As classes foram definidas conforme proposto por Costa (2004) e suas definições são apresentadas na Figura 27.

Faixa Epar	Quantidade de dados	Classe e (conjuntos de dados na classe)				
		A	B	C	D	E
0 - 50	481	1 (481)	1 (481)	1 (481)	1 (2236)	
50 - 100	1755	2 (1755)	2 (1755)	2 (1755)		1 (3184)
100 - 150	710	3 (710)	3 (710)	3 (710)	2 (948)	
150 - 200	238	4 (238)	4 (238)			
200 - 250	106	5 (147)				
250 - 300	41			4 (435)		
300 - 350	25	6 (34)	5 (197)		3 (197)	2 (197)
350 - 400	9					
>400	16	7 (16)				

Figura 27: Definição das classes estabelecidas e estudadas na classificação do número de partículas.

No caso “A”, os dados foram distribuídos em 7 grupos de acordo com o valor de Epart: a classe 1 agrupa valores de Epart de até 50 partículas por minuto; a classe 2 valores de Epart entre 50 e 100; a classe 3 de 100 a 150; a classe 4 de 150 a 200; a classe 5 de 200 a 300; a classe 6 de 300 a 400 e a classe 7 valores de Epart superiores a 400. No caso “B”, os dados foram separados em 5 classes, sendo as primeiras classes iguais às do caso A e a classe 5 inclui valores de Epart superiores a 200. No caso “C”, os dados são separados em quatro classes: as três primeiras classes possuem a mesma definição que

nos casos “A” e “B” e a classe 4 agrupa dados com Epart superior a 150. No caso “D”, os dados são divididos em três classes: a classe 1 inclui pontos com Epart inferiores a 100; a classe 2 inclui valores de Epart entre 100 e 200 e a classe 3 agrupa valores de Epart superiores a 200. No caso “E”, os dados são agrupados em apenas duas classes: a classe 1 agrupa dados com Epart inferior a 200 e a classe 2, valores de Epart superiores a 200 partículas por minuto.

Assim como na primeira abordagem, também foram exploradas redes MLP e RBF de classificação com 3 camadas (de entrada, oculta e de saída). Os primeiros parâmetros de análise fixados foram o número de redes geradas a cada análise, a distribuição dos 3.381 conjuntos de dados em amostras de treinamento e testes, a faixa de número de neurônios ocultos, o decaimento dos pesos, as funções de ativação e o número de redes a serem mostradas após cada análise. Estes parâmetros e suas definições são mostrados na Tabela 11. É importante ressaltar que a saída da rede corresponde ao número da classe à qual o conjunto de dados pertence, que é uma variável do tipo categórica (ou discreta) e não mais ao valor de Epart, variável contínua.

Tabela 11: Parâmetros iniciais de geração de redes de classificação.

Parâmetros	Valores definidos
Número de redes geradas a cada análise	1.000
Amostra de treinamento	80% do conjunto de dados
Amostra de teste	20% do conjunto de dados
Amostra de validação	0% do conjunto de dados
Número de neurônio ocultos para redes MLP	De 4 a 14 (faixa sugerida pelos desenvolvedores do software)
Número de neurônio ocultos para redes RBF	De 10 a 100 e de 144 a 252 (faixa sugerida pelos desenvolvedores do software)
Decaimento dos pesos	0,0001 (mínimo) e 0,001 (máximo)
Funções de ativação	Inicialmente todas as disponíveis conforme seção 4.4
Entradas	Vvlu, x, Tlu, Var1, Var2, Var3, Par1, Par2, Tar1, Tar2, PI
Número de redes a serem mostradas após cada análise (seleção automática do software)	10

O procedimento utilizado na geração de redes e na busca daquela com melhor desempenho seguiu o padrão estabelecido para o estudo de predição. Inicialmente, foram

testadas diferentes funções de ativação para as camadas oculta e de saída, com o número de neurônios na camada oculta variando de acordo com as faixas descritas na Tabela 11. A cada teste, correspondendo à geração de 1.000 novas redes, as arquiteturas e funções de ativação apresentando os melhores desempenhos eram selecionadas e utilizadas para o refinamento da busca. O processo foi repetido até que não houvesse uma melhora significativa do desempenho das redes. Foi realizada também a análise de sensibilidade das redes geradas para verificar a influência de cada entrada no desempenho da rede e caso alguma entrada não fosse relevante, ela era retirada nos testes seguintes.

Para redes de classificação, além da análise do desempenho geral da rede, é necessária uma avaliação do percentual de classificações corretas referente a cada classe. Isto é importante principalmente quando os dados não se encontram uniformemente distribuídos nas classes estabelecidas, o que pode prejudicar o treinamento e, conseqüentemente, a capacidade de generalização da rede. Se, por exemplo, uma das classes apresenta poucos dados em relação às demais, a rede pode não ser capaz de classificar estes dados corretamente. Entretanto, como o número de casos desta classe é insignificante frente à totalidade dos dados, o desempenho geral pode ser elevado e pouco afetado pelos erros de classificação referentes a tal classe. Desta forma, mesmo com um bom desempenho geral, uma determinada rede pode não conseguir classificar todas as classes de forma satisfatória.

Além da geração de redes para classificação do número de partículas, foi proposta e estudada uma abordagem “híbrida” adicional, que une a predição do número de partículas por redes neurais (seção 5.3) à estratégia de separar os dados em classes diferentes. Assim, a rede de predição de melhor desempenho foi utilizada para predição do número de partículas. A partir dos valores preditos calculados pela rede tanto no teste quanto no treinamento, os resultados foram tratados posteriormente de forma a serem classificados e separados nas classes estabelecidas e discutidas anteriormente. Foi realizada uma comparação entre as classificações feitas a partir dos valores preditos pela rede e dos valores reais, avaliando-se, então, a eficiência da classificação deste método. Isso foi feito verificando a porcentagem de acerto na classificação dos dados de acordo com cada classe. Esta estratégia será melhor detalhada na seção 6 durante a discussão dos resultados, para que seja possível compreender os procedimentos realizados e as razões que levaram ao desenvolvimento de tal abordagem “híbrida”.

5.5 Análise de Cluster

A terceira abordagem proposta foi estudar o problema utilizando redes neurais de mapeamento auto-organizável para realizar análises do tipo *cluster*. O objetivo foi verificar se estas redes conseguem modelar o número de partículas, gerando agrupamentos com características em comum. Em outras palavras, buscou-se testar a capacidade de redes auto-organizáveis para o reconhecimento de padrões nos dados disponíveis e comparar os agrupamentos feitos pelas redes com a classificação proposta anteriormente (seção 5.4).

Foram testadas redes bidimensionais de mapeamento auto-organizável com diferentes topologias utilizando o software STATISTICA®. Variou-se o número de neurônios de 4 a 180, assim como a sua disposição na rede e foram testadas redes com as seguintes topologias: 2x2, 3x3, 1x100, 3x40, 3x60, 4x10, 5x7, 5x8, 5x20, 6x6, 10x10 e 20x5 (em que o primeiro número indica o número de linhas e o segundo o número de colunas da matriz formada pelos neurônios, e o produto destes valores fornece o total de neurônios da rede). As diversas topologias das redes testadas estão representadas na Figura 28, que relaciona o número de neurônios e o número de linhas.

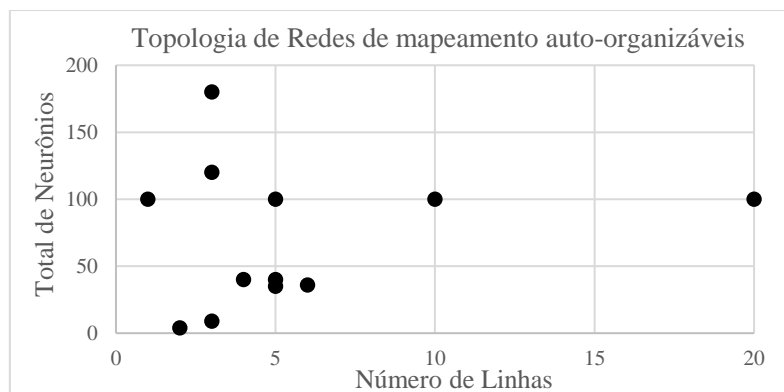


Figura 28: Representação das topologias das redes testadas.

Foram adotados os valores recomendados pelos desenvolvedores do software para os demais parâmetros da rede, que se encontram resumidos na Tabela 12. É importante ressaltar que a variável *Epart* não foi utilizada, já que a rede utiliza treinamento não supervisionado.

Tabela 12: Parâmetros utilizados para a geração de redes de mapeamento auto-organizável.

Parâmetros	Valores definidos
Número de redes geradas a cada análise	1
Amostra de treinamento	80% do conjunto de dados
Amostra de teste	20% do conjunto de dados
Amostra de validação	0% do conjunto de dados
Número de neurônios	De 4 a 180
Entradas	Vvlu, x, Tlu, Var1, Var2, Var3, Par1, Par2, Tar1, Tar2, PI
Ciclos (épocas) de treinamento	1.000
Taxas de aprendizado	De 0,1 a 0,02
Condições de parada de treinamento	A partir de 10 ciclos, quando a melhora na performance é de 10^{-7}
Vizinhança	Inicia-se com 3 neurônios e finaliza com 0
Inicialização dos pesos	Distribuição normal, valores aleatórios

A rede gera um mapa topológico, aqui escolhido retangular, em que cada neurônio corresponde a um elemento, e cada conjunto de dados é associado a um neurônio.

Em um primeiro momento, para a avaliação de cada rede, foi realizada uma análise baseada na divisão dos dados em duas classes: a classe 1 inclui os conjuntos de dados em que E_{part} é menor ou igual a 200 e a classe 2 reúne os conjuntos de dados em que E_{part} é maior que 200 (ver caso “E” apresentado na Figura 27). A análise consistiu em identificar regiões em que há uma predominância de conjuntos de dados de classe 2, ou seja, de operação anormal da caldeira.

Para tal, foram identificados os neurônios vencedores para cada conjunto de dados (vetor de entrada). Para cada neurônio, foi calculado o percentual de casos de cada classe que nele foram agrupados. Assim, foi possível buscar regiões em que houve grande concentração de determinada classe, identificando regiões do mapa que agrupam dados de cada classe.

Por exemplo, na rede com neurônios dispostos em uma matriz 20x5, o neurônio 1 foi o vencedor 42 vezes, das quais 17 correspondem a dados de classe 1 e 25 da classe 2. A partir destes resultados, calculou-se a porcentagem de dados de classe 2 que foram

agrupados no neurônio 1. Neste caso, 60% dos dados agrupados no neurônio 1 são de classe 2 e, portanto, este neurônio foi considerado como sendo majoritariamente uma região “classe 2”.

A partir das porcentagens calculadas para os dados de classe 2 em cada neurônio, foram construídos mapas evidenciando estes valores na tentativa de identificação visual de regiões com maior incidência de casos de classe 2. Uma escala de cor foi utilizada para facilitar a visualização, cores mais escuras correspondem a zonas com mais casos de classe 2. A Figura 29 apresenta o mapa criado para a rede 20x5, em que cada valor percentual corresponde a um neurônio e a contagem do número do neurônio é realizada a partir da primeira linha (na primeira linha estão localizados os neurônios de 1 a 5, na segunda, de 6 a 10, e assim por diante).

Os neurônios com porcentagens nulas indicam que eles foram ativados somente por dados de classe 1. Percebe-se que existem 4 neurônios que agrupam majoritariamente dados de classe 2 e que existem pelo menos 3 regiões que agrupam porcentagens superiores a 20%. Isso indica que, no uso desse classificador, caso um dado operacional ative um neurônio presente nestas regiões, há chances significativas de que ele pertença à classe 2. Assim, apesar de não ser um indicativo preciso, pode dar um alerta de que as condições operacionais podem estar fugindo à normalidade. A mesma análise foi feita com as demais configurações da rede. Os demais mapas e resultados obtidos serão apresentados na seção 6.4.

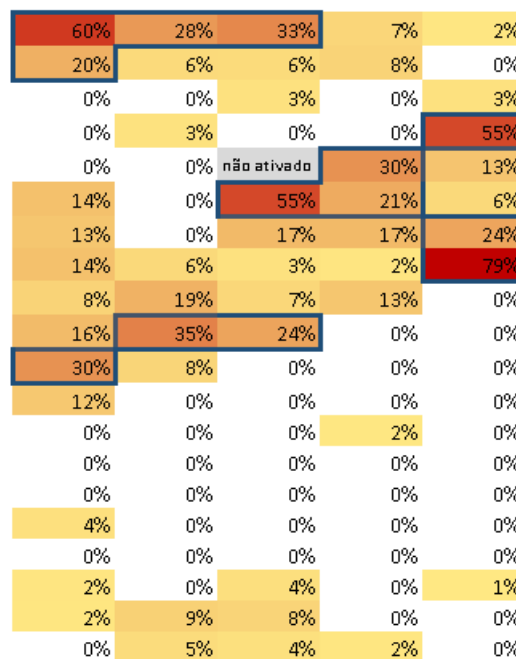


Figura 29: Mapa auto-organizável com 100 neurônios dispostos em uma matriz 20x5, indicando regiões de grande incidência de casos de classe 2.

Em seguida, a rede que apresentou os resultados mais satisfatórios teve seu comportamento avaliado quando os dados são divididos nas outras classes tratadas anteriormente, ou seja, quando eles são separados em 3, 4, 5 e 7 classes. O procedimento utilizado para este estudo foi similar ao reportado acima, entretanto, como a complexidade da análise aumenta com o número de classes, ao invés de serem criados mapas contendo os valores percentuais de uma determinada classe na posição de cada neurônio, optou-se por indicar qual a classe que ativou majoritariamente aquele neurônio. Foi estabelecido um valor percentual mínimo de 50% para que uma classe fosse considerada majoritária em um dado neurônio. Caso nenhuma classe fosse preponderante, o neurônio recebia o valor “0”. Essa metodologia é chamada de “proporção mínima com vizinhança por Voronoi” (DE SOUZA Jr., 2003).

6. RESULTADOS

Nesta seção, serão apresentados os principais resultados obtidos a partir da metodologia descrita anteriormente.

6.1 Predição

Adotou-se a nomenclatura do software STATISTICA® para descrição de cada rede que segue o seguinte formato: (Tipo de rede) (Número de neurônios de entrada) – (Número de neurônios ocultos) – (Número de neurônios de saída). Por exemplo, uma rede MLP 11-14-1 consiste de uma rede do tipo MLP, com 11 neurônios de entrada, 14 ocultos e 1 de saída.

6.1.1 Redes MLP

O teste inicial de geração de redes mostrou que as redes MLP com melhor desempenho, ou seja, com maior coeficiente de correlação apresentavam como funções de ativação para a camada oculta as funções tangente hiperbólica e logística, e para a camada de saída as funções tangente hiperbólica e função exponencial. O desempenho da rede é fornecido pela correlação entre as saídas calculadas pela rede e aquelas desejadas. O melhor desempenho no primeiro teste foi de 0,7446 para os dados de treinamento e de 0,7321 para os dados de teste. As três melhores redes obtidas são detalhadas na Tabela 13.

Tabela 13: Resultados do teste inicial com redes MLP para predição.

Rede	Performance Treinamento/ Teste	Erro Treinamento/ Teste	Função de Ativação	
			Camada Oculta	Camada Saída
MLP 11-13-11	0,7446/0,7321	0,0017/0,0023	Logística	Exponencial
MLP 11-14-1	0,7440/0,7309	0,0017/0,0023	Tangente hiperbólica	Tangente hiperbólica
MLP 11-10-1	0,7256/0,7283	0,0018/0,0023	Tangente hiperbólica	Exponencial

Além das performances, foi realizada, também, uma análise da sensibilidade das entradas. Observou-se que todas as sensibilidades foram superiores a 1, indicando que todas as variáveis são relevantes para o modelo proposto, conforme explicado na seção

5.2.2. As quatro variáveis que se mostraram mais relevantes para cada rede serão destacadas e, para a rede de melhor desempenho, todas as variáveis serão ordenadas segunda sua relevância.

Com base nesses resultados, outras 1.000 redes foram geradas restringindo as funções de ativação e mantendo os demais parâmetros constantes. A restrição das funções de ativação mostrou que as redes com melhor desempenho continuavam apresentando tanto a função tangente hiperbólica quanto a logística como funções de ativação para a camada oculta, entretanto, para a camada de saída, verificou-se que as melhores redes usavam a função exponencial.

A melhor rede apresentou como desempenho de treinamento e teste, respectivamente, os valores de 0,7515 e 0,7561. Assim como no primeiro teste, a análise de sensibilidade indicou que todas as entradas são importantes, entretanto é possível destacar as entradas mais relevantes. A Tabela 14 resume os resultados.

Tabela 14: Resultados do segundo teste com redes MLP para predição.

Rede	Performance	Erro	Função de Ativação		Principais Entradas por relevância
	Treinamento / Teste		Treinamento/ Teste	Camada Oculta	
MLP 11-11-1	0,7515/ 0,7561	0,0016/ 0,0021	Tangente hiperbólica	Exponencial	Var2, Var1, Par2, Tar2
MLP 11-12-1	0,7397/ 0,7367	0,0017/ 0,0023	Logística	Exponencial	Var2, Var3, Par2, Var1
MLP 11-12-1	0,7181/ 0,7324	0,0018/ 0,0023	Tangente hiperbólica	Exponencial	Var3, Var1, Var2, Par2

Desta forma, foram propostos testes independentes para avaliar o desempenho de redes com funções de ativação tangente hiperbólica (teste 3) e logística (teste 4) para a camada oculta, com função exponencial para a camada de saída. Os resultados das três melhores redes encontradas em cada teste são fornecidos na Tabela 15 e na Tabela 16. Verificou-se que, no geral, os melhores desempenhos foram obtidos para o teste 3 e a melhor rede encontrada nestes testes apresenta como função de ativação da camada oculta a função tangente hiperbólica e teve um desempenho, para o teste, de 0,7575.

A partir destes resultados, concluiu-se que as melhores funções de ativação para a camada oculta e para a camada de saída são, respectivamente, tangente hiperbólica e função exponencial.

Tabela 15: Resultados do terceiro teste com redes MLP para predição.

Rede	Performance	Erro	Função de Ativação		Principais Entradas por relevância
	Treinamento / Teste	Treinamento / Teste	Camada Oculta	Camada Saída	
MLP 11-11-1	0,7470/ 0,7575	0,0017/ 0,0017	Tangente hiperbólica	Exponencial	Var3, Var1, Tar2, Var2
MLP 11-14-1	0,7391/ 0,7492	0,0017/ 0,0022	Tangente hiperbólica	Exponencial	Var1, Tar2, Var2, Var3
MLP 11-12-1	0,7340/ 0,7457	0,0017/ 0,0022	Tangente hiperbólica	Exponencial	Var1, Var2, Var3, Par1

Tabela 16: Resultados do quarto teste com redes MLP para predição.

Rede	Performance	Erro	Função de Ativação		Principais Entradas por relevância
	Treinamento / Teste	Treinamento / Teste	Camada Oculta	Camada Saída	
MLP 11-13-1	0,7434/ 0,7548	0,0017/ 0,0021	Logística	Exponencial	Var3, Var1, Par2, Var2
MLP 11-9-1	0,7120/ 0,7380	0,0018/ 0,0022	Logística	Exponencial	Var3, Var1, Par2, Var2
MLP 11-11-1	0,7281/ 0,7347	0,0018/ 0,0023	Logística	Exponencial	Var2, Var3, Tar2, Var1

Outros testes foram realizados gerando redes com 11, 12 e 13 neurônios na camada oculta para verificação da melhor arquitetura. De todas as redes testadas, aquela que apresentou melhor desempenho possui 11 neurônios ocultos e será usada como base para as próximas análises. Os resultados obtidos são apresentados na Tabela 17. Todas as variáveis, assim como em todos os testes anteriores, apresentaram sensibilidade superior a 1 e foram, portanto, mantidas como entradas relevantes. Os pesos da rede são fornecidos no apêndice 2.

O desempenho da rede selecionada (de 0,7464 e 0,7612 para o treinamento e para o teste) foi considerado satisfatório. Desempenhos superiores são improváveis de serem alcançados já que os dados apresentam ruídos de medição e a própria variável E_{part} é resultado da média de duas medições distintas que podem diferir de até 100 partículas/min.

Tabela 17: Resultados para melhor rede MLP para predição.

Rede	Performance Treinamento/ Teste	Erro Treinamento/ Teste	Função de Ativação	
			Camada Oculta	Camada Saída
MLP 11-11-1	0,7464/ 0,7612	0,0017/ 0,0021	Tangente Hiperbólica	Exponencial
Entradas por Relevância	Var2, Var1, Tar2, Tar1, Par2, Var3, Par1, Vvlu, Pl, x, Tlu			

Entretanto é necessário verificar se a predição é satisfatória na totalidade dos dados ou apenas em pontos correspondentes à operação normal da caldeira, isto é, com Epart inferior a 200. O gráfico da Figura 30 mostra os valores preditos e medidos em função dos conjuntos de dados usados no desenvolvimento da rede. O gráfico da Figura 31 apresenta uma comparação entre os valores de Epart estimados pela rede neuronal e os valores reais.

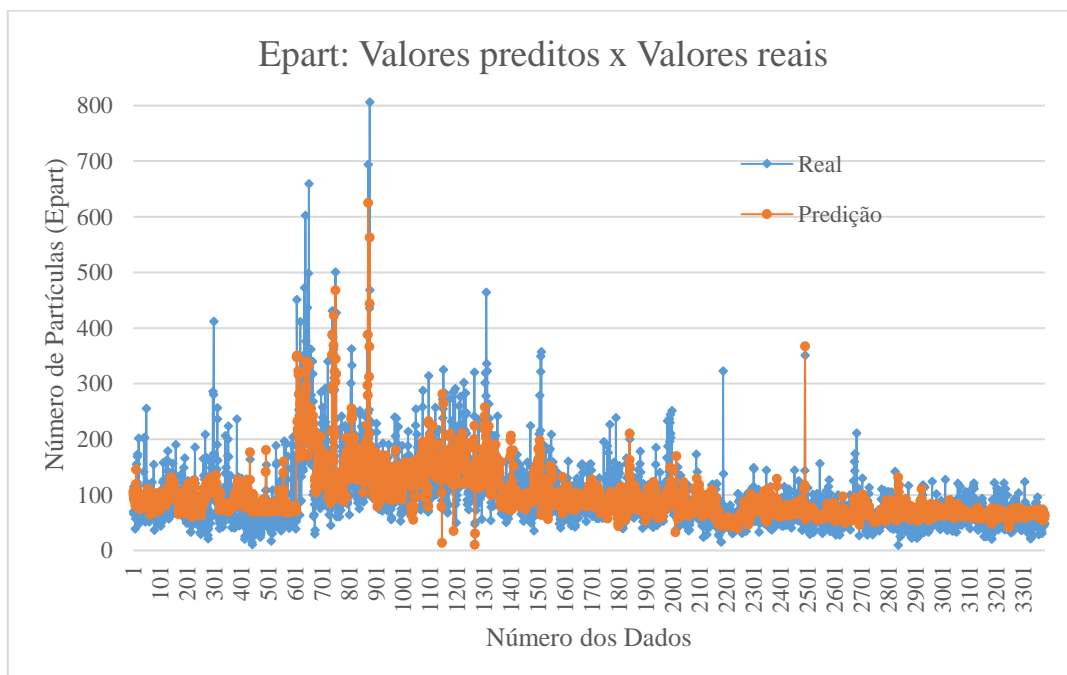


Figura 30: Comparação entre os valores preditos pela melhor rede neuronal e os valores reais para cada conjunto de dados.

Percebe-se que a predição é satisfatória para a maioria dos pontos que correspondem às condições normais de operação (em relação ao valor de Epart). Entretanto, a rede é incapaz de prever corretamente valores elevados de Epart. Isto se justifica pela má distribuição dos dados em relação aos valores de Epart: apenas 197 conjuntos de dados, dos 3.381, apresentam valores de Epart superiores a 200

partículas/min. Esta quantidade de dados é insuficiente para o bom treinamento da rede em relação a esta situação.

O ponto mais importante da predição da quantidade de partículas formadas na caldeira é identificar se o limite de 200 partículas/min foi ultrapassado e, portanto, se a operação passou a ser considerada a anormal. Como a rede não é capaz de prever corretamente esses pontos, há chances significativas de que a rede indique que a operação está normal mesmo quando o real valor de Epart é superior ao limite de operação padrão. Desta forma, a rede não foi considerada satisfatória apesar de seu bom desempenho geral.

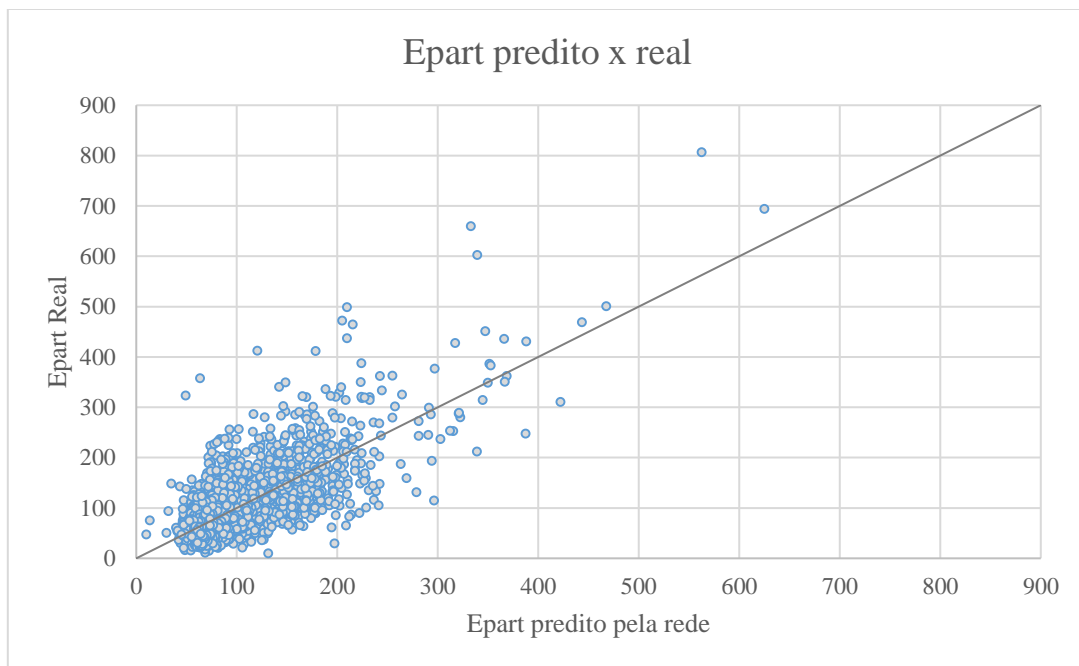


Figura 31: Comparação entre os valores de Epart preditos e os reais para todos os dados.

6.1.2 Redes RBF

O estudo da utilização de redes do tipo RBF para predição do número de partículas seguiu uma metodologia similar à do estudo de redes MLP. Foram testadas as funções de ativação disponíveis no software e diferentes números de neurônios na camada oculta. Além disso, as entradas que não se mostraram relevantes para a melhora do desempenho da rede foram descartadas. Uma vez determinados os melhores parâmetros para as redes, estes foram fixados e outras redes foram geradas para buscar obter melhores resultados.

Primeiramente, foram geradas 1.000 redes com camadas ocultas contendo de 10 a 100 neurônios. As funções de ativação que apresentaram os melhores desempenhos foram a função gaussiana para a camada oculta e a função identidade para a camada de

saída. Entretanto, os desempenhos das redes, ou seja, os coeficientes de correlação foram muito inferiores àqueles encontrados nas redes MLP. Os coeficientes obtidos para as melhores 10 redes variaram entre 0,2705 e 0,3875. A melhor rede encontrada no primeiro teste apresentou 10 neurônios escondidos e apenas 4 variáveis de entrada relevantes de acordo com a análise de sensibilidade: x, Tar2, Var3 e Vvlu. A Tabela 18 resume os resultados obtidos para a melhor rede.

Tabela 18: Resultados do primeiro teste para a melhor rede RBF para predição.

Rede	Performance Treinamento/ Teste	Erro Treinamento/ Teste	Função de Ativação	
			Camada Oculta	Camada Saída
RBF 11-10-1	0,3688/ 0,3875	27/ 27	Gaussiana	Identidade
Entradas por Relevância	x, Tar2, Var3 e Vvlu			

Como os resultados não foram satisfatórios, decidiu-se explorar redes com mais neurônios na camada oculta, utilizando a recomendação do próprio software de 144 a 252 neurônios. Entretanto, os coeficientes de correlação resultantes foram próximos de zero ou mesmo negativos e os valores de erro encontrados foram extremamente elevados, sendo, em alguns casos, superiores a 10^{20} . Assim, a estratégia de aumentar o número de neurônios na camada oculta foi descartada e optou-se por explorar a rede com melhor desempenho anteriormente encontrada.

Para tal, foram geradas outras 1.000 redes com 10 neurônios na camada oculta, com as funções de ativação gaussiana e identidade para as camadas oculta e de saída, respectivamente, e com apenas as quatro entradas consideradas relevantes (x, Tar2, Var3 e Vvlu). Verificou-se um aumento de desempenho, refletido no coeficiente de correlação que chegou a 0,4351 para os dados de teste. A Tabela 19 fornece os resultados correspondentes à melhor rede do tipo RBF encontrada para a predição do Epart.

O desempenho da melhor rede encontrada foi de 0,4216 e 0,4351, respectivamente, para os dados de treinamento e de teste. Estes valores foram considerados baixos e não satisfatórios e, como as redes MLP apresentaram melhores desempenhos, concluiu-se que as melhores redes para predição do número de partículas são redes do tipo MLP. Assim, o uso de redes RBF para a predição de Epart foi descartado e os testes relativos a este tipo de rede foram interrompidos. Este resultado era esperado,

porque as redes RBF são redes com funções de ativação locais na camada escondida e, portanto, tipicamente, empregadas para classificação.

Tabela 19: Resultados para melhor rede RBF para predição obtida.

Rede	Performance Treinamento/ Teste	Erro Treinamento/ Teste	Função de Ativação	
			Camada Oculta	Camada Saída
RBF 4-10-1	0,4216/ 0,4351	0,0031/ 0,0040	Gaussiana	Identidade
Entradas por Relevância	Var3, Tar2, x, Vvlu			

6.2 Redes de Classificação

Assim como na primeira abordagem, os primeiros testes usando redes neuronais para classificação da quantidade de partículas foi iniciado explorando as funções de ativação disponíveis no software utilizado. Para redes MLP foram testadas as funções identidade, logística, tangente hiperbólica, exponencial e seno. Já para redes RBF, a função de ativação para a camada oculta foi a função gaussiana e para a camada de saída foram testadas as funções identidade e Softmax. Foram analisadas, na ordem, as divisões propostas em 7, 5, 4, 3 e 2 classes, conforme apresentado na Figura 27.

As melhores redes MLP e RBF encontradas para cada número de classes, assim como seus desempenhos e características são apresentados na Tabela 20 e na Tabela 21. Nas redes MLP, todas as variáveis de entrada foram relevantes para o desempenho da rede e somente as que possuem maior influência são listadas na tabela. Nas redes RBF, todas as variáveis que se mostraram relevantes são listadas na tabela.

A análise dos resultados das redes MLP e RBF torna evidente que há uma melhora do desempenho das redes conforme o número de classes consideradas é reduzido, mostrando que uma classificação mais complexa leva a um aumento de erros na associação dos dados a uma classe. Por exemplo, para classificações em duas classes (caso “E”), as redes são capazes de classificar corretamente mais de 90% dos dados, o que é satisfatório.

Como a quantidade de dados em cada uma das classes é muito diferente, o desempenho geral não é suficiente para determinar a adequação da rede ao problema proposto, sendo recomendável, então, analisar separadamente o resultado obtido para cada classe. Assim, avaliou-se a quantidade de dados corretamente classificados em cada classe, gerando os resultados apresentados nas Tabela 22 e Tabela 23.

Tabela 20: Melhores redes MLP para classificações em 7,5,4,3 e 2 classes.

Número de Classes (Caso)	Rede	Performance Treinamento/ Teste	Função de Ativação		Principais Entradas por relevância
			Camada Oculta	Camada Saída	
7 (A)	MLP 11-9-7	57,3752/ 57,8402	Tangente Hiperbólica	Softmax	Var2, Tar2, Tar1, Var3
5 (B)	MLP 11-10-5	57,5601/ 58,7278	Tangente Hiperbólica	Softmax	Tar2, Var1, Var3, Var2
4 (C)	MLP 11-12-4	61,5157/ 60,3550	Logística	Softmax	Par2, Tar2, Var2, Var1
3 (D)	MLP 11-11-3	77,8928/ 80,1775	Tangente Hiperbólica	Softmax	Var1, Par2, Var2
2 (E)	MLP 11-11-2	96,1183/ 95,2663	Tangente Hiperbólica	Softmax	Par2, Var1, Tar1, Var2

Tabela 21: Melhores redes RBF para classificações em 7,5,4,3 e 2 classes.

Número de Classes	Rede	Performance Treinamento/ Teste	Função de Ativação		Variáveis de Entradas por relevância
			Camada Oculta	Camada Saída	
7 (A)	RBF 10-9-7	55,5638/ 54,5858	Gaussiana	Softmax	Par2, Var1, Par1, Var2, Tar2, Tar1, Pl, x, Var3, Tlu
5 (B)	RBF 11-10-1	56,0074/ 54,7337	Gaussiana	Softmax	Par2, Var1, Var3, Tar2, Pl, Tar1, x, Var3, Par1, Tlu, Vvlu
4 (C)	RBF 11-80-4	51,3494/ 53,9941	Gaussiana	Softmax	Par1, Var1, Var2, Par2, Var3, Tar1, Tar2, Vvlu, Pl, Tlu, x
3 (D)	RBF 11-12-3	73,6784/ 75,2959	Gaussiana	Softmax	Var1, Par2, Var2, x, Tar2, Par1, Tar1, Var3, Pl, Vvlu, Tlu
2 (E)	RBF 11-15-2	94,5656/ 93,7870	Gaussiana	Softmax	Var2, x, Pl, Tar1, Vvlu, Var3, Tlu, Par1, Tar2, Var1, Par2

Tabela 22: Melhores redes MLP para classificações em 7,5,4,3 e 2 classes.

Desempenho individual de cada classe: Porcentagem de classificações corretas								
Número de classes	Rede	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6	Classe 7
7	MLP 11-9-7	1%	90%	42%	1%	14%	3%	0%
5	MLP 11-10-5	1%	90%	38%	0%	31%	-	-
4	MLP 11-12-4	8%	89%	30%	55%	-	-	-
3	MLP 11-11-3	91%	56%	33%	-	-	-	-
2	MLP 11-11-2	99%	42%	-	-	-	-	-

Tabela 23: Melhores redes RBF para classificações em 7,5,4,3 e 2 classes.

Desempenho individual de cada classe: Porcentagem de classificações corretas								
Número de classes	Rede	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6	Classe 7
7	RBF 10-9-7	1%	90%	42%	1%	14%	3%	0%
5	RBF 11-10-1	0%	94%	29%	0%	1%	-	-
4	RBF 11-80-4	2%	85%	26%	23%	-	-	-
3	RBF 11-12-3	92%	44%	8%	-	-	-	-
2	RBF 8-13-2	99%	14%	-	-	-	-	-

Percebe-se claramente que os modelos com uma quantidade maior de classes apresentam classes em que o índice de acerto é próximo de zero, o que é consistente com seu baixo desempenho geral.

Os modelos com número menor de classes, cujos desempenhos gerais são elevados, apresentam desempenhos individuais muito heterogêneos. Para a classe com maior quantidade de dados, o índice de acerto é muito elevado. Não obstante, o mesmo não ocorre para as demais classes. Isto mostra que a rede não é capaz de apresentar um desempenho equilibrado em todas as classes quando a distribuição dos dados não é uniforme. Conseqüentemente, o desempenho geral, por si só, não é uma boa métrica para a determinação da qualidade de uma rede.

A rede que apresentou os melhores resultados foi a MLP 11-11-2 (com duas classes). Seu desempenho geral foi superior a 95% e os desempenhos individuais das classes 1 e 2 foram, respectivamente, 99% e 42%. Fica claro que o desempenho geral reflete basicamente o índice de acerto da classe 1, que contém 94% dos dados disponíveis. Esta rede não foi considerada satisfatória para o monitoramento do número de partículas, pois o maior interesse é justamente identificar situações anormais de operação, correspondentes à classe 2.

Costa et al. (2005) propuseram, para o caso “E” com separação dos dados em duas classes, o uso de uma rede neuronal do tipo RBF apresentando 3 camadas e 32 neurônios na camada oculta, com 5 variáveis de entrada (Par2, Var1, Par1, Var3, Tlu) conforme mencionado na seção 3.4.1. O desempenho geral foi de cerca de 80,0% e classificou corretamente 80,6% dos dados de classe 1 e 80,2% dos de classe 2. Tentou-se obter uma rede com as mesmas características descritas por Costa et al. (2005), entretanto as redes obtidas apresentaram melhores desempenhos gerais, mas não com o mesmo equilíbrio entre os desempenhos individuais de cada classe.

Vale ressaltar que os dados de Costa et al. (2005) apresentam uma taxa muito elevada - cerca de 80% - de falsos alarmes, ou seja, pontos incorretamente classificados como classe 2 e conseqüentemente como operação anormal. Somente a título de comparação, a rede MLP 11-11-2 obtida neste estudo apresenta um índice menor que 30% de falsos alarmes. Em resumo, o aumento do desempenho individual na classe 2 é obtido às custas de um aumento da quantidade de falsos positivos.

6.3 Abordagem “híbrida”: Predição e Classificação

Tendo em vista a impossibilidade de obter resultados melhores para a classificação da variável Epart uma nova estratégia foi proposta. Decidiu-se testar a eficiência de uma abordagem “híbrida” que aliasse os resultados obtidos pelas redes MLP de predição à estratégia de classificar o número de partículas.

Dessa forma, a rede MLP de melhor desempenho (ver Tabela 17) foi usada para predição do número de partículas. A partir dos valores preditos tanto no teste quanto no treinamento, os resultados foram classificados e separados nas classes estabelecidas e discutidas anteriormente. Foi realizada uma comparação entre a classificação feita a partir dos valores preditos e os reais, avaliando-se, então, a eficiência da classificação deste método, ou seja, a capacidade de classificar corretamente cada caso. Isso foi feito verificando a porcentagem de acerto na classificação dos dados de acordo com cada classe.

Por exemplo, no caso de duas classes, em que valores de Epart superiores a 200 pertencem à classe 2, 98,79% dos dados pertencentes à classe 1 foram classificados corretamente, no entanto, somente 39,74% dos dados de classe 2 foram corretamente identificados. A Tabela 24 apresenta os resultados obtidos para classificação do número de partículas em 7, 5, 4, 3 e 2 classes.

Tabela 24: Resultados classificação após predição do número de partículas.

Número de classes	%Classe 1	%Classe 2	%Classe 3	%Classe 4	%Classe 5	%Classe 6	%Classe 7	% Total
2	98,79%	39,74%	-	-	-	-	-	95,49%
3	85,41%	58,31%	39,74%	-	-	-	-	75,27%
4	3,21%	80,87%	44,88%	52,19%	-	-	-	58,96%
5	3,21%	80,87%	44,88%	29,17%	39,74%	-	-	56,64%
7	3,21%	80,87%	44,88%	29,17%	25,89%	25,93%	25,00%	55,86%

Percebe-se que as classes que agrupam valores elevados de Epart (em negrito na tabela acima) apresentam uma baixa eficiência de classificação. Essa tendência já havia sido verificada anteriormente no estudo das redes de classificação, entretanto, observou-se uma melhora significativa para algumas classes, principalmente para a classificação em sete classes. Para os valores de Epart entre 0 e 50, que correspondem à classe 1 dos modelos com números de classes iguais a 4, 5 e 7, percebe-se que foi mantida a dificuldade de modelagem.

A baixa eficiência para a classificação, e, portanto, para a predição de valores elevados de Epart, conforme explicado anteriormente, é consequência da má distribuição dos dados disponíveis e da pequena quantidade de casos em que eles ocorrem. Dessa forma, a rede recebe, para seu treinamento, milhares de dados referentes a condições normais de operação, tendo exemplos suficientes para ser bem treinada e conseguir, assim, fazer a predição com bom desempenho para esses casos.

No entanto, ela recebe uma quantidade significativamente inferior de dados com valores elevados de Epart (menos de 6% dos dados correspondem a Epart maior que 200) e a rede é, então, incapaz de aprender e de identificar estes padrões com o mesmo índice de acerto. Entretanto, como são poucos os casos de operação anormal, a eficiência geral da rede é pouco afetada pelos erros correspondentes a estes casos. O impacto desta heterogeneidade dos dados para classificação em duas classes, assim como a importância de falsos alarmes serão tratados a seguir.

6.3.1 Impacto da distribuição não equilibrada dos dados para classificação em duas classes - Relação entre desempenho de redes neuronais de classificação e falsos alarmes

Como este trabalho se insere no conceito de Indústria 4.0, a ocorrência de alarmes falsos não pode ser desconsiderada. O gerenciamento de alarmes é hoje considerado crucial para a operação segura e produtiva (WRIGHT, 2011) e é mandatório pela norma ANSI/ISA-18.2-2009. Entre os objetivos do gerenciamento de alarmes encontram-se a quantidade e a relevância dos alarmes apresentados ao operador.

À primeira vista, o desempenho de uma rede neuronal de classificação definido pela percentagem de classificações acertadas parece ser uma boa métrica de qualidade. No entanto, quando a rede é treinada para identificar em algumas classes as situações normais (ou seguras) de operação, e em outras classes as condições anormais, é muito provável que existam mais dados para as primeiras situações.

Considerando uma classificação em duas classes, em que a classe 1 corresponde a operações normais e a classe 2 a operações anormais, é possível encontrar uma relação entre o desempenho e o número de falsos alarmes.

Sejam:

p_i percentual de acerto da classe i , ou seja, quantos dados da classe i que são corretamente classificados na classe i ; este valor foi também chamado de desempenho individual da classe;

n'_i número de dados da classe i ;

m_i quantidade de dados que foram classificados pela rede em cada classe.

A partir destes dados pode-se determinar a quantidade de dados que foram classificados pela rede em cada classe (m_i) pelas Equações (28) e (29).

$$m_1 = p_1 n'_1 + (1 - p_2) n'_2 \quad (28)$$

$$m_2 = p_2 n'_2 + (1 - p_1) n'_1 \quad (29)$$

O desempenho geral da rede, p_r , também pode ser calculado como a relação entre os dados corretamente classificados e o total de dados pelas Equações (30) e (31).

$$p_r = \frac{p_1 n'_1 + p_2 n'_2}{n'_1 + n'_2} = \frac{p_1 + p_2 (n'_2/n'_1)}{1 + (n'_2/n'_1)} \quad (30)$$

$$p_r = \frac{p_1 + p_2(n'_2/n'_1)}{1 + (n'_2/n'_1)} \quad (31)$$

A equação acima comprova que o índice de acerto da classe 2 (situação anormal) tem pouco efeito sobre o desempenho global quando n_2 é muito menor que n_1 . Nos dados utilizados neste trabalho, a razão n'_2/n'_1 é igual a 0,062.

Quando a rede neuronal é utilizada para alertar o operador do risco de operar, a porcentagem de alarmes falsos, p_{af} , é dada pela relação entre o número de conjuntos de dados **incorretamente** classificados na classe 2 e o número **total** de conjuntos de dados classificados na classe 2 (Equação (32)).

$$p_{af} = \frac{m_2 - p_2 n'_2}{m_2} = 1 - \frac{p_2 n'_2}{m_2} \quad (32)$$

Usando a equação acima para m_2 , têm-se as Equações (33) e (34).

$$p_{af} = 1 - \frac{p_2 \cdot n'_2}{p_2 \cdot n_2 + (1 - p_1) \cdot n'_1} \quad (33)$$

$$p_{af} = 1 - \frac{1}{1 + \frac{(1 - p_1) \cdot n'_1}{p_2 \cdot n'_2}} \quad (34)$$

A Tabela 25 abaixo mostra o percentual de falsos alarmes obtidos nas condições citadas neste trabalho para o método de predição e posterior classificação com e sem o uso de repetições e no trabalho de Costa et al. (2004).

Tabela 25: Comparação entre os desempenhos e porcentagens de alarmes falsos dos métodos de classificação tratados neste estudo.

Redes	n₁	n₂	p₁	p₂	p_r	p_{af}
Costa et al. (2005)	3184	197	80,6%	80,2%	80,6%	79,6%
Abordagem híbrida - Rede MLP predição e classificação	3184	197	99,0%	43,0%	95,7%	27,3%
Abordagem híbrida - Rede MLP com dados adicionais predição e classificação	3184	3152	94,6%	77,6%	86,1%	6,6%

Percebe-se que os resultados de Costa et al. apresentam valores satisfatórios e homogêneos para os desempenhos geral da rede e individuais de cada classe, todos de aproximadamente 80%. Entretanto, a porcentagem de alarmes falsos é significativamente alta (79,6%), o que não é recomendado para um modelo a ser usado no controle de processos. O resultado da inclusão de dados nas redes de abordagem híbrida, que será explicada na seção 6.3.2, mostra que é possível obter um modelo com uma porcentagem de alarmes falsos muito inferior, de apenas 6,6%. Isso se justifica pela melhor distribuição dos dados nas classes 1 e 2.

A Figura 32 ilustra a relação entre o número de alarmes falsos e os índices de desempenho local para uma razão n_2/n_1 igual a 0,062 (curva azul) e para valores iguais de n_1 e n_2 (curva verde). A curva azul corresponde à relação entre os dados utilizados neste trabalho.

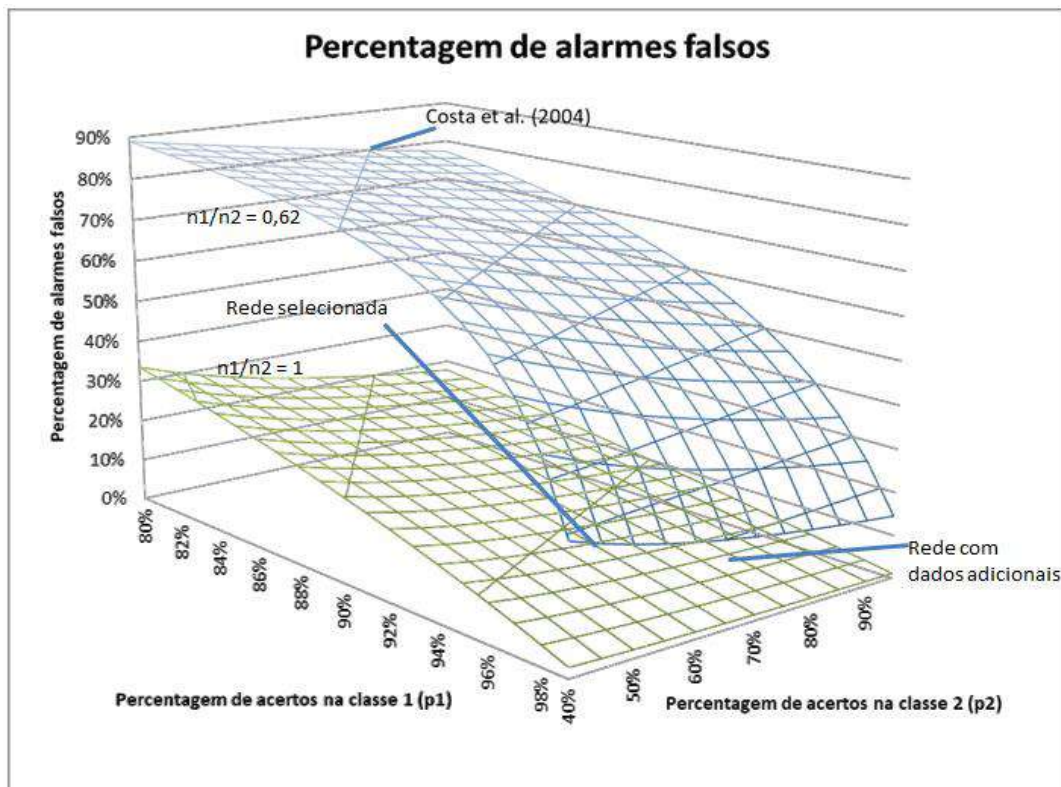


Figura 32: Relação entre o número de alarmes falsos e os índices de desempenho local para uma razão n_2/n_1 igual a 0,062 (curva azul) e para valores iguais de n_1 e n_2 (curva verde).

6.3.2 Adição de dados ruidosos

Decidiu-se, então, introduzir uma maior quantidade de dados correspondentes a E_{part} maior que 200 com o objetivo de confirmar a hipótese de que a eficiência de predição e de classificação melhora com uma distribuição de dados mais equilibrada.

Devido à indisponibilidade de dados reais da planta, propôs-se simular a medição de novos dados em condições similares a partir dos existentes. Levando em consideração que as medições das variáveis operacionais estão sujeitas a ruídos, os novos dados foram gerados pela adição de um ruído aleatório a cada uma das variáveis do conjunto de dados referentes a E_{part} superior a 200. A geração dos ruídos foi realizada com auxílio da função “*randn*” (*normally distributed random numbers*) do Matlab[®], que retorna um valor aleatório de uma distribuição normal. O ruído adicionado era normal, com média zero e desvio-padrão igual a 0,001. Esses novos dados podem ser interpretados como novas medidas de condições próximas àquelas verificadas durante a aquisição de dados. A adição de ruído gaussiano evita, também, que os dados sejam simplesmente repetidos e que a rede seja treinada diversas vezes com valores iguais, o que poderia prejudicar a sua capacidade de generalização.

A sensibilidade do método híbrido de predição com posterior classificação em relação à quantidade de dados de E_{part} elevado acrescentados (após adição de ruído) foi testada. Os dados originais de classe de operação anormal após adição de ruído gaussiano foram repetidos 2, 4, 6, 8, 10, 12, 14 e 16 vezes para cada teste. A adição de 16 repetições (com ruído gaussiano) permite que a distribuição seja mais equilibrada com número praticamente idêntico de dados nas classes 1 e 2.

Assim, foram sendo adicionados novos dados de forma a aumentar a frequência de casos em que E_{part} é elevado. A cada adição, foram geradas 1.000 novas redes apresentando as mesmas características da rede de predição mostrada na Tabela 17 e a melhor rede foi selecionada. A mesma metodologia descrita acima foi usada e os valores de E_{part} preditos foram classificados. Esta classificação foi comparada à classificação do E_{part} real. A influência da adição dos novos dados na predição é apresentada na Tabela 26.

Tabela 26: Desempenhos das redes de predição de acordo com o número de repetições.

Repetições	Performance Treinamento	Performance Teste
0	0,7464	0,7612
2	0,8243	0,7991
4	0,8642	0,8279
6	0,8733	0,8562
8	0,8783	0,8704
10	0,8937	0,8734
12	0,8926	0,8823
14	0,8994	0,8858
16	0,8940	0,8862

Percebe-se, claramente, que o desempenho da rede aumenta com o número de repetições até se aproximar de uma assíntota. A partir de 12 repetições a melhora do desempenho se torna menor e o valor do coeficiente de correlação tende a estabilizar. Esta tendência é mostrada na Figura 33.

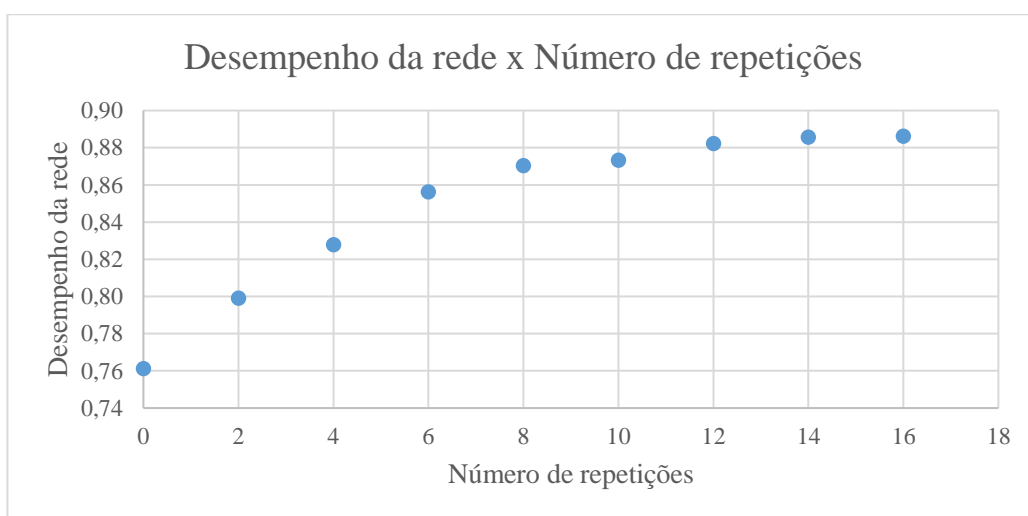


Figura 33: Influência do número de repetições no desempenho da rede.

As Tabelas 27, 28, 29, 30 e 31 apresentam os resultados da adição de novos dados ruidosos, indicando os desempenhos gerais de classificação (total) e os desempenhos individuais de cada classe obtidos pela abordagem híbrida.

Tabela 27: Comparação do desempenho de classificação de cada classe para divisão dos dados em 7 classes de acordo com o número de repetições.

7 CLASSES								
Repetições	Classen1 (%)	Classe 2 (%)	Classe 3 (%)	Classe 4 (%)	Classe 5 (%)	Classe 6 (%)	Classe 7 (%)	Total (%)
0	3,21	80,87	44,88	29,17	25,89	25,93	25,00	55,86
2	1,87	79,70	35,30	37,24	41,14	25,49	29,17	52,13
4	4,57	76,34	31,93	29,29	52,86	37,65	47,50	51,52
6	7,69	68,87	39,24	25,83	58,91	42,44	42,86	51,28
8	8,94	65,45	37,27	29,88	56,82	41,83	47,22	49,91
10	10,40	67,22	33,61	27,39	64,87	41,71	53,98	52,92
12	14,55	62,66	37,55	22,82	65,32	47,85	50,00	53,14
14	9,36	67,39	28,41	25,73	66,42	53,94	54,58	54,73
16	2,70	62,37	32,63	26,97	65,65	43,58	51,84	52,59

Para o caso de 7 classes, o desempenho das classes de 5 a 7, correspondentes aos dados adicionados, melhora com o número de repetições, enquanto o desempenho da classe 2 é reduzido. O desempenho geral da classificação não se altera significativamente.

Tabela 28: Comparação do desempenho de classificação de cada classe para divisão dos dados em 5 classes de acordo com o número de repetições.

5 CLASSES						
Repetições	Classe 1 (%)	Classe 2 (%)	Classe 3 (%)	Classe 4 (%)	Classe 5 (%)	Total (%)
0	3,21	80,87	44,88	29,17	39,74	56,64
2	1,87	79,70	35,30	37,24	57,46	55,26
4	4,57	76,34	31,93	29,29	67,58	55,72
6	7,69	68,87	39,24	25,83	72,69	56,70
8	8,94	65,45	37,27	29,88	73,33	57,01
10	10,40	67,22	33,66	27,39	77,13	59,86
12	14,55	62,66	37,55	22,82	77,64	60,52
14	9,36	67,39	28,41	25,73	77,64	61,62
16	2,70	62,37	32,63	26,97	77,64	61,26

Para a classificação em 5 classes, a influência da adição de dados é verificada, claramente, na classe 5, correspondente aos novos dados. O desempenho geral da classificação não se altera significativamente.

Tabela 29: Comparação do desempenho de classificação de cada classe para divisão dos dados em 4 classes de acordo com o número de repetições.

4 CLASSES					
Repetições	Classe 1 (%)	Classe 2 (%)	Classe 3 (%)	Classe 4 (%)	Total (%)
0	3,21	80,87	44,88	52,19	58,96
2	1,87	79,70	35,30	71,53	59,63
4	4,57	76,34	31,93	78,41	61,09
6	7,69	68,87	39,24	82,13	62,50
8	8,94	65,45	37,27	85,08	63,89
10	10,40	67,22	33,61	87,98	66,98
12	14,55	62,66	37,55	88,67	68,20
14	9,36	67,39	28,41	90,95	70,58
16	2,70	62,37	32,63	89,99	69,91

O modelo com 4 classes se comporta de forma similar ao de 5 classes, com a melhora do desempenho concentrada na classe 4.

Tabela 30: Comparação do desempenho de classificação de cada classe para divisão dos dados em 3 classes de acordo com o número de repetições.

3 CLASSES				
Repetições	Classe 1 (%)	Classe 2 (%)	Classe 3 (%)	Total (%)
0	85,41	58,31	39,74	75,27
2	83,36	54,58	57,53	72,08
4	81,74	52,84	67,58	71,82
6	76,33	54,89	72,69	70,76
8	75,66	57,56	73,33	71,35
10	77,40	53,99	77,13	73,13
12	73,96	53,68	77,64	72,24
14	76,24	51,37	77,64	73,06
16	68,41	53,78	77,64	71,01

No modelo de 3 classes, a adição de novos dados permite uma distribuição mais equilibrada que nos modelos anteriores. Houve uma melhora significativa na classe 3 às custas do desempenho da classe 1. O desempenho geral não foi muito alterado.

Tabela 31: Comparação do desempenho de classificação de cada classe para divisão dos dados em 2 classes de acordo com o número de repetições.

2 CLASSES			
Repetições	Classe 1 (%)	Classe 2 (%)	Total (%)
0	98,79	39,74	95,49
2	97,83	57,46	91,52
4	96,83	67,58	89,93
6	95,73	72,69	88,78
8	95,70	73,33	87,71
10	95,17	77,08	87,85
12	94,89	77,64	87,21
14	95,36	77,64	86,84
16	94,57	77,64	85,90

No modelo de 2 classes, obtêm-se os melhores resultados. O desempenho da classe 2 melhora de forma significativa, enquanto o desempenho da classe 1 sofre uma redução pequena. O desempenho geral é reduzido. Percebe-se que a partir de 10 repetições não há mudanças significativas nos desempenhos. A Figura 34 ilustra a evolução do desempenho geral e do da classe 2 com o número de repetições.

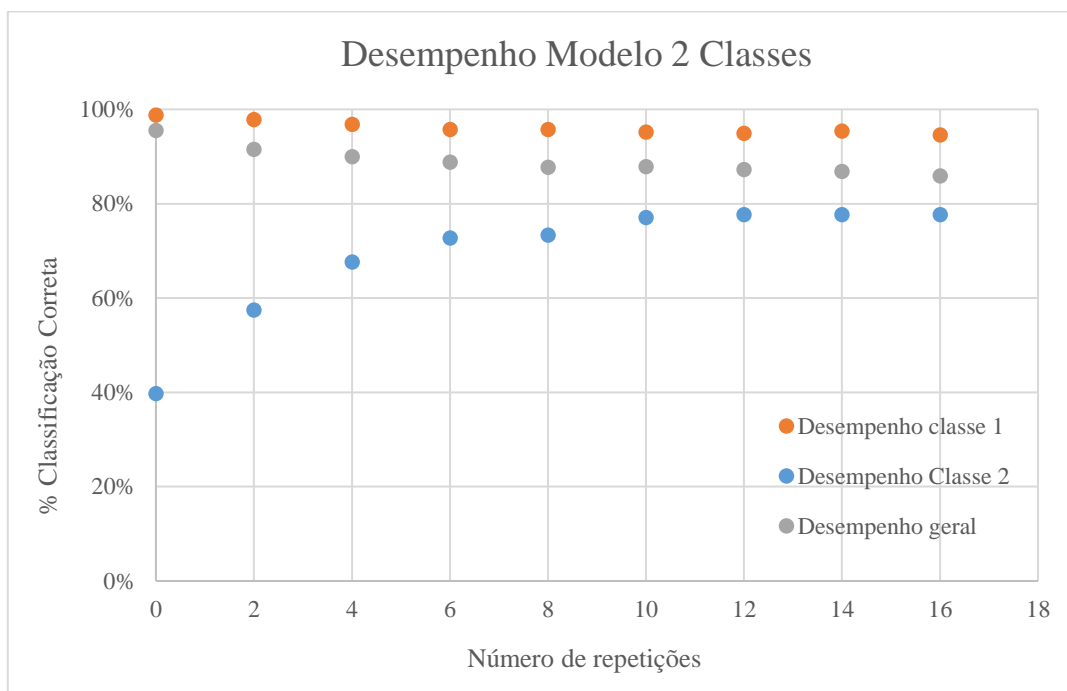


Figura 34: Desempenhos geral e individual da classe 2 com o aumento do número de repetições.

Comprova-se, portanto, que a qualidade da rede pode ser melhorada quando se dispõe de uma distribuição equilibrada de pontos em cada classe. A Figura 35 mostra que quando os dados estão igualmente distribuídos obtém-se o melhor desempenho para a classificação desta classe. Entretanto, a partir de 40% a melhora no desempenho passa ser insignificante.

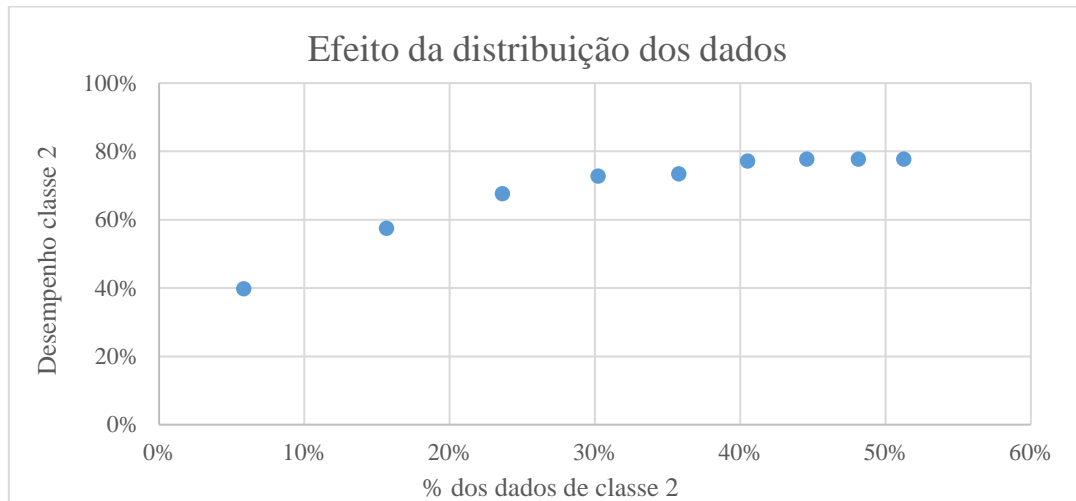


Figura 35: Efeito do equilíbrio da distribuição dos dados em duas classes no desempenho de classificação da classe 2.

Conclui-se, portanto, que para a modelagem do problema em duas classes, é desejável a obtenção de dados bem distribuídos em cada classe. Além disso, a metodologia de inclusão de dados ruidosos proposta foi bem sucedida ao acarretar melhorias significativas no desempenho das redes neurais para o problema de predição e classificação da quantidade de partículas formadas.

Os resultados acima permitem concluir que foi possível obter uma modelagem satisfatória do número de partículas formadas na caldeira de recuperação Kraft a partir de um modelo híbrido de predição e classificação, baseado em uma rede MLP de 3 camadas, com 11 neurônios na camada oculta, utilizando conjuntos adicionais de dados obtidos por adição de ruído gaussiano aos dados originais. Os melhores resultados foram obtidos com adição de 10 a 16 repetições.

6.4 Análise Cluster – Mapas Auto-Organizáveis

A abordagem final foi estudar o problema utilizando redes neurais de mapas auto-organizáveis em duas etapas. A análise das redes geradas se baseou na capacidade de agrupar dados com características similares em regiões visualmente identificáveis.

Na primeira etapa, o critério usado foi a classificação dos dados em duas classes, avaliando-se a capacidade de cada rede de apresentar regiões majoritariamente correspondentes a dados de classe 2. Na segunda etapa, a rede selecionada foi testada para os demais modelos com maior número de classes.

Foram geradas redes de topologia 2x2, 3x3, 1x100, 3x40, 3x60, 4x10, 5x7, 5x8, 5x20, 6x6, 10x10 e 20x5.

As redes 2x2 e 3x3 não apresentaram regiões claras e representativas de dados de classe 2 e não houve, portanto, uma separação satisfatória dos dados. Elas possuem poucos neurônios e nenhum deles conseguiu de fato agrupar uma quantidade significativa de dados de classe 2, que ficaram distribuídos nos neurônios da rede sem nenhum padrão notável.

A rede 1x100 apresentou dois neurônios agrupando majoritariamente dados de classe 2, porém não foi possível verificar a existência de uma região com grande concentração de dados desta classe.

As redes 6x6, 5x7 e 4x10 apresentaram apenas um neurônio majoritariamente de classe 2, entretanto apenas uma pequena parcela do total de dados de classe 2 foi efetivamente agrupada nestes neurônios. Dessa forma, estas redes permitiram apenas a identificação de, respectivamente, 6,59%, 6,60% e 5,58% dos dados de classe 2. Portanto, considerou-se que o agrupamento realizado por estas redes não é satisfatório.

As redes 3x60, 3x40, 5x20, 5x8 já apresentam uma melhor separação dos dados, mas geram regiões difusas. A Figura 36 apresenta como exemplo o resultado obtido para a rede 5x8. O mapa foi modificado de forma a indicar em cada neurônio a porcentagem de dados de classe 2 (operação anormal) que foram agrupados em tal neurônio. Foi usada uma escala de cores para destacar os neurônios com maiores concentrações de dados de classe 2. Nesta figura, os dados de classe 2 aparecem em praticamente metade das regiões sem ter predominância (porcentagem superior a 50%) em nenhuma delas, e uma região onde não há dados da classe 2.

0%	0%	0%	0%	2%	1%	31%	17%
0%	0%	0%	0%	1%	2%	13%	33%
1%	0%	0%	0%	1%	1%	4%	10%
0%	0%	0%	5%	20%	7%	6%	3%
2%	0%	0%	39%	17%	8%	32%	3%

Figura 36: Rede 5x8.

As redes 10x10 e 20x5 apresentaram os melhores resultados, com boa separação de regiões representativas de cada classe. Os mapas correspondentes a estas redes são

fornecidos na Figura 29, mostrada anteriormente na descrição da metodologia, e na Figura 37. Nestas figuras estão destacados os neurônios com mais de 20% de dados de classe 2, delimitando regiões com maior concentração de dados de classe 2.

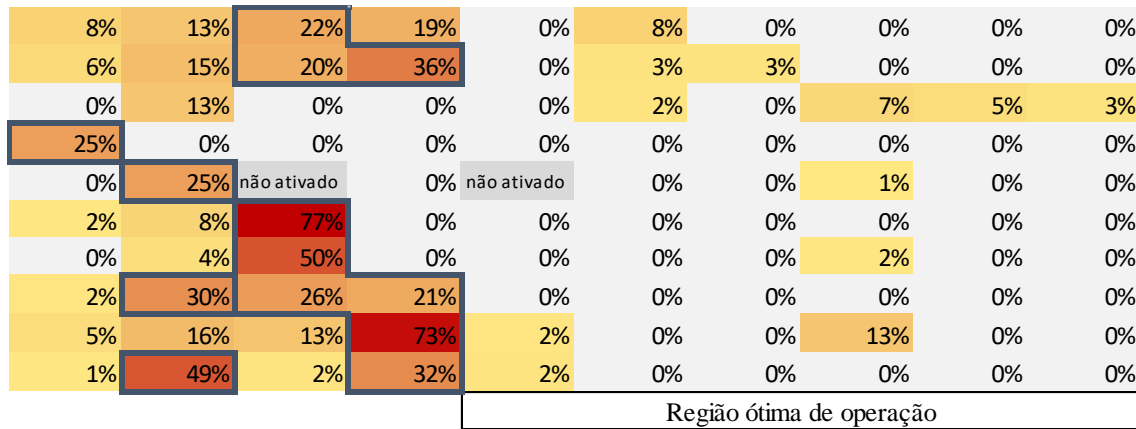


Figura 37: Rede 10x10.

É possível observar uma região de operação dentro do limite de formação de partículas desejado, visualmente do lado direito do mapa 10x10 ou na parte inferior do mapa 20x5. Estas duas topologias apresentam neurônios que possuem mais de 50% de dados de classe 2: três neurônios na rede 10x10 e quatro na rede 20x5. Para essas duas redes, calculou-se a quantidade de dados da classe 2 que são agrupados nestes neurônios majoritários. Na rede 10x10, os três neurônios majoritários capturam 10% dos dados de classe 2, enquanto os quatro neurônios majoritários da rede 20x5 capturam 27% destes dados. Em função disso, foi selecionada a rede 20x5 para a segunda etapa. A Figura 38 resume os resultados obtidos.

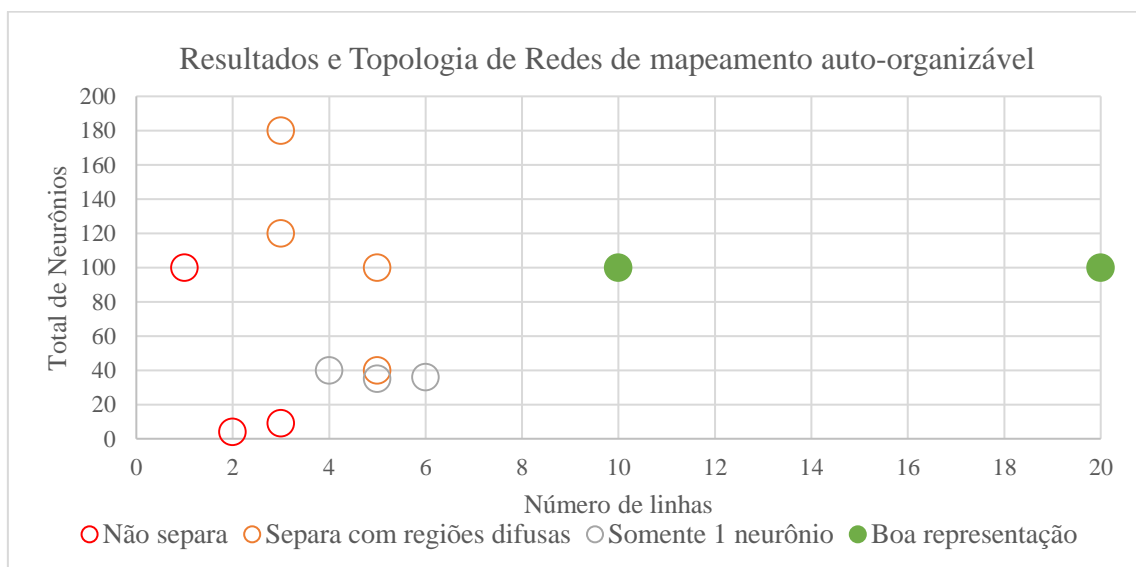


Figura 38: Resumo dos resultados da primeira etapa da análise de *clusters*.

A rede selecionada, com topologia 20x5, permite a identificação de regiões de operação normal e regiões de maior probabilidade de operação anormal. A Figura 39 é uma visualização do mapa gerado pela rede, em que as diferentes regiões são indicadas com cores diferentes. As regiões em verde indicam elevada probabilidade de operação sem problemas de emissão de partículas. As regiões em vermelho, pelo contrário, indicam maior probabilidade de emissão acima de 200 partículas/min. Os quatro neurônios majoritários estão destacados com vermelho mais escuro. As regiões amarelas podem ser entendidas como faixas de transição.



Figura 39: Mapa referente à rede 20x5.

A região em verde da Figura 39 concentra 72% de todos os conjuntos de dados, dos quais apenas 0,6% pertencem à classe 2. A região em vermelho agrupa somente 9% dos dados de operação, 61% dos quais pertencem à classe 2. Isto mostra que se um dado inédito for agrupado pela rede nesta região, haverá grande probabilidade de ele pertencer à classe 2. A região em amarelo agrupa 19% dos dados operacionais, dos quais 10% pertencem à classe 2. Percebe-se, então, que o mapa permite classificar os dados, porém existe uma faixa de indeterminação. Por um lado, uma parte dos dados de classe 2 não são indicados na região em vermelho (falsos negativos) e, por outro, alguns dados de classe 1 são classificados como operação anormal (falsos positivos).

Desta forma, o modelo gerado a partir de uma rede de mapeamento auto-organizável não foi capaz de classificar todos os dados operacionais em classes de operação normal e anormal, gerando um número significativo de alarmes falsos. Ou seja, o mapa não segregou totalmente os dados em regiões unicamente de operação normal ou anormal. Entretanto, o modelo gera um mapa com regiões com grandes probabilidades de operação normal (e com poucos falsos negativos) e de operação anormal, podendo ser utilizado para geração de recomendações operacionais.

Na segunda etapa da análise, explorou-se o comportamento da topologia 20x5 para um maior número de classes, obtendo-se os resultados apresentados na Figura 40, na qual são indicadas, no mapa, as classes predominantes em cada neurônio.

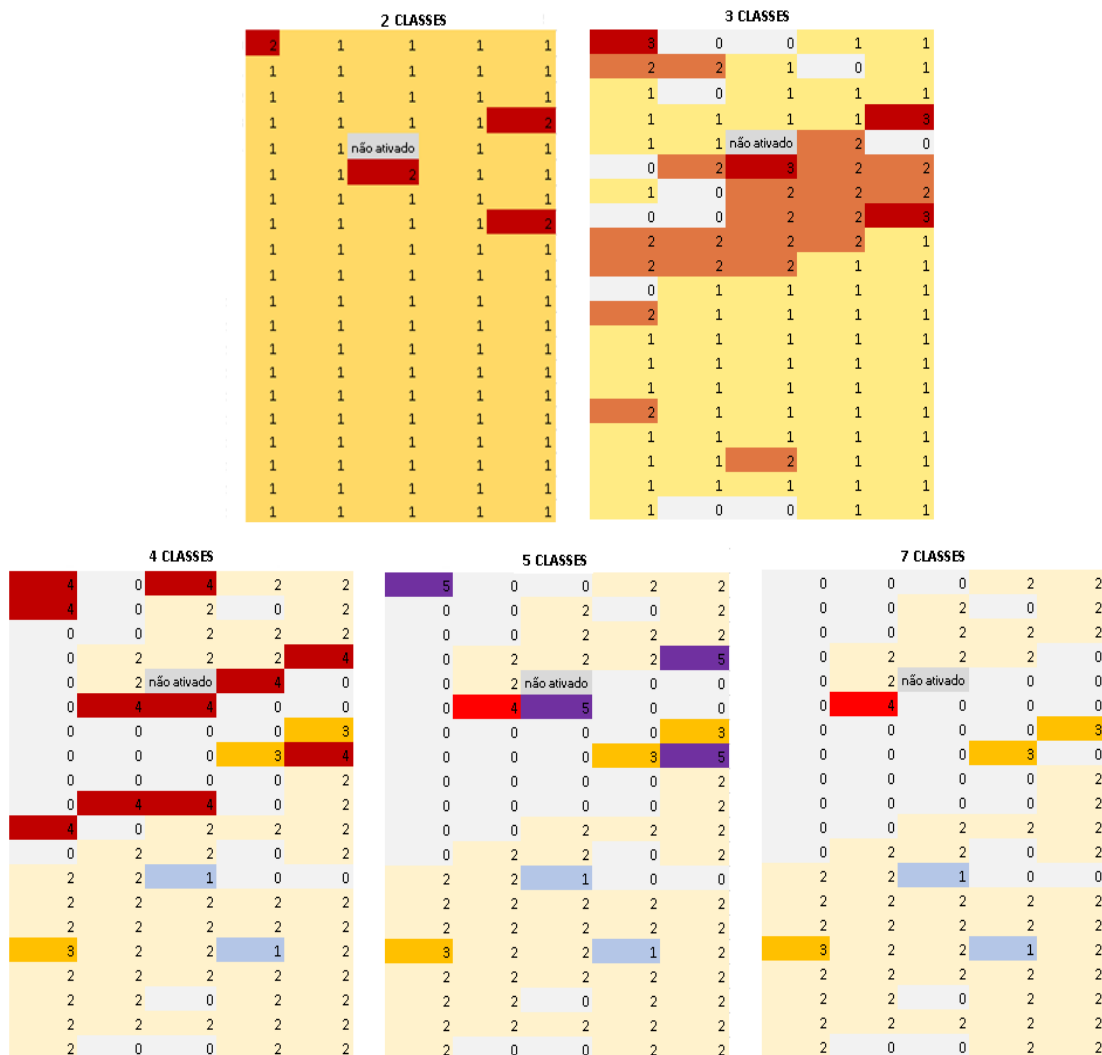


Figura 40: Comparação da rede de topologia 20x5 para análise em 2, 3, 4, 5 e 7 classes.

Como o critério utilizado para a seleção da classe predominantes foi a presença de mais de 50% de pontos da classe, apenas os quatro neurônios majoritários de classe 2 aparecem no mapa referente à classificação em duas classes.

Comparando o mapa obtido para classificação em três classes ao mapa da Figura 39, percebe-se que a região de transição (em amarelo) é bastante semelhante à região do mapa de três classes, onde predomina a classe 2. Portanto, o resultado para duas e três classes pode ser entendido como o mesmo resultado expresso de forma diferente. Entretanto, a análise com base em três classes apresenta uma desvantagem pela presença de zonas indefinidas, indicadas com o número zero na Figura 40.

Para a análise em quatro, cinco e sete classes, há um aumento da complexidade do mapa, com maior número de regiões de indefinição, sem que novas informações sejam obtidas. Para a classificação em sete classes, percebe-se, ainda, que as classes 5, 6 e 7 não conseguem ser representadas por nenhum neurônio, não sendo possível identificar no mapa nenhuma região de anormalidade operacional. A presença de mais regiões de indefinição é justificada pelo fato de que a separação em mais classes implica um maior espalhamento da distribuição dos dados em cada neurônio.

A modelagem do problema com base em mapas auto-organizáveis não foi capaz de segregar eficientemente os pontos anormais de operação, o que torna não recomendável sua utilização para o controle do processo. Entretanto, este método pode ser usado para gerar recomendações referentes à segurança operacional, indicando que a operação pode estar comprometida. A análise não mostrou benefícios em trabalhar com modelos com muitas classes; os melhores resultados foram efetivamente obtidos pela separação em duas ou três classes.

7. CONSIDERAÇÕES FINAIS

A caldeira de recuperação é um equipamento central nas plantas de produção de papel e celulose pelo processo Kraft e o controle de sua operação é essencial. As condições operacionais que levam à formação de material particulado na caldeira devem ser evitadas, pois a deposição nas superfícies do equipamento provoca perda de eficiência de troca térmica e paradas operacionais não programadas, representando prejuízos econômicos significativos. Dados da literatura sugerem que para uma operação adequada e segura da caldeira, a emissão de partículas deve ser inferior a 200 partículas/min (COSTA, 2004). A natureza complexa dos processos físicos e químicos que ocorrem durante a operação da caldeira de recuperação dificulta o desenvolvimento de modelos matemáticos para a modelagem da formação de partículas. Por isto, o presente trabalho propõe o uso de técnicas de inteligência artificial e de “*Machine Learning*” para a abordagem deste problema.

A formação de material particulado em uma caldeira de recuperação Kraft foi modelada pela aplicação de redes neurais dos tipos MLP, RBF e de mapeamento auto-organizável para o desenvolvimento de modelos preditivos e classificatórios a partir de dados históricos operacionais de uma planta industrial de polpa celulósica.

Nos estudos, foram adotadas três abordagens diferentes para a modelagem da formação de partículas visando ao controle e monitoramento na caldeira: a predição do número de partículas formadas na caldeira por redes preditivas; a classificação em classes pré-estabelecidas em função da quantidade de partículas por redes classificatórias; e a análise de agrupamentos, ou análise *cluster*, com mapas auto-organizáveis. Uma abordagem suplementar foi proposta para uma modelagem “híbrida”, combinando os conceitos de predição e de classificação, a partir da classificação dos valores estimados por uma rede de predição. Também foi avaliada a sensibilidade deste método à inclusão de novos dados de forma a reduzir o impacto de uma distribuição pouco equilibrada dos dados de entrada.

Os melhores resultados para as redes preditivas foram obtidos para uma rede de três camadas do tipo MLP com 11 variáveis de entrada e 11 neurônios ocultos, cujas funções de ativação da camada oculta e da camada de saída são, respectivamente, as funções tangente hiperbólica e exponencial. Seu desempenho, correspondente ao coeficiente de correlação entre os valores reais da variável predita e os valores calculados pela rede, foi superior a 0,76 para os dados de teste e seu erro foi de cerca de 0,2%. A

predição se mostrou satisfatória para a maioria dos pontos que correspondem às condições normais de operação da caldeira, ou seja, para formação de partículas inferior a 200 partículas/min. Entretanto, condições em que a emissão é superior a 200 partículas/min não apresentaram uma predição satisfatória, o que é justificado pela distribuição desequilibrada dos dados disponíveis: apenas 197 conjuntos de dados, dos 3.381, apresentam valores superiores a 200 partículas/min, quantidade insuficiente para o bom treinamento da rede. Como o objetivo principal é garantir que a rede seja capaz de prever se a operação está dentro da normalidade, identificando se o limite de 200 foi ultrapassado, a rede não foi considerada satisfatória apesar de seu bom desempenho geral.

As redes RBF também foram testadas para a predição da emissão de partículas, porém apresentaram desempenhos inferiores aos das redes MLP.

Na segunda abordagem, testou-se a capacidade de redes MLP e RBF de classificar o número de partículas em classes pré-estabelecidas. Estudou-se a divisão dos dados em modelos com quantidade de classes entre dois e sete. Os modelos com uma quantidade maior de classes apresentam índices de acerto de classificação (desempenho individual) para determinadas classes próximos a zero e conseqüentemente baixos desempenhos gerais. Os modelos com número menor de classes apresentam desempenhos gerais melhores, porém os desempenhos individuais de cada classe são muito heterogêneos: apenas a classe com maior quantidade de dados possui um alto índice de acerto. Isto mostra que a rede não consegue ter um desempenho equilibrado em todas as classes quando a distribuição dos dados não é uniforme.

A rede que apresentou os melhores resultados de classificação foi uma rede do tipo MLP com 11 variáveis de entrada, 11 neurônios na camada oculta e dois neurônios na camada de saída correspondendo a duas classes. Seu desempenho geral foi superior a 95% e os desempenhos individuais das classes 1 e 2 foram, respectivamente, 99% e 42%. Fica claro que o desempenho geral reflete basicamente o índice de acerto da classe 1, que contém 94% dos dados disponíveis. Assim como na abordagem anterior, esta rede não foi considerada satisfatória, pois o maior interesse é justamente a identificação de situações anormais de operação, correspondentes à classe 2.

Na abordagem “híbrida”, a metodologia de predição por redes seguida de classificação apresentou melhores resultados que os obtidos para as redes de classificação da segunda abordagem. Entretanto, nem todas as classes exibiram desempenhos satisfatórios, novamente por causa da má distribuição dos dados. Para testar a influência desta distribuição de dados, foram adicionados novos dados obtidos a partir da adição de

ruído gaussiano às variáveis dos conjuntos de dados correspondentes a emissões superiores a 200 partículas/min. A inclusão destes novos dados aumentou significativamente a eficiência de predição da rede, assim como o desempenho da classificação, comprovando que a qualidade da rede pode ser melhorada quando se dispõe de uma distribuição equilibrada de pontos em cada classe. O melhor modelo encontrado foi obtido a partir da classificação em duas classes, com adição de novos dados até que a classe com menor quantidade de dados possuísse, pelo menos, 40% do total de dados. Neste método, o desempenho da rede de predição foi superior a 0,87; o desempenho de classificação geral foi superior a 87% e o desempenho individual das classes 1 e 2 foram superiores, respectivamente, a 94% e 77%.

Os testes mostraram que é possível obter uma modelagem satisfatória do número de partículas formadas na caldeira de recuperação Kraft a partir de um modelo “híbrido” de predição e de classificação, baseado em uma rede MLP de 3 camadas, com 11 neurônios na camada oculta, utilizando conjuntos adicionais de dados obtidos por adição de ruído gaussiano aos dados originais que correspondiam à operação anormal. Os melhores resultados foram obtidos com a inclusão de 10 a 16 repetições de todos os dados de operação anormal após adição de ruído gaussiano.

Nos estudos com aplicação de redes de mapeamento auto-organizável, dentre as várias topologias testadas, foi possível obter uma rede de topologia 20x5 que permite a identificação de regiões de operação normal, regiões de transição e regiões de maior probabilidade de operação anormal, ou seja, com emissão de material particulado acima de 200 partículas/min. A região de operação normal concentra 72% de todos os conjuntos de dados, dos quais apenas 0,6% apresentam valores acima de 200 partículas/min. A região de maior probabilidade de operação anormal agrupa somente 9% dos dados de operação, 61% dos quais apresentam emissão de partículas superior ao limite operacional. A existência de dados de operação anormal localizados em regiões de operação normal gera falsos negativos, enquanto a situação inversa gera falsos positivos. Embora a modelagem do problema com base em mapas auto-organizáveis não tenha sido capaz de segregar eficientemente todos os pontos anormais de operação, o que limita severamente seu uso para o controle do processo, este método pode ser usado como um procedimento para gerar recomendações referentes à segurança operacional, indicando ao operador momentos em que a operação pode estar comprometida. A análise dos resultados não mostrou benefícios em se trabalhar com modelos com muitas classes, sendo os melhores resultados obtidos para classificação dos dados em duas ou três classes.

7.1 Sugestões para trabalhos futuros

A modelagem do complexo processo de formação de partículas em caldeiras de recuperação Kraft é ainda um desafio para a indústria de papel e celulose e, por isso, apresenta um grande potencial para futuras pesquisas, que poderão auxiliar na melhor compreensão deste processo.

Com base nos resultados obtidos no presente trabalho, que desenvolveu modelos com potencial aplicação industrial, foram identificadas oportunidades para aperfeiçoamento destes modelos em trabalhos futuros. Sugere-se, primeiramente, a aplicação da metodologia utilizando novos conjuntos de dados operacionais com uma distribuição mais uniforme em relação à quantidade de partículas formadas na caldeira.

Sugere-se também estudar a influência de outras variáveis operacionais que não estavam disponíveis nos dados aqui tratados. A formação de partículas pode ser influenciada de forma significativa, por exemplo, pela composição química do licor negro e pelas condições operacionais no interior da caldeira, tais como a temperatura e a pressão em diferentes pontos do equipamento. Sugere-se, também, que pesquisas futuras incluam na modelagem a análise da distribuição granulométrica das partículas, já que a deposição do material particulado segue mecanismos distintos e causa problemas operacionais específicos para diferentes faixas de granulometria.

Outra sugestão é o emprego da metodologia proposta neste trabalho para a modelagem de outros parâmetros de controle importantes para a indústria de papel e celulose, tais como a aplicação de redes neuronais na predição do número Kappa e do *Freeness*, que são parâmetros relevantes para o controle da digestão do material celulósico e do refino da polpa e que são usualmente determinados por métodos laboratoriais, o que dificulta seu controle em tempo real.

Recomenda-se, ainda, a exploração de outras características dos mapas auto-organizáveis, que mostraram um grande potencial para a identificação de padrões presentes nos dados operacionais.

8. REFERÊNCIAS BIBLIOGRÁFICAS

ADLER, E., Lignin chemistry: past present and future, **Wood Science and Technology**, v. 11, n. 3, p. 169-218, 1977.

ALVAREZ, A. B.; LUQUE, B. **Rede Neural de Kohonen e Outras Técnicas para Treinamento Não-Supervisionado**. Universidade Estadual de Campinas - Faculdade de Engenharia Elétrica e Computação, Campinas, p. 1-49, 2003.

AMERICAN NATIONAL STANDARDS INSTITUTE. International Society of Automation, **ANSI/ISA 18.2-2009: Management of alarm systems for the process industries**. 2009.

ANDRITZ OY. Helsinki, Finlândia. Kari Saviharju; Jorma Simonen; Liisa Simonen; Esa Vakkilainen; Esko Mattelmaki. **Combustion air system for recovery boilers, burning spent liquors from pulping processes**. US 7 207 280 B2, 30 dez. 2003, 24 abr. 2007.

ANTSAKLIS, P. J. Neural Networks in control systems. **IEEE Control Systems Magazine**, v. 10, n. 3, p. 3-5, 1990.

BAUGHMAN, D. R.; LIU, Y. A. **Neural Networks in Bioprocessing and Chemical Engineering**. Academic Press, Inc. San Diego, California 1995, 1-487 p.

BOUCARD, H. **Contributions to the understanding of the hydrothermal processes: application to black liquor**. 2014. 309 f. Tese (Doutorado em Engenharia Química e de Processo) - École des Mines d'Albi-Carmaux, Universidade de Toulouse, Toulouse, França, 2014.

CARDON, A; MÜLLER, D. N. **Introdução às Redes Neurais Artificiais**. 1994, 32 f. Dissertação (Pós-Graduação em Ciência da Computação) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 1994.

CASTRO F. C. C.; CASTRO M. C. F. Mapas Auto-organizados de Kohonen - SOM. In: **Redes Neurais Artificiais: Curso de Engenharia Elétrica**. PUCRS, Faculdade de Engenharia, Departamento de Engenharia Elétrica, Rio Grande do Sul, 2014. cap. 7, p. 1-19.

CASTRO, H. F. Papel e Celulose. In: **Processos químicos industriais II**. Universidade de São Paulo, Escola de Engenharia de Lorena - EEL, Lorena, São Paulo, 2009, apostila 4, p. 1-30.

CENTRO DE GESTÃO E ESTUDOS ESTRATÉGICOS (CGEE). **Eficiência Energética: recomendações de ações de CT&I em segmentos da indústria selecionados - Celulose e Papel**. Série Documentos Técnicos, n. 20, Brasília, DF, set. 2013.

CORREIA, F. M. **Métodos estatísticos e redes neurais aplicados a modelos preditivos em digestor contínuo de celulose kraft de eucalipto**. 2016, 183 f. Tese (Doutorado em Engenharia Química) - Universidade Estadual de Campinas, Departamento de Engenharia de Sistemas Químicos, Campinas, São Paulo, 2016.

COSTA, A. O. S. **Descrição matemática da caldeira de recuperação Kraft**. 2004, 150 f. Tese (Doutorado em Engenharia Química) - Coordenação dos Programas de Pós-Graduação de Engenharia da UFRJ (COPPE/UFRJ) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2004.

COSTA, A. O. S.; SOUZA Jr., M. B.; BISCAIA Jr., E. C; LIMA, E. L. Monitoring particulate material formation in a kraft furnace recovery boiler. **Tappi Journal**, Atlanta, Estados Unidos da América, v. 4, n. 4, p. E33-47, 2005.

DAVIN, L. B.; JOURDES, M.; PATTEN, A. M.; KIM, K.; VASSÃO, D. G.; LEWIS, N. G. Dissection of lignin macromolecular configuration and assembly: Comparison to related biochemical process in allyl/propenyl phenol and lignan biosynthesis. **Natural Products Reports**, v. 25, n. 6, p. 1015-1090, Royal Society of Chemistry, 2008. Disponível em: <<http://pubs.rsc.org/doi:10.1039/B510386J>> e <<http://public.wsu.edu/~lange-m/Documnets/Teaching2011/Davin2008.pdf>>. Acesso em: 19 maio 2017.

DE SOUZA Jr., M. B. **Redes neuronais multicamadas aplicadas à modelagem e controle de processos químicos**. 1993, 194 f. Tese (Doutorado em Engenharia Química) - Coordenação dos Programas de Pós-Graduação de Engenharia da UFRJ (COPPE/UFRJ) - Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, 1993.

DE SOUZA Jr., M. B. **Redes neurais com Statistica®**. In: Curso de treinamento para a Versão 6. Escola de Química da UFRJ. Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2003, apostila, 39 p.

DE SOUZA Jr., M. B. Introdução e Fundamentos das redes. In: **Redes neurais artificiais na engenharia química**. Universidade Federal do Rio de Janeiro, Escola de Química, Rio de Janeiro, 2005, apostila, cap. 1-2, p. 1-19.

DOPICO, M.; GOMEZ, A.; DE LA FUENTE, D.; GARCÍA, N.; ROSILLO, R.; PUCHE, J. A vision of industry 4.0 from an artificial intelligence point of view. In: The 18th International Conference on Artificial Intelligence, 2016, Las Vegas. **Proceedings**, Nevada, Estados Unidos da América, p. 407-413, jul. 2016. ISBN: 1-60132-438-3, CSREA Press.

DUFOUR, P.; BHARTIYA, S.; DHURJATI, P.S.; DOYLE III, F. J. Neural network-based software sensor: Data set design and application to a continuous pulp digester. **Control Engineering Practice, International Federation of Automatic Control**, Elsevier, v.13, n. 2, p. 135-143, 2005. ISSN: 0967-0661.

FERREIRA, D. J. O.; CARDOSO, M.; PARK, S. W. Gas flow analysis in a Kraft recover boiler. **Fuel processing Technology**, v. 91, p. 789-798, 2010.

FREDERICK, W. J., HUPA, M. Black Liquor Droplet Burning Processes. Adams, T. N. (ed), In: **Kraft Recovery Boilers**, 1^a Ed., Atlanta, Estados Unidos da América, TAPPI Press, cap. 5, p. 131-160, 1997, 381 p. ISBN 0-9625985-9-3.

GALDIANO, G. P. **Inventário do ciclo de vida do papel offset produzido no Brasil**. 2006. 280 f. Dissertação (Mestrado em Engenharia) - Escola Politécnica da Universidade de São Paulo, Departamento de Engenharia Química, São Paulo, 2006.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Introduction. In: **Deep learning**. Cambridge, MA, Estados Unidos da América, MIT Press, cap. 1, p. 1-28, 2016. 800 p. ISBN: 978-0262035613. Disponível em: <<http://www.deeplearningbook.org/contents/intro.html>>. Acesso em 25 julho 2017.

GRACE, T. M., CAMERON, J. H., CLAY, D. T. **Char Burning, Project 3473-6:** Summary Technical Report para American Paper Institute - Recovery Boiler Committee, The Institute of Paper Chemistry - Appleton, Wisconsin, 1985.

GRACE, T. M.; FREDERICK, W. J.; SALCUDEAN, M.; WESSEL, R. A. **Black Liquor Combustion:** Validated Recovery Boiler Modeling. The Institute of Paper Science and Technology, Oregon State University, University of British Columbia, Babcock & Wilcox Company. Final Year Report, vol. 1, 1998, 248 p. Disponível em: <<https://smartech.gatech.edu/bitstream/handle/1853/1760/F016-vol1--1998-09.pdf>>. Acesso em 30 maio 2017.

HATTON, J. V. Development of yield prediction equations in Kraft pulping. **Tappi Journal**, v. 56, n. 7, p. 97-111, 1973.

HOCHULI, A. G. **Redes Neurais Convolucionais.** 2016. Trabalho apresentado como atividade do Doutorado. Programa de Pós-Graduação em Informática - Departamento de Informática. Universidade Federal do Paraná, Curitiba, 2016. Disponível em: <<http://www.inf.ufpr.br/aghochuli/Café/>> e <http://www.inf.ufpr.br/aghochuli/café/CNN_PPT.pdf>. Acesso em 15 maio 2017.

HOLMBLAD, H. F. **Low temperature corrosion in black liquor recovery boilers due to hygroscopic salts.** 2015. 108 f. Dissertação (Mestrado em Química Inorgânica) - Abo Akademi University, Faculdade de Ciência e Engenharia, Turku, Finlândia, 2015.

INDÚSTRIA BRASILEIRA DE ÁRVORES (IBÁ). Relatório Anual 2016, Brasília, 2016, 100 p. Disponível em: <http://iba.org/images/shared/Biblioteca/IBA_RelatorioAnual2016_.pdf>. Acessado em 01 maio 2017.

JARVINEN, M.; ZEVENHOVEN, R.; VAKKILAINEN, E. **Implementation of a detailed physical black liquor combustion model into furnace calculations.** IFRF Combustion Journal, International Flame Research Foundation. n. 200206, p. 1-34, 2002. ISSN 1562-479X. Disponível em: <http://www.industrial.combustion.ifrf.net/paper_download.html?paperId=36>. Acesso em: 20 maio 2017.

LIMA, A. F., RÓDES, L., PHILLIPP, P. **Celulose e Papel: Tecnologia de Fabricação de Pasta Celulósica**, v. 1, 2ª Ed., IPT/SENAI, São Paulo, 1988, 559 p.

MACUKOW B. Neural Networks - State of art, brief history: Basic Models and Architecture. In: The 15th IFIP TC8 International Conference on Computer Information Systems and Industrial Management, CISIM 2016, LNCS 9842, Vilnius. **Proceedings**, Lituânia, 14-16 set. 2016, p. 3-14, 752 p. ISBN: 978-3-319-45377-4. DOI 10.1007/978-3-319-45378-1.

MIKKANEN, P. **Fly ash particle formation in Kraft recovery boilers**. 2000. 69 f. Tese (Doutorado em Tecnologia) - Helsinki University of Technology, Technical Research Centre of Finland, VTT Publications 421, Espoo, Finlândia, 2000.

MINISTÉRIO DO TRABALHO E EMPREGO (MTE). Norma Reguladora **NR-13: Caldeiras, vasos de pressão e tubulações**, 2014.

MORRISON, R. T.; BOYD R. N. **Química Orgânica**. 5ª ed., Tradução M. Alves Silva, Lisboa, Ed. Fundação Calouste Gulbenkian, Revista e alterada de acordo com a 2ª Ed. original, 1966, 1394 p.

MORRISON, R. T.; BOYD R. N. **Química Orgânica**. 13ª ed., Tradução M. Alves Silva, Lisboa, Ed. Fundação Calouste Gulbenkian, 1996, Tradução da 6ª Ed., 1992, 1506 p.

NILSSON, N. J. **Introduction to machine learning**: An early draft of a proposed textbook. Robotics Laboratory, Department of Computer Service. Stanford University, Stanford, Califórnia, Estados Unidos da América, 2005, p. 188. Disponível em: <<http://robotics.stanford.edu/people/nilsson/mlbook.html>>. Acesso em: 10 jul. 2017.

OTERO, F. Use of neural networks in process engineering: Thermodynamics, diffusion, and process control and simulation applications. **CT&F - Ciência, Tecnología y Futuro**, v.1, n. 4, p. 49-63, 1998.

PANDEY, K. K. A study of chemical structure of soft and hardwood and wood polymers by FTIR spectroscopy. **Journal of Applied Polymer Science**, v. 71, p. 1969-1975, John Wiley & Sons, Inc., 1999.

PAOLIELLO, F. A.; ALVES, A. C.; MAESTRI, E.; FELICÍSSIMO NETO, J.; ASPGREN, L. P. **Guia de inspeção de caldeiras de recuperação**. Comitê de Segurança em Caldeiras de Recuperação do Brasil - CSCRB. Associação Brasileira Técnica de Celulose e Papel - ABTCP, São Paulo, 2002.

PRASZKIER, S. **Influence of design and materials with consideration to the annual maintenance works in recovery boilers**. 2011. 104 f. Dissertação (Mestrado em Engenharia de Materiais) - Chalmers University of Technology, Gothenburg, Suécia, 2011.

RANTANEN, R. **Modelling and Control of Cooking Degree in Conventional and Modified Continuous Pulping Processes**. 2006. 140 p. Dissertação (Graduação em Tecnologia). Faculdade de Tecnologia - Departamento de Engenharia de Processo e Ambiental. Universidade de Oulu, Finlândia, 2006.

RENOVA. FPA, S.A. **Introdução aos Processos de Produção de Celulose e Papel**. 34 p. Disponibilizado pela Tecnicelpa - Associação Portuguesa dos Técnicos das Indústrias de Celulose e Papel, Portugal. 2000. Disponível em: <<http://www.tecnicepa.com/files/CelulosePapel.doc>>. Acesso em: 05 jan. 2017.

ROBERT, N. T. F. **Produção de Embalagem de Papel**. Serviço Brasileiro de Respostas Técnicas (SBRT). Dossiê técnico. Rede de Tecnologia e Inovação do Rio de Janeiro - REDETEC, 2007.

ROCHA, C.; ARMANI, E.; PIIRAINEN, E. Controle de Refinação de Celulose com Medição Contínua do Freeness. **O Papel**, v. 67, n. 7, p. 79-93, jul. 2006.

ROSA, R. A. F.; VALLE, M. E. Um estudo da aplicação de redes neurais auto-organizáveis para a identificação autônoma de fonemas portugueses. **Revista Ciências Exatas e Naturais**, v. 17, n. 2, p. 199-218, jul./dez. 2012.

RUMELHART, D. E.; MCCLELLAND, J. L. Learning internal representations by error propagation. **Parallel Distributed Processing**. Rumelhart, D. E. e McClelland (ed), J. L. MIT Press, Cambridge, 1986.

RÜSSMANN, M.; LORENZ, M.; GERBERT, P.; WALDNER, M.; JUSTUS, J.; ENGEL P.; HARNISCH, M. **Industry 4.0: The future of productivity and growth in**

manufacturing Industries, 2015. Disponível em: <http://www.inovasyon.org/pdf/bcg.perspectives_Industry.4.0_2015.pdf>. Acesso em: 20 maio 2017.

SILVA, W. L.; RIBEIRO, C. T.; COSTA, E, F. Reduction efficiency prediction of CENIBRA's recovery boiler by direct minimization of Gibbs free energy. **Brazilian Journal of Chemical Engineering**, v. 25, n. 3, São Paulo, 2008.

STATISTICA: STATISTICA[®] software. Version 8. Neural Networks. Statsoft, 2004.

THE MEAD CORPORATION. Hayton, Ohio. Michael A. Fornetti; Timothy A. Legault. **Combustion system for a black liquor recovery boiler**. US 5 715 763, 11 set. 1995, 10 fev. 1998.

TRAN, H. N. Recovery Boiler Fireside Deposits and Plugging Prevention. Tappi Kraft Recovery Course 2007, St. Petersburg, Flórida, Estados Unidos da América. **Proceedings**. Tappi Press, v. 2, p. 537-572, 8-11 jan. 2007. ISBN: 978-1-60560-063-5. Disponível em: <<http://www.tappi.org/content/events/08kros/manuscripts/4-7.pdf>> e <<http://toc.proceedings.com/02549webtoc.pdf>>. Acesso em 01 jun. 2017.

VAKKILAINEN, E. K. **Kraft recovery boilers: High dry solids firing**, Helsinki, Finlândia, 2006. 254 p. Disponível em: <https://www.researchgate.net/profile/Esä_Vakkilainen/publication/279871546_Kraft_recovery_boilers_-_Principles_and_practice/links/561a6dba08ae6d1730898b68/Kraft-recovery-boilers-Principles-and-practice.pdf>. Acesso em: 10 jun. 2017.

VAKKILAINEN, E. K. **Kraft recovery boilers: Principles and Practice**, In: Suomen Soodakattilayhdistys R. Y. (Ed.), Valopaino Oy, Helsinki, Finlândia, 2005, 246 p. ISBN 9529186037. Disponível em: <<http://www.doria.fi/handle/10024/111915>>. Acessado em: 19 jun. 2017.

VROOM, K. The H factor: A means of expressing cooking times and temperatures as a single variable. **Pulp and Paper Magazine of Canada**, v. 58, n. 3, p. 228-231, 1957.

WALTZ, F.; DE SOUZA Jr. M. B.; MORADA, T.; FILHO L. Indústria 4.0 no aumento dos resultados da indústria de papel e celulose. **Webinar Associação Brasileira Técnica de Celulose e Papel - ABTCP: Indústria 4.0**. 2017. Disponível em:

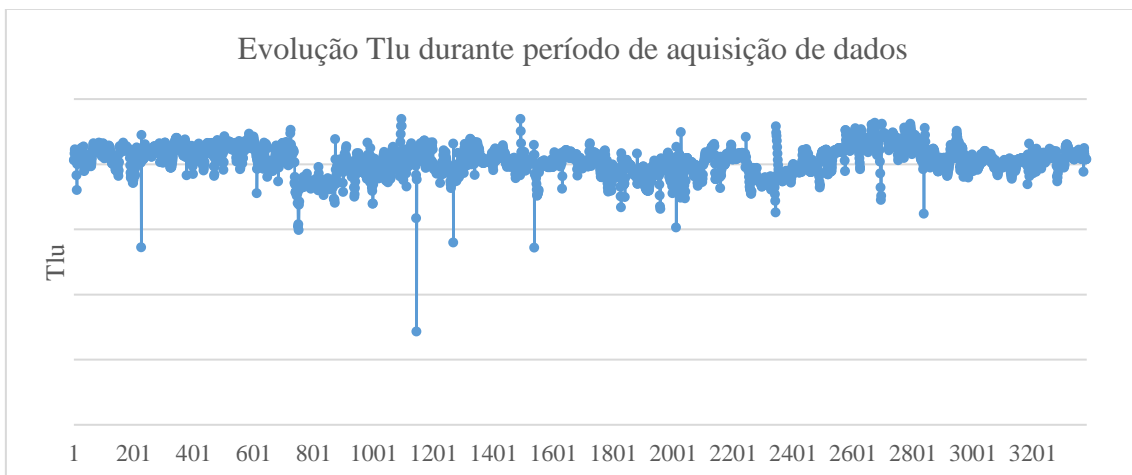
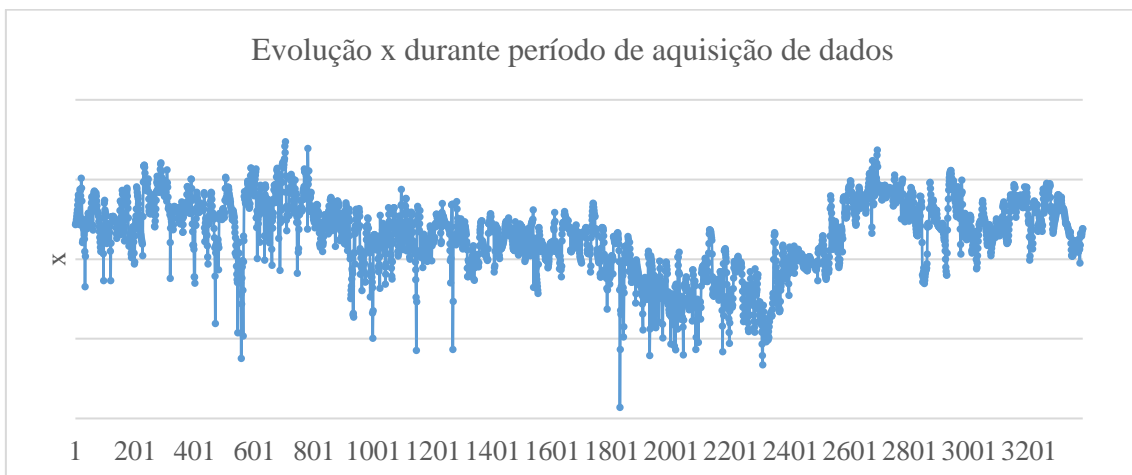
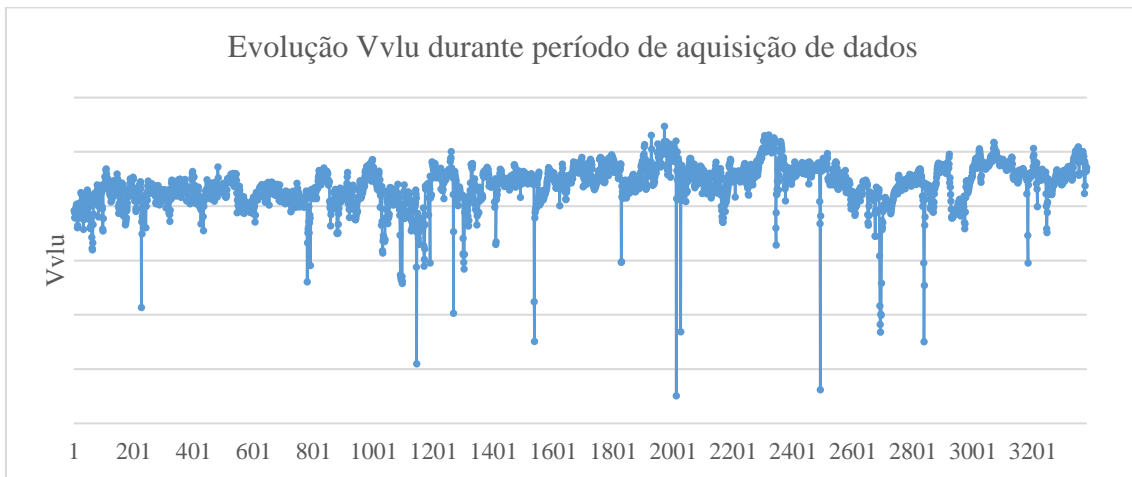
<<http://www.radixeng.com.br/docs/news/ApresentacaoABTCP.pdf> >. Acesso em: 20 jun. 2017.

WANG, K. Intelligence predictive maintenance (IPdM) system - Industry 4.0 scenario. Advanced Manufacturing and Automation V. **WIT Transactions on Engineering Sciences**, v. 113, p. 259-268, ISSN 1743-3533 (WIT Press on-line), 2016.

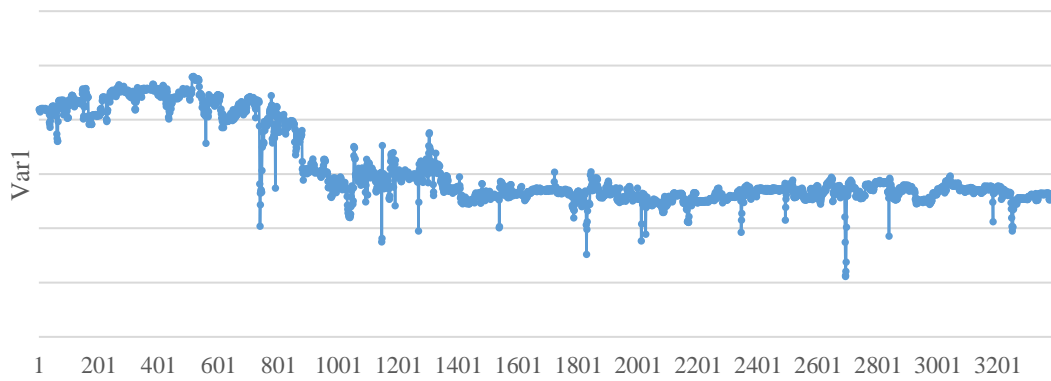
WRIGHT, J. P. Alarm Management Standards and Best Practices. In: Rockwell Automation: Process Solutions User Group (PSUG) Meeting, Chicago, Illinois, Estados Unidos da América. **Proceedings**. 14-15 nov. 2011, 51 p. Disponível em: <https://www.rockwellautomation.com/resources/downloads/rockwellautomation/pdf/events/automation-fair/2011/psug/afpsug11_ed16.pdf>. Acesso em: 30 jun. 2017.

APÊNDICES

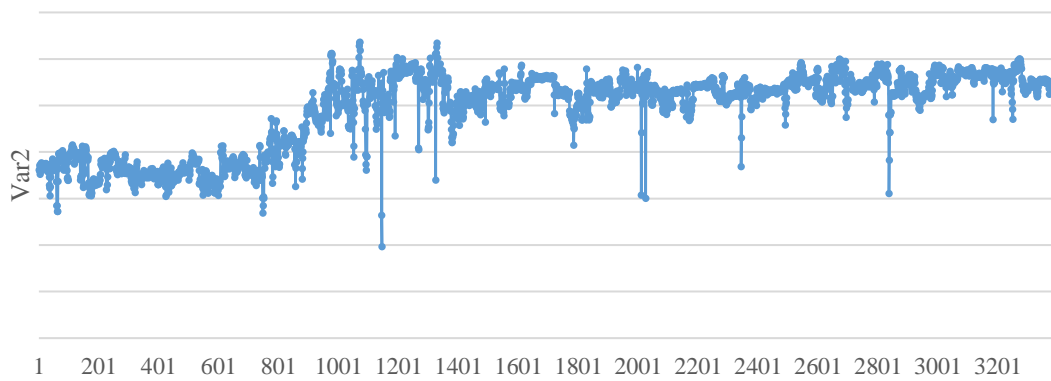
APÊNDICE 1: Gráficos da evolução das variáveis de entrada durante a aquisição de dados. Os valores das variáveis foram omitidos por razões de confidencialidade.



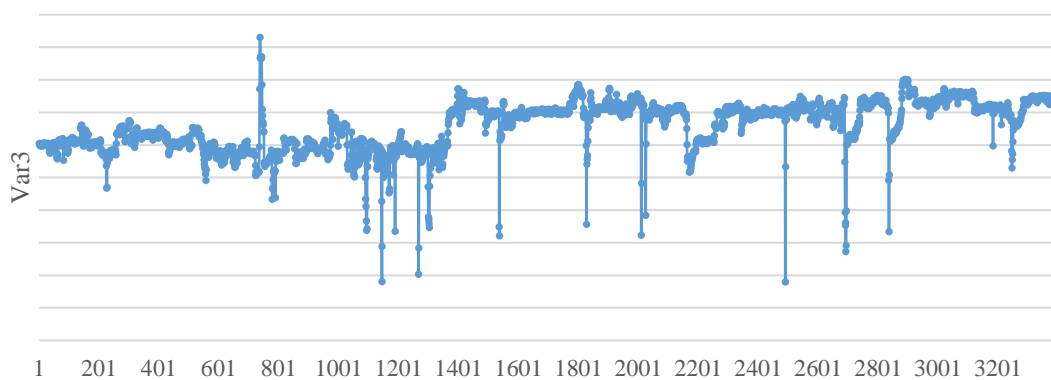
Evolução Var1 durante período de aquisição de dados



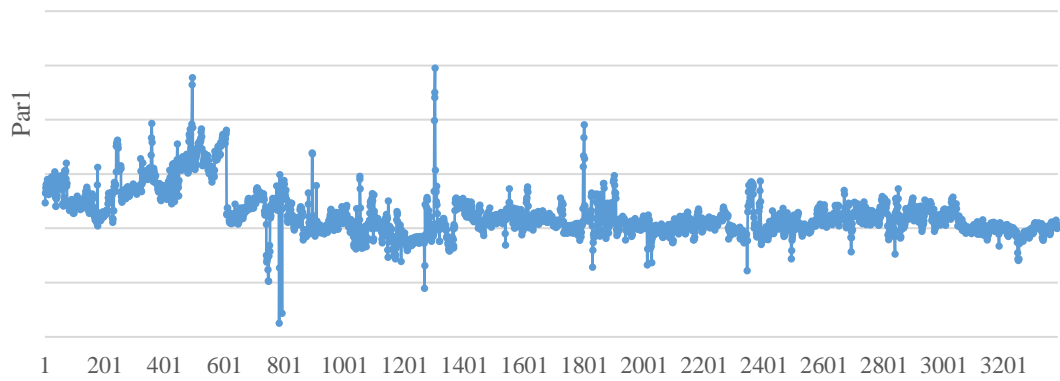
Evolução Var2 durante período de aquisição de dados



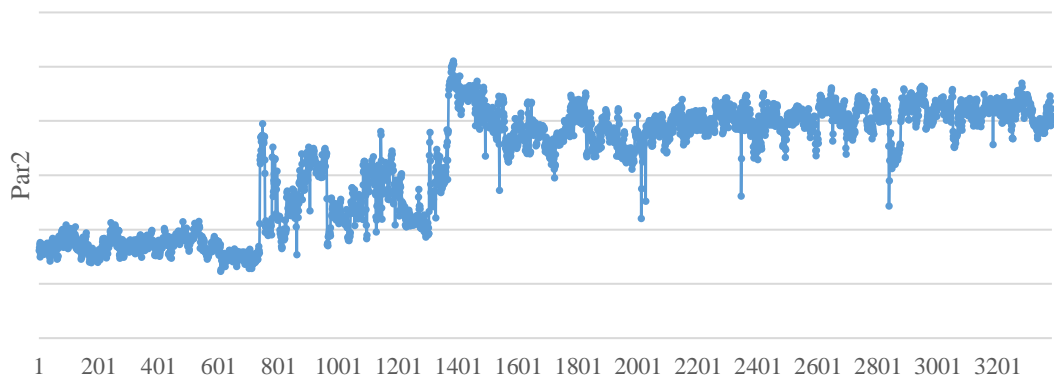
Evolução Var3 durante período de aquisição de dados



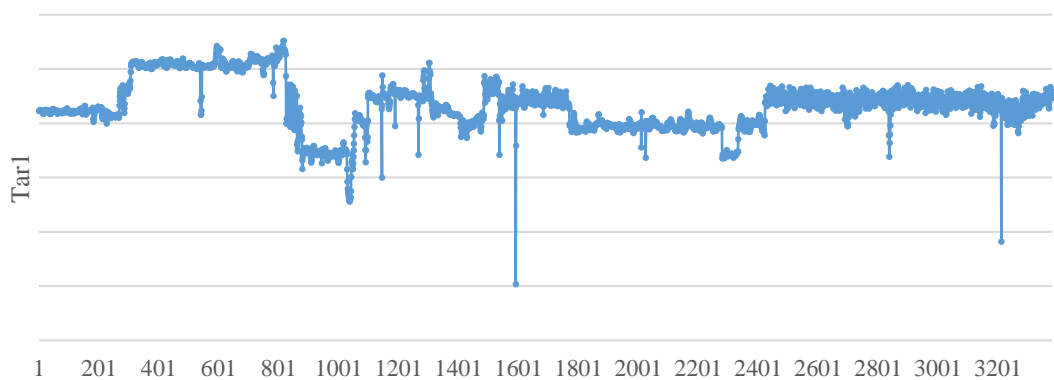
Evolução Par1 durante período de aquisição de dados



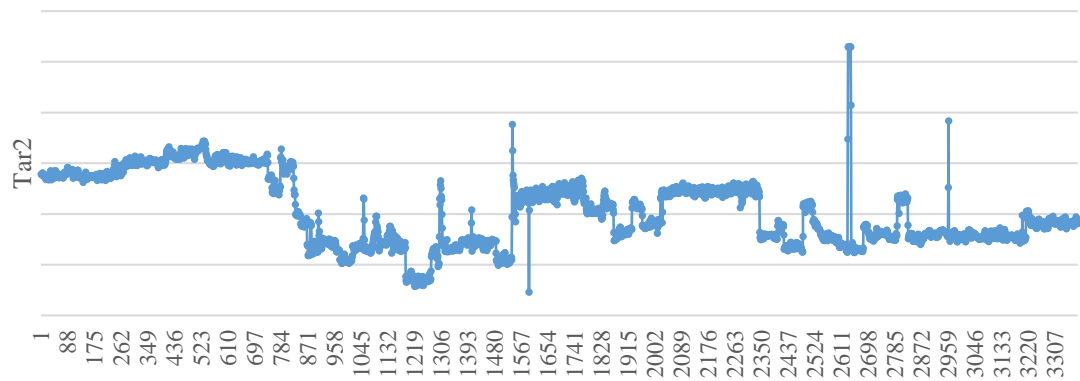
Evolução Par2 durante período de aquisição de dados



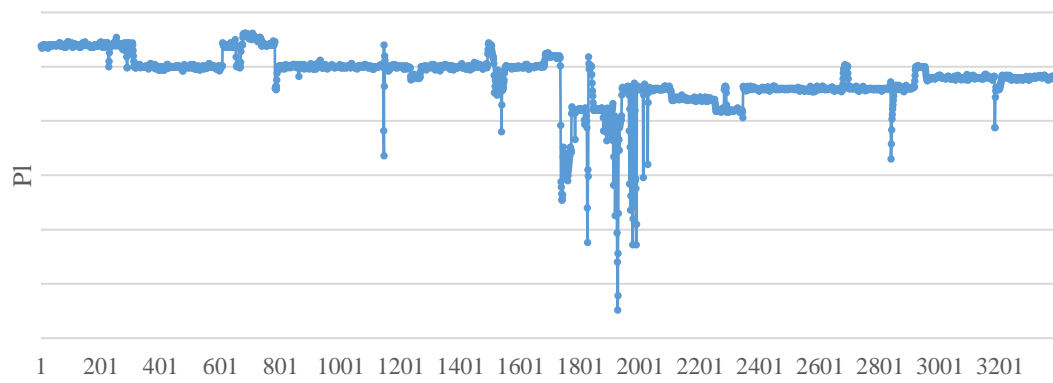
Evolução Tar1 durante período de aquisição de dados



Evolução Tar2 durante período de aquisição de dados



Evolução PI durante período de aquisição de dados



APÊNDICE 2: Pesos das conexões da rede escolhida MLP 11-11-1.

Conexões Rede Predição MLP 11-11-1	Pesos	Conexões Rede Predição MLP 11-11-1	Pesos
Vvul --> neurônio oculto 1	-0,86568	Var1 --> neurônio oculto 11	-7,93682
Vvul --> neurônio oculto 2	1,01250	Var2 --> neurônio oculto 1	1,81061
Vvul --> neurônio oculto 3	-0,60195	Var2 --> neurônio oculto 2	1,92958
Vvul --> neurônio oculto 4	0,26271	Var2 --> neurônio oculto 3	-3,00589
Vvul --> neurônio oculto 5	0,93251	Var2 --> neurônio oculto 4	0,14464
Vvul --> neurônio oculto 6	-0,31147	Var2 --> neurônio oculto 5	0,56674
Vvul --> neurônio oculto 7	0,50739	Var2 --> neurônio oculto 6	-3,82325
Vvul --> neurônio oculto 8	-0,19555	Var2 --> neurônio oculto 7	-2,34668
Vvul --> neurônio oculto 9	-2,63391	Var2 --> neurônio oculto 8	0,79257
Vvul --> neurônio oculto 10	0,89564	Var2 --> neurônio oculto 9	5,31891
Vvul --> neurônio oculto 11	0,92451	Var2 --> neurônio oculto 10	-4,97094
x --> neurônio oculto 1	-0,97197	Var2 --> neurônio oculto 11	0,08601
x --> neurônio oculto 2	0,08442	Var3 --> neurônio oculto 1	1,70662
x --> neurônio oculto 3	1,19649	Var3 --> neurônio oculto 2	1,11275
x --> neurônio oculto 4	-1,36766	Var3 --> neurônio oculto 3	-0,62048
x --> neurônio oculto 5	-1,43873	Var3 --> neurônio oculto 4	-1,43900
x --> neurônio oculto 6	2,01086	Var3 --> neurônio oculto 5	-0,16955
x --> neurônio oculto 7	0,88608	Var3 --> neurônio oculto 6	1,34586
x --> neurônio oculto 8	-2,05753	Var3 --> neurônio oculto 7	1,61309
x --> neurônio oculto 9	1,37506	Var3 --> neurônio oculto 8	2,12182
x --> neurônio oculto 10	-1,04111	Var3 --> neurônio oculto 9	-1,22561
x --> neurônio oculto 11	-2,86958	Var3 --> neurônio oculto 10	-0,91028
Tlu --> neurônio oculto 1	-2,75520	Var3 --> neurônio oculto 11	-1,14843
Tlu --> neurônio oculto 2	-1,94631	Par1 --> neurônio oculto 1	-1,78052
Tlu --> neurônio oculto 3	1,59748	Par1 --> neurônio oculto 2	0,62042
Tlu --> neurônio oculto 4	0,01139	Par1 --> neurônio oculto 3	0,08319
Tlu --> neurônio oculto 5	-1,81236	Par1 --> neurônio oculto 4	2,58023
Tlu --> neurônio oculto 6	5,60897	Par1 --> neurônio oculto 5	-1,23198
Tlu --> neurônio oculto 7	2,33692	Par1 --> neurônio oculto 6	-1,28403
Tlu --> neurônio oculto 8	-0,28319	Par1 --> neurônio oculto 7	1,17690
Tlu --> neurônio oculto 9	-4,03319	Par1 --> neurônio oculto 8	0,48122
Tlu --> neurônio oculto 10	2,14148	Par1 --> neurônio oculto 9	1,06520
Tlu --> neurônio oculto 11	-0,46282	Par1 --> neurônio oculto 10	0,78794
Var1 --> neurônio oculto 1	2,54831	Par1 --> neurônio oculto 11	2,66651
Var1 --> neurônio oculto 2	0,28943	Par2 --> neurônio oculto 1	-3,23577
Var1 --> neurônio oculto 3	-0,82550	Par2 --> neurônio oculto 2	-1,10874
Var1 --> neurônio oculto 4	-2,88415	Par2 --> neurônio oculto 3	2,15523
Var1 --> neurônio oculto 5	-2,24112	Par2 --> neurônio oculto 4	1,79925
Var1 --> neurônio oculto 6	2,75106	Par2 --> neurônio oculto 5	2,60173
Var1 --> neurônio oculto 7	5,85645	Par2 --> neurônio oculto 6	0,00554
Var1 --> neurônio oculto 8	3,45594	Par2 --> neurônio oculto 7	0,10985
Var1 --> neurônio oculto 9	0,64736	Par2 --> neurônio oculto 8	-2,40499
Var1 --> neurônio oculto 10	0,29549	Par2 --> neurônio oculto 9	2,12860

Conexões Rede Predição MLP 11-11-1	Pesos	Conexões Rede Predição MLP 11-11-1	Pesos
Par2 --> neurônio oculto 10	3,65324	input bias --> neurônio oculto 1	-1,22762
Par2 --> neurônio oculto 11	-2,43551	input bias --> neurônio oculto 2	3,54903
TAr1 --> neurônio oculto 1	0,77543	input bias --> neurônio oculto 3	1,48097
TAr1 --> neurônio oculto 2	-1,36079	input bias --> neurônio oculto 4	3,12188
TAr1 --> neurônio oculto 3	-1,02234	input bias --> neurônio oculto 5	-0,05506
TAr1 --> neurônio oculto 4	0,22091	input bias --> neurônio oculto 6	-0,12491
TAr1 --> neurônio oculto 5	0,72283	input bias --> neurônio oculto 7	-3,06149
TAr1 --> neurônio oculto 6	0,39206	input bias --> neurônio oculto 8	0,66586
TAr1 --> neurônio oculto 7	-1,02325	input bias --> neurônio oculto 9	1,18632
TAr1 --> neurônio oculto 8	-0,70138	input bias --> neurônio oculto 10	0,98100
TAr1 --> neurônio oculto 9	-0,09088	input bias --> neurônio oculto 11	4,02222
TAr1 --> neurônio oculto 10	-0,16909	neurônio oculto 1 --> Epart	-3,20921
TAr1 --> neurônio oculto 11	0,48094	neurônio oculto 2 --> Epart	-3,08521
TAr2 --> neurônio oculto 1	-0,38941	neurônio oculto 3 --> Epart	-2,49336
TAr2 --> neurônio oculto 2	-0,38252	neurônio oculto 4 --> Epart	-1,18192
TAr2 --> neurônio oculto 3	-1,09046	neurônio oculto 5 --> Epart	-2,11998
TAr2 --> neurônio oculto 4	0,07478	neurônio oculto 6 --> Epart	3,84516
TAr2 --> neurônio oculto 5	-0,71879	neurônio oculto 7 --> Epart	2,22412
TAr2 --> neurônio oculto 6	-2,13085	neurônio oculto 8 --> Epart	-1,67131
TAr2 --> neurônio oculto 7	-0,61377	neurônio oculto 9 --> Epart	0,58345
TAr2 --> neurônio oculto 8	-0,82072	neurônio oculto 10 --> Epart	1,43415
TAr2 --> neurônio oculto 9	3,05568	neurônio oculto 11 --> Epart	2,97214
TAr2 --> neurônio oculto 10	0,77428	Bias --> Epart	-2,40167
TAr2 --> neurônio oculto 11	-1,24572		
PI --> neurônio oculto 1	-3,55479		
PI --> neurônio oculto 2	-1,29927		
PI --> neurônio oculto 3	0,97540		
PI --> neurônio oculto 4	-0,35638		
PI --> neurônio oculto 5	-0,27824		
PI --> neurônio oculto 6	4,18523		
PI --> neurônio oculto 7	0,28056		
PI --> neurônio oculto 8	-2,37878		
PI --> neurônio oculto 9	1,26809		
PI --> neurônio oculto 10	-1,06480		
PI --> neurônio oculto 11	-3,35770		