UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
PROGRAMA DE PÓS-GRADUAÇÃO EM TECNOLOGIA DE PROCESSOS QUÍMICOS E BIOQUÍMICOS

MESTRADO PROFISSIONAL EM ENGENHARIA DE COMBUSTÍVEIS E PETROQUÍMICA - EQ/UFRJ

SOBRE A TRANSPORTABILIDADE DE MODELOS PREDITIVOS BASEADOS EM REDES NEURONAIS: APLICAÇÕES A UMA INDÚSTRIA QUÍMICA

Dissertação de Mestrado

Renê Landim de Oliveira

Rio de Janeiro,

Dezembro de 2016

SOBRE A TRANSPORTABILIDADE DE MODELOS PREDITIVOS BASEADOS EM REDES NEURONAIS: APLICAÇÕES A UMA INDÚSTRIA QUÍMICA

Renê Landim de Oliveira

Dissertação submetida ao Corpo Docente do Curso de Pós-Graduação em Tecnologia de Processos Químicos e Bioquímicos da Escola de Química da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências.

Aprovado por:

Prof. Maurício B. de Souza Jr., D.Sc.
(orientador)

Profa. Andrea Valdman, D.Sc.
(orientadora)

Prof. Carlos André Vaz Junior, D.Sc.

Giane Ribeiro Stuart, D.Sc.

Profa. Vivianni Marques Leite dos Santos, D.Sc.

Rio de Janeiro, RJ - Brasil
Dezembro de 2016

SOBRE A TRANSPORTABILIDADE DE MODELOS PREDITIVOS BASEADOS EM REDES NEURONAIS: APLICAÇÕES A UMA INDÚSTRIA QUÍMICA

Renê Landim de Oliveira

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Tecnologia de Processos Químicos e Bioquímicos da Escola de Química da Universidade Federal do Rio de Janeiro, para a obtenção do grau de Mestre em Ciências.

Orientadores: Prof. Maurício B. de Souza Jr., D.Sc.
Profa. Andrea Valdman, D. Sc.

Rio de Janeiro, RJ - Brasil
Dezembro de 2016

Landim de Oliveira, Rene
LO48s Sobre a transportabilidade de modelos preditivos baseados em redes neuronais: aplicações a uma indústria química / Rene Landim de Oliveira. -- Rio de Janeiro, 2016.
108 f.

Orientador: Maurício Bezerra de Souza Júnior.
Coorientadora: Andrea Valdman.
Dissertação (mestrado) - Universidade Federal do Rio de Janeiro, Coordenação dos Programas de Pós Graduação em Engenharia, Programa de Pós-Graduação em Engenharia Química, 2016.

1. Machine learning. 2. Sensor virtual. 3. Geração de vapor. 4. Recuperação de amônia. 5. Ferramentas de apoio operacional. I. Bezerra de Souza Júnior, Maurício , orient. II. Valdman, Andrea, coorient. III. Título.

AGRADECIMENTOS

A Deus pelo dom da vida e tudo o que me concedeu de graça e de paz para iniciar e prosseguir na busca deste conhecimento. Foram vários os momentos em que a lembrança da presença de Deus me confortava e assim continuava perseverante.

Aos meus pais, Maria Rosalina e Claudionor (*in memorian*) por me concederem a vida, por me darem o melhor de si e me ensinarem que as dificuldades podem ser enfrentadas e superadas quando se caminha na verdade. Obrigado por entenderem minha distância e estarem sempre comigo.

Aos meus irmãos, Nei e Ronaldo, e suas famílias, pelo apoio e por confiarem em mim, me apoiarem nos meus estudos, desde o inicio na graduação e agora nessa etapa da dissertação. O que fez a diferença por essa e outras etapas terem se concluído é ter tido de vocês o companheirismo e a amizade.

Aos meus professores Maurício Bezerra e Andrea Valdman, pela orientação e dedicação em me transmitir o melhor dos seus conhecimentos para essa dissertação. Suas grandes ideias e colocações tornaram esse estudo prazeroso, o que me engrandeceu o suficiente para dar o melhor de mim a esse e a outros trabalhos.

Á FCCSA, meus gestores Alcir Camargo e Roberto Carvalho, e todos meus amigos da área de Otimização de Processos, pela compreensão da importância dessa dissertação para mim e confiança no desenvolvimento desse trabalho. Graças a essa empresa e pessoas como vocês, tomei conhecimento dessa área da ciência e ao me decidir por aprofundá-lo estiveram comigo.

A todos os meus amigos, aos que tomaram conhecimento desse estudo e de maneira atenciosa me apoiaram e a quem de modo muito especial apareceu nesse momento em minha vida. Agradeço pelas palavras inspiradoras e sugestões engrandecedoras que me motivaram na busca desse aprendizado. Aos amigos e professores, com os quais passei pelas oito disciplinas do mestrado, foi um prazer as aulas, trabalhos, provas e dividir conhecimentos, tornando esse momento agradável.

“Se futuro e passado existem, quero saber onde estão. Se ainda não consigo compreender, todavia sei que, onde quer que estejam, não serão futuro nem passado, mas presente. Se aí fosse futuro, não existiria ainda; e se fosse passado, já não existiria. Por conseguinte, em qualquer parte onde estiverem, seja o que for, não podem existir senão no presente.”

Santo Agostinho

Portanto, para quem vê, não se trata do futuro, mas do presente, do qual é tirada a predição de um futuro concebido na mente.

Santo Agostinho

“Qualquer que seja a natureza dessa misteriosa previsão do futuro, não podemos ver senão o que existe. Mas o que existe não é futuro e sim presente. Por conseguinte, quando dizemos que vemos o futuro, não se veem os próprios acontecimentos ainda inexistentes – isto é, o futuro –, mas sim as causas ou os sinais precursores que já existem.”

Santo Agostinho

Resumo da Dissertação de Mestrado apresentada ao Curso de Pós-Graduação em Tecnologia de Processos Químicos e Bioquímicos da Escola de Química/UFRJ como parte dos requisitos necessários para obtenção do grau de Mestre em Ciências.

SOBRE A TRANSPORTABILIDADE DE MODELOS PREDITIVOS BASEADOS EM REDES NEURONAIS: APLICAÇÕES A UMA INDÚSTRIA QUÍMICA

Renê Landim de Oliveira
Dezembro, 2016

Orientadores: Prof. Maurício B. de Souza Jr., D.Sc.
Profa. Andrea Valdman, D.Sc.

O desenvolvimento de modelos preditivos de processo baseados em princípios básicos é tipicamente uma atividade complexa e cara na indústria química, dada a própria complexidade fenomenológica e operacional dos processos. Contudo, na era do "*big data*", as indústrias modernas são tipicamente um ambiente rico em dados pelo uso de historiadores de processos. Tais dados necessitam ser analisados para que essa informação possa ser utilizada de forma proveitosa. Nesse contexto, o uso de algoritmos de aprendizagem de máquina, que são abordagens baseadas em dados para o desenvolvimento de modelos, se revela extremamente promissor. Dentre esses algoritmos destacam-se, pela sua capacidade, as redes neuronais artificiais (RNAs). Esta dissertação propõe uma metodologia para levantamento de dados e projeto de redes neuronais para desenvolvimento de modelos de regressão de processos industriais, que possa ser facilmente transportada – ou seja, aplicada, com pequenas adaptações – a diversos processos industriais. Para tanto, a metodologia proposta foi aplicada a dois processos distintos e bem típicos de plantas químicas industriais: produção de vapor em caldeira e separação em uma coluna. A variável de saída estudada na caldeira foi o teor de oxigênio em função de seis variáveis operacionais. Na coluna *stripper*, o teor de amônia recuperado no fundo foi estudado em função de 10 variáveis operacionais (reduzidas a partir de 16 iniciais, usando algoritmos de redução de dimensionalidade). Os modelos desenvolvidos apresentaram capacidade preditiva (coeficiente de correlação de 0,99 para a caldeira e de 0,83 para a coluna) superior à de modelos de regressão linear múltipla. Ambos os modelos podem ser usados como sensores virtuais em esquemas de monitoramento em controle e otimização. Os resultados evidenciam que a metodologia apresentada pode ser facilmente transportada para gerar modelos de diferentes processos baseados em RNAs, com boa capacidade preditiva.

Abstract of a Thesis presented to Graduate Program on Technology of Chemical and Biochemical Processes - EQ/UFRJ as partial fulfillment of the requirements for the degree of Master of Sciences.

ON THE TRANSPORTABILITY OF PREDICTIVE MODELS BASED ON NEURAL NETWORKS: CHEMICAL INDUSTRY APLICATIONS

Renê Landim de Oliveira
December, 2016

Supervisors: Prof. Maurício B. de Souza Jr., D.Sc.
Profa. Andrea Valdman, D.Sc.

The development of predictive models based on basic principles is typically a complex and expensive activity in the chemical industry, given the phenomenological and operational processes complexity. However, in the "big data" era, modern industries are normally an environment rich in process history data. This information can only be useful if such data is extensively analyzed. In this context, the use of machine-learning algorithms, focused on the development of data based models, turns out to be extremely promising. Among these algorithms, artificial neural networks (ANNs) stand out due to their flexibility. This dissertation proposes a methodology for data collection and design of neural networks to develop industrial processes regression models, which can be easily applied with small adaptations to various industrial processes. The methodology was applied to two distinct and very typical processes available on chemical plants: steam production through a boiler and process separation through a stripper column. The output variable studied in the boiler was the oxygen content as a function of six operational variables. In the stripper column, the ammonia content on the bottom was studied as a function of 10 operational variables (reduced from 16 initially, using dimensionality reduction algorithms). The developed models have a good prediction performance (correlation coefficient of 0.99 for the boiler and of 0.83 for the column) and are better than multiple linear regression models. Both models can be used as virtual sensors in control and optimization monitoring schemes. The results show that a methodology can be easily transposed to generate models of different processes based on ANNs, with good predictive capacity.

Sumário

1 INTRODUÇÃO 1
2 FUNDAMENTOS DE REDES NEURONAIS 3
2.1 Introdução 3
2.2 Histórico e conceitos em RNA 5
2.3 Principais expressões matemáticas 8
2.3.1 O neurônio e a conexão entre camadas 8
2.3.2 Funções de ativação 9
2.3.2.1 Função Sigmoidal 9
2.3.2.2 Função tangente hiperbólica 9
2.3.2.3 Função de base radial (RBF) 10
2.3.3 O treinamento BP 11
2.3.4 Aperfeiçoamento do BP 14
2.3.4.1 O algoritmo de treinamento CG 14
2.3.4.2 O algoritmo de treinamento BFGS 16
2.3.5 Overfitting na etapa de treinamento 17
2.4 Aplicação de RNA em caldeira e coluna 18
2.4.1 RNA aplicadas a caldeiras 19
2.4.2 RNA aplicadas a colunas 26
3 METODOLOGIA 29
Levantamento de dados 30
Tratamento de dados 31
4 ESTUDO DE CASO 1: O PROCESSO DE GERAÇÃO DE VAPOR 34
4.1 Introdução 34
4.2 Passo 1: Estudo inicial do processo de geração do vapor 34
4.2.1 Princípios básicos de funcionamento 34
4.2.2 Descrição da geração de vapor e suas variáveis 35
4.2.3 Instrumentação e controle da caldeira 37
4.2.4 Ganhos previstos no estudo da caldeira 38
4.3 Passo 2: Levantamento e tratamento dos dados da caldeira 39
4.3.1 Levantamento de dados 39
4.3.2 Tratamento dos dados 41
4.4 Passo 4: Projeto dos modelos para a caldeira 43
4.4.1 Regressão por RLM na caldeira 43
4.4.2 Verificação da RLM para a caldeira 45
4.4.3 Regressão na caldeira: Treinamento das redes MLP 46

4.4.4 Análise de Sensibilidade da rede MLP 48
4.4.5 Verificação da MLP projetada para a caldeira 49
4.4.6 Conclusões sobre a modelagem 51
5 ESTUDO DE CASO 2: PROCESSO DE RECUPERAÇÃO DE AMÔNIA NA COLUNA *STRIPPER* 52
5.1 Introdução 52
5.2 Passo 1: Estudo inicial da coluna 53
5.2.1 Descrição da coluna e suas variáveis 53
5.2.2 Instrumentação e controle na coluna 56
5.2.3 Ganhos previstos no estudo da coluna 57
5.3 Passo 2: Levantamento e tratamento dos dados da coluna 58
5.3.1 Levantamento dos dados 58
5.3.2 Tratamento dos dados 59
5.4 Passo 3: definição das variáveis de entrada para o modelo da coluna 62
5.4.1 Redução de dimensionalidade por PCA 62
5.4.2 Redução de dimensionalidade por matriz de correlação e multicolinearidade 64
5.5 Passo 4: Projeto dos modelos para a coluna 66
5.5.1 Regressão na coluna: Treinamento das redes MLP 66
5.5.2 Verificação da MLP do treinamento da coluna 70
5.5.3 Retreinamento da rede MLP 72
5.5.4 Conclusões sobre a modelagem na coluna 74
6 CONCLUSÕES 75
REFERÊNCIAS BIBLIOGRÁFICAS 77
APÊNDICE A.1 80
Técnicas auxiliares para tratamento de dados 80
APÊNDICE A.2 87
Pesos e biases nos modelos da caldeira e coluna 87
APÊNDICE A.3 89
Regressão na coluna: RLM 89
Validação da RLM da coluna 91

LISTA DE FIGURAS

Figura 1 Rede MLP com três camadas 5
Figura 2 Esquema dos constituintes da célula neuronal 6
Figura 3 Características de um neurônio artificial 8
Figura 4 Ilustração da regra do ponto de parada 18
Figura 5 Rede neuronal dinâmica para a predição de NOx e O_2 20
Figura 6 Configuração típica de uma rede RBF 23
Figura 7 Representação da unidade de debutanização 26
Figura 8 RNA da coluna debutanizadora. 27
Figura 9 Fluxograma do desenvolvimento da rede 29
Figura 10 Perda de eficiência pelos gases de combustão 35
Figura 11 Digrama de processo das unidades 36
Figura 12 Representação esquemática da caldeira 37
Figura 13 Sistema de aquisição de dados para a caldeira 39
Figura 14 Padrões de teor de oxigênio na caldeira usados na estimação e verificação 41
Figura 15 Histograma dos padrões de teor de oxigênio da caldeira usados na estimação e verificação 42
Figura 16 Valores de oxigênio reais e preditos por RLM na estimação 44
Figura 17 Valores residuais de oxigênio por RLM 44
Figura 18 Valores preditos e observados na verificação do modelo de RLM da caldeira 45
Figura 19 Valores preditos e observados na verificação do modelo de RLM da caldeira 45
Figura 20 Análise de resíduos na validação RLM da caldeira 46
Figura 21 Dados reais e preditos de oxigênio da caldeira no treinamento 47
Figura 22 Análise dispersão dos resíduos do teor de oxigênio da caldeira no treinamento 48
Figura 23 Dados reais e preditos do teor de oxigênio da caldeira no treinamento 48
Figura 24 Relação entre teor de oxigênio e vazão de gás 49
Figura 25 Dados preditos do oxigênio em relação aos observados na etapa de verificação 50
Figura 26 Dispersão dos resíduos do teor de oxigênio da caldeira na verificação 50
Figura 27 Dados reais e preditos do teor de oxigênio da caldeira na verificação 51
Figura 28 Esquema geral da coluna 53
Figura 29 Representação esquemática da coluna 55
Figura 30 Resultados de NH3 Fundo com os outliers 59
Figura 31 Resultados de NH3 Fundo sem os outliers 60

Figura 32 Dados de NH_3 no fundo da coluna na estimação e verificação da RNA e RLM 60
Figura 33 Distribuição dos dados NH_3 no fundo da coluna na estimação e verificação da RNA e RLM 61
Figura 34 Análise dos componentes principais na coluna 63
Figura 35 Importância das variáveis da coluna por PCA 63
Figura 36 Dados reais de NH_3 fundo em relação aos preditos no treinamento 68
Figura 37 Resíduos da NH_3 fundo em relação aos dados reais de treinamento 69
Figura 38 Dispersão dos resíduos de NH_3 fundo da coluna no treinamento 69
Figura 39 Dados reais e preditos de NH_3 fundo no treinamento 70
Figura 40 Dados reais de NH_3 fundo em relação aos preditos na verificação 71
Figura 41 Resíduos de NH_3 fundo em relação aos dados reais de verificação 71
Figura 42 Dispersão dos resíduos de NH_3 fundo na verificação 72
Figura 43 Dados reais e preditos de NH_3 fundo na verificação 72
Figura 44 Dados reais e preditos de NH_3 fundo no retreinamento 73
Figura 45 Resíduos de NH3 fundo em relação aos dados reais no retreinamento 73
Figura 46 Representação da transformação de três variáveis em dois CP 82
Figura 47 Dados reais de NH_3 fundo em relação aos preditos na RLM com 10 variáveis 90
Figura 48 Dados reais e preditos de NH_3 fundo na RLM com 10 variáveis 91
Figura 49 Resíduos de NH_3 fundo pela RLM em relação aos dados reais com 10 variáveis... 91
Figura 50 Valores preditos e observados na validação da RLM na coluna 92
Figura 51 Valores preditos e observados com os dados de validação da RLM na coluna 92

LISTA DE TABELAS

Tabela 1 Principais funções de ativação utilizadas em RNA do tipo MLP 10
Tabela 2- Comparação das principais variaveis utilizadas em RNA nos trabalhos aplicados a caldeiras. 25
Tabela 3 Resultados da RNA BP de 3 algoritmos de treinamento na predição do butano em uma coluna. 28
Tabela 4 Variáveis de operação, utilizadas como entradas e saída da caldeira 36
Tabela 5 t-test com os dados do sistema supervisório Scada em 6 tempos de coleta 40
Tabela 6 Estatística dos dados de estimação e verificação da caldeira 42
Tabela 7 Resultados dos coeficientes da RLM na caldeira 43
Tabela 8 Coeficientes de correlação na estimação dos parâmetros da caldeira 46
Tabela 9 Análise de sensibilidade da MPL 6-10-1 49
Tabela 10 Variáveis de processo da coluna de recuperação de amônia 54
Tabela 11 Estatística dos dados de estimação e de verificação da coluna 61
Tabela 12 Matriz de correlação das variáveis da coluna 65
Tabela 13 Coeficientes de correlação no treinamento da coluna 67
Tabela 14 Análise de sensibilidade nas variáveis da coluna 67
Tabela 15 Tabela de frequência dos dados da variável de entrada da coluna 68
Tabela 16 Pesos e biases na estimação dos parâmetros no modelo da caldeira 87
Tabela 17 Pesos e biases na estimação dos parâmetros no modelo da coluna 88
Tabela 18 Resultados dos coeficientes da RLM na coluna com as 10 variáveis de entrada 89
Tabela 19 Resultados da RLM nas 7 variáveis de entrada 90

LISTA DE ABREVIAÇÕES

BFGS – algoritmo de treinamento proposto por Broyden–Fletcher–Goldfarb–Shann

BP – Backpropagation

BPNN – Backpropagation Neural Network

CG – Conjugate Gradiente. Algoritmo de treinamento

DNN – Deep Neural Network

DTM – Distance to model.

FB – Feedback

FCCSA – Fábrica Carioca de Catalisadores

FF – Feedforward

GDR – Generalized Delta Rule. Algoritmo de treinamento.

LM – Levenberg-Marquardt. Algoritmo de treinamento.

MLP – *multi-layer perceptrons* (preceptrons de múltipla camada)

PC – *principal component* (componente principal)

PCA – *principal component analysis* (análise de componentes principais)

PIC – controlador e indicador de pressão

RBF – *radial basis function* (função de base radial)

RLM – Regressão linear múltipla

RNA – Redes neuronais artificiais

SOS – sum of squares (função erro por soma dos quadrados)

SP – Set Point

SVR – support vector regression

LISTA DE VARIÁVEIS

α – termo de momento na equação do treinamento BP.

η – termo da taxa de treinamento na equação do treinamento BP.

x – para denotar uma variável de entrada

y – para denotar uma variável de saída

$\hat{x}$ – para denotar uma variável predita

$\bar{x}$ – para representar a média de uma população de dados

$\bar{\bar{x}}$ – para representar a média de uma população de dados preditos

N – número de dados em uma população amostral

s – desvio padrão dos dados amostrais

s^2 – variância dos dados amostrais

w – para denotar o peso associado a um neurônio

a – para denotar a atividade de um neurônio

β – coeficiente de regressão

f – para denotar uma função de transferência

E – erro entre a variável desejada e a predita

R – coeficiente de correlação

t – para denotar os *scores*

S – vetor direção de busca (do inglês search)

1 INTRODUÇÃO

Neste início do século XXI, está-se assistindo ao que pode ser considerado o anúncio da quarta revolução industrial, com o surgimento das chamadas fábricas inteligentes. Neste tipo de fábrica, as máquinas, dispositivos de campo e até mesmo os produtos têm capacidade computacional e de comunicação em rede, conferindo-lhe grande flexibilidade para se adaptar a diferentes demandas (James, 2012). No Brasil, indústrias de manufaturas (de aeronaves, automotiva e outras) já combinam sistemas integrados, automatizados e interconectados. Também plantas de produção – de petroquímicas (Li, 2016) a indústrias de cerveja (O Globo, 2016) – vem se beneficiando de novas tecnologias digitais disruptivas, como internet das coisas (IoT, Internet of Things), computação nas nuvens e tecnologias de *big data*.

Muitos desafios aparecem nesse novo ambiente industrial. Um deles é como usar a farta gama de dados agora disponíveis para, de maneira eficiente, monitorar, controlar e otimizar os processos (Venkatasubramanian, 2009; Qin, 2014; Shu *et al*., 2016). Ferramentas de aprendizagem de máquina podem exercer um importante papel neste contexto. Com estas ferramentas, um computador aprende a executar uma tarefa estudando um conjunto de dados de treinamento. Na sequência, esse computador executa essa mesma tarefa, mas para dados aos quais não havia sido necessariamente apresentado antes. Dentre essas ferramentas, destacam-se as redes neuronais artificiais (RNA), uma vez que elas podem ser usadas para classificação, regressão, clusterização e redução de dimensionalidade (Louridas, & Ebert, 2016).

Em um trabalho elaborado há um quarto de século atrás, Aström & McAvoy (1992) argumentam que os modelos fenomenológicos são caros e não podem ser facilmente transportados, ou seja, tendo sido desenvolvidos para uma planta, serem usados em outra. Por sua vez, as redes neuronais, projetadas a partir de dados, poderiam ser facilmente transportadas.

O momento atual no qual – através de sistemas de gerenciamento de informações da planta (PIMS ou "*Plant Information Management System*") – a indústria dispõe de dados históricos para o desenvolvimento de modelos baseados nessas informações, revela-se oportuno para estudar as características de transportabilidade de redes neuronais artificiais.

Nesta dissertação, o conceito de transportabilidade é estendido pela obtenção de modelos preditivos de diferentes unidades dentro de uma mesma planta industrial usando uma

mesma metodologia. Para tanto, é proposta uma metodologia para levantamento, análise e tratamento de dados e seleção, treinamento e validação de redes neuronais.

A planta escolhida para a realização do estudo foi a da Fábrica Carioca de Catalisadores (FCCSA), localizada no Distrito Industrial de Santa Cruz no Rio de Janeiro – RJ. Esta fábrica produz catalisadores de FCC (*Fluid Catalytic Cracking*) e outros aditivos para a indústria de refino, sendo líder na América do Sul na produção de tais catalisadores. Dentro da planta da FCCSA, foram escolhidos dois estudos de caso: 1. a geração de vapor (de uma caldeira água tubular) e 2. a separação de amônia residual (numa coluna *stripper* de pratos e recheios). Esses processos foram escolhidos por sua representatividade. A caldeira é um equipamento amplamente usado no setor de utilidades das indústrias. Já a coluna também é ilustrativa de equipamentos multiestágios largamente empregados para separação de componentes em misturas miscíveis.

Em ambos os casos, os modelo de regressão têm o objetivo de gerar sensores virtuais ou *soft sensors*, nos quais variáveis *online* de processo (chamadas neste contexto de "variáveis secundárias"), fáceis de medir, são usadas para predizer *online* variáveis alvo, de difícil medição (Kadlec *et al*., 2009). Assim, nos dois processos é feita a predição, respectivamente, das variáveis teor de oxigênio e teor de amônia em função de respectivas variáveis secundárias.

A dissertação foi desenvolvida na FCCSA e na UFRJ. Inicialmente foram estudados *in loco* os processos: os equipamentos envolvidos, seu funcionamento, variáveis de entrada e saída, instrumentação e controle, importância na planta e ganhos previstos pela aplicação da metodologia. Na sequência, foram levantados dados para a modelagem. Esses dados foram analisados – usando estatística descritiva, teste t de hipótese etc. – e tratados – por técnicas de redução de dimensionalidade como análises de componentes principais (PCA), matriz de correlação e multicolinearidade, etc. Em seguida, foram selecionadas e projetadas as RNAs, as quais foram comparadas estatisticamente com métodos de regressão linear múltipla (RLM).

A organização desse texto está como segue. É apresentada a fundamentação teórica sobre RNA, assim como uma revisão de aplicações a caldeiras e a colunas. Na sequência, introduz-se o método proposto nesta dissertação. Em seguida, mostram-se as aplicações da metodologia aos estudos de caso 1 e 2, respectivamente. Por fim, são apresentadas as principais conclusões da pesquisa e sugestões para trabalhos futuros. O Apêndice 1 apresenta alguns métodos estatísticos empregados na análise dos dados e dos modelos obtidos.

2 FUNDAMENTOS DE REDES NEURONAIS

2.1 Introdução

As RNAs são uma ferramenta cada vez mais utilizada especialmente na Engenharia Química em que os fenômenos de não linearidade são comuns. Na indústria química, muitos processos não lineares não são facilmente mapeados devido à sua complexidade.

A inteligência artificial é a área da ciência que estuda a capacidade de se realizar funções associadas com a inteligência humana. Várias são as técnicas utilizadas que tentam implementar no computador essa capacidade do cérebro humano. Uma dessas técnicas é a RNA, que possui como característica a habilidade de generalizar o que foi aprendido.

Uma RNA é um processador paralelamente distribuído composto de unidades simples de processamento para armazenar conhecimento por experiência e torná-lo disponível para uso, assemelhando-se ao cérebro humano em dois aspectos: o conhecimento é adquirido a partir de seu ambiente através de um processo de aprendizagem e as forças de ligação, conhecidas, como pesos sinápticos, são usadas para armazenar o conhecimento (Haykin, 2009). Com essas características, dentre outras possibilidades, uma RNA torna-se uma técnica capaz de correlacionar uma variável de saída em função dos dados de entrada e condições de um dado processo, fornecendo um modelo baseado em dados.

As RNAs podem ser usadas para fins de inferência (ou predição), diagnóstico de falhas e controle avançado. As técnicas de inferência são capazes de inferir o valor de uma variável continuamente e em tempo real sem modificar as condições operacionais, permitindo acompanhar (gerenciar) o desempenho de um processo e, a qualquer alteração nas condições operacionais, informar a anormalidade ao acompanhante do processo, tornando-se uma solução bastante adequada para redução de custos operacionais e aumento da produtividade.

Através desta capacidade de inferir, algumas vantagens são apresentadas em relação à medição convencional: maior durabilidade, visto que os medidores, quando em contato com o meio agressivo, podem se danificar e, com a eliminação ou redução de tubulações e cabeamentos, otimizam-se espaços e diminuem-se custos. Comumente os sistemas de medição, além de terem altos custos, exigem frequentes manutenções e calibrações e um bom estudo do ponto de colocação do instrumento para que a medição possa ser bem realizada. Das desvantagens apresentadas pela RNA na aplicação para predição de variáveis pode-se citar a necessidade de refazer a correlação quando os dados atuais sofrerem um desvio em relação àqueles usados na construção do modelo.

Técnicas baseadas em RNA, portanto, são bastante empregadas na Engenharia Química, visto que neste segmento é bastante comum a geração de informações do processo, através de medições (sejam por instrumentação de campo ou em laboratório) e, muitas vezes, não se dispõe de modelo para o processo.

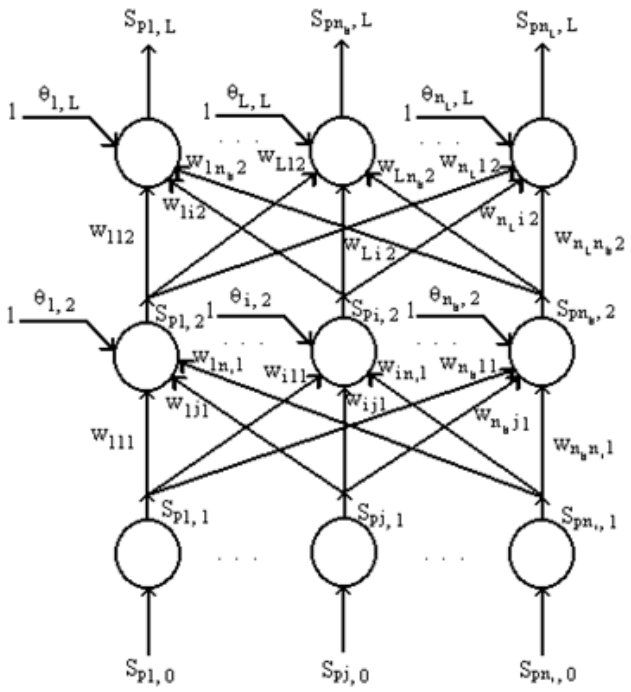
As RNA visam também suprir a limitação de monitoramento e controle de uma planta devido à falta de equipamentos e instrumentos capazes de medir uma determinada variável. Nesse sentido, pode ser aplicada quando uma determinada variável de interesse não é medida instantaneamente ou quando o instrumento de medida necessita ser retirado de operação para manutenção ou sofreu algum dano.

Há diversos paradigmas de RNA, com diferentes topologias e métodos de aprendizagem. Uma RNA típica é composta por uma densa distribuição de unidades, chamadas de neurônios, conectados entre si e que passam sinais de um para outro. Cada interconexão tem um peso associado que expressa a importância de cada entrada do neurônio.

A arquitetura de uma rede pode ser de três tipos: redes *feedforward* (FF) de camada única, redes FF multicamadas e redes recorrentes ou redes *feedback* (FB), (Haykin, 2009). A diferença dessas arquiteturas é a distribuição dos sinais: na rede FF os sinais apenas se movem para frente, enquanto que na FB há realimentações, o que lhe confere uma natureza dinâmica. A estrutura geralmente empregada é a FF em função da dinâmica da rede com realimentação poder assumir comportamento espúrio, diferente do processo modelado, exigindo com isso maiores cuidados na etapa de treinamento.

Um modelo típico de RNA FF é o modelo *multilayer perceptron* (MLP) ou *perceptron* multicamadas, uma rede com pelo menos uma camada de neurônios escondidos e uma camada de saída de neurônios. Segundo Haykin (2009), a função da camada escondida é intervir entre os dados externos de entrada e a saída da rede e se há mais de uma camada escondida, a saída de uma é a entrada da outra imediatamente posterior.

A representação de uma rede MLP pode ser vista na Figura 1, com três camadas A, B e C (camada de entrada, camada escondida e camada de saída); três variáveis de entrada I_1, I_j e I_L; três neurônios na camada escondida e três variáveis de saída c_1, c_2 e c_3 (Baughman e Liu, 1995).



Fonte: De Souza Jr. (1993)

Figura 1 Rede MLP com três camadas

No treinamento a rede aprende as características fundamentais do problema. Há vários algoritmos de aprendizado, sendo o *backpropagation* (BP) o mais conhecido. Nesse algoritmo, um conjunto de dados de entrada é apresentado à camada de entrada da rede. A rede propaga esses dados para as demais camadas até que os dados de saída sejam gerados. Se os dados de saída são diferentes dos desejados, um erro é calculado retornando da camada de saída para a camada de entrada, sendo os pesos modificados conforme o erro é propagado.

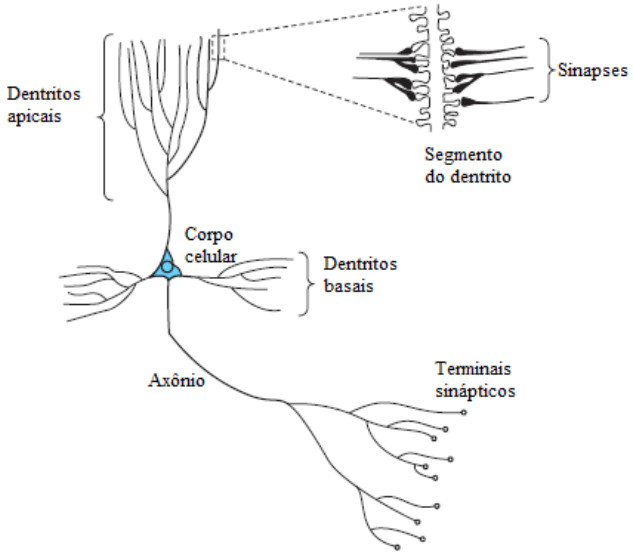
A arquitetura MLP das redes BP geralmente é obtida por tentativas variadas do número de camadas e número de neurônios escondidos em cada camada, diferentes funções de transferência e algoritmos de aprendizagem até a minimização do erro. Uma vez obtida a rede, seus pesos e biases estão definidos.

2.2 Histórico e conceitos em RNA

O objetivo deste tópico é uma apresentação resumida da evolução do estudo de RNA e dos descobrimentos mais marcantes desta técnica e dos conceitos e definições a que chegaram os principais autores e usuários de RNA em seus estudos, desenvolvimentos e aplicações.

O estudo das RNA tem inspiração no cérebro humano e para ilustrar como funciona o sistema nervoso é apresentada a Figura 2, onde está representado um conjunto complexo de células, os neurônios. O papel dos neurônios é fundamental no funcionamento e comportamento do corpo humano e do raciocínio.

Os neurônios possuem uma ampla variedade de formas e tamanhos em diferentes partes do cérebro. O neurônio recebe a maior parte das suas entradas através de espinhas dendríticas, onde se lê na Figura 2 “segmento do dentrito”. A célula piramidal pode receber 10.000 ou mais contatos sinápticos e projetá-los em milhares de células alvo (Haykin, 2009).



Fonte: Haykin (2009) – Adaptado

Figura 2 Esquema dos constituintes da célula neuronal

O primeiro modelo de RNA foi do ano de 1943, apresentado por McCulloch e Pitts, no qual a saída do neurônio levava o valor de 1,0 se o campo local do neurônio era positivo e 0 (zero) se negativo. Em 1949, Hebb postulou a primeira regra para treinamento auto-organizável e, em 1958, Rosenblat propôs o *perceptron* como o primeiro modelo de treinamento supervisionado, apresentando também as primeiras contribuições em RNA. As limitações computacionais do *perceptron* de Rosenblatt foram posteriormente colocadas em uma sólida base matemática por Minsky e Papert (1969), mostrando que o *perceptron* como definido por Rosenblatt era inerentemente incapaz de fazer algumas generalizações globais com base em exemplos localmente aprendidos. Rumelhart e McClelland apresentaram em 1986 o uso do método BP, iniciando o conceito para treinamento de RNA (Haykin, 2009).

Evoluindo em eficiência, as RNAs provaram sua capacidade na solução de desafios complexos em aplicação a problemas de engenharia química, onde a abordagem tradicional na solução de problemas nem sempre dava bons resultados devido a sua alta complexidade e não linearidade (Himmelblau, 2000).

Hoskins e Himmelblau (1988) realizaram a primeira aplicação de RNA para a engenharia química em um problema de detecção e diagnóstico de falhas. Desde então as redes se tornaram uma ferramenta amplamente usada na área.

Um novo modelo de redes conhecido como *Deep Neural Network* (DNN) despertou interesse a partir dos anos 1990. Essas redes são construídas a partir de uma cascata de camadas ocultas de unidades entre as camadas de entrada e de saída. As características representadas por cada camada de processamento definem uma hierarquia de níveis de abstração, da mais baixa para a mais alta, que infere representações úteis através de um processo de aprendizagem a partir de dados não marcados em larga escala. Diferentes modelos de rede são adequados para formar um DNN, tanto redes FF como FB, esta para altas velocidades de processamento. Os pesos e biases podem ser determinados tanto por treinamentos supervisionados como não supervisionados. O alto custo computacional para treinamento e teste é um dos itens mais relevantes na DNN, e um importante desenvolvimento nessa área é a elaboração de métodos de pré-treinamento sem supervisão, usando um algoritmo baseado em máquinas Boltzmann para aprendizado da rede. Um campo em exploração em DNN é a neurociência que busca a interpretação de padrões de processamento e a comunicação da informação dentro do sistema nervoso, tentando estabelecer uma relação entre a estimulação e a resposta com ativações elétricas do neurônio no cérebro (Prieto *et al*, 2016).

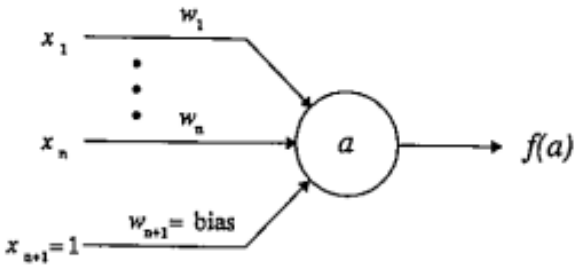
Nos documentos do *Science Direct*, pesquisados por "*neural networks*" em resumos, títulos e palavras chave de livros e jornais nos últimos 30 anos, são encontrados 47.520 trabalhos em todas as áreas da ciência, sendo a maioria na área da Neuro Computação. Para o mesmo período, encontra-se para a área da Engenharia Química 2.641 documentos, representando 5,6% dos trabalhos publicados, sendo as áreas de Computação e Engenharia Química, aplicação na Engenharia térmica, Engenharia de alimentos e Controle de Processos as que mais demandaram trabalho, (*Science Direct*, 2016).

2.3 Principais expressões matemáticas

A presença do recurso matemático nas RNA é indispensável no desenvolvimento da modelagem de uma RNA. Através de funções matemáticas as sinapses que ocorrem no cérebro humano poderão ser imitadas e assim fazer o papel do processamento de um sinal de entrada recebido para gerar um sinal de saída objetivo.

2.3.1 O neurônio e a conexão entre camadas

Duas características básicas definem o neurônio apresentado na Figura 3: a atividade *a*, resultante das entradas, ou *inputs*; e a saída, ou *output*. A atividade a é calculada pelo somatório das multiplicações das entradas x_i e pesos w_i, Equação 2-1, e em seguida a atividade é aplicada à função de transferência $f\ (a)$ (Hoskins e Himmelblau, 1988).



Fonte: Baughman e Liu (1995) – Adaptado

Figura 3 Características de um neurônio artificial.

$$a_i = \sum x_i . w_i \qquad (2\text{-}1)$$

A Equação 2-2, uma função de transferência sigmoidal, representa a saída do neurônio *i* na camada escondida, calculada com a atividade *a*.

$$f_i(a_i) = \frac{1}{1+e^{-a_i}} \qquad (2\text{-}2)$$

A Equação 2-3 é a que descreve, portanto, a relação entre as entradas e a camada escondida com os pesos w_i. E a Equação 2-4 a que representa a conexão da camada escondida com a de saída onde os pesos são dados por w_o.

$$a_i = f(w_i . x_i) \quad (2\text{-}3)$$

$$y_i = f(a_i . w_o) \quad (2\text{-}4)$$

2.3.2 Funções de ativação

A função de ativação, ou também chamada de função de transferência, é a responsável, dentro de um neurônio, por receber um sinal de entrada vindo da camada imediatamente anterior e emitir um sinal de saída que será a entrada do neurônio da camada posterior. A função de transferência se encontra nas camadas escondidas da rede e não nas camadas de entrada ou saída.

Sobre as funções de transferência, as mais comumente encontradas para aplicação em RNA são a sigmoidal, a tangente hiperbólica e a gaussiana de base radial.

2.3.2.1 Função Sigmoidal

A função sigmoidal, representada pela Equação 2-5, é aquela que tem boa *performance* para predição. Matematicamente é mais suavizada quando comparada com a de base radial. Trata-se de uma função com valores limites entre 0 e 1 para valores de entrada de $-\infty$ a $+\infty$, respectivamente (Baughman e Liu, 1995). Conforme Hoskins e Himmelblau (1988), a função sigmoidal é diferenciável e não decrescente, características que possibilitam que essa função possa ser usada em conjunto no treinamento BP.

2.3.2.2 Função tangente hiperbólica

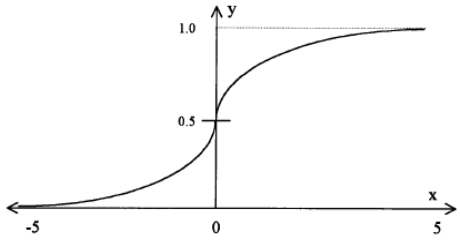
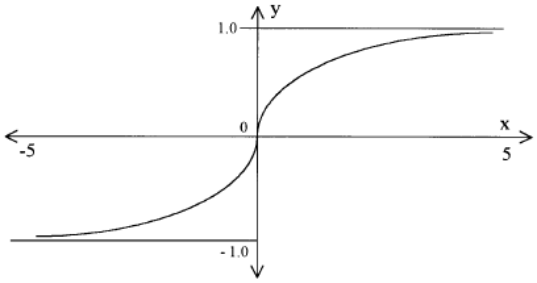
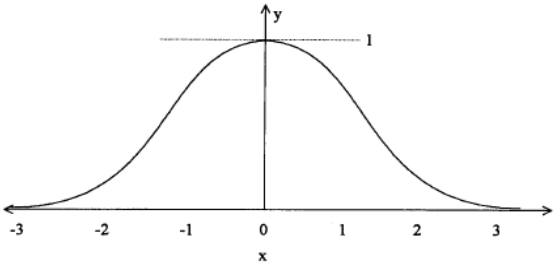
A função tangente hiperbólica, representada pela Equação 2-6, opera com valores limites entre – 1 e +1 quando a entrada é, respectivamente, de $-\infty$ a $+\infty$. Assim como a função sigmoidal, a tangente hiperbólica tem boa *performance* nos trabalhos voltados para predição. Comparando a função sigmoidal com a tangente hiperbólica, o que se observa é maior sensibilidade aos dados de entrada da função tangente hiperbólica, ou seja, para pequenas variações na entrada é maior a variação de saída para a tangente hiperbólica.

2.3.2.3 Função de base radial (RBF)

Representada pela Equação 2-7, trata-se de uma função com natureza local, pois tem um máximo de resposta em 1 (um) quando a entrada é 0 (zero), decaindo para 0 (zero) quando a entrada se distancia do zero. Devido a essa característica, as redes com neurônios na camada interna com função de ativação radial (chamadas de redes RBF) são mais indicadas para aplicações em estudos de diagnóstico de falha e categorização de características que para regressão (Baughman e Liu, 1995).

A Tabela 1 apresenta as equações de 2-5, 2-6 e 2-7 das principais funções de ativação utilizadas em redes multicamadas.

Tabela 1 Principais funções de ativação utilizadas em RNA do tipo MLP

Função de transferência	Representação gráfica	Equação
Função sigmoidal		$f(x) = \frac{1}{1+e^{-x}}$ (2-5)
Função tangente hiperbólica		$f(x) = tan\square\, x = \frac{e^{x} - e^{-x}}{e^{x}+e^{-x}}$ (2-6)
Função de base radial		$f(x) = exp\ \frac{-x^2}{2}$ (2-7)

2.3.3 O treinamento BP

As Equações de 2-1 a 2-4 mostram como funciona uma rede, ou seja, como, conhecidos os seus parâmetros e dados de entrada, são calculadas as saídas. Essas saídas podem apresentar erros em relação aos valores desejados ou alvos (experimentais). Para que os parâmetros da rede sejam obtidos, é necessária uma abordagem que minimize o erro entre as predições da rede e os valores desejados. Na chamada aprendizagem supervisionada, o procedimento iterativo de obtenção dos pesos e biases que resultam na minimização desse erro é chamado de treinamento da rede. Segundo De Souza Jr. (1993), os pesos iniciais devem ser arbitrados entre –1 e +1, de modo a evitar problemas numéricos.

Existem vários tipos de algoritmos de treinamento específicos para determinados modelos de RNA, estes algoritmos diferem entre si, principalmente, pelo modo como os pesos são modificados.

O algoritmo *backpropagation* (BP) ou retropropagação, introduzido na década de 1980, tornou-se, desde então, um dos mais poderosos tipos de treinamento. Ele se baseia na minimização de uma função objetivo dos erros de predição da rede. Tipicamente, é usada a função de soma do erro quadrado, definida pela Equação 2-8 (De Souza Jr.,1993).

$$E = \sum_{p=1}^{N_p} E_p = \frac{1}{2} \sum_{i=1}^{n_l} \left(s_{pi,L} - t_{piL}\right)^2 \qquad (2\text{-}8)$$

Em que:

E = Erro quadrático total da camada de saída;

$s_{pi,L}$ = valores desejados para o i^{th} neurônio na camada de saída;

t_{pi1} = valores preditos para o i^{th} neurônio na camada de saída.

A minimização de E é alcançada, variando-se os pesos e os biases. O algoritmo de otimização do gradiente descendente é empregado. Para tanto, o vetor gradiente da função E em relação aos pesos e biases deve ser calculado. Dessa forma, no treinamento BP, duas fases são necessárias para encontrar a minimização da função erro: a primeira fase *forward* que significa encontrar o primeiro valor de erro com os primeiros valores de pesos e a segunda fase *retropropagação* que significa propagar para trás o erro encontrado até o seu ajuste. A primeira fase – descrita pelas Equações 2-9 a 2-13, – mais a segunda fase – correspondente à Equação 2-14 – se repetem até que o erro seja pequeno (De Souza Jr., 1993).

a) Produção da função de ativação na passagem pela primeira camada.

$$S_{pj,1} = f\ (S_{pj,0}) \tag{2-9}$$

Onde:

p = cada padrão de dados p (p = 1, 2, ... N_p);

$S_{pj,0}$ = sinais de entrada, (j = 1, 2, ... n_i), em que n_i corresponde ao número de neurônios na camada de entrada;

$S_{pj,1}$ = sinais após passagem pela primeira camada de neurônios, com (j = 1, 2, ... n_i).

b) Produção da função de ativação e sua aplicação na função de transferência da camada escondida.

$$\gamma_{pj,2} = \left[\sum_{i=1}^{n_i} w_{ji1}\ s_{pi1}\right] + \theta_{j2} \tag{2-10}$$

$$S_{pj,2} = f\ (\gamma_{pj,2}) \tag{2-11}$$

Onde:

$\gamma_{pj,2}$ = função de ativação;

$S_{pj,2}$ = sinais obtidos após a passagem pela camada escondida;

f = função de transferência;

j = 1, 2, ... n_h;

n_h corresponde ao número de neurônios na camada escondida;

θ_{j2} = limite interno de ativação, ou *bias*.

c) Produção da função de ativação e sua aplicação na função de transferência na saída da rede, camada L.

$$\gamma_{pj,L} = \left[\sum_{i=1}^{n_h} w_{ji2}\ s_{pi2}\right] + \theta_{jL} \tag{2-12}$$

$$S_{pj,L} = f\ (\gamma_{pj,L}) \tag{2-13}$$

Onde:

$\gamma_{pj,L}$ = função de ativação;

$S_{pj,L}$ = sinais obtidos após passagem pela camada de saída, ou seja, dados preditos;

j = 1, 2, ... n_L;

θ_{jL} = limite interno de ativação, ou *bias*.

Na segunda fase é usado o método do gradiente descendente para a correção dos pesos, primeiro entre a camada escondida e a camada de saída e em seguida entre a camada escondida e a camada de entrada (De Souza Jr., 2015). O método BP convencional possui tipicamente dois parâmetros: a taxa de treinamento e termo de momento, representados respectivamente por η e α (De Souza Jr., 2015):

$$\Delta w_{k+1} = -\eta \nabla E(w_k) + \alpha \Delta w_k \tag{2-14}$$

Onde:

w_{k+1} = novo peso calculado;

w_k = peso anterior, ou peso velho;

$\nabla E(w_k)$ = vetor gradiente do erro em função dos pesos, calculado pelo algoritmo envolvendo as equações acima.

As expressões para o cálculo do vetor gradiente se encontram na literatura (De Souza Jr., 1993). A taxa de treinamento é uma constante positiva que controla a taxa na qual o novo peso é ajustado, baseando-se no termo do gradiente descendente calculado (Baughman e Liu, 1995), seu valor deve ser pequeno para evitar oscilações em torno do mínimo. O termo de momento suaviza o efeito de mudanças drásticas do peso pela adição de uma fração de mudança de peso proveniente da apresentação anterior (Hoskins e Himmelblau, 1988).

O erro diminui com um aumento do número de iterações, começando com um valor alto e diminuindo rapidamente, e, em seguida diminui mais lentamente à medida que a rede procura um mínimo local na superfície do erro (Haykin, 2009). Esse algoritmo é convergente quando a taxa de variação do erro por iteração é suficientemente pequena sendo essa taxa tipicamente considerada pequena no intervalo de 0,1 a 1 por cento por iteração (Haykin, 2009).

2.3.4 Aperfeiçoamento do BP

2.3.4.1 O algoritmo de treinamento CG

Apresentado por Leonard e Kramer (1990), o algoritmo de treinamento CG é um avanço no treinamento BP e emprega direções de busca conjugadas, pertencendo assim à classe de métodos de otimização de segunda ordem (Haykin, 2009).

Esse método pode ser visto como um BP batelada com ajustes dinâmicos de α e η. No BP, α e η são geralmente constantes ou alterados em intervalos frequentes, geralmente por uma intervenção. η fixo é indesejado, pois um passo de comprimento arbitrário em uma direção local não necessariamente resulta na redução da função objetivo. Por essa razão, aumentos na função objetivo geralmente ocorrem no BP. O uso de um α fixo pode causar ciclos em relação a um mínimo. Assim, a proposta dos autores foi a aplicação da direção de busca, Equação 2-15, no passo de otimização na iteração q, como dado por (Leonard e Kramer, 1990):

$$w_{q+1} = w_q + M'_p \tag{2-15}$$

Onde:

w_{q+1} = peso ajustado, novo;

w_q = peso anterior ao ajuste, calculado por algoritmo específico;

η = taxa de treinamento

O tamanho do passo é determinado em cada iteração realizando a linha de busca que minimiza a função objetivo na direção de busca.

A proposta do método CG, ao invés da Equação apresentada pelo GDR, é apresentada a Equação 2-16, complementada pelas Equações 2-17 a 2-21.

$$M'_p = M_p + \alpha M'_{p-1} \tag{2-16}$$

$$M_p = \sum_{p=1}^{p} m_p \tag{2-17}$$

$$\alpha = \frac{\|M_p\|^2}{\|M_{p-1}\|^2} \tag{2-18}$$

Onde:

α = termo de momento;

m_p = gradiente que representa a direção descendente mais íngreme em relação à função objetivo para o padrão de treinamento p;

$$m_{ijk} = \delta_{j,k+1} * z_{i,k} \tag{2-19}$$

Onde:

m_{ijk} = componente do gradiente descendente associado com o peso w_{ijk};

$$\delta_{j,L} = \left(d_j - y_j\right) * y_j * (1 - y_j) \tag{2-20}$$

Onde:

$z_{i,k}$ = função de transferência sigmoidal, Equação 3;

w_{ijk} = ligação do neurônio i na camada k para o neurônio j na camada $k + 1$

$1 \leq j \leq N_L$

E para as outras camadas:

$$\delta_{i,k} = z_{ik}(1 - z_{ik}) * \sum_{j=1}^{N_{k+1}} w_{ijk} * \delta_{j,k+1} \tag{2-21}$$

Onde:

$1 < k < L - 1$ e $1 \leq i \leq N_k$

$z_{i,k}$ = função de transferência sigmoidal.

O fator de momento é calculado passo a passo da iteração ao invés de ser mantido constante. Usando o α dado pela Equação 18 resulta em uma busca em direções conjugadas o que torna o método mais eficiente do que aquele baseado na direção descendente quando as superfícies da solução são pobremente dimensionadas. Da mesma forma, o tamanho de passo

a cada iteração, η, é ajustado ao invés de ser fixo e isso garante uma redução na função objetivo e previne ciclos e *overfitting* (Leonard e Kramer, 1990).

Ao aplicar o algoritmo CG em comparação ao BP tradicional em um problema "OR" Leonard e Kramer (1990) obtiveram que o CG foi duas vezes mais rápido, convergiu em todos os casos alcançando o mínimo global em 52% dos casos, enquanto o BP não convergiu em 24% dos casos.

Para o problema de diagnóstico de falhas em três reatores em série Leonard e Kramer (1990) utilizaram o método CG comparado com o BP de Hoskins e Himmelblau (1988) e obtiveram melhores resultados em termos do número de iterações e falhas na convergência com poucas camadas escondidas.

2.3.4.2 O algoritmo de treinamento BFGS

O método dos autores Broyden–Fletcher–Goldfarb–Shanno (BFGS) é considerado atualmente uma das melhores formas do método quasi-Newton, um método de segunda ordem. O método quasi-Newton usa as informações de uma curva de segunda ordem sobre a superfície do erro sem a necessidade do conhecimento da Hessiana. Isso por usar duas iterações sucessivas, juntamente com os respectivos vetores gradiente. Para o quasi-Newton, quando a linha de busca é precisa e estamos na proximidade de um mínimo local com uma Hessiana positiva, o método tende a se aproximar do método de Newton e com isso atingir a convergência mais rápida do que o CG. Comparado ao CG, o método quasi-Newton não é tão sensível na precisão da linha de busca na etapa de otimização e requer armazenamento da matriz dos vetores gradientes sobrecarregando os cálculos e por isso ser preferível para redes de menor escala, Haykin (2009).

Na atualização dos pesos, w, o método quasi-Newton, usa basicamente a Equação 2-22, seguida da Equação 2-23, Haykin (2009).

$$w\ (n+1) = w(n) + \ (n)s(n) \tag{2-22}$$

$$s\ (n) = -S(n)g(n) \tag{2-23}$$

Em que:

$s\ (n)$ é o vetor direção

$S\ (n)$ é uma matriz positiva ajustada a cada iteração

$g\ (n)$ é o vetor gradiente

$w\ (n+1)$ é o peso novo, na iteração *n+1*

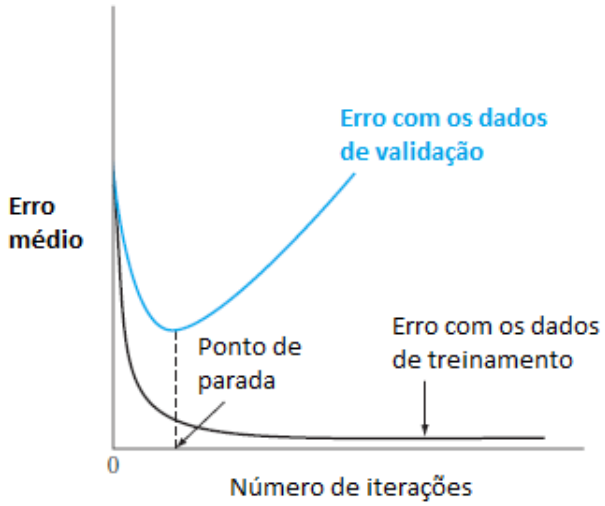
$w(n)$ é o peso anterior, na iteração *n*

A expressão da matriz S é apresentada em Haykin (2009). A menos que o peso já seja conhecido, a inicialização do método se dá escolhendo-se um peso inicial usando os mesmos procedimentos do algoritmo BP (Haykin, 2009).

2.3.5 Overfitting na etapa de treinamento

Um dos problemas encontrados na etapa de treinamento é o *overfitting*. Quando uma rede aprende com muitos exemplos de entrada e saída ela pode acabar "memorizando" os dados de treinamento e isso pode fazer com que a rede perca capacidade de generalização (Haykin, 2009). De modo a evitar isso, através de um procedimento conhecido como validação cruzada, o conjunto de dados para estimação dos parâmetros da rede é tipicamente dividido aleatoriamente em padrões de treinamento, de validação e de teste, conforme denominados por De Souza Jr.,(2015).

Os padrões de treinamento são aqueles efetivamente empregados na atualização dos pesos e biases. Na medida em que se aumenta o número de iterações, a tendência é que esse erro diminua. O erro para o segundo conjunto de dados, os padrões de validação (que não são empregados na atualização dos parâmetros da rede), é também calculado ao longo do treinamento, com o objetivo de monitorá-lo. Quando esse erro aumenta, devido à perda de generalização por *overfitting*, o treinamento é parado em um procedimento chamado por De Souza Jr. (2015) de "parada brusca" ou "*early stopping*". De modo a verificar o desempenho da rede, o erro para um outro conjunto independente de padrões, os padrões de teste, é então calculado para novamente verificar a capacidade de generalização da rede. A Figura 4 ilustra esse procedimento de "parada brusca". Nela, o erro de treinamento diminui com o aumento do número de iterações e a curva de erro de validação (em azul) diminui para um mínimo e, em seguida, aumenta. Neste procedimento, o ponto de mínimo da curva de erro de validação é o ponto de parada do treinamento (De Souza Jr., 2015).



Fonte: Haykin (2009) - Adaptado

Figura 4 Ilustração da regra do ponto de parada

2.4 Aplicação de RNA em caldeira e coluna

Como será apresentado nesse tópico, muitos trabalhos têm reportado a habilidade das RNAs em estabelecer uma conexão entre dados de entrada e dados de saída para interpretar o comportamento do fenômeno envolvido no processo estudado.

Dos objetivos bastante encontrados na literatura usando esta técnica de inferência encontram-se a predição dos componentes dos gases de combustão em um processo de queima (Tronci *et al*, 2002; Iliyas *et al*, 2013) ou a composição de um produto no fundo de uma coluna (Pani *et al*, 2015).

Nos trabalhos encontrados durante essa dissertação, os tópicos geralmente abordados e apresentados pelos autores que se utilizaram da ferramenta de RNA, são:

- Quanto às definições, ou seja, como cada autor entende e define a ferramenta;
- Quanto ao processo e o equipamento, apresentando suas principais características e o seu contexto na indústria;
- Quanto às variáveis, apresentando quais foram os dados de saída investigados, os dados de entrada utilizados, bem como o tamanho da amostra dos dados de entrada e como foram aquisitados;
- Quanto à estrutura da rede, FF ou FB, sendo a rede FF a mais encontrada;
- Quanto aos tipos de rede, em que o MLP e RBF são as mais utilizadas;
- Quanto às funções de ativação, onde para o tipo MLP as mais encontradas são as funções sigmoidal e tangente hiperbólica;

- Quanto ao algoritmo de treinamento, em que para as redes MLP o mais encontrado é o BP;
- Quanto ao uso de técnicas complementares, tais como a *Principal Component Analysis* (PCA) ou a *Genetic Algorithm* (GA);
- Quanto ao percentual de dados utilizado na validação, o número de camadas escondidas e o número de neurônios em cada camada escondida;
- Quanto à comparação entre duas técnicas, normalmente através do coeficiente R^2.

2.4.1 RNA aplicadas a caldeiras

Os principais gases medidos na exaustão de uma caldeira são o óxido de nitrogênio e o oxigênio (NOx e O_2). A medição de NOx é essencial para monitorar o nível de emissão, enquanto a medição de O_2 é usada para otimizar a eficiência de combustão.

A previsão do teor de oxigênio em gases de combustão em uma caldeira foi estudada por Valdman (2010) para o monitoramento *on line* de gases residuais e diagnóstico de uma caldeira operada com misturas de diesel e biodiesel. Sua rede preliminar era formada por nove variáveis de entrada, referentes ao combustível (vazão, temperatura e regulagem do excesso de ar), água (nível e temperatura de alimentação), vapor (vazão e pressão na linha), eficiência (da caldeira, calculada) e gases residuais (temperatura) e seis variáveis de saída, todas referentes à concentração, % ou ppm, dos gases residuais (O_2, CO, CO_2, NO, NO_2 e SO_2). Após uma análise dos dados brutos de processo foram desenvolvidos filtros para o tratamento dos mesmos, gerando uma matriz de dados para a definitiva aplicação. Através da análise de sensibilidade, três variáveis de entrada foram retiradas, devido a pouca influência (vazão de combustível, eficiência e nível) na capacidade preditiva da rede.

Shakil *et al*. (2009) usaram a técnica de RNA para a inferência de NO_X e O_2 para o monitoramento contínuo desses gases de combustão em uma caldeira industrial com capacidade acima de 100 MW. Seu estudo foi comparar as duas categorias de RNA (estáticas e dinâmicas) escolhendo aquela que melhor se adequasse à predição dos dados. Para as redes dinâmicas foram utilizadas duas estruturas básicas para a sua construção: *Input–output recurrent neural network (IORNN)* e *focused-time lagged feed forward neural network (FTLNN)*. As redes IORNN têm conexões de realimentação e as redes FTLFNN têm apenas entradas atrasadas.

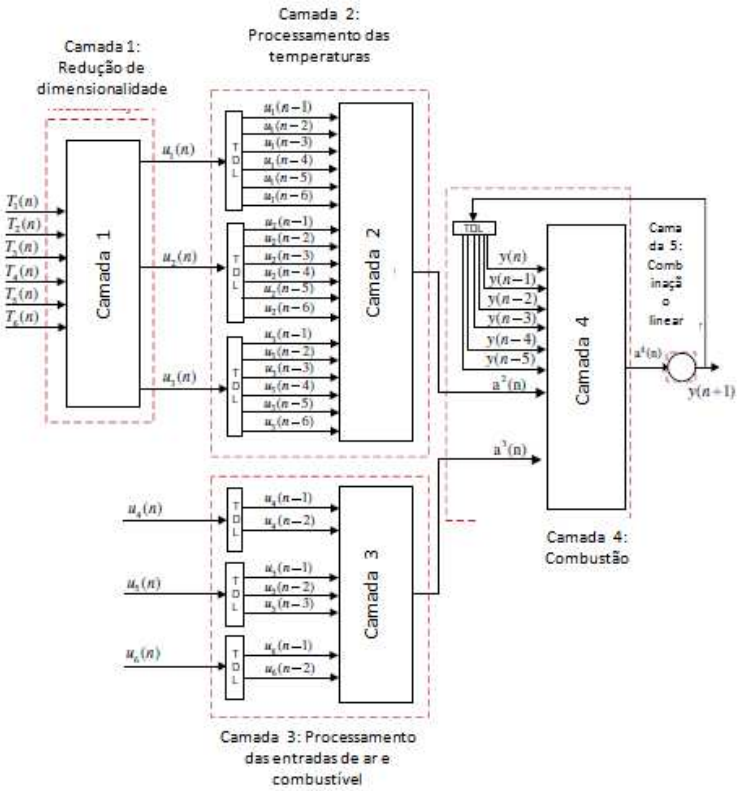
A rede estática MLP de Shakil *et al* (2009) apresentou 9 variáveis de entrada: seis temperaturas, vazão de ar, vazão de gás e relação ar/combustível. O tamanho da amostra foi

representado por 8700 padrões. O algoritmo de treinamento utilizado foi o BP Levenberg-Marquardt (LM). O número de camadas escondidas foi de 1 (um) com 50 neurônios nessa camada, 50% dos dados foram utilizados no treinamento e os outros 50% na validação. Devido ao comportamento dinâmico do NO_X a rede estática não conseguiu bons resultados na etapa de treinamento em algumas regiões dos dados.

Para a rede dinâmica, Shakil *et al*. (2009) utiliaram como dados de entrada seis temperaturas, vazão de ar, vazão de combustível e a relação ar/combustível. O tamanho desta rede foi representado pelos mesmos 8700 padrões da rede estática. Além disso, a Equação 2-24 representa o modelo dinâmico, com saída *y*, funções não lineares *g(.)* e *h(.)* e *k-1* valores imediatamente anteriores aos *k* valores, para o sistema apresentado na Figura 5. Também a técnica PCA foi implementada para a redução de dimensionalidade, onde as seis variáveis de temperatura foram reduzidas para três.

$$y(k) = f(g^{k-1}(.),\ h^{k-1}(.),\ y^{k-1}) \qquad (2\text{-}24)$$

Uma representação esquemática da rede dinâmica é apresentada na Figura 5, usada para os modelos de NO_X e O_2.



Fonte: *Shakil et al. (2009)* – Adaptado

Figura 5 Rede neuronal dinâmica para a predição de NOx e O_2

Ao total foram 4 camadas: duas camadas envolvendo as temperaturas (Camadas 1 e 2), sendo a saída da primeira camada a entrada da segunda (a primeira com seis variáveis de temperatura para aplicação do PCA e a segunda com três variáveis de temperatura, resultantes da primeira) e uma camada de processamento (Camada 3) com três variáveis (ar, combustível e razão ar/combustível) onde é aplicado o módulo FTLFNN com oito neurônios escondidos. As saídas dessas duas camadas são as entradas de uma camada de combustão (Camada 4) que consiste de um módulo IORNN (*Input-output recurrent neural network*). Essa camada de combustão é seguida por outra camada chamada de saída (*linear combiner node).*

Os resultados de Shakil *et al*. (2009) apontam para uma melhor *performance* da rede dinâmica. Na etapa de treinamento, mais de 99% dos resultados, tanto para o O_2 como para o NO_X, ficaram com um erro dentro de apenas $\pm$ 0,1% para as redes dinâmica e estática. Na etapa de validação, tanto para O_2 como para NO_X, a rede dinâmica continuou com a mesma *performance* do treinamento mas a rede estática ficou com um percentual abaixo de 90% com erro dentro de $\pm$ 0,1%.

Zhou *et al*. (2012) reportam o grande uso de BP na estrutura de uma rede neuronal MLP e FF para o mapeamento não linear de variáveis, comparando essas estruturas com outro modelo de regressão para a inferência dos gases NO_X e O_2 de uma caldeira. As técnicas utilizadas foram o *support vector regression* (SVR) e duas RNA: a rede *backpropagation neural network* (BPNN) e a rede neuronal *generalized regression neural network* (GRNN). O objetivo do estudo foi a predição do teor de NO_X de uma caldeira a carvão. Em suas referências apresenta as vantagens da SVR como, por exemplo, o treinamento resulta em um mínimo global e os parâmetros do modelo são menores do que para o BPNN (apenas dois parâmetros). Em relação à BPNN, foi utilizada a estrutura de MLP do tipo FF multicamada, para o mapeamento não linear dos padrões de entrada. A rede MLP tem 23 neurônios na camada escondida. Como algoritmos de treinamento foram citados o CG e o método de Newton, mas o utilizado foi o LM, tendo em vista que este foi bastante citado em outros trabalhos. Para comparação com o LM foi também empregado o método *steepest descent momentum*. Além da SVR e BPNN foi também utilizada a GRNN cuja equação fundamental se baseia em parâmetros estatísticos.

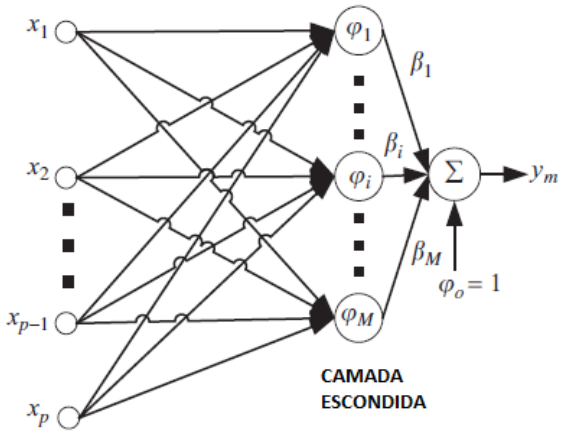
Os dados experimentais foram obtidos de uma caldeira de 300 MW, onde o carvão é queimado. Os valores de NO_X utilizados foram a média de várias horas em condição estável de operação. Os dados utilizados foram representados por 670 padrões, variando os parâmetros da caldeira para analisar as características do NO_X. A rede é composta por 21

variáveis de entrada: 10 níveis de velocidade, rotação para quatro moinhos, carga, % O_2 na vazão gás, vazão de ar, qualidade do carvão, teor de cinzas, capacidade calorífica e teor de umidade.

Primeiramente, a partir de uma equação de reta fundamental, Zhou *et al*. (2012) verificaram que a predição obtida através de um modelo *Multiple Linear Regretion* (MLR) obteve um erro de 37,7% e R^2 de 0,77, o qual não é adequado na predição, motivando o estudo de modelos não lineares SVR e RNA. Para os modelos SVR e RNA, as 21 variáveis de entrada foram utilizadas e dos 670 padrões, 446 (67%), foram para a etapa de treinamento e o restante 221 (33%) para a validação. Comparando os métodos LM e GDM, os autores encontraram uma melhor resposta preditiva para o LM, porém com tempo computacional mais longo. Outros métodos também comparados foram o Quasi-Newton e o CG. Os resultados dos modelos demonstraram na etapa de validação que para a predição do NO_X o melhor desempenho foi para o SVR (R=0,9355), seguido da rede GRNN (R=0,9267) e da rede BPNN (R=0,8890).

Iliyas *et al*. (2013) compararam diferentes estruturas de RNA para a predição de NO_X e O_2 de uma caldeira industrial. Tiveram como objetivo a predição e implementação de um sensor virtual para os dados de saída NO_X e O_2 em uma caldeira industrial com capacidade de 4000 t/dia de vapor e 160 MW de potência. As técnicas empregadas foram a RNA e o modelo 3D fluido dinâmico computacional. Antes de usar a rede para treinar os dados reais da caldeira, foi feita uma comparação das redes RBF e MLP para encontrar aquela com o melhor desempenho na predição dos gases.

Simulando a rede RBF, conforme representada na Figura 6, diferentes estruturas de rede foram testadas por Iliyas *et al*. (2013), variando de 5 a 30 o número de neurônios. A rede mais adequada foi com 20 neurônios. Na Figura 6 apresenta-se a rede RBF com x_p entradas, função de ativação φ_i gaussiana e pesos β para produzir o sinal y_m de saída. Os dados de entrada foram: razão ar/combustível, vazão mássica de combustível, vazão mássica de ar, temperatura máxima da câmara de combustão, temperatura média da câmara de combustão e temperatura dos gases de combustão. A rede foi treinada com dois conjuntos de dados. Um com as seis variáveis de entrada acima, a etapa de treinamento utilizou 10 padrões ou 60 vetores e para a validação foram utilizados 2 padrões, ou 12 vetores. No outro foram utilizados 114 vetores no treinamento e 24 na validação, com 8 variáveis de entrada incluindo, além das acima mencionadas, a temperatura de entrada do ar e ângulo de rotação. Em ambos os casos, o coeficiente de correlação foi 0,998 ou maior, o que indicou que apenas as seis variáveis eram suficientes para uma boa performance na predição.



Fonte: Iliyas *et al*. (2013) – Adaptado

Figura 6 Configuração típica de uma rede RBF

Simulando a rede MLP, Iliyas *et al*. (2013) utilizaram a estrutura FF. A melhor rede foi conseguida com uma camada escondida com oito neurônios e o método BP LM para o algoritmo de treinamento. A função de ativação utilizada para todos os neurônios foi a log-sigmoidal. Os dados utilizados no treinamento e validação foram os mesmos da rede RBF com 6 variáveis de entrada.

Após a simulação dos dados acima mencionados o modelo foi proposto por Iliyas *et al* (2013) para a caldeira em questão. Para os dados reais da caldeira, foi utilizado um total de 3000 padrões amostrados a cada um minuto e retirados os *outliers*. Os dados de entrada foram: duas temperaturas da superfície do tubo, vazão de combustível, vazão de ar e razão ar / combustível. O modelo RBF apresentou o melhor resultado e 85% dos dados foram utilizados para treinamento e 15% dos dados foram utilizados para validação.

Tronci *et al*. (2002) propuseram o desenvolvimento de um sensor virtual como ferramenta capaz de estimar variáveis de um processo de exaustão com pretensões de monitoramento destes gases no meio ambiente, tendo em vista a liberação de NO_X e CO, e também do ponto de vista econômico, tendo em vista o O_2.

Para estimar as variáveis significativas para o nível de poluição foi proposto um estudo de inferência dos gases da câmara de combustão, cujos dados foram obtidos de uma planta piloto, para testar diferentes condições operacionais, alteradas a cada duas horas e, portanto não representando o estado estacionário. A planta piloto, com capacidade de 4,8 MW, construída no centro de pesquisa de Santa Cecília (Cagilari, Itália) mantendo-se as condições térmicas e cinéticas de uma planta real em Vado Ligure. Os autores basearam-se em trabalhos anteriores, de Lu e Hogg no ano de 2000 e de Ikonen *et al.* nesse mesmo ano,

que obtiveram êxito na aplicação das RNAs FF no desenvolvimento de sensores para sistemas dinâmicos. A partir de todos os dados, foram retirados os *outliers* de NO_X, quando o combustível queimado era carvão fazendo o uso de *nozzles*, não especificados no projeto do equipamento. As variáveis de entrada foram carga térmica, vazão do carvão pulverizado, vazão do combustível óleo, vazão de ar e relação ar / combustível. Os dados de saída foram NO_X, O_2 e CO. Foram realizados dois testes com janela operacional de 100 minutos. O primeiro operou a carga térmica constante com mudanças no total de ar de entrada e na relação ar combustível, mantendo o carvão como combustível. Comparando os resultados preditos com os reais, com 18 h de operação, o erro foi pequeno , o que possibilitou dizer que a correlação e o sensor virtual são uma alternativa válida como sistema de monitoramento.

Com pretensões de racionamento de energia em um processo de combustão em caldeiras, Kljajic *et al.* (2012) utilizaram as RNAs como ferramenta para a predição da eficiência de um conjunto de 65 caldeiras, selecionadas aleatoriamente e localizadas em 50 *sites* ao norte da Sérvia, incluindo usuários industriais, comerciais e públicos. Os autores utilizaram uma rede MLP com estrutura FF e como algoritmo de treinamento, o BP *steepest descent*. As variáveis de entrada foram o tipo de combustível, tempo de operação em anos, tipo de caldeira, capacidade, range da carga e teor de oxigênio. A variável de saída estudada foi a eficiência. O número de neurônios na camada escondida foi variado de 5 a 30, encontrando que 10 neurônios eram suficientes. A função de ativação utilizada foi a função de transferência sigmoidal. A validação foi feita com 60 padrões de entrada e saída (10% dos dados segundo pesquisa realizada). O erro foi menor que 0,1%.

Os maiores desvios entre os valores reais e preditos foram encontrados para maiores valores do teor de oxigênio e para menores valores de carga da caldeira. O sistema estudado trabalha com oxigênio entre 3% e 5% e carga entre 60% e 90%, obtendo para essas faixas melhores predições. A análise de sensibilidade mostrou que para uma caldeira água tubular a gás natural, a maior variação de eficiência foi obtida em função da variação do teor de oxigênio dos gases de combustão, seguido do percentual de carga e da capacidade instalada. Para esse tipo de caldeira, a rede conseguiu predizer 90% dos casos. Outras caldeiras também foram estudadas encontrando que melhorias na eficiência podem ser conseguidas de 0,5 a 1,0% para caldeiras a gás natural e de 0,5 a 1,5% para aquelas que queimam óleo pesado. Como resultado do estudo, a modelagem passou a ser capaz de revelar as fontes de variação da alta eficiência auxiliando nas recomendações de operação do equipamento.

A Tabela 2 mostra resumidamente os principais dados das RNAs para os estudos de caso apresentados neste capítulo.

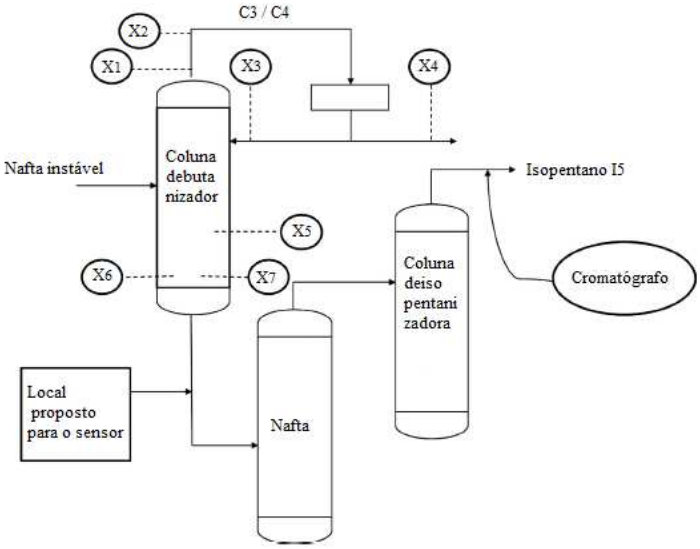
Tabela 2- Comparação das principais variaveis utilizadas em RNA nos trabalhos aplicados a caldeiras.

	Técnica(s)	Dados investigados	N° variáveis de entrada	N° de padrões	Estrutura	Tipo de rede	Função de ativação	Alg de treinamento	% no treinamento	% de dados valida	N° de neur cam. Esc.	R^2
Shakil *et al.* (2009)	RN estática	NOx e O_2	9	8700		MLP		LM e BP	50%	50%	50	
	RN dinâm.								50%	50%	25	
Valdman (2010)	RNA	O_2, CO, CO_2, NO, NO_2 e SO_2	9	614		MLP	Tangente hiperbólic		80%	20%	10	0,95
Iliyas *et al.* (2013)	RN. Comparou a MLP com a RBF.	NOx e CO_2	6	12	FF	MLP (simulação)	log-sigmoidal	LM BP	83%	17%	8	
			6	12		RBF (simulação)	Gaussiana		83%	17%	20	0,99
			8	17		RBF (simulação)	Gaussiana		83%	17%	20	0,99
			5	3000		RBF (dados reais)	Gaussiana		85%	15%	20	
Zhou *et al.* (2012)	RN e SVR. Comparadas	NOx e CO_2	21	670	FF	MLP		GDM LM	67%	33%	23	
Tronci *et al.* (2002)	RN	NOx, O_2 e CO	6	100 minutos de coleta	FF	MLP						
Kljajic *et al.* (2012)	RN	Eficiência	6	325	FF	MLP	Sigmoidal	BP steepest descend	90%	10%	10	Err0 < 0,1

2.4.2 RNA aplicadas a colunas

Pani *et al.* (2015) desenvolveram um estudo em uma coluna debutanizadora para a inferência do teor de butano no produto de fundo da coluna. A estrutura da rede apresentou sete neurônios na camada de entrada e um neurônio na camada de saída. Os dados foram igualmente divididos em dois conjuntos, um para treinamento e outro para validação do modelo. A representação da unidade onde a aplicação da RNA foi desenvolvida pode ser visualizada na Figura 7. Devido à longa distância do cromatógrafo em relação ao ponto de interesse na medição do butano, o analisador virtual foi proposto.

A coluna em estudo é utilizada para a separação de gases leves e GLP, provenientes de uma nafta instável, contendo o butano a ser recuperado no topo. O controle da qualidade da coluna é dificultado devido à falta de monitoramento *on line*.



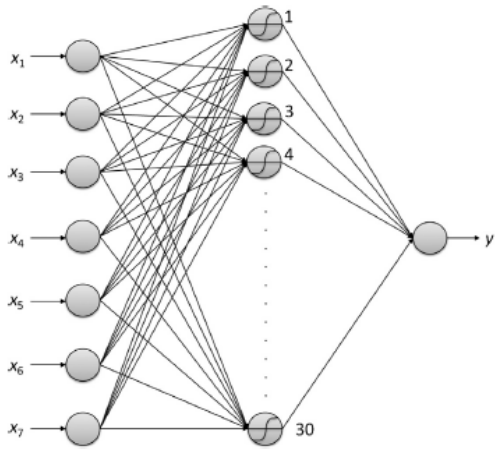
Fonte: Pani *et al.* (2015) – Adaptado

Figura 7 Representação da unidade de debutanização

As técnicas de predição utilizadas por Pani foram a Regressão Linear Múltipla (RLM), a Regressão por Componentes Principais e a RNA BP com o algoritmo LM. O modelo obtido por RNA foi o de melhor performance.

As sete variáveis de entrada para a inferência do teor de butano no fundo foram: Temperatura de topo; Pressão de topo; Vazão de refluxo; Vazão de retirada; Temperatura do sexto prato; e duas temperaturas de fundo.

A Figura 8 representa o modelo de RNA da debutanizadora, com 7 variáveis de entrada, 1 variável de saída e 30 neurônios na camada escondida.



Fonte: Pani *et al*. (2015)

Figura 8 RNA da coluna debutanizadora.

Os testes foram realizados a partir de 2394 dados coletados e o *Software Matlab* foi utilizado para implementação do mesmo. Na divisão dos dados para treinamento e para validação foi utilizado o algoritmo Kennard–Stone. A RLM apresentou um coeficiente de correlação R igual a 0,313 para os dados de treinamento e 0,395 para os dados de validação.

O modelo usando a técnica de componentes principais foi encontrado calculando os auto vetores e auto valores da matriz de covariância. O resultado desse modelo indicou 5 componentes principais, englobando mais de 95% da variância. No entanto, o coeficiente de correlação obtido foi muito baixo.

O modelo de RNA BP foi desenvolvido otimizando o número de neurônios na camada escondida, variando de 3 a 40, obtendo-se 30 neurônios como o melhor resultado. As funções de ativação da camada escondida e da camada de saída foram respectivamente a tangente hiperbólica e a linear. O algoritmo de treinamento utilizado foi o GD e o LM. Os resultados, em termos do coeficiente R para os três diferentes algoritmos de treinamento, encontram-se na tabela 3, onde se vê os melhores resultados de regressão para a técnica RNA empregando o algoritmo de treinamento LM. O trabalho de Pani demonstra a capacidade da RNA para aplicações em colunas, com o objetivo da inferência de uma variável de interesse.

Tabela 3 Resultados da RNA BP de 3 algoritmos de treinamento na predição do butano em uma coluna.

	Algoritmo de treinamento da BPNN	R
Dados de treinamento	GD	0,537
	CG	0,757
	LM	0,925
Dados de validação	GD	0,553
	CG	0,664
	LM	0,856

Fonte: Pani *et al*. (2015)

Ramli *et al*. (2014) fizeram a aplicação de RNA em uma coluna debutanizadora de uma indústria de refino de petróleo. O objetivo foi melhorar o monitoramento da qualidade dos efluentes, através das inferências das frações de topo e fundo do n-butano. Antes do emprego do modelo, o monitoramento era feito por dados de laboratório com o qual havia um atraso, próprio da natureza deste tipo de acompanhamento. As 3 técnicas de modelagem utilizadas e comparadas foram: RLM, Partial least square (PLS) e RNA.

As variáveis de entrada foram: Temperatura de topo; temperatura de fundo; temperatura no fundo do condensador; temperatura da nafta leve depois do condensador; temperatura de saída do *reboiler* para a coluna; temperatura de alimentação da debutanizadora; nível da coluna; nível do condensador; indicador de nível da coluna; indicador de nível do condensador; vazão de nafta leve para estocagem; vazão de GLP para estocagem e pressão no topo da coluna.

Os dados utilizados na modelagem foram obtidos num intervalo de amostragem a cada 1 minuto por 540 minutos. Desse conjunto de dados, 65% foram usados para treinamento e 35% para validação. As técnicas PCA e PLS foram utilizadas para determinação das principais variáveis do modelo. Para a RNA o número de neurônios na camada escondida foi de 8 a 40 e, através do erro calculado a cada simulação, obteve-se um melhor resultado para 10 neurônios.

Os três métodos conseguiram um elevado coeficiente de correlação: o R^2 da RNA foi igual a 1, tanto para o n-butano de topo como o de fundo; o R^2 da PLS foi 0,985 para o topo e 0,812 para o fundo; e a RLM chegou ao R^2 de 0,989 para o n-butano de topo e 0,815 para o de fundo. Os resíduos também apresentados para cada técnica mostraram os menores valores para a RNA ratificando sua superioridade no estudo de predição da coluna.

3 METODOLOGIA

A Figura 9 resume os passos da metodologia proposta para o projeto de redes neuronais transportáveis. Na sequência, cada passo é detalhado.

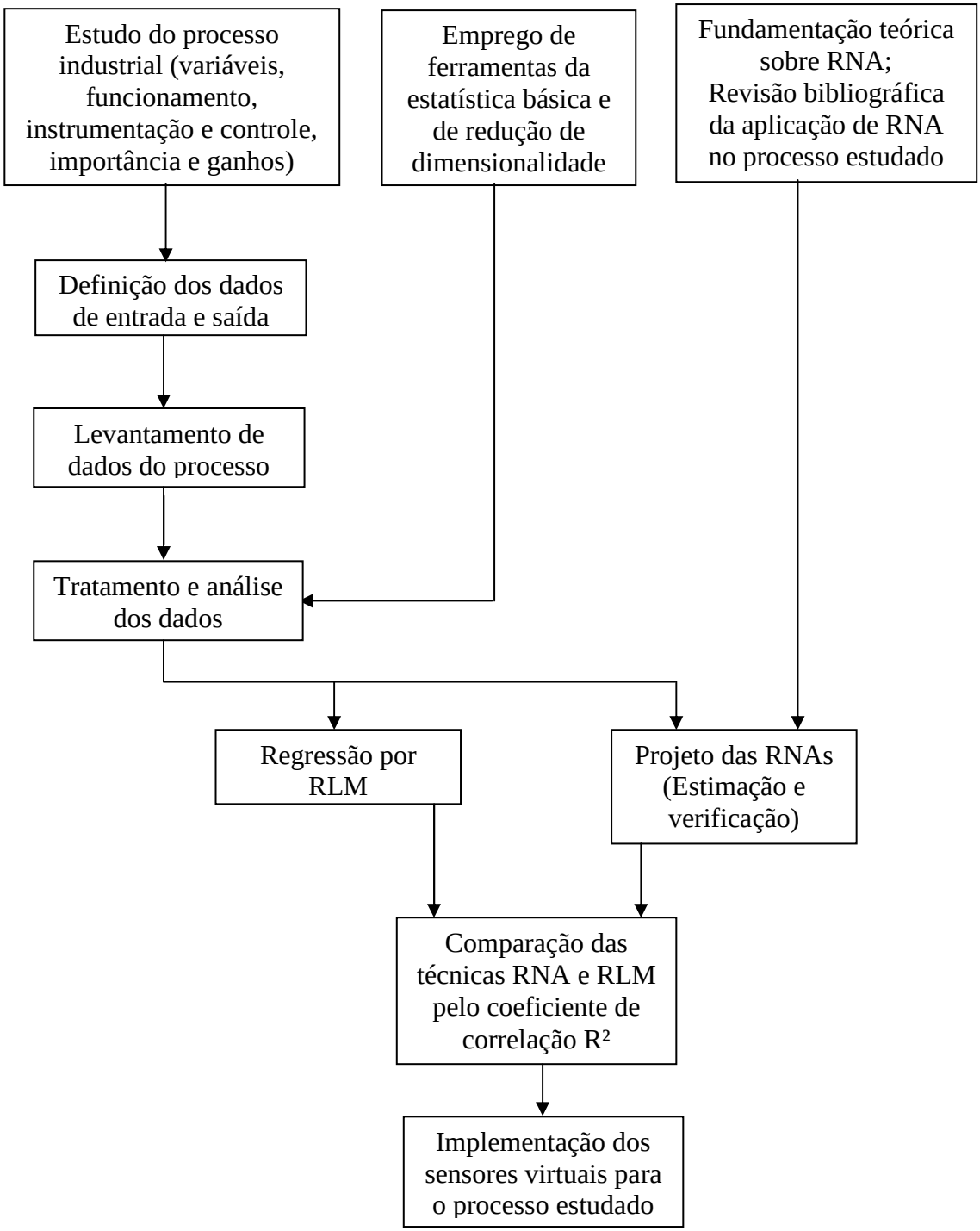


Figura 9 Fluxograma do desenvolvimento da rede

PASSO 1: ESTUDO INICIAL DO PROCESSO

O fato de as RNAs serem projetadas a partir de dados e não de modelagem fenomenológica não elimina a necessidade de se conhecer o processo a ser modelado, ou seja, conhecer seu funcionamento, instrumentação e controle, variáveis de entrada e de saída, condições operacionais, tipos de dados disponíveis, etc.. Somente assim, pode-se elaborar uma definição adequada do objeto de estudo e mesmo do objetivo da modelagem. Para tanto, um estudo sobre o processo deve ser realizado, de preferência *in loco* e envolvendo equipes multidisciplinares, incluindo engenheiros de processo e de controle, por exemplo. Informações de especialistas, de manuais de fabricantes dos equipamentos, da literatura, de processos similares (reais e simulados) devem ser levadas em conta nesta etapa inicial.

PASSO 2: LEVANTAMENTO E TRATAMENTO DOS DADOS

Na abordagem aqui proposta, parte dos dados será reservada para estimação dos parâmetros do modelo e outra parte para verificação do modelo, equivalendo a uma etapa de validação independente (Hair *et al*., 2005). O conjunto dos dados de estimação são aqueles usados para a determinação dos parâmetros da rede e são divididos aleatoriamente em padrões de treinamento (70% dos dados), padrões de validação (15% dos dados) e padrões de teste (15% dos dados), conforme explicado no item 2.3.5. O s dados de verificação serão usados para verificar a eficiência dos parâmetros obtidos na etapa de estimação.

Levantamento de dados

Esta etapa pode ser considerada uma extensão da anterior. Conhecido o processo, deve-se decidir como realizar o levantamento de dados para o desenvolvimento do modelo, analisar e tratar tais dados.

Dados levantados a partir de historiadores de processo estão sujeitos a tratamentos de compressão e exceção e esses quando utilizados devem estar desligados para melhorar a qualidade dos dados coletados. Dessa forma, é preferível obter os dados diretamente a partir de sistemas supervisórios sem terem sido submetidos a tais tratamentos.

O tamanho do conjunto de padrões é importante para RNA. Uma recomendação para o tamanho amostral usado para estimação dos parâmetros do modelo de RNA é dada por Hair *et al*. (2005), os quais indicam que o tamanho da amostra seja de 10 a 30 casos no conjunto de dados para cada peso e nunca ter muitos casos (o contrário de testes estatísticos em que o poder de inferência é maior para maior quantidade de dados). Segundo os autores um modelo pode ser ajustado com tamanho da amostra maior que o número de pesos, e perde generalização ao se tornar muito específico à amostra quando o número de casos é alto.

O número de padrões está ligado ao tamanho da janela temporal usada para o levantamento de dados e ao valor do período de amostragem, Análises estatísticas (função distribuição de probabilidade dos dados; média; variância; *t-test* para comparação de amostras etc.) são empregadas para a definição de tais parâmetros. No Apêndice A.1 essas ferramentas estatísticas são brevemente revistas. Essas ferramentas estão tipicamente disponíveis em softwares de análise, como se verá adiante.

Tratamento de dados

Os dados devem ser tratados para a retirada de *outliers*. Os métodos DMT e Hotteling são aqui empregados para este fim. Tais métodos são brevemente descritos no Apêndice A.1. Adicionalmente, nesta metodologia, consideram-se *outliers* dados acima ou abaixo de três desvios-padrão. Outras regras adicionais podem ser necessárias para uma filtragem eficiente dos dados como será visto no Estudo de Caso da coluna (Capítulo 5).

PASSO 3: DEFINIÇÃO DE VARIÁVEIS DE ENTRADA PARA O MODELO

O aumento no número de variáveis de entrada levará a redes maiores, com mais parâmetros, e que, portanto, demandarão mais padrões para seu treinamento (o chamado fenômeno de “maldição da dimensionalidade”) (De Souza Jr., 1993). Para evitar essa situação, quando o número de variáveis de entrada é grande, técnicas de redução da dimensionalidade podem ser aplicadas, tais como análise de componentes principais (PCA), análise de multicolinearidade, (que permite a identificação de duas ou mais variáveis que possuem o mesmo nível de informação), análise de sensibilidade etc. Tais técnicas são

brevemente revistas no Apêndice A.1 e também estão largamente disponíveis em softwares de análise.

PASSO 4: PROJETO DOS MODELOS DE RNA E RLM

Esta etapa é aquela em que a característica de transportabilidade das redes fica mais evidente, uma vez que o paradigma de redes é o mesmo para as diferentes aplicações, ou seja, a metodologia aqui proposta adota redes do tipo MLP com uma camada escondida. Os números de neurônios na camada de entrada e na de saída são definidos, respectivamente, pelo número de variáveis de entrada e de saída do sistema sendo modelado.

Para a estimação dos parâmetros da rede, adotam-se as seguintes diretrizes:

- Treinamento de um conjunto de algumas dezenas (a depender da complexidade do sistema a ser representado) de redes com combinações de diferentes funções de ativação nos neurônios das camadas escondida e de saída (identidade, logística, tangente hiperbólica, exponencial e seno) e algoritmos de treinamento (BFGS, gradiente descendente e CG);
- Função soma dos quadrados para a função erro, ou *sum of squares* (SOS), para o procedimento de minimização;
- Número de ciclos na etapa de treinamento igual a 200;
- Taxa de treinamento igual a 0,1 e fator de momento igual a 0,1 no método BP;
- 70% dos dados são usados para treinamento; 15% dos dados para teste e 15% para validação.

Como forma de se avaliar a eficiência do resultado de predição da RNA, analisa-se o coeficiente de correlação R^2 (um parâmetro conhecido no meio acadêmico e industrial e de fácil interpretação) para os diferentes conjuntos de padrões na etapa de estimação. A equação de determinação deste coeficiente de comparação dos valores preditos e reais se encontra no apêndice A.1. É ainda sugerida uma verificação independente do modelo com dados independentes. Finalmente, um modelo linear de regressão múltipla RLM é usado para comparação, uma vez que pela Lei da Parcimônia (Hair *et al*., 2005), deve-se sempre adotar o modelo mais simples que realiza a mesma função.

Rigorosamente, os estudos propostos nesta metodologia podem ser feitos em diferentes softwares (MATLAB, R, etc.) ou mesmo fora deles, por programação em

FORTRAN, C, Python, etc. Contudo, aproveitando-se de uma pareceria da UFRJ com a DELL, representada no Brasil pela BDA Solutions, utilizou-se o módulo Statistica Neural Networks do Statistica, versão 12.0.

PASSO 5: RETREINAMENTO DA REDE

A proposta desse passo é usar todos os dados disponíveis (de estimação e de verificação) para realizar a construção de uma rede e verificar se esta tem a mesma capacidade de predição daquela obtida no passo anterior com uma quantidade menor de dados.

PASSO 6: IMPLEMENTAÇÃO DO SENSOR VIRTUAL

Sugere-se a implementação no sistema supervisório, como procedido por Valdman *et al*. (2010), ou no sistema Plant Information (Salvatore *et al*., 2009).

4 ESTUDO DE CASO 1: O PROCESSO DE GERAÇÃO DE VAPOR

4.1 Introdução

Este tópico apresenta a aplicação da metodologia de desenvolvimento de redes transportáveis a um processo de geração de vapor em caldeira. Para tanto, inicialmente se estuda o processo, conforme PASSO 1 da metodologia proposta. Assim, no item 4.2, são apresentadas as informações levantadas nesse estudo feito na FCCSA: o princípio de funcionamento dos equipamentos, as variáveis de entrada e a variável de saída usada na construção do modelo, a instrumentação e controle presentes, sua importância na planta e os ganhos previstos com o estudo. O item 4.3 descreve a implementação do PASSO 2, de coleta e manipulação dos dados. Finalmente, o item 4.4 apresenta o desenvolvimento dos modelos de rede e regressão linear e a comparação entre ambas.

4.2 Passo 1: Estudo inicial do processo de geração do vapor

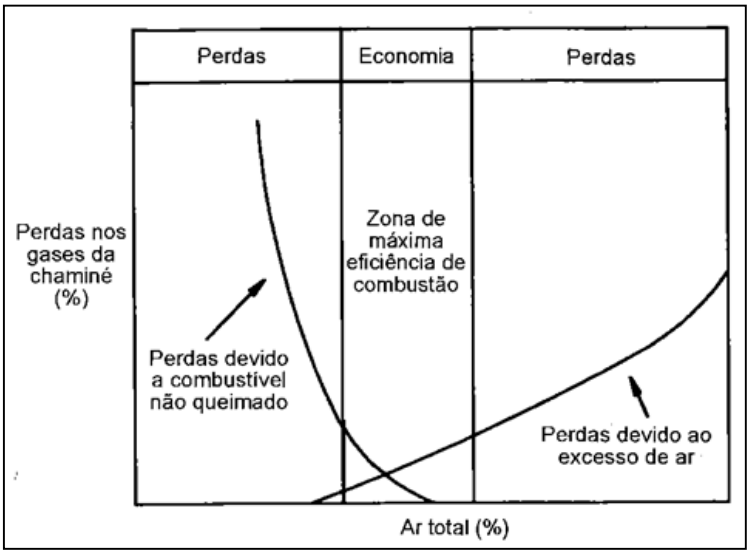
4.2.1 Princípios básicos de funcionamento

Durante o processo de geração do vapor ocorre a queima de um determinado combustível através da qual são gerados os gases de combustão, cuja alta temperatura atingida promove o aquecimento da água líquida até a formação de vapor.

Comumente, em sistemas de combustão, estão presentes os aspectos de custo, devido à energia despendida no processo, e o risco da operação, devido à presença de altas temperaturas e pressão. Nesse sentido, é importante obter a informação da composição dos gases gerados na combustão. Sendo esses gases uma mistura de vários componentes, o conhecimento do teor de oxigênio é o que permite tirar conclusões importantes sobre o funcionamento deste equipamento. O alto teor de oxigênio significa excesso de ar, o que compromete a eficiência energética da operação. Por outro lado, caso se utilize somente o ar estequiométrico, parte do combustível não será queimada e isso representa uma situação de risco da operação, proporcionando uma condição favorável à explosão do equipamento, uma vez que o combustível não queimado se acumulará no interior da câmara de combustão.

A Figura 10 mostra que existe um ponto ótimo de operação para a quantidade de ar e que tanto o seu excesso como a sua falta provocam perdas nos gases da chaminé. As perdas

por excesso aumentam em proporção muito menor que as perdas com combustível não queimado.



Fonte: Bega (2003)

Figura 10 Perda de eficiência pelos gases de combustão

4.2.2 Descrição da geração de vapor e suas variáveis

Na FCCSA, há duas caldeiras água tubulares em operação para atender à produção que demanda vapor em diversos processos, tanto na forma contínua como na forma em batelada, incidindo no equipamento uma variação na sua vazão de vapor e, consequentemente, uma perturbação no processo de geração de vapor.

Essas caldeiras são constituídas de dois *tubulões*, um superior e outro inferior, denominados tubulão de vapor e de lama, respectivamente. Interligando os tubulões há tubulações dispostas na forma de feixes tubulares e paredes de água. Entre esses tubos se encontra a fornalha. A queima de combustível ocorre na fornalha, onde o calor é transferido por meio de irradiação para as paredes da mesma, que são resfriadas com água, exceto a parede do queimador, que é revestida com refratário. Com a finalidade de melhorar o rendimento do equipamento, os gases quentes exauridos da zona de convecção são aproveitados para o pré-aquecimento do ar necessário à combustão.

Em relação à água de alimentação das caldeiras, o sistema possui uma unidade de desmineralização, que remove todos os íons presentes na água através de um processo de troca iônica, composto de um vaso desclorador, um vaso de troca catiônica e um vaso de troca

aniônica. Antes do processo de desmineralização, a água bruta captada no Canal São Francisco é tratada em uma Estação de Tratamento. O tratamento interno da água no interior da caldeira é feito pela adição de produtos químicos: inibidores de incrustações, sequestrantes de oxigênio e agentes passivadores.

A Figura 11 apresenta um diagrama de processo dos equipamentos e das etapas citadas.

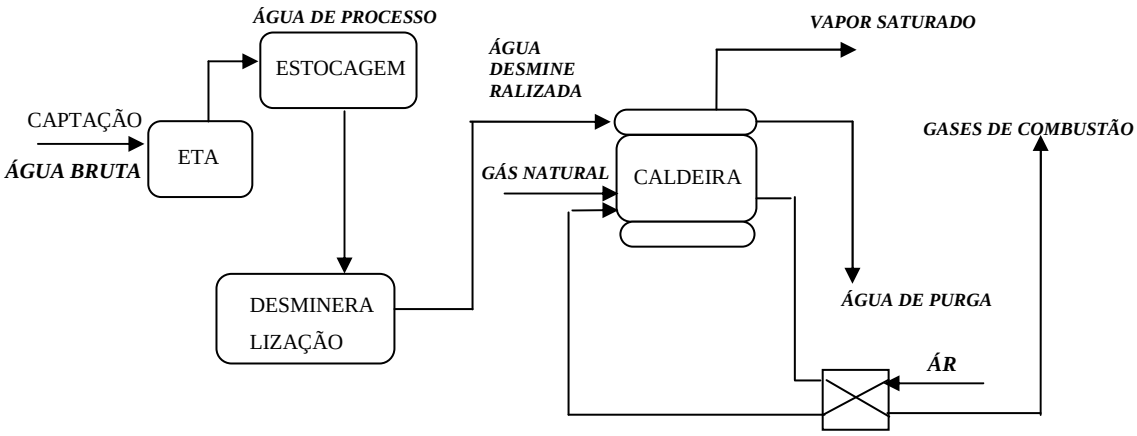


Figura 11 Digrama de processo das unidades

Na Tabela 4, encontram-se a variável de saída (Teor de oxigênio, %) e as variáveis de entrada, disponíveis no sistema de aquisição de dados e seus valores de projeto. A eficiência de projeto da caldeira é de 78,3%, calculada com o poder calorífico inferior do gás combustível.

Tabela 4 Variáveis de operação, utilizadas como entradas e saída da caldeira

Variáveis de entrada e saída do estudo	Valores de projeto
1- Teor de Oxigênio, %	12 % (excesso de ar nos gases de combustão)
2- Pressão de operação, kgf/cm²	13,5 kgf/cm^2
3- Vazão de gás natural, kg/h	1.158 Nm^3/h
4- Vazão de ar, kg/h	18.258 kg/h
5- Vazão de vapor, t/h	15.000 kg/h
6- Vazão de água, m³/h	16.000 kg/h
7- Temperatura dos gases de combustão, °C	272 °C

4.2.3 Instrumentação e controle da caldeira

Há na FCCSA em torno de 8 processos em batelada, demandando vapor separadamente em intervalos de tempo que ficam entre 10 a 25 minutos. Ou seja, em um determinado processo, admite-se vapor por 15-20 minutos e somente após 30-40 minutos, uma nova admissão é realizada. Há também consumidores contínuos, sendo um deles, o maior de todos os consumidores, cuja demanda é de 40% do total de vapor gerado. A variação de pressão se dá à medida que consumidores demandam vapor em processos em batelada. Como não há retorno de condensado, à medida que o vapor é solicitado na planta, o nível do tubulão de vapor também sofre variações.

A Figura 12 apresenta o esquema simplificado da unidade estudada. A numeração entre parênteses refere-se à variável de entrada ou de saída do estudo, conforme Tabela 4.

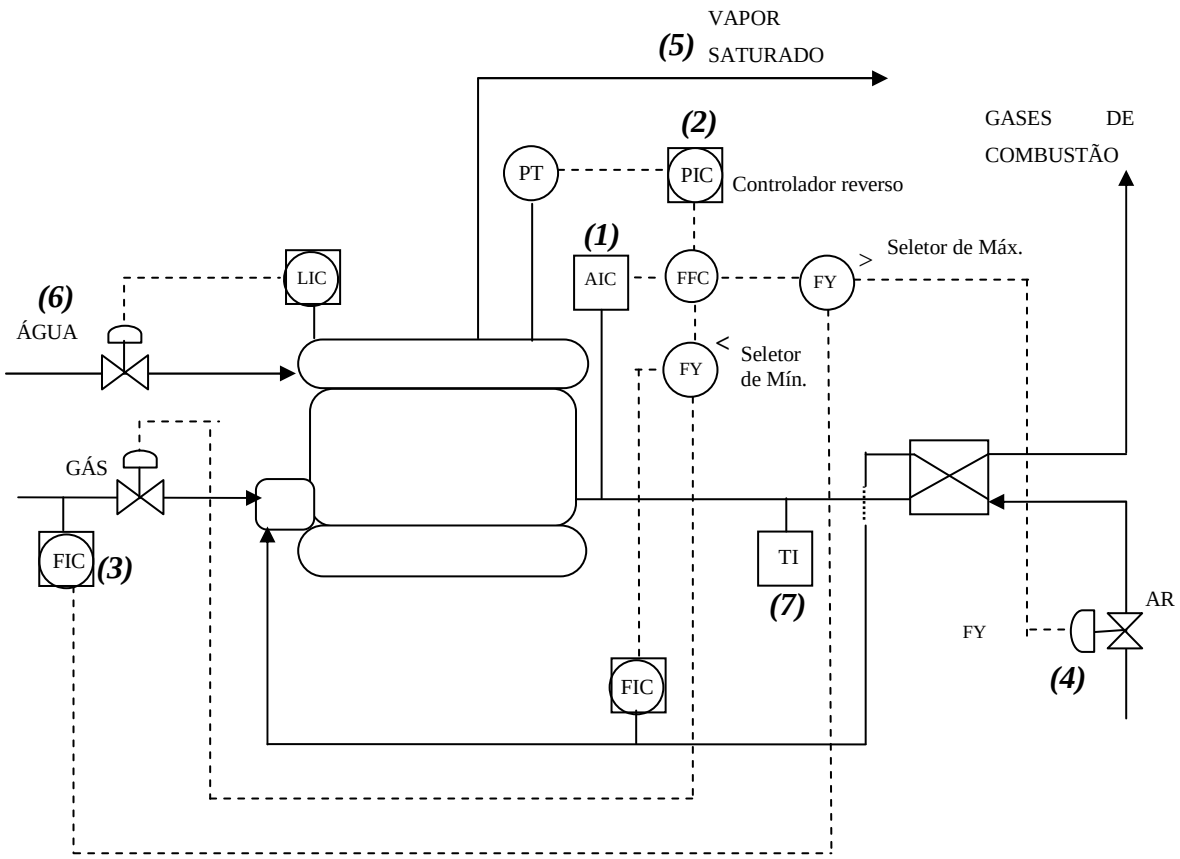


Figura 12 Representação esquemática da caldeira

Para o controle da produção de vapor, é necessário saber a pressão de operação, a vazão das correntes de entrada e a composição dos gases de combustão. Geralmente, no caso dos gases de combustão, é mais usual utilizar a medida do teor de oxigênio.

O controle das caldeiras da FCCSA é feito por uma malha de controle conhecida como limites cruzados. Nesse sistema, um controlador e indicador de pressão (PIC), ou mestre, opera em cascata com os controladores de vazão de gás e ar (*servos* ou secundários), ou seja, a saída do controlador de pressão é o *Set Point* (SP) das malhas de vazão de gás e ar. Pela estratégia de limites cruzados, o sinal de saída do PIC passa por seletores (FY) de máximo e de mínimo e após as comparações o sistema determinará qual será o SP a ser utilizado nas malhas de vazão de ar e combustível.

No controlador de pressão, com atuação reversa, a saída passa por um bloco de cálculo onde, através da relação ar/gás são calculadas as vazões requeridas de ar e de gás pelo controlador mestre.

No controlador de excesso de oxigênio, com atuação reversa, a saída é utilizada para promover pequenas oscilações (com batentes) na relação AR/GÁS.

No controlador da vazão de ar (FIC), com atuação reversa, o SP do controlador é obtido a partir da verificação do maior valor entre a vazão de ar requerida pelo PIC e a vazão de ar requerida pelo transmissor de vazão de gás. Esta condição faz com que, com a diminuição da pressão da caldeira, ocorra primeiro um aumento na vazão de ar. Já no FIC, com atuação reversa, o SP do controlador é o menor valor entre a vazão de gás requerida pelo PIC e a vazão de gás requerida pelo transmissor de vazão de ar. Esta condição faz com que, com o aumento da pressão da caldeira, ocorra primeiro uma diminuição na vazão de gás.

Para o controle de condutividade no interior da caldeira, o sistema possui o valor de condutividade medido e, a partir de um valor de arbitrado atua na abertura e fechamento da válvula que regula a vazão de purga.

4.2.4 Ganhos previstos no estudo da caldeira

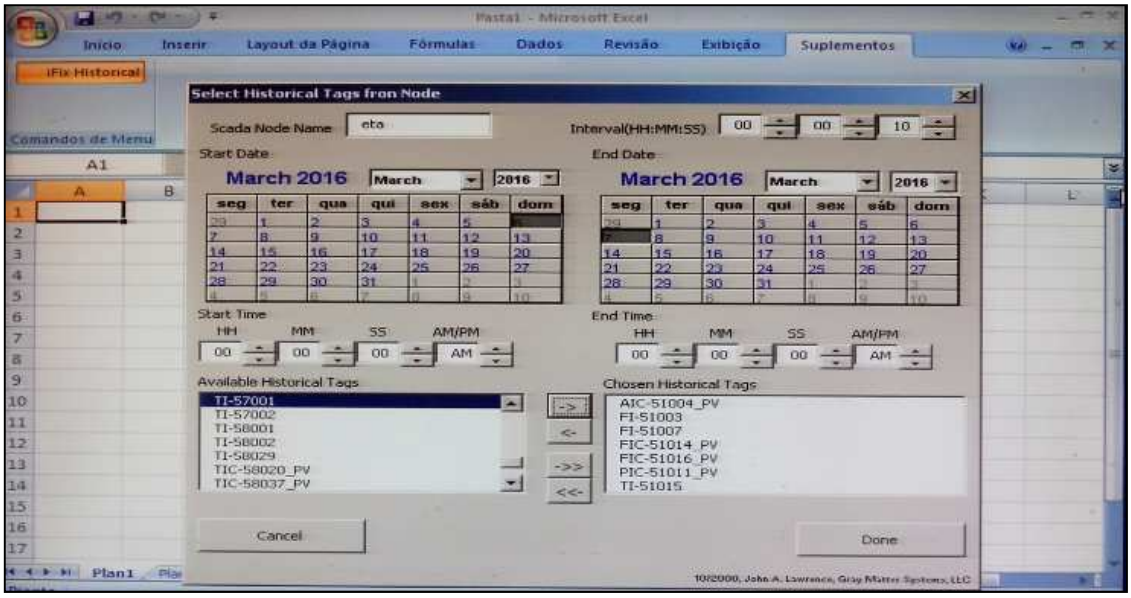
Desde a partida do sistema de medição do teor de oxigênio nos gases de combustão, há aproximadamente 7 anos atrás, tem-se registro de uma única vez em que esse analisador ficou fora de operação por motivos de manutenção externa. Nessa ocasião, operou-se a caldeira com a razão Ar / Gás maior que o típico, com o impacto do maior excesso de ar e consumo de combustível. Obter um analisador virtual que monitore o teor de oxigênio na ausência do

instrumento terá como benefício a operação em valores típicos de relação Ar / Gás. Por esse motivo o teor de oxigênio é uma importante variável a ser estudada.

4.3 Passo 2: Levantamento e tratamento dos dados da caldeira

4.3.1 Levantamento de dados

Os dados foram extraídos diretamente do sistema supervisório Scada da unidade, através de uma conexão com a Microsoft Excel, possibilitando a aquisição e análise de dados da planta pelo seu histórico. Para isso, uma sequência de dados é inserida em uma planilha do Excel, como: o tag ou identificação dos instrumentos de medida das variáveis de interesse, o nó onde os tags se encontram, o período inicial, o período final e o intervalo de tempo em que um dado e outro serão aquisitados. Com estas ferramentas, uma planilha é montada listando-se todos os dados correspondentes das variáveis escolhidas para o estudo. A Figura 13 apresenta a interface com o sistema supervisório Scada para a coleta de dados do sistema.



Fonte: Supervisório Scada

Figura 13 Sistema de aquisição de dados para a caldeira

Das duas caldeiras em operação, a escolha daquela que teria os dados utilizados foi em função da maior demanda, pois na segunda caldeira somente há geração de vapor quando a

necessidade no processo é tão alta que somente a primeira não consegue atender. Assim foi escolhida a caldeira cujo tag é BO-5101 B.

Na caldeira, escolheu-se para estimação dos parâmetros (incluindo treinamento, validação e teste) das redes um período de 24 h de operação, tempo considerado pelos operadores suficiente para que o processo apresentasse toda a sua variabilidade. Para a verificação (um segundo teste) da rede já projetada, escolheu-se um período de 24 h de outro dia.

Para que uma quantidade adequada de dados fosse coletada, dentro do período de 24 h, foi realizado um estudo comparativo entre diversos tempos de aquisição de dados. O tempo de amostragem de 10 s geraria uma alta quantidade de padrões. A esse tempo de amostragem foram comparados outros: 30 s, 60 s, 120 s, 240 s e 480 s.

Para essa comparação e definição do melhor tempo de amostragem, foi usada a técnica estatística *t-test* para amostras independentes, conforme mostrado na Tabela 5. Nessa tabela, o Grupo 1 é aquele obtido para tempo de amostragem de 10 s e o Grupo 2, para os demais tempos. A análise revela não haver diferença estatisticamente significativa, na coleta de dados em 10, 30 ou 60 s. O *p* valor reportado (> 0,99) indica que ao aceitar a hipótese de diferença entre os grupos comparados (10 s versus 30 s e 10 s versus 60 s), a probabilidade de erro nessa afirmação é maior do que 99%. Ao comparar o tempo de 10 s com os demais tempos (120 s, 240 s e 480 s), o que se verifica é uma maior probabilidade de erro.

O resultado desse estudo é que a aquisição de dados seja a cada 60 s, visto que não houve diferença estatisticamente significativa (demonstrado pelo *p* valor, do *t-test*) para os dados coletados a cada 10 s quando comparados com tempos de 30 s e 60 s; assim, para um período de 24 h, gerou-se um banco com 1.440 padrões para estimação dos parâmetros (incluindo treinamento, validação e teste) das redes. Adotou-se o mesmo tempo de amostragem para os dados de verificação (um segundo teste) da rede selecionada na estimação, obtendo-se também 1.440 padrões.

Tabela 5 *t-test* com os dados do sistema supervisório Scada em 6 tempos de coleta

Tempo de amostragem	Média grupo 1	Média grupo 2	p valor	N grupo 1	N grupo 2	Desv pad grupo 1	Desv pad grupo 2
10 s vs. 30 s	4,675	4,675	0,992	8641	2881	2,771	2,772
10 s vs. 60 s	4,675	4,676	0,998	8641	1441	2,771	2,768
10 s vs. 120 s	4,675	4,688	0,909	8641	721	2,771	2,760
10 s vs. 240 s	4,675	4,671	0,978	8641	361	2,771	2,760
10 s vs. 480 s	4,675	4,725	0,813	8641	181	2,771	2,759

4.3.2 Tratamento dos dados

Dando sequência ao tratamento dos dados da caldeira, primeiramente foram verificados e avaliados no intuito de retirar aqueles que fossem *outliers*. Por análise temporal dos dados não se verificou a necessidade de retirada de algum dado, embora haja essa tendência para quando o resultado de oxigênio está no patamar de 12%. No entanto, esses dados são verdadeiros por se tratarem de medição no limite máximo do instrumento.

Nas Figuras 14 e 15, faz-se uma apresentação dos dados utilizados para modelar a caldeira. Nas Figuras, os dados de estimação estão designados na legenda por EST e os de verificação por VER. Os padrões usados na RLM foram aqueles usados na estimação dos parâmetros da RNA. Os dados de verificação foram usados para testar tanto o modelo linear quanto o da RNA. Constata-se que os dados de verificação estão mais distribuídos entre o máximo e o mínimo e exibem mais amostras com maiores teores que os de treinamento. Como o objetivo destes dados é justamente o de verificar os modelos desenvolvidos, eles foram mantidos.

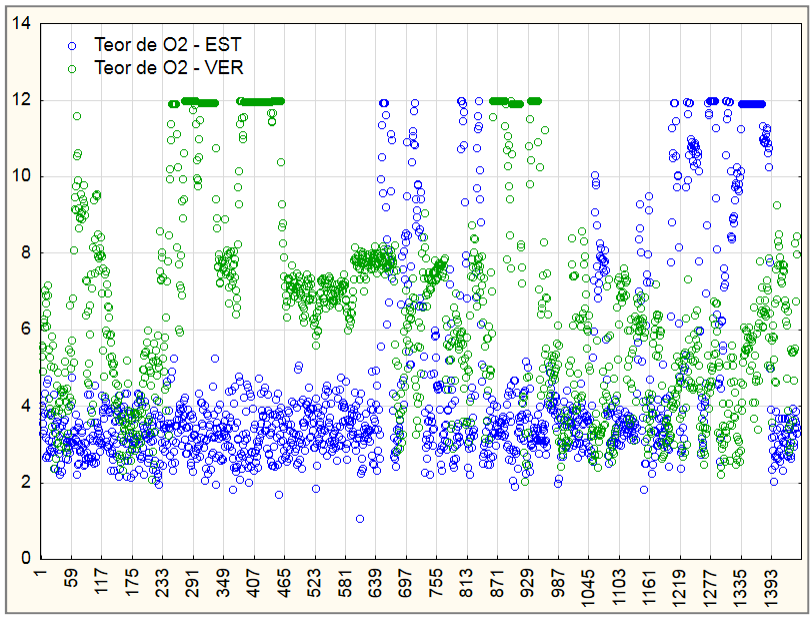


Figura 14 Padrões de teor de oxigênio na caldeira usados na estimação e verificação

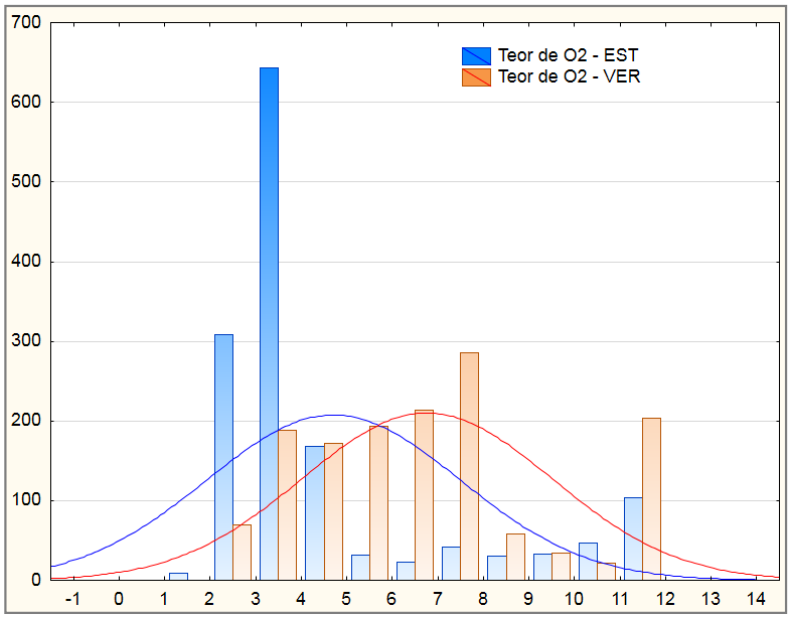


Figura 15 Histograma dos padrões de teor de oxigênio da caldeira usados na estimação e verificação

Tabela 6 Estatística dos dados de estimação e verificação da caldeira

Variável (unidade da medição)	Precisão dos instrumentos (%)	Dados de estimação (N=1.441)				Dados de verificação (N=1.441)			
		Média	Mínimo	Máximo	Desvio padrão	Média	Mínimo	Máximo	Desvio padrão
Teor de Oxigênio %	1,00	4,68	1,09	12,00	2,77	6,72	2,03	12,00	2,74
Vazão de Vapor t/h	0,50	5,83	0,95	8,92	1,63	4,35	0,00	7,18	1,50
Vazão de Água m³/h	0,50	6,89	0,00	13,91	2,25	5,07	0,00	11,16	1,82
Vazão de Gás kg/h	0,50	444,48	81,74	721,33	114,40	343,33	74,20	585,09	92,69
Vazão de Ar kg/h	0,50	10.819,68	9.291,34	15.007,02	1.321,25	9.822,48	9.486,66	12.270,04	252,40
Pressão kgf/cm²	0,50	13,40	13,16	13,73	0,07	13,40	13,04	13,73	0,07
Temperatura gases °C	1,00	248,61	221,91	281,56	12,38	236,04	210,34	259,27	6,83

Decidiu-se não aplicar técnicas de redução de dimensionalidade, uma vez que era pequeno o número de entradas. No entanto, após o treinamento das redes, foi feita análise de sensibilidade para verificar a importância das variáveis de entrada na predição, como se verá no próximo item.

4.4 Passo 4: Projeto dos modelos para a caldeira

4.4.1 Regressão por RLM na caldeira

Com os dados de estimação, um modelo de regressão linear múltipla foi desenvolvido para comparação com a RNA. Na Tabela 7 os resultados da regressão por RLM na caldeira, cujos coeficientes dados na coluna B possibilitam obter a correlação para a predição do teor de oxigênio. Nessa tabela, tem-se que b* é o coeficiente de regressão que seria obtido se todas as variáveis tivessem sido padronizadas de modo a se ter média igual a zero e desvio-padrão igual a 1. A vantagem de se trabalhar com o coeficiente b* está no fato de se poder comparar a contribuição relativa de cada variável independente na previsão da variável dependente.

Tabela 7 Resultados dos coeficientes da RLM na caldeira

	b*	Erro padrão de b*	B	Erro padrão de B	p valor
			2,550882	4,231299	0,546697
Vazão de vapor	-0,504933	0,030039	-0,860116	0,051170	0,000000
Vazão de água	-0,116832	0,012318	-0,143782	0,015160	0,000000
Vazão de gás	-0,728461	0,032872	-0,017629	0,000796	0,000000
Vazão de ar	0,949991	0,028572	0,001991	0,000060	0,000000
Pressão	0,024691	0,010622	0,915892	0,394010	0,020236
Temperatura dos gases	-0,320909	0,047640	-0,071778	0,010656	0,000000

O coeficiente de correlação, R^2, encontrado foi 0,94, um valor que para dados de processos industriais é muito bom. A disposição dos resultados preditos e observados da RLM no mesmo gráfico, como apresentado na Figura 16, mostra que em poucas regiões a inferência não é tão precisa, como pode ser destacado para valores de teor de oxigênio próximos a 12%. Uma explicação para tal falta de precisão é o alcance do limite superior de leitura do instrumento, não permitindo que outros resultados de teor de oxigênio entrem no modelo.

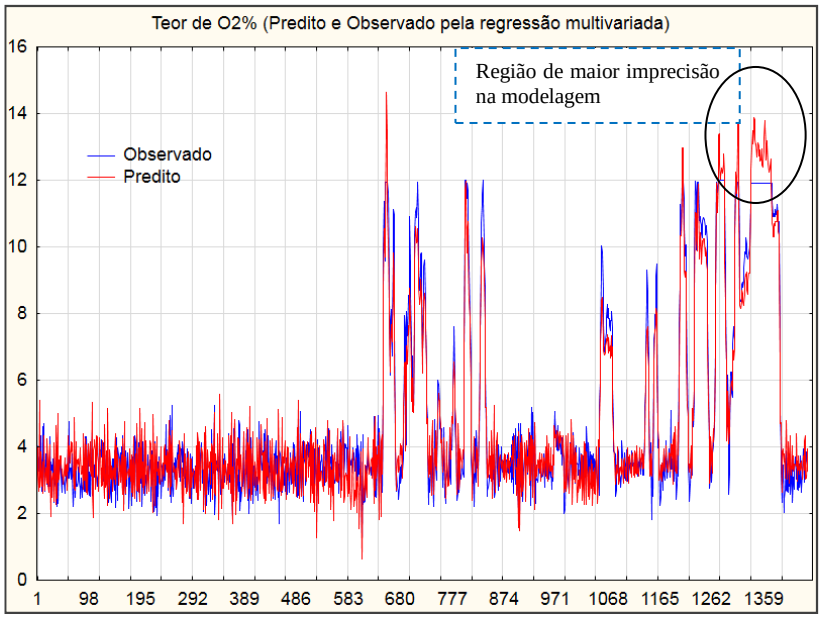


Figura 16 Valores de oxigênio reais e preditos por RLM na estimação

A Figura 17 mostra a disposição dos resíduos, dados pela diferença entre os valores observados e preditos. A distribuição dos dados segue uma curva aproximadamente normal, com a maior frequência de valores (em torno de 55%) iguais a 0 (zero), mais ou menos 0,5. A região de maior imprecisão, acima mencionada, no gráfico de resíduos mostra que a diferença entre os valores observados e preditos é na ordem de 2,5.

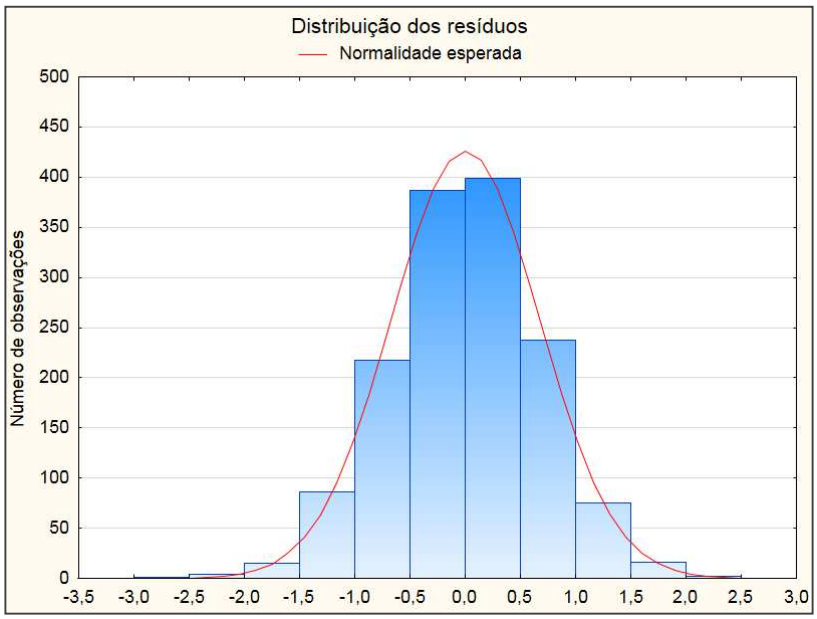


Figura 17 Valores residuais de oxigênio por RLM

4.4.2 Verificação da RLM para a caldeira

Nessa etapa, são empregados os padrões de verificação, diferentes dos usados na estimação dos parâmetros da RLM. O coeficiente de correlação foi de 0,87 (menor que na construção do modelo, que foi de 0,94), e a apresentação gráfica dos resultados se encontra nas Figuras 18, 19 e 20. O teor de oxigênio real tem um limite em 12% dado pelo instrumento, e com isso esse patamar quando estabelecido não é bem representado pelo modelo e, portanto, é a região em que se verifica a maior diferença entre o predito e o observado, como apresentam as figuras.

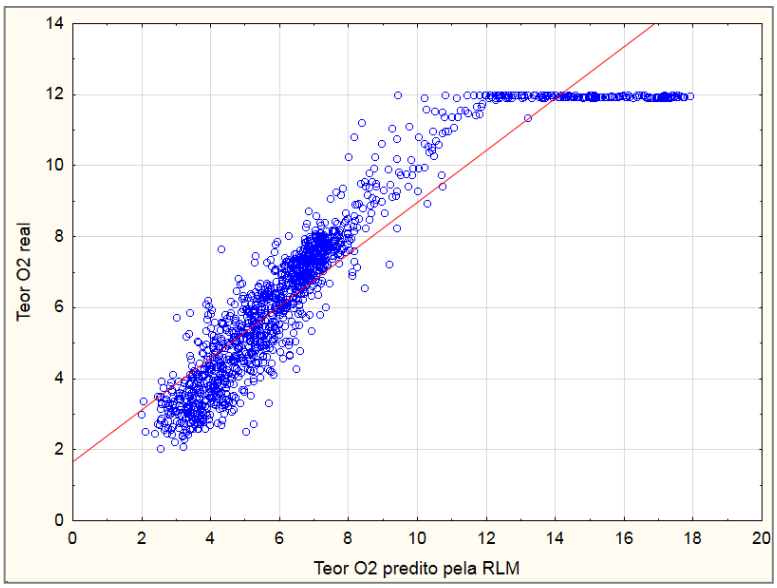


Figura 18 Valores preditos e observados na verificação do modelo de RLM da caldeira

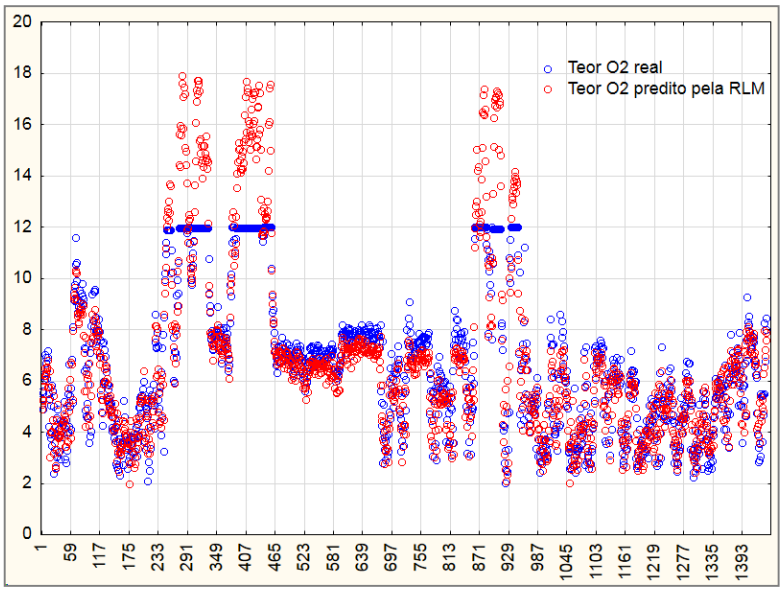


Figura 19 Valores preditos e observados na verificação do modelo de RLM da caldeira

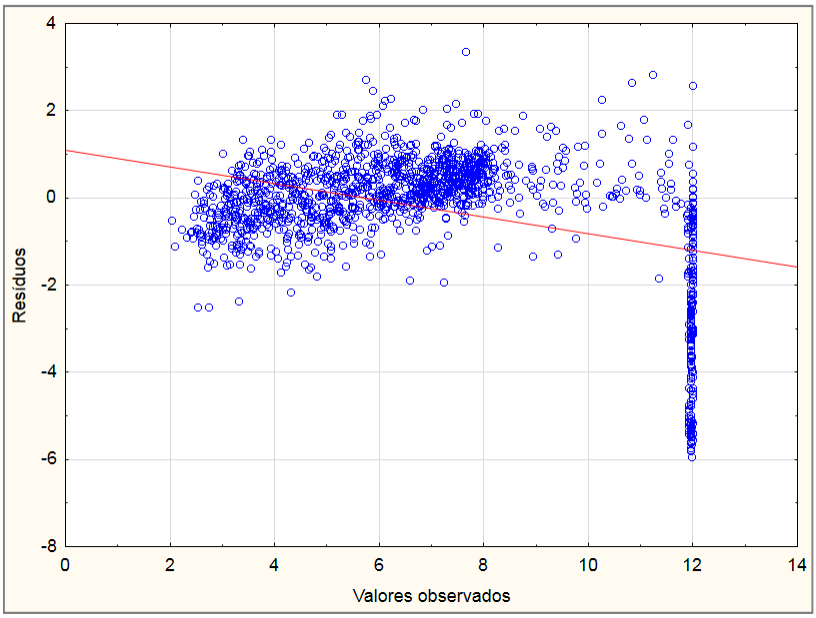


Figura 20 Análise de resíduos na validação RLM da caldeira

4.4.3 Regressão na caldeira: Treinamento das redes MLP

Essa etapa consiste da estimação dos parâmetros da rede da caldeira em que 70% dos dados são usados para treinamento, 15% para validação e 15% para teste da rede.

As variáveis de entrada apresentadas à rede foram seis (vazão de vapor, vazão de água, vazão de gás natural, vazão de ar, temperatura dos gases e pressão) sem redução de dimensionalidade. Nesse procedimento, 20 redes foram colocadas para treinar e cinco delas foram retidas para que uma fosse escolhida para melhor representar a inferência da variável de saída, o teor de oxigênio na caldeira.

Das cinco redes retidas, outras, além da escolhida, apresentaram um elevado coeficiente de correlação. Assim, foi utilizado como critério o número adequado de neurônios na camada escondida, pois muitos neurônios para o mesmo nível de resposta não é necessário.

A Tabela 8 apresenta os resultados da rede escolhida, uma MLP 6-10-1.

Tabela 8 Coeficientes de correlação na estimação dos parâmetros da caldeira

Rede	Coeficiente do treinamento	Coeficiente do teste	Coeficiente da validação
MLP 6-10-1	0,990	0,993	0,990

Nessa rede, o número de padrões para estimação dos parâmetros foi de 1440; o algoritmo de treinamento o BFGS; a função erro SOS; a função de ativação da camada escondida Tangente hiperbólica; e a função de ativação logística na camada de saída. Foram gerados 70 pesos e 11 biases, o que atende ao critério do número de padrões ser entre 10 a 30 vezes o número de pesos. Os pesos e biases dessa rede estão apresentados no Apêndice A.2. O coeficiente de correlação da rede MPL 6-10-1 superou o encontrado pela RLM.

A Figura 21 apresenta a disposição dos resultados preditos e observados, em que a localização dos dados bem próximos à reta é uma informação da boa precisão do modelo. Outro resultado dessa figura é que o erro entre os valores preditos e observados são similares ao longo de toda a extensão de resultados do teor de oxigênio.

Na Figura 22 a disposição dos resíduos ratifica a boa precisão do modelo. A maior diferença observada entre os valores preditos e observados é de mais ou menos 1,2, um erro menor que o encontrado pela RLM. A Figura 23, em complemento às anteriores mostra os valores preditos e observados no mesmo gráfico apresentando que os resultados seguem a mesma tendência e que a imprecisão observada na RLM quando o teor de oxigênio atinge 12% não é verificada com o modelo RNA.

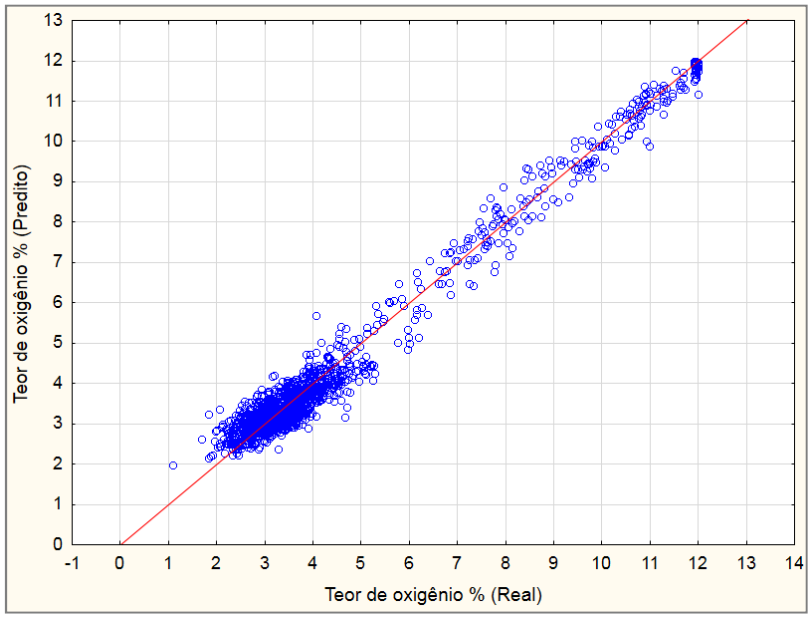


Figura 21 Dados reais e preditos de oxigênio da caldeira no treinamento

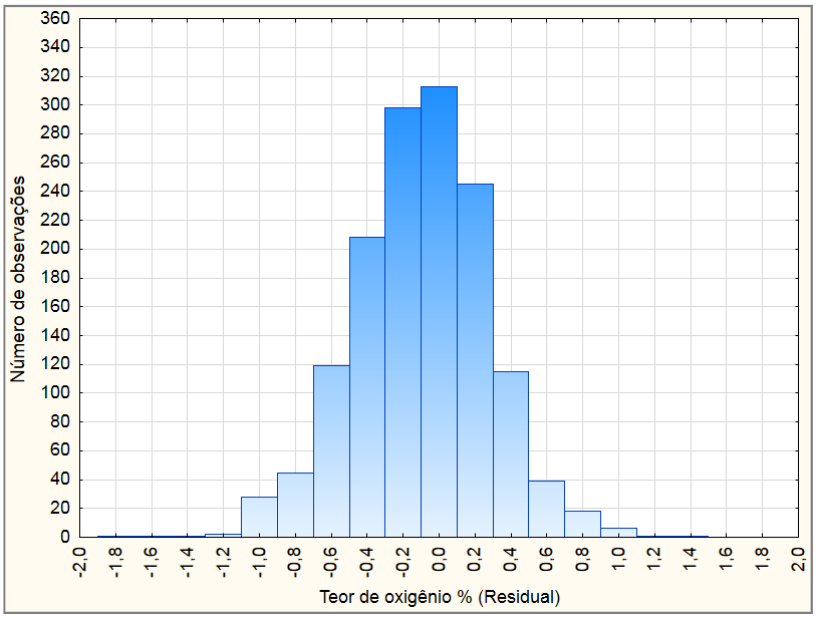


Figura 22 Análise da dispersão dos resíduos do teor de oxigênio da caldeira no treinamento

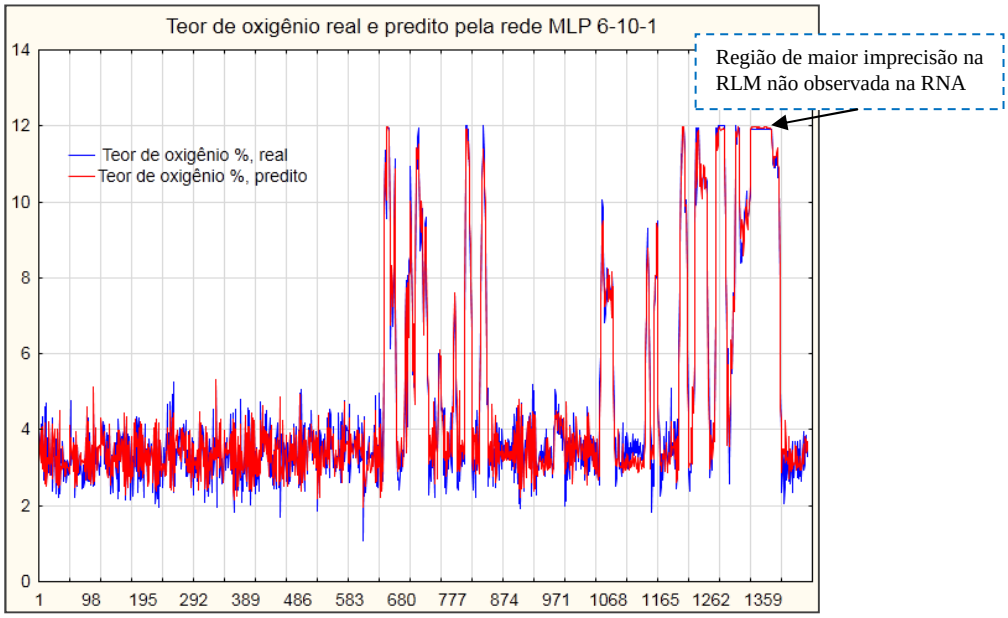


Figura 23 Dados reais e preditos do teor de oxigênio da caldeira no treinamento

4.4.4 Análise de Sensibilidade da rede MLP

Após a estimação dos parâmetros da rede foi feita a análise de sensibilidade, exibida na Tabela 9, demonstrando que a variável de maior valor nesse cálculo foi a vazão de gás seguida da temperatura dos gases.

Tabela 9 Análise de sensibilidade da MPL 6-10-1

Vazão de gás	Temperatura dos gases	Vazão de ar	Vazão de vapor	Pressão	Vazão de água
48,07	30,68	9,89	1,52	1,19	1,07

Como mostrado no Apêndice A.1, o cálculo da análise de sensibilidade é uma razão entre o erro da rede quando a variável está indisponível, substituída pela sua média, pelo erro total da rede quando os valores da variável são usados como entradas na rede. Dessa definição, resulta que a presença da variável vazão de gás tem correlação com a variável teor de oxigênio para valores maiores da vazão de gás. A relação entre essas duas variáveis pode ser verificada na Figura 24, confirmando a informação da análise de sensibilidade, para valores maiores da vazão de gás. Baseado na estequiometria de combustão e na teoria de eficiência de equipamentos térmicos, estas também ajudam a explicar e ratificar que a presença da variável vazão de gás é importante no modelo de determinação do teor de oxigênio. Quando a demanda de vapor é baixa, haverá uma redução na vazão de gás, acompanhada pela vazão de ar.

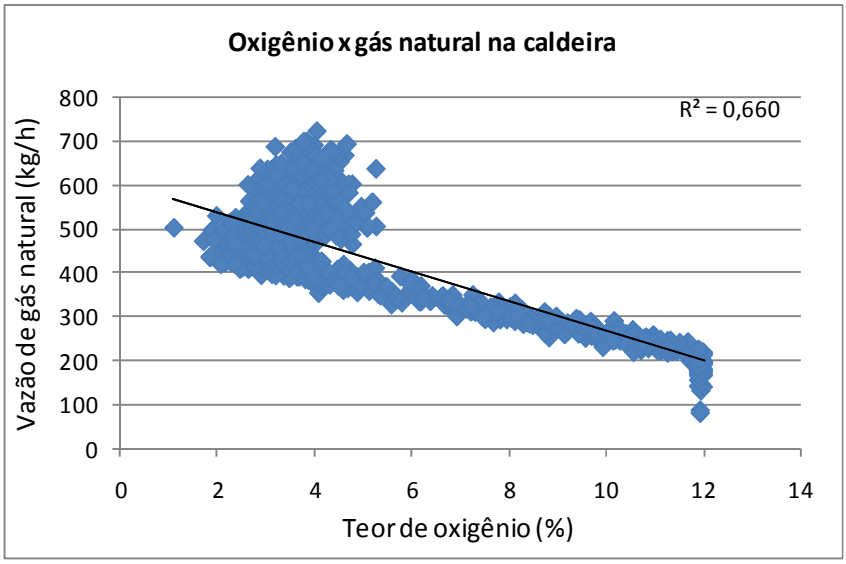


Figura 24 Relação entre teor de oxigênio e vazão de gás

4.4.5 Verificação da MLP projetada para a caldeira

Os padrões de verificação foram apresentados à rede MPL 6-10-1 e os coeficientes de correlação obtidos foram analisados. O coeficiente de correlação R^2 foi de 0,973. As Figuras

25, 26 e 27 exibem os resultados para estes dados. Ao contrário da verificação na RLM, a RNA apresentou bom comportamento para os valores reais informados pelo instrumento que tem um limite no teor de oxigênio em 12%.

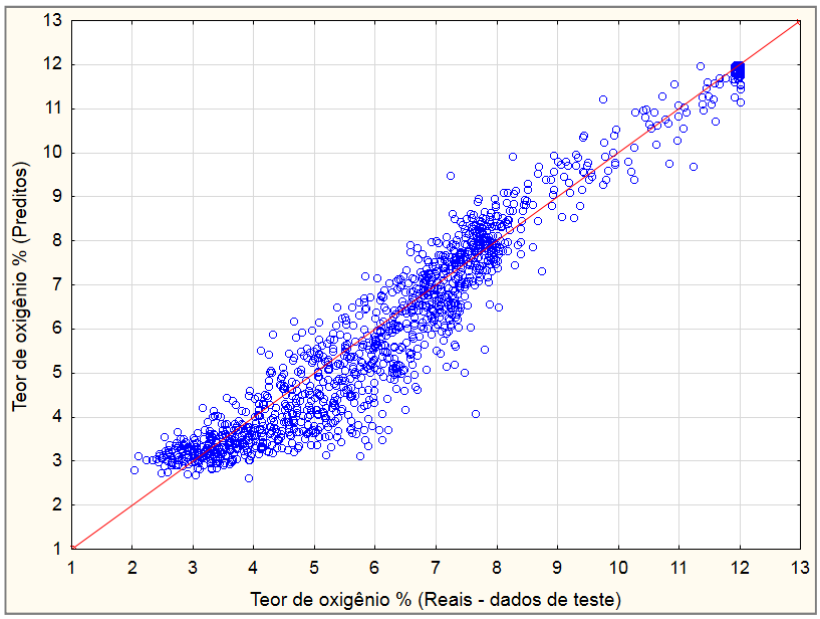


Figura 25 Dados preditos do oxigênio em relação aos observados na etapa de verificação

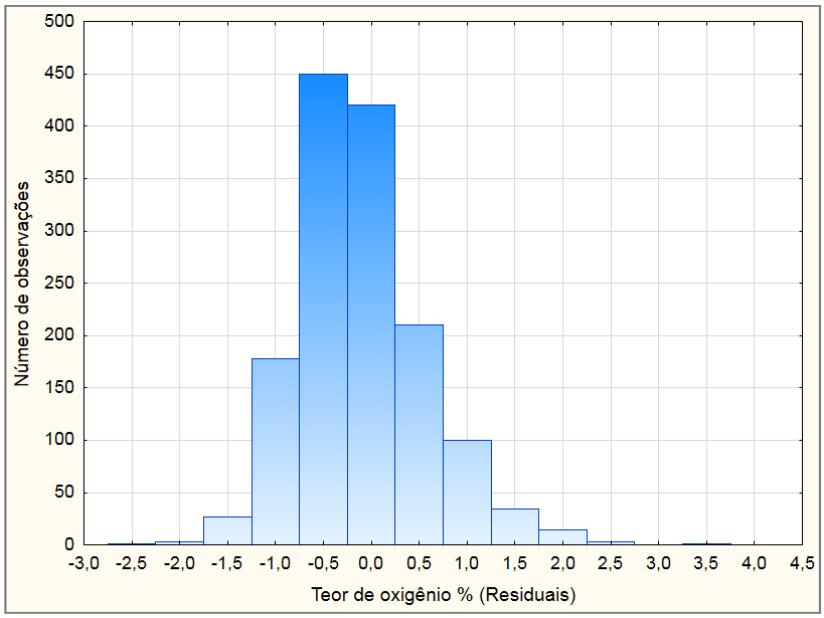


Figura 26 Dispersão dos resíduos do teor de oxigênio da caldeira na verificação

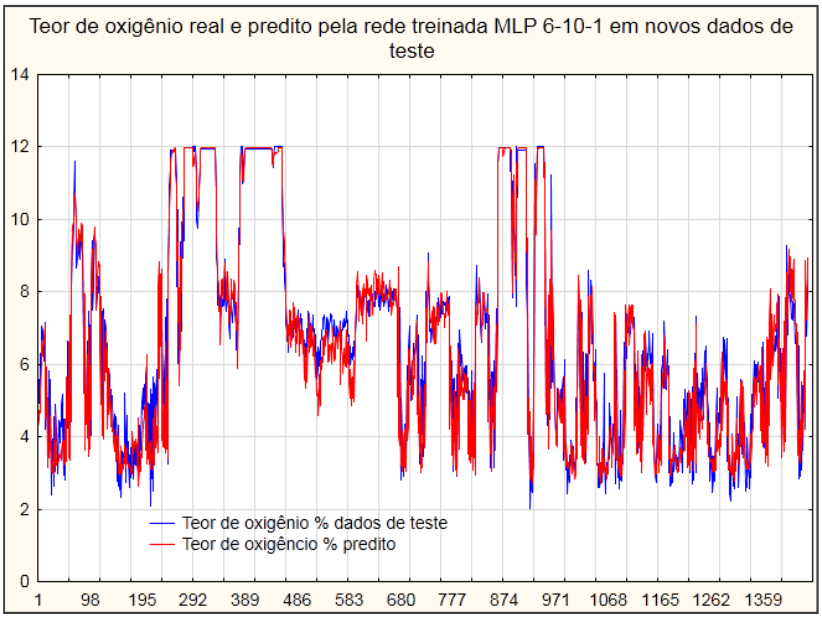


Figura 27 Dados reais e preditos do teor de oxigênio da caldeira na verificação

Considera-se que a etapa de implementação do sensor virtual no PI ou no sistema supervisório é uma etapa tecnologicamente dominada e que foge ao escopo desta dissertação. Por esta razão, não será apresentada e não implementado na FCCSA durante a dissertação.

Devido ao elevado resultado do coeficiente de correlação do modelo, não foi realizada a etapa de retreinamento, Passo 5 da metodologia.

4.4.6 Conclusões sobre a modelagem

Embora a RLM tenha conseguido um elevado coeficiente de correlação, a modelagem por RNA conseguiu superá-la, passando o R^2 de 0,94 para 0,99. As Figuras 19 (modelo RLM) e 27 (modelo RNA) demonstram bem a melhor qualidade da modelagem por redes, pois na comparação dessas Figuras valores preditos maiores que o limite de 12% no teor de oxigênio (conforme medição e informação do instrumento) não são dados pela RNA, mas são encontrados pelo modelo RLM. Também a verificação de ambos os modelos apresentou uma melhor previsibilidade da RNA (R^2 igual a 0,97) frente à RLM (R^2 igual a 0,87).

A boa *performance* do modelo de RNA foi importante para que seu uso efetivo como *Soft sensor*, ou até mesmo para estudos mais sofisticados de controle, possa se realizar sem maiores preocupações de seus usuários. Essa alta qualidade da previsão é necessária devido se tratar de uma variável mensuradora da eficiência de um equipamento que possui considerável atenção nos aspectos de segurança operacional.

5 ESTUDO DE CASO 2: PROCESSO DE RECUPERAÇÃO DE AMÔNIA NA COLUNA *STRIPPER*

5.1 Introdução

Este tópico apresenta a segunda aplicação da metodologia de desenvolvimento de redes transportáveis. A proposta é seguir os mesmos passos, com as adaptações exigidas pelas características do sistema.

Esta aplicação é a um processo em estágios numa coluna, outro equipamento também muito comum nas indústrias de processo e, portanto, interessante para testar a metodologia. Especificamente, o processo estudado é a recuperação de amônia em coluna *stripper*. O objetivo é recuperar amônia na corrente de topo, proveniente dos processos de produção, e ao mesmo tempo especificar que o teor de amônia no fundo esteja abaixo de um limite superior, para não causar impactos negativos nos processos subsequentes.

Do ponto de vista ambiental, a variável mais importante é o teor de amônia no fundo da coluna que deve ser mantido abaixo de 15 ppm. Para monitorar essa variável, a unidade possui um analisador de amônia instalado na corrente de fundo, após passagem pelo pré-aquecedor da carga. O analisador é configurado para fornecer um resultado a cada 20 minutos e entre dois resultados informados são efetuadas as etapas de limpeza e calibração automatizadas. Além do analisador, amostras são enviadas ao laboratório a cada 4 h para análise deste mesmo parâmetro. Embora a freqüência seja reduzida, quando o analisador apresenta problemas e necessita sair de operação somente a análise de laboratório auxilia no monitoramento e controle desta variável.

Os PASSOS 2 a 5 da metodologia aplicados a este caso são descritos sequencialmente nos itens 5.2 a 5.5.

5.2 Passo 1: Estudo inicial da coluna

5.2.1 Descrição da coluna e suas variáveis

Uma visão geral de onde se localiza a coluna a montante e a jusante das demais unidades na FCCSA é apresentada na Figura 28. O efluente amoniacal, oriundo de 4 unidades, passa por um processo de clarificação dos sólidos presentes por sedimentação e é estocado para alimentar a coluna.

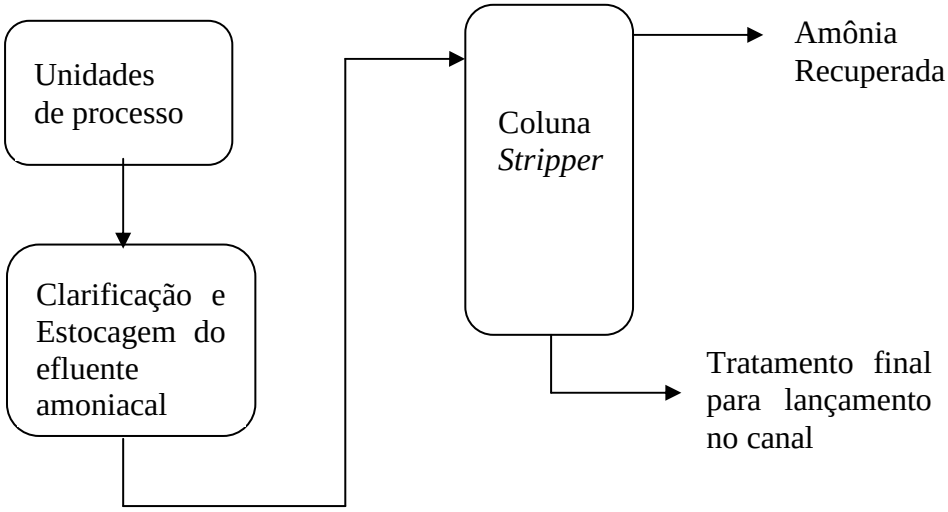


Figura 28 Esquema geral da coluna

Na Tabela 10, são apresentadas a variável de saída (Teor de amônia no fundo da coluna) e as variáveis de entrada (demais variáveis) que irão compor o conjunto de variáveis do estudo da coluna e que estão disponíveis no sistema de aquisição de dados da empresa. Na segunda coluna da Tabela, o nome simplificado das variáveis, pelo qual serão chamadas na parte de resultados. A variável vazão de vapor vivo é aquela proveniente da caldeira e a variável vazão de arraste na entrada da coluna é a soma do vapor vivo com o arraste do vaso *flash*.

Tabela 10 Variáveis de processo da coluna de recuperação de amônia

Variáveis de entrada e saída no estudo da coluna	Nome da variável simplificada
1- Teor de amônia no fundo da coluna (ppm)	NH_3 fundo
2- Vazão de alimentação da coluna (Kg/h)	Carga
3- Vazão de vapor vivo (Kg/h)	Vapor
4- Vazão de arraste na entrada da coluna (Kg/h)	Vazão arraste
5- Relação das vazões de alimentação e vapor vivo	Razão carga vapor
6- Relação das vazões de arraste e vapor vivo	Arraste
7- Vazão de soda cáustica (L/h)	Soda
8- Vazão de refluxo no topo (L/h)	Refluxo
9- Temperatura da carga (°C)	T carga
10- Temperatura de topo (°C)	T topo
11- pH da alimentação	pH carga
12- Temperatura do fundo (°C)	T fundo
13- Temperatura do vaso de *flash* (°C)	T *flash*
14- Temperatura do fundo após o pré aquecedor (°C)	T saída trocador
15- Pressão do vaso *flash* (Kgf/cm²)	P *flash*
16- Nível no fundo da coluna (%)	Nível fundo
17- Diferencial de pressão do leito no recheio superior (mm Água)	DP recheio sup

A Figura 29 apresenta com maiores detalhes as correntes dos produtos e insumos além das partes que compõem a coluna: o ponto de alimentação entre as seções de enriquecimento (composta por 4 pratos) e esgotamento (composta por 2 leitos de recheio), o condensador dos vapores que saem pelo topo que se utiliza de água de resfriamento como fluido frio, o vaso coletor da amônia recuperada, de onde parte é enviada para o tanque de estocagem e parte retorna à coluna como razão de refluxo, a admissão de vapor no fundo da coluna interligado ao vaso *flash* por onde amônia ainda é retirada da corrente de fundo e retorna ao processo por uma depressão causada por um ejetor. A numeração indica onde as variáveis apresentadas na Tabela 10 estão sendo medidas. As variáveis 5 e 6 (Relação das vazões de alimentação e vapor vivo e Relação das vazões de arraste e vapor vivo) não aparecem na Figura 29 por serem valores obtidos de uma relação e não através de uma medição.

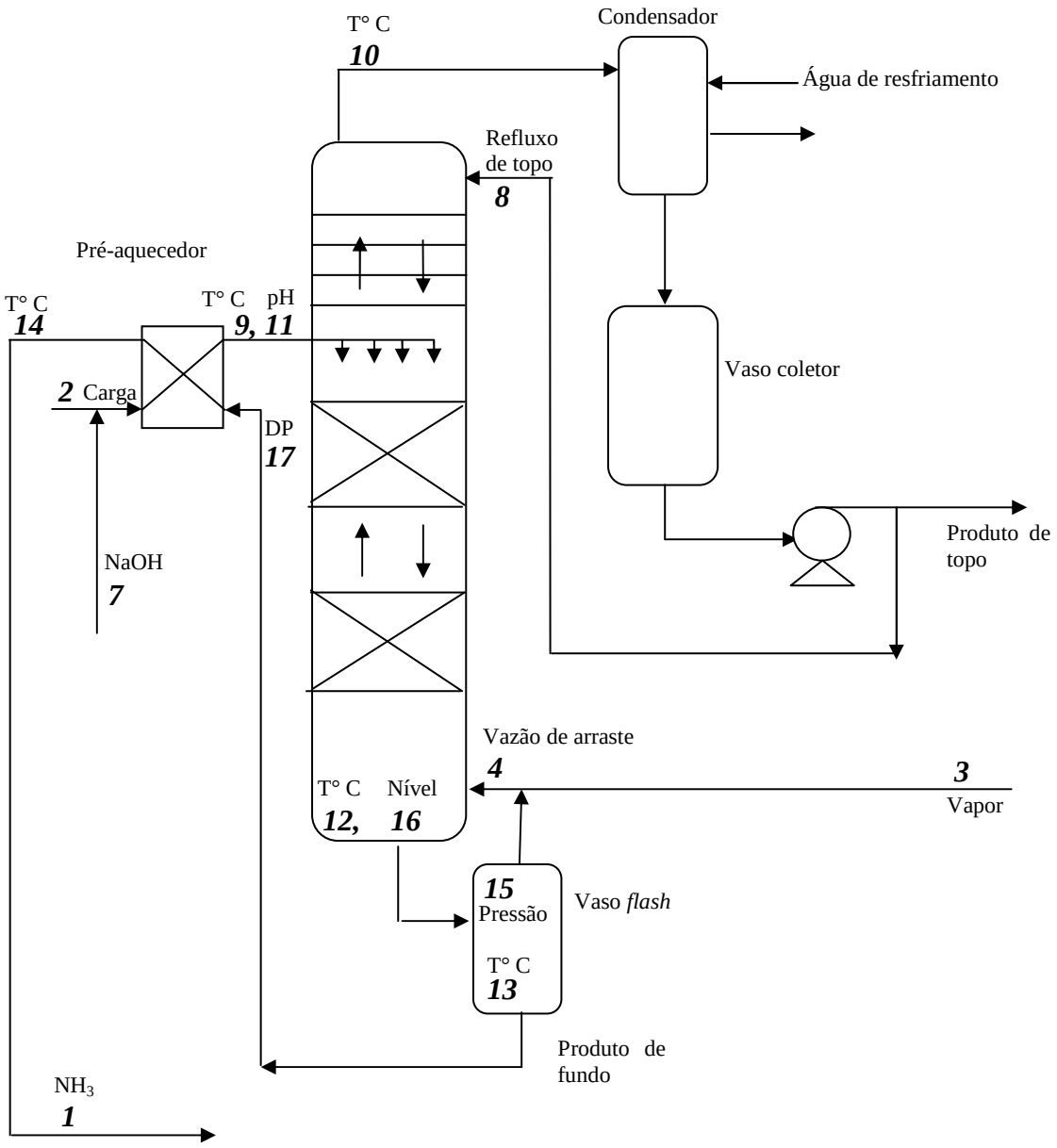


Figura 29 Representação esquemática da coluna

A coluna é dividida em duas seções: enriquecimento e esgotamento.

A seção de enriquecimento fica situada entre o ponto de entrada da alimentação e o topo e seu objetivo é concentrar a amônia. Essa seção é constituída de 4 pratos com borbulhadores e um “ladrão”, cuja finalidade é manter o nível nos pratos e escoar o excedente de líquido para o prato inferior.

A seção de esgotamento fica situada entre o ponto de entrada da alimentação e o fundo da coluna, em que o objetivo é o esgotamento da amônia para valores abaixo do limite de 15 ppm. Esta seção é constituída de dois leitos de recheios com um redistribuidor de líquidos entre eles. Cada leito de recheio utiliza anéis tipo Pall Rings colocados de forma aleatória.

A alimentação da coluna passa por um trocador de calor e por ajuste de pH. A fonte quente desse trocador é proveniente do fundo da própria coluna e o insumo para correção de pH é soda cáustica a 50% de concentração em NaOH.

A amônia recuperada no topo da coluna é transferida para o tanque de estocagem, onde o teor da solução amoniacal é obtida pela reposição de solução de amônia 20-25%.

A coluna apresenta uma diferença de temperatura decrescente do fundo para o topo, porque os vapores que sobem entram em contato com o líquido descendente cedendo-lhe calor. O aquecimento da alimentação libera NH_3 existente, o qual sobe junto com o vapor. Parte do vapor ascendente se condensa, com a troca de calor, saindo com o produto de fundo, e o restante sai pelo topo, junto com a amônia, condensando-se como solução de amônia.

O fundo da coluna tem calor fornecido por vapor vivo, cuja vazão é relacionada com a vazão de alimentação. O produto de fundo sofre uma expansão no vaso *flash* que opera com pressão negativa. Através dessa expansão o vapor gerado é retornado à coluna através de um ejetor instalado na linha do vapor vivo proveniente da área de utilidades.

O diferencial de pressão da coluna é mantido constante pelo alívio dos não condensáveis direcionados para uma coluna absorvedora.

O nível de obstrução na coluna, devido à presença de sólidos na carga, é monitorado por instrumentos diferenciais de pressão nas seções de pratos e recheios.

5.2.2 Instrumentação e controle na coluna

Na região de topo da coluna as variáveis controladas são: vazão de alimentação, pH e temperatura de topo.

A vazão de alimentação é controlada em modo cascata com a vazão de vapor. O SP é dado na vazão de vapor e por um relacional de vazão é dado o SP da vazão de alimentação.

O controle de pH na alimentação da coluna é realizado por adição de soda cáustica, sendo a malha composta por um elemento sensor de pH e uma válvula de controle de alimentação de soda cáustica, cujo consumo é proporcional a quantidade do efluente a ser tratado e do teor de amônia. Assim, não se deve operar a vazão de soda em manual ou automático, mas sim em cascata com o elemento sensor de pH.

O controle de temperatura no topo é para mantê-la constante através de ajuste na vazão de refluxo do topo, sendo o controle em modo cascata. Elevando-se o SP da temperatura ocorre uma redução no SP da vazão de refluxo. Uma redução nessa temperatura implica no

aumento da concentração de amônia no topo. Uma elevação na vazão de refluxo de topo implica em redução de temperatura do prato onde entra, num aumento do refluxo interno dos pratos inferiores e menor subida de vapor. Assim, no prato de entrada do refluxo de topo sai menor quantidade de vapor d'água.

Na região de fundo da coluna as variáveis controladas são: vazão de vapor, pressão do vaso *flash*, nível do vaso coletor e do fundo da coluna.

A vazão de vapor é controlada com o seu SP definido pelo operador. A saída do controlador será o SP do controlador da vazão de alimentação.

A pressão do vaso *flash* é mantida sob controle através de medição de pressão e uma válvula de controle.

O nível do vaso coletor do produto de topo é mantido constante através da retirada para o tanque de estocagem. O nível de fundo da coluna é controlado por sua medição e uma válvula de controle entre o fundo e o vaso *flash*.

As demais variáveis listadas na Tabela 10 não são controladas, mas suas informações são imprescindíveis ao monitoramento operacional do equipamento.

O instrumento de pressão diferencial, por exemplo, tem a função de monitorar indiretamente o grau de obstrução que bandejas e recheios sofrem pela deposição e ou cristalização de materiais insolúveis e aderentes. Esta obstrução afeta negativamente a eficiência da coluna.

5.2.3 Ganhos previstos no estudo da coluna

A primeira instalação do analisador de amônia na FCCSA ocorreu há aproximadamente 15 anos atrás e desde então necessitou ser substituído uma vez, há aproximadamente 4 anos, cujo valor de compra, comparado a outros instrumentos em operação é bastante elevado. Obter um analisador virtual que consiga prever o teor de amônia no fundo da coluna tem, portanto os seguintes ganhos e motivações:

- Aumento na frequência da informação do teor de amônia no fundo da coluna, possibilitando otimizar o tempo em que o instrumento e o laboratório informam os seus resultados, reduzindo custos dos reagentes e mão de obra do operador e do analista.

- Estimativa *on line* do teor de amônia no período que o analisador apresentar problema de funcionamento.

- Possibilidade de manter somente o analisador virtual na ocasião futura de perda do instrumento, por fim de vida útil, devido o seu alto custo de compra.

Observação: nesse caso, do ponto de vista ambiental, o resultado do analisador virtual não será a informação enviada ao órgão. No processo de tratamento do efluente, após a coluna há uma unidade de recuperação de sólidos após a qual amostras são retiradas e levadas ao laboratório para análise do teor de amônia na freqüência determinada pelo órgão ambiental.

5.3 Passo 2: Levantamento e tratamento dos dados da coluna

5.3.1 Levantamento dos dados

Devido o sistema de supervisão e controle para a coluna ser o SDCD, o mesmo recurso utilizado para a aquisição de dados da caldeira não pôde ser usado para a coluna. A aquisição dos dados da coluna foi através de um dos sistemas para a monitoração das unidades da FCCSA que é o PI System da OSI Software, o qual arquiva dados das variáveis que possuem medição. Uma das funcionalidades deste sistema é o *PI-DataLink* que é capaz de estabelecer uma conexão direta entre o PI System e o Microsoft Excel, possibilitando a análise de dados da planta em tempo real ou pelo seu histórico. Uma planilha em Excel foi feita listando-se todas as variáveis escolhidas para o estudo.

A variável que definiu o tempo de amostragem foi a própria variável de saída, o teor de amônia no fundo da coluna. Isso porque o analisador, ao informar um resultado, passa por processos de limpeza e calibração para a informação do resultado seguinte, consumindo nessas etapas reagentes. O tempo atualmente definido para essa operação é de 20 min. Assim, todas as demais variáveis tiveram seus valores obtidos no mesmo instante de tempo que o teor de amônia, um valor pontual a cada 20 min.

O levantamento dos dados da coluna ocorreu entre o período de 04/06/2016 até 27/07/2016 gerando 1903 padrões, dos quais 137 foram *outliers*, 1177 foram usados para estimação dos parâmetros dos modelos (treinamento, validação e teste) e 589 para verificação.

5.3.2 Tratamento dos dados

Os *outliers* nos dados da coluna foram avaliados e tratados empregando-se inicialmente as técnicas DMT e Hotteling. Adicionalmente, foram considerados *outliers* dados acima ou abaixo de três desvios-padrão em relação à média ou acima do limite superior de especificação de 14 pmm para o resultado do teor de amônia.

Dos 1903 padrões da coluna, 7,2%, foram encontrados como *outliers*, sendo que 24 foram retirados por DTM ou Hotteling e 113 pelos outros critérios acima mencionados. Os 24 dados por DTM e Hotelling foram similarmente distribuídos com as variáveis T topo, Refluxo e P flash. Dos 113 dados retirados pelos critérios, a maioria foi de resultados de NH_3 fundo (64%), seguido por T topo (17%), carga (8%), Razão carga vapor (7%) e as demais, Refluxo, T flash e pH carga (4%). Os resultados de NH_3 fundo encontrados acima de 14 ppm e retirados do banco de dados variavam de 14,1 ppm a 200 ppm, uma variação muito alta para pouca quantidade de dados.

Analisando a variável (NH_3 fundo) que possui o maior número de *outliers* tem-se que 9 resultados estão entre 50 ppm a 200 ppm e os demais, conforme Figura 30, entre 0 ppm a 50 ppm.

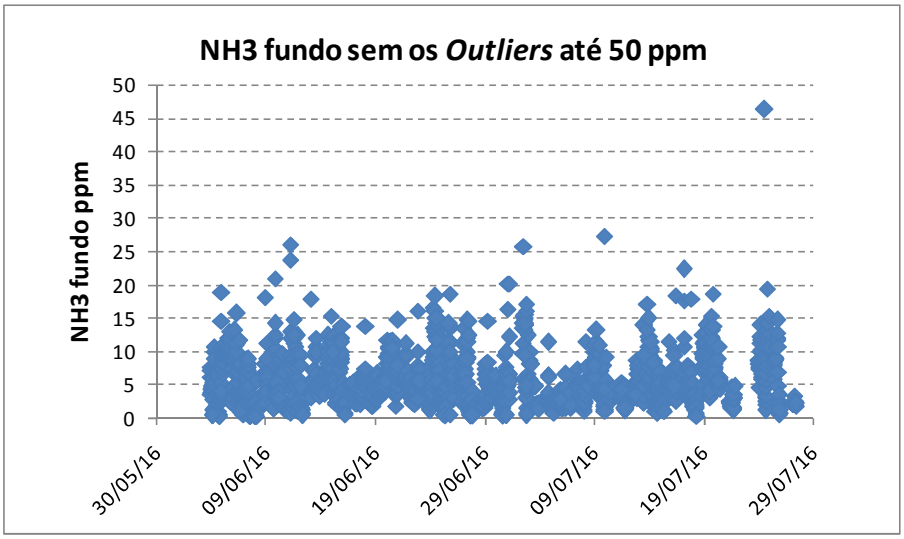


Figura 30 Resultados de NH3 Fundo com os outliers

Após a retirada dos *outliers* acima de 14 ppm de NH_3 fundo a distribuição dos dados desta variável ficam conforme Figura 31.

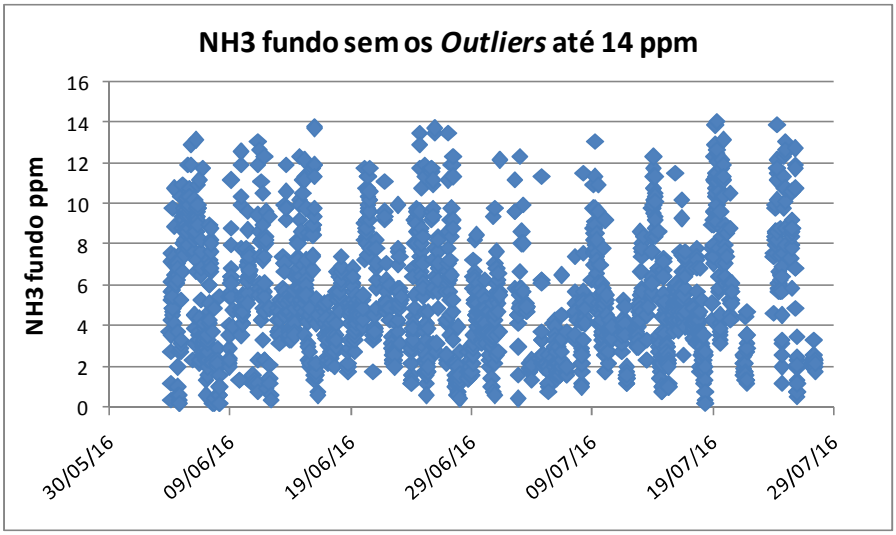


Figura 31 Resultados de NH3 Fundo sem os outliers

Nas Figuras 32 e 33 e Tabela 11 são apresentados os dados da variável de saída utilizados na construção dos modelos de RNA e RLM da coluna. Os dados usados na estimação dos parâmetros da RLM foram os mesmos empregados na estimação dos parâmetros da RNA. Distribuição dos padrões da variável estudada no treinamento RNA: teor de amônia no fundo da coluna. Os dados de estimação estão designados nas legendas das Figuras por EST e os de verificação por VER.

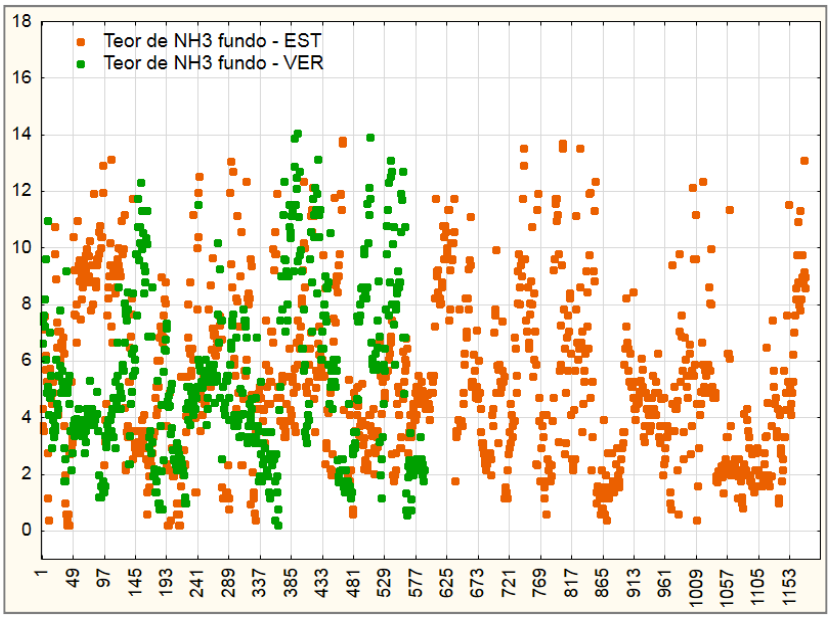


Figura 32 Dados de NH_3 no fundo da coluna na estimação e verificação da RNA e RLM

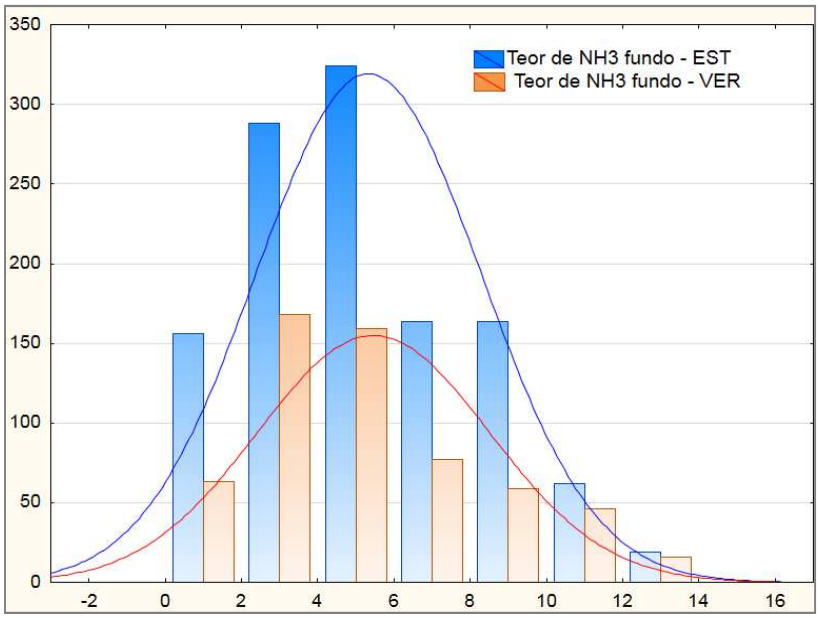


Figura 33 Distribuição dos dados NH_3 no fundo da coluna na estimação e verificação da RNA e RLM

Tabela 11 Estatística dos dados de estimação e de verificação da coluna

Variável (unidade da medição)	Precisão dos instrumentos (%)	Dados da estimação (N=1177)				Dados da verificação (N=589)			
		Média	Mínimo	Máximo	Desvio padrão	Média	Mínimo	Máximo	Desvio padrão
NH_3 fundo (ppm)	2,00	5,29	0,20	13,82	2,94	5,42	0,20	14,06	2,97
Carga (Kg/h)	1,00	44.090,30	32.839,53	49.876,66	3.606,05	44.530,17	32.839,53	49.876,66	3.130,22
Vapor (Kg/h)	1,00	3.622,90	3.097,34	3.959,09	202,02	3.652,54	3.097,34	3.980,34	178,32
Vazão arraste (Kg/h)	0,50	5.143,18	4.274,44	5.665,77	297,17	5.152,76	4.274,44	5.665,77	263,67
Razão carga Vapor		12,16	9,65	14,00	0,56	12,20	9,65	14,00	0,53
Arraste		1,42	1,32	1,51	0,03	1,41	1,32	1,51	0,03
Soda (L/h)	0,50	186,47	0,22	605,02	119,43	191,76	0,22	605,02	113,77
Refluxo (L/h)	0,50	1.035,98	11,00	3.619,41	568,63	1.858,28	11,00	4.121,94	801,02
T carga (°C)	1,00	63,72	53,11	79,16	4,94	64,99	48,82	79,16	4,78
T topo (°C)	1,00	93,14	91,18	96,08	0,55	93,65	91,18	96,08	0,66
pH carga	0,50	11,74	10,95	12,56	0,22	11,85	10,95	12,56	0,21
T fundo (°C)	1,00	103,35	101,94	105,78	0,92	104,44	101,94	105,90	1,06
T *flash* (°C)	1,00	86,56	82,72	92,92	1,34	87,89	82,72	92,92	1,53
Tsaída Trocador (°C)	1,00	66,38	51,61	73,60	3,54	67,58	51,61	77,66	3,42
P *flash* (Kgf/cm²)	0,50	- 0,44	- 0,60	- 0,21	0,08	- 0,32	- 0,60	- 0,20	0,09
Nível fundo (%)	0,50	34,90	24,10	44,78	4,52	30,07	24,10	44,78	4,38
DP recheio (mm Água)	0,50	315,22	130,72	572,31	75,83	360,11	130,72	591,15	79,15

5.4 Passo 3: definição das variáveis de entrada para o modelo da coluna

Redução da dimensionalidade é estudada a seguir usando recursos e informações das técnicas PCA, matriz de correlação e multicolinearidade.

5.4.1 Redução de dimensionalidade por PCA

Essa técnica busca reduzir o número de variáveis por medição do poder de cada variável no seu respectivo componente. A técnica é entendida como um método de transformação das variáveis originais em novas variáveis não correlacionadas. Cada componente principal é uma combinação linear das variáveis originais.

Os componentes principais são ordenados em ordem decrescente de sua variância, ou seja, o componente principal que contém mais informação é o primeiro, sendo o ultimo aquele componente principal com menos informação.

A importância da variável é uma informação útil na análise de PCA. Ela mede o quão bem uma variável é representada pelos componentes principais e esse valor é conhecido como potência, ou "*Power*", como se vê na Figura 34. Esse valor varia de 0 a 1. Para um modelo que tenha um número suficiente de componentes, as demais variáveis não serão bem representadas e mais propensas de serem insignificantes.

Na Figura 34, o histograma mostra que R^2X e Q^2 melhoram, ou seja, tendem a 1, quanto mais componentes principais são dados. Essa análise gráfica, cujas Equações são apresentadas no Apêndice A.1, mostra que este modelo tem um número suficiente de componentes principais, no caso 4, capazes de modelar suficientemente as variáveis, bastando o complemento de mais algumas variáveis para melhorar a construção do modelo.

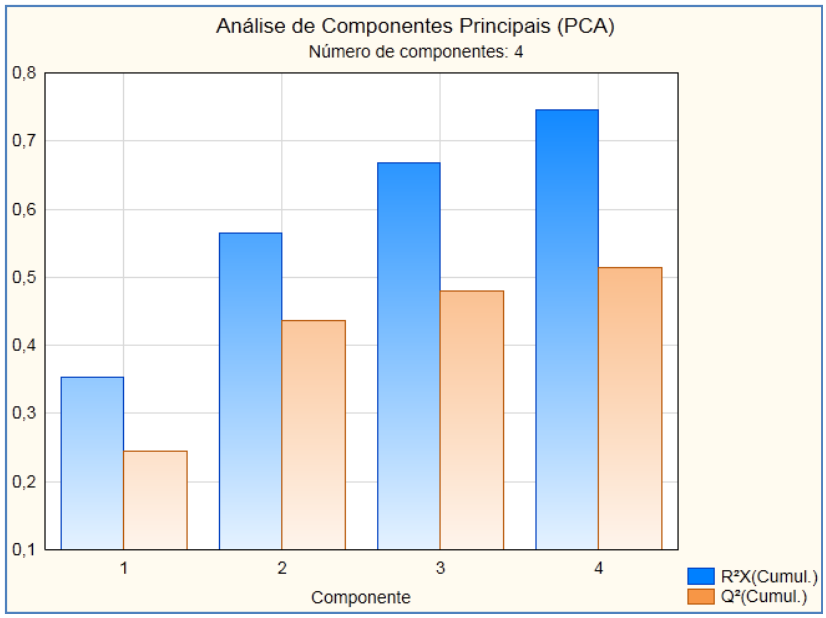


Figura 34 Análise dos componentes principais na coluna

Teoricamente o que se deve observar é que, quanto mais componentes principais o modelo apresentar, maior será o poder de modelagem das variáveis individuais. Examinando a sequência de histogramas gerado, Figura 35, o que se observa, em particular, é que as variáveis Carga, Vazão de arraste, Vapor e T fundo, são as mais importantes e que as variáveis menos importantes são: T topo, Nível fundo e T saída trocador, sendo portanto as variáveis que ao serem retiradas no modelo RNA não causam prejuízo para o mesmo.

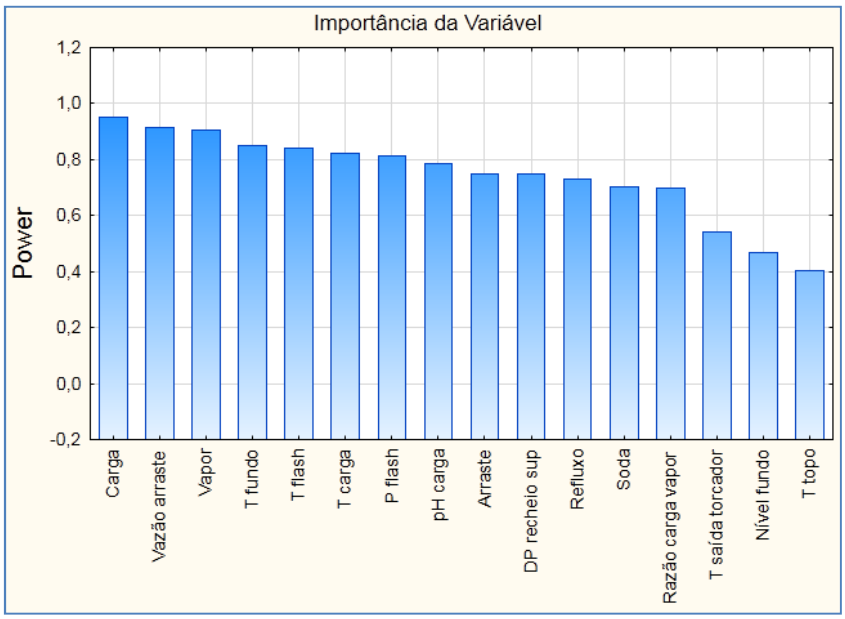


Figura 35 Importância das variáveis da coluna por PCA

5.4.2 Redução de dimensionalidade por matriz de correlação e multicolinearidade

A matriz de correlação mostra quais variáveis possuem relação insignificante com o teor de NH_3 no fundo da coluna e que, portanto, poderão ser retiradas e a multicolinearidade apresenta quais variáveis têm o mesmo nível de informação, podendo deixar que apenas uma fique no modelo.

A matriz de correlação, Tabela 12, apresenta que as variáveis T topo, T fundo e Nível fundo têm correlação insignificante com o teor de amônia no fundo. Correlações significativas, p valor > 0,05, estão marcadas em vermelho.

A multicolinearidade, Tabela 12, mostra que três variáveis possuem alta correlação ou mesmo nível de informação e que, portanto duas destas variáveis podem se removidas por terem correlação maior ou igual a 0,85 com a variável Carga. São elas as variáveis vapor e vazão de arraste. O resultado da Multicolinearidade também ajuda a explicar o resultado da PCA, visto que as três variáveis mais correlacionadas são as mesmas a apresentarem os três componentes principais.

Tabela 12 Matriz de correlação das variáveis da coluna

	Média	NH_3 fundo	Carga	Vapor	Vazão arraste	Razão carga vapor	Arraste	Soda	Refluxo	T carga	T topo	pH carga	T fundo	T *flash*	T saída trocador	P *flash*	Nível fundo	DP recheio sup
NH_3 fundo (ppm)	5,29	1,00	0,53	0,33	0,35	0,56	0,07	0,18	- 0,30	- 0,22	- 0,06	- 0,14	- 0,01	0,08	0,16	- 0,11	- 0,02	0,26
Carga (Kg/h)	44.090,30	0,53	1,00	0,85	0,85	0,77	0,07	0,36	- 0,31	- 0,24	- 0,20	- 0,20	0,21	- 0,06	0,22	- 0,22	0,09	0,71
Vapor (Kg/h)	3.622,90	0,33	0,85	1,00	0,92	0,33	- 0,11	0,37	- 0,28	- 0,34	- 0,24	- 0,14	0,27	- 0,21	0,30	- 0,38	0,06	0,76
Vazão arraste (Kg/h)	5.143,18	0,35	0,85	0,92	1,00	0,42	0,28	0,41	- 0,50	- 0,46	- 0,19	- 0,21	0,04	- 0,42	0,36	- 0,44	0,24	0,74
Razão carga Vapor	12,16	0,56	0,77	0,33	0,42	1,00	0,25	0,20	- 0,20	- 0,02	- 0,08	- 0,19	0,05	0,16	0,02	0,07	0,09	0,35
Arraste	1,42	0,07	0,07	- 0,11	0,28	0,25	1,00	0,14	- 0,60	- 0,34	0,12	- 0,19	- 0,59	- 0,58	0,18	- 0,20	0,45	0,02
Soda (L/h)	186,47	0,18	0,36	0,37	0,41	0,20	0,14	1,00	- 0,39	- 0,19	- 0,07	- 0,54	0,03	- 0,14	0,19	- 0,13	0,27	0,26
Refluxo (L/h)	1.035,98	- 0,30	- 0,31	- 0,28	- 0,50	- 0,20	- 0,60	- 0,39	1,00	0,64	- 0,01	0,29	0,53	0,65	- 0,48	0,49	- 0,42	- 0,31
T carga (°C)	63,72	- 0,22	- 0,24	- 0,34	- 0,46	- 0,02	- 0,34	- 0,19	0,64	1,00	0,10	0,13	0,64	0,74	- 0,50	0,84	- 0,37	- 0,24
T topo (°C)	93,14	- 0,06	- 0,20	- 0,24	- 0,19	- 0,08	0,12	- 0,07	- 0,01	0,10	1,00	0,09	- 0,09	- 0,01	- 0,27	0,15	0,10	- 0,24
pH carga	11,74	- 0,14	- 0,20	- 0,14	- 0,21	- 0,19	- 0,19	- 0,54	0,29	0,13	0,09	1,00	0,15	0,14	- 0,23	0,14	- 0,25	0,00
T fundo (°C)	103,35	- 0,01	0,21	0,27	0,04	0,05	- 0,59	0,03	0,53	0,64	- 0,09	0,15	1,00	0,71	- 0,28	0,59	- 0,42	0,22
T *flash* (°C)	86,56	0,08	- 0,06	- 0,21	- 0,42	0,16	- 0,58	- 0,14	0,65	0,74	- 0,01	0,14	0,71	1,00	- 0,43	0,74	- 0,34	- 0,20
T saída Trocador (°C)	66,38	0,16	0,22	0,30	0,36	0,02	0,18	0,19	- 0,48	- 0,50	- 0,27	- 0,23	- 0,28	- 0,43	1,00	- 0,48	0,19	0,35
P *flash* (Kgf/cm²)	- 0,44	- 0,11	- 0,22	- 0,38	- 0,44	0,07	- 0,20	- 0,13	0,49	0,84	0,15	0,14	0,59	0,74	- 0,48	1,00	- 0,31	- 0,23
Nível fundo (%)	34,90	- 0,02	0,09	0,06	0,24	0,09	0,45	0,27	- 0,42	- 0,37	0,10	- 0,25	- 0,42	- 0,34	0,19	- 0,31	1,00	0,03
DP recheio (mm Água)	315,22	0,26	0,71	0,76	0,74	0,35	0,02	0,26	- 0,31	- 0,24	- 0,24	0,00	0,22	- 0,20	0,35	- 0,23	0,03	1,00

5.5 Passo 4: Projeto dos modelos para a coluna

5.5.1 Regressão na coluna: Treinamento das redes MLP

Essa etapa consiste da estimação dos parâmetros da rede desenvolvida para a coluna em que 70% dos dados são usados para treinamento, 15% para validação e o restante dos dados (15%) para teste da rede.

Devido às características mais complicadas do processo, os resultados obtidos pelo modelo RLM para a coluna foram insatisfatórios, exibindo coeficiente de correlação R^2 = 0,452. Por isso, passa-se diretamente ao modelo de redes neuronais neste item, sendo o detalhamento do modelo linear apresentado no Apêndice A.3.

O numero de variáveis de entrada apresentadas à rede foram reduzidas de 16 para 10 com base nos resultados do item anterior. Pela técnica de PCA as variáveis com baixa importância foram: T saída do trocador, Nível e T topo. Pela Matriz de Correlação as variáveis com insignificante relação com a variável de saída teor de amônia foram: T fundo, Nível e T topo, e por Multicolinearidade, as variáveis com o mesmo nível de informação da variável carga foram: vapor e vazão de arraste. Serão, portanto reduzidas do modelo RNA 6 variáveis: T saída do trocador, Nível, T topo, T fundo, vapor e vazão de arraste.

Essa etapa consiste em apresentar o melhor resultado obtido na regressão por redes MLP nas 10 variáveis de entrada para a inferência do teor de amônia.

No treinamento, outras redes retidas apresentaram coeficiente ligeiramente superior, sem, no entanto apresentar uma boa correlação na etapa de validação e por isso a rede MLP10-12-1 cujos coeficientes, desta e das demais redes retidas, são dados na Tabela 13 foi a escolhida no estudo.

A análise de sensibilidade, avaliada após o treinamento das redes para verificar se haveria variável desnecessária no modelo não indicou alguma com razão muito superior às demais e apresentou para todas as variáveis um valor maior que 1 na razão entre o erro da rede quando a variável está indisponível, substituída pela sua média, pelo erro total da rede quando os valores da variável são usados como entradas na rede, Tabela 14. Por essa análise, as 10 variáveis de entrada no modelo foram mantidas.

Tabela 13 Coeficientes de correlação no treinamento da coluna

Índice	Rede	Coeficiente do treinamento	Coeficiente do teste	Coeficiente da validação	Algoritmo de treinamento	Função de ativação da camada escondida	Função de ativação da camada de saída
1	MLP 10-9-1	0,887183	0,802583	0,788445	BFGS 89	Tang. hip.	Exponencial
2	MLP 10-5-1	0,825761	0,754436	0,765307	BFGS 101	Logística	Tang. hip.
3	MLP 10-11-1	0,843031	0,790732	0,780867	BFGS 71	Logística	Exponencial
4	MLP 10-5-1	0,842173	0,769249	0,761946	BFGS 93	Logística	Exponencial
5	MLP 10-10-1	0,865160	0,775409	0,770167	BFGS 114	Logística	Tang. hip.
6	MLP 10-13-1	0,861869	0,790638	0,759870	BFGS 97	Tang. hip.	Identidade
7	MLP 10-13-1	0,879803	0,800593	0,780376	BFGS 121	Logística	Logística
8	MLP 10-9-1	0,883579	0,796121	0,782416	BFGS 104	Tang. hip.	Logística
9	MLP 10-8-1	0,851397	0,787926	0,764321	BFGS 107	Exponencial	Logística
10	MLP 10-13-1	0,872518	0,786694	0,771371	BFGS 102	Logística	Logística
11	MLP 10-8-1	0,866929	0,803197	0,771446	BFGS 133	Logística	Exponencial
12	MLP 10-12-1	0,834799	0,769304	0,765508	BFGS 49	Tang. hip.	Exponencial
13	MLP 10-10-1	0,861339	0,768167	0,771697	BFGS 105	Logística	Identidade
14	MLP 10-8-1	0,863023	0,817153	0,762687	BFGS 153	Exponencial	Exponencial
15	MLP 10-11-1	0,831875	0,780010	0,766311	BFGS 84	Exponencial	Exponencial

Tabela 14 Análise de sensibilidade nas variáveis da coluna

P *flash*	T carga	Razão carga vapor	Carga	T *flash*	Refluxo	DP recheio sup	Arraste	Soda	pH carga
2,631	2,397	1,971	1,429	1,811	1,512	1,247	1,239	1,199	1,075

Nessa rede, o número de padrões apresentados foram 1177, conforme determinado em materiais e métodos, o algoritmo de treinamento foi o BFGS, a função erro foi a SOS, as funções de ativação utilizadas nas camadas escondida e de saída foi a Exponencial.

A RNA foi uma rede MLP 10-12-1 gerando, portanto 132 pesos e 13 biases, indicando que os 1177 padrões do treinamento e embora não atenda planamente ao critério da relação de 10 a 30 vezes a quantidade de pesos, está bem próximo.

O coeficiente de correlação R^2 obtido foi igual a 0,834 demonstrando boa capacidade da técnica com o objetivo de inferência e satisfatório visto que essa é a ordem de grandeza obtida no trabalho de Pani *et al*. (2015), para uma coluna debutanizadora, para a inferência do teor de butano no produto de fundo.

Uma apresentação dos resultados e interpretação dos mesmos segue pelas Figuras, 36, 37, 38 e 39.

A disposição dos resultados preditos e observados no mesmo gráfico, Figura 36, mostra que a inferência é mais precisa para valores de NH_3 fundo até 8 ppm. Para valores de 9 a 14 ppm há uma redução na capacidade da rede em explicar o fenômeno. A hipótese mais provável desta diferença de precisão é devido à diferença na quantidade de dados apresentados à rede. Como mostrados na Tabela 15, 79% dos dados da variável de entrada estão entre 0 ppm a 8 ppm.

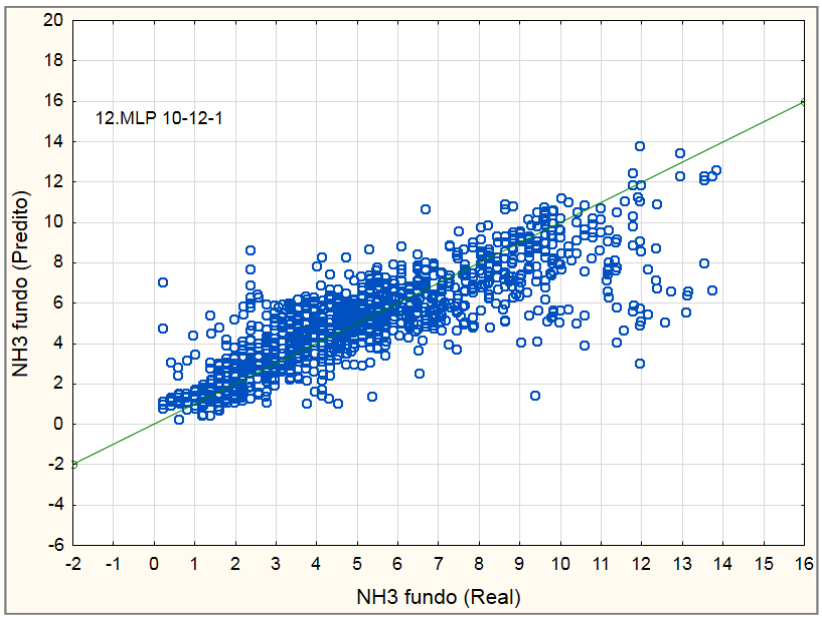


Figura 36 Dados reais de NH_3 fundo em relação aos preditos no treinamento

Tabela 15 Tabela de frequência dos dados da variável de entrada da coluna

Categoria	N	N acumulado	%	% acumulado	% de todos os casos	% de todos os casos, acumulado
-2 < x <= 0	0	0	0,00000	0,0000	0,00000	0,0000
-0 < x <= 2	156	156	13,25404	13,2540	13,25404	13,2540
2 < x <= 4	288	444	24,46899	37,7230	24,46899	37,7230
4 < x <= 6	324	768	27,52761	65,2506	27,52761	65,2506
6 < x <= 8	164	932	13,93373	79,1844	13,93373	79,1844
8 < x <= 10	164	1096	13,93373	93,1181	13,93373	93,1181
10 < x <= 12	62	1158	5,26763	98,3857	5,26763	98,3857
12 < x <= 14	19	1177	1,61427	100,0000	1,61427	100,0000

Também a Figura 37 dá a explicação do efeito da quantidade de dados na precisão do modelo. Na análise do gráfico de resíduos, o pior desempenho é para os mais altos teores de amônia. Lembrando que o resíduo é a diferença entre os valores preditos e observados e que, portanto quanto mais próximos de 0 (zero) melhor.

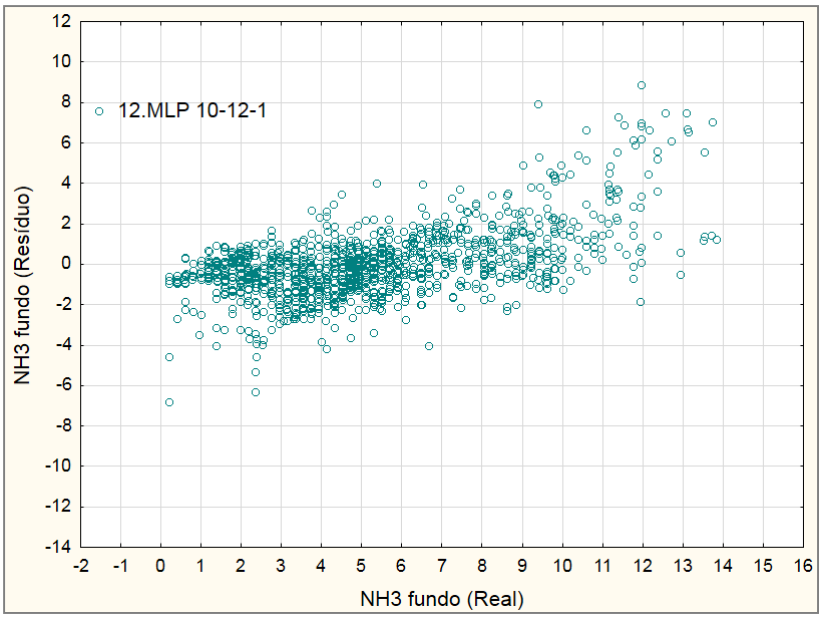


Figura 37 Resíduos da NH_3 fundo em relação aos dados reais de treinamento

Na Figura 38 a disposição do resultado de resíduos. A distribuição dos dados segue uma curva normal, com a maior frequência de valores iguais ou próximos a 0 (zero), mais ou menos 1.

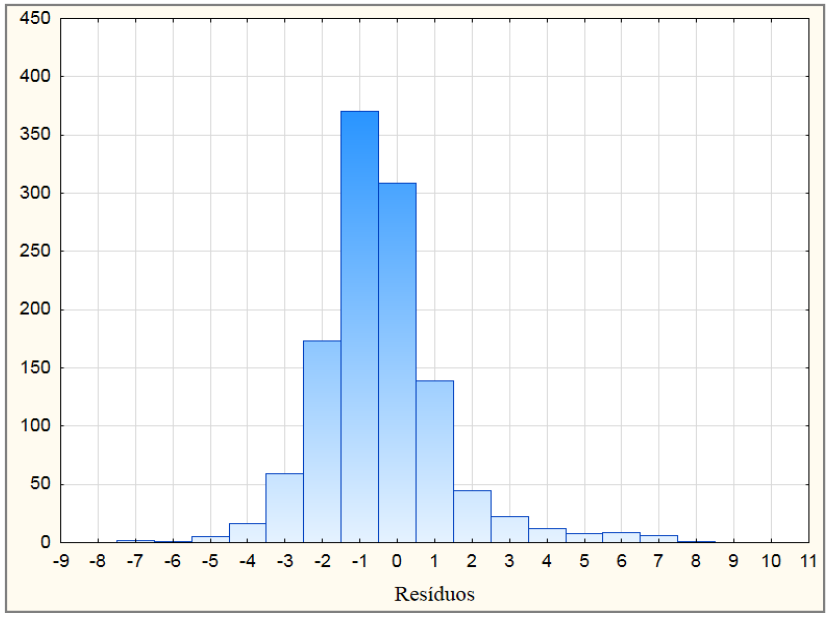


Figura 38 Dispersão dos resíduos de NH_3 fundo da coluna no treinamento

A Figura 39 apresenta a disposição dos resultados preditos e observados e embora haja a presença de erro, como apresentado pelo gráfico de resíduos, é verificado o acompanhamento da tendência dos valores, ou seja, sempre que o valor observado é aumentado ou reduzido essa mesma curva é dada pelos valores preditos. Isso demonstra a boa capacidade do modelo em representar o processo para a maior parte dos dados.

Os pontos de maiores desvios entre o real e o predito estão destacados na Figura 39. A ordem de grandeza do erro nessas regiões destacadas está entre 30% e 50%.

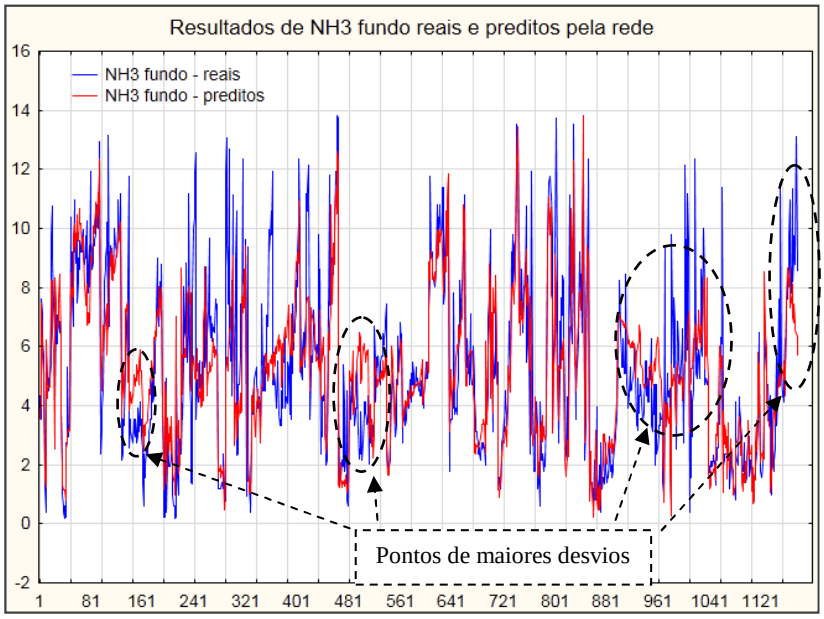


Figura 39 Dados reais e preditos de NH_3 fundo no treinamento

5.5.2 Verificação da MLP do treinamento da coluna

Na verificação novos padrões são apresentados à rede MPL 10-12-1 conseguida no treinamento e os coeficientes de correlação foram analisados.

O número de padrões nessa etapa foram 589 (1/3 dos dados totais disponíveis);

O coeficiente de correlação R^2 foi de 0,690, um valor abaixo do encontrado no treinamento (0,834). Uma hipótese para esse menor coeficiente é a menor quantidade de dados (1/3 do total) utilizados na verificação.

O resultado de R^2 igual a 0,690 na etapa de verificação pode-se considerar satisfatório, visto ser bastante semelhante ao que foi encontrado pelo trabalho de Pani *et al.* (2015), citado nessa dissertação, para a medição de butano no fundo de uma debutanizadora, em que R^2 foi de 0,855 no treinamento e R^2 igual a 0,733 na verificação.

As Figuras 40, 41, 42 e 43 apresentam os resultados sendo a interpretação similar à dada aos mesmos gráficos da etapa de treinamento, ou seja, o modelo é mais preciso para menores valores da variável de entrada, a maior parte dos resíduos se situam perto de 0 (zero) e os dados preditos e observados, embora apresente diferenças é notado, na maior parte dos dados, que seguem a mesma tendência, quando em redução ou quando em subida.

Os pontos de falhas do modelo nessa etapa da verificação, onde ocorrem os maiores desvios entre o real e o predito, estão destacados na Figura 43. A ordem de grandeza do erro nessas regiões destacadas é maior que na etapa de treinamento e está entre 50% e 100%.

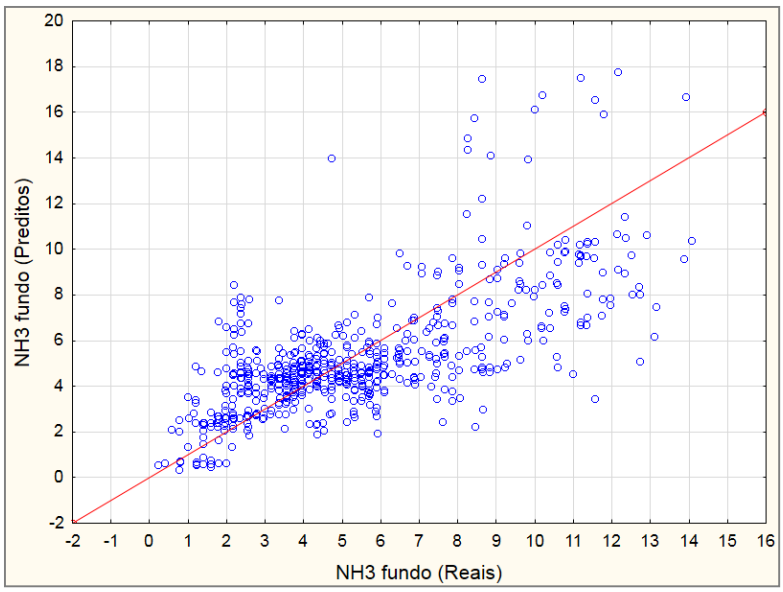


Figura 40 Dados reais de NH_3 fundo em relação aos preditos na verificação

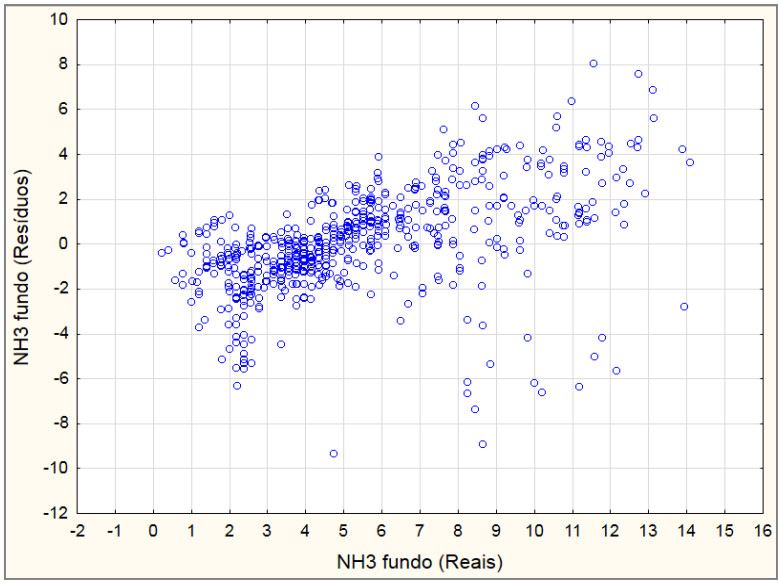


Figura 41 Resíduos de NH_3 fundo em relação aos dados reais de verificação

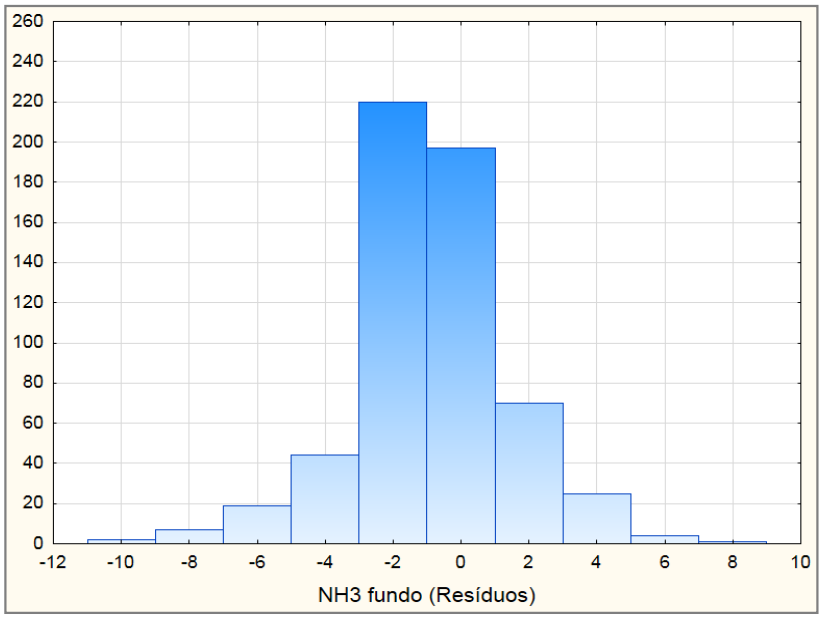


Figura 42 Dispersão dos resíduos de NH_3 fundo na verificação

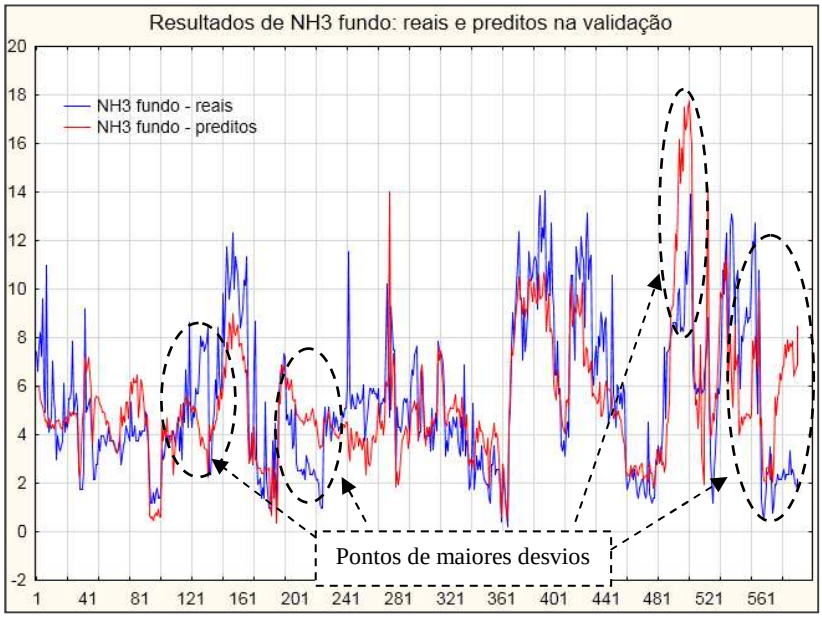


Figura 43 Dados reais e preditos de NH_3 fundo na verificação

5.5.3 Retreinamento da rede MLP

Realizando o passo 5 da metodologia, foram usados todos os dados disponíveis para a construção de um novo modelo.

O coeficiente de correlação dessa avaliação foi de 0,86, um valor próximo ao obtido na etapa de treinamento com uma quantidade menor de dados Essa avaliação confere a boa

capacidade de correlação das variáveis na predição daquela de interesse. Na Figura 44 o resultado dessa avaliação apresentando os valores reais e preditos da variável estudada e na Figura 45 o gráfico de resíduos que tem o mesmo comportamento do treinamento anteriormente realizado em que a diferença entre o real e predito aumenta para valores maiores da variável de saída.

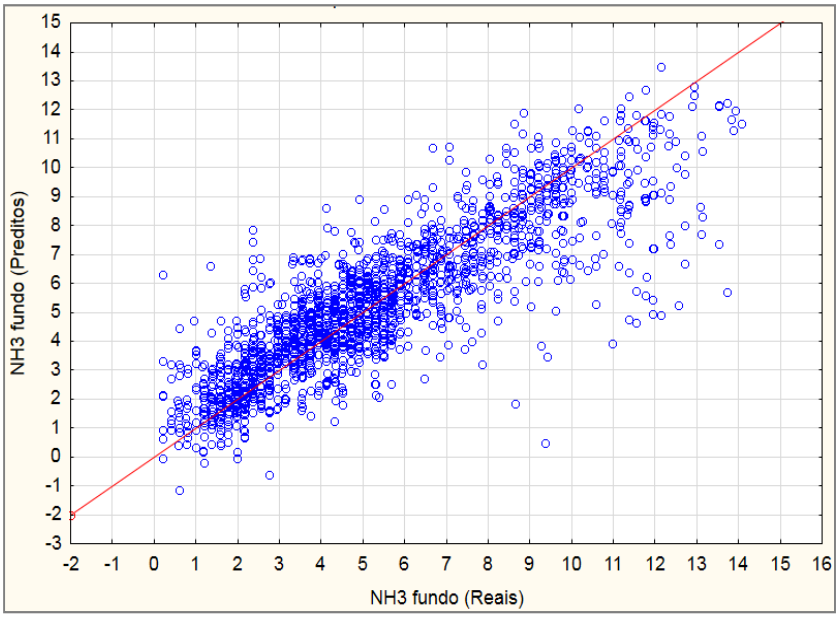


Figura 44 Dados reais e preditos de NH_3 fundo no retreinamento

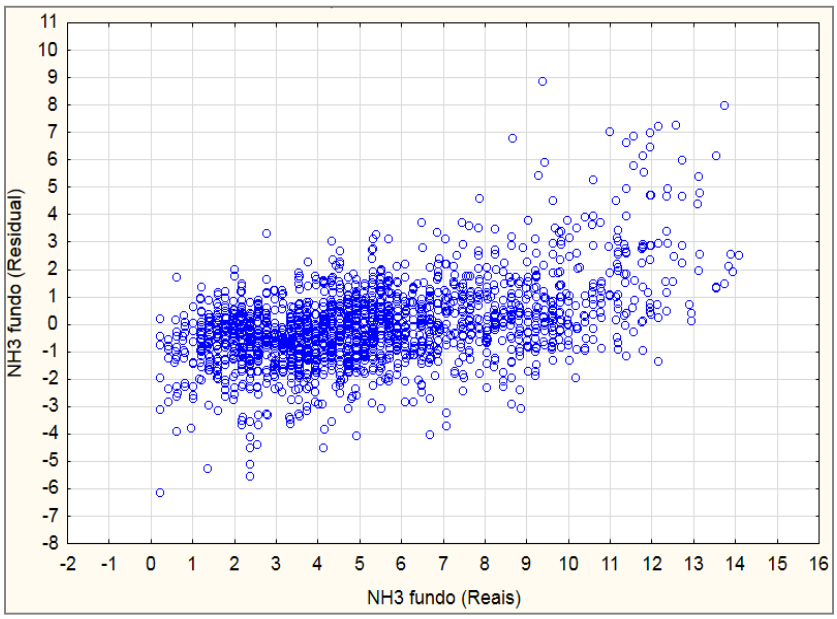


Figura 45 Resíduos de NH3 fundo em relação aos dados reais no retreinamento

5.5.4 Conclusões sobre a modelagem na coluna

O coeficiente de correlação da modelagem por RNA conseguiu uma boa superação daquele encontrado por RLM, passando o R^2 de 0,452 para 0,834. Na verificação de ambos os modelos, com novos dados sendo apresentados, também se constata a maior capacidade do modelo RNA, em que o R^2 foi de 0,69 e na RLM 0,45 (Apêndice A.3).

O valor do coeficiente de correlação indica que o modelo RNA obtido na coluna tem boa capacidade de previsão da variável de interesse, e isso possibilita seu uso tanto como *Soft sensor*, ou para fins de controle mais avançado do processo. Para a aplicação industrial a que se propõe o estudo, a boa qualidade da inferência é necessária, devido se tratar de uma variável monitorada para a qualidade de lançamento de efluente, cujo limite superior de especificação não deve ser atingido.

6 CONCLUSÕES

Sobre a transportabilidade de modelos preditivos baseados em redes neuronais:

Comparativamente a modelos fenomenológicos, o modelo de RNA não demanda, para cada aplicação, o emprego de equações tão específicas como são as equações de balanço de massa e energia e as equações constitutivas dos modelos fundamentais, baseando-se em um modelo "plástico" de elementos (os neurônios), cujo número e funções de ativação simples podem ser adaptados. O fato de as redes serem considerados aproximadores universais para mapeamentos não lineares (Funahashi, 1989) também assegura a possibilidade de usar as redes para diferentes aplicações.

Pode-se afirmar que conceitos de "transportabilidade" de modelos baseados em redes neuronais está presente em trabalhos desde fins do século passado (Aström & McAvoy, 1992). Em alguns casos, a transportabilidade é relativa a diferentes pontos operacionais (Walker *et al*., 1999). Noutros, como no próprio trabalho de Aström & McAvoy (1992), ao mesmo processo em diferentes plantas.

Nesta dissertação, estendeu-se a qualificação de transportável para equipamentos diferentes. Mostrou-se que, para equipamentos diferentes de uma indústria química, uma metodologia padronizada proposta pode ser usada no desenvolvimento dos modelos, fornecendo evidências da pertinência da proposição. Essa metodologia, que agrega ferramentas estatísticas ao projeto de redes neuronais, pode ser classificada como uma abordagem de aprendizagem de máquina. Tais desenvolvimentos, no contexto atual de riqueza de dados, devido á existência de sistemas PI, etc., confere originalidade e contribuição à proposta e resultados obtidos.

Sobre aplicações a uma indústria química

Na caldeira o coeficiente de correlação do modelo RNA (0,99) superou o obtido por RLM (0,94) e esse elevado valor, próximo a um, dá considerável garantia de sua aplicação, mesmo sendo um equipamento que envolve aspectos de segurança, tanto para inferência como para aplicação a modelos avançados.

Na coluna, o coeficiente de correlação R^2 obtido por RNA foi igual a 0,834 superou o obtido por RLM que foi igual a 0,452, demonstrando um ganho considerável da técnica, sendo esta pela primeira vez empregada na FCCSA para uso em predição de uma variável.

O resultado desse estudo possibilitou ampliar o conhecimento dos sistemas avaliados. Com isso novas otimizações passam a ser possíveis bem como os ganhos previstos nesse estudo, os quais têm um aspecto em redução de custos relacionados à substituição de um instrumento e aumento na quantidade de informações do processo.

Sugestões para trabalhos futuros

A motivação para trabalhos futuros é dar continuidade ao estudo desenvolvido nessa dissertação visto as possibilidades de ganhos da aplicação da técnica de RNA na indústria. Dentre as principais sugestões pode-se citar:

- Testar o uso das componentes principais como entradas das redes neuronais;
- Implementar as redes desenvolvidas no sistema supervisório e usá-las como sensores virtuais;
- Usar os modelos desenvolvidos também para otimização e controle dos processos;
- Incorporar à metodologia desenvolvida, etapas de avaliação do modelo de rede e recalibração do mesmo, se necessário;
- Modelar outros equipamentos da planta usando a metodologia proposta;
- Modelar a planta completa usando redes neuronais, através dos novos tipos de redes “deep learning”.

REFERÊNCIAS BIBLIOGRÁFICAS

Åström, K.J.; McAvoy, T. J. Intelligent control. Full papers, vol 2, n° 3, p. 115-127, 1992.

Baughman, D. R.; Liu, Y. A. Neural Networks in Bioprocessing and Chemical Engineering. Academic Press, INC. 1995.

Bega, E. A. Instrumentação aplicada ao controle de caldeiras. Interciência, 3ª edição, 2003.

De Souza Jr., M.B. Material do curso de Redes Neuronais aplicados a processos químicos e bioquímicos. Escola de Química, UFRJ, 2015.

De Souza Jr., M.B. Redes neuronais multicamadas aplicadas a modelagem e controle de processos químicos, Tese de doutorado, Programa de Pós Graduação de Engenharia da Universidade Federal do Rio de Janeiro, COPPE/UFRJ, 1993.

Ekonomou, L. Greek long-term energy consumption prediction using artificial neural networks. Energy, vol 35, p. 512-517, 2010.

Hair, J.F.J.; Anderson, R.E.; Tatham, R.L.; Black, W.C. Análise Multivariada de dados. Bookman, 2005.

Haykin, S. Neural Networks and Learning Machines, Third Edition, McMaster University, Hamilton, Ontario, Canada, 2009.

Himmelblau, D.M. Applications of Artificial Neural Networks in Chemical Engineering. Korean J. Chem. Eng, vol 17, p. 373-392, 2000.

Funahashi, K.I. On the Approximate Realization of Continuous Mappings by Neural Networks. Neural Networks, Vol. 2, pp. 183-192, 1989.

Hoskins, J.C.; Himmelblau, D.M. Artificial Neural Network of Knowledge Representation in Chemical Engineering. Computers Chem. Engng, vol 12, n° 9/10, p. 881-890, 1988.

Iliyas, S.A.; Elshafei, M.; Habib, M.A.; Adeniran, A.A. RBF neural network inferential sensor for process emission monitoring. Control Engineering Practice, vol 21, p. 962–970, 2013.

James, T. Smart Factories. Engineering & Tecnology, 2012.

Kadlec, P.; Gabrys, B.; Strandt, S.. Data-driven Soft Sensors in the process industry. Computers and Chemical Engineering, vol 33, p. 795-814, 2009.

Kljajic, M.; Gvozdenac, D; Vukmirovi, S. Use of Neural Networks for modeling and predicting boiler's operating performance. Energy, vol 45, p. 304–311, 2012.

Leonard, J; Kramer, M.A. Improvement of the backpropagation algorithm for training neural network. Computers Chem. Engng, vol 14, n° 3, p. 337-341, 1990.

Li, Defang. Perspective for smart factory in petrochemical industry. Computers and Chemical Engineering, vol 91, p. 136-148, 2016.

Louridas, P.; Ebert, C. Machine Learning. Software technology, p. 110-115, 2016.

Montgomery, D. C. Introdução ao controle estatístico da qualidade. LTC, 2004.

Pani, A.K.; Amin, K.G.; Mohanta, H.K. Soft sensing of product quality in the debutanizer column with principal component analysis and feed-forward artificial neural network. Alexandria Engineering Journal, 2016.

Prieto, A.; Prieto, B.; Ortigosa, E.M.; Ros, E.; Pelayo, F.; Ortega, J.; Rojas, I. Neural networks: An overview of early research, current frameworks and new challenges. Neurocomputing, vol 214, p. 242-268, 2016.

Qin, S. J. Process Data Analytics in the Era of Big Data. AIChE Journal, vol 60, nº 9, p. 3092-3100, 2014.

Ramli, N.M.; Hussain, M.A.; Jan, B.M.; Abdullah, B. Composition Prediction of a Debutanizer Column using Equation Based Artificial Neural Network Model. Neurocomputing, 131 59-76 (2014)

Salvatore, L ; De Souza Jr., M.B. ; Campos, M.C.M.M. Design and Implementation of a Neural Network Based Soft Sensor to Infer Sulfur Content in a Brazilian Based Diesel Hydrotreating Unit. Chemical Engineering Transactions, v. 17, p. 1389-1394, 2009.

Science Direct. Disponível em http://www.sciencedirect.com/science/search, Acesso em 23 de novembro de 2016.

Scrivano, R. Grande Indústria já adota 'fábrica inteligente'. Jornal O Globo, 2016. Acessado em 21 de novembro de 2016.

Shakil, M.; Elshafei, M.; Habib, M.A.; Maleki, F.A. Soft sensor for NOx and O2 using dynamic neural networks. Computers and Electrical Engineering, vol 35, p. 578–586, 2009.

Shu, Y.; Ming, L.; Cheng, F.; Zhang, Z.; Zhao, J. Abnormal situation management: Challenges and opportunities in the big data era. Computers and Chemical Engineering, vol 91, p. 104-113, 2016.

Statistica, Statistica Eletronic Manual – Statistica 12.7, Dell, USA, 2016.

Statsoft. Representante Dell. Material do curso de estatística. Fábrica Carioca de Catalisadores, FCCSA, 2016.

Tronci, S.; Baratti, R.; Servida, A. Monitoring pollutant emissions in a 4:8 MW power plant through neural network. Neurocomputing, vol 43, p. 3–15, 2002.

Valdman, A., Sistema de automação para monitoramento on-line de gases residuais e diagnóstico de uma caldeira operada com misturas diesel/biodiesel, Dissertação de Mestrado, Programa de Pós Graduação em Tecnologia de Processos Químicos e petroquímicos, Escola de Química, UFRJ, 2010.

Valdman, A.; De Souza JR., M. B. ; Folly, R. O. M. ; Valdman, B. Implementação em Sistema SCADA de um Sensor Virtual baseado em Redes Neuronais. Revista C & I. Controle & Instrumentação, v. 161, p. 28-31, 2010.

Venkatasubramanian, V. Drowning in data: Informatics and Modeling Challenges in a Data-Rich Networked World. AIChE Journal, vol 55, nº 1, p. 2-8, 2009.

Walker, D.M; Brown, R; Tufillaro, N.B. Constructing transportable behavioural models for nonlinear electronic devices. Physics Letter, vol. 255 p. 236–242, 1999

Zhou, H.; Zhao, J.P.; Zheng, L.G.; Wang, C.L.; Cen, K.F. Modeling NOx emissions from coal-fired utility boilers using support vector regression with ant colony optimization. Engineering Application of Artificial Intelligence, 2012.

APÊNDICE A.1

Técnicas auxiliares para tratamento de dados

As técnicas descritas aqui, usadas em complemento ao estudo de RNA, têm por finalidade examinar os dados, visto que ao preparar e avaliar os dados para uso em modelo de regressão deve-se considerar o tamanho da amostra, os *outliers*, a distribuição e a transformação dos dados bem como a possibilidade da redução na quantidade de variáveis que devem compor a construção do modelo. A abordagem para o desenvolvimento de modelo de regressão linear múltipla (RLM) é aqui incluída para prover um modelo mais simples para comparação com as RNA.

A.1 O Coeficiente de correlação R^2

O objetivo do coeficiente de correlação R, definido pela Equação A.1.1, é medir a relação entre duas variáveis aleatórias. Com esse resultado é possível avaliar a eficiência de uma determinada técnica bem como compará-la com outra.

Na prática, o R^2 fornece informações sobre a força e direção da correlação variando, portanto de –1 a +1, em que 0 (zero) indica que não há relação, 1 (um) indica uma relação positiva perfeita (quando uma variável aumenta em valor, a outra variável também aumenta em valor), e –1 (menos um) indica uma relação linear negativa perfeita (se uma variável aumenta em valor, a outra variável diminui em valor). Elevando o coeficiente de correlação ao quadrado, o valor resultante (R^2) irá representar a força, ou amplitude, da correlação, (Dell, 2016).

$$R = \left(\frac{\sum(x - \bar{x})(\hat{x} - \bar{\hat{x}})}{\sqrt{\sum(x - \bar{x})^2 \sum(\hat{x} - \bar{\hat{x}})^2}} \right) \qquad \text{(A.1.1)}$$

Em que:

x = valores reais

$\bar{x}$ = média dos valores reais

$\hat{x}$ = valores preditos

$\bar{\hat{x}}$ = média dos valores preditos

A.2 *O t-test*

O objetivo deste teste é avaliar as diferenças das médias entre dois grupos sendo esses independentes ou dependentes entre si.

Os valores de *p*, reportados no *t-test*, são comumente usados para testar uma "hipótese nula", que afirma em geral que não existe qualquer diferença entre dois grupos, ou que não há nenhuma correlação entre um par de dados.

O valor *p* reportado no *t-test* representa a probabilidade de erro envolvido em aceitar a hipótese de investigação sobre a existência de uma diferença. Ou seja, esta é a probabilidade de erro associado com a rejeição da hipótese de nenhuma diferença entre as duas categorias de observações na população quando, na verdade, a hipótese é verdadeira (Dell, 2016).

A.3 *Redução na quantidade de variáveis*

Três recursos foram utilizados nessa dissertação com o intuito de apresentar ao modelo somente as variáveis mais significantes na sua construção ou retirar alguma variável que possa ter o mesmo nível de informação que outra e com isso se tornando desnecessária. Os métodos utilizados foram a Análise de Sensibilidade, a Análise de componentes principais (PCA) e a Multicolinearidade.

A.4 Analise de sensibilidade

O cálculo da análise de sensibilidade é uma razão entre o erro da rede quando a variável está indisponível, substituída pela sua média, pelo erro total da rede quando os valores da variável são usados como entradas na rede, Dell (2016). Essa ferramenta possibilita avaliar se alguma variável contribui ou prejudica no modelo. Variáveis que possuem análise de sensibilidade igual a 1,0 ou menos não contribuem para o modelo e, portanto devem ser retiradas do treinamento da rede. Se a razão é maior que 1 a variável ajuda na solução do problema.

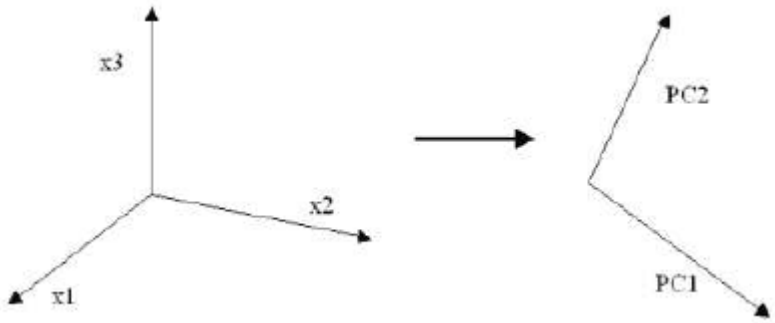
A.5 Análise de componentes principais

A técnica PCA tem como objetivo reduzir a dimensionalidade de um conjunto de variáveis mantendo a máxima variabilidade em termos da estrutura de variância-covariância através de um novo conjunto de sistemas de coordenadas que é menor em dimensão do que o número de variáveis originais. Cada coordenada no novo sistema é conhecido como um componente principal (CP), (Dell, 2016).

Na Equação A.1.2, c_{ij} são as constantes a serem determinadas para os componentes principais $z_{1,}, z_2, \dots, z_p$ que representarão as variáveis $x_{1,}, x_2, \dots, x_p$, (Montgomery, 2004).

$$\begin{cases} z_1 = c_{11}x_1 + c_{12}x_2 + \cdots + c_{1p}x_p \\ z_2 = c_{21}x_1 + c_{22}x_2 + \cdots + c_{2p}x_p \\ \quad . \\ \quad . \\ z_p = c_{p1}x_1 + c_{p2}x_2 + \cdots + c_{pp}x_p \end{cases} \tag{A.1.2}$$

Um diagrama esquemático de um modelo, de dois componentes principais, usado para representar três variáveis é mostrado na Figura 46.



Fonte: Dell (2016)

Figura 46 Representação da transformação de três variáveis em dois CP

A centralização de X como parte da etapa de pré-processamento, cria um ponto de origem para o novo sistema de coordenadas. A orientação de PC1 é escolhida de maneira a capturar a máxima variabilidade nos dados. Cada observação do conjunto de dados originais $x_i = (x_1, x_2)$ é projetada em PC1. A esta projeção dá-se o nome de *score*, representado por t_i , sendo esse valor o que representará o novo conjunto de dados $t = (t_i, \dots, t_N)$ (Dell, 2016).

Para modelos multidimensionais, o conjunto de componentes principais forma um sistema ortogonal com o ponto de origem dado pela média dos dados originais e orientação definida por fatores (designado como *loading*) (Dell, 2016).

Duas medidas úteis para diagnóstico do modelo de componentes principais são a Fração da variação explicada R^2X e a Fração da variação prevista Q^2X (Dell, 2016).

R^2X é medido em toda a amostragem dado pela Equação A.1.3:

$$R^2X = 1 - \frac{somda\ dos\ quadrados\ dos\ resíduos}{soma\ do\ quadrado} \qquad \text{(A.1.3)}$$

Para um modelo de componentes principais com *A* componentes, o R^2X acumulado é dado pela Equação A.1.4:

$$R^2X\ (Acum) = 1 - \frac{\sum_{i,j}\left(x_{i,j} - \hat{x}_{i,j}^{A}\right)^2}{\sum_{i,j} x_{i,j}^2} \qquad \text{(A.1.4)}$$

Em que:

$x_{i,j}^{A}$ é o valor predito do modelo de componentes principais, com *A* componentes, para a *i-ésima* observação da *j-ésima* variável.

Quanto mais significante for a componente principal, maior será R^2X. Um modelo com suficientes componentes principais tem o valor R^2X próximo de 1,0.

Q^2 dado pela Equação A.1.5 é matematicamente idêntico a R^2X, diferindo-se que se trata de uma medição usando amostras que não foram usadas na construção do modelo de componentes principais.

$$Q^2 = 1 - \frac{soma\ dos\ quadrados\ dos\ resíduos\ preditos}{soma\ dos\ quadrados\ dos\ resíduos\ para\ o\ componente\ anterior} \qquad \text{(A.1.5)}$$

Para um modelo com A componentes principais Q^2 acumulado é definido pela Equação A.1.6:

$$Q^2(Acum) = 1 - \prod_{i=1}^{A} \frac{soma\ dos\ quadrados\ dos\ resíduos\ preditos}{soma\ dos\ quadrados\ dos\ resíduos\ para\ o\ componente\ anterior} \quad (A.1.6)$$

A.6 *Multicolinearidade*

Entende-se por multicolinearidade a melhor correlação linear entre variáveis independentes. Estimativas dos coeficientes ficam indeterminados na presença de multicolinearidade e seus erros-padrões são infinitos. Para multicolinearidade inferior, as estimativas dos coeficientes pode até ser possível, Statsoft (2016).

A.7 *Outliers*

Outliers são observações que não se encaixam na tendência normal dos dados. A permanência desses valores pode prejudicar a construção do modelo.

Usando técnica estatística, os *outliers* podem ser detectados pelo valor dos *scores*, cujos valores atípicos são visualizados pelo valor chamado de Hotelling, representado por T^2, dado pela Equação A.1.7, Dell (2016).

$$T_i^2 = \sum_{k=1}^{N} \frac{t_{ik}^2}{s_{ik}^2} \quad (A.1.7)$$

Em que:

T_i^2 é o valor Hotelling;

t_{ik} é o *score* da $i\ ésima$ observação para o $k\ ésimo$ componente;

s_{ik}^2 é a variância estimada de t_k

Outliers moderados, os quais têm pouca capacidade de "puxar" o modelo não são facilmente observados pelo valor de Hotelling, sendo necessário calcular um valor chamado Distância do Modelo, ou D-To-Model (DTM), dado pela Equação A.1.8, complementada pela Equação A.1.9.

$$DTM = \sqrt{\frac{\sum_{j=1}^{M} (x_{ij} - \hat{x}_{ij})^2}{DF}} \quad (A.1.8)$$

$$DF = M - C \tag{A.1.9}$$

Em que:

x_{ij} são os elementos das variáveis originais;

$\hat{x}_{ij}$ são os elementos das variáveis preditas;

DF é o grau de liberdade;

M é o número de variáveis originais;

C é o número de componentes principais;

Outro parâmetro que pode ser obtido é chamado *Power*, sendo seu valor dado pela Equação A.1.10, e exprime a importância de uma dada variável no modelo.

$$Power = 1 - \frac{SV_j}{SV_j^o} \tag{A.1.10}$$

Em que:

SV_j é a variação padrão dos resíduos;

SV_j^o é o desvio padrão inicial.

As técnicas Hotelling, DTM e Power são encontradas no software Statistica no módulo PCA.

Nesta dissertação, a verificação dos *outliers* empregando as técnicas acima foi feita nos dados originais com o objetivo de sua retirada do banco de dados. O valor de Power foi usado quando da aplicação da técnica de PCA com o objetivo da redução de dimensionalidade nas variáveis de entrada.

A.8 A *Regressão* linear múltipla (RLM)

A RLM, assim como a RNA, tem por objetivo representar uma correlação entre variáveis de modo que uma delas possa ser calculada em função das demais. O resultado de sua eficiência é dado através de um valor de coeficiente de correlação (R^2). Quanto mais próximo de 1 (um) melhor a correlação estabelecida entre as variáveis.

O principio de resolução deste problema é ajustar uma linha reta em um espaço n-dimensional (onde n é o número de variáveis independentes) para certo número de pontos. No

caso mais simples (uma variável dependente e uma variável independente) visualiza-se um gráfico de dispersão onde os *scatterplots* são parcelas bidimensionais dos escores em um par de variáveis, (Dell, 2016).

Quando na avaliação da correlação entre variáveis está presente apenas uma variável independente para relacionar-se a uma variável dependente, a equação A.1.11 é a que define essa abordagem. Nesse caso, uma linha pode ser visualizada no espaço bidimensional.

$$y = \beta_0 + \beta_1 x_1 + \varepsilon \qquad \text{(A.1.11)}$$

Onde:

y = variável de saída, dependente;

β_0 = constante ou interceptação da reta;

β_1 = inclinação da curva;

x = variável de entrada, independente;

ε = erro

β_0 e β_1 são também chamados de coeficientes de regressão.

No caso da análise multivariada, quando há mais de uma variável independente, a linha de regressão não pode ser visualizada no espaço bidimensional, mas pode ser calculada, sendo o cálculo mais complexo. Um modelo de RLM descreve uma relação entre um conjunto de variáveis quantitativas independentes x_j, para $j \in \{1, \dots, k\}$ e uma variável quantitativa dependente *y* na forma da Equação A.1.12, Statsoft (2016).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \qquad \text{(A.1.12)}$$

O parâmetro β_j representa a variação esperada na resposta y por unidade de variação unitária em x_j, quando todos os outros regressores (ou variáveis independentes) x_i $(i \neq j)$ forem mantidos constantes.

A resolução da Equação A.1.12 é feita pelo método dos mínimos quadrados (Statsoft, 2016). A linha de regressão expressa a melhor previsão da variável dependente y, dadas as variáveis independentes x. Quanto menor for a variabilidade dos valores residuais em torno da linha de regressão em relação à variabilidade global, melhor é a previsão, Dell (2016).

APÊNDICE A.2

Pesos e biases nos modelos da caldeira e coluna

Tabela 16 Pesos e biases na estimação dos parâmetros no modelo da caldeira

	Pesos dos neurônios entre a camada de entrada e a camada escondida									
Variável	1º Neurônio	2º Neurônio	3º Neurônio	4º Neurônio	5º Neurônio	6º Neurônio	7º Neurônio	8º Neurônio	9º Neurônio	10º Neurônio
Vazão de vapor t/h	0,18514	-0,52487	0,01513	2,83291	-0,52954	-4,01053	-0,50739	-3,08112	-1,86964	1,44653
Vazão de água m³/h	-0,28668	-0,65178	0,21042	1,12357	0,02544	-1,83718	0,05884	-1,01491	-0,71745	1,12922
Vazão de gas natural kg/h	0,77648	1,23358	-0,70111	3,51948	0,20296	-1,20836	1,05588	-5,73269	-5,45486	4,61751
Vazão de ar kg/h	2,95139	-2,64227	-1,85047	4,33727	-2,02744	-2,15356	-0,10401	-2,54601	-3,54617	-0,18567
Pressão kgf/cm²	-1,01367	-0,06245	1,53013	-2,63596	-0,44967	1,35886	0,26165	1,13033	0,5454	-0,81888
T gases de combustão °C	-0,07095	-0,30184	-0,5722	3,65318	4,41297	-2,80871	-1,69163	-4,6301	-3,83469	1,51209

Bias de entrada	-0,0195	-1,14925	1,37953	-4,12859	0,24794	2,5881	0,64974	1,98155	4,49487	-2,6498

Pesos dos neurônios entre a camada escondida e a camada de saída									
1º Neurônio	2º Neurônio	3º Neurônio	4º Neurônio	5º Neurônio	6º Neurônio	7º Neurônio	8º Neurônio	9º Neurônio	10º Neurônio
0,74909	-4,15357	-0,63688	-0,56973	-0,99609	-0,75956	1,40045	2,40715	0,26755	-1,55311

Bias de entrada	-0,91921

Tabela 17 Pesos e biases na estimação dos parâmetros no modelo da coluna

	Pesos dos neurônios entre a camada de entrada e a camada escondida											
Variável	1º Neurônio	2º Neurônio	3º Neurônio	4º Neurônio	5º Neurônio	6º Neurônio	7º Neurônio	8º Neurônio	9º Neurônio	10º Neurônio	11º Neurônio	12º Neurônio
Carga	-1,40121	1,45940	2,50964	-0,59222	-2,41352	-0,46445	0,76770	-0,01849	0,28757	-2,19953	1,80404	-0,08056
Razão carga vapor	-0,13691	1,96583	-0,07053	1,49174	-1,99838	0,06717	0,82086	-0,04445	0,12681	0,86720	-0,22070	-0,22184
Arraste	1,06713	-0,18390	-1,37325	0,64951	-1,49647	0,26905	-0,26986	0,71068	-0,31166	1,42902	0,62609	-0,11224
Soda	0,31300	1,90994	0,79085	-0,71957	-1,74751	0,14846	-0,01145	-0,10811	0,09103	-1,47905	0,49251	-0,03264
Refluxo	-0,76425	0,83154	-0,49610	-0,46430	-1,27342	0,44882	-0,20664	-0,03311	0,06108	-0,56071	1,79226	0,13279
T carga	1,36334	-1,78166	-0,00373	-0,66338	-1,22766	0,42324	0,63872	0,22929	-0,22571	5,48920	0,00403	-0,01562
pH carga	-0,56572	-2,17227	-0,81313	0,69073	2,54818	0,38159	-1,05941	0,20862	0,21011	0,43461	-1,40674	0,06184
T flash	1,15506	-0,81558	1,27184	1,62554	0,15610	-0,08317	1,59255	0,29223	-0,07463	1,86087	0,69156	0,23526
P flash	-0,70798	-0,49479	-0,85261	1,35753	2,12572	0,05224	-1,42081	-0,48417	0,43749	-2,66727	1,03811	0,08389
DP recheio sup	-1,10877	0,45626	0,59832	3,04906	0,87497	0,16282	-0,81900	-0,10790	0,27887	1,94076	1,59523	-0,06265
Bias de entrada	-0,44930	-0,93743	-0,39202	-2,26343	-0,02657	0,77295	-0,36408	0,33579	0,11474	-0,71829	-1,03524	-0,11921

Pesos dos neurônios entre a camada escondida e a camada de saída											
1º Neurônio	2º Neurônio	3º Neurônio	4º Neurônio	5º Neurônio	6º Neurônio	7º Neurônio	8º Neurônio	9º Neurônio	10º Neurônio	11º Neurônio	12º Neurônio
1,08696	-0,37071	0,69620	1,81603	-2,32245	0,17668	0,65303	0,36560	-0,14648	-1,15276	-2,83058	-0,40696

Bias de entrada	-0,70318

APÊNDICE A.3

Regressão na coluna: RLM

Usando as 10 variáveis de entrada, os resultados da RLM se encontram na Tabela 18. O coeficiente de correlação R^2 ajustado foi de 0,452.

As Tabelas 18 e 19 mais as Figuras 45, 46 e 47 apresentam os resultados da predição de NH_3 no fundo da coluna pela técnica RLM.

Tabela 18 Resultados dos coeficientes da RLM na coluna com as 10 variáveis de entrada

N = 1177	b*	Erro padrão de b*	B	Erro padrão de b	p valor
Interseção			-78,2453	14,13409	0,000000
Carga	0,266422	0,059394	0,0002	0,00005	0,000008
Razão carga vapor	0,280118	0,050575	1,4650	0,26451	0,000000
Arraste	-0,072342	0,039844	-6,6544	3,66502	0,069681
Soda	-0,075028	0,028904	-0,0018	0,00071	0,009556
Refluxo	-0,343344	0,037123	-0,0018	0,00019	0,000000
T carga	-0,304619	0,046896	-0,1815	0,02794	0,000000
pH carga	-0,008189	0,027105	-0,1109	0,36700	0,762597
T *flash*	0,428844	0,053253	0,9453	0,11738	0,000000
P *flash*	- ,007119	0,048352	-0,2645	1,79633	0,882982
DP recheio sup	-0,106482	0,035946	-0,0041	0,00139	0,003116

b* = coeficiente de regressão que seria obtido se todas as variáveis tivessem sido padronizadas de modo a se ter média igual a zero e desvio-padrão igual a 1. A vantagem de se trabalhar com o coeficiente b* está no fato de se poder comparar a contribuição relativa de cada variável independente na previsão da variável dependente.

Excluindo da análise os regressores sem significância (p valor > 0,05) outra regressão foi feita cujos resultados se encontram na Tabela 19. O coeficiente de correlação R^2 ajustado foi também de 0,452. Dessa maneira prossegui-se na validação com a curva obtida pelos coeficientes da Tabela 18.

Tabela 19 Resultados da RLM nas 7 variáveis de entrada

N = 1177	b*	Erro padrão de b*	B	Erro padrão de b	p valor
Interseção			-95,7975	5,999842	0,000000
Carga	0,310812	0,054638	0,0003	0,000045	0,000000
Razão carga vapor	0,224683	0,041558	1,1751	0,217353	0,000000
Soda	-0,073280	0,025255	-0,0018	0,000622	0,003782
Refluxo	-0,320150	0,034761	-0,0017	0,000180	0,000000
T carga	-0,329814	0,035458	-0,1965	0,021122	0,000000
T *flash*	0,479573	0,037404	1,0571	0,082446	0,000000
DP recheio sup	-0,107160	0,034759	-0,0042	0,001349	0,002098

Com os dados da Tabela 18, coluna B, é obtida a curva que dá os valores preditos de NH_3 fundo em função das 10 variáveis de entrada. Nas Figuras 47 e 48 os valores preditos e observados, na Figura 49 o gráfico de resíduos em função dos valores observados.

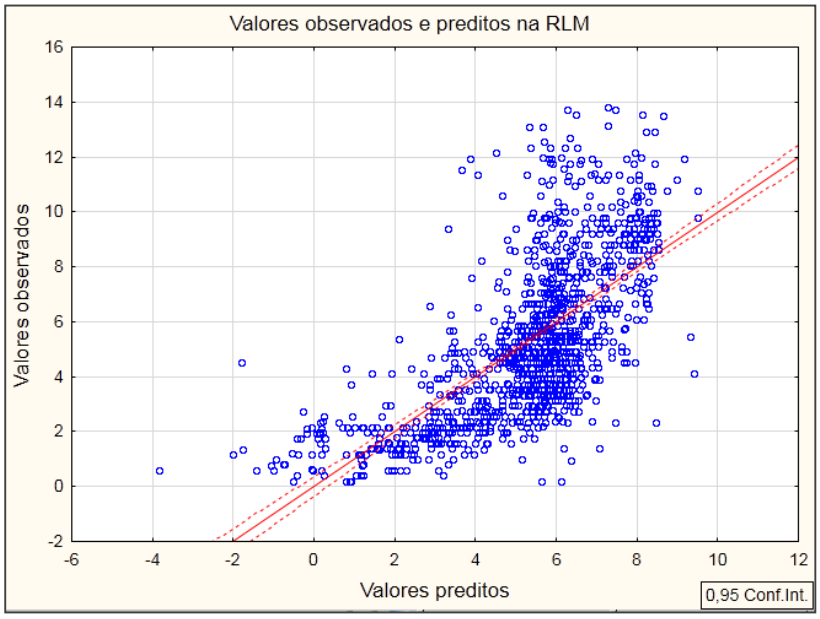


Figura 47 Dados reais de NH_3 fundo em relação aos preditos na RLM com 10 variáveis

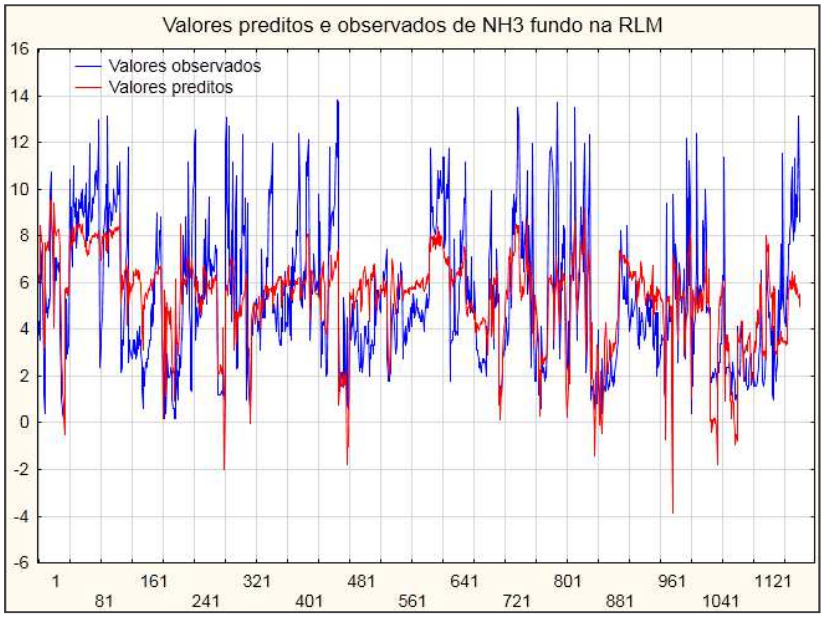


Figura 48 Dados reais e preditos de NH_3 fundo na RLM com 10 variáveis

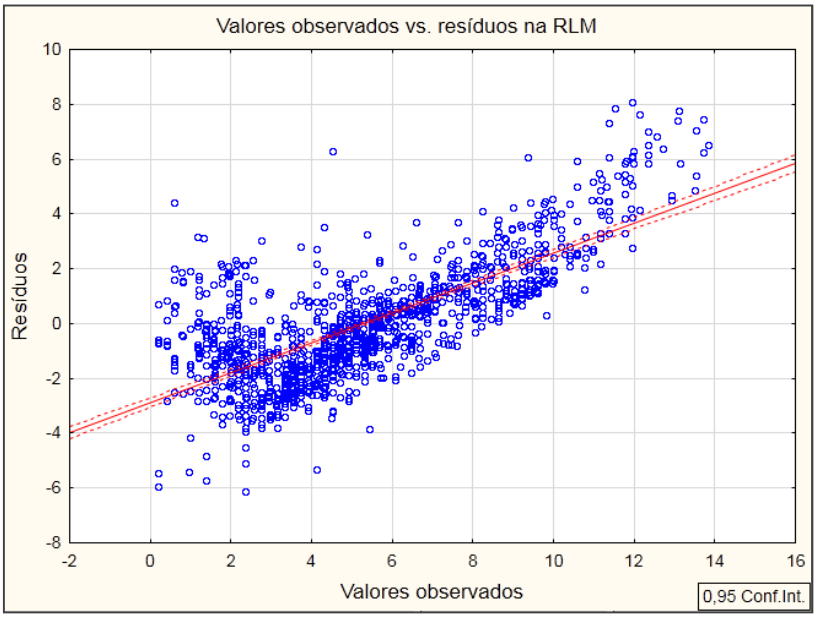


Figura 49 Resíduos de NH_3 fundo pela RLM em relação aos dados reais com 10 variáveis

Validação da RLM da coluna

Na validação da RLM, os mesmos 589 dados da validação da RNA foram apresentados aos coeficientes da curva obtida pela Tabela 18, coluna B. O coeficiente de correlação dessa análise foi de 0,485, semelhante ao obtido na construção do modelo. As Figuras 50 e 51 apresentam os valores preditos juntamente com os dados reais.

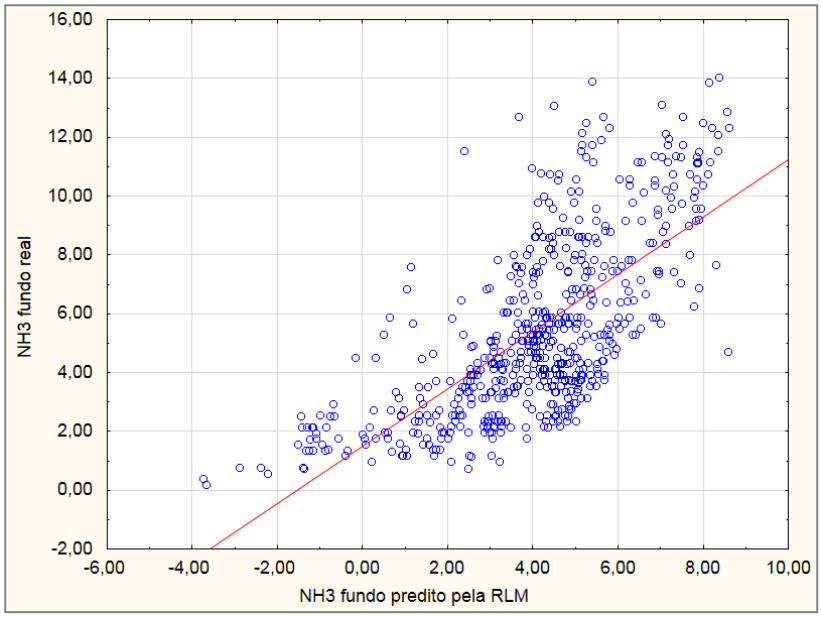


Figura 50 Valores preditos e observados na validação da RLM na coluna

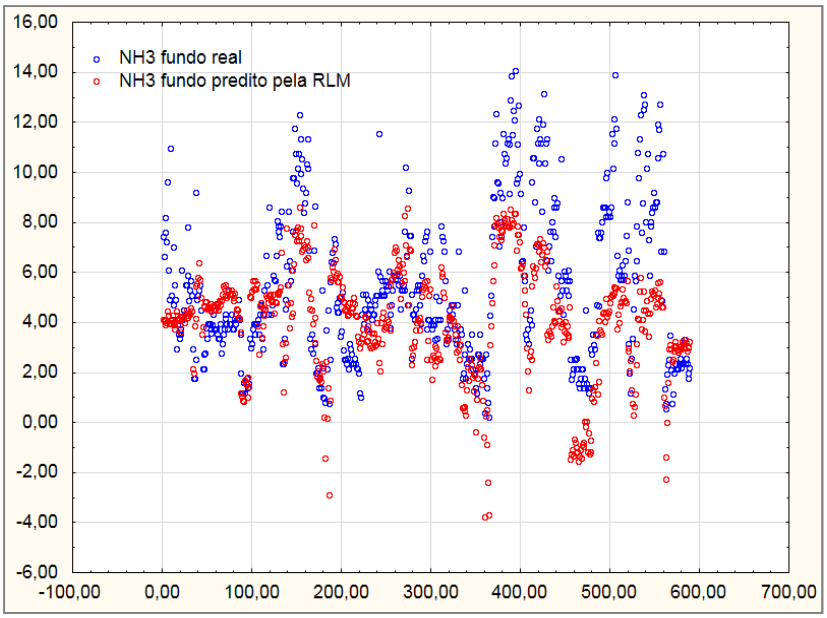


Figura 51 Valores preditos e observados com os dados de validação da RLM na coluna