

UNIVERSIDADE ESTADUAL DE CAMPINAS

FACULDADE DE ENGENHARIA QUÍMICA

ÁREA DE CONCENTRAÇÃO

SISTEMAS DE PROCESSOS QUÍMICOS E INFORMÁTICA

**APLICAÇÃO DE REDES NEURAIS PARA O AJUSTE
OPERACIONAL DO PERÍODO FINAL DE SOPRO DE UM
PROCESSO DE ACIARIA A OXIGÊNIO**

Autora: Tatiana Aparecida Pacianotto
Orientadora: Prof^a Dra. Ana Maria Frattini Fileti
Co-orientador: Eng. Dr. André Pitasse da Cunha

Dissertação de Mestrado apresentada à Faculdade de Engenharia Química como parte dos requisitos exigidos para a obtenção do título de Mestre em Engenharia Química.

Campinas – São Paulo

Outubro de 2003

UNIDADE	BC
Nº CHAMADA	
	F/UNICAMP
	P.118a
V	
TÓMBO BC	59532
PROC.	16-P.117/04
C	<input type="checkbox"/>
D	<input checked="" type="checkbox"/>
PREÇO	11,00
DATA	14/04/2004
Nº CPD	

Bib Id 321323

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA ÁREA DE ENGENHARIA - BAE - UNICAMP

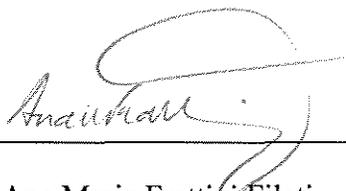
P118a Pacianotto, Tatiana Aparecida
 Aplicação de redes neurais para o ajuste operacional do período final de sopro de um processo de aciaria a oxigênio / Tatiana Aparecida Pacianotto.--Campinas, SP: [s.n.], 2003.

Orientadores: Ana Maria Frattini Fileti e André Pitasse da Cunha.

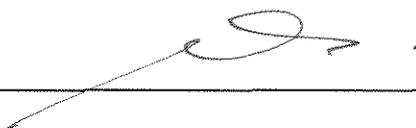
Dissertação (mestrado) - Universidade Estadual de Campinas, Faculdade de Engenharia Química.

1. Redes neurais (Computação). 2. Industria siderurgia. 3. Controle de processo. I. Fileti, Ana Maria Frattini. II. Cunha, André Pitasse da. III. Universidade Estadual de Campinas. Faculdade de Engenharia Química. IV. Título.

Dissertação de Mestrado defendida por Tatiana Aparecida Pacianotto e aprovada em 31 de outubro de 2003 pela banca examinadora constituída pelos doutores:



Prof^a. Dra. Ana Maria Frattini Fileti – Orientadora



Prof. Dr. José Claudio Geromel



Prof. Dr. José Vicente Hallak d'Angelo

Este exemplar corresponde à versão final da Dissertação de Mestrado em Engenharia
Química.

A handwritten signature in black ink, appearing to read 'Ana Maria Frattini Fileti', is written over a horizontal line. The signature is stylized and somewhat cursive.

Ana Maria Frattini Fileti

A Deus.

À minha família, pelo amor, carinho e apoio durante toda a minha vida. E `a memória de meu tio Quim, pelo incentivo e força que sempre recebi.

Agradecimentos

À Prof. Dr^a. Ana Maria Frattini Fileti, pela orientação, dedicação, paciência e amizade durante o desenvolvimento deste trabalho.

Ao Eng. André Pitasse da Cunha, pela co-orientação, incentivo e atenção.

À Faculdade de Engenharia Química da UNICAMP, através do Departamento de Sistemas de Processos Químicos e Informática, pela infra-estrutura oferecida.

À CAPES pela bolsa de estudos e apoio financeiro que permitiram a execução deste trabalho.

À CSN pela disposição dos dados e informações técnicas de seu processo utilizadas neste trabalho.

A meus pais, Carlos e Eloisa, pelo amor, pela credibilidade e apoio incondicional. Aos meus irmãos, Júnior e Eduardo, pelo otimismo e carinho.

Às minhas grandes amigas Ana e Priscila, pela convivência, constante ajuda tanto na parte profissional quanto pessoal, além de estarem sempre ao meu lado.

Aos meus amigos e colegas de laboratório por estarem sempre dispostos a ajudar, pelas dicas diante das dificuldades, e pelos momentos de descontração.

A todos os amigos que me acompanharam nesta jornada, nos momentos de alegria e naqueles em que precisei de atenção e carinho.

Ao Pervane e Gustavo, pelo suporte e paciência dedicados.

Ao Pierre, pelo incentivo, credibilidade e companheirismo que tornaram mais suaves a realização deste trabalho.

Aos professores de graduação da UEM, que foram responsáveis pela minha formação de graduação.

Fica registrado o meu reconhecimento a todos que contribuíram direta ou indiretamente.

“É pela lógica que provamos,
mas é pela intuição que descobrimos.”

Henri Poincaré – matemático francês.

RESUMO

O presente trabalho apresenta a aplicação de redes neurais artificiais ao processo de produção de aço líquido com o objetivo de elaborar dois modelos: rede direta, para predição de temperatura e porcentagem de carbono ao final do processo, e rede inversa, para ajuste operacional das quantidades de oxigênio para o término do sopro e de minério de ferro a ser adicionado no período final da batelada. Neste processo, o metal a 1400°C é convertido ao aço a aproximadamente 1650°C pela oxidação exotérmica de carbono, manganês, fósforo, silício e enxofre dissolvidos no ferro. Oxigênio é soprado no conversor LD usado na etapa de refino do aço através de uma lança e nenhuma fonte de calor externa é requerida. Este processo é transiente, altamente complexo e sofre freqüentes oscilações na composição da matéria-prima. No processo de refino da Cia. Siderúrgica Nacional (CSN-Volta Redonda/RJ), as quantidades das matérias-primas e de oxigênio para iniciar o sopro em um conversor LD são definidas através de um modelo estático, em que os cálculos são realizados com informações prévias ao início da batelada. Em um ponto pré-determinado do sopro de oxigênio, medidas de temperatura e de porcentagem de carbono são obtidas usando uma sub-lança. O tempo entre o ponto de amostragem e o final do processo é conhecido como o período de final de sopro. Durante este período, um modelo convencional é usado para predizer quanto de oxigênio e de minério de ferro deverá ser adicionado a fim de obter a porcentagem de carbono e a temperatura do aço desejadas no final do sopro. Este último modelo é chamado de modelo dinâmico por ser alimentado com informações obtidas durante o processo e por determinar ações corretivas. O modelo neural direto proposto neste trabalho é composto das seguintes variáveis de entrada: a temperatura e o percentual de carbono medidos através da sub-lança, as quantidades de minério de ferro e de oxigênio determinadas pelo modelo convencional vigente e, opcionalmente, as quantidades de ferro-gusa, sucata e escória alimentadas. As variáveis de saída são as mesmas do modelo dinâmico comercial utilizado atualmente na CSN: a porcentagem de carbono e a temperatura do aço no período de final de sopro. Para a simulação do modelo dinâmico direto foram consideradas 5030 bateladas (filtradas de 15.000 bateladas) para treinamento e 200 bateladas para testes. Este conjunto de dados equivale à cerca de 4 anos de processamento na aciaria da CSN. As melhores topologias encontradas foram 4x9x2 e 4x3x6x2, que alcançaram um acerto simultâneo das duas variáveis de saída de 76,5%. Cunha (2001) obteve um acerto simultâneo de 73,3% com um modelo dinâmico neural treinado com este mesmo banco de dados. O desempenho atingido pelo modelo convencional da CSN encontra-se em torno de 66%. Com a adição de variáveis de entrada opcionais um maior acerto simultâneo (77%) foi obtido com a arquitetura 7x4x4x2. Através de um banco de dados de operações mais recentes, o melhor acerto simultâneo foi obtido com a topologia 7x5x3x2, a qual foi mantida para a estrutura da rede inversa. Nesta estrutura, as variáveis de saída passam a ser as quantidades manipuláveis de oxigênio e minério de ferro para o final do sopro. A temperatura e a porcentagem de carbono desejadas ao final da batelada passam a compor o vetor de entrada juntamente com as demais variáveis citadas anteriormente. Testes industriais foram realizados com a rede inversa 7x5x3x2 treinada. As medições de temperatura e de porcentagem de carbono ao final de 40 bateladas mostraram que em 82,5% destas bateladas as ações corretivas, indicadas pela rede inversa e implementadas pelo operador, conduziam o processo às especificações desejadas de temperatura e porcentagem de carbono simultaneamente. A especificação de temperatura é atingida em 97,5% das bateladas. Observou-se ainda que a especificação de porcentagem de carbono é mais facilmente alcançada (90,6%) quando situa-se próxima a 0,04% de carbono. Conclui-se, portanto, que os modelos neurais desenvolvidos apresentaram desempenho de predição superior ao modelo dinâmico convencional utilizado atualmente pela CSN e que a rede inversa está apta a fornecer informações corretivas para o ajuste operacional deste processo.

ABSTRACT

The present work describes the application of neural networks in the basic oxygen steelmaking process. Two neural models were developed: the prediction model and the inverse model. The first one aims to predict end-blow temperature and carbon percentage. The inverse model was built to provide end-blow corrective actions in oxygen and iron ore additions. In this process, hot metal at about 1400°C is converted to steel at 1650°C by exothermic oxidation of metalloids dissolved in the iron. Oxygen is blown into the LD converter through a lance and no external heat source is required. This process is transient, highly complex and is subject to oscillations in raw material composition. In the refining process of the *Cia. Siderúrgica Nacional* (CSN-Volta Redonda/RJ), raw material proportions and the oxygen flow rate required to start blowing in the LD converter are defined through a conventional static model (calculations are performed based on feed informations). At a pre-determined point in the blow, temperature and carbon percentage are measured using the sub-lance. The time between the sample point and the end of the process is known as the end-blow period. During this period, a conventional model is used to predict how much oxygen to blow and how much coolant to add in order to achieve the desired end-point temperature and carbon content in liquid steel. This model is called dynamic model since process measurements are employed in its calculations and it determines corrective actions. The end-blow neural model proposed in the present work is composed of the following input variables: carbon (sub-lance), temperature (sub-lance), coolant (iron ore) and oxygen. Informations on amount of cast iron, scrap iron and slag can be optionally inputted. The output variables are end-blow carbon percentage and end-blow temperature. CSN's historical databases (4 years) were used in training procedures. Approximately 15000 records were collected. Records containing spurious data were filtered out and the new database with 5230 vectors were split in two sets, one for training and other for testing. The best architectures obtained were 4x9x2 and 4x3x6x2, that present 76,5% of simultaneous agreement in temperature and carbon percentage. Cunha's neural model (CUNHA, 2001) achieved 73,3% of simultaneous agreements in temperature and carbon outputs against 66% obtained with the conventional dynamic model. Adding the optional variables to the input vector, a hitting-rate (simultaneous agreement) of 77% was achieved through the 7x4x4x2 architecture. A recent data-base was also employed in training procedures. The architecture 7x5x3x2 presented the best result and it was chosen to be the architecture of the inverse model. In this case, the output variables are the oxygen volume and the amount of iron ore to be added in the end-blow period. Set-points of temperature and carbon percentage at the end of the run are joined to the other input variables. The inverse neural model 7x5x3x2 was employed in the industrial investigation. Temperature and carbon percentage measurements at the end of 40 industrial runs presented a 82,5% hitting-rate. It could be observed that temperature set-point was achieved in 97,5% of the cases. Carbon percentage set-point around 0,04% was easily achieved: 90,5% of the cases. In conclusion, the performance of the end-blow model predictions has increased substantially and the inverse neural model may be employed to predict oxygen and coolant flow rates without the requirement of a steelmaking expert.

III.3.1 Normalização.....	31
III.3.2 Topologia.....	31
III.3.3 Função de ativação.....	31
III.3.4 Algoritmo de treinamento.....	32
III.3.5 Conjunto de dados.....	32
III.4 Conclusão.....	33
IV: Resultados e discussões.....	35
IV.1 Dados 1.....	35
IV.1.1 Normalização.....	35
IV.1.2 Topologia.....	36
IV.1.3 Função de ativação.....	40
IV.1.4 Algoritmo de treinamento.....	41
IV.1.5 Banco de dados.....	41
IV.2 Dados 2.....	45
IV.2.1 Vetores das variáveis de entrada e saída.....	46
IV.2.2 Topologia.....	48
IV.2.3 Análise de componente principal (“prepca”).....	55
IV.3 Comparação final de desempenho.....	56
IV.4 Conclusões.....	60
V: Teste Industrial.....	62
V.1 Metodologia.....	62
V.2 Resultados.....	67
VI: Conclusões e Sugestões.....	68
VI.1 Conclusões.....	70
VI.2 Sugestões.....	72

Referências Bibliográficas	74
Anexo 1	76
Anexo 2	78

NOMENCLATURA

a = conjunto dos vetores de saída da RNA;

aC = acerto do teor de carbono (%);

a_j = vetor de saída da última camada da rede da variável de saída *j*;

a_{kp} = valor de saída do *k*-ésimo neurônio do *p*-ésimo par entrada-saída calculado pela rede;

aS = acerto simultâneo (%);

aT = acerto da temperatura no fim do sopro (%);

b = vetor de *bias* da RNA;

Backpro. = *Backpropagation* puro;

BOS = *Basic Oxygen Steelmaking*;

BR = algoritmo de treinamento *Levenberg-Marquardt* com Regularização Bayesiana;

CFS = teor de carbono no período de fim de sopro;

CIB = porcentagem de carbono com 88% de sopro (*in blow*);

dXprev = é a variação anterior dos pesos e *bias*;

e = função-objetivo;

Esc = quantidade de escória carregada no início do sopro;

f = função de transferência;

Gusa = quantidade de ferro gusa carregada no início do sopro;

H = $J^T J$ = aproximação da matriz de Hessian, contém as derivadas segundas dos erros da rede com respeito aos pesos e *bias*;

J = matriz Jacobiana, contém as derivadas primeiras dos erros da rede com respeito aos pesos e *bias*;

Linear = função de transferência linear;

LM = algoritmo de treinamento *Levenberg-Marquardt* puro;

Logsig = função de transferência sigmoial;

lr = taxa de aprendizagem;

M = quantidade de neurônios na camada intermediária;

mc = Constante de momento;

meanp = média de p;

MFe = quantidades de minério de ferro utilizadas para o restante do sopro;

MnFS = porcentagem de manganês medida no fim de sopro;

MnIB = porcentagem de manganês medida com 88% de sopro;

MSE = a média do somatório do quadrado do erro;

N = número atual de parâmetros da RNA;

Oxig = quantidade de oxigênio utilizado para o restante do sopro;

p = conjunto de vetores dos dados de entrada da RNA;

PFS = porcentagem de fósforo medida no fim de sopro;

pi = vetor dos dados entrada da variável de entrada i da RNA;

PIB = porcentagem de fósforo medida com 88% de sopro;

pn = Conjunto p normalizado;

Prepca = Análise do componente principal;

RNA = Rede Neural Artificial;

SFS = porcentagem de enxofre medida no fim de sopro;

SIB = porcentagem de enxofre medida com 88% de sopro;

SSE = Somatório do quadrado dos erros;

SSW = Somatório do quadrado dos pesos e *bias*;

stdp = desvio padrão de p;

Suc = quantidade de sucata carregada no início do sopro;

Tansig = função de transferência tangente hiperbólica;

TFS = e a temperatura do aço no período de fim de sopra;

TIB = temperatura medida com 88% de sopra (*in blow*);

t_{kp} = valor desejado;

w = vetor peso da RNA;

X = vetor coluna de todos valores de pesos e *bias*;

y_j = vetor de saída da rede da variável de saída j ;

α e β = parâmetros da função-objetivo;

γ = número de parâmetros efetivos da RNA.

CAPÍTULO I: INTRODUÇÃO

O aço é um produto siderúrgico definido como uma liga metálica composta de ferro e carbono. É um insumo básico na economia mundial que está presente em quase todos os bens de consumo, na construção civil e em obras de infra-estrutura.

Segundo Fiora (2003) na revista CREA-SP (Jul/Ago 2003), em 2002, foram produzidas no Brasil, de acordo com os dados do Instituto Brasileiro de Siderurgia, mais de 29,6 milhões de toneladas de aço bruto (contra 26,7 milhões em 2001), o que coloca o país entre os oitos maiores produtores mundiais. O total das exportações de produtos siderúrgicos em 2002 ultrapassou US\$ 2,9 bilhões, com um crescimento de 28,3% em relação a 2001. As importações atingiram a casa de US\$ 2,5 bilhões. Para 2003, a expectativa é de que sejam produzidas mais de 31 milhões de toneladas. A capacidade instalada da siderurgia nacional é de 32 milhões de toneladas/ano. Com a projeção de novos investimentos superiores a US\$ 3 bilhões, até 2005, o país terá capacidade de produzir quase 37 milhões de toneladas em 2007. Ainda, segundo a revista CREA-SP (Jul/Ago 2003), a siderurgia brasileira está entre as melhores do mundo em competitividade. Nos últimos 10 anos, foram investidos no setor US\$ 10 bilhões, o que possibilitou uma total modernização das usinas. Como resultado, o país tem o menor custo de produção de placas de aço. É com esta filosofia de modernização e redução de custos que se propõe a realização deste trabalho.

Com base no livro “Aço e Siderurgia” organizado pela Comissão de comunicação social do IBS (Instituto Brasileiro de Siderurgia, 1984) e no trabalho de Cunha (2001), é possível resumir o processo siderúrgico em algumas etapas:

- A primeira delas é o preparo das matérias-primas, ou seja, uma mistura de carvão mineral é transformada em coque na **coqueria**, com a eliminação de compostos voláteis e impurezas, como o enxofre, produzindo o produto coque com granulometria e propriedades químicas e metalúrgicas adequadas ao consumo para as reações de redução dentro do alto-forno. Além disso, as partículas de ferro passam por um processo de **sinterização**, ou seja, aglomeração e sinterização de finos de minério de ferro, de fundentes e outros aditivos, utilizando como combustível finos de coque;

- As etapas seguintes são a redução do minério e o refino do aço, ou seja, a produção de ferro gusa (ferro líquido que contém um alto índice de carbono – mais ou

menos 4,0%) e a produção do aço. A tecnologia mais empregada nesta etapa, no Brasil, é a usina integrada que compreende o **alto-forno** e um **conversor a oxigênio**. O **alto-forno** é responsável pela produção de ferro gusa (ferro líquido), a partir de minério de ferro, fundentes, agente redutor (coque) e outros aditivos. O **conversor a oxigênio** é um forno forrado com tijolos refratários básicos (magnésia-carbono), que transforma, com sopro de oxigênio, o ferro gusa e a sucata, em aço, através de reações químicas de oxidação de elementos como Si, Mn, P e C. A sucata é composta por fontes de ferro diversas, tais como refugos do processo de laminação, lingotamento, escória beneficiada, etc. Neste processo de refino, onde a principal matéria-prima é o ferro gusa, aditivos para o ajuste da composição química da escória e do aço são alimentados, tais como cal calcítica, cal dolomítica, minério de ferro, FeSi, etc.;

- Na continuidade do processo ocorre a conformação mecânica, ou seja, o **lingotamento**. Nesta etapa, o aço líquido é transformado em placas pelo processo de solidificação de forma contínua.

A figura 1.1 apresenta o fluxograma básico do processo siderúrgico de aço:

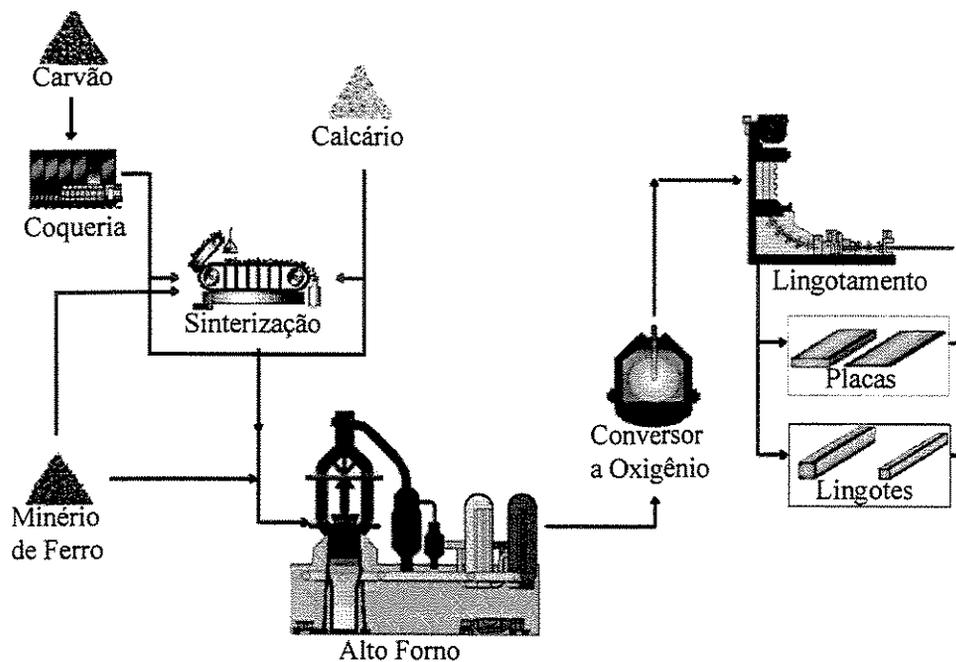


Figura 1.1: Fluxograma Básico do processo siderúrgico de aço-carbono.

Neste sistema de produção descrito acima, o processo produtivo que possui maior controle é a produção de aço. Isto se deve principalmente ao fato de que a aciaria recebe, em seus conversores, matérias-primas com maior frequência de amostragem e caracterização, como o ferro gusa, que compõe de 75 a 100% a carga do conversor (dependendo do tipo de aço requisitado). Desta forma a necessidade de automação deste processo conduziu ao desenvolvimento de modelos de controles estático e dinâmico, visando garantir a qualidade do aço. Esses modelos são classificados genericamente de **estático**, quando os cálculos são realizados a partir de dados levantados antes do início do sopro; e **dinâmico**, quando utilizam informações colhidas durante o refino e determinam ações corretivas. Se a temperatura estiver abaixo do previsto e/ou a porcentagem de carbono acima do especificado, o banho é resoprado. Caso contrário, o banho poderá ser resfriado com adições de minério de ferro. Entretanto, apesar da existência destes modelos, a previsão da qualidade do aço ainda é bastante prejudicada pela falta de informação referente às demais matérias-primas e aos gases e fumos gerados pelo processo (CUNHA, 2001; MALYNOWSKYJ, 1997).

Processos industriais normalmente são caracterizados por inúmeros fenômenos que, se tratados individualmente, não descrevem com precisão o modelamento como um todo. E a interação de vários fenômenos num mesmo processo leva a um alto nível de complexidade de modelagem. Como se isso não bastasse, os conhecidos problemas de variação dos resultados de qualidade de matérias-primas e do produto adicionam ainda maiores dificuldades de validação de modelos matemáticos para um sistema complexo, como a produção do aço (CUNHA, 2001).

Como alternativa à modelagem rigorosa de processos, surgiu uma das tendências mais recentes em software a serviço da automação industrial siderúrgica: a implantação de redes neurais artificiais. É possível encontrar muitos trabalhos e aplicações de redes neurais em diversas etapas do processo siderúrgico buscando eficiência operacional, porém um número bastante limitado é aplicado aos conversores a oxigênio, objeto de estudo do presente trabalho.

Por ser uma técnica de aplicação em base de dados (informações disponíveis nas empresas), as redes neurais artificiais se adequam perfeitamente às aplicações industriais, pois as combinações de várias experiências industriais das aleatoriedades citadas possuem uma faixa de atuação que é possível rastrear (dentro de limites de especificação) . E uma

grande vantagem da rede neural é a possibilidade de melhoria contínua do modelo de previsão, mediante o aumento da apresentação de casos reais de processos ao sistema de aprendizado artificial.

Tendo em vista que, em conversores a oxigênio de aciarias integradas, há um problema complexo de baixa previsibilidade da porcentagem de carbono e a temperatura do aço no período de final do sopro, o objetivo desta dissertação de mestrado é desenvolver modelos via redes neurais artificiais para a predição destas variáveis, com base em dados industriais do processo de refino. Uma maior previsibilidade da porcentagem de carbono no final do sopro do processo de obtenção do aço se mostra economicamente importante, pois implica em diminuir o tempo das bateladas do conversor, aumentando assim o número de bateladas diárias.

Na prática, o objetivo da variável temperatura de final de sopro é mais difícil de ser atingido que o objetivo da variável % de C, primeiramente, porque o desvio máximo permitido é de ordem bastante inferior ao permitido para a porcentagem de Carbono. Isto se deve também ao fato de que a principal ação corretiva para o ajuste da temperatura é implementada através da adição de minério de ferro. Entretanto, este minério de ferro apresenta fortes oscilações granulométrica e de composição química, o que não mantém padronizado o efeito refrigerante da adição desta variável. Além disto, o acerto da temperatura tem uma implicação significativa na fase subsequente do processo, que é o lingotamento contínuo.

A proposta estabelecida neste trabalho é motivada pela tese de doutorado de Cunha (2001), que desenvolveu modelos neurais para o processo de refino de aço. Os resultados obtidos foram superiores aos modelos convencionais da CSN. O resultado obtido pelo modelo dinâmico via redes neurais para prever o acerto simultâneo da porcentagem de carbono e da temperatura do aço no período final do sopro no processo de aciaria, foi de 73,3% contra 44% do modelo convencional.

Com base nestes resultados, o objetivo é encontrar novos modelos neurais que elevem a porcentagem de acerto simultâneo do modelo dinâmico encontrada por Cunha, através da adição de novas variáveis ao modelo e através de aplicações de diferentes arquiteturas de redes neurais artificiais.

Este trabalho está dividido em capítulos, sendo que, no Capítulo 2, é desenvolvida uma revisão bibliográfica sobre o processo siderúrgico e sobre redes neurais artificiais. O Capítulo 3 trata da metodologia empregada, no Capítulo 4 estão apresentados os resultados experimentais, o Capítulo 5 traz o teste industrial e o Capítulo 6 as conclusões as sugestões para trabalhos futuros.

CAPÍTULO II: REVISÃO BIBLIOGRÁFICA

Neste capítulo são apresentados alguns conceitos relativos ao processo em estudo, seu controle e modelagem, além da técnica de redes neurais artificiais. Alguns trabalhos de aplicação de redes neurais em aciaria/siderurgia também são expostos.

II.1 Conversor a oxigênio

No processo de elaboração do aço a aciaria assume um papel importante, pois é nesta etapa do processo que é estabelecida a qualidade do aço.

A aciaria básica a oxigênio (também conhecido como conversor LD) foi desenvolvida no início da década de 1950, pela siderúrgica austríaca Vöest-Alpine em Linz – Donawitz (LD). O principal insumo deste processo é o ferro-gusa, produzido em altos-fornos (PAULA, 2001).

O princípio químico empregado no conversor é a oxidação. O oxigênio se combina com elementos como: carbono, manganês, silício e fósforo (exceto o enxofre) e, inevitavelmente, com parte do ferro, formando óxidos que podem deixar o sistema como gases, ou passar à escória. A remoção do enxofre não envolve oxidação, mas depende da basicidade da escória e da temperatura atingida no processo (CUNHA, 2001).

A descarburização é a mais importante reação no conversor porque é ela que controla a taxa de produção. O oxigênio e o carbono, ambos dissolvidos no meio, e o carbono do banho e o FeO, reagem e formam em ambas as reações o monóxido de carbono (CUNHA, 2001; HERSKOVIC, 1976).

A Figura 2.1 apresenta um conversor a oxigênio que é um reator que torna o processo de refino simples e eficiente.

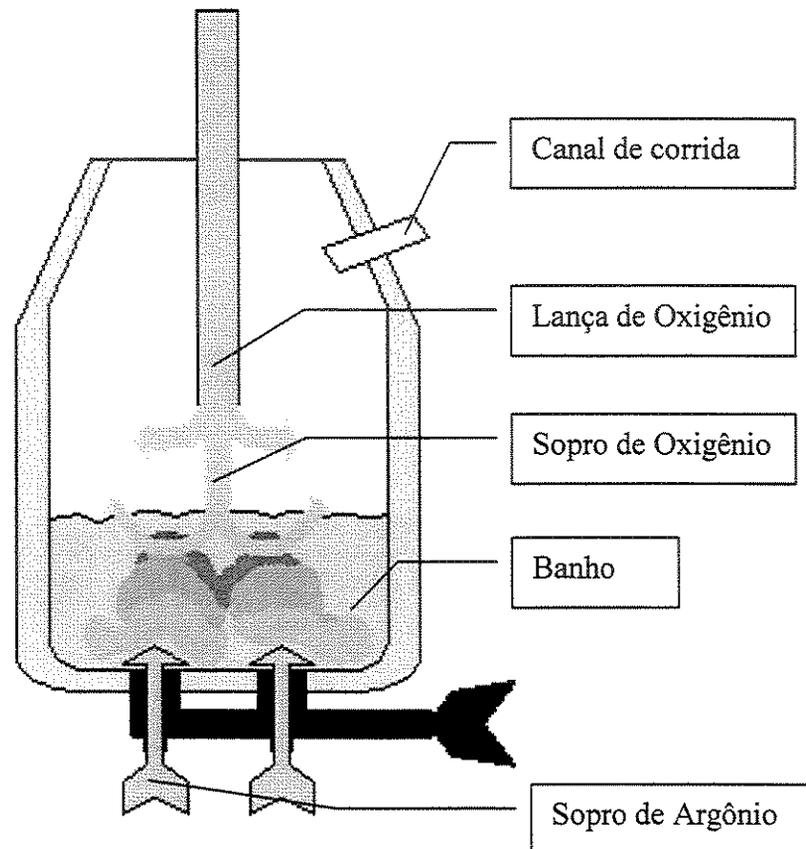


Figura 2.1: Conversor a Oxigênio

Fonte: CUNHA (2001)

As etapas do processo do conversor foram descritas abaixo (CUNHA, 2001):

1. Carregamento de matérias-primas: materiais ferrosos reciclados, tais como sucatas de ferro, sucatas mistas (sucatas de ferro sem e com revestimento de zinco e estanho) e sucatas oriundas de escórias de bateladas anteriores;
2. Carregamento de ferro gusa, a principal matéria-prima, cerca de 80% em média da carga metálica, que é um ferro líquido com 94,5% de ferro e 4,3% de carbono. Este produto de altos fornos possui importantes impurezas, tais como silício, fósforo, manganês e enxofre, que são controladas para a produção do aço;
3. Carregamento de fundentes, para ajuste da composição da escória;
4. Sopros de oxigênio, realizado no banho através de uma lança, oxidando silício, carbono, manganês e fósforo. Os óxidos de silício, manganês e fósforo, aliados aos óxidos de cálcio e de magnésio oriundos dos fundentes, formam junto ao óxido de ferro gerado no sopro, uma escória altamente básica. A escória tem papel fundamental no processo e sua composição química, além de proteger a região refratária dos conversores

que tem contato com a sílica e o FeO da escória, funciona estabilizando fósforo e enxofre, evitando que estas impurezas retornem ao aço;

5. Amostragem do banho e medição de temperatura e de percentual de carbono por inferenciação da temperatura de solidificação da amostra. Nesta etapa, se a temperatura do banho estiver abaixo do previsto e/ou o carbono estiver acima do projetado, o modelo dinâmico calcula a quantidade de oxigênio necessária ao resopro. Mas se a temperatura estiver acima do previsto pelo modelo estático, o modelo dinâmico indica o quanto de minério (refrigerante) será necessário para o resfriamento para atingir a temperatura desejada;

6. Vazamento do aço para processamentos posteriores;

7. Vazamento da escória. Esta escória é beneficiada por terceiros, dentro das próprias instalações da CSN, e retorna como sucatas (ricas em ferro metálico).

A Figura 2.2 apresenta o esquema operacional do conversor a oxigênio:

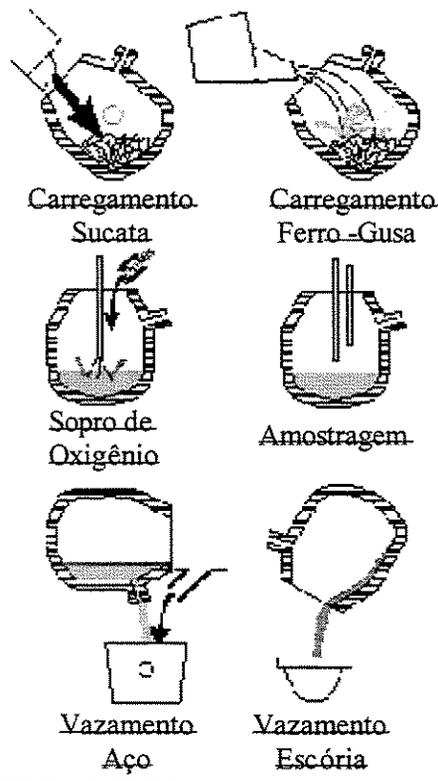


Figura 2.2: Esquema dos passos operacionais do conversor a oxigênio

Fonte: MILLER et al. (1998)

II.1.1 Controle Operacional (CUNHA, 2001; HERSKOVIC, 1976).

Segundo Herskovic (1976), um dos parâmetros que pode avaliar os resultados operacionais de um conversor a oxigênio é o acerto da composição química e da temperatura do aço no final do sopro. Este acerto é importante para possibilitar uma prática de desoxidação e adição de ligas em condições reprodutivas, evitando variações no rendimento. O acerto no conversor define a composição final do aço e condiciona a qualidade do produto final. Além de estar intimamente ligado à produtividade, pois o resopro de bateladas por alto carbono ou baixa temperatura, e o seu resfriamento, além de prejudicarem a qualidade do aço, podem consumir um tempo operacional adicional.

As condições operacionais afetam o acerto da composição e da temperatura no final do sopro na medida em que elas controlam a formação da escória e o desenrolar do refino, afetando nitidamente o grau de oxidação do metal no final do sopro.

Mais comumente utiliza-se um modelo matemático mais ou menos complexo, que abrange um número variável de parâmetros, e que fornece ao soprador subsídios mais ou menos completos para a condução do sopro. Esses modelos são classificados genericamente de **estático**, quando os cálculos são realizados a partir de dados levantados antes do início do sopro; e **dinâmico**, quando utilizam informações colhidas durante o refino e determinam ações corretivas.

Desta forma, para a determinação das matérias-primas, dos insumos e das condições de operação, são utilizados três tipos de modelo: o modelo manual, baseado no ajuste do operador, o modelo estático e o modelo dinâmico. O modelo estático é o modelo inicial utilizado para a definição de todos os aditivos (materiais reciclados, sucatas e ferro gusa) e da quantidade de minério e de oxigênio necessários ao processo para a composição química e temperatura desejadas.

Os **modelos estáticos** são baseados em balanços de massa e de energia e em correções a partir de análises estatísticas de dados operacionais e experimentais. Qualquer modelo tem o seu acerto altamente dependente das limitações de medidas, diferenciação da ação humana (que varia de acordo com a padronização das operações da empresa), simplificações teóricas dos modelos e procedimentos ainda não equalizados (como tempo entre bateladas e de transporte de ferro gusa). Estas e outras limitações tornam pobre a

resposta do modelo estático, justificando o uso de modelos dinâmicos e/ou modelos alternativos de inteligência artificial.

Já o **modelo dinâmico** funciona como um ajuste final para a obtenção das características desejadas, onde a coleta de amostras ao final do sopro indica qual a porcentagem de carbono (CIB) e de temperatura (TIB) naquele momento. Com este modelo calculam-se as necessidades de refrigerantes e de oxigênio necessários para atingir as características do aço.

II.1.1.1 Modelo Dinâmico de aciaria

O modelo dinâmico utiliza as quantidades de aditivos estipuladas pelo modelo estático e, através das indicações de sensores, possibilita ações corretivas durante o sopro. O ponto fundamental neste controle é o acompanhamento da velocidade de descarburação, avaliada através de análise de gases, métodos audiométricos ou utilização de sublanças. Este tipo de controle requer alto investimento e necessita alto grau de padronização da aciaria (CUNHA, 2001).

A coleta de informações durante o sopro é um ponto crítico do controle dinâmico, uma vez que exige equipamentos de alto grau de confiabilidade, reprodutividade e precisão, além de um tempo de resposta mínimo a um custo compatível (MALYNOWSKYJ, 1997).

Dos métodos de controle dinâmico disponíveis, a sublança é a técnica mais conhecida e difundida. Ela é dotada de um dispositivo que simultaneamente analisa o teor de carbono e a temperatura do banho a 88% do sopro de oxigênio. O objetivo é atingir a trajetória do banho dentro da faixa de acerto objetivado com as informações obtidas das amostragens realizadas durante o sopro. O teor de carbono instantâneo é inferido através da temperatura de solidificação da amostra (CUNHA, 2001; MALYNOWSKYJ, 1997).

O modelo existente na CSN considera as seguintes variáveis operacionais: capacidade dos conversores, altura da lança para ressopro, vazão de oxigênio, pressão de oxigênio, inclinação dos furos e pesos da escória. Algumas variáveis são consideradas constantes, como altura da lança e outras inferidas, como é o caso do peso da escória. Assim estas simplificações geram limitações no acerto do modelo dinâmico. Este modelo é um modelo cinético/estático desenvolvido pela Nippon Steel. É baseado em equações exponenciais de taxa de descarburação e taxa de acréscimo de temperatura.

Como a nova rota de automação industrial seguida pela siderurgia é a implantação de rede neural artificial (RNA), alternativa de inteligência artificial que representa a tendência mais recente de software a serviço da competitividade (JOCKYMAN e SILVA, 2002), foi desenvolvido por Cunha (2001) um modelo dinâmico para o processo de refino via redes neurais artificiais.

Em especial, no caso do conversor a oxigênio da CSN, há algumas particularidades que justificam a utilização de redes neurais artificiais para o controle do modelo dinâmico. Pode-se citar a disponibilidade de um banco de dados razoavelmente grande, uma quantidade de ruídos considerável, a grande variação química do ferro gusa e das demais matérias-primas do conversor e principalmente a dificuldade de utilizar um modelo matemático para realizar este controle.

Os resultados obtidos por Cunha (2001) com o modelo dinâmico via redes neurais foram superiores aos modelos convencionais da CSN. O acerto simultâneo do modelo neural foi de 73,3% contra 44% do modelo convencional. Este excelente resultado motivou testes industriais. Para tanto, combinaram-se os modelos estático convencional e dinâmico via redes neurais. Após 30 bateladas, o acerto simultâneo foi de 83,3%.

Cox et al. (2002) desenvolveram uma investigação para o uso de redes neurais artificiais (RNAs) para a prever o oxigênio e o refrigerante (minério de ferro) necessários para o período final do sopro no processo de aciaria. Os dados para o treinamento e teste do modelo neural foram obtidos na base de dados históricos da planta BOS (*Basic Oxygen Steelmaking*) da Port Talbot (empresa CORUS), usina siderúrgica do reino Unido. Após uma cuidadosa seleção dos dados, o modelo neural obteve um acerto de 85% para o conjunto de treinamento e 80,8% para o conjunto de teste. Mas, o modelo neural não foi aplicado na planta para verificar a sua performance.

II.2 Modelagem de processos

Segundo Shaw e Simões (1999) a modelagem de situações reais pode ser baseada em (1) Método experimental, (2) Modelagem matemática (ou analítica) e (3) Método heurístico:

- O método experimental é equivalente ao de se graficar pontos discretos de uma curva característica de entradas e saídas obtidas experimentalmente. Geralmente, os equipamentos físicos de que se constitui o processo não estão facilmente disponíveis.

Conseqüentemente, não se podem efetuar experimentos para determinar como o processo reage às diversas entradas, não se podendo portanto projetar o sistema de controle apropriado. Mesmo que os equipamentos de processo sejam disponíveis para experimentação, o procedimento experimental pode ser caro. Um estudo cuidadoso também se faz necessário para avaliar se as faixas de valores de entrada e saída são cobertas pelos instrumentos de medição. É muito mais desejável prever os possíveis valores de entrada, sem a necessidade de usar medições. Assim, é preciso uma descrição de como o processo reage para várias entradas, sendo exatamente isso o que os modelos matemáticos fazem;

- O método de modelagem matemática implica em se obter um modelo idealizado do processo a ser controlado, geralmente em forma de equações diferenciais ou algébricas. Para que o modelo seja razoavelmente simples são necessárias certas restrições. Como exemplo, é comum assumir que o processo seja linear, ou seja, que variações na entrada produzem variações proporcionais na saída. Devido à necessidade de tais simplificações, o projetista, em geral, encontra sérias dificuldades no desenvolvimento de uma descrição matemática significativa e realista de um processo industrial. É claro que, ao assumir simplificações em modelos matemáticos, podem-se ignorar informações importantes, e esta perda de informações resulta em reajustes e ressintonias do controlador na planta real, na operação do processo. Isso é o que ocorre na prática, e pode ser até bem sucedido, todavia, quando a complexidade do processo excede um determinado grau, os modelos matemáticos não se tornam questionáveis. Os engenheiros voltados às aplicações práticas já chegaram a conclusão que, devido às diversas simplificações feitas devido às limitações físicas, torna-se contraditório a busca cada vez maior de precisão na modelagem matemática;

- O método heurístico consiste em realizar uma tarefa de acordo com a experiência prévia, com regras práticas ou “dicas” e estratégias já freqüentemente utilizadas. Uma regra heurística é uma implicação lógica ou uma regra típica de uma situação de controle. O método heurístico é similar ao método experimental de construir uma tabela de entradas e saídas. No entanto, a restrição de linearidade não é mais necessária ou mesmo relevante, uma vez que a função entrada-saída é descrita ponto-a-ponto exatamente como no método experimental. Em outras palavras, o poder do método heurístico está em sua habilidade de possibilitar a construção de uma função de entrada versus saída, não matemática, e útil na descrição de uma planta ou processo.

Recentemente um interesse ativo no desenvolvimento e aplicação de metodologias não-lineares tem emergido. Controle não-linear agora ocupa uma crescente posição de importância na área de engenharia de controle de processo. No entanto, em muitos sistemas não-lineares é extremamente difícil e caro obter um modelo preciso do processo com princípios anteriores (HUSSAIN, 1999). Ao mesmo tempo, uma técnica nova de modelagem tem se evidenciado: a rede neural artificial. Ela é basicamente uma operação “caixa –preta” relacionando dados de entrada com dados de saída com muita habilidade, mas por um caminho indefinido. Uma vez que redes neurais artificiais são técnicas estatísticas não-lineares, elas podem ser usadas para resolver problemas que não são adequados para os métodos estatísticos convencionais (MALINOVA, 2001). Em muitos problemas da prática industrial os engenheiros iniciam com dados de entrada-saída e não há modelo matemático. Uma importante tarefa é desenvolver um modelo consistente. Por outro lado, não é possível um modelo não-linear, detalhado. Uma saída para resolver este problema é usar as redes neurais artificiais, que desenvolvem modelos não-lineares a partir de dados de entrada-saída, e esta propriedade tem profundas implicações, tanto teóricas como práticas (RAUCH, 1998). Em particular, as redes neurais têm muito a oferecer quando a solução de um problema de interesse é dificultada por alguns dos seguintes pontos como falta de entendimento físico/estatístico do problema, variações estatísticas nos dados observáveis, mecanismo não linear responsável pela geração de dados.

II.2.1 Redes Neurais

Como a proposta deste trabalho é desenvolver modelos neurais para a previsão do acerto simultâneo da porcentagem de carbono e da temperatura do aço no período de fim de sopro, a seguir, são apresentados alguns conceitos relativos a esta ferramenta.

II.2.1.1 Definição de redes neurais artificiais

Uma rede neural artificial (RNA) é uma poderosa ferramenta que se baseia no processamento paralelo e distribuído de informações (CUNHA, 2001). O paralelismo não é somente um conceito essencial ao processamento de informação em redes neurais, mas é também a fonte de flexibilidade. Além disso, o paralelismo pode ser maciço, o que dá à rede uma notável forma de robustez (pequenas incertezas do modelo e pequenas perturbações resultam apenas em pequenos erros de estimativa). Entradas ruidosas ou

incompletas podem ainda ser reconhecidas, uma rede danificada pode ainda ser capaz de funcionar satisfatoriamente, e a aprendizagem não precisa ser perfeita (HAYKIN, 2001).

É evidente que uma rede neural extrai seu poder computacional através, primeiro, de sua estrutura maciça paralelamente distribuída e segundo de sua habilidade de aprender e, portanto generalizar. A generalização se refere ao fato da rede neural produzir saídas adequadas para entradas que não estavam presentes durante o treinamento (aprendizagem). Estas duas capacidades de processamento de informação tornam possível para as redes neurais resolver problemas complexos (de grande escala) que são atualmente intratáveis. Na prática, contudo, as redes neurais não podem fornecer uma solução trabalhando individualmente. Em vez disso, elas precisam ser integradas em uma abordagem consistente de engenharia de sistemas. Suas aplicações são encontradas em campos muito diversos, como modelagem, análise de séries, processamento de sinais e controle, em virtude de uma importante propriedade: a habilidade de aprender a partir de dados de entrada (HAYKIN, 2001).

II.2.1.2 Estrutura (HAYKIN, 2001; DEMUTH e BEALE, 2002)

A estrutura de uma rede neural artificial consiste em um conjunto de neurônios (processadores matemáticos) em camadas e na conexão entre eles (sinapses) através de funções de ativação. A primeira camada recebe as variáveis de entrada e a última fornece as de saída. As camadas internas são chamadas de camadas intermediárias. A função dos neurônios da camada intermediária é intervir entre a entrada externa e a saída da rede de uma maneira útil. Adicionando-se uma ou mais camadas intermediárias, tornamos a rede capaz de extrair estatísticas de ordem elevada. Esta habilidade é valiosa quando o tamanho da camada é grande.

A Figura 2.3 apresenta um esquema da estrutura simplificada de uma rede neural artificial com uma única camada intermediária.

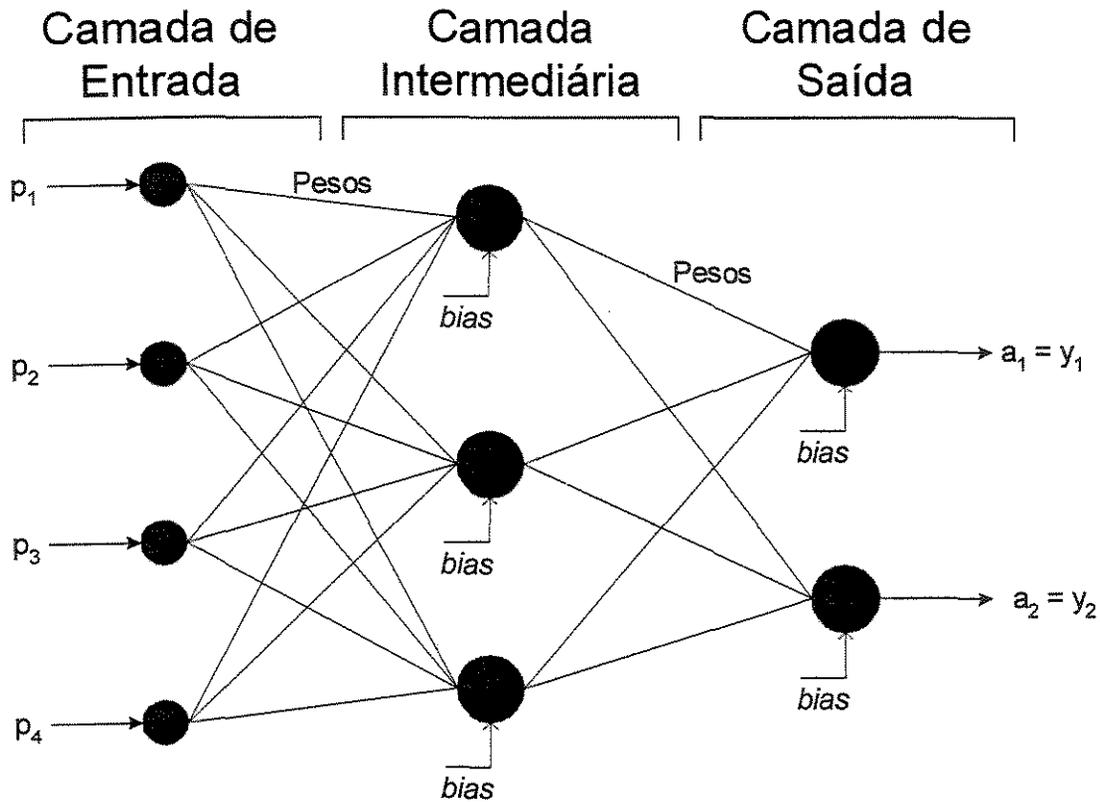


Figura 2.3: Exemplo de uma estrutura simplificada de uma rede neural artificial com uma única camada intermediária.

Onde:

p_i = vetor dos dados de entrada da variável de entrada i

a_j = vetor de saída da última camada da rede da variável de saída j

y_j = vetor de saída da rede da variável de saída j

Cada **neurônio artificial** é uma unidade de processamento de informação que é fundamental para a rede neural. O neurônio recebe um valor de entrada p que é multiplicado por um **peso** w , o somatório deste produto com o **bias** b forma o argumento da função de transferência f . Os neurônios podem usar qualquer função de transferência diferenciável f para gerar suas saídas a . A Equação 2.1 apresenta um modelo de um neurônio artificial.

$$a = f(wp+b) \tag{2.1}$$

O *bias* é um parâmetro externo do neurônio artificial que tem o efeito de aumentar ou diminuir a entrada líquida da função de ativação, dependendo se ele for positivo ou negativo, respectivamente.

A **função de transferência** restringe a amplitude de saída do neurônio, podendo também ser referida como função restritiva já que limita o intervalo permissível de amplitude do sinal de saída a um valor finito. Os tipos básicos de funções de transferência são linear, logaritmo-sigmoidal e tangente hiperbólica:

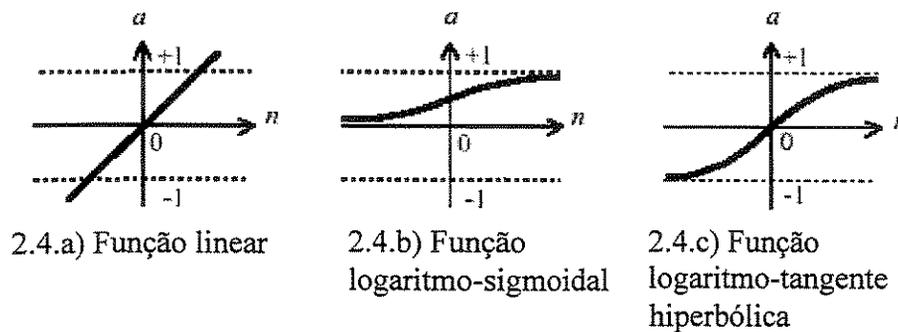


Figura 2.4: Comportamento dos tipos básico de funções de transferência
 Fonte: DEMUTH E BAELE. (2002)

O comportamento da função linear está representado na Figura 2.4.a e na Equação 2.2. Esta função de transferência gera a saída como a entrada. A função logaritmo-sigmoidal gera saídas entre 0 e 1, e é definida como uma função estritamente crescente que exibe um balanceamento adequado entre comportamento linear e não linear. Seu comportamento está representado na Figura 2.4.b e na Equação 2.3. Algumas vezes é interessante que a função de transferência se estenda de -1 a +1, assumindo neste caso uma forma anti-simétrica em relação à origem. Neste caso, a função de transferência mais adequada é a tangente hiperbólica, a qual o comportamento está representado na Figura 2.4.c e na Equação 2.4.

$$y(x) = x \tag{2.2}$$

$$y(x) = \frac{1}{1 + \exp(-x)} \tag{2.3}$$

$$y(x) = \tanh(x) = \frac{2}{1 + \exp(-2 \cdot x)} - 1 \quad (2.4)$$

A aplicação de algoritmos de treinamento ou aprendizagem às redes, ajusta os pesos e os *bias* entre as conexões dos neurônios, podendo assim estabelecer as relações existentes entre variáveis de entrada e variáveis de saída, associadas a processos ou funções em estudo.

II.2.1.3 Arquiteturas da RNA (HAYKIN, 2001)

Segundo Haykin, a maneira pela qual os neurônios estão estruturados está intimamente ligada com o algoritmo de aprendizagem usado para treinar a rede. Em geral, podem-se identificar três classes de arquiteturas de rede fundamentalmente diferentes:

- **Redes *Feedforward* com Camada Única:** tem-se uma camada de nós de fonte que se projeta sobre uma camada de saída de neurônios (nós computacionais), mas não vice-versa. Em outras palavras, esta rede é estritamente do tipo alimentada adiante ou acíclica. Está ilustrada na Figura 2.5a uma rede *feedforward* ou acíclica com uma única camada de neurônios.

- **Redes *Feedforward* com Múltiplas Camadas:** A segunda classe de uma rede neural *feedforward* se distingue pela presença de uma ou mais camadas intermediárias, cujos nós computacionais são chamados correspondentemente de neurônios ocultos ou unidades intermediárias.

Os nós de fonte da camada de entrada da rede fornecem os respectivos elementos do padrão de ativação (vetor de entrada), que constituem os sinais de entrada aplicados aos neurônios na segunda camada. Os sinais de saída da segunda camada são utilizados como entradas da terceira camada, e assim por diante para o resto da rede. Tipicamente, os neurônios em cada camada da rede têm como suas entradas apenas os sinais de saída da camada precedente. O conjunto de sinais de saída dos neurônios da camada final da rede constitui a resposta global da rede para o padrão de ativação fornecido pelos nós de fonte da camada de entrada. A Figura 2.5b apresenta um exemplo de rede *feedforward* com uma camada intermediária.

- **Redes recorrentes:** Uma rede neural recorrente se distingue de uma rede *feedforward* por um laço de realimentação. Este laço tem um impacto profundo na capacidade de aprendizagem da rede e no seu desempenho. Além disso, os laços de

realimentação envolvem o uso de ramos particulares compostos de elementos de atraso unitário, o que resulta em um comportamento dinâmico não-linear, admitindo-se que a rede neural contenha unidades não-lineares. A Figura 2.5c traz uma rede recorrente com neurônios ocultos.

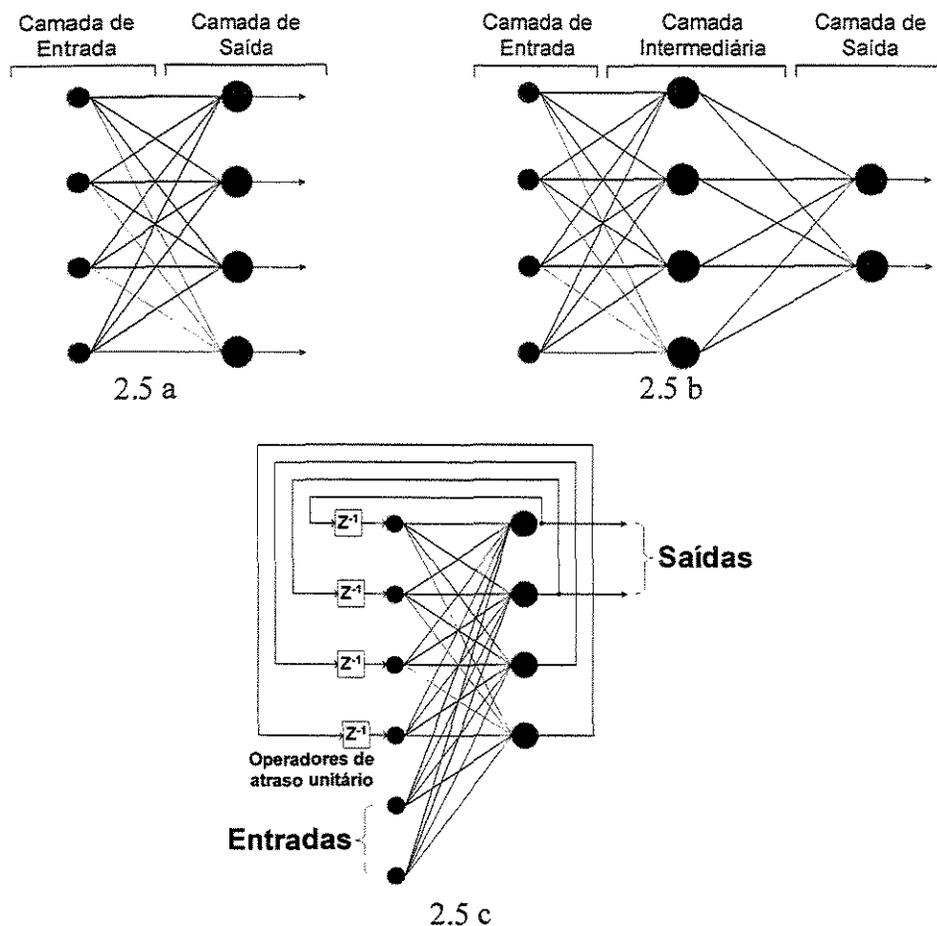


Figura 2.5: Exemplo de estruturas de redes neurais: a) Rede *feedforward* ou acíclica com uma camada de neurônios. b) Rede *feedforward* com uma camada intermediária. c) Rede recorrente com neurônios ocultos.

Fonte: HAYKIN (2001)

II.2.1.4 Processo de Treinamento (HAYKIN, 2001; BRAGA et al., 2003)

A propriedade que é de importância primordial para uma rede neural é a sua habilidade de aprender a partir de exemplos. Para um determinado conjunto de dados, o algoritmo de treinamento deve ser responsável pela adaptação dos parâmetros da rede, de maneira que, em um número finito de interações do algoritmo, haja convergência para uma

solução. O critério de convergência varia de acordo com o algoritmo e com o paradigma de treinamento, mas pode envolver, por exemplo, a minimização de uma função-objetivo, a variação do erro de saída ou mesmo a variação das magnitudes dos vetores de peso da rede. Basicamente, considera-se que o processo de aprendizado tem como característica a ocorrência de estímulo da rede pelo meio externo através da apresentação do conjunto de dados. Como consequência deste estímulo, o algoritmo de treinamento provoca mudanças nos parâmetros da rede e, finalmente, a mudança nos parâmetros acarreta em uma mudança no comportamento da rede. Espera-se que a mudança gradual no comportamento da rede resulte em uma melhoria gradativa de seu desempenho.

O objetivo do processo de aprendizado é então encontrar o ajuste do vetor de pesos para que o objetivo do treinamento que visa a convergência seja atingido. Os algoritmos de treinamento se diferem então pela maneira de obter o ajuste do vetor de pesos. Podem existir classificados basicamente três paradigmas distintos: aprendizado supervisionado, aprendizado não supervisionado e aprendizado por reforço.

O aprendizado supervisionado caracteriza-se pela existência de um professor, ou supervisor, externo à rede que tem a função de monitorar a resposta da mesma para cada valor de entrada. O aprendizado não-supervisionado caracteriza-se pela não existência de saídas desejadas para as entradas, sendo o conjunto de treinamento formado apenas pelos vetores de entrada. Esta forma de aprendizado aplica-se, tipicamente, a problemas de caracterização de dados. O aprendizado por reforço pode ser considerado um paradigma intermediário. O conjunto de treinamento é formado apenas pelos vetores de entrada, mas há existência de um crítico externo em substituição ao supervisor do aprendizado supervisionado. Como não há valores de saída desejados para a saída, o crítico em vez de retornar o erro de saída da rede, retorna um sinal de reforço ou penalidade associado à última ação da rede. Caso a ação tenha acarretado uma degradação no desempenho, ela será penalizada, ocorrendo com menor chance em situações futuras. Caso a ação tenha como resultado uma melhora de desempenho, ela deverá ser reforçada, ocorrendo com maior probabilidade em situações futuras. Os algoritmos de aprendizado por reforço visam então a maximização do reforço e a consequente melhora do desempenho.

Há muitos algoritmos de treinamento de redes *feedforward* com aprendizado supervisionado, mas alguns são mais utilizados, como os descritos abaixo.

II.2.1.4.1 Algoritmos de Treinamento

O algoritmo *backpropagation* é um algoritmo básico de treinamento, trata-se de uma técnica de gradiente e não de otimização. Consiste de dois passos através das diferentes camadas da rede: um passo para frente, a propagação, e um passo para trás, a retropropagação. No passo para frente, um vetor de entrada é aplicado aos neurônios da rede e seu efeito se propaga através da rede, camada por camada. Finalmente, um conjunto de saídas é produzido como resposta real da rede. Durante o passo de propagação, os pesos sinápticos são todos fixos. Durante o passo para trás, por outro lado, os pesos sinápticos são todos ajustados de acordo com uma regra de correção de erro. Especificamente, a resposta real da rede é subtraída de uma resposta desejada (alvo) para produzir um sinal de erro. Este sinal de erro é então propagado para trás através da rede, contra a direção das conexões sinápticas. Os pesos sinápticos são ajustados para fazer com que a resposta real da rede se mova para mais perto da resposta desejada, em um sentido estatístico (HAYKIN, 2001).

O ajuste dos pesos e das *bias* é obtido na direção em que a função-objetivo decresce mais rapidamente, ou seja, o negativo do gradiente (DEMUTH e BAELE, 2002). A equação de iteração deste algoritmo pode ser escrita como segue nas equações 2.5 e 2.6:

$$X_{k+1} = X_k + dX_k \quad (2.5)$$

Em que,

$$dX_k = mc \cdot dX_{prev} + lr \cdot (1 - mc) \cdot \frac{de}{dX} \quad (2.6)$$

onde:

k = iteração atual

X = vetor coluna de todos valores de pesos e *bias*.

dX_{prev} = é a variação anterior dos pesos e *bias*.

mc = Constante de momento.

lr = taxa de aprendizagem.

e = função-objetivo.

Tipicamente, a função-objetivo utilizada com o algoritmo *backpropagation* é a média do quadrado dos erros (e), apresentada na equação 2.6 (DEMUTH e BEALE, 2002):

$$e = \frac{1}{N \cdot n} \cdot \sum_{p=1}^N \sum_{k=1}^n (t_k^p - a_k^p)^2 \quad (2.7)$$

onde:

N = número de pares de vetores de entrada-saída

n = número de neurônios da camada de saída

t_k^p = valor desejado

a_k^p = valor de saída do k -ésimo neurônio do p -ésimo par entrada-saída calculado pela rede

Este algoritmo geralmente é lento para a aplicação em problemas práticos. Foram então desenvolvidos métodos de alto desempenho que convergem com uma velocidade muito maior que o algoritmo *backpropagation* (DEMUTH e BEALE, 2002).

Um exemplo destes métodos numéricos de otimização para treinamento da rede neural é o **algoritmo de Levenberg-Marquadt**, que foi feito para aproximar a velocidade de treinamento de segunda ordem sem ter que computar a matriz de *Hessian* (DEMUTH e BEALE, 2002). Sua equação de iteração pode ser escrita como apresentado na equação 2.8:

$$X_{k+1} = X_k - [J^T J + \mu I]^{-1} J^T e \quad (2.8)$$

onde:

J = matriz Jacobiana, contém as derivadas primeiras dos erros da rede com respeito aos pesos e *bias*.

$H = J^T J$ = **aproximação** da matriz de *Hessian*, contém as derivadas segundas dos erros da rede com respeito aos pesos e *bias*.

Quando o escalar μ é zero, o método de *Levenberg-Marquadt* é apenas o método de Newton, usando a aproximação da função de *Hessian*. Quando μ é grande, o método se torna o de gradiente descendente com um pequeno passo. O método de Newton é mais rápido e mais exato, próximo a um erro mínimo, assim o objetivo é deslocar para o método de Newton tão rápido quanto possível. Desta forma, o μ é diminuído depois de cada etapa

bem sucedida (redução na função-objetivo) e é aumentado apenas quando um passo de tentativa iria aumentar a função-objetivo. Desta forma, a função-objetivo será sempre reduzida em cada iteração do algoritmo. Este algoritmo geralmente é o mais rápido para treinamentos de redes neurais *feedforward* de tamanho moderado (DEMUTH e BEALE, 2002).

Um dos problemas que acontece durante o treinamento de uma rede neural é denominado **sobre-ajuste**. O erro obtido com o conjunto de treinamento é um valor muito pequeno, mas quando um conjunto de dados novo é apresentado à rede o erro obtido é grande. A rede memorizou os exemplos de treinamento, mas não aprendeu generalizar em situações novas (DEMUTH e BEALE, 2002).

Um método para melhorar generalização é usar uma rede que seja grande o suficiente para garantir um ajuste adequado. Quanto maior for a rede usada, mais complexas são as funções que a rede pode criar. Se a rede for suficientemente pequena ela não será capaz de fazer um bom ajuste dos dados (DEMUTH e BEALE, 2002). O objetivo do treinamento deve ser então o de encontrar o número de parâmetros ideal. No entanto, estimar o número de parâmetros é uma tarefa árdua que requer conhecimento sobre a complexidade do problema. Este conhecimento normalmente não está disponível, principalmente tratando-se de problemas multidimensionais, pois muitas vezes é este conhecimento que se deseja obter por meio do processo de modelagem (BRAGA et al., 2003).

Uma abordagem para melhorar a generalização, evitando o sobre-ajuste, é usar a **regularização**, que envolve a modificação da função-objetivo, que é normalmente escolhida para ser média dos erros quadrados da rede (eq 2.7) no conjunto de treinamento. No algoritmo de **Regularização Bayesiana** a função-objetivo assume a forma descrita na equação 2.9 (DEMUTH e BEALE, 2002):

Assumindo F como a função-objetivo:

$$F = \alpha \cdot SSE + \beta \cdot SSW \quad (2.9)$$

Onde:

SSE = somatório dos erros quadrados

SSW = somatórios do quadrado dos pesos e bias

α e β = parâmetros da função-objetivo

Os parâmetros de regularização são obtidos com a estrutura de *Bayesian*, que estima estes parâmetros usando técnicas estatísticas. Como a técnica de regularização bayesiana requer o cálculo da matriz de Hessian, o algoritmo de treinamento associado à regularização é o de *Levenberg – Marquardt* (HAGAN e FORESSE, 1997).

Uma das características deste algoritmo é dar uma medida de quantos parâmetros da rede (pesos e *bias*) estão efetivamente sendo usados por ela. Este número efetivo de parâmetros permanece aproximadamente constante, não importando quão grande é o número total de parâmetros da rede. (Isto assume que a rede tem sido treinada para um número de iterações suficientes para assegurar a convergência) (DEMUTH e BAELE, 2002).

Segundo Hagan e Foresse (1997), ao utilizar o algoritmo de regularização *bayesiana*, há algumas verificações que devem ser administradas. Primeiramente, se o valor de gama (número efetivo de parâmetros) for muito próximo ao valor de N (número atual de parâmetros), então provavelmente a rede não é grande o bastante para representar adequadamente a função real. Neste caso, é necessário aumentar o número de neurônios da camada intermediária e refazer o treinamento da rede. Se a rede maior tiver o mesmo valor de gama, então a rede menor é grande o suficiente. Senão, mais neurônios na camada intermediária devem ser adicionados. A outra verificação é de consistência de resultados. Se a rede é suficientemente grande, então uma segunda rede maior irá apresentar valores próximos de SSE (Somatório do quadrado dos erros), SSW (Somatório do quadrado dos pesos e *bias* da rede) e gama.

II.2.1.4.2 Critérios de parada (HAYKIN, 2001; DEMUTH e BEALE, 2002)

Em geral, não se pode demonstrar que o algoritmo de treinamento convergiu e não existem critérios bem-definidos para encerrar a sua operação. Em vez disso, há alguns critérios razoáveis, que podem ser usados para encerrar o ajuste dos pesos.

Cada apresentação de todo o conjunto de treinamento durante o processo de aprendizagem é chamada de **época**. O processo de aprendizagem é repetido época após época, até que ocorra qualquer destas condições:

- o conjunto de pesos e *bias* estabilize e a função-objetivo convirja para um valor mínimo;

- o número máximo de épocas seja atingido;
- o gradiente da função-objetivo caia abaixo de um valor mínimo;

No caso dos algoritmos Levenberg-Marquardt e Regularização Bayesiana há um critério de parada adicional:

- O parâmetro μ exceda a um valor máximo (este é um bom indicativo de que o algoritmo tenha verdadeiramente convergido)

II.2.1.4.3 Preparação dos dados de treinamento para o projeto de uma rede neural

A aplicação bem sucedida de uma rede neural requer um conjunto de dados apropriados. Há alguns critérios heurísticos para o pré-processamento dos dados que podem ser aplicados (SHAW e SIMÕES, 1999).

Uma avaliação importante é definir quais as variáveis de processo serão as entradas e saídas da rede. Além disso, os dados coletados devem ser significativos e cobrir amplamente o domínio do problema.

Em muitas situações as variáveis de entrada podem estar altamente correlacionadas sendo interessante diminuir a dimensão do modelo (SHAW e SIMÕES, 1999). Um processo efetivo para executar esta operação é a **análise do componente principal**. Esta técnica tem três efeitos: ortogonaliza os componentes do vetor de entrada (de forma que eles não sejam correlacionados com qualquer outro), ordena o resultado dos componentes ortogonalizados (componentes principais) de forma que aqueles com maior variação venham primeiro, e elimina aqueles componentes que contribuem menos no conjunto de dados (DEMUTH e BAELE, 2002).

Devido ao fato das redes neurais serem sensíveis à escala das variáveis, se a magnitude dos valores de entrada diferem entre si, a rede pode erroneamente atribuir uma maior importância a valores maiores (SHAW e SIMÕES, 1999), é usual normalizar as variáveis do conjunto de treinamento da rede neural dentro de um intervalo específico, como [-1;1] ou [0,1;0,9]. Há ainda outra maneira de determinar o intervalo de normalização definindo o valor zero para a média dos dados e o valor um para o desvio padrão (Média/DP), segundo a equação 2.10 (DEMUTH e BAELE, 2002):

$$pn = \frac{p - \text{mean}p}{\text{std}p} \quad (2.10)$$

onde:

p = Conjunto de dados

mean_p = média de p

std_p = desvio padrão de p

p_n = Conjunto p normalizado

Se este pré-processamento é aplicado, os valores da saída da rede treinada estarão dentro do intervalo de normalização, logo, precisam ser desnormalizados (DEMUTH e BAELE, 2002).

II.2.1.5 Aproximação de funções (HAYKIN, 2001)

Uma rede neural multicamadas treinada pode ser vista como um veículo prático para realizar um mapeamento não-linear de entrada-saída de natureza geral. Há um **Teorema da Aproximação Universal** que afirma que uma camada intermediária é suficiente para que uma rede neural multicamadas consiga uma aproximação uniforme para um dado conjunto de treinamento. Entretanto, o teorema não diz que a única camada intermediária é ótima no sentido do tempo de aprendizagem, facilidade de implementação ou (mais importante) generalização.

O teorema da aproximação universal assume que a função contínua a ser aproximada é dada e que está disponível uma camada intermediária de tamanho ilimitado para a aproximação. Estas duas suposições são violadas na maioria das aplicações práticas de RNAs multicamadas.

O problema com RNAs que usam uma única camada intermediária é que os neurônios desta camada tendem a interagir entre si globalmente. Em situações complexas, esta interação torna difícil melhorar a aproximação em um ponto sem piorá-la em algum outro ponto. Por outro lado, com duas camadas intermediárias o processo de aproximação (ajuste de curva) se torna mais gerenciável. Em particular, podemos proceder como segue (FUNAHASHI, 1989; CHESTER, 1990) :

- As *características locais* são extraídas na primeira camada intermediária. Especificamente, alguns neurônios da primeira camada intermediária são usados para dividir o espaço de entrada em regiões e outros neurônios naquela camada aprendem as características locais daquelas regiões;

- As *características globais* são extraídas na segunda camada intermediária. Especificamente, um neurônio na segunda camada intermediária combina as saídas de neurônios da primeira camada intermediária operando em uma região particular do espaço de entrada, e com isso aprende as características globais daquela região e fornece zero como saída nas outras regiões.

II.3 Conclusão

Uma das tendências mais recentes em *software* a serviço da automação industrial siderúrgica é a implantação de redes neurais artificiais, e há um número bastante limitado desta aplicação aos conversores a oxigênio. Tendo em vista que em conversores a oxigênio de aciarias integradas, há um problema complexo de baixa previsibilidade do teor de carbono e a temperatura do aço no final do sopro, o objetivo desta dissertação de mestrado é desenvolver modelos via redes neurais artificiais para a predição do período final de sopro de um conversor a oxigênio no processo de fabricação do aço, baseados em dados industriais do processo de refino. Já que uma maior previsibilidade do teor de carbono no final de sopro do processo de obtenção do aço implica em diminuir o tempo das bateladas do conversor, aumentando o número de bateladas diárias.

Em especial, no caso do conversor a oxigênio da CSN, há algumas particularidades que justificam a utilização de redes neurais artificiais como modelo dinâmico para o ajuste do oxigênio do resopro e/ou a quantidade de minério de ferro adicionada. Pode-se citar a disponibilidade de um banco de dados razoavelmente grande, uma quantidade de ruídos considerável, a grande variação química do ferro gusa e das demais matérias-primas do conversor e principalmente a dificuldade de utilizar um modelo matemático fenomenológico para realizar este controle.

CAPÍTULO III: METODOLOGIA

Este capítulo apresenta a estrutura do modelo neural para cada banco de dados (“Dados 1” e “Dados 2”), o critério de acerto para o modelo e também a metodologia dos testes que são aplicados aos bancos de dados para determinar o modelo neural mais adequado.

III.1 Estrutura do Modelo

O problema a ser estudado é o baixo acerto simultâneo de composição química e de temperatura do aço ao final do processo produtivo. O acerto simultâneo do modelo convencional utilizado pela CSN estava em torno de 44% e foi melhorado, recentemente para 66%, e com a aplicação do modelo dinâmico via redes neurais desenvolvido por Cunha (2001), o acerto simultâneo foi de 73,3%.

Como a proposta deste trabalho é tentar melhorar o modelo dinâmico para o processo de refino via redes neurais obtido por Cunha, a princípio, o banco de dados (Dados I) utilizado neste trabalho é exatamente o mesmo utilizado por Cunha(2001).

Para a simulação do modelo dinâmico de Cunha(2001) foram consideradas 5030 bateladas (filtradas de 15.000 bateladas) para treinamento e 200 bateladas para testes. Este conjunto de dados equivale à cerca de 4 anos de processamento na aciaria da CSN. São três conversores a oxigênio trabalhando continuamente, resultando em aproximadamente 70 bateladas por dia. Cada conversor apresenta um volume de 190 m³. Em cada batelada o conversor recebe uma carga total (gusa e sucata) de 245 toneladas para produzir 220 toneladas de aço.

As variáveis de entrada deste modelo são determinadas a partir da disponibilidade dos dados à época. Desta forma são definidas quatro variáveis de entrada: a temperatura do banho medida pela sublança (TIB – temperatura *in blow*), a porcentagem de carbono do banho calculado pela temperatura de solidificação, através do diagrama Fe-C, obtida pela amostra da sublança (CIB – porcentagem de carbono *in blow*) e as quantidades de minério de ferro (MFe) e de oxigênio (Oxig) determinadas pelo modelo vigente e implementadas pelo operador para o restante do sopro. As variáveis de saída são as mesmas do modelo dinâmico comercial já existentes na CSN: a porcentagem de carbono (CFS) e a temperatura do aço (TFS) no período de fim de sopro. A Figura 3.1 ilustra a estrutura neural utilizada. Importantes variáveis, como altura da lança ao longo do sopro, a composição química dos

gases e fumos e a quantidade e composição química da escória e do banho não estão disponíveis para a modelagem. Isto introduz limitações às previsões por redes neurais ou por qualquer modelo convencional alternativo.

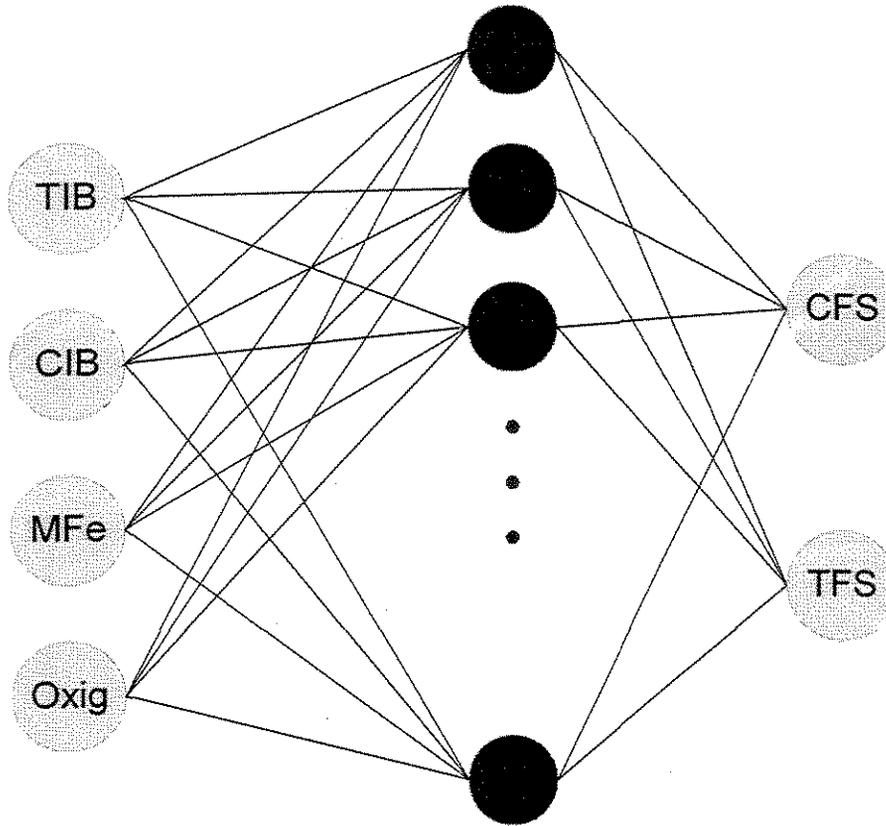


Figura 3.1: Estrutura Neural de previsão de C e T para o Modelo Dinâmico – Dados 1
Fonte: CUNHA (2001)

Este banco de dados ('Dados 1') apresenta valores de bateladas em que ocorreu a adição ou não de minério de ferro.

Também foi obtido através da CSN um outro banco de dados (Dados 2). Os dados aqui apresentados constitem como entrada da rede: a porcentagem de carbono (CIB), e a temperatura (TIB) medidas com 88% de sopro, as quantidades de minério de ferro (MFe) e de oxigênio (Oxig) utilizadas para o restante do sopro, as quantidades de ferro-gusa (Gusa), de sucata (Suc), e de escória (Esc) carregados no início do sopro, e as porcentagens de manganês (MnIB), fósforo (PIB) e a de enxofre (SIB) tomadas para medição com 88% de sopro. Já a saída da rede foi composta pelas seguintes variáveis: a porcentagem de carbono (CFS), a temperatura do aço (TFS), a porcentagem de manganês (MnFS), de fósforo (PFS) e a de enxofre (SFS) medidas no fim de sopro.

Este novo banco de dados é dividido em 1844 vetores entrada-saída para o conjunto de treinamento e 400 vetores entrada-saída para o conjunto de teste. E só contém valores de bateladas em que houve a adição de minério de ferro. A Figura 3.2 ilustra a nova estrutura de rede neural para o modelo dinâmico:

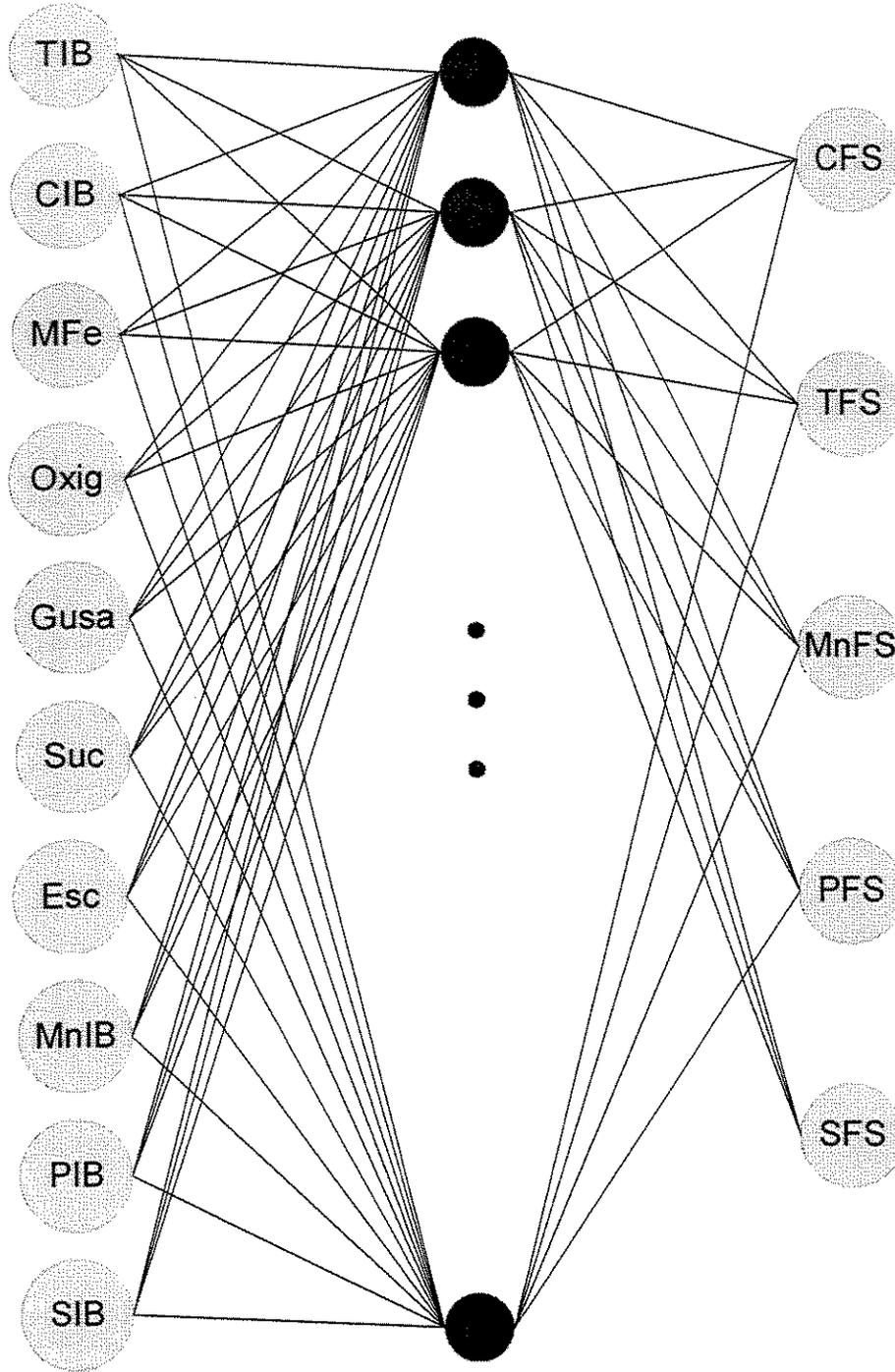


Figura 3.2: Estrutura Neural de previsão de C e T para o Modelo Dinâmico – Dados 2

O algoritmo de treinamento utilizado por Cunha (2001) foi o *backpropagation*. No entanto, a proposta deste trabalho é testar outros métodos de treinamento, como o Método de *Levenberg-Marquardt* puro e o Método de *Levenberg-Marquardt* com o algoritmo de Regularização Bayesiana. Além de testar outros detalhes da rede, como função de transferência, o intervalo de normalização e a aleatoriedade dos dados, entre outros.

A estrutura de rede que Cunha obteve o melhor resultado para o modelo dinâmico foi uma rede *feedforward* multicamada, com duas camadas escondidas e topologia 4x9x3x2 e 4x3x3x2. As funções de ativação utilizadas nos cálculos internos dos neurônios foram do tipo sigmoidal. Assim sendo, como o objetivo do presente trabalho é encontrar uma estrutura que apresente um melhor resultado de acerto simultâneo de porcentagem de carbono e temperatura no fim de sopro, são realizados testes alterando as características da rede neural obtida por Cunha. Estes testes são propostos devido às dificuldades de trabalhar com um banco de dados industriais, que normalmente passam despercebidas em estudos de redes neurais, onde os banco de dados são "bem comportados".

A ferramenta computacional para modelagem do modelo dinâmico via RNAs utilizada é a *Neural Network Toolbox*, do *Software MATLAB*[®], uma ferramenta computacional que compreende um conjunto de funções implementadas para o desenvolvimento de redes neurais. O objetivo do trabalho não é a criação de uma nova ferramenta (programa computacional) para aplicação de redes neurais, mas sim sua aplicação direta ao processo de aciaria. O código da RNA desenvolvido com o *Software MATLAB*[®] segue disposto no Anexo 1.

III.2 Critério de acerto

O critério de acerto da rede neural aqui adotado é o mesmo utilizado pela CSN, ou seja, o intervalo de acerto para o desvio da variável de saída referente à porcentagem de carbono no final do sopro é $[-0,01; 0,02]\%$ de carbono e para o desvio da variável referente à temperatura do aço no final do sopro é $\pm 15^{\circ}\text{C}$. Desta forma, o acerto simultâneo ocorre quando a resposta do modelo dinâmico de ambas as variáveis estão dentro de seus respectivos intervalos de acerto citados acima. Este critério é aplicado aos resultados da generalização, ou seja, da simulação do conjunto de teste. E então é calculada a porcentagem dos acertos do conjunto de teste.

III.3 Testes

Neste item são apresentados os testes referentes às diferentes formas de normalização, topologia, funções de ativação, algoritmos de treinamento para diferentes conjuntos de dados, com a avaliação direta do critério de acerto já apresentado.

III.3.1 Normalização

Como já foi mencionado no capítulo anterior, é usual normalizar as variáveis do conjunto de treinamento da rede neural dentro de um intervalo específico. Assim sendo, são testados intervalos de $[-1; 1]$, $[0,1;0,9]$ e o intervalo no qual a média dos dados de cada vetor receba o valor zero e o desvio padrão o valor 1 (Média/DP - como descrito na equação 2.10). A estrutura da rede testada apresenta como algoritmo de treinamento o Método de *Levenberg-Marquardt* com o algoritmo de Regularização Bayesiana, e como funções de ativação as funções “logsig” e linear para as camadas intermediárias e de saída, respectivamente. A escolha deste algoritmo é feita desta forma, pois segundo DEMUTH e BEALE (2002), este algoritmo melhora a generalização da rede, evitando um sobre-ajuste devido a uma modificação na função-objetivo. A literatura também faz referência de que para a camada de saída a melhor função de ativação é a linear.

III.3.2 Topologia

Este teste tem como finalidade determinar a topologia mais adequada para a rede. A princípio varia-se o número de neurônios de uma única camada intermediária. Numa segunda etapa o teste é realizado com duas camadas intermediárias. A rede neural testada é treinada com o algoritmo de *Levenberg-Marquardt* com Regularização Bayesiana, usando o melhor intervalo de normalização obtido no teste acima e as funções de ativação “logsig”/linear ou “logsig”/“logsig”/linear, para uma ou duas camadas intermediárias, respectivamente.

III.3.3 Função de ativação

Para encontrar a função de ativação que melhor se adapta aos dados nas camadas intermediária e final são realizados testes com as funções de transferência log-sigmoidal, tangente-sigmoidal e linear como segue na tabela 3.1. A estrutura de rede utilizada é com algoritmo de *Levenberg-Marquardt* com Regularização Bayesiana como algoritmo de treinamento, o intervalo de normalização e a topologia dos melhores resultados obtidos com os testes 3.1 e 3.2, respectivamente.

Tabela 3.1: Combinação das várias opções de funções de ativação aplicadas nas camadas da rede neural para o modelo dinâmico.

Combinação	Camada Intermediária	Camada de Saída
1	Logsig	Linear
2	Tansig	Linear
3	Logsig	Logsig
4	Tansig	Tansig

Onde:

Logsig = função de transferência log-sigmoidal – Eq 2.3

Tansig = função de transferência tangente-sigmoidal – Eq 2.4

Linear = função de transferência linear Eq 2.2

III.3.4 Algoritmo de treinamento

Para avaliar o algoritmo de treinamento que resulta em um melhor resultado para a RNA é realizado um teste com os algoritmos de *Levenberg-Marquardt* com Regularização Bayesiana, *Levenberg-Marquardt* e *Backpropagation* puros. A estrutura de rede testada apresenta as características de normalização, topologia e ativação que obtiveram um maior acerto simultâneo de porcentagem de carbono e temperatura do aço, obtido nos testes anteriores.

Para a função de treinamento *Backpropagation* a função-objetivo aplicada é a média do somatório do quadrado do erro (MSE) e não o somatório do quadrado dos erros (SSE), como nos demais treinamentos algoritmos deste trabalho.

III.3.5 Conjunto de dados

Uma vez determinada a melhor estrutura da rede (intervalo de normalização, função de ativação, topologia e algoritmo de treinamento), pode-se explorar diferentes opções de trabalho para os conjuntos de treinamento e teste:

1) Um teste interessante é avaliar o acerto da rede tornando os dados do conjunto de treinamento e de teste aleatórios, sem seqüência cronológica;

2) Outra opção é redividir os conjuntos de treinamento e teste, diminuindo a quantidade de vetores entrada-saída do primeiro e aumentando a quantidade de vetores do segundo. Como para este teste é necessário variar a quantidade de vetores dos conjuntos, é

importante verificar se o conjunto de teste representa bem, e em toda sua extensão, o conjunto de treinamento;

3) Outro teste que pode ser realizado é aplicar a **análise do componente principal (Prepca)** para avaliar se é possível reduzir a dimensão do modelo neural;

4) Uma técnica muito empregada em trabalhos experimentais é a análise estatística via desvio padrão (σ). Desta maneira, é sugerido um teste com conjuntos de dados que são compostos por valores que estejam dentro dos intervalos de $[m\acute{e}dia \pm 2\sigma]$ e $(m\acute{e}dia \pm 1\sigma)$. No entanto, como o objetivo deste trabalho é a aplicação industrial e os dados industriais não se comportam desta forma idealizada, o modelo neural obtido não teria aplicação prática confiável, uma vez que a rede neural não realiza extrapolações seguras;

5) Como o novo conjunto de dados ‘Dados 2’ apresenta seis novas variáveis de entrada (total de dez variáveis) e três novas de saída (total de cinco), foi proposto um teste para avaliar a influência da adição destas novas variáveis na generalização. Com este propósito, são realizados treinamentos e simulações de modelos neurais que apresentavam as quatro variáveis do primeiro conjunto de dados (CIB, TIB, MFe e Oxig) e as duas variáveis de saída (CFS e TFS) acrescidas das novas variáveis do novo conjunto de dados. É definida uma única camada intermediária com uma quantidade de neurônios (M) fixa. Os modelos testados estão descritos na Tabela 3.2.

III.4 Conclusão

A partir dos bancos de dados ‘Dados 1’ e ‘Dados 2’, são propostos testes para o modelo neural de normalização, topologia, função de ativação, algoritmo de treinamento, alterações no banco de dados e modificação dos vetores de entrada e saída. A decisão do modelo mais apropriado é realizada a partir de um critério que avalia o acerto simultâneo de porcentagem de carbono e de temperatura, como é feito na CSN. Os resultados destes testes são apresentados no capítulo seguinte.

Tabela 3.2: Modelos neurais testados para avaliar a influência da adição das novas variáveis do conjunto de dados ‘Dados 2’

Topologia	Vetores de entrada da RNA	Vetores de saída da RNA
4xMx2	TIB, CIB, MFe, OIB	
5xMx2	TIB, CIB, MFe, OIB, Gusa	
6xMx2	TIB, CIB, MFe, OIB, Gusa, Suc	
7xMx2	TIB, CIB, MFe, OIB, Gusa, Suc, Esc	CFS,TFS
8xMx2	TIB, CIB, MFe, OIB, Gusa, Suc, Esc, MnIB	
9xMx2	TIB, CIB, MFe, OIB, Gusa, Suc, Esc, MnIB, PIB	
10xMx2	TIB, CIB, MFe, OIB, Gusa, Suc, Esc, MnIB, PIB, SIB	
8xMx3	TIB, CIB, MFe, OIB, Gusa, Suc, Esc, MnIB	
9xMx3	TIB, CIB, MFe, OIB, Gusa, Suc, Esc, MnIB, PIB	CFS,TFS,MnFS
10xMx3	TIB, CIB, MFe, OIB, Gusa, Suc, Esc, MnIB, PIB, SIB	
9xMx4	TIB, CIB, MFe, OIB, Gusa, Suc, Esc, MnIB, PIB	CFS,TFS,MnFS,PFS
10xMx4	TIB, CIB, MFe, OIB, Gusa, Suc, Esc, MnIB, PIB, SIB	
10xMx5	TIB, CIB, MFe, OIB, Gusa, Suc, Esc, MnIB, PIB, SIB	CFS,TFS,MnFS,PFS,SFS

CAPÍTULO IV: RESULTADOS E DISCUSSÕES

Utilizando a metodologia proposta no Capítulo 3, este capítulo apresenta uma avaliação dos resultados dos testes aplicados aos conjuntos de dados 1 e 2. A avaliação é realizada com base no acerto simultâneo da porcentagem de carbono e da temperatura do aço na simulação dos conjuntos de teste.

IV.1 Dados 1

Conforme descrito no Capítulo 3, o primeiro banco de dados a ser trabalhado é o mesmo utilizado por Cunha, logo o conjunto de treinamento é constituído de 5030 vetores entrada-saída e o conjunto de teste de 200 vetores entrada-saída. São definidas quatro variáveis de entradas: a temperatura do banho medida pela sublança (TIB – temperatura *in blow*), o percentual de carbono do banho calculado pela temperatura de solidificação, através do diagrama Fe-C, obtida pela amostra da sublança (CIB – porcentagem de carbono *in blow*) e as quantidades de minério de ferro (MFe) e de oxigênio (Oxig) utilizadas para o restante do sopro. As variáveis de saída são as mesmas do modelo dinâmico já existentes na CSN, o teor de carbono (CFS) e a temperatura do aço (TSF) no período de fim de sopro (Figura 3.1).

Ao banco de dados 1 são aplicados os testes descritos no capítulo anterior.

IV.1.1 Normalização

Para encontrar o intervalo de normalização mais adequado para os dados é realizado um teste de normalização com uma rede de topologia 4x3x3x2, uma das melhores topologias de Cunha. Os resultados obtidos neste teste se encontram na Tabela 4.1.

Tabela 4.1: Resultados do teste do intervalo de normalização para o pré-processamento dos dados da RNA do modelo dinâmico para período de fim de sopro na aciaria.

Teste	Resultados do Treinamento						Resultados do Teste			
	SSE	SSW	N	γ	aC(%)	aT(%)	aS(%)	aC(%)	aT(%)	aS(%)
-1 a 1	1220,37	112,92	35	30,7	83,92	79,72	67,46	88,00	85,50	75,00
0,1 a 0,9	194,83	301,69	35	29,9	83,92	79,72	67,46	87,00	85,50	74,00
Média/DP	7213,29	76,55	35	30,1	84,10	79,56	67,28	88,00	84,00	73,50

Legenda:

SSE = Somatório do quadrado dos erros

SSW = Somatório do quadrado dos pesos e *bias*

N = número atual de parâmetros da RNA

γ = número de parâmetros efetivos da RNA

aC = acerto do teor de carbono (%)

aT = acerto da temperatura no fim do sopro (%)

aS = acerto simultâneo (%)

Como o critério para avaliar o desempenho da rede é o acerto simultâneo da porcentagem de carbono e da temperatura do aço no final do sopro, avaliando os resultados apresentados na Tabela 4.1, é possível observar que o intervalo de normalização [-1;1] é o que apresenta um acerto simultâneo maior (aS = 75,00%) para as condições trabalhadas.

IV.1.2 Topologia

Com o objetivo de confirmar a topologia mais adequada para este banco de dados são realizados testes com uma e duas camadas intermediárias na estrutura da rede. Os resultados estão apresentados na Tabela 4.2.1 para uma camada intermediária.

Tabela 4.2.1: Resultados do teste para definir a melhor topologia com uma camada intermediária para a RNA de previsão do acerto simultâneo da porcentagem de carbono e temperatura para período de fim de sopro na aciaria.

Teste	Resultados do Treinamento								Resultados do Teste				
	Topologia_1	SSE	MSE	SSW	N	γ	aC(%)	aT(%)	aS(%)	SSE	MSE	aC(%)	aT(%)
4x1x2	1339,15	0,27	27,90	9	8,90	82,66	79,46	65,59	262,11	1,31	85,00	86,00	73,00
4x2x2	1235,39	0,25	72,88	16	15,70	84,00	79,80	67,50	281,97	1,41	87,00	87,00	76,00
4x3x2	1223,89	0,24	70,69	23	21,50	84,16	79,66	67,51	285,57	1,43	87,00	84,50	73,00
4x4x2	1220,84	0,24	108,67	30	26,80	84,16	79,64	67,51	285,30	1,43	87,00	85,00	73,50
4x5x2	1219,33	0,24	91,94	37	32,80	84,12	79,76	67,51	285,31	1,43	88,00	84,50	74,00
4x6x2	1217,52	0,24	101,72	44	38,00	84,27	79,84	67,71	285,33	1,43	86,50	86,00	74,50
4x7x2	1215,71	0,24	97,74	51	43,20	84,21	79,74	67,50	285,97	1,43	87,00	86,00	74,50
4x8x2	1212,84	0,24	118,48	58	49,90	84,21	80,12	67,87	285,19	1,43	88,00	86,50	76,00
4x9x2	1211,97	0,24	112,65	65	54,00	84,27	79,98	67,81	284,78	1,42	88,50	86,50	76,50
4x10x2	1210,77	0,24	129,86	72	57,50	84,19	79,86	67,63	284,57	1,42	88,00	86,00	75,50
4x11x2	1208,54	0,24	148,92	79	64,00	84,21	80,10	67,83	284,19	1,42	88,00	85,50	75,00
4x12x2	1208,51	0,24	150,12	86	64,60	84,18	80,10	67,81	283,67	1,42	88,00	85,50	75,00

Analisando os resultados de acerto simultâneo (aS) apresentados na Tabela 4.2.1, é possível verificar que o acerto simultâneo do conjunto de teste apresentou-se maior do que o acerto simultâneo do conjunto de treinamento. Desta forma, é proposta uma análise da média do somatório do quadrado do erro (MSE), que seria uma medida do desvio dos resultados obtidos pelo modelo neural e dos resultados esperados. Esta análise se baseia no fato de que o MSE do conjunto de treinamento deve ser menor que o MSE do conjunto de teste, ou seja, o desvio dos resultados deve ser maior para o conjunto de teste do que o desvio do treinamento. Avaliando os resultados de MSE apresentados na tabela 4.2.1 pode-se verificar, como era esperado, que o MSE do treinamento é menor que o MSE do teste, logo, há uma coerência nos desvios dos resultados. Diante desta evidência, verifica-se que a diferença no comportamento destes critérios, o acerto simultâneo e o MSE, deve-se ao fato de que o primeiro faz uma análise vetor a vetor se houve ou não acerto e o segundo mede qual é a dimensão média do desvio dos resultados.

Também são avaliados os valores do acerto simultâneo na Tabela 4.2.1, sendo verificado que a melhor topologia para a rede com uma camada intermediária é $4 \times 9 \times 2$ (aS = 76,50%). No entanto, ao observar os critérios sugeridos por Hagan (1997), já mencionados no Capítulo 2, a melhor topologia seria a $4 \times 11 \times 2$ (aS = 75,00%) pois é com esta topologia que os parâmetros efetivos da rede (γ) começam a se tornar constantes.

Para encontrar a melhor topologia com duas camadas intermediárias são realizados testes em que o número de neurônios da primeira camada intermediária é mantido fixo e variou-se o número de neurônios da segunda camada intermediária até ser atingido o critério sugerido por Hagan, ou seja, até a estabilidade dos valores de γ , SSE e SSW. Os resultados se encontram na Tabela 4.2.2.

Tabela 4.2.2: Resultados do teste para definir a melhor topologia com duas camadas intermediárias para a RNA do modelo dinâmico

Teste	Resultados do Treinamento						Resultados do Teste				
	Topologia_2	SSE	SSW	γ	gama	aC(%)	aT(%)	aS(%)	aC(%)	aT(%)	aS(%)
4x1x1x2	1339,2	35,6	11	9,02	82,66	79,38	65,53	85,00	86,00	73,00	
4x1x2x2	1337,7	57,4	15	10,80	82,66	79,54	65,63	84,50	85,50	72,00	
4x1x3x2	1337,9	49,9	19	10,60	82,68	79,58	65,71	84,50	86,00	72,50	
4x2x1x2	1333,2	49,4	17	15,30	82,66	79,46	65,73	85,00	86,00	73,00	
4x2x2x2	1230,9	91,0	22	20,30	83,96	79,66	67,30	87,00	85,50	74,50	
4x2x3x2	1228,9	147,5	27	22,30	84,18	79,66	67,46	87,00	85,00	74,00	
4x2x4x2	1226,1	197,5	32	25,10	84,21	79,60	67,44	86,50	85,00	73,50	
4x2x5x2	1227,8	145,3	37	25,60	84,18	79,76	67,51	86,50	86,00	74,50	
4x3x1x2	1331,0	64,8	23	20,20	82,56	79,54	65,65	85,00	85,50	72,50	
4x3x2x2	1222,1	123,6	29	25,50	84,14	79,92	67,67	87,50	85,00	74,00	
4x3x3x2	1218,7	178,0	35	27,80	84,10	79,78	67,55	87,50	85,50	74,50	
4x3x4x2	1218,8	209,2	41	33,50	84,29	79,94	67,89	87,50	84,50	73,50	
4x3x5x2	1216,3	291,9	47	40,00	84,14	79,90	67,65	87,00	85,50	74,50	
4x3x6x2	1213,1	263,4	53	45,50	84,21	80,16	68,09	88,50	86,50	76,50	
4x3x7x2	1214,9	230,8	59	44,00	84,19	79,92	67,75	87,50	86,00	75,00	
4x4x1x2	1329,5	64,4	29	24,30	82,58	79,76	65,92	85,00	86,00	73,00	
4x4x2x2	1218,7	144,8	36	30,50	84,18	79,72	67,53	87,50	85,00	74,00	
4x4x3x2	1217,7	184,6	43	36,00	84,06	80,26	67,93	87,50	85,00	74,00	
4x4x4x2	1216,9	157,8	50	40,10	84,14	79,94	67,77	88,00	85,50	75,00	
4x4x5x2	1213,3	242,9	57	45,00	84,19	79,84	67,79	88,00	86,00	75,50	
4x4x6x2	1211,2	250,4	64	53,50	84,21	79,90	67,83	88,00	86,00	75,50	
4x4x7x2	1212,3	237,8	71	52,00	84,25	80,00	67,81	88,00	85,00	74,50	
4x5x1x2	1326,9	82,4	35	29,60	82,58	79,88	65,98	85,00	85,50	72,50	
4x5x2x2	1216,7	138,5	36	43,00	84,19	79,92	67,71	87,50	86,50	75,50	
4x5x3x2	1214,1	194,9	42	51,00	84,18	79,64	67,46	88,00	85,00	74,50	
4x5x4x2	1208,4	304,8	59	51,20	83,98	80,06	67,85	87,50	87,00	76,00	
4x5x5x2	1208,0	316,6	67	55,20	84,45	79,90	67,81	88,00	84,50	74,00	
4x5x6x2	1204,8	381,7	75	64,80	84,33	80,04	67,93	88,00	85,50	75,00	
4x5x7x2	1204,7	349,6	83	65,30	84,25	80,10	67,89	88,00	86,50	76,00	
4x6x1x2	1325,6	109,1	41	33,60	82,60	79,60	65,79	85,00	86,50	73,50	
4x6x2x2	1215,9	145,0	50	41,00	84,29	79,76	67,55	88,00	86,50	76,00	
4x6x3x2	1210,9	194,7	59	50,50	84,21	79,80	67,67	88,00	85,50	75,00	
4x6x4x2	1207,5	315,5	68	57,20	84,23	80,00	67,97	88,50	86,00	76,00	
4x6x5x2	1205,0	308,4	77	60,50	84,18	80,36	68,13	87,50	86,00	75,00	
4x6x6x2	1203,1	340,7	86	74,70	84,16	80,22	67,99	88,50	85,50	75,50	
4x6x7x2	1202,3	370,3	95	81,40	84,08	80,26	67,91	87,50	84,50	73,50	
4x6x8x2	1204,4	318,1	104	82,80	84,25	80,12	67,93	88,00	84,50	74,00	
4x7x1x2	1324,8	104,7	47	37,00	82,56	79,82	65,90	85,00	85,50	72,50	
4x7x2x2	1212,9	206,0	57	48,50	84,16	79,98	67,71	87,50	86,00	75,00	
4x7x3x2	1211,1	178,9	67	54,40	84,21	80,06	67,87	87,50	85,50	74,50	
4x7x4x2	1203,7	385,3	77	64,90	84,29	79,90	67,83	87,50	86,00	75,50	
4x7x5x2	1204,0	329,6	87	69,70	84,10	80,30	68,03	87,00	84,50	73,00	
4x7x6x2	1206,1	254,3	97	73,90	84,29	80,08	67,91	88,00	85,00	74,50	
4x7x7x2	1200,0	381,8	107	89,50	84,18	80,26	67,95	87,50	85,00	74,00	
4x7x8x2	1200,0	372,6	117	90,30	84,29	79,96	67,79	88,00	85,00	74,50	
4x8x1x2	1323,1	146,0	53	40,40	82,60	79,88	66,06	85,00	85,50	72,50	
4x8x2x2	1211,9	169,8	64	52,40	84,14	80,16	67,75	87,50	85,00	74,00	
4x8x3x2	1210,1	185,9	75	58,60	84,21	79,82	67,55	88,00	86,00	75,50	
4x8x4x2	1209,8	185,1	86	61,80	84,19	80,00	67,81	88,00	86,00	75,50	
4x8x5x2	1205,4	285,3	97	75,70	84,08	80,34	67,99	87,50	84,00	73,50	
4x8x6x2	1205,1	244,4	108	76,20	84,19	79,98	67,61	88,00	85,00	74,50	
4x9x1x2	1324,6	105,5	59	37,70	82,56	79,90	65,92	85,00	85,50	72,50	
4x9x2x2	1210,0	187,1	71	58,10	84,14	80,04	67,67	88,00	84,50	74,00	
4x9x3x2	1207,5	241,0	83	65,60	84,25	79,98	67,71	88,00	86,00	75,50	
4x9x4x2	1205,7	267,6	95	71,90	84,25	79,94	67,73	87,50	85,50	74,50	
4x9x5x2	1202,7	309,1	107	80,70	84,19	79,92	67,79	88,50	85,00	75,00	

A Tabela 4.2.2 mostra que a topologia do modelo neural com duas camadas intermediárias que apresenta o maior acerto simultâneo de porcentagem de carbono e temperatura é $4 \times 3 \times 6 \times 2$ ($aS = 76,50\%$) e esta topologia também coincide a melhor topologia segundo o critério de Hagan. Assim, a Figura 4.1 mostra um gráfico que compara os valores do número total de parâmetros da rede (N) e o número de parâmetros efetivos (γ) em função do aumento do número de neurônios da segunda camada intermediária, quando há 3 neurônios na primeira camada intermediária da rede.

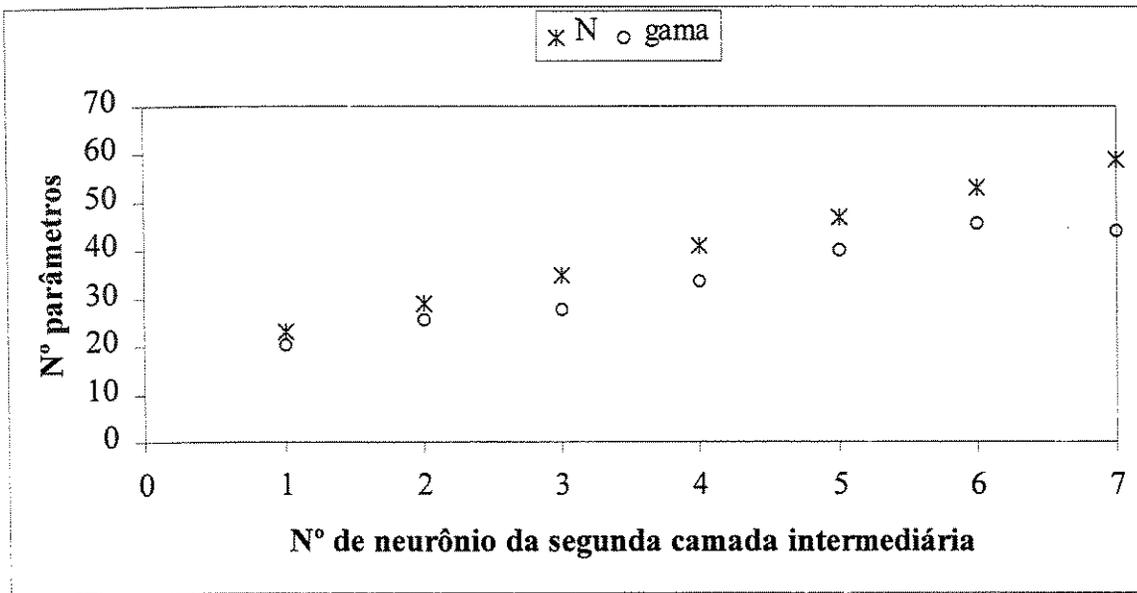


Figura 4.1: Comparação da quantidade de parâmetros totais da rede com o número de parâmetros efetivos a partir da variação da quantidade de neurônios da segunda camada intermediária, quando há 3 neurônios na primeira camada intermediária do modelo neural.

Analisando a Figura 4.1 é possível verificar que com um número baixo de neurônios na camada intermediária o valor de N e γ se apresentam muito próximos. Na medida em que o número de neurônios na camada intermediária aumenta, a diferença entre N e γ também aumenta até que o valor de γ permaneça constante. Assim, a topologia que apresenta uma maior consistência nos resultados é a $4 \times 3 \times 6 \times 2$. Esta topologia também é aquela que apresenta o maior acerto simultâneo ($aS = 76,50\%$).

Comparando esta topologia ($4 \times 3 \times 6 \times 2$) em relação à melhor topologia obtida por Cunha ($4 \times 3 \times 3 \times 2$), observa-se uma grande proximidade entre as duas topologias e consideram-se os resultados coerentes. Levando-se em conta que os vetores de *bias* são

ajustados no treinamento e são inicializados de forma randômica no presente trabalho e que no trabalho de Cunha estes vetores foram mantidos constantes e iguais a um, alguma discrepância seria esperada.

Observa-se também que a soma do número de neurônios intermediários da melhor rede com duas camadas intermediárias é similar ao número de neurônios intermediários da melhor rede com uma camada intermediária e portanto, a rede com apenas uma camada é escolhida para os testes que se seguem.

IV.1.3 Função de ativação

Com o intuito de verificar qual a melhor combinação de funções de ativação que fornece um maior acerto simultâneo para o modelo dinâmico, é realizado um teste com uma rede de topologia 4x9x2 (melhor resultado obtido com o teste anterior). Os resultados estão expostos na Tabela 4.3.

Tabela 4.3: Resultados do teste de função de ativação para as camadas intermediária e final da RNA do modelo dinâmico para período de fim de sopro na aciaria.

Teste	Resultados do Treinamento						Resultados do Teste			
	SSE	SSW	N	γ	aC(%)	aT(%)	aS(%)	aC(%)	aT(%)	aS(%)
logsig/linear	1212,02	111,78	65	53,90	84,31	79,94	67,79	88,50	86,50	76,50
tansig/linear	1211,48	34,22	65	53,90	84,23	80,16	68,01	88,00	86,50	76,00
logsig/logsig	1897,30	404,54	65	32,20	83,06	75,19	62,72	81,00	82,50	68,50
tansig/tansig	1209,33	54,72	65	56,90	84,12	79,86	67,48	87,50	85,00	74,50

Onde:

Logsig = função de transferência log-sigmoidal – Eq. 2.3

Tansig = função de transferência tangente hiperbólica – Eq. 2.4

Linear = função de transferência linear – Eq. 2.2

Os resultados apresentados na Tabela 4.3 sugerem que a combinação de funções de transferência “logsig” e linear para as camadas intermediária e de saída, respectivamente, é aquela que fornece um maior acerto simultâneo de porcentagem de carbono e temperatura (aS = 76,50%).

IV.1.4 Algoritmo de treinamento

Para avaliar o algoritmo de treinamento que apresenta um melhor resultado para a RNA foi feito um teste com as funções de treinamento *Levenberg-Marquardt* com Regularização Bayesiana (BR), *Levenberg-Marquardt* puro (LM) e *Backpropagation* puro (Backpro.). A topologia 4x9x2, as funções de ativação “logsig” e linear para as camadas intermediária e final, respectivamente, são utilizadas na rede testada. Para a função de treinamento *Backpropagation* são utilizados os melhores parâmetros empregados por Cunha de taxa de aprendizagem e momento igual a 0,7 e a função-objetivo aplicada é a média do somatório do quadrado do erro (MSE) e não o somatório do quadrado dos erros (SSE), como nos demais treinamentos apresentados neste trabalho. Os resultados deste teste estão expostos na Tabela 4.4.

Tabela 4.4: Resultados do teste para definir o algoritmo de treinamento que resulta em um maior acerto simultâneo para o modelo neural.

Teste	Resultados do Treinamento						Resultados do Teste				
	Algor. Trein.	MSE	SSE	SSW	N	γ	aC(%)	aT(%)	aS(%)	aC(%)	aT(%)
BR	---	1212,00	111,90	65	53,70	84,31	79,90	67,75	88,50	86,50	76,50
LM	---	1206,41	---	65	---	84,08	80,18	67,73	87,50	84,00	73,00
Backpro.	0,1223	---	---	65	---	83,82	79,50	67,20	87,50	86,50	75,50

A Tabela 4.4 apresenta resultados que indicam que o algoritmo de treinamento que apresenta o maior acerto simultâneo de porcentagem de carbono e temperatura do aço é o algoritmo de treinamento de *Levenberg-Marquardt* com Regularização Bayesiana (aS = 76,50%).

Com base na literatura, este resultado era esperado uma vez que este algoritmo de treinamento de *Levenberg-Marquardt* com Regularização Bayesiana apresenta uma modificação na função-objetivo que procura evitar o sobreajuste dos dados. Assim sendo, verifica-se um acerto simultâneo bastante similar para os dados de treinamento e uma significativa melhora de generalização (acerto simultâneo do banco de teste) para a rede treinada com a função *Levenberg-Marquardt* com Regularização Bayesiana.

IV.1.5 Banco de dados

Com as melhores condições do modelo neural já definidas (o algoritmo de treinamento de *Levenberg-Marquardt* com Regularização Bayesiana, topologia 4x9x2,

intervalo de normalização de $[-1;1]$, e funções de ativação “logsig” e linear para as camadas intermediária e final, respectivamente), a próxima etapa é explorar diferentes maneiras de trabalhar com o banco de dados.

Aleatoriedade

Uma das opções de alterar o banco de dados é verificar o que ocorre com o acerto simultâneo quando a rede é treinada com dados dispostos em ordem aleatória. A Tabela 4.5.1 apresenta os resultados deste teste.

Tabela 4.5.1: Resultados do teste para verificar o efeito da aleatoriedade dos dados de treinamento no acerto simultâneo para o modelo dinâmico.

Teste	Resultados do Treinamento						Resultados do Teste			
	SSE	SSW	N	γ	aC(%)	aT(%)	aS(%)	aC(%)	aT(%)	aS(%)
Aleatórios	1203,56	107,82	65	53,50	84,21	80,26	67,95	83,00	78,50	65,50
Não-Aleatórios	1212,02	111,78	65	53,90	84,31	79,94	67,76	88,50	86,50	76,50

Os resultados apresentados na Tabela 4.5.1 indicam que o modelo neural apresenta um resultado melhor com os dados de treinamento em ordem seqüencial (aS = 76,50%).

Variação da quantidade dos vetores que compõem os conjuntos de treinamento e de teste

A quantidade de vetores entrada-saída dos conjuntos de treinamento e de teste foi alterada para verificar qual a relação que fornece um maior acerto simultâneo de porcentagem de carbono e de temperatura de aço para o modelo neural dinâmico. Os resultados estão apresentados na Tabela 4.5.2.

Tabela 4.5.2: Resultados do teste para avaliar qual a relação de quantidade de vetores entrada-saída que fornece o maior acerto simultâneo para o modelo dinâmico.

Teste	Resultados do Treinamento						Resultados do Teste			
	SSE	SSW	N	γ	aC(%)	aT(%)	aS(%)	aC(%)	aT(%)	aS(%)
5030/200	1212,00	111,90	65	53,70	84,31	79,90	67,75	88,50	86,50	76,50
4830/400	1166,52	110,21	65	53,60	84,14	79,90	67,68	88,50	82,50	73,25
4630/600	1113,19	134,76	65	54,20	83,89	80,24	67,69	87,33	80,00	71,00
4030/1200	976,69	151,91	65	55,90	83,62	80,45	67,59	86,33	78,92	69,08
2615/2615	634,73	146,76	65	52,60	83,94	81,15	68,45	83,83	78,78	66,40

A Tabela 4.5.2 mostra que a melhor relação das quantidades dos vetores de entrada-saída entre os conjuntos de treinamento e teste é a 5030/200, ou seja, 5030 vetores

de entrada-saída para o conjunto de treinamento e 200 vetores para o conjunto de teste (aS = 76,50%).

Este teste necessita de uma verificação do comportamento dos dados do conjunto de teste em relação ao de treinamento, ou seja, verifica-se se os dados do conjunto de teste são representativos e atingem toda a extensão do conjunto de treinamento. Desta forma, para cada relação de vetores apresentada na Tabela 4.5.2 foram construídos gráficos que comparam as variáveis de saída destes dois conjuntos. Um destes gráficos está representado na Figura 4.2.

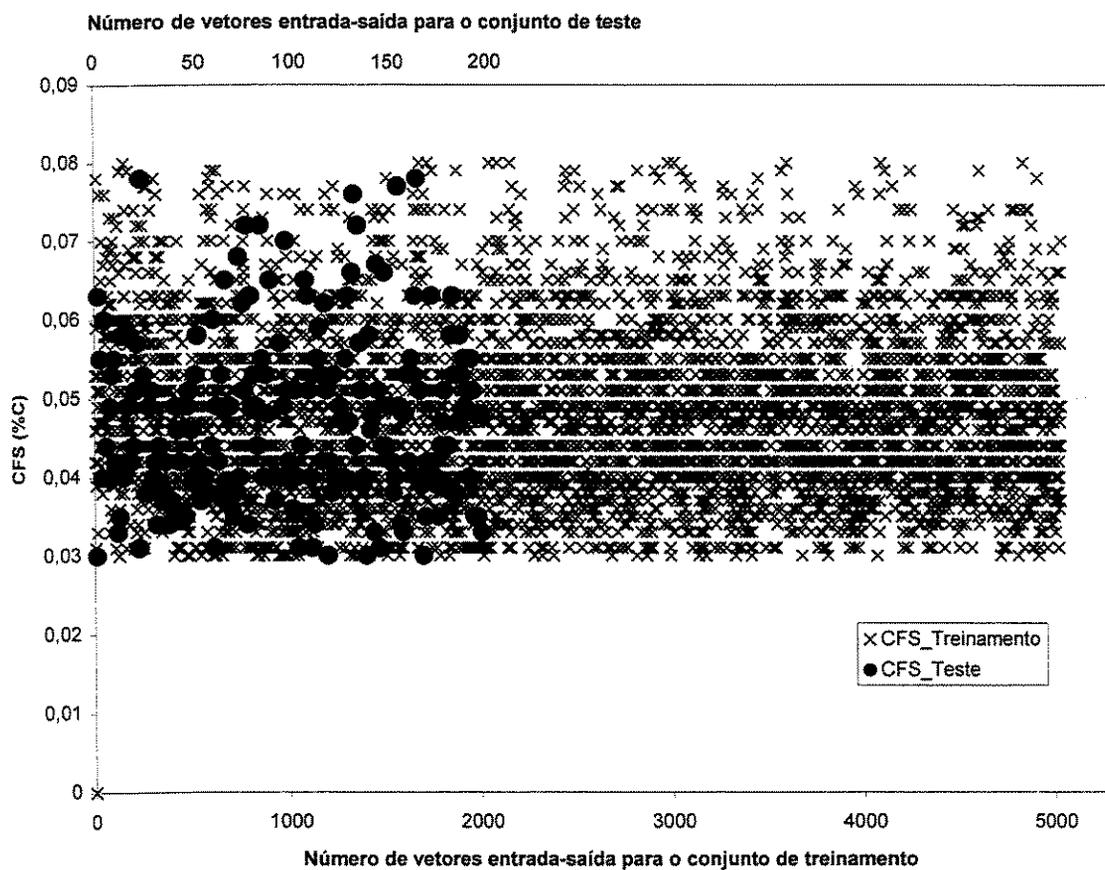


Figura 4.2: Comparação do comportamento da variável de saída porcentagem de carbono no final do sopro (CFS) para os dados de treinamento (5030) e teste (200).

Analisando a Figura 4.2 é possível avaliar que a dispersão dos dados do conjunto de teste com 200 vetores entrada-saída é compatível com a dos dados do conjunto de treinamento com 5030 vetores entrada-saída.

Análise do componente principal

A análise do componente principal (“prepca”) é uma ferramenta que pode reduzir a dimensão do modelo neural. Desta forma, é aplicada ao banco de dados esta ferramenta. Como o tamanho do banco de dados pode ser reduzido retendo apenas aqueles componentes que contribuem mais que uma fração especificada da variação total do banco de dados, são testadas frações mínimas de discrepância. Os resultados se encontram na Tabela 4.5.3.

Tabela 4.5.3: Resultados do teste de aplicação da ferramenta análise do componente principal (“prepca”) para verificar seu efeito nos dados de treinamento para atingir um maior acerto simultâneo para a RNA .

Teste	Resultados do Treinamento						Resultados do Teste			
	Fração Mín. Disc.	SSE	SSW	N	γ	aC(%)	aT(%)	aS(%)	aC(%)	aT(%)
0,001	7155,83	68,86	65	56,10	84,18	80,02	67,83	88,00	85,50	75,00
0,01	7161,33	56,49	65	54,80	84,16	79,82	67,46	88,00	86,00	75,50
0,05	7157,33	59,24	65	55,90	84,25	79,96	67,67	88,00	85,50	75,00
0,1	8793,89	146,16	47	39,60	82,98	70,68	59,07	83,50	73,50	61,50

Em 0,1 => 2 variáveis do vetor de entrada.

A Tabela 4.5.3 traz os resultados da aplicação do “prepca” ao banco de dados e indica que com uma fração mínima de discrepância de 0,1 a dimensão do banco de dados se reduz de 4 para 2 e, no entanto, o acerto simultâneo obtido é menor (aS = 61,50%) que o maior resultado obtido nos testes anteriores (aS = 76,50%). Com as outras frações mínimas de discrepância testadas não há redução da dimensão do problema.

Desvio Padrão

Ao banco de dados inicial foi realizada uma triagem de acordo com os desvios padrões (σ) e foram obtidos conjuntos reduzidos de treinamento e de teste dentro dos intervalos [média $\pm 1\sigma$] e [média $\pm 2\sigma$]. Foram então realizados testes com estes conjuntos para avaliar se ocorre o aumento do acerto simultâneo. Os resultados estão apresentados na Tabela 4.5.4.

Tabela 4.5.4: Resultados do teste a separação do banco de dados com 2 e 1 desvios padrões para analisar o acerto simultâneo da porcentagem de carbono e temperatura do aço do modelo dinâmico via RNA.

Teste	Resultados do Treinamento						Resultados do Teste				
	σ	SSE	SSW	N	γ	aC(%)	aT(%)	aS(%)	aC(%)	aT(%)	aS(%)
2		1322,90	55,57	65	45,60	86,36	82,74	71,54	89,63	87,20	78,05
1		579,32	9,94	65	128,00	97,40	96,37	93,77	97,30	97,30	94,59
Integral		1212,02	11,78	65	539,00	84,31	79,94	67,79	88,50	86,50	76,50

O resultado observado na Tabela 4.5.4 para 1 desvio padrão ($aS = 94,59\%$) apresenta-se elevado, no entanto, é importante salientar que este modelo neural não tem aplicação prática confiável, pois a rede neural não tem a propriedade de extrapolação. A restrição dos dados para uma faixa bem comportada do processo dificilmente gera um modelo aplicável no chão-de-fábrica.

IV.2 Dados 2

De acordo com o Capítulo 3, um outro banco de dados (Dados 2) também foi obtido pela CSN. Dez variáveis de entrada são definidas (Figura 3.2): a porcentagem de carbono (CIB), e a temperatura (TIB) medidas com 88% de sopro, as quantidades de minério de ferro (MFe) e de oxigênio (Oxig) utilizadas para o restante do sopro, as quantidades de gusa (Gusa), de sucata (Suc), e de escória (Esc) carregados no início do sopro, e as porcentagens de manganês (MnIB), fósforo (PIB) e a de enxofre (SIB) amostradas com 88% de sopro. Já a saída da rede é composta pelas seguintes variáveis: a porcentagem de carbono (CFS), a temperatura do aço (TFS), as porcentagem de manganês (MnFS), de fósforo (PFS) e a de enxofre (SFS) medidas ao fim do sopro.

Este novo banco de dados é dividido em 1844 vetores entrada-saída para o conjunto de treinamento e 400 vetores entrada-saída para o conjunto de teste.

As melhores condições de modelagem neural obtidas com o banco de dados anterior (Dados 1) são mantidas, ou seja, o intervalo de normalização de $[-1;1]$, o algoritmo de treinamento de *Levenberg-Marquardt* com Regularização Bayesiana, e as funções de transferência “logsig” e linear para as camadas intermediária e final, respectivamente.

Desta forma, os testes que se seguem, aplicados no novo banco de dados, são realizados para avaliar somente a topologia de rede que resulta em um maior acerto simultâneo de porcentagem de carbono e de temperatura do aço no modelo dinâmico.

IV.2.1 Vetores das variáveis de entrada e saída

Como é descrito no Capítulo 3 (Tabela 3.2), é aplicado ao banco de dados ‘Dados 2’ um teste para avaliar a influência da adição das novas variáveis deste banco na generalização. A princípio, é determinado, aleatoriamente, que os modelos neurais testados apresentariam oito neurônios na camada intermediária. A Tabela 4.6 apresenta os resultados obtidos com este teste.

Tabela 4.6: Resultados do teste para avaliar a influência da adição das novas variáveis deste banco de dados na generalização com uma topologia Ex8xS

Teste	Resultados do Treinamento							Resultados do Teste		
	Topologia	SSE	SSW	N	γ	aC(%)	aT(%)	aS(%)	aC(%)	aT(%)
4x8x2	389.96	144.62	58	47.30	85.20	87.26	74.78	87.25	86.25	74.50
5x8x2	386.72	131.57	66	54.20	85.30	87.26	75.00	88.50	86.50	76.00
6x8x2	383.92	135.66	74	61.00	85.03	87.26	74.62	87.25	86.50	75.00
7x8x2	375.61	148.33	82	71.60	85.85	87.26	75.33	88.00	86.50	75.50
8x8x2	370.46	149.89	90	78.40	85.41	87.20	74.84	88.50	85.75	75.75
9x8x2	348.25	159.03	98	87.60	86.61	87.80	76.36	89.50	85.50	76.00
10x8x2	349.56	159.97	106	92.80	86.82	87.64	76.46	88.50	85.50	75.00
8x8x3	406.69	167.98	99	89.30	85.09	87.31	74.62	89.00	87.00	77.25
9x8x3	383.62	245.40	107	100.00	86.50	87.69	76.25	88.00	86.25	75.50
10x8x3	384.68	168.12	115	106.00	86.44	88.02	76.68	88.75	86.00	75.50
9x8x4	503.43	186.34	116	107.00	86.55	88.12	76.57	88.75	85.25	75.75
10x8x4	429.64	163.90	124	115.00	86.23	88.45	76.41	87.25	86.00	75.00
10x8x5	484.68	201.41	133	125.00	86.61	87.96	76.25	89.25	86.00	76.50

Observando os resultados da Tabela 4.6 é possível verificar que o maior acerto simultâneo é obtido com a topologia 8x8x3 (aS = 77,25%), ou seja, quando são adicionados os vetores das variáveis Gusa, Suc, Esc e MnIB à entrada da rede e o vetor MnFS à saída da rede.

Como a melhor topologia verificada para o banco de dados ‘Dados 1’ (4x9x2) apresentava nove neurônios na camada intermediária, foi repetido este teste com nove neurônios na camada intermediária. Os resultados estão dispostos na Tabela 4.7.

Tabela 4.7: Resultados do teste com uma topologia Ex9xS para avaliar a influência da adição das novas variáveis desde banco no acerto simultâneo do modelo dinâmico.

Teste	Resultados do Treinamento						Resultados do Teste			
	SSE	SSW	N	γ	aC(%)	aT(%)	aS(%)	aC(%)	aT(%)	aS(%)
4x9x2	390.35	131.29	65	49.00	84.76	87.31	74.46	88.25	86.50	75.75
5x9x2	384.71	158.02	74	61.30	85.14	87.36	74.84	89.00	86.50	76.25
6x9x2	382.96	149.83	83	65.90	84.92	87.20	74.62	88.00	86.50	75.75
7x9x2	374.90	147.10	92	76.90	85.03	87.36	74.89	87.25	86.25	74.75
8x9x2	368.26	146.80	101	86.60	85.30	86.93	74.67	88.50	87.00	76.25
9x9x2	346.64	196.37	110	97.70	86.50	87.47	76.25	90.50	86.25	77.50
10x9x2	343.66	175.32	119	106.00	86.93	87.91	76.90	88.75	86.25	76.25
8x9x3	403.05	172.35	111	99.80	85.03	87.31	74.73	88.75	86.75	76.50
9x9x3	379.85	211.26	120	110.00	86.50	87.58	76.03	89.25	86.50	76.75
10x9x3	380.79	205.14	129	118.00	86.93	87.91	76.68	90.00	86.25	77.25
9x9x4	500.92	199.93	130	120.00	86.55	87.91	76.46	89.00	86.50	76.75
10x9x4	424.78	207.17	139	129.00	86.55	87.69	76.14	88.75	86.25	76.00
10x9x5	479.82	200.79	149	139.00	86.23	87.80	76.08	89.25	86.00	76.25

O maior resultado simultâneo verificado com os resultados apresentados na Tabela 4.7 é com a topologia 9x9x2 (aS = 77,50%). Esta topologia indica que a adição das variáveis de entrada Gusa, Suc, Esc, MnIB e PIB resulta numa melhor generalização.

Como as melhores estruturas encontradas tanto para 8 (8x8x3) quanto para nove (9x9x2) neurônios na camada intermediária, não coincidem, foi realizado o mesmo teste com um número significativamente maior de neurônios na camada intermediária (M = 15). Os resultados se encontram na Tabela 4.8.

Tabela 4.8: Resultados do teste com uma topologia Ex15xS para avaliar a influência da adição das novas variáveis desde banco na generalização

Teste	Resultados do Treinamento						Resultados do Teste			
	Topologia	SSE	SSW	N	gama	aC(%)	aT(%)	aS(%)	aC(%)	aT(%)
4x15x2	389.10	157.71	107	57.00	84.87	86.98	74.24	87.50	86.25	74.75
5x15x2	380.78	200.02	122	85.00	85.25	87.80	75.33	88.75	85.75	75.25
6x15x2	369.12	333.18	137	114.00	85.20	87.58	75.05	87.25	86.50	74.50
7x15x2	364.21	228.08	152	122.00	85.57	87.96	75.98	87.25	86.00	74.00
8x15x2	352.74	261.69	167	141.00	85.74	88.45	76.30	88.00	85.25	74.50
9x15x2	329.15	283.73	182	155.25	87.20	88.07	77.17	89.75	85.50	76.25
10x15x2	326.23	299.85	197	168.00	87.42	88.02	77.28	88.50	86.25	76.00
8x15x3	383.98	327.21	183	164.00	85.95	87.69	75.81	87.75	86.50	75.50
9x15x3	360.50	380.19	198	180.00	87.26	87.47	76.68	88.00	86.25	75.50
10x15x3	360.84	329.58	192	213.00	86.88	88.12	76.95	88.00	85.75	74.50
9x15x4	477.23	317.92	214	193.00	86.66	88.23	76.95	88.50	85.75	75.25
10x15x4	404.56	286.73	229	206.00	86.98	88.29	77.28	88.25	85.25	74.50
10x15x5	456.16	342.37	245	225.00	86.71	88.61	77.12	89.50	85.25	75.50

Avaliando os resultados apresentados na Tabela 4.8, é possível observar que os valores do acerto simultâneo obtidos com a topologia Ex15xS são menores do que os apresentados nas Tabela 4.6 e 4.7, que apresentam os resultados do teste com as topologias Ex8xS e Ex9xS, respectivamente. De qualquer forma, a topologia que apresenta o maior acerto simultâneo é a 9x15x2, (aS = 76,25%), incluindo a maioria das variáveis de entrada do novo banco de dados ('Dados 2') e mantendo as variáveis de saída as CFS e TFS.

Como não foi possível obter uma estrutura consistente que satisfaça as diferentes quantidades de neurônios intermediários testadas, foi determinado que para os próximos testes, todos os vetores das variáveis de entrada e saída do banco de dados 'Dados 2' seriam utilizados. Desta maneira, o modelo neural estaria trabalhando com mais informações do sistema.

IV.2.2 Topologia

Com o objetivo de determinar a topologia que fornece o maior acerto simultâneo foram aplicados, ao banco de dados 'Dados 2', testes (Capítulo 3 item III 3.2) com uma e duas camadas ocultas. As estruturas de topologia testadas foram 10xMx5 e 10xMxNx5, respectivamente. Os resultados obtidos com o teste para uma camada intermediária estão na Tabela 4.9.1.

Tabela: 4.9.1: Resultados do teste para definir a melhor topologia com uma camada intermediária para a RNA ('Dados 2').

Teste	Resultados do Treinamento								Resultados do Teste				
	Topologia_1	SSE	MSE	SSW	N	γ	aC(%)	aT(%)	aS(%)	SSE	MSE	aC(%)	aT(%)
10x1x5	785,0	0,43	14,42	21	20,40	84,00	78,90	66,65	1241,0	3,10	87,25	81,75	71,00
10x2x5	659,3	0,36	24,86	37	35,30	83,41	83,35	69,79	1262,0	3,16	87,25	84,75	73,25
10x3x5	574,1	0,31	42,33	53	50,30	83,24	86,77	72,34	1274,7	3,19	87,75	86,50	75,00
10x4x5	527,6	0,29	67,32	69	65,70	85,36	87,58	74,89	1291,4	3,23	89,25	86,25	76,50
10x5x5	509,0	0,28	102,75	85	81,20	85,79	87,04	74,95	1296,1	3,24	88,75	86,75	76,25
10x6x5	504,8	0,27	106,54	101	94,50	85,74	87,26	75,16	1303,5	3,26	89,25	86,00	76,25
10x7x5	492,2	0,27	173,33	117	111,00	86,01	87,80	75,71	1306,2	3,27	89,25	87,50	77,50
10x8x5	486,2	0,26	175,34	133	124,43	86,50	87,58	75,81	1304,0	3,26	89,50	86,00	76,75
10x9x5	479,8	0,26	200,79	149	139,00	86,23	87,80	76,08	1299,0	3,25	89,25	86,00	76,25
10x10x5	475,4	0,26	224,43	165	154,00	86,17	87,85	75,98	1293,4	3,23	89,00	86,50	76,75
10x11x5	468,9	0,25	275,40	181	171,00	86,61	88,02	76,36	1286,2	3,22	88,00	85,00	74,75
10x12x5	465,9	0,25	266,63	197	183,00	86,82	88,02	76,57	1287,1	3,22	89,50	85,75	76,00
10x13x5	462,5	0,25	293,18	213	197,00	86,82	87,80	76,52	1293,8	3,23	89,50	86,00	76,50
10x14x5	452,9	0,25	379,97	229	216,00	86,71	87,64	76,19	1299,4	3,25	86,75	85,25	73,50
10x15x5	456,2	0,25	342,37	245	225,00	86,71	88,61	77,12	1297,0	3,24	89,50	85,25	75,50
10x16x5	450,5	0,24	404,65	261	243,00	86,88	87,80	76,36	1294,4	3,24	88,00	85,75	75,25
10x17x5	447,8	0,24	381,95	277	255,00	86,55	88,07	76,52	1295,3	3,24	87,50	84,00	73,00
10x18x5	441,2	0,24	486,02	293	273,00	87,20	87,69	76,84	1296,7	3,24	87,25	86,00	74,50
10x19x5	438,4	0,24	516,46	309	288,00	87,69	88,07	77,39	1303,0	3,26	87,25	85,00	74,25
10x20x5	440,2	0,24	483,98	325	299,00	87,80	88,29	77,82	1308,0	3,27	86,50	85,25	73,00
10x21x5	432,5	0,23	562,49	341	317,00	87,58	88,39	77,60	1293,7	3,23	86,75	85,75	73,75
10x22x5	433,4	0,24	551,76	357	330,00	87,36	88,88	77,87	1298,9	3,25	86,25	85,25	72,75

Como já foi realizado para o banco de dados 'Dados 1', para este novo banco é realizada uma análise da média do somatório do quadrado do erro (MSE), que será uma medida do desvio dos resultados obtidos pelo modelo neural e dos resultados esperados. E de fato, os resultados apresentados pela Tabela 4.9.1 mostram que o valor do MSE da simulação do conjunto de teste é maior que o MSE para de treinamento, ou seja, o desvio dos resultados deve ser maior para conjunto de teste do que o desvio do conjunto de treinamento. E também, como já havia sido verificado no banco de dados anterior, a Tabela 4.9.1 mostra que o acerto simultâneo do conjunto de teste é maior que o acerto simultâneo do conjunto de treinamento. Esta diferença de comportamento deve-se ao fato de que o critério de acerto simultâneo faz uma análise vetor a vetor se há ou não acerto e o MSE mede qual foi a dimensão média do desvio dos resultados.

Ainda verificando os resultados da Tabela 4.9.1, nota-se que a topologia de rede que apresentou o maior acerto simultâneo é a 10x7x5 (aS = 77,50%). Entretanto, segundo

os critérios de Hagan, a rede que faria uma melhor generalização seria a de topologia 10x21x5 (aS = 73,75%).

Para encontrar a melhor topologia com duas camadas intermediárias são realizados testes em que o número de neurônios da primeira camada intermediária permanece fixo e o número de neurônios da segunda camada intermediária é variado até ser atingido o critério sugerido por Hagan, ou seja, até a estabilidade dos valores de γ , SSE e SSW. Os resultados se encontram na Tabela 4.9.2.

Tabela 4.9.2: Resultados do teste para definir a melhor topologia com duas camadas intermediárias para a modelo neural via RNA ('Dados 2').

Teste	Resultados do Treinamento							Resultados do Teste		
	SSE	SSW	N	γ	aC(%)	aT(%)	aS(%)	aC(%)	aT(%)	aS(%)
10x1x1x5	785,06	38,55	23	20,70	84,00	79,18	66,92	87,25	81,50	70,75
10x1x2x5	781,83	52,00	30	24,90	84,00	79,18	66,76	87,25	82,25	71,50
10x1x3x5	781,84	51,18	37	24,90	84,00	79,12	66,70	87,25	82,00	71,25
10x1x4x5	781,84	50,67	44	24,90	84,00	79,18	66,76	87,25	82,25	71,50
10x2x1x5	777,66	56,71	35	32,00	84,00	78,96	66,65	87,25	82,00	71,50
10x2x2x5	659,30	63,13	43	38,50	83,35	83,62	70,01	87,25	85,00	73,50
10x2x3x5	651,09	114,81	51	45,90	83,35	83,79	70,12	87,25	85,00	73,50
10x2x4x5	649,32	105,42	59	50,20	83,35	83,95	70,28	87,25	85,25	74,00
10x2x5x5	649,27	100,28	67	50,20	83,46	83,89	70,34	87,25	85,00	73,75
10x3x1x5	774,76	65,43	47	40,50	84,00	79,61	67,14	87,25	81,25	70,75
10x3x2x5	652,14	72,31	56	49,90	83,51	83,30	70,01	87,25	84,75	73,25
10x3x3x5	573,51	84,62	65	56,40	83,30	86,88	72,45	87,75	86,50	75,00
10x3x4x5	570,95	112,26	74	64,80	83,19	87,31	72,78	87,75	86,50	75,00
10x3x5x5	570,11	117,00	83	68,80	83,51	87,53	73,21	87,75	86,25	74,75
10x3x6x5	569,58	120,73	92	71,20	83,51	87,42	72,99	87,75	86,50	75,00
10x3x7x5	568,98	132,44	101	75,40	83,46	87,15	72,83	87,75	86,50	75,00
10x3x8x5	568,79	134,14	110	76,10	83,57	87,36	73,05	87,75	86,50	75,00
10x3x9x5	568,84	131,20	119	75,50	83,51	87,26	72,99	87,75	86,50	75,00
10x4x1x5	773,93	68,99	59	49,90	84,00	79,07	66,81	87,25	80,75	70,25
10x4x2x5	647,34	98,34	69	61,50	83,57	84,27	70,88	87,25	84,75	73,25
10x4x3x5	566,70	122,70	79	70,50	83,35	87,42	72,78	87,50	85,25	73,50
10x4x4x5	524,39	146,59	89	79,50	85,52	87,85	75,27	89,50	86,00	76,25
10x4x5x5	518,88	196,92	99	87,40	85,30	87,47	74,57	89,25	86,50	76,75
10x4x6x5	517,27	217,48	109	93,80	85,52	87,53	74,73	89,25	86,50	76,75
10x4x7x5	515,85	274,60	119	104,00	85,47	87,31	74,73	89,00	87,00	77,00
10x4x8x5	512,38	397,10	129	113,00	85,63	87,64	75,16	88,75	86,50	76,25
10x4x9x5	514,07	291,76	139	115,00	391,00	85,63	88,12	89,63	86,00	76,50
10x4x10x5	514,61	270,41	149	116,00	85,47	87,74	74,95	89,00	86,00	76,00
10x5x1x5	769,80	111,88	71	60,90	84,00	78,63	66,21	87,25	80,75	70,00
10x5x2x5	643,58	112,32	82	72,40	83,46	84,00	70,77	87,25	84,75	73,50
10x5x3x5	561,11	140,95	93	82,00	83,24	87,74	72,94	87,75	85,25	74,00
10x5x4x5	511,87	235,30	104	93,80	85,90	87,85	75,49	90,25	86,50	77,25
10x5x5x5	499,21	240,78	115	104,00	85,68	87,15	75,05	90,00	86,50	77,50
10x5x6x5	497,42	227,37	126	114,00	85,95	87,47	75,16	89,75	86,00	77,25
10x5x7x5	490,65	322,37	137	124,00	86,06	87,53	75,43	88,50	86,75	76,50
10x5x8x5	489,77	365,44	148	131,00	85,95	87,58	75,43	89,00	85,25	75,75
10x5x9x5	486,47	408,03	159	141,00	86,55	88,12	76,25	89,00	86,00	76,50
10x5x10x5	487,30	422,75	170	148,00	86,33	87,80	75,81	88,75	86,00	76,50
10x5x11x5	482,55	541,79	181	162,00	86,55	88,29	76,46	89,26	85,00	75,50
10x5x12x5	481,62	569,34	192	170,00	86,77	88,02	76,41	89,50	87,25	78,25
10x5x13x5	483,29	458,17	203	172,00	86,39	88,45	76,46	89,25	87,00	77,50
10x6x1x5	769,00	101,55	83	67,60	84,00	79,23	66,87	87,25	80,25	69,75
10x6x2x5	640,91	129,29	95	82,60	83,68	84,65	71,53	87,25	85,25	73,75
10x6x3x5	558,48	158,73	107	93,80	83,19	87,74	73,21	87,25	84,50	72,50
10x6x4x5	503,76	255,43	119	109,00	86,23	87,53	75,76	90,25	85,75	77,50
10x6x5x5	494,17	291,99	131	119,00	85,95	87,69	75,65	89,75	85,75	76,50
10x6x6x5	484,64	471,95	143	133,00	86,17	88,12	76,25	88,75	85,50	76,25
10x6x7x5	480,74	394,36	155	143,00	86,98	88,07	76,57	89,50	86,50	78,00
10x6x8x5	481,56	417,21	167	151,00	86,55	88,45	76,68	90,00	85,75	77,50
10x6x9x5	477,91	431,75	179	162,00	86,44	87,96	76,14	88,75	86,50	77,00
10x6x10x5	474,78	445,75	191	170,00	86,93	87,74	76,30	89,00	86,50	76,25
10x6x11x5	473,11	681,53	203	183,00	86,61	88,67	76,95	89,25	85,25	76,75
10x6x12x5	467,09	650,66	215	197,00	86,77	88,23	76,68	88,75	86,75	77,00

A topologia com duas camadas intermediárias que apresenta o maior acerto simultâneo, como está mostrado na Tabela 4.9.2, é a 10x5x12x5 (aS = 78,25%).

Baseado na análise do processo e na dificuldade de obter *on line* as variáveis como porcentagem de manganês, fósforo e enxofre, sete variáveis de entrada são definidas como mais importantes: a porcentagem de carbono (CIB), e a temperatura (TIB) medidas com 88% de sopro, as quantidades de minério de ferro (MFe) e de oxigênio (Oxig) utilizadas para o restante do sopro, e as quantidades de gusa (Gusa), de sucata (Suc), e de escória (Esc) carregadas no início do sopro. Já a saída da rede é composta pelas seguintes variáveis: a porcentagem de carbono (CFS) e a temperatura do aço (TFS) medidas no fim de sopro.

Com esta estrutura dos vetores de entrada e saída, é sugerido um teste de topologia, para averiguar a quantidade de neurônios intermediários, para uma (7xMx2) e duas camadas (7xMxNx2), que resulta num acerto simultâneo maior. A Tabela 4.10.1 traz os resultados do teste com uma camada intermediária.

Tabela: 4.10.1: Resultados do teste para definir a melhor topologia com uma camada intermediária para a RNA – Estrutura 7xMx2.

Teste	Resultados do Treinamento							Resultados de Teste			
	Topologia_1	SSE	SSW	N	γ	aC(%)	aT(%)	aS(%)	SSE	aC(%)	aT(%)
7x3x2	392,38	73,14	32	29,00	84,60	86,98	73,81	608,08	89,00	86,5	76,75
7x4x2	389,07	97,03	42	37,10	84,54	87,04	73,86	608,76	87,75	86,75	75,25
7x5x2	384,82	104,17	52	45,40	84,49	87,42	74,30	602,57	87,75	86,50	75,50
7x6x2	380,33	120,82	62	54,40	84,92	87,47	74,62	601,33	87,00	86,50	75,00
7x7x2	377,49	144,85	72	62,70	85,09	87,36	75,11	600,30	87,50	86,50	75,25
7x8x2	375,32	174,01	82	70,60	84,76	87,58	74,67	602,05	86,50	85,50	73,50
7x9x2	374,02	157,04	92	79,10	85,30	87,20	74,73	597,13	88,25	86,00	75,00
7x10x2	372,66	161,40	102	84,30	85,47	87,31	75,27	598,61	88,25	85,75	74,75
7x11x2	370,24	175,67	112	93,50	85,52	87,47	75,43	598,65	88,50	85,50	75,00
7x12x2	367,25	203,23	122	102,00	85,41	87,91	75,87	603,58	87,00	86,25	74,25
7x13x2	367,61	187,91	132	107,00	85,57	87,85	75,81	598,87	87,75	86,00	74,75
7x14x2	365,39	211,76	142	115,00	85,57	88,29	76,19	604,87	87,75	85,25	74,00
7x15x2	363,71	239,10	152	122,00	85,52	88,45	76,36	605,76	87,50	85,00	73,50
7x16x2	362,88	362,88	162	129,00	85,63	88,02	75,98	603,59	87,50	85,50	74,00
7x17x2	361,33	274,24	172	136,00	85,79	88,50	76,57	606,55	87,50	85,75	74,25
7x18x2	317,95	317,92	182	148,00	85,85	88,39	76,63	600,63	87,00	85,50	73,50
7x19x2	357,48	327,07	192	152,00	85,90	88,29	76,52	605,98	86,50	85,00	73,00
7x20x2	353,12	395,19	202	167,00	86,01	88,45	76,84	601,04	86,50	86,00	73,50

A Tabela 4.10.1 apresenta resultados que indicam que a topologia que fornece um maior acerto simultâneo ($aS = 76,75\%$) é a $7 \times 3 \times 2$. Também é realizado este teste para duas camadas intermediárias. Os resultados estão dispostos na Tabela 4.10.2.

Tabela 4.10.2: Resultados do teste de topologia com duas camadas intermediárias – estrutura 7xMxNx2.

Teste	Resultados do Treinamento							Resultados de Teste			
	Topologia_2	SSE	SSW	N	γ	aC(%)	aT(%)	aS(%)	SSE	aC(%)	aT(%)
7x2x1x2	424,37	49,77	23	195,00	83,62	86,98	72,83	596,07	87,75	86,50	75,00
7x2x2x2	401,22	80,45	28	23,70	84,49	86,61	73,26	609,43	89,50	86,00	76,50
7x2x3x2	397,31	130,58	33	26,20	84,82	86,98	73,97	603,55	88,25	87,00	76,25
7x2x4x2	396,45	123,27	38	26,80	84,54	86,77	73,59	606,95	88,75	86,00	75,75
7x3x1x2	422,92	56,83	32	25,10	83,62	86,98	72,83	595,34	87,25	87,25	75,25
7x3x2x2	391,94	116,32	38	31,70	84,87	87,20	74,30	607,81	89,25	86,50	77,00
7x3x3x2	389,67	157,28	44	37,10	84,92	86,88	74,02	608,38	87,75	86,00	75,00
7x3x4x2	388,51	207,70	50	41,90	84,60	87,42	74,24	596,15	87,75	87,25	76,25
7x3x5x2	387,64	179,38	56	42,90	84,76	87,09	74,02	608,92	88,00	86,50	75,75
7x4x1x2	420,64	66,08	41	31,90	83,68	87,04	73,05	598,86	87,75	86,50	75,00
7x4x2x2	392,68	131,27	48	40,10	84,87	86,66	74,02	606,71	89,25	85,75	76,50
7x4x3x2	384,59	179,85	55	46,30	84,54	87,69	74,24	609,60	88,25	86,25	75,75
7x4x4x2	386,95	149,36	62	47,10	84,87	87,20	74,46	607,13	89,50	86,75	77,00
7x4x5x2	379,75	255,65	69	56,70	84,82	86,93	73,92	603,51	87,75	87,50	76,25
7x4x6x2	381,30	225,60	76	58,80	85,14	87,04	74,57	610,31	89,00	86,50	76,25
7x5x1x2	419,61	79,18	50	36,90	83,62	87,47	73,48	596,31	87,00	85,50	73,50
7x5x2x2	384,33	148,64	58	49,90	84,71	87,47	74,51	604,02	88,50	86,75	76,25
7x5x3x2	380,45	196,95	66	55,60	85,09	87,42	74,57	605,63	88,25	86,75	76,25
7x5x4x2	378,99	224,26	74	60,30	84,49	87,64	74,24	604,54	87,75	86,75	75,75
7x5x5x2	377,69	249,23	82	65,80	84,49	87,20	74,19	609,85	87,75	87,50	76,25
7x5x6x2	374,58	347,62	90	76,80	85,09	87,91	75,11	600,68	88,25	86,25	75,25
7x5x7x2	372,46	299,16	98	77,40	84,87	87,64	74,95	602,48	87,50	85,50	74,00
7x6x1x2	418,99	81,04	59	42,10	83,68	87,80	73,75	599,43	87,75	86,00	74,50
7x6x2x2	379,76	207,15	68	58,10	84,60	87,80	74,46	598,78	87,00	86,25	75,00
7x6x3x2	377,66	208,49	77	64,70	85,03	87,96	75,33	601,51	87,50	86,75	75,50
7x6x4x2	374,19	288,80	86	72,00	85,25	87,80	75,27	603,90	88,00	87,50	76,75
7x6x5x2	371,89	265,94	95	77,40	84,87	87,69	74,84	600,47	88,50	86,25	75,50
7x6x6x2	373,64	247,09	104	80,70	85,03	87,69	75,00	602,98	87,75	86,25	75,25
7x6x7x2	367,53	361,30	113	92,50	84,98	87,74	74,95	601,49	88,00	86,00	75,25
7x6x8x2	367,41	376,75	122	98,70	85,47	87,96	75,54	602,54	88,00	86,00	74,75
7x6x9x2	367,91	351,67	131	101,00	85,20	88,02	75,38	605,36	87,75	86,25	75,25
7x6x10x2	364,92	433,07	140	114,00	85,25	88,18	75,71	605,31	87,00	85,50	73,50
7x6x11x2	351,05	927,55	149	133,00	85,90	88,18	76,14	606,31	87,00	86,75	75,25
7x6x12x2	357,97	651,43	158	134,00	85,79	87,53	75,43	617,78	88,00	85,50	74,75
7x7x1x2	417,59	93,54	68	49,40	83,68	87,85	73,86	597,26	87,25	86,00	74,00
7x7x2x2	376,80	177,63	78	67,00	85,36	87,36	75,05	601,22	87,50	86,25	75,25
7x7x3x2	375,80	204,14	88	73,20	85,14	87,47	74,73	600,89	87,75	86,50	75,25
7x7x4x2	371,81	236,85	98	81,10	84,87	87,47	74,95	606,45	87,75	86,50	75,25
7x7x5x2	235,00	246,24	108	85,56	85,20	88,02	75,16	600,74	88,00	85,25	74,75
7x7x6x2	366,95	429,21	118	100,00	85,20	87,96	75,33	600,35	87,00	87,00	75,50
7x7x7x2	358,35	521,17	128	113,00	85,57	87,15	74,89	601,68	87,75	86,00	75,00
7x7x8x2	360,72	437,56	138	116,00	85,74	88,12	76,25	601,98	87,25	86,25	74,50

Os resultados apresentados pela Tabela 4.10.2 indicam dois modelos neurais que apresentam um maior acerto simultâneo: a 7x3x2x2 (aS = 77,00%) e a 7x4x4x2 (aS =

77,00%). Segundo Hagan e Foresse (1997), se o valor de γ (número efetivo de parâmetros) for muito próximo ao valor de N (número atual de parâmetros), então provavelmente a rede não é grande o bastante para representar adequadamente a função real. Desta forma, a rede selecionada para as considerações seguintes é a de topologia $7 \times 4 \times 4 \times 2$ ($\gamma = 47,10$ e $N = 62$), pois a de topologia $7 \times 3 \times 2 \times 2$ ($\gamma = 31,70$ e $N = 38$) apresenta valores muito próximos de N e γ .

IV.2.3 Análise de componente principal (“prepca”)

A ferramenta de análise do componente principal (“prepca”) é aplicada ao banco de dados ‘Dados 2’ na tentativa de melhorar o acerto simultâneo com a redução da dimensão do modelo neural. A topologia escolhida para o teste é a de $10 \times 5 \times 12 \times 5$, pois é a que apresenta o maior acerto nos testes anteriores ($aS = 78,25\%$), apesar destes vetores de entrada e saída conterem variáveis que não são mensuráveis *on-line* (porcentagem de manganês, fósforo e enxofre). Os resultados deste teste estão dispostos na Tabela 4.11

Tabela 4.11: Resultados do teste de aplicação do “prepca” para verificar seu efeito no acerto simultâneo do modelo neural (‘Dados 2’)

Teste		Resultados do Treinamento					Resultados do Teste				
Fração	Min. Disc.	SSE	SSW	N	γ	aC(%)	aT(%)	aS(%)	aC(%)	aT(%)	aS(%)
0,001		3865,99	680,31	192	167,00	86,39	87,85	75,87	89,50	86,00	77,00
0,01		3868,63	694,31	192	166,00	85,47	88,02	75,38	89,00	85,25	75,75
0,05		4616,51	667,82	177	156,00	83,95	75,22	62,69	89,75	71,50	65,25
0,1		6463,86	488,68	162	133,00	83,84	68,49	57,65	88,00	66,75	58,80

Em 0,05 \Rightarrow 7 variáveis no vetor de entrada e em 0,1 \Rightarrow 4 variáveis no vetor de entrada

Os resultados da aplicação da ferramenta “prepca” ao banco de dados ‘Dados2’ mostram que com as frações mínimas de discrepância de 0,05 e 0,1 a dimensão do vetor de entrada se reduz de 10 para 7 e 10 para 4, respectivamente. Mas, o acerto simultâneo obtido com estes modelos neurais ($aS = 65,25\%$ e $aS = 58,80\%$, respectivamente) são menores que o maior resultado obtido com testes anteriores ($aS = 78,25\%$). Com as outras frações mínimas de discrepância testadas não houve redução da dimensão do problema.

IV.3 Comparação final de desempenho

Após a realização dos testes apresentados até aqui, é possível realizar uma análise comparativa para avaliar o desempenho dos modelos neurais obtidos. Para tal fim foram selecionados os modelos que apresentam um maior acerto simultâneo e cujas variáveis são medidas *on-line*. Estes modelos são os de topologia 4x9x2 (aS = 76,50%) e 4x3x6x2 (aS = 76,50%) para o banco de dados 'Dados 1' e para o 'Dados 2' é o de topologia 7x4x4x2 (aS = 77,00%).

Uma comparação entre a porcentagem de carbono no final do sopro de um conversor a oxigênio industrial e a porcentagem de carbono, calculada pela RNA com as topologias acima citadas, assim como a comparação obtida por Cunha a partir do banco de dados 'Dados 1' com a topologia 4x9x3x2 estão apresentadas na Figura 4.3.

Comparando os gráficos da Figura 4.3a, da Figura 4.3b e da Figura 4.3c com o da Figura 4.3d, é possível verificar que há uma certa similaridade entre eles.

Também foi verificada a comparação do comportamento da porcentagem de carbono no final do sopro de um conversor a oxigênio industrial e da mesma variável obtida pelos modelos neurais em discussão. A Figura 4.4 apresenta estas comparações.

Com a Figura 4.4, é possível verificar que há uma melhora significativa na generalização da estrutura de rede 7x4x4x2 (Fig 4.4c) em relação às outras apresentadas (Fig 4.4a, 4.4b e 4.4c). Esta estrutura adiciona as quantidades de ferro-gusa, sucata e escória às variáveis de entrada do modelo neural.

A outra variável de saída da rede, a temperatura do aço no período do final de sopro, também foi avaliada. A comparação dos valores da temperatura obtida no processo com os valores obtidos com o processo de generalização dos modelos neurais em estudo está representada na Figura 4.5.

Analisando a Figura 4.4, é possível verificar que há uma melhora significativa na generalização da estrutura de rede 7x4x4x2 (Fig 4.4c) em relação às outras apresentadas (Fig 4.4a, 4.4b e 4.4c). Esta estrutura adiciona as quantidades de ferro-gusa, sucata e escória às variáveis de entrada do modelo neural.

A outra variável de saída da rede, a temperatura do aço no período do final de sopro, também foi avaliada. A comparação dos valores da temperatura obtida no processo

com os valores obtidos com o processo de generalização dos modelos neurais em estudo está representada na Figura 4.5.

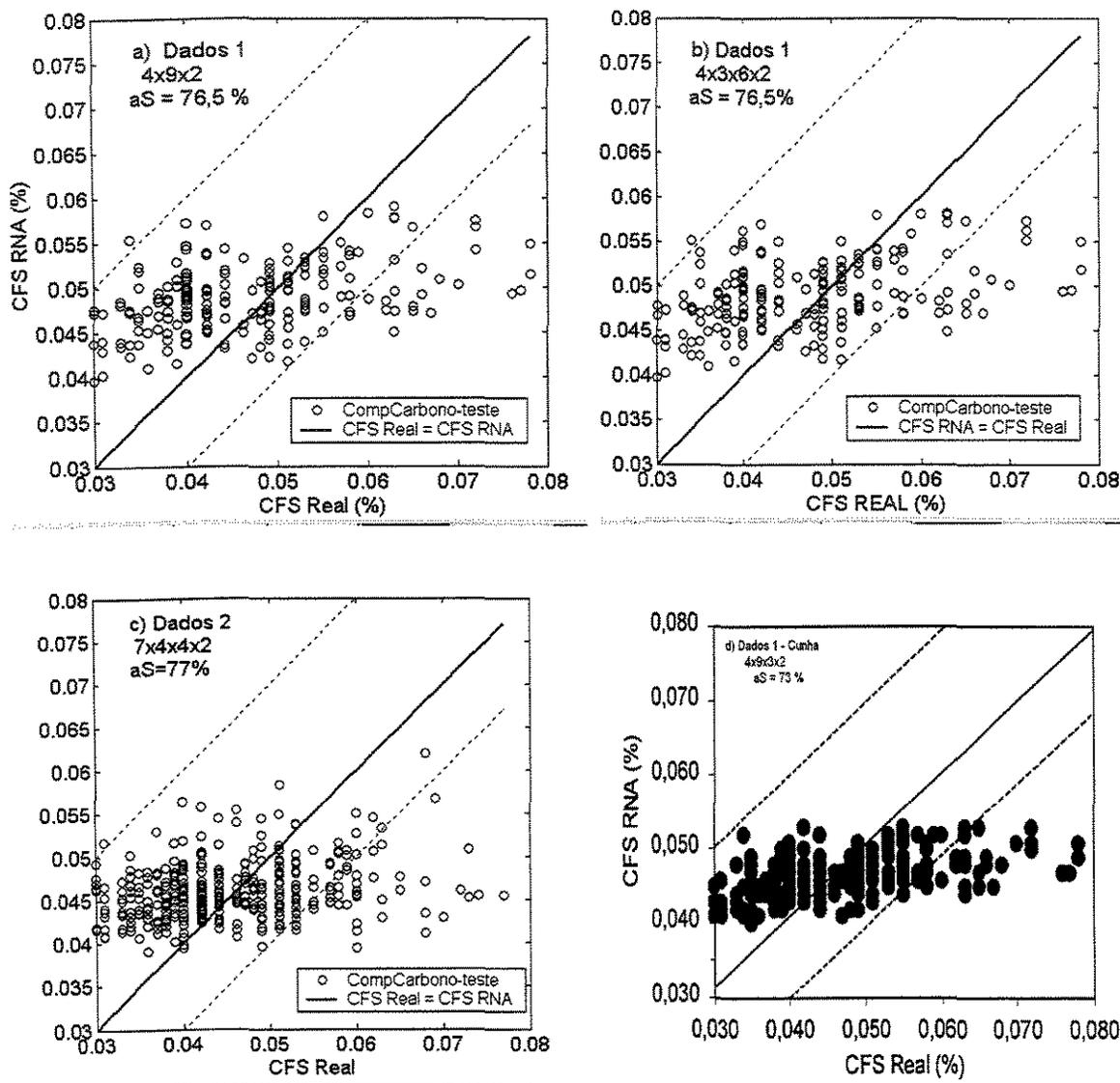


Figura 4.3: Comparação da porcentagem de carbono no final do sopro de um conversor a oxigênio industrial e da porcentagem de carbono obtida na generalização dos modelos neurais com as seguintes particularidades: a) Banco de dados ‘Dados 1’, topologia 4x9x2, aS = 76,50%; b) Banco de dados ‘Dados 1’, topologia 4x3x6x2, aS = 76,50%; c) Banco de dados ‘Dados 2’, topologia 7x4x4x2, aS = 77,00%; d) Banco de dados ‘Dados 1’, topologia 4x9x3x2, aS = 73,30% obtido por Cunha.

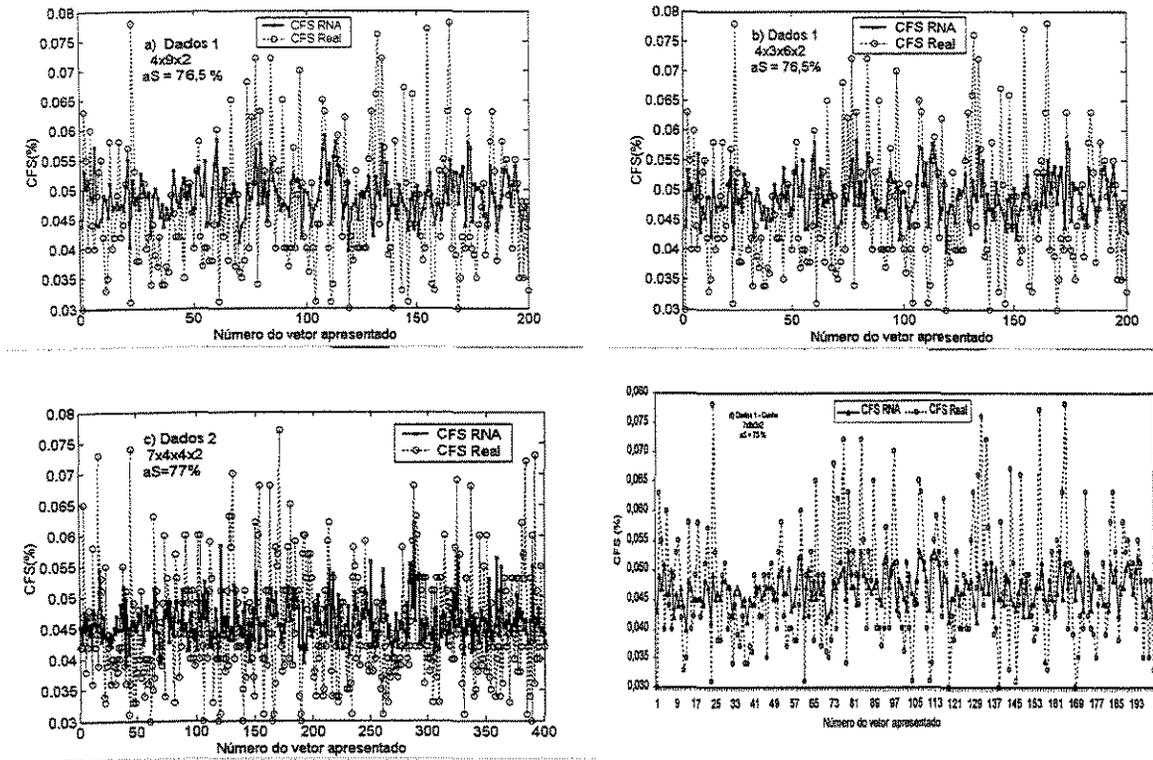


Figura 4.4: Comparação entre o comportamento da porcentagem de carbono no final do sopro de um conversor a oxigênio industrial e o comportamento da porcentagem de carbono obtida pelos modelos neurais com as seguintes particularidades: a) Banco de dados ‘Dados 1’, topologia 4x9x2, aS = 76,50%; b) Banco de dados ‘Dados 1’, topologia 4x3x6x2, aS = 76,50%; c) Banco de dados ‘Dados 2’, topologia 7x4x4x2, aS = 77,00%; d) Banco de dados ‘Dados 1’, topologia 4x9x3x2, aS = 73,30% obtido por Cunha

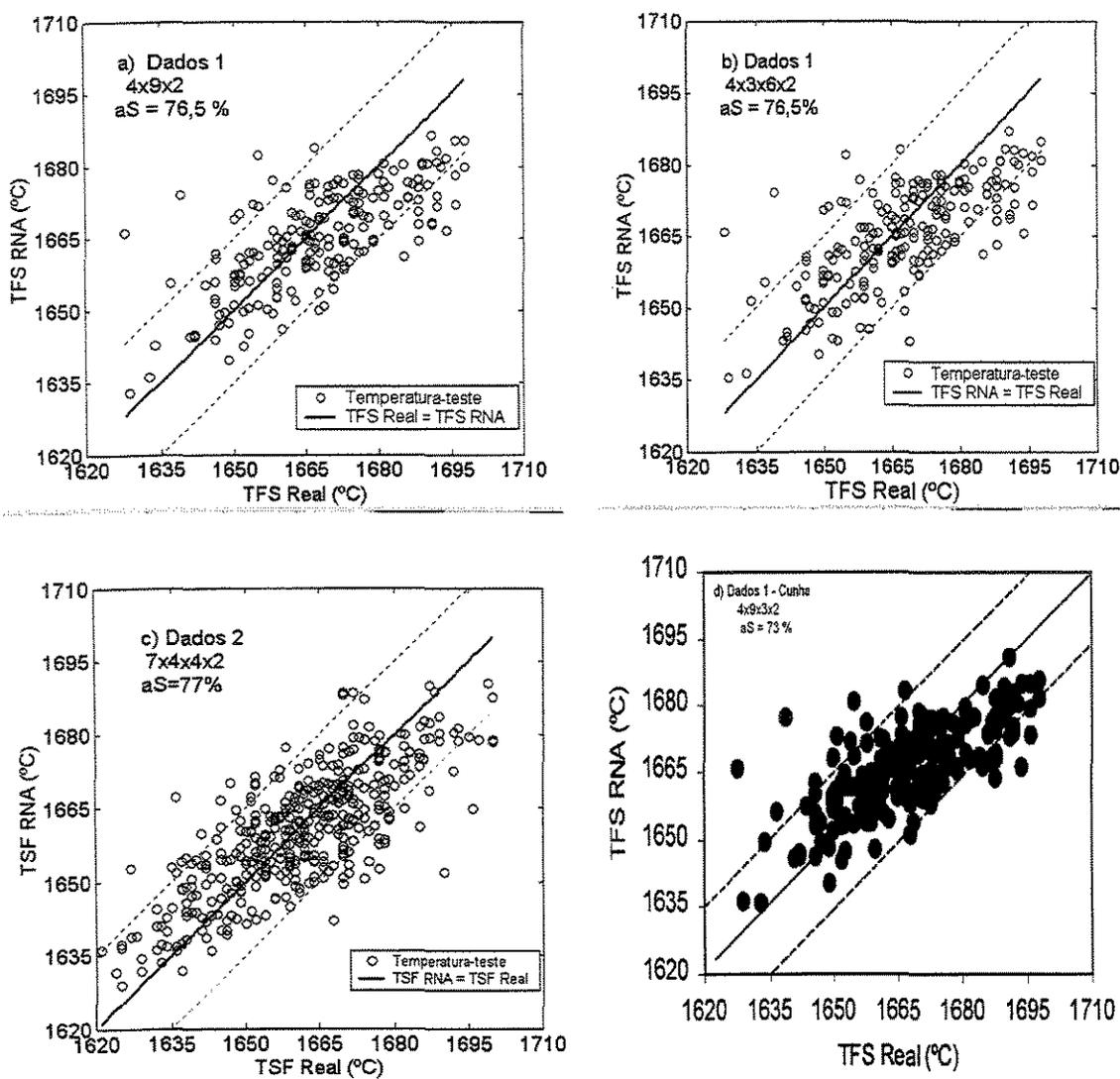


Figura 4.5: Comparação entre a temperatura do aço no final do sopro de um conversor a oxigênio industrial e a temperatura obtida pelos modelos neurais com as seguintes particularidades: a) Banco de dados ‘Dados 1’, topologia 4x9x2, aS = 76,50%; b) Banco de dados ‘Dados 1’, topologia 4x3x6x2, aS = 76,50%; c) Banco de dados ‘Dados 2’, topologia 7x4x4x2, aS = 77,00%; d) Banco de dados ‘Dados 1’, topologia 4x9x3x2, aS = 73,30% obtido por Cunha

Observando a Figura 4.5, assim como a Figura 4.3, é possível avaliar uma similaridade entre as estruturas apresentadas.

Para uma melhor avaliação do desempenho das estruturas neurais, foi realizada a comparação do comportamento da temperatura real de final de sopro industrial com a obtida através dos modelos neurais em evidência neste capítulo. A comparação está exposta na Figura 4.6.

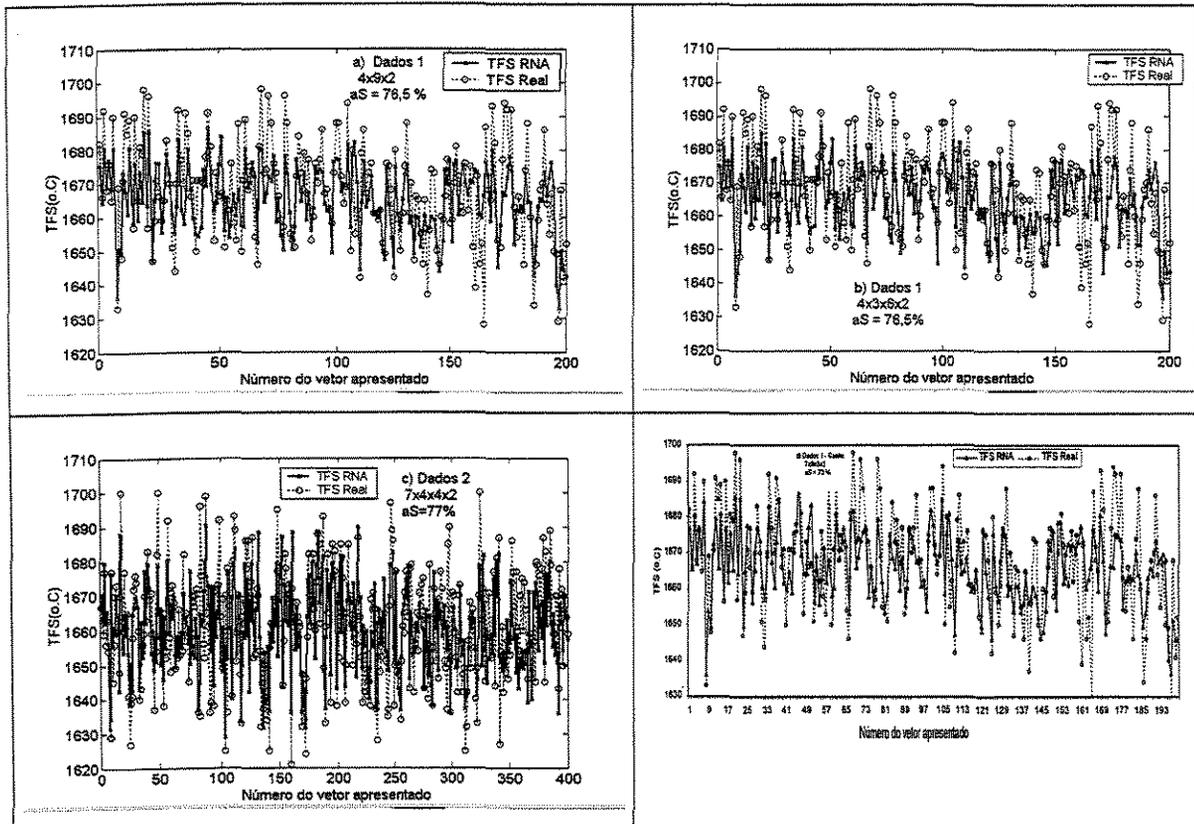


Figura 4.6: Comparação entre comportamento da temperatura de carbono no final do sopro de um conversor a oxigênio industrial e da mesma variável obtida pelos modelos neurais com as seguintes particularidades: a) Banco de dados ‘Dados 1’, topologia 4x9x2, aS = 76,50%; b) Banco de dados ‘Dados 1’, topologia 4x3x6x2, aS = 76,50%; c) Banco de dados ‘Dados 2’, topologia 7x4x4x2, aS = 77,00% d) Banco de dados ‘Dados 1’, topologia 4x9x3x2, aS = 73,30% obtido por Cunha

Avaliando os gráficos das Figuras 4.6a, 4.6b, 4.6c em relação ao gráfico da Figura 4.6d, chega-se facilmente a constatação de que a estrutura 7x4x4x2 é a que apresenta uma melhor generalização, concordando com a avaliação realizada a partir da Figura 4.4.

IV.4 Conclusões

Diante de todos os testes aqui apresentados, pode-se concluir que a estrutura neural que fornece maiores acertos simultâneos é: o intervalo de normalização de [-1;1], o algoritmo de treinamento de *Levenberg-Marquardt* com Regularização Bayesiana, e as funções de transferência “logsig” e linear para as camadas intermediária e final, respectivamente.

As tentativas de alterar o banco de dados com a aleatoriedade dos dados, ou a redução da dimensionalidade do problema através da ferramenta “prepca”, ou ainda a variação da quantidade dos vetores que compõem os conjuntos de treinamento e de teste, não aumentam o acerto simultâneo da porcentagem de carbono e da temperatura do aço.

Quanto a topologia, para o banco de dados ‘Dados 1’, a que oferece um maior acerto é a $4 \times 9 \times 2$ (aS = 76,50%) com uma cada intermediária e a $4 \times 3 \times 6 \times 2$ (aS = 76,50%) com duas camadas intermediárias. Já com o banco de dados ‘Dados 2’, a topologia que resulta num maior acerto simultâneo é a $7 \times 4 \times 4 \times 2$ (aS = 77,00%), salientando a importância da adição na entrada do modelo neural das variáveis de quantidade de ferro-gusa, sucata e escória. Os pesos e os *bias* destes modelos neurais estão dispostos no Anexo 2.

O capítulo seguinte apresenta um teste industrial para avaliar a real eficiência do modelo neural dinâmico neural. A estrutura de rede aplicada é aquela ($7 \times 4 \times 4 \times 2$ – aS = 77,00%) que apresenta na generalização o maior acerto simultâneo de porcentagem de carbono e de temperatura do aço no período de final de sopro e cujas variáveis são medidas *on-line*.

CAPÍTULO V: TESTE INDUSTRIAL

O presente trabalho apresenta como objetivo principal a aplicação industrial de um modelo dinâmico via redes neurais no controle de um conversor a oxigênio. Desta maneira, é proposto um teste industrial para avaliar a real eficiência do modelo dinâmico neural que apresenta na generalização o maior acerto simultâneo de porcentagem de carbono e de temperatura de aço no período de final de sopro.

O **modelo dinâmico** funciona como um ajuste final para a obtenção das características desejadas do aço. A coleta de amostras ao final do sopro indica qual a porcentagem de carbono (CIB) e o valor de temperatura (TIB) naquele momento. Com o modelo convencional baseado em equações diferenciais de variação da composição química e da temperatura ao longo do tempo, além das informações de adição (matérias reciclados, ferro-gusa e sucata) definidos com o modelo estático, calculam-se as necessidades de refrigerantes e de oxigênio necessários para atingir as características do aço.

Diante da necessidade de trabalhar com as variáveis manipuladas deste processo, ou seja, a quantidade de minério de ferro e de oxigênio utilizadas no resopro, o modelo dinâmico neural desenvolvido neste trabalho teve que sofrer algumas modificações. Assim, foi desenvolvido um modelo neural inverso, ou seja, as variáveis de entrada passam a ser aquelas medidas pela sublança e os valores que são os objetivos pré-estabelecidos de porcentagem de carbono e temperatura para o fim de sopro e as variáveis de saída são as quantidades de minério de ferro e de oxigênio necessárias para atingir os valores objetivados.

V.1 Metodologia:

Uma vez que o banco de dados 'Dados 2' não apresenta valores de bateladas em que não houve a necessidade de adição de minério de ferro, foi percebido que tal fato acarretaria em um problema operacional. Este problema se deve ao fato que a rede neural não faz extrapolações confiáveis.

Desta maneira, a partir do banco de dados que deu origem ao "Dados 2", foi realizada uma filtragem para selecionar os vetores das bateladas em que não houve a adição de minério de ferro. Destes dados selecionados, 2000 vetores foram adicionados ao conjunto de treinamento e outros 2000 foram adicionados ao conjunto de teste. Formou-se

um novo banco de dados “Dados 3”. Com este novo banco de dados foi desenvolvido o modelo neural utilizado no testes industrial.

Avaliando os resultados obtidos no Capítulo 4 verificou-se que o modelo neural com 2 camadas intermediárias é o mais adequado para o desenvolvimento do modelo neural. Tal fato se deve a análise dos maiores acertos simultâneos encontrados nas tabelas das Figuras 4.10.1 que apresenta a estrutura 7x3x2 (aS = 76,75%) e da Figura 4.10.2 que apresenta a estrutura 7x4x4x2 (aS = 77,00%). Desta forma, foi aplicado ao banco de dados “Dados 3” o teste de topologia (Capítulo 3 – item III 3.2) com duas camadas intermediárias e a estrutura 7xMxNx2 de uma rede direta. A Tabela 5.1 apresenta os resultados do teste de topologia com a rede neural direta.

Tabela 5.1: Resultados do teste de topologia aplicado ao banco de dados ‘Dados 3’ com 2 camadas intermediárias e a estrutura 7xMxNx2 de uma rede neural direta.

Topologia_2	SSE	SSW	N	γ	aC(%)	aT(%)	aS(%)	SSE	aC(%)	aT(%)	aS(%)
7x2x2x2	796,60	84,06	28	25,20	82,86	87,35	72,65	2867,9	86,18	87,42	75,50
7x2x3x2	795,45	85,82	33	26,90	82,92	87,40	72,80	2855,2	86,03	87,21	75,19
7x3x2x2	788,48	111,54	38	33,90	82,86	87,82	72,95	2877,2	86,44	87,42	75,71
7x3x3x2	786,70	152,59	44	38,30	83,27	87,91	73,27	2889,2	86,29	87,78	75,81
7x3x4x2	783,49	140,20	50	41,70	83,24	87,49	73,04	2875,0	86,44	87,42	75,76
7x3x5x2	783,43	142,45	56	42,70	83,16	87,55	72,95	2873,7	86,39	87,57	75,81
7x4x2x2	781,02	113,90	48	40,20	82,77	88,05	73,24	2877,7	86,59	87,78	76,48
7x4x3x2	778,13	144,28	55	47,30	83,04	87,96	73,27	2832,3	86,39	88,14	76,58
7x4x4x2	775,57	145,10	62	50,70	83,10	87,96	73,01	2913,9	86,70	87,78	76,12
7x4x5x2	774,36	145,71	69	54,30	83,16	87,91	73,19	2922,5	86,90	87,88	76,43
7x4x6x2	774,30	142,80	76	54,60	83,16	87,99	73,22	2922,3	86,90	88,03	76,48
7x5x2x2	771,13	205,52	58	51,70	83,13	88,29	73,57	2905,0	86,80	87,98	76,37
7x5x3x2	771,01	150,73	66	55,30	83,19	87,85	73,22	2864,9	86,90	87,67	76,58
7x5x4x2	766,73	206,23	74	61,90	82,98	87,85	72,92	2841,9	86,34	87,83	76,12
7x5x5x2	769,64	158,64	82	62,30	83,24	87,94	73,27	2857,9	86,34	87,98	76,27

Os resultados apresentados na Tabela 5.1 mostram que a topologia da rede direta que apresenta o maior acerto simultâneo é a 7x5x3x2 (aS = 76,58%).

Como foi necessário o desenvolvimento de uma rede neural inversa, foi realizado o treinamento da rede inversa com a melhor estrutura encontrada para a rede direta, ou seja, 7x5x3x2. O critério de acerto para avaliar a generalização do modelo é um intervalo de acerto para o desvio da variável de saída referente ao minério de ferro de $\pm 500\text{kg}$ e para o desvio da variável referente ao oxigênio adicionado no período de resopro de $\pm 300\text{Nm}^3$. A Figura 5.1 apresenta a estrutura neural inversa empregada.

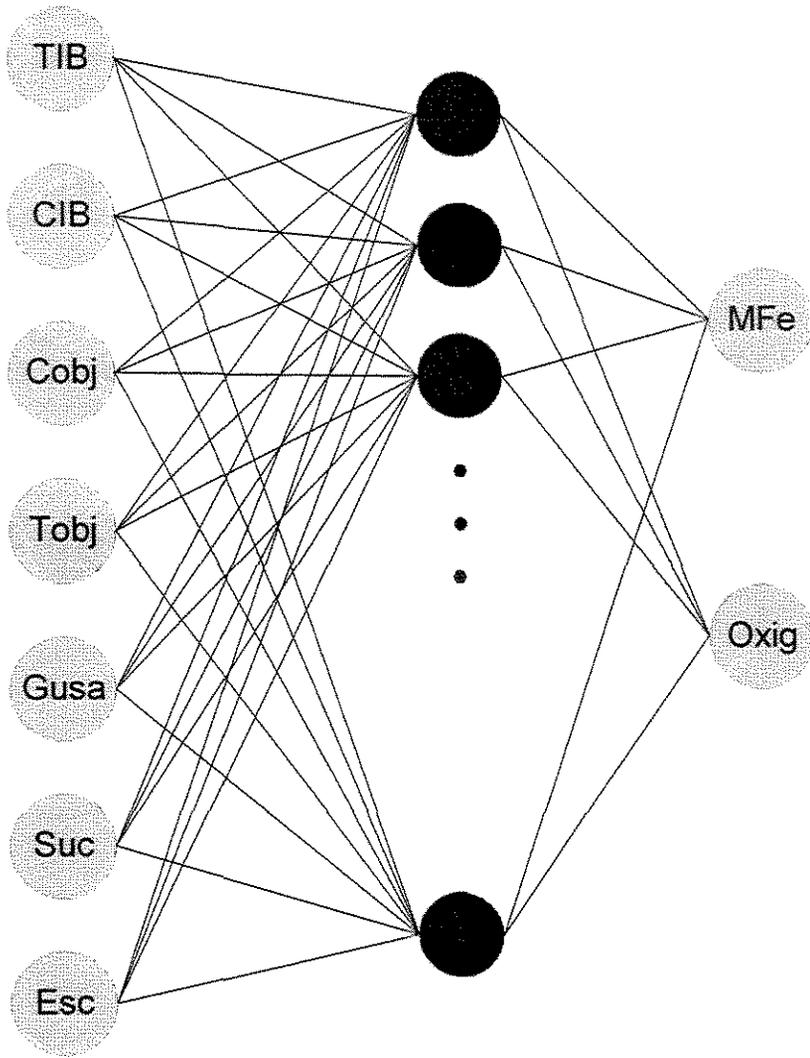


Figura 5.1: Rede Neural Inversa para o modelo dinâmico

A Tabela 5.2 apresenta os resultados do treinamento e da generalização da rede inversa com a topologia 7x5x3x2.

Tabela 5.2: Resultados do treinamento e da generalização da rede inversa com topologia 7x5x3x2

Teste	Resultados do Treinamento							Resultados do Teste			
	SSE	SSW	N	γ	aMFe(%)	aOxig(%)	aS(%)	SSE	aMFe(%)	aOxig(%)	aS(%)
7x5x3x2	265,08	308,65	66	59,8	86,31	94,63	81,89	1860,2	93,37	94,76	88,70

A tabela 5.2 mostra que a rede neural inversa fornece na generalização um acerto da quantidade de minério de ferro de $aMFe = 93,37\%$ e de $aOxig = 94,76\%$ para a quantidade de oxigênio utilizada para resopro. Cunha obteve para o mesmo modelo, utilizando a topologia $4 \times 9 \times 3 \times 2$ para o banco de dado 'Dados 1', um acerto de 80% para a quantidade de minério de ferro e de 97% para a quantidade de oxigênio utilizada para resopro.

Desta forma, foi montada no Microsoft Excel uma planilha que executa os cálculos da rede neural inversa, com estrutura $7 \times 5 \times 3 \times 2$, a partir dos pesos e dos bias obtidos no treinamento. O operador digita nas células especificadas as entradas do modelo diante das medições e especificações do processo no instante *in blow* (descida da sublança) e obtém, de forma instantânea, as saídas, ou seja, as quantidades de minério de ferro e de oxigênio a serem adicionadas para o resopro. A Figura 5.2 apresenta a planilha desenvolvida para o teste industrial.

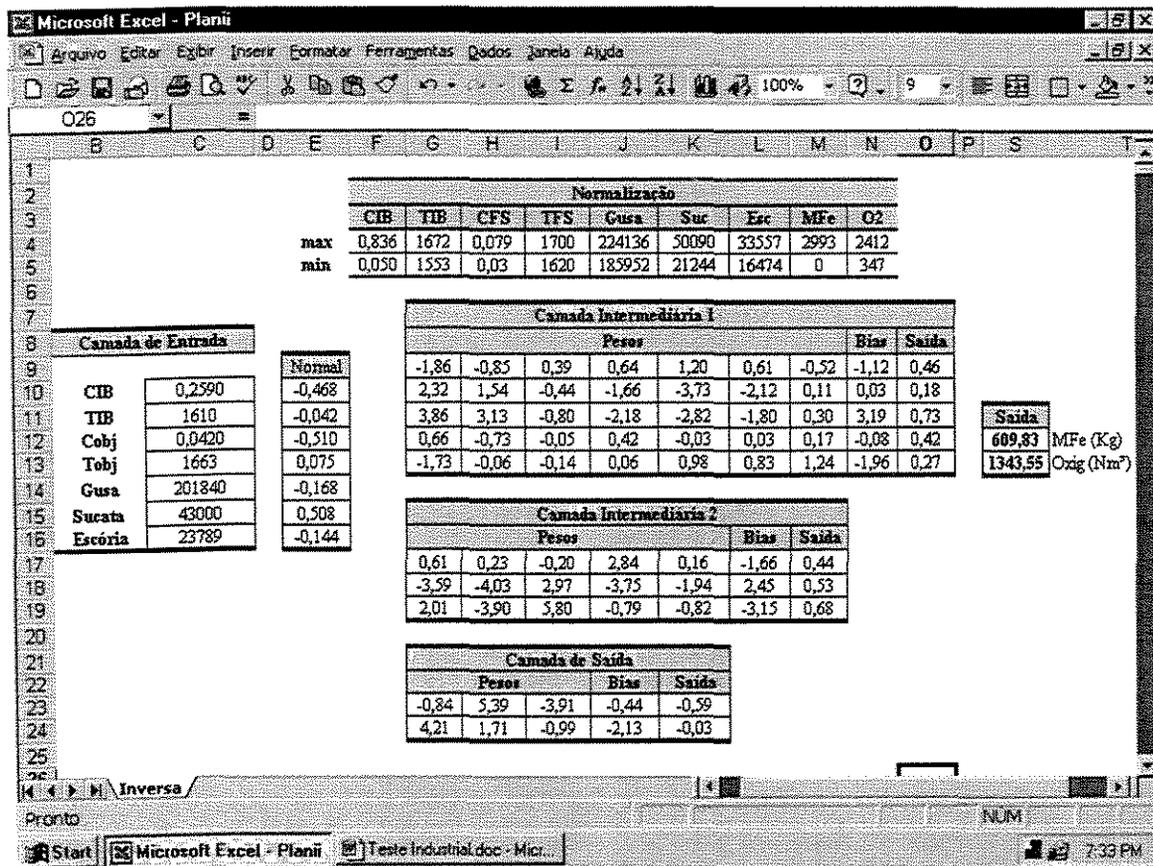


Figura 5.2: Planilha do Excel construída para realizar os cálculos da rede neural inversa com topologia $7 \times 5 \times 3 \times 2$

Com a planilha devidamente preparada determinou-se uma seqüência de atividades para o teste industrial com o modelo de rede neural inversa. A Figura 5.3 apresenta esta seqüência.

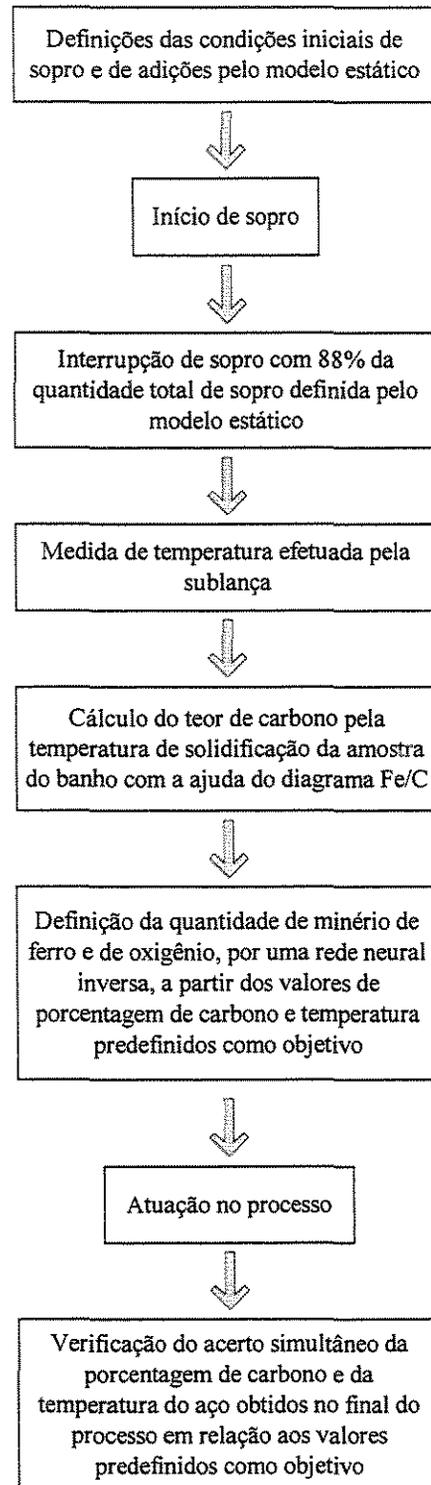


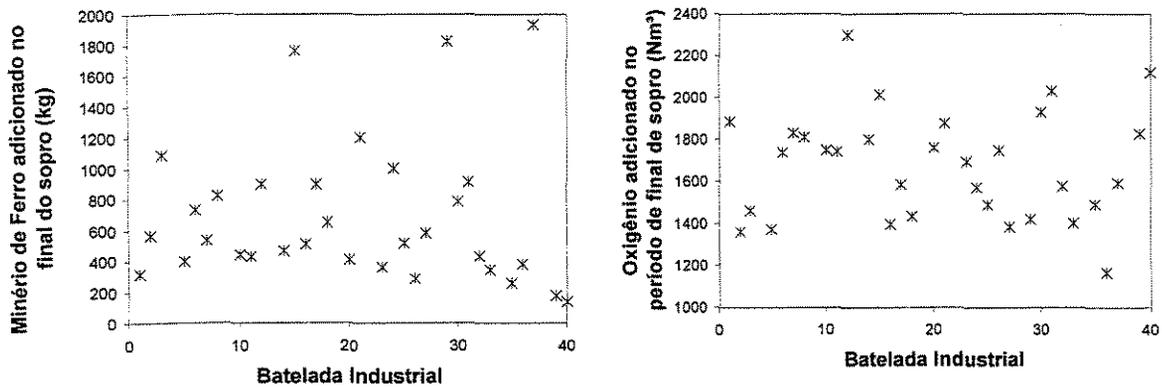
Figura 5.3: Seqüência de atividades para teste industrial com modelo de rede neural inversa.

V.2 Resultados

Seguindo a seqüência de atividades propostas acima foi realizado o teste industrial no processo de aciaria da CSN.

De início verificou-se a ocorrência de um desvio sistemático de temperatura e porcentagem de carbono. Isto levou o operador a corrigir os valores de adição de minério de ferro e de oxigênio previstos pela rede inversa através de um acréscimo de 200kg de minério e de 150 Nm³ de oxigênio.

Isto posto, foram avaliadas 40 bateladas realizadas no mesmo conversor. A Figura 5.4 apresenta os valores de volume de oxigênio e de quantidade de minério de ferro adicionadas em cada batelada a partir dos valores obtidos com a rede inversa.



a)

b)

Figura 5.4: Valores de a) Minério de Ferro e b) Oxigênio adicionados no final de sopro a partir dos resultados obtidos com a rede neural inversa no teste industrial

O valor de acerto de porcentagem de carbono obtido com este teste foi de 82,50% e o de temperatura foi de 97,50%, resultando num acerto simultâneo de 82,50%. A Figura 5.5 apresenta os gráficos dos desvios (valor real – valor objetivado) de porcentagem de carbono e de temperatura obtidos com os valores das bateladas do teste industrial.

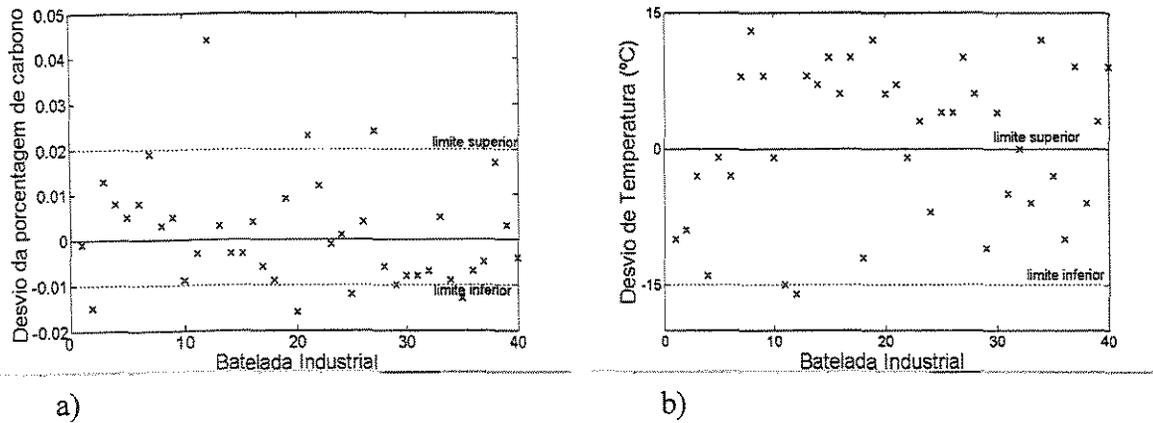


Figura 5.5: Gráficos de desvios de: a) porcentagem de carbono; b) Temperatura, obtidos a partir do teste industrial.

Observando os vetores que a rede não consegue prever com sucesso, verificou-se que a maior incidência de erros ocorre quando a porcentagem de carbono objetivada é de 0,05 %. Diante disto, foi realizada uma avaliação em separado das bateladas em que os valores de carbono objetivado são 0,04% e 0,05%. Estes valores são os que aparecem com maior frequência no período de teste.

Desta forma, para a porcentagem objetivada de carbono de 0,04% o acerto da temperatura é de 96,9% e o da porcentagem de carbono é de 90,6% resultando em um acerto simultâneo de 90,6%. Já para a porcentagem de 0,05 % de carbono, o acerto de temperatura é de 100,00% e os acertos de carbono e o simultâneo são 50,00% cada. Portanto, sugere-se que se trabalhe com duas redes distintas, uma com 0,04% e outra com 0,05% de carbono.

V.3 Conclusões:

Realizado o teste industrial com 40 bateladas obteve-se um acerto simultâneo e de porcentagem de carbono de 82,5%. Já para a variável temperatura do aço o acerto é de 97,5%.

Na prática, o objetivo da variável temperatura de final de sopro é mais difícil de ser atingido que o objetivo da variável porcentagem de carbono, primeiramente, porque o desvio máximo permitido é de ordem bastante inferior ao permitido para a porcentagem de carbono. Isto se deve também ao fato de que a principal ação corretiva para o ajuste da

temperatura é implementada através da adição de minério de ferro. Entretanto, este minério de ferro apresenta fortes oscilações granulométricas e de composição química, o que **não** mantém padronizado o efeito refrigerante da adição desta variável. Além disto, o equipamento industrial que efetua a adição da quantidade calculada de minério é bastante impreciso, introduzindo erros que parecem ter sido capturados durante o treinamento do modelo neural.

Apesar do acerto simultâneo se apresentar similar, sob este aspecto, o presente modelo apresenta resultados superiores que os obtido por Cunha (2001), já que este último obteve acerto de 100,0% para a porcentagem de C e de 83,3% para a temperatura final de sopro e o modelo desenvolvido no presente trabalho apresenta 82,5% para o carbono e 97,5% de acerto para a temperatura.

O capítulo seguinte apresenta as conclusões finais deste trabalho.

CAPÍTULO VI: CONCLUSÕES E SUGESTÕES

Diante dos resultados apresentados nos capítulos anteriores, este capítulo apresenta as conclusões finais e sugestões deste trabalho.

VI.1 Conclusões

Esta dissertação aplica redes neurais artificiais ao processo de produção de aço líquido com o objetivo de elaboração de dois modelos: a rede direta, para predição de temperatura e porcentagem de carbono ao final do processo, e a rede inversa, para ajuste operacional das quantidades de oxigênio para o término do sopro e de minério de ferro a ser adicionado no período final da batelada.

O modelo neural direto proposto neste trabalho é composto das seguintes variáveis de entrada: a temperatura e o percentual de carbono medidos através da sub-lança, as quantidades de minério de ferro e de oxigênio determinadas pelo modelo convencional vigente e, opcionalmente, as quantidades de ferro-gusa, sucata e escória alimentadas, e as porcentagens de manganês, fósforo e enxofre tomadas para medição com 88% de sopro. As variáveis de saída são as mesmas do modelo dinâmico comercial utilizado atualmente na CSN: a porcentagem de carbono e a temperatura do aço no período de final de sopro e opcionalmente as porcentagens de manganês, fósforo e enxofre de final de sopro.

A fim de encontrar novos modelos neurais que elevem a porcentagem de acerto simultâneo do modelo dinâmico encontrada por Cunha, foram realizados testes para definir as melhores condições para os modelos neurais. Avaliando os resultados que apresentaram o maior acerto simultâneo de carbono e de temperatura do aço definiram-se as seguintes características para o modelo neural: intervalo de normalização de $[-1;1]$; funções de ativação log-sigmoïdal e linear para as camadas intermediária e final, respectivamente; e o algoritmo de treinamento de *Levenberg-Marquardt* com regularização bayesiana. Como era esperado, este algoritmo apresentou os melhores resultados já que possui uma modificação na função-objetivo que procura evitar o sobreajuste dos dados.

As tentativas de alterar o banco de dados com a aleatoriedade dos dados, ou a redução da dimensionalidade do problema através da ferramenta de análise de componentes principais não aumentaram o acerto simultâneo de porcentagem de carbono e de temperatura do aço na generalização.

Os testes de topologia para o banco de dados ‘Dados1’ (mesmo de Cunha) apresentaram as estruturas 4x9x2 (aS = 76,50%) e 4x3x6x2 (aS = 76,50%) com os maiores acertos simultâneos.

Já com a adição de novas variáveis ao modelo (‘Dados 2’) é possível verificar um maior acerto simultâneo nos resultados do modelo neural que apresenta todas as variáveis disponíveis tanto para entrada quanto para a saída (10x5x12x5 – aS = 78,25%). No entanto, como os vetores de entrada e saída de porcentagem de manganês, fósforo e enxofre não são mensuráveis *on-line*, foi realizado um teste de topologia sem estas variáveis. Assim sendo, a topologia que apresentou na generalização um maior acerto simultâneo foi a 7x4x4x2 (aS = 77,00%).

Com as melhores condições do modelo neural definidas, foi proposto um teste industrial, para verificar a aplicação do modelo obtido.

Diante da necessidade de trabalhar com as variáveis manipuladas deste processo, ou seja, a quantidade de minério de ferro e de oxigênio utilizadas no resopro, o modelo dinâmico neural desenvolvido neste trabalho teve que sofrer algumas modificações. Assim, foi desenvolvido um modelo neural inverso, ou seja, as variáveis de entrada passam a ser aquelas medidas pela sublança e os valores que são os objetivos pré-estabelecidos de porcentagem de carbono e temperatura para o fim de sopro. As variáveis de saída são as quantidades de minério de ferro e de oxigênio necessárias para atingir os valores objetivados.

Adicionando-se mais informações de operação ao banco de dados ‘Dados2’, realizou-se um novo teste de topologia. Estes dados se referem a condições operacionais de não-adição de minério de ferro. O melhor resultado foi obtido com a topologia 7x5x3x2, a qual foi mantida para a estrutura da rede inversa.

As medições de temperatura e de porcentagem de carbono ao final de 40 bateladas mostraram que em 82,5% destas bateladas as ações corretivas, indicadas pela rede inversa e implementadas pelo operador, conduziam o processo às especificações desejadas de temperatura e porcentagem de carbono simultaneamente. A especificação de temperatura é atingida em 97,5% das bateladas. Observou-se ainda que a especificação de porcentagem de carbono é mais facilmente alcançada (90,6%) quando situa-se próxima a 0,04% de carbono. Conclui-se, portanto, que os modelos neurais desenvolvidos apresentaram

desempenho de predição superior ao modelo dinâmico convencional utilizado atualmente pela CSN e que a rede inversa está apta a fornecer informações corretivas para o controle operacional deste processo.

Na prática, o objetivo da variável temperatura de final de sopro é mais difícil de ser atingido que o objetivo da variável % de C, primeiramente, porque o desvio máximo permitido é de ordem bastante inferior ao permitido para a porcentagem de Carbono. Isto se deve também ao fato de que a principal ação corretiva para o ajuste da temperatura é implementada através da adição de minério de ferro. Entretanto, este minério de ferro apresenta fortes oscilações granulométrica e de composição química, o que não mantém padronizado o efeito refrigerante da adição desta variável. Além disto, o equipamento industrial que efetua a adição da quantidade calculada de minério é bastante impreciso, introduzindo erros que parecem ter sido capturados durante o treinamento do modelo neural. Apesar do acerto simultâneo se apresentar similar, sob este aspecto, o presente modelo apresenta resultados superiores que os obtido por Cunha (2001), já que este último obteve acerto de 100% para a porcentagem de C e de 83,3% para a temperatura final de sopro e o modelo desenvolvido no presente trabalho apresenta 82,6% para o carbono e ~100% de acerto para a temperatura.

Desta forma, o modelo neural desenvolvido foi de suma importância para o ajuste fino do processo siderúrgico da CSN. Quando se trata de siderurgia, qualquer resultado por menor que possa parecer, tem um grande impacto no resultado técnico-econômico das empresas, que produzem milhões de toneladas de aço por ano.

VI.2 Sugestões

O modelo dinâmico atua após a medida de porcentagem de carbono e de temperatura pela sublança, assim, a altura da lança é uma informação importante que pode elevar o acerto simultâneo do modelo neural. Seria interessante gerar um registro desta altura ou ainda padronizar a operação de forma que esta altura seja mantida constante para todas as bateladas.

Como o modelo dinâmico neural desenvolvido neste trabalho não foi treinado com dados referentes à produção de aço com carga líquida (sem adição de sucata), seria interessante desenvolver um modelo só com este tipo de dados.

Outra sugestão seria trabalhar com bancos de dados particulares para cada conversor, pois é notada na prática a diferença de comportamento entre eles. E também pode-se trabalhar com bancos de dados diferenciados para cada grau de aço objetivado diferente, gerando modelos neurais dinâmicos específicos.

REFERÊNCIAS BIBLIOGRÁFICAS

BRAGA, A.P.; CARVALHO, A. C. P. L. F. C; LUDERMIR, T. B. *Sistemas Inteligentes: fundamentos e aplicações*. 1ª edição Reading: Rezende, S.O., 2003. Capítulo 6: Redes Neurais Artificiais, p. 141-168.

CHESTER, D.L; *Why two hidden layers are better than one*. In *International Joint Conference on Neural Network I*, 1990 vol.1, 265-268, Washington, D.C., apud HAYKIN, S. *Redes Neurais – Princípios e práticas*. Porto Alegre-RS-Brasil: Bookman 2001.

COX, I. J.; LEWIS, R. W.; RANSING, R. S.; LASZCZEWSKI, H; BERNI G. *Application of neural computing in basic oxygen steelmaking*. *Journal of Materials Processing Technology* 120, p. 310–315, 2002.

CUNHA, A. P. *Redes Neurais em processos siderúrgicos: analisador virtual de propriedades metalúrgicas do sinter e modelo de predição de qualidade do aço*. Campinas: Faculdade de Engenharia Química, Universidade Estadual de Campinas – UNICAMP, 2001. Tese (Doutorado)

DEMUTH,H.; BEALE,M.; *Neural Network Toolbox User's Guide for Use with MATLAB®*. Version 4, Reading: Hagan,M., 2002 Cap. 5: Backpropagation. URL: <http://www.mathworks.com/support/product/NN/>

FIORA, E. *Aço – Radiografia do setor siderúrgico* CREA – SP, Ano III, nº 10, Jul/Ago 2003

FRUEHAN, R.J. *The Making, Shaping and Treating of Steel - 11th Edition* Steelmaking and Refining Volume, Reading: MILLER, T. W.; JIMENEZ, J.; SHARAN, A.; GOLDSTEIN, D. A. Cap. 9: Oxygen Steelmaking Processes, p.475-478, *The AISE Steel Foundation*.,1998.

FUNAHASHI,K.; *On the approximate realization of continuous mappings by neural networks*. *Neural Networks*, vol.2, p. 183-192, 1989 apud HAYKIN, S. *Redes Neurais – Princípios e práticas*. Porto Alegre-RS-Brasil: Bookman 2001

HAGAN, M.T.; FORESSE,F,D *Gauss-Newton Approximation to Bayesian Learning*. In *Proceedings of the 1997 International Joint Conference on Neural Networks*, 1997 URL: <http://elec-engr.okstate.edu/mhagan/>

HAYKIN, S. *Redes Neurais – Princípios e práticas*. Porto Alegre-RS-Brasil: Bookman 2001

HERSKOVIC, J. *Elaboração do aço: fusão e refino*. Reading: Mundim, M. J., 1976 Capítulo IV: O Processo LD Associação Brasileira de Metais, ABM; São Paulo, SP; 1976.

HUSSAIN, M.A.; *Review of the applications of neural networks in chemical process control – simulation and online implementation*. Artificial Intelligence in Engineering, 13, p. 55-68, 1999.

INSTITUTO BRASILEIRO DE SIDERURGIA – IBS – COMISSÃO DE COMUNICAÇÃO SOCIAL *Aço e Siderurgia*. IBS, 1984.

JOCKYMAN, A.; SILVA, M. J. *Rede neural é mais um salto da siderurgia brasileira, seguindo a tendência mundial da atualização tecnológica*. Revista Metalurgia & Metais, nº 526, Outubro de 2002.

MALINOVA,T.; MALINOV,S.; PANTEV,N. *Simulation of microhardness profiles for nitrocarburized surface layers by neural network*. Surface and Coatings Technology 135 p.258-267, 2001.

MALYNOWSKYJ, A. *Aplicação de redes neurais artificiais ao modelo físico-químico do conversor LD na fabricação de aço líquido*. São Paulo: Escola Politécnica da USP, 1997. Tese (Doutorado)

PAULA, G. M. *Desenvolvimento Tecnológico, Alianças Estratégicas e Apoio Governamental: o caso Tecnored*. Instituto de Economia - Universidade Federal de Uberlândia - Dezembro de 2001

RAUCH, H.E. *A control engineer's use of artificial intelligence*. Control Engineering Practice 6 p.249-258, 1998

SHAW, I. S.; SIMÕES, M. G. *Controle e modelagem FUZZY*. São Paulo: Editora Edgard Blücher Ltda; 1ª edição, 1999

ANEXO I

Este anexo contém o código para treinamento e generalização da RNA desenvolvido com o *Software* MATLAB[®].

```
%_____Código de treinamento e generalização da RNA _____
arq=load('Seltr_D.txt');
p=arq(:,1:7); t=arq(:,8:9);
[pn,minp,maxp] = premnmx(p);
[tn,mint,maxt] = premnmx(t);
net=newff([minmax(pn)], [9 2], {'logsig' 'purelin'}, 'trainbr');
net=init(net);
[net,tr] = train(net,pn,tn);
pr=load('Sel_D.txt');
p2=pr(:,1:7); t2=pr(:,8:9);
nt=1947;
[p2n] = tramnmx(p2,minp,maxp);
[ps]=sim(net,p2n);
[psim] = postmnmx(ps,mint,maxt); a=psim';
ac=a(:,1); aT=a(:,2); t2c=t2(:,1); t2T=t2(:,2);
ec=ac-t2c; eT=aT-t2T;
[t2n,mint2,maxt2]=premnmx(t2); an=ps';
for i=1:nt
    if (ec(i)>=(-0,01)&ec(i)<=0,02)
        yc(i)=1;
    else
        yc(i)=0;
    end
end
acerto_c=sum(yc)*100/nt;
for i=1:nt
    if (eT(i)>=(-15)&eT(i)<=15)
        yT(i)=1;
    else
        yT(i)=0;
    end
end
acerto_T=sum(yT)*100/nt;
acerto(:,1)=yc'; acerto(:,2)=yT';
for i=1:nt
    if (acerto(i,1)==1&acerto(i,2)==1)
        yS(i)=1;
    else
        yS(i)=0;
    end
end
acerto_simultaneo=sum(yS)*100/nt ;
nt_tr=3390;
```

```
[pstr]=sim(net,pn);
[psimtr] = postmnmx(pstr,mint,maxt); atr=psimtr';
ac_tr=atr(:,1); aT_tr=atr(:,2);
t3c=t(:,1); t3T=t(:,2);
ec_tr=ac_tr-t3c; eT_tr=aT_tr-t3T;
for i=1:nt_tr
    if (ec_tr(i)>=(-0,01)&ec_tr(i)<=0,02)
        yc_tr(i)=1;
    else
        yc_tr(i)=0;
    end
end
acerto_c_tr=sum(yc_tr)*100/nt_tr;
for i=1:nt_tr
    if (eT_tr(i)>=(-15)&eT_tr(i)<=15)
        yT_tr(i)=1;
    else
        yT_tr(i)=0;
    end
end
acerto_T_tr=sum(yT_tr)*100/nt_tr;
acerto_tr(:,1)=yc_tr'; acerto_tr(:,2)=yT_tr';
for i=1:nt_tr
    if (acerto_tr(i,1)==1&acerto_tr(i,2)==1)
        yS_tr(i)=1;
    else
        yS_tr(i)=0;
    end
end
acerto_simultaneo_tr=sum(yS_tr)*100/nt_tr;
```

ANEXO II

Este anexo contém os pesos e os *bias* dos modelos neurais selecionados neste trabalho.

Para o banco de dados ‘Dados 1’ os modelos selecionados foram: 4x9x2 para uma camada escondida e 4x3x6x2 para duas camadas escondidas.

A Tabela A1 apresenta a matriz pesos e os *bias* da camada intermediária obtida a partir do treinamento do banco de dados ‘Dados 1’ com o modelo neural 4x9x2.

Tabela A1: Matriz de pesos e *bias* da camada intermediária do modelo neural com topologia 4x9x2 – ‘Dados1’

Matriz de pesos				<i>bias</i>
-0,9704	-1,4847	-2,0667	0,5713	0,7548
0,5287	-1,3232	-0,6377	-0,0896	-0,5482
0,9535	-0,748	1,5423	0,8956	-0,2915
-2,7349	0,3088	-1,7783	1,4949	1,2036
0,8127	-3,1172	2,9646	-0,2555	-0,4605
2,0262	0,6533	-0,8386	-0,0925	-0,9498
-1,8093	-3,1858	0,388	0,9322	0,9032
-1,1217	-1,9916	-0,4408	-0,0363	1,2782
0,0062	0,0918	-0,2669	1,3448	-0,6046

A matriz de pesos e os *bias* da camada de saída do modelo neural 4x9x2 obtida no processo de treinamento está exposta na Tabela A2.

Tabela A2: Matriz de pesos e os *bias* da camada de saída do modelo neural com topologia 4x9x2 – ‘Dados 1’

Matriz de pesos									<i>bias</i>
1,1121	-1,8028	-1,7353	-0,5098	1,8214	0,4550	0,3290	-1,3719	0,7663	0,4594
0,2311	1,1785	0,8939	-1,1733	-0,4411	-1,6142	-2,7439	1,4101	1,4164	0,5182

As Tabelas A3, A4 e A5 apresentam as matrizes de pesos e os *bias* resultantes do treinamento do banco de dados ‘Dados 1’ com o modelo neural 4x3x6x2.

Tabela A3: Matriz de pesos e *bias* da primeira camada intermediária do modelo neural com topologia 4x3x6x2 – ‘Dados1’

Matriz de pesos				<i>bias</i>
1,1538	5,0031	1,3350	-0,2335	-0,7552
-1,8872	1,6985	-4,1743	1,3186	0,5253
-2,7462	-3,3490	-0,2534	1,1086	0,8357

Tabela A4: Matriz de pesos e os *bias* da segunda camada intermediária do modelo neural com topologia 4x3x6x2 – ‘Dados1’

Matriz de pesos			<i>bias</i>
-3,1691	-3,9768	3,6480	-0,1300
-0,6983	-0,1088	2,8988	0,4537
2,4544	0,0487	-0,1396	-0,7405
-2,4304	3,3779	0,3376	-0,8651
-2,8351	-2,5890	5,1752	1,8439
-2,1041	-2,3362	-1,9568	2,4640

Tabela A5: Matriz de pesos e os *bias* da camada de saída do modelo neural com topologia 4x3x6x2 – ‘Dados 1’

Matriz de pesos						<i>bias</i>
1,4425	-2,2214	3,0280	1,1884	1,2501	0,9001	-2,4027
-1,4715	3,1439	-0,5291	-0,1721	-2,0735	2,5387	-1,1880

Com o banco de dados ‘Dados 2’, a topologia que resultou num maior acerto simultâneo foi a 10x5x12x5. As Tabelas A6, A7 e A8 apresentam as matrizes de pesos e os *bias* ajustados no processo de treinamento para cada camada da rede neural.

Tabela A6: Matriz de pesos e *bias* da primeira camada intermediária do modelo neural com topologia 7x4x4x2 – ‘Dados 2’

Matriz de Pesos							<i>bias</i>
0,2930	-2,0872	1,0929	-3,2524	0,5493	0,2619	0,0092	1,1227
-3,3138	0,7844	1,7926	0,7102	0,5118	-0,6807	-0,3996	-1,1595
4,5862	-1,6777	-2,5518	0,3890	-1,1055	2,5894	-0,9255	-0,2418
-1,4303	-1,5470	-0,1088	1,4361	-1,1208	-0,1173	-0,2702	-0,9065

Tabela A7: Matriz de pesos e *bias* da segunda camada intermediária do modelo neural com topologia 7x4x4x2 – ‘Dados 2’

Matriz de Pesos				<i>bias</i>
-3,6306	0,6878	0,0354	-0,0681	2,8185
-1,0950	-0,9250	1,3093	-1,4769	-0,6748
-1,8411	-0,8664	-2,6083	4,7164	-1,7438
-0,9680	1,6895	0,9353	3,1285	-0,2331

Tabela A8: Matriz de pesos e *bias* da camada de saída do modelo neural com topologia 7x4x4x2 – ‘Dados 2’

Matriz de Pesos				<i>bias</i>
-0,8851	2,3887	3,4655	-2,7305	1,7449
2,8984	0,1398	-1,7364	-0,6656	-1,2772