



UNICAMP

**UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA QUÍMICA
DESENVOLVIMENTO DE PROCESSOS QUÍMICOS**

**MODELAGEM DE FORNO PARA A PRODUÇÃO DE
CIMENTO ATRAVÉS DA APLICAÇÃO DE TÉCNICAS DE
REDES NEURAIIS**

Edson Guaracy Lima Fujita

Autor

Prof. Dr. Milton Mori

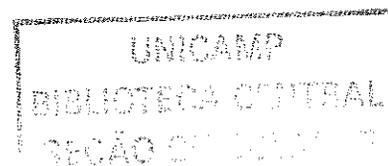
Orientador

**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 - SP

Outubro / 2001

UNICAMP
BIBLIOTECA CENTRAL



UNIDADE	30
Nº CHAMADA	T/UNICAMP
	F955m
V	EX
TOMBO BC.	48512
PREC.	16.837/02
PREÇO	R\$ 11,00
DATA	
Nº CPD	

CM00166500-4

1B ID 237920

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

<p>F954m F955m</p>	<p>Fujita, Edson Guaracy Lima. Modelagem de forno para a produção de cimento através da aplicação de técnicas de redes neurais / Edson Guaracy Lima Fujita.--Campinas, SP: [s.n.], 2001.</p> <p>Orientador: Milton Mori. Dissertação (mestrado) - Universidade Estadual de Campinas, Faculdade de Engenharia Química.</p> <p>1. Cimento - industria. 2. Cimento portland -industria. 3. Fornos. 4. Redes neurais (Computação). I. Mori, Milton. II. Universidade Estadual de Campinas. Faculdade de Engenharia Química. III. Título.</p>
-----------------------------------	--

Dissertação de Mestrado defendida por Edson Guaracy Lima Fujita e aprovada em 04 de outubro de 2001 pela banca examinadora constituída pelos doutores:

Milton Mori

Prof. Dr. Milton Mori

Aline e da Costa

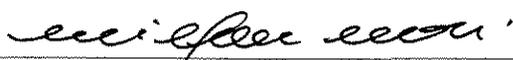
Dra. Aline Carvalho da Costa

Lincoln Moro

Dr. Lincoln Fernando Lautenschlager Moro

2002.17651

Esta versão corresponde à redação final da Dissertação de Mestrado em Engenharia Química.



Prof. Dr. Milton Mori

Orientador

DEDICATÓRIA

Este trabalho é dedicado aos meus pais, Takeshi e Circe, por seu amor e seus ensinamentos, tão valiosos em toda a minha vida.

Dedico à minha esposa, por seu amor, carinho, atenção e compreensão, especialmente nas horas difíceis.

Dedico aos meus filhos, realização e continuidade de nossa vida, com especial amor e carinho, para que, com perseverança, nunca deixem de buscar a concretização de seus sonhos.

E também dedico a todos meus familiares e amigos, que sempre me apoiaram e constantemente vibram com todas as minhas conquistas.

AGRADECIMENTOS

Agradeço em primeiro lugar a Deus, por todos os dons que diariamente me concede.

Agradeço ao meu orientador e amigo, Prof. Dr. MILTON MORI, que me deu todo apoio e suporte para execução deste trabalho.

Agradeço à VOTORANTIM S. A., representada pelos engenheiros HERBERT FOSTER e JOÃO JAYME IEISS, que forneceu os dados para a realização deste trabalho.

Agradeço aos professores da UNICAMP, que contribuíram com valiosas orientações e conhecimentos para a elaboração deste trabalho e aos meus colegas do laboratório – LMSPQ, pelo constante apoio, pelo seu companheirismo, e especialmente à Karla pelas preciosas observações a respeito deste trabalho.

RESUMO

O cimento é um dos mais importantes produtos disponíveis para a sociedade, atualmente. Sua versatilidade permite as mais diversas aplicações. O cimento é utilizado para a fabricação de postes, cisternas, valas, tubos de drenagem, assim como para a construção de muros, calçadas, casas, prédios, pontes e barragens. Também pode ser utilizado na preservação do meio ambiente seja pelo aproveitamento, no seu processo produtivo, de pneus velhos, como fonte de ferro e de energia, ou para a imobilização de produtos tóxicos ou radioativos.

Para ter essa flexibilidade de uso, são necessários rigorosos controles na sua fabricação e rigidez na manutenção de suas especificações.

O modelo desenvolvido neste trabalho, com base em técnicas de redes neurais, tem por objetivo inferir propriedades de qualidade do clínquer na saída do forno, de forma rápida e precisa, antecipando o resultado das análises de laboratório, e com isso, possibilitando a correção das condições de operação do forno num menor tempo. Assim, há a redução da variabilidade operacional do produto, e com isso a garantia de melhor qualidade do produto final. O modelo do forno é composto de duas redes neurais, uma predizendo a Cal Livre e outra predizendo o Peso Litro. Ambas redes possuem vinte entradas e dois nós na camada intermediária. O total de pesos de cada rede foi otimizado com a aplicação da técnica “*Optimal Brain Surgeon*”, ficando reduzidos a 38 e 31 pesos, sem contar com o *bias*, para predição de Cal Livre e Peso Litro, respectivamente.

Neste trabalho é efetuada, também, uma análise de sensibilidade das variáveis de entrada, identificando seu efeito nas variáveis inferidas, a Cal Livre e o Peso Litro do clínquer, de modo a priorizá-las quanto à sua importância, o que permite a sua manipulação quando da definição da estratégia de controle do forno, seja por meio da ação operador, ou de sistema avançado de controle, por computador.

ABSTRACT

CEMENT PRODUCTION FURNACE MODELING: A NEURAL NETWORK TECHNIC APPLICATION.

Cement is one of the most important products available nowadays. Its wide versatility permits different applications. It's used for manufacturing poles, draining tubes, cisterns and ditches, for building sidewalks, houses, bridges and dams. It can be helpful for preserving the environment consuming used tires for iron and energy supply in its production, or for toxic or radioactive products immobilization.

For these wide flexibility, rigorous manufacturing controls, and hard tying to the specifications, are needed.

The proposed neural artificial model objective is to give to the operator quick and precise clinker quality information, at the output of the rotary kiln, anticipating the laboratory analysis sampling results, minimizing the operational furnace conditions corrections, reducing the product operational variability, assuring a better final product quality. The model consists of two neural nets, one for Free Lime and the other for the clinker Density prediction. Both have twenty inputs and two neurons in the hidden layer. The total number of connections were optimized by "Optimal Brains Surgeon" technique, after what the number of weights were reduced to 38 and 31, without taking in count the bias, for Free Lime and for Density prediction, respectively. An input variables sensibility analysis was performed for checking their importance in Free Lime and in Density prediction. The results can be used for the priority manipulation by the operational strategy: the operator or the computer advanced control.

An additional analysis is done for ranking the input variables concerning their effect on the studied variables, the clinker's Free Lime and Density, helping the operators to better understand and define the furnace operation strategy.

SUMÁRIO

DEDICATÓRIA.....	v
AGRADECIMENTOS.....	vi
RESUMO	vii
ABSTRACT.....	viii
SUMÁRIO	ix
LISTA DE FIGURAS.....	xii
GLOSSÁRIO.....	15
Capítulo 1	16
INTRODUÇÃO.....	16
Capítulo 2	17
DESCRIÇÃO DO PROCESSO	17
Capítulo 3.....	24
MATERIAIS E MÉTODOS.....	24
Local e programas utilizados	24
Modelagem utilizando Redes Neurais.....	24
Tipos de Redes Neurais.....	30
Tipos de Redes Neurais, quanto à sua estrutura	30
Redes Recorrentes	32
Redes Feedforward.....	32
Rede MLP.....	33
Tipos de Redes quanto ao algoritmo de treinamento.....	34

Algoritmos de Aprendizagem Supervisionada	34
Algoritmo Backpropagation (BP).....	35
Metodologia para otimização dos pesos.....	37
Algoritmo “Optimal Brain Surgeon”.....	42
Análise de Significância das Variáveis de Entrada	43
Levantamento dos Dados de Entrada	47
Normalização das Variáveis de Entrada	48
Rede e treinamento iniciais.....	49
Análise de Sensibilidade da Rede Treinada	51
Capítulo 4.....	52
RESULTADOS E DISCUSSÕES	52
Dados de Entrada	52
Treinamento da Rede: 25 nós intermediários, 2 saídas	54
Treinamento da Rede: 11 nós intermediários, 2 saídas	65
Treinamento da Rede: 9 nós intermediários, 2 saídas	66
Número ideal de pesos da rede.....	73
Treinamento da Rede: 2 nós intermediários, 1 saída.....	73
Análise de Sensibilidade dos dados de entrada.....	82
Análise de Sensibilidade: rede inicial, com dois nós intermediários, predizendo Cal Livre	82
Análise de Sensibilidade: rede inicial, com dois nós intermediários, predizendo Peso Litro.....	84
Análise de Sensibilidade: rede com redução dos pesos, predizendo Cal Livre... 	85
Análise de Sensibilidade: rede com redução dos pesos, predizendo Peso Litro. 	87

Capítulo 5.....	89
CONCLUSÕES E RECOMENDAÇÕES	89
Quanto à estrutura de rede	89
Quanto ao procedimento de otimização dos pesos da rede	90
Quanto ao modelo obtido.....	90
Quanto à análise de sensibilidade das entradas para a rede predizendo Cal Livre	90
Quanto à análise de sensibilidade das entradas para a rede predizendo Peso Litro.....	92
Recomendações.....	93
REFERÊNCIAS BIBLIOGRÁFICAS.....	94
Anexos.....	96
DISTRIBUIÇÃO DOS DADOS DE ENTRADA vs SAÍDA	96

LISTA DE FIGURAS

<i>Número</i>	<i>Página</i>
Figura 1 - Processo de produção de cimento mostrando os processos de moagem e de clinquerização.....	17
Figura 2 – Clínquer fluidizado na câmara de combustão do forno rotativo.....	18
Figura 3 - Esquema de Torre de Ciclones.....	19
Figura 4 - Torre de Ciclones e Forno Rotativo	20
Figura 5- Esquema de um neurônio	26
Figura 6 - Se a soma dos sinais de entrada excederem um valor de excitação o neurônio "dispara".....	27
Figura 7 - Os neurônios estão interconectados. Quando um neurônio "dispara", outros neurônios recebem seu sinal e podem disparar também.....	27
Figura 8 - Representação matemática de um neurônio	28
Figura 9 - Esquemas de ligações entre nós de uma Rede Neural.....	31
Figura 10 - Estrutura de uma Rede Neural tipo MLP.....	33
Figura 11 - Rede com treinamento adequado: a função permite boa interpolação entre os dados usados para treinamento (consegue “generalizar”).....	38
Figura 12 – A função obtida da rede com treinamento insuficiente não consegue modelar bem os dados	38
Figura 13 - Rede sobre treinada: identifica os pontos usados para treinamento, mas não permite a interpolação entre os dados usados para treinamento (não consegue “generalizar”).....	39
Figura 14 - Rede inicialmente proposta, com 25 nós intermediários, predizendo Cal Livre e Peso Litro.....	50
Figura 15 – Os dados do teor de oxigênio se distribuem bem em relação à Cal Livre (o – azul), e também e em relação ao Peso Litro (* - vermelho). O valor médio dos dados dessa entrada está relacionado com a média da Cal Livre, marcada com (x) e com a média do Peso Litro, marcada com (+).....	52

Figura 16 - Os dados de emissão de particulados se distribuem mal em relação às respectivas saídas de Cal Livre, marcadas em azul (o) e às respectivas saídas de Peso Litro, marcadas em vermelho (*). O valor médio dos dados dessa entrada está relacionado com a média da Cal Livre, marcada com (x) e com a média do Peso Litro, marcada com (+).	53
Figura 17 - Saída da rede com 25 nós intermediários, predizendo Cal Livre. No segundo gráfico, mostrando os valores calculados e os dados reais, os pontos (*) indicam os dados utilizados para validação. No terceiro gráfico, relacionando os valores calculados e os valores desejados, os pontos marcados com (*) são os dados utilizados para treinamento, enquanto que os dados marcados com (o) são os dados de validação.	55
Figura 18 - Saída da rede com 25 nós intermediários, predizendo Peso Litro. A representação dos pontos é a mesma da Figura 17.	56
Figura 19 - Curva dos erros após retreinamento da rede, depois da redução de pesos com a metodologia OBS	57
Figura 20 – Rede com 25 nós intermediários, após a redução dos pesos para 117	58
Figura 21 - Saída da rede reduzida para 117 pesos, predizendo Cal Livre	59
Figura 22 - Saída da rede reduzida para 117 pesos, predizendo Peso Litro	60
Figura 23 - Comparação entre as saídas da rede inicial e da rede com o número de pesos reduzido para 117	61
Figura 24 - Rede com número de pesos reduzido para 28	62
Figura 25 - Saída da rede reduzida para 28 pesos, predizendo Cal Livre	63
Figura 26 - Saída da rede reduzida para 28 pesos, predizendo Peso Litro	64
Figura 27 - Rede com 11 nós intermediários e 2 nós de saída	65
Figura 28 - Rede inicial com 9 nós na camada intermediária e 2 na camada de saída ..	66
Figura 29 - Saída da rede com 9 nós intermediários predizendo Cal Livre	67
Figura 30 - Saída da rede com 9 nós intermediários predizendo Peso Litro	68
Figura 31 - Rede com 9 nós intermediários e pesos reduzidos para 81	69
Figura 32 - Saída da rede com 9 nós intermediários e 81 pesos, predizendo Cal Livre	70
Figura 33 - Saída da rede com 9 nós intermediários e 81 pesos, predizendo Peso Litro	71

Figura 34 - Comparação entre saídas da rede de 9 nós intermediários, inicial e com pesos reduzidos para 81	72
Figura 35 – Estrutura da rede com 2 nós intermediários, predizendo somente a Cal Livre	74
Figura 36 – Estrutura da rede com 2 nós intermediários, predizendo somente o Peso Litro.....	74
Figura 37 - Resultado inicial da rede com 2 nós intermediários e uma saída, predizendo a Cal Livre	75
Figura 38 – Resultado inicial da rede com 2 nós intermediários e uma saída, predizendo o Peso Litro.....	76
Figura 39 - Rede com 2 nós intermediários e uma saída, após redução de pesos, predizendo Cal Livre.....	77
Figura 40 - Rede com 2 nós intermediários e uma saída, após redução de pesos, predizendo Peso Litro	77
Figura 41 - Saídas da rede com 2 nós intermediários e 1 saída, predizendo Cal Livre, após redução de pesos.....	78
Figura 42 - Saídas da rede com 2 nós intermediários e 1 saída, predizendo Peso Litro, após redução de pesos.....	79
Figura 43 - - Comparação dos resultados da rede inicial, com dois nós na camada intermediária, e a rede depois da redução de pesos, predizendo Cal Livre.....	80
Figura 44 - Comparação dos resultados da rede inicial, com dois nós na camada intermediária, e a rede depois da redução de pesos, predizendo Peso Litro.....	81

GLOSSÁRIO

Backpropagation. Metodologia utilizada para treinamento de redes neurais, corrigindo os pesos da mesma, por um procedimento de comparação do erro médio quadrático, entre o resultado obtido pela rede e o valor desejado, e que é propagado, corrigindo os pesos a partir da camada de saída até a camada de entrada.

Cal Livre. Qualidade do clínquer que expressa o teor de óxido de cálcio presente no mesmo.

Clínquer. Consiste de um produto granulado, obtido por tratamento térmico de uma mistura adequada de calcário e argila (farinha) até fusão parcial e posterior reação química entre os óxidos de sílica, cálcio, ferro e alumínio em forno rotativo. O clínquer, após moagem e adição de aditivos torna-se o cimento.

C2S, C3A, C4AF são os outros principais óxidos presentes no clínquer, cujas composições são, respectivamente, 2CaO.SiO_2 , $3\text{CaO.Al}_2\text{O}_3$ e $4\text{CaO.Al}_2\text{O}_3.\text{Fe}_2\text{O}_3$.

C3S. É o silicato de trióxido de cálcio (3CaO.SiO_2). É o principal dos silicatos de cálcio presentes no clínquer, por ser o maior responsável na contribuição do desenvolvimento das propriedades de resistência do clínquer.

Farinha. Mistura homogeneizada de minerais, argila, calcário e silicatos, que ao sofrer transformações físico-químicas em forno rotativo torna-se o clínquer.

FSC. É o Fator de Saturação que relaciona a concentração de CaO com os demais óxidos, de Silício, de Alumínio e de Ferro.

MA. É o módulo de Alumínio. Relaciona a concentração de alumínio e de ferro na farinha, ou seja, a relação entre seus principais fundentes.

MLP. Tipo de rede neural “Multi Layer Perceptron” onde as camadas de nós da rede são interligadas seqüencialmente pelos seus pesos, a partir da camada de entrada até a(s) camada(s) intermediária(s), e dela(s) até a camada de saída, sem que qualquer nó de uma camada posterior seja interligado a qualquer nó de uma camada anterior.

MS. É o Módulo de Sílica, que relaciona a concentração de sílica com a soma dos óxidos de Alumínio e de Ferro, ou seja, a proporção de materiais não fundentes e fundentes do clínquer.

OBS – OPTIMAL BRAIN SURGEON. É uma metodologia para redução de pesos, visando a otimização de redes neurais treinadas, tanto para a melhora do tempo de processamento, como para uma melhor predição e uma boa generalização.

Peso Litro. Qualidade do clínquer que expressa a sua densidade aparente.

Capítulo 1

INTRODUÇÃO

O cimento é normalmente definido como um aglomerante hidráulico constituído basicamente de óxidos de cálcio, silício, alumínio e ferro. Quando misturado com água, em adequadas proporções, endurece, tanto exposto ao ar como submerso em água.

O cimento, como produto final para aplicação, se apresenta sob forma de diversas misturas, gerando com isso diferentes tipos de produtos. Estes produtos têm, cada um, características específicas de aplicação. O componente principal para a obtenção desses diferentes tipos de cimento é o clínquer Portland, de cuja composição química e cristalina derivam as propriedades aglomerantes e hidráulicas dos diferentes tipos de cimento.

Este trabalho focalizou a modelagem do forno de cimento, utilizando técnicas de redes neurais artificiais, para prever a Cal Livre e o Peso Litro, que são dados de qualidade final do clínquer gerado. Além da modelagem do forno, foram aplicadas técnicas de otimização de redes neurais, para minimização de sua estrutura de pesos e uma análise de sensibilidade, para identificar as principais entradas do modelo, ou seja, aquelas que têm a maior influência na qualidade inferida dos produtos.

Capítulo 2

DESCRIÇÃO DO PROCESSO

O processo industrial para a produção de cimento divide-se basicamente em seis grandes etapas: a mineração, a pré-homogeneização, a moagem de cru, a clinquerização, a moagem de cimento, e, o ensacamento e expedição do produto.

Segundo Votorantim (2000), os minérios de argila e calcário, após extração nas minas, são britados, homogeneizados e moídos, compondo, em adequadas proporções, a farinha, que alimentará o forno rotativo. Neste a farinha é transformada em clínquer. O clínquer, após moagem e recebendo, se necessário, aditivos complementares, torna-se o cimento, seguindo para as etapas de ensacamento e expedição.

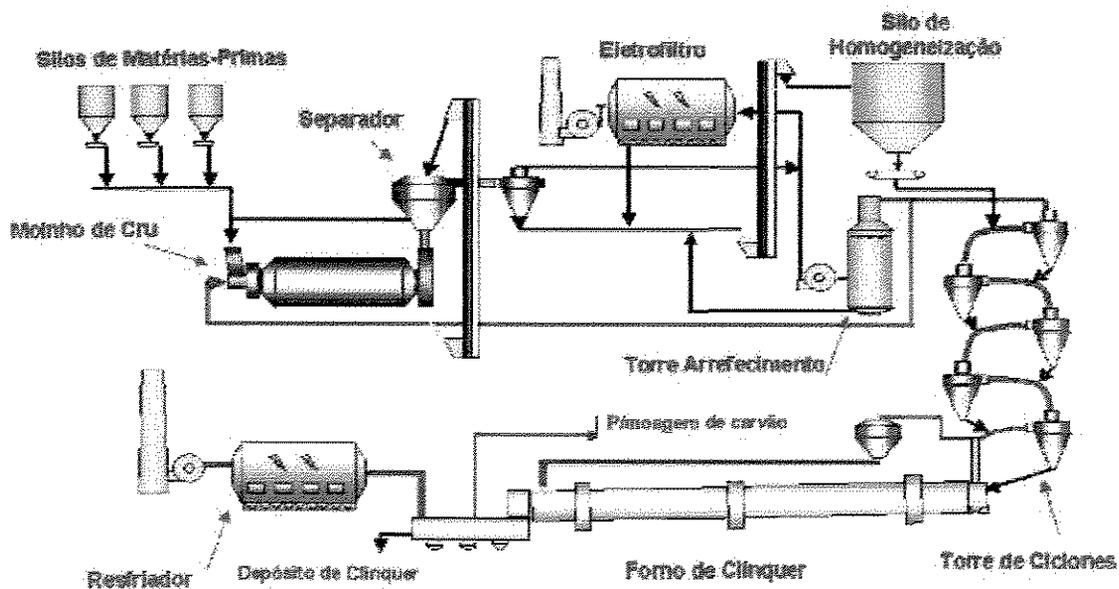


Figura 1 - Processo de produção de cimento mostrando os processos de moagem e de clinquerização

Este trabalho focaliza a etapa de clinquerização, modelando o forno rotativo 07 da Fábrica de Cimento de Rio Branco – PR, pertencente ao Grupo Votorantim Cimentos, cujo esquema está apresentado na Figura 1.

A clinquerização consiste na etapa que ocorre nos fornos rotativos e onde se dá a transformação química dos minerais naturais (farinha) em minerais sintéticos (clínquer).

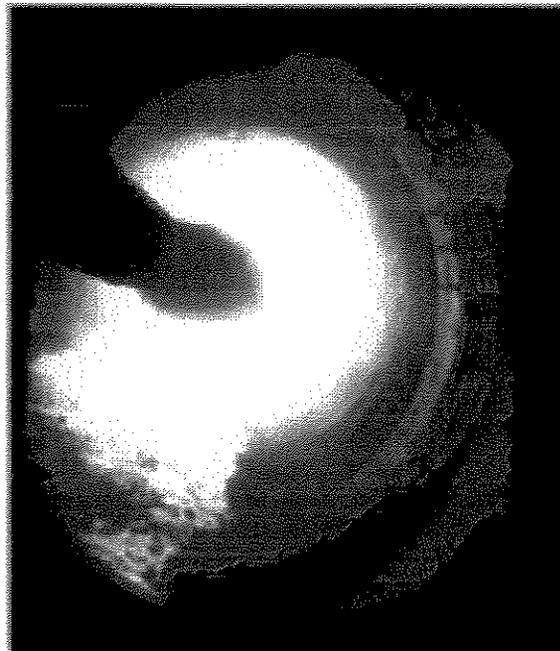


Figura 2 – Clínquer fluidizado na câmara de combustão do forno rotativo

O clínquer é um produto granulado, resultante do tratamento térmico de uma mistura de calcário e argila, até fusão parcial com posterior reação química entre os óxidos de sílica, cálcio, ferro e alumínio. Na sua maior parte o clínquer é constituído por silicatos (75%) e numa proporção menor, de aluminatos e ferro-aluminatos cálcicos. Os principais componentes do clínquer são o **C3S** ($3\text{CaO}\cdot\text{SiO}_2$), **C2S** ($2\text{CaO}\cdot\text{SiO}_2$), **C3A** ($3\text{CaO}\cdot\text{Al}_2\text{O}_3$) e **C4AF** ($4\text{CaO}\cdot\text{Al}_2\text{O}_3\cdot\text{Fe}_2\text{O}_3$), sendo o **C3S** o principal destes, já que é o grande responsável pelo desenvolvimento das propriedades de resistência do clínquer. Dessa maneira, podemos caracterizar o processo de clinquerização como aquele que fabrica C3S, pois se deve garantir que esta substância esteja presente em quantidade e qualidade adequadas.

Na produção do clínquer as reações ocorrem por difusão iônica entre sólidos. Esta difusão é acelerada com o aparecimento de uma fase líquida, localizada na interface dos

grãos, e que começa a se formar à temperatura próxima de 1300°C, como pode ser visto na Figura 2. A velocidade de difusão e a capacidade de clinquerização estão relacionadas à quantidade e à viscosidade desta fase líquida.

O clínquer sai do forno a uma temperatura aproximada de 1200°C e sofre um resfriamento para reduzir a sua temperatura para cerca de 100°C.

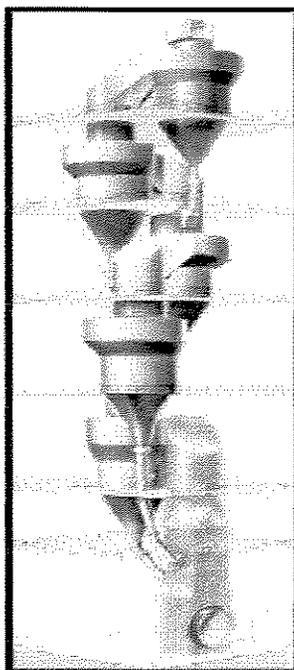


Figura 3 - Esquema de Torre de Ciclones

Os dados utilizados para a modelagem foram gerados num forno do tipo SP – “Suspension Preheater”, que não tem pré-calcinação, possui duas torres de ciclones de quatro estágios cada, e capacidade máxima de processar 2.400 t/d de farinha, obtendo uma descarbonatação de até 50%. Um esquema da Torre de Ciclones é apresentado na Figura 3, e a fotografia do conjunto Torre de Ciclones – Forno Rotativo está na Figura 4.

Como já foi dito, é de grande importância que as matérias primas, utilizadas na fabricação do cimento, estejam presentes numa proporção bem definida na farinha, garantindo assim a obtenção das propriedades desejadas do clínquer, entre elas a Cal Livre e o Peso Litro.

Com a finalidade de garantir a proporção dos diferentes componentes químicos, os operadores utilizam algumas relações que orientam a produção da farinha e que definem tanto suas características como seu comportamento durante a fase de clínquerização.

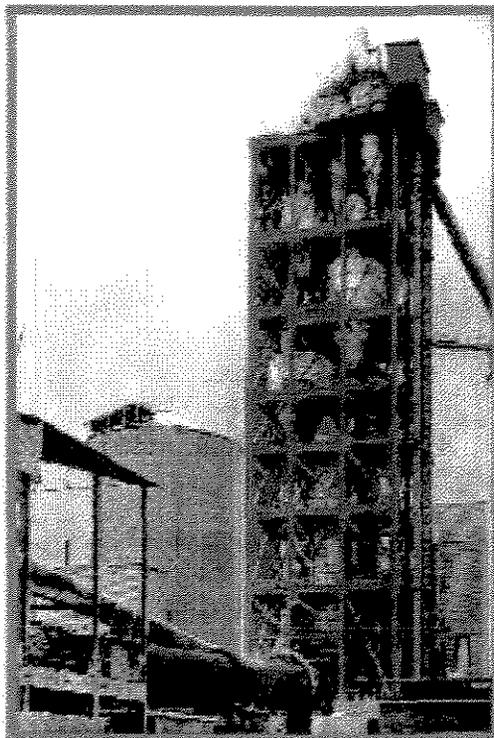


Figura 4 - Torre de Ciclones e Forno Rotativo

As principais relações utilizadas são:

- Fator de Saturação (FSC);
- Módulo de Sílica (MS);
- Módulo de Alumínio (MA).

Fator de saturação (FSC):

Relaciona a concentração de CaO presente na farinha/clínquer com os demais componentes com os quais ele reage. Normalmente seu valor está entre 92 e 96%.

$$FSC = \frac{CaO}{2,8SiO_2 + 1,18Al_2O_3 + 0,65Fe_2O_3} * 100 \quad [1]$$

Um alto valor de **FSC**, acima de 96%, calculado pela Equação 1, provoca queima difícil da farinha, devido ao grande conteúdo de CaO; tende a causar expansibilidade do cimento, e com isto há tendência a maior quantidade de CaO livre no clínquer, aumenta o conteúdo de C3S, ocorrendo maior disponibilidade de CaO, e reduz o conteúdo de C2S por efeito contrário ao anterior.

Módulo de Sílica (MS)

Relaciona a concentração de sílica (SiO_2) com a concentração de Al_2O_3 e Fe_2O_3 , ou seja, a proporção de materiais não fundentes (silicatos) e fundentes. Os valores normalmente encontrados nas fábricas estão em torno de 2,3 a 2,7.

$$MS = \frac{\text{SiO}_2}{\text{Al}_2\text{O}_3 + \text{Fe}_2\text{O}_3} \quad [2]$$

Um alto valor de **MS** (acima de 2,7), calculado pela Equação 2, tem como consequência uma elevada carga térmica do forno, devido à maior presença de materiais não fundentes na farinha. Com isso, torna-se necessário queimar uma quantidade maior de combustível no forno para obter a quantidade de fase líquida desejada. Esse alto valor gera a redução na quantidade de fase líquida porque, para uma mesma quantidade de combustível queimado, a porcentagem de fase líquida no forno será menor, já que a farinha possui menos fundentes. O alto **MS** produz um clínquer pulverulento, devido à menor quantidade de fase líquida. O valor elevado leva a um alto consumo de combustível e uma queima difícil. O alto **MS** tem a tendência a causar expansibilidade do clínquer, sendo a queima mais difícil, e assim, há maior probabilidade da ocorrência de CaO livre; isso também gera dificuldade na formação de colagens pela menor quantidade de fundentes; e leva à deterioração do revestimento refratário pela dificuldade de formação de colagens.

Um baixo valor de **MS** (abaixo de 2,3) ocasiona um excesso de fase líquida e menor consumo térmico pela maior presença de fundentes; a deterioração do revestimento refratário, devido à possibilidade de formação de uma colagem de baixo ponto de fusão, que facilmente se desfaz e o favorecimento à formação de bolas de clínquer de alta dureza. Também o cimento resultante desse clínquer terá baixa resistência.

Módulo de Alumínio (MA)

Relaciona a concentração de alumínio e de ferro na farinha, ou seja, a relação entre os dois principais fundentes. Seu valor tem influência direta na viscosidade da fase líquida. Como referência para o valor de MA podemos citar o intervalo de 1,3 a 1,6.

$$MA = \frac{Al_2O_3}{Fe_2O_3} \quad [3]$$

Um alto valor de **MA** (acima de 1,6), conforme Equação 3, resulta em queima mais difícil e maior consumo de combustível devido a uma fase líquida mais viscosa; num aumento da proporção de C3A e redução do C4AF pela maior disponibilidade de Al_2O_3 ; no aumento da viscosidade da fase líquida a uma temperatura constante devido à maior proporção de óxido de alumínio; num alto calor de hidratação devido à presença de C3A; e numa tendência à pega rápida do cimento e às altas resistências a pequenas idades, devido à presença de C3A.

Um baixo valor de **MA** (abaixo de 1,3), ao contrário, resulta em uma fase líquida mais fluida e num baixo calor de hidratação.

Uma outra característica importante no processo de clínquerização é a da **QUEIMABILIDADE DA FARINHA**, que consiste na predisposição de uma dada farinha a se transformar em clínquer. A queimabilidade é influenciada principalmente pelos seguintes fatores:

- Tempo de reação;
- Temperatura;
- Finura da farinha;
- Teor de CaO;
- Velocidade do forno;
- Presença de fundentes.

A maioria dos clínqueres industriais contém cal livre em quantidades até 2% do peso, cuja presença pode ser atribuída aos seguintes fatores:

- Farinha não homogênea ou muito grossa;

- Deficiência de queima (CaO não combinado);
- Resfriamento muito lento (decomposição parcial de C2S e C3A);
- Fator de saturação de cal muito alto (excesso de CaO na mistura crua).

Um teor de cal livre acima de 2,5% é indesejável, uma vez que pode causar expansibilidade na argamassa ou no concreto.

A reação química entre o cimento e a água produz inicialmente um enrijecimento da pasta (estágio plástico) e a seguir um endurecimento (estágio sólido). Nos cimentos considerados normais, estas reações químicas provocam ligeiras mudanças no volume da pasta de cimento (o volume final pode ser até 6% menor que o inicial, dependendo das condições sob as quais se desenvolve a hidratação). Nos cimentos que podem ser denominados de anormais, além da redução de volume descrita pode existir simultaneamente um aumento considerável do volume da pasta ao longo do tempo, devido às reações químicas de alguns compostos com água de amassamento. As substâncias que podem provocar estas instabilidades de volume indesejáveis são principalmente: a cal livre (óxido de cálcio não combinado), o sulfato de cálcio em dosagens muito acima do necessário para regularizar o tempo de pega e o óxido de magnésio em forma cristalizada (periclásio).

Dessa maneira, podemos entender a importância da análise do teor de CaO no clínquer, como um amplo indicador de qualidade, tanto do processo de clínquerização da farinha no forno rotativo, como do produto final, o cimento. Esta é a principal razão para que o modelo neural do forno calcule essa qualidade.

O Peso Litro constitui-se no ensaio de determinação da densidade aparente do clínquer, sendo um indicativo muito importante para a operação do forno.

De forma similar à Cal Livre, indica a qualidade da clínquerização. Um teor de Cal Livre alta no clínquer, e um valor de Peso Litro baixo, são indicações de que o clínquer não foi devidamente queimado. Por outro lado, um teor baixo de Cal Livre associado a um valor alto de Peso Litro, indica que o clínquer foi queimado em demasia.

Capítulo 3

MATERIAIS E MÉTODOS

Local e programas utilizados

O trabalho foi realizado no Laboratório de Modelagem e Simulação de Processos Químicos – LMSPQ, da Faculdade de Engenharia Química – FEQ, da Universidade Estadual de Campinas – UNICAMP, em Campinas.

A coleta de dados foi efetuada em unidade industrial (Forno 7) da VOTORANTIM Cimentos S.A., situada na cidade de Rio Branco – PR, dotada de computador de processo, supervisorio, com sistema de controle baseado em Regra “Fuzzy”.

Para os testes preliminares e tratamento de dados foram utilizados os programas: NeuralWorks da NeuralWare Inc., Excel para Windows, da Microsoft Corporation, NeuroSolutions da NeuroDimension, Inc., e, para o treinamento final da rede neural e a geração de dados e gráficos, foi utilizado o Toolbox NNET, para redes neurais, do programa Matlab da The MathWorks, Inc. Foi utilizado, também, o Toolbox NNTools, NEURAL NETWORK BASED SYSTEM IDENTIFICATION TOOLBOX, beta version 1.0, elaborado pelo Institute of Automation da Technical University of Denmark, disponibilizado na internet pelo Prof. Dr. Magnus Norgaard em 1995.

Modelagem utilizando Redes Neurais

As Redes Neurais, também conhecidas como RNA – Redes Neurais Artificiais (ou ANN – Artificial Neural Network), foram inspiradas a partir do trabalho que um neuro-biologista, Warren McCulloch e um estatístico, Walter Pitts, publicaram, “A Logical Calculus of Ideas Imminent in Nervous Activity” em 1943.

Em 1957 Frank Rosenblatt, publicou em Cornell o primeiro grande projeto em computação neural, o desenvolvimento de um elemento chamado “perceptron”, que desencadeou uma série de pesquisas em computação neural.

Este tipo de rede despertou grande interesse tanto no início como em meados dos anos 60. No final dos anos 60, Marvin Minsky e Seymour Papert, ambos do Laboratório de Pesquisas Eletrônicas do MIT, publicaram um livro, “Perceptrons”, criticando profundamente o perceptron, e praticamente liquidaram este tópico de pesquisa ao mostrar as limitações do perceptron em termos de poder de expressão e habilidade de aprendizagem.

Em meados da década de 80 as Redes Neurais tornaram a reviver por causa de melhores algoritmos de treinamento, com o uso de sigmóides. Outra razão para o renascimento das Redes Neurais reside no fato de que *nossas expectativas também mudaram* em relação às suas possibilidades e suas aplicações. Esse desenvolvimento explosivo ocorreu após a publicação do livro “Parallel Distributed Processing”, editado em 1986 por Rumelhart e McClelland.

Pela sua habilidade em identificar sistemas complexos, fortemente não lineares, as Redes Neurais têm sido utilizadas, com muito êxito, em aplicações na área da Engenharia Química. Há vários trabalhos publicados hoje em dia, como, por exemplo, na revista *Computers & Chemical Engineering*¹ com uma edição exclusiva, apresentando treze artigos sobre redes neurais, além de revistas contendo artigos especializados em Redes Neurais, como “Neural Networks”, “AI Expert”, entre outras.

Em todas as referências consultadas há vários detalhes a respeito do desenvolvimento das redes neurais, incluindo outras referências sobre a sua fundamentação teórica. Em especial NeuralWare (1995) que em vários capítulos descreve o desenvolvimento dos vários tipos de redes, e Kovács (1996). Além disso, todas apresentam extensa bibliografia sobre o assunto.

¹ *Computers & Chemical Engineering*. Oxford, v. 16, n. 4, 423p. April 1992. (Neural Network Applications in Chemical Engineering).

O modelo das Redes Neurais foi desenvolvido a partir do cérebro humano. No cérebro, bilhões de células nervosas, chamadas de *neurônios*, interligam-se através de sinapses e dendritos. A representação esquemática de um neurônio biológico está na Figura 5. Quando um estímulo, ou uma combinação de estímulos atinge um neurônio, através dos dendritos de outros neurônios, conectados neste por sinapses, dependendo da qualidade, ou intensidade desse estímulo o neurônio receptor pode tornar-se excitado, e, por seu axônio, repassar esse estímulo a seus dendritos, que irão estimular os milhares de neurônios a eles conectados.

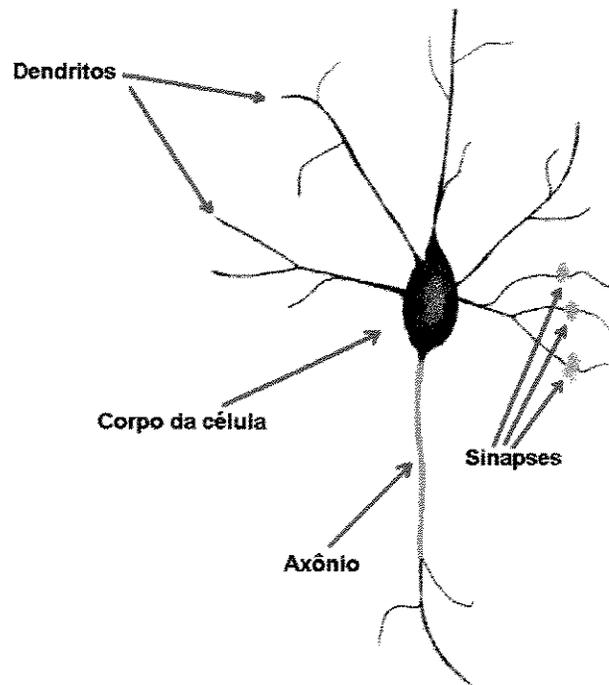


Figura 5- Esquema de um neurônio

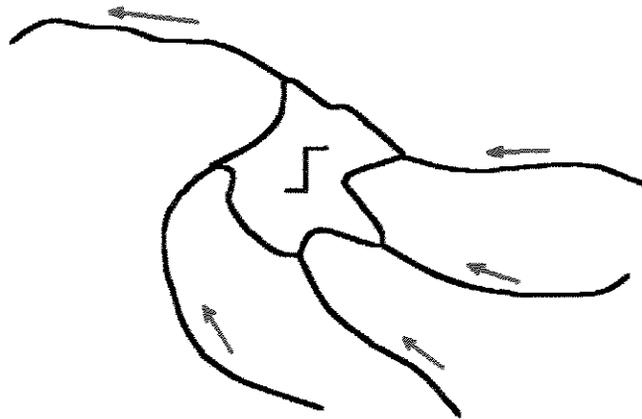


Figura 6 - Se a soma dos sinais de entrada excederem um valor de excitação o neurônio "dispara"

Um neurônio pode ser considerado como uma unidade limiar – um elemento de processamento que recolhe as entradas e produz uma saída somente se o somatório das entradas exceder a um valor de limite interno, como ilustrado na Figura 6.

A estrutura de interligação: corpo do neurônio – axônio – dendritos / sinapses – outros neurônios, tem o formato de uma rede interligada, como pode visto na Figura 7.



Figura 7 - Os neurônios estão interconectados. Quando um neurônio "dispara", outros neurônios recebem seu sinal e podem disparar também.

O modelo de neurônio mostrado na Figura 8 corresponde a um modelo matemático de um neurônio humano e possui como sinal de entrada, conforme Equação 4, o vetor:

$$X = (x_1, x_2, \dots, x_i, \dots, x_n), \quad [4]$$

onde x_i representa a ativação do i -ésimo neurônio.

Associado a cada neurônio existe um valor ajustável, w_{ij} (peso). Os pesos ligados ao j -ésimo neurônio (j), formam o vetor: $w_j = (w_{1j}, w_{2j}, \dots, w_{ij}, \dots, w_{nj})$, onde o peso w_{ij} representa o valor da ligação do neurônio i para o neurônio j . Ocasionalmente existe um parâmetro adicional (bias) θ_j modulado pelo peso w_{0j} compondo as entradas dos neurônios, que serve para dar estabilidade aos cálculos, especialmente durante a fase de treinamento da rede. Este modelo é também referenciado como Elemento de Processamento – PE.

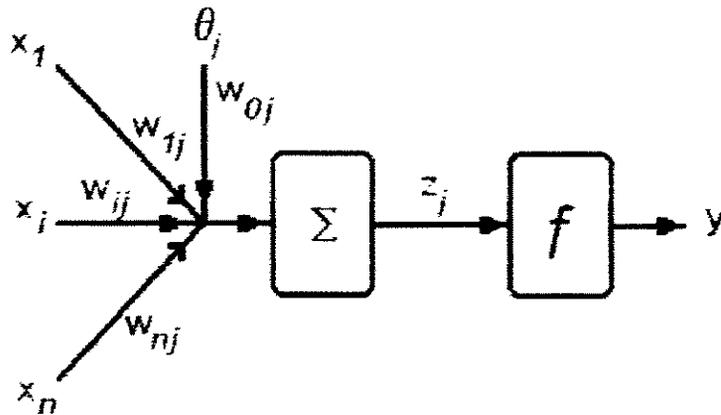


Figura 8 - Representação matemática de um neurônio

A somatória das entradas de um neurônio (que também é referenciado como nó da rede, ou elemento de processamento) balanceadas por seus respectivos pesos e mais um bias, também ponderado, resulta no valor de entrada do j -ésimo neurônio, z_j . Este valor sofre uma função de ativação, gerando a saída efetiva do neurônio, y_i , como representado na Equação 5.

$$\begin{aligned} z_j &= \left(\sum x_i \cdot w_{ij} \right) + \theta_j \cdot w_{0j} \\ y_j &= f(z_j) \end{aligned} \quad [5]$$

Existem várias funções de ativação propostas e em uso [Pham,1995], como na Tabela 1, que apresenta as funções comumente mais utilizadas.

Tabela 1 - Funções de Ativação

Tipo de Funções	Funções
Linear	$f(s) = s$
De Ativação	$f(s) = +1$ se $s > s_i$ $f(s) = -1$ nos demais casos
Sigmóide	$f(s) = 1/(1 + \exp(-s))$
Tangente Hiperbólica	$f(s) = (1 - \exp(-2s))/(1 + \exp(2s))$
Função base radial	$f(s) = \exp(-s^2/\beta^2)$

O Toolbox da MATLAB tem funções para cálculo da tangente hiperbólica, *tansig* e *pmntanh* cujas respectivas equações estão mostradas nas seguintes equações:

$$f(x) = \frac{2}{(1 + e^{(-2*x)})} - 1 \quad (\text{tansig})$$

$$f(x) = 1 - \frac{2}{(e^{2*x} + 1)} \quad (\text{pmntanh})$$
[6]

Pela sua capacidade de “aprender” com exemplos apresentados à Rede Neural, estas têm grande habilidade na **Classificação de Padrões** e em **Síntese de Funções**.

A Classificação de Padrões pode dar-se em dois níveis: “baixo” e “alto”. Na atividade de classificação de “baixo nível” podemos citar a habilidade em detectar características de objetos, diferenciar sons, entre outras, enquanto que na de “alto nível”, podemos incluir a habilidade em julgar qualidades de investimentos, riscos envolvidos em determinadas atividades, como definição do prêmio de seguros, etc.

Podemos definir a Síntese de Funções como a habilidade em estabelecer relações entre valores de entrada e seus correspondentes valores de saída, e com isso, a eventual necessidade do aprendizado de complexas relações não lineares. Nesta habilidade incluímos a elaboração de modelos de processos.

Uma Rede Neural é, então, um conjunto de nós, denominados de neurônios, ou elementos processadores, interligados e interagindo entre si. Como existem diversas maneiras de interligar os nós das redes, inúmeras estruturas têm sido estudadas e publicadas, porém, somente algumas realmente têm tido aplicações práticas em processos químicos industriais, e é destas que falaremos com mais detalhes.

Em processos químicos industriais, que apresentam forte não linearidade, dificultando a obtenção de modelos fenomenológicos, a ponto de, em alguns casos, conseguir-se apenas um modelo extremamente simplificado, válido em condições muito específicas, as Redes Neurais colocam-se como uma técnica muito interessante de ser empregada, pela habilidade em contornar essa dificuldade em obter um modelo fenomenológico representativo do processo. O sucesso na aplicação dependerá, entretanto, do adequado conhecimento do processo e de suas variáveis significativas, além da disponibilidade de uma boa base de dados, em quantidade e qualidade, representando corretamente o domínio onde a solução será modelada, uma vez que as redes não possuem habilidade para extrapolar adequadamente para fora do domínio dos dados a ela apresentados para treinamento.

Como o problema a ser resolvido enquadra-se nas condições aqui explanadas, a Rede Neural é forte candidata para solucioná-lo.

Tipos de Redes Neurais

A Rede Neural é um sistema de computação que consiste de um grande número de elementos processadores (nós), que processam a informação determinando o valor do sinal de saída, baseando-se nos valores dos sinais de entrada, e pode ser classificada quanto à *estrutura* da ligação entre seus nós, e também ser diferenciada pelo *tipo de algoritmo de treinamento* utilizado.

Tipos de Redes Neurais, quanto à sua estrutura

Cada elemento de processamento (nó/neurônio) executa um processo de cálculo em suas entradas para gerar sua saída. As entradas de um neurônio geralmente são tratadas

uniformemente, e muitas vezes, as saídas de um elemento de processamento são idênticas. Cada conexão pode, também, executar um processo de cálculo simples, mas opera somente um sinal.

Fica claro que o resultado calculado por uma Rede Neural irá depender da topologia (estrutura) dessa rede, da natureza dos cálculos efetuados em cada neurônio, da natureza dos cálculos realizados em cada conexão e dos valores de cada parâmetro ajustável nos neurônios e nas conexões, como pode ser observado na Figura 9. Assim, as Redes Neurais podem ser totalmente projetadas, desde seu início, para executar um cálculo muito específico.

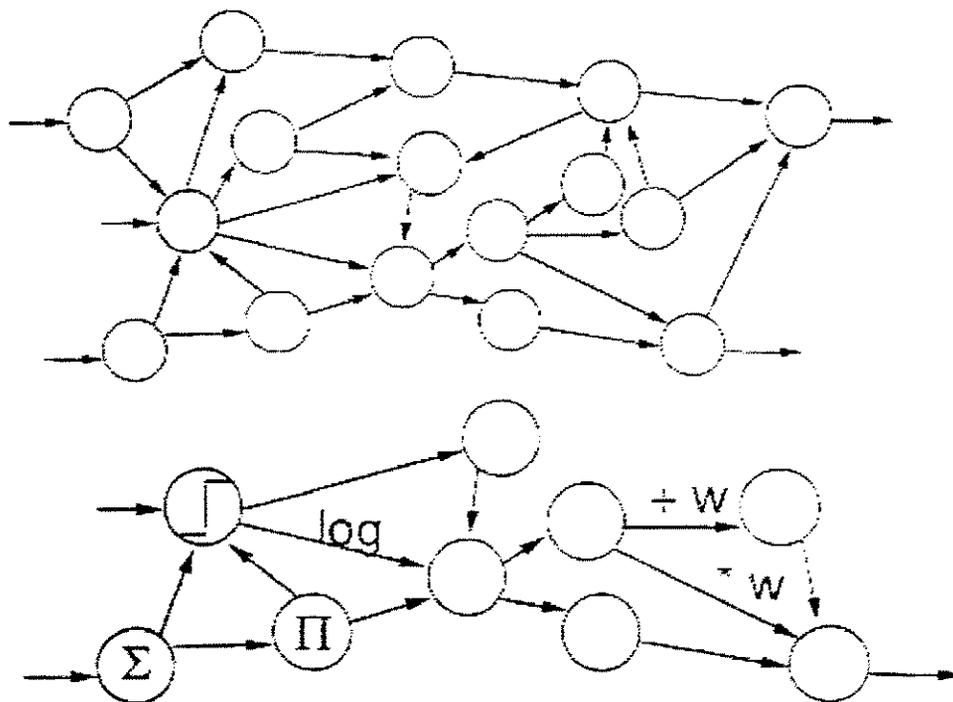


Figura 9 - Esquemas de ligações entre nós de uma Rede Neural

Considerando uma estrutura em camadas, caso todos os neurônios de uma camada, de entrada, por exemplo, que fazem a interface com o ambiente externo, recebendo os dados, estiverem ligados unicamente com os neurônios da camada seguinte, ou seja, a camada intermediária, que alguns autores preferem denominar de camada escondida

(*hidden*), e todos os neurônios desta camada intermediária somente estiverem ligados aos neurônios da camada de saída, e estes só ao ambiente externo, essa estrutura é conhecida como rede *feedforward*, ou seja, as ligações só permitem a passagem de estímulos aos neurônios da camada seguinte.

No caso de existirem ligações que permitam que os estímulos possam ser distribuídos para os neurônios da mesma camada, ou retornando para os neurônios da camada anterior, este tipo de estrutura é conhecido como rede *recorrente*.

Redes Recorrentes

As redes recorrentes possuem conexões que permitem que o fluxo de sinais possa ir tanto em direção aos neurônios de saída, quanto em direção aos neurônios de entrada, ou lateralmente, em direção aos neurônios da mesma camada.

Esse tipo de estrutura faz com que a rede tenha uma memória dinâmica: suas saídas, em determinado instante, refletem a entrada atual da rede, assim como entradas e saídas anteriores. Por isso, este tipo de rede é aplicável onde são necessárias suas propriedades de “memória”, de curto ou longo prazo. Dentre as diversas possibilidades de estruturas de rede recorrente, as mais conhecidas e utilizadas são: rede Hopfield, rede Elman, rede Jordan e rede Kohonen.

Redes Feedforward

Nestas redes os neurônios geralmente são agrupados em camadas, e os sinais de entrada evoluem dentro da rede, camada por camada, até a camada de saída, de forma unidirecional. As redes feedforward podem, muito naturalmente, executar o mapeamento entre o espaço de entrada e o espaço de saída. Desta maneira, num dado instante, a saída é somente função da entrada naquele instante. Entre os tipos de rede feedforward estão as MLP – “Multilayer Perceptron”, as redes LVQ – “Learning Vector Quantization” e as redes GMDH – “Group Method of Data Handling”. Entre essas redes, a que melhor se aplica aos processos químicos industriais é a MLP, que será utilizada na modelagem objeto desta dissertação.

Rede MLP

A MLP é, talvez, o tipo de rede *feedforward* melhor conhecido. A principal razão para isso é a sua habilidade tanto em modelar relações funcionais simples como relações muito complexas. Os neurônios da camada de entrada agem somente como estocagem para distribuição dos sinais de entrada x_i aos neurônios da camada intermediária. Na estrutura MLP cada neurônio j da camada intermediária efetua o somatório de seus sinais de entrada, após serem ponderados com o peso de suas respectivas conexões w_{ij} com a camada de entrada, e calcula a saída y_j , como uma função f desse somatório, como pode ser visto na Figura 10.

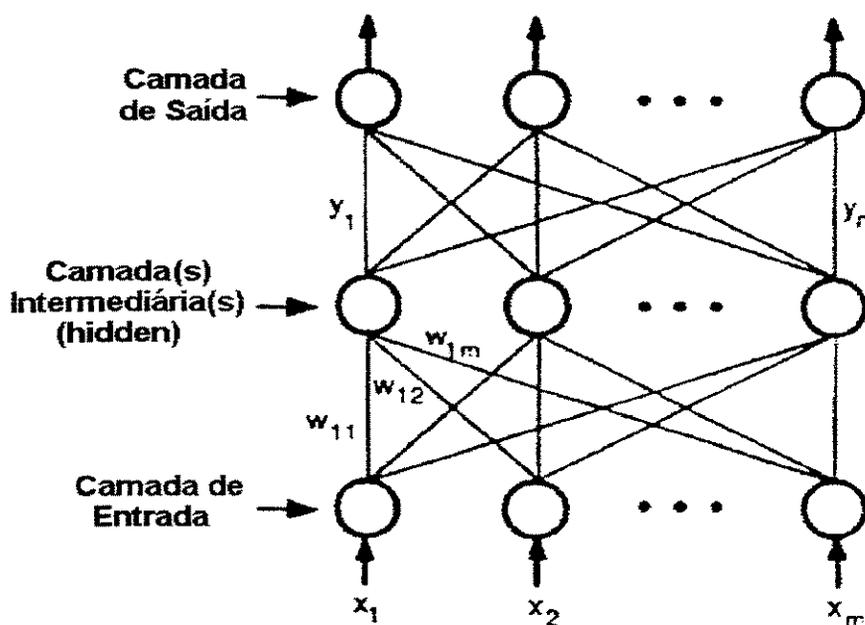


Figura 10 - Estrutura de uma Rede Neural tipo MLP

Assim como apresentado na Figura 10, a saída dos neurônios da camada de saída pode ser calculada pela Equação 7, a seguir:

$$y_i = f\left(\left(\sum w_{ji} x_j\right) + w_{0i}\right) \quad [7]$$

onde:

x são as conexões de entrada da rede;

y são as saídas calculadas de cada nó (elemento de processamento) de uma camada, considerando todas entradas do nó, com seus respectivos pesos, o *bias* e respectivo peso, e a função de ativação definida para essa camada;

w_{ji} são os pesos conectando o *i-ésimo* nó de uma camada ao *j-ésimo* nó da camada seguinte;

w_{0j} são os pesos conectando o *bias* ao *j-ésimo* nó da camada seguinte.

Tipos de Redes quanto ao algoritmo de treinamento

As Redes Neurais conseguem estabelecer suas relações entre entrada e saída a partir de dados de treinamento. Com isso, fica bem clara a necessidade de bons dados, em número e quantidade, para que essas relações identificadas sejam as mais representativas possíveis do processo que está sendo analisado. É através do treinamento que os valores dos pesos das conexões entre os neurônios de uma rede são calculados, e esses pesos é que caracterizam e definem essas relações entre entrada e saída.

Segundo Pham (1995) há dois algoritmos básicos para o treinamento de Redes Neurais: algoritmos de aprendizagem supervisionada e algoritmos de aprendizagem não supervisionada.

Há um tipo de algoritmo, chamado de aprendizagem reforçada, mas que pode ser considerado uma forma especial de aprendizagem supervisionada.

Algoritmos de Aprendizagem Supervisionada

Um algoritmo de aprendizagem supervisionada ajusta os pesos das interconexões entre os neurônios, de acordo com a diferença entre o valor calculado pela rede, para um

dado valor de entrada, e o respectivo valor esperado de saída. Dessa maneira, o algoritmo de aprendizagem supervisionada requer um “professor”, ou “supervisor”, para fornecer o valor desejado do sinal de saída (“objetivo”). Entre os algoritmos de aprendizagem supervisionada estão: o método do gradiente (“delta rule”), o método generalizado do gradiente, também conhecido como algoritmo *backpropagation*, e o algoritmo LVQ.

O método de aprendizagem reforçada, como dito anteriormente, é um caso especial de aprendizagem supervisionada. Ao invés de utilizar um “professor” para fornecer as saídas desejadas, o algoritmo de aprendizagem reforçada utiliza uma análise crítica para avaliar quão bem a rede consegue gerar uma saída correspondente à entrada. Um exemplo de aprendizagem reforçada é o do Algoritmo Genético (GA).

Algoritmo Backpropagation (BP)

Este algoritmo, também conhecido como algoritmo do *gradiente descendente* é o algoritmo mais comumente utilizado para o treinamento de redes MLP.

A correção do peso, Δw_{ji} na conexão entre os neurônios i e j é dada pela Equação 8 a seguir:

$$\Delta w_{ji} = \eta \cdot \delta_j \cdot x_i \quad [8]$$

onde:

η é um parâmetro chamado de *taxa de aprendizagem*;

δ_j é um fator cujo valor vai depender da camada que o neurônio pertence: camada intermediária ou camada de saída.

O fator de correção para os neurônios de saída, δ_j , é calculado pela Equação 9:

$$\delta_j = \left(\frac{\partial f}{\partial \text{net}_j} \right) (y_j^{(r)} - y_j) \quad [9]$$

onde:

f é a função de ativação;

net_j é o somatório total dos sinais de entrada para o neurônio j , ponderados por seus respectivos pesos;

y_j é a saída calculada do neurônio j ;

$y_j^{(t)}$ é a saída desejada para o neurônio j .

Como não há saídas desejadas para os neurônios da camada intermediária, a diferença entre a saída desejada e a saída calculada do neurônio j da camada intermediária é substituída pelo somatório ponderado dos termos de δ_q , já calculados para os neurônios q conectados à saída de j , e assim, o fator de correção para os neurônios da camada intermediária é calculado pela Equação 10:

$$\delta_j = \left(\frac{\partial f}{\partial net_j} \right) \sum_q w_{qj} \cdot \delta_q \quad [10]$$

Os algoritmos de treinamento geralmente utilizam o método de minimização do erro global, que é dado pela Equação 11:

$$E^q = \frac{1}{2} \sum_{j=1}^n [y_j^{(t)} - o^{(t)}]^2$$
$$E = \sum_{q=1}^n E^q \quad [11]$$

onde q = número de casos apresentados para treinamento

Bufo (2000), lista uma série de técnicas de minimização do erro global, aplicadas às redes neurais tipo *backpropagation*.

O modelo do forno em estudo é complexo e não linear. As redes *feedforward* têm sido aplicadas com muito sucesso para a solução desse tipo de problema: problemas desafiadores, com grandes dimensões dos espaços de entrada e saída, com forte inter correlação entre as variáveis de processo, com grande lapso de tempo e comportamento não linear (Kim et al.,1998).

Dias (2000) utilizou uma rede neural com duas camadas intermediárias porque uma das entradas não era contínua.

Lai & Serra (1997) comentam que o teorema bem conhecido de Hecht-Nielsen (1989)² conclui que qualquer vetor $y = f(x)$ pode ser calculado usando uma adequada rede neural, com três níveis, e totalmente conectada entre um nível e outro, demonstrando que a rede neural pode resolver problemas de absoluta generalidade, porém não define quantos neurônios são requeridos para a camada intermediária. Apontam, também que o sucesso das redes neurais pode ser atribuído a certas grandes vantagens, tais como sua habilidade de formar associações, de generalizar, sua alta estabilidade quando operando com dados incorretos, e sua rápida implementação. Por outro lado, suas desvantagens incluem aritmética e capacidade lógica limitadas, baixa precisão em certos casos, e inabilidade em explicar os resultados.

Di Massimo et al. (1992) informam que, embora a rede neural *feedforward* possua uma estrutura muito elementar, há trabalhos mostrando que qualquer função não linear contínua pode ser representada por rede neural *feedforward* com uma topologia consistindo de duas camadas intermediárias.

Segundo Oliveira (2000), na modelagem de processos químicos o uso de uma única camada interna tem-se mostrado suficiente. Quando há necessidade de modelos mais complexos, o ajuste do número de neurônios da camada intermediária geralmente é suficiente.

Metodologia para otimização dos pesos

Uma rede neural, para conseguir modelar um sistema e conseguir generalizar, ou seja, também representá-lo quando apresentada a outros dados, não utilizados para o aprendizado, deve possuir um número adequado de pesos e ser bem treinada. Nas figuras apresentadas a seguir, obtidas de exemplos do MATLAB, vemos o resultado de uma rede

² HECHT-NIELSEN, R. Theory of back propagation neural networks in *International Joint Conference on Neural Networks*, 1989.

bem treinada, de uma rede com treinamento insuficiente (sub treinada) e uma rede treinada em demasia (sobre treinada, ou com “*overtraining*”).

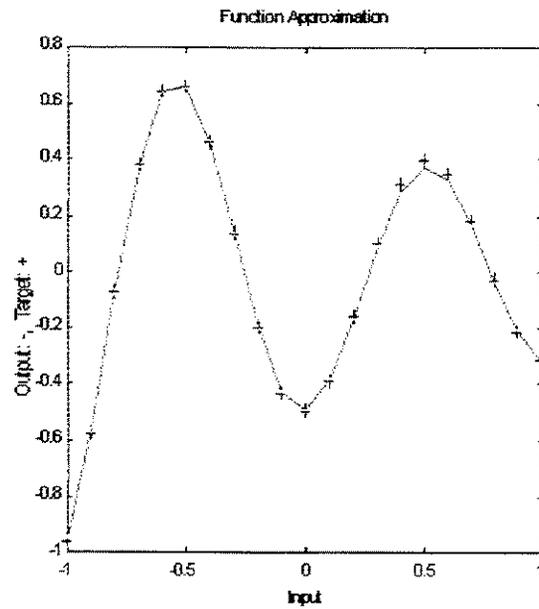


Figura 11 - Rede com treinamento adequado: a função permite boa interpolação entre os dados usados para treinamento (consegue “generalizar”)

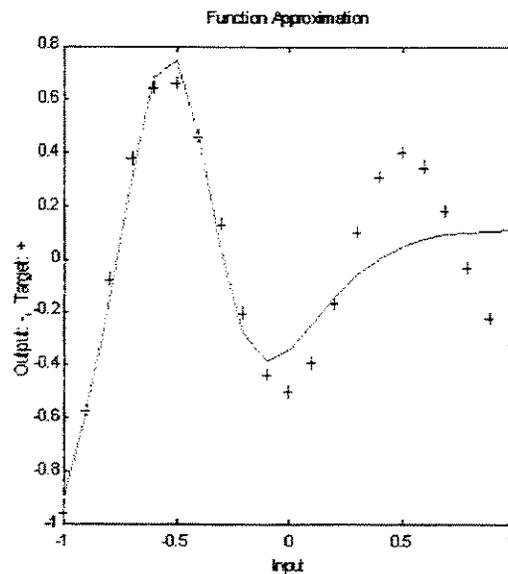


Figura 12 – A função obtida da rede com treinamento insuficiente não consegue modelar bem os dados

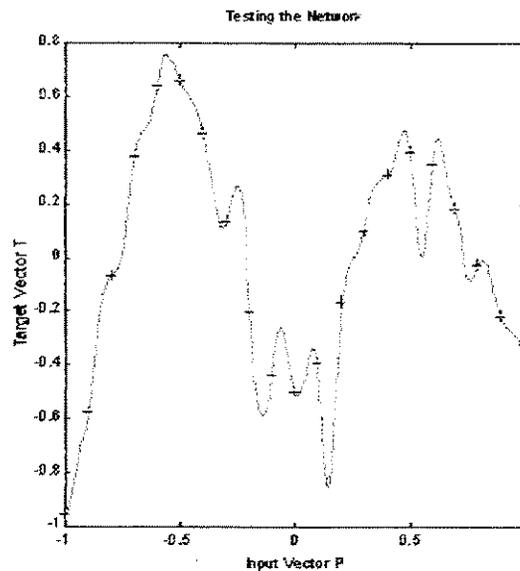


Figura 13 - Rede sobre treinada: identifica os pontos usados para treinamento, mas não permite a interpolação entre os dados usados para treinamento (não consegue “generalizar”)

Normalmente os dados de entrada, usados para treinamento da rede, são gerados com ruídos inerentes do próprio processo. No caso da rede ser sobre treinada, os pesos identificam e armazenam esses ruídos, e assim, com a introdução de novos dados, ela não consegue representa-los adequadamente.

Segundo Bhat & McAvoy (1992), o sobre-ajuste é uma condição que ocorre quando a razão entre os parâmetros do modelo e os dados do conjunto de testes é grande. Com um grande número de parâmetros, o erro da predição no conjunto de treinamento pode ser grandemente minimizado, porque cada parâmetro é ajustado para reproduzir parte dos dados de treinamento. Dessa maneira, o modelo reproduz os dados de treinamento com precisão, mas não necessariamente captura o comportamento global do processo. Estes modelos sobre-ajustados devem ser evitados a todo custo. Atualmente, as redes *backpropagation* são os paradigmas de redes mais utilizados para a resolução de problemas de engenharia química. Apesar de sua ampla aplicação, não há uma metodologia para a determinação da estrutura da rede para uma particular aplicação de mapeamento. A falta de um procedimento tem resultado numa tendência do uso de redes muito maiores que o necessário. Esses modelos possuem parâmetros excessivos, ou pesos, e muitas vezes memorizam os dados de treinamento, causando dificuldades no manuseio de dados não

apresentados à rede. Desta maneira, é importante o uso de redes que tenham estrutura mais simples, ou seja, utilize o mínimo de pesos e nós, e sugerem que as redes BP com menor complexidade podem ser obtidas de 2 maneiras:

1. utilizando redes com o menor número possível de nós;
2. utilizando redes com conexões esparsas.

Bhat & McAvoy (1992) citam que Hagiwara (1990)³ sugere que o número de nós da camada intermediária pode ser reduzido pela identificação de nós “ruins”. Um fator de “baixa representatividade” associado a cada nó depende do erro propagado através desse nó. Com isso, os nós com altos fatores de “baixa representatividade” são excluídos. Nesse mesmo artigo, são citados Mozer & Smolensky (1989)⁴ sugerindo que os nós da camada intermediária podem ser excluídos por um processo de “definição de relevância”. A relevância de cada nó “ r ” é calculada pela Equação 12:

$$\partial_i = E_{\text{sem nó } i} - E_{\text{com nó } i} \quad [12]$$

onde E_i é o erro médio quadrático da rede com dados de treinamento.

A rede é reduzida pela remoção de nós irrelevantes. Entretanto, esses métodos testam o fator de relevância que os nós individuais possuem na rede. Considerando o formato paralelo que os modelos de rede BP possuem, é possível que um nó seja irrelevante por si só, mas em conjunto com outros nós, o efeito pode ser construtivo ou destrutivo. Dessa maneira é mais importante visualizar o problema de uma perspectiva mais global e testar como os nós operam juntos. Em segundo lugar, é muito “comum” que apenas alguns pesos conectados a um nó possam ser irrelevantes e não todo o nó. Niida et al (1992)⁵,

³ HAGIWARA, M. Novel back-propagation algorithm for reduction of hidden units and acceleration of convergence using artificial selection. *Proc IJCNN*. San Diego, 1990, p. 1625-1630.

⁴ MOZER, M.C. & SMOLENSKY, P. Skeletonization: a technique for trimming the fat from a network via relevance assessment. *Adv. Neural Network Inform. Process. Syst.* 1989.

⁵ NIIDA, K.; TANI, J. & KOSHIJIMA, I. Application of neural network to rule extraction from operation data. *Computers chem. Engng.* In press (1992).

também citados no mesmo artigo, apresentam um método que não exclui os nós, mas os pesos individuais. Essa técnica foi baseada na minimização de uma função objetivo dual, durante a fase de treinamento.

A idéia básica é reduzir o tamanho da rede pela identificação e exclusão de pesos redundantes na rede. Pesos redundantes são definidos como aqueles que, dada uma entrada não causam nem efeito excitatório nem inibitório na saída do nó. Se uma função sigmóide de cada nó é centrada em torno de zero (como é o caso da tangente hiperbólica), então uma saída positiva pode ser chamada de excitação, e uma saída negativa um sinal inibitório. É simples de se notar que, para tal sigmóide, um peso com magnitude zero vai resultar em uma saída, nem excitatória nem inibitória da sigmóide. Em outras palavras, o peso associado com uma entrada redundante terá magnitude zero, ou seja, é redundante. Uma aproximação é usada uma vez que pesos pequenos em valor absoluto são categorizados como redundantes e excluídos. Bhat & McAvoy (1992) propõem então usar a função sigmóide, conforme Equação 13, modificada de:

$$\sigma(x) = \frac{1}{1+e^{-x}} \quad \text{para} \quad \sigma(x) = \frac{1}{1+e^{-x}} - 0,5 \quad [13]$$

que foi re-escalada para a faixa de $-0,5$ a $+0,5$.

Hassib & Stork (1993), por outro lado, investigaram o uso de todas as derivadas de segunda ordem da função erro para redução da rede neural (ou seja, removendo os pesos sem importância, de uma rede treinada) de modo a melhorar a generalização, simplificar as redes, reduzir as necessidades computacionais, de memória e de armazenamento, aumentar a velocidade de retreinamentos das redes, e em alguns casos, permitir a extração de regras. Para isso propuseram o método “*Optimal Brain Surgeon*” – OBS, afirmando ser melhor que os métodos baseados em magnitude e que o método “*Optimal Brain Damage*”, proposto por Le Cun, Denker & Solla (1990)⁶.

⁶ LE CUN, Y.; DENKER, J. S. & SOLLA, S. A. Optimal Brain Damage in *Proceedings of the Neural Information Processing Systems-2*, Touretzky (ed.), p. 598-605. 1990.

Algoritmo "Optimal Brain Surgeon".

Para a aplicação do método considera-se que a rede foi treinada até que o erro atinja um mínimo local. A função do erro, explícita em termos de séries de Taylor, com relação ao erro, é dada pela Equação 14

$$\partial E = \left(\frac{\partial E}{\partial \mathbf{w}} \right)^T \cdot \hat{\mathbf{c}}\mathbf{w} + \frac{1}{2} \hat{\mathbf{c}}\mathbf{w}^T \cdot \mathbf{H} \cdot \hat{\mathbf{c}}\mathbf{w} + O(\|\hat{\mathbf{c}}\mathbf{w}\|^3) \quad [14]$$

onde $\mathbf{H} \equiv \partial^2 E / \partial \mathbf{w}^2$ é a matriz Hessiana (contendo todas as derivadas de segunda ordem);

\mathbf{w} são os pesos;

e o sobrescrito \mathbf{T} indica a transposta do vetor.

Para uma rede treinada até um erro em mínimo local, o primeiro termo (linear) é desprezível, e os termos de terceira ordem, e superiores, são ignorados. O objetivo é então ajustar um dos pesos a zero (chamado de w_q) para minimizar o aumento do erro, dado na Equação 14.

A eliminação de w_q é expressa pela Equação 15 como:

$$\hat{\mathbf{c}}\mathbf{w}_q + w_q = 0 \quad [15]$$

ou mais genericamente, conforme Equação 16:

$$\mathbf{e}_q^T \cdot \hat{\mathbf{c}}\mathbf{w} + w_q = 0 \quad [16]$$

onde \mathbf{e}_q é o vetor unitário no espaço dos pesos, correspondente ao peso (unitário) w_q . O objetivo é, então, resolver a Equação 17:

$$\text{Min}_q \left\{ \text{Min}_{\hat{\mathbf{c}}\mathbf{w}} \left\{ \frac{1}{2} \hat{\mathbf{c}}\mathbf{w}^T \cdot \mathbf{H} \cdot \hat{\mathbf{c}}\mathbf{w} \right\} \quad \text{tal que} \quad \mathbf{e}_q^T \cdot \hat{\mathbf{c}}\mathbf{w} + w_q = 0 \right\} \quad [17]$$

Para resolver, forma-se uma Lagrangeana a partir da Equação 14 e da Equação 15, gerando a Equação 18:

$$L = \frac{1}{2} \hat{\mathbf{c}}\mathbf{w}^T \cdot \mathbf{H} \cdot \hat{\mathbf{c}}\mathbf{w} + \lambda (\mathbf{e}_q^T \cdot \hat{\mathbf{c}}\mathbf{w} + w_q) \quad [18]$$

onde λ é o multiplicador Lagrangeano indeterminado.

A mudança ótima nos pesos é obtida pelas derivadas das funções, utilizando as restrições da eliminação de w_q , e invertendo a matriz, cuja mudança no erro resulta na Equação 19:

$$\hat{c}w = -\frac{w_q}{[H^{-1}]_{qq}} H^{-1} \cdot e_q \quad \text{e} \quad L_q = \frac{1}{2} \frac{w_q^2}{[H^{-1}]_{qq}} \quad [19]$$

Segundo os autores, nem \mathbf{H} , nem \mathbf{H}^{-1} precisa ser diagonal (como é o caso assumido por Le Cun et al., 1990). Além disso, o método recalcula a magnitude de todos os pesos da rede, através da parte esquerda das equações descritas logo acima. O termo L_q , acima, é a “saliência” do peso q – o aumento no erro que resulta quando este peso é eliminado.

Nørgaard (1995) desenvolveu e disponibilizou um conjunto de “ferramentas” (toolbox) para o software MATLAB, com um grande número de funções para o treinamento e avaliação de redes neurais tipo MLP. Entre essas funções está a OBS, utilizada neste trabalho.

Análise de Significância das Variáveis de Entrada

Garson (1991) propõe o particionamento dos pesos da rede neural para torna-la boa, tanto para modelagem, como para a identificação de padrões. Embora seja possível a comparação entre os resultados das predições e classificações gerados pelas redes neurais com outras técnicas, tais como regressão múltipla, ou análise discriminatória, ainda não há uma metodologia aceita que relacione a importância relativa dos fatores de entrada, usados pela rede para obter suas conclusões. Ao invés disso, as redes neurais têm sido apresentadas aos seus usuários como um tipo de “caixa preta”, cujo interno, inimaginavelmente complexo, opera de alguma forma, um tanto quanto mágica, transformando as entradas em saídas preditas. Atualmente a melhor compreensão das redes neurais tem minimizado essa imagem de “caixa preta”. Os pesos que conectam, tanto os nós de entrada aos nós

intermediários, como os nós intermediários até a saída, podem ser utilizados para identificar o grau relativo da predição da saída, associado a cada variável de entrada.

$$\frac{\sum_j^{n_H} \left(\frac{I_{Vj}}{\sum_k^{n_V} I_{Vj.k}} O_j \right)}{\sum_i^{n_V} \left(\sum_j^{n_H} \left(\frac{I_{Vj.i}}{\sum_k^{n_V} I_{Vj.k.i}} O_j \right) \right)} \quad [20]$$

onde, na Equação 20,

I é o peso que conecta o V -ésimo nó de entrada ao j -ésimo nó da camada intermediária;

n_H é o total de nós da camada intermediária (“*hidden*”);

O é o peso que conecta o j -ésimo nó da camada intermediária ao nó de saída;

n_V é o total de variáveis de entrada (nós de entrada).

Para cada um dos j nós n_H da camada intermediária, a soma dos produtos formados pela multiplicação dos pesos que conectam os nós de entrada I , relativos à variável V , e o nó j da camada intermediária, com o peso que conecta o nó intermediário j à saída O , é dividida pela soma de todas essas quantidades, para todas as variáveis. O resultado é a porcentagem de todos os pesos de saída, atribuíveis a uma dada variável independente, excluindo o peso do *bias*, proveniente do algoritmo *backpropagation*. Garson (1991) explica o porque da exclusão do *bias* da sua equação, e utiliza os valores absolutos de cada peso. Assim a equação estabelece uma relação entre as relações do valor absoluto do peso de cada entrada e a soma dos valores absolutos dos demais pesos que chegam àquele nó

hidden, vezes o peso de saída desse nó hidden, e o somatório dessas relações, para cada nó hidden, gerando a importância relativa de cada entrada na saída gerada.

Elgibaly & Elkamel (1998) indicam as vantagens na aplicação de redes neurais, sobre os métodos estatísticos, onde cada nó não é totalmente responsável pelo mapeamento de uma saída para uma dada entrada, e somente quando o efeito de todos os nós é combinado, a rede torna-se capaz de prever uma saída correspondente a dadas entradas. Apresentam uma aplicação, na qual denominam a equação de Garson (Equação 20) de Medida de Importância – IM.

Achamos questionável, nos trabalhos de Garson (1991) e Elgibaly & Elkamel (1998), a utilização dos valores absolutos dos pesos. Ao calcularmos a relação entre o peso de uma variável de entrada, e a soma dos pesos de todas as entradas em determinado nó da camada intermediária, sem considerar os respectivos sinais correspondentes (utilizando os seus valores absolutos), estamos desconsiderando as inter-relações que as variáveis de entrada têm entre si, na entrada de cada nó da camada intermediária, quando são somadas, conforme a Equação 5, antes de sofrerem a transformação pela função de transferência. Neste trabalho efetuaremos os cálculos de ambas as formas, considerando tanto o valor natural dos pesos, como de seus valores absolutos, comparando os resultados obtidos.

Para Gontarski (2000), na aplicação da análise de significância, a rede deve estar devidamente treinada e gerando bons resultados mesmo para condições não utilizadas no treinamento. Ele utilizou o critério de acrescentar e descontar 10% ao valor da média de cada valor de entrada para se ter os níveis, inferior e superior, ao aplicar a metodologia de Planejamento de Experimentos, considerando um dos procedimentos fundamentais para o sucesso da análise a escolha dos fatores a analisar e os respectivos níveis de variação. Aplicou a metodologia de experimento fatorial fracionado, para reduzir o número de experimentos necessários à metodologia de Planejamento de Experimentos, uma vez que estudou 9 fatores a dois níveis, requerendo 2^9 experimentos, e conseguiu reduzir para $2^{9/2}$ experimentos, utilizando o programa MINITAB para auxiliar na execução da análise.

Gontarski et al. (2000) propuseram uma análise de sensibilidade baseada no índice de correlação do modelo, R_j , entre os valores experimentais e os valores calculados para a variável j , conforme a Equação 21 a seguir:

$$R_j = \frac{N \left[\sum_1^N (x_i - \mu_x)(y_i - \mu_y) \right]}{\sqrt{\left[N \sum_1^N x_i^2 - \left(\sum_1^N x_i \right)^2 \right] \left[N \sum_1^N y_i^2 - \left(\sum_1^N y_i \right)^2 \right]}} \quad [21]$$

$(j = 1, N_V)$

onde

N é o número de dados utilizados;

x_i é o valor calculado para a j -ésima variável;

y_i é o valor experimental da j -ésima variável;

μ_x é a média dos valores calculados pela rede neural;

μ_y é a média dos valores experimentais;

N_V o número de nós de entrada da rede.

A análise de sensibilidade é efetuada pelo estudo do efeito da eliminação de uma variável de entrada no índice de correlação do modelo, em relação ao índice de correlação da rede treinada com todas as variáveis de entrada, conforme Equação 22:

$$(\% \text{ influência })_j = \left(1 - \frac{R_j}{R_{CB}} \right) \cdot 100, \quad (j = 1, N_V) \quad [22]$$

onde R_{CB} é o índice de correlação entre os valores experimentais e os valores calculados, para o caso base.

Como essa análise não permitiu uma conclusão bem definida sobre a importância das variáveis de entrada no resultado desejado, foi efetuada uma nova análise de sensibilidade, agora baseada em pequenos incrementos dados a cada variável de entrada, conforme a Equação 23 a seguir:

$$V_j^a = V_j^0 \cdot w_j, \quad (j = 1, N_v) \quad [23]$$

onde

V_j^a é o valor da j -ésima variável utilizado no modelo para obter o índice R_j ;

V_j^0 é o valor da j -ésima variável utilizado no modelo para obter o índice R_j ;

w_j é o fator aplicado a cada j -ésima variável.

Levantamento dos Dados de Entrada

Os dados foram coletados a partir de relatórios gerados automaticamente pelo sistema supervisor existente na fábrica. Essa coleta foi feita em época em que houve muitos feriados, por isso algumas análises não foram efetuadas e alguns conjuntos de dados foram rejeitados. Além disso, os dados operacionais eram coletados de hora em hora, enquanto que os de laboratório a cada quatro horas, o que contribuiu para a eliminação de outro grande conjunto de dados. Dessa maneira, ficaram somente trinta conjuntos de dados para treinamento e validação da rede.

Yeh (1998) recomenda o uso de setenta e cinco por cento dos dados para treinamento e os restantes vinte e cinco por cento para validação da rede.

Dias (2000), que possuía um conjunto com vinte e oito dados experimentais, utilizou-o todo para o treinamento da rede. Embora não pudesse testar a rede treinada, estudou a sensibilidade da saída da rede a mudanças em cada entrada, de seu valor máximo para um valor mínimo, mantendo as demais constantes.

Gagné & Blaise (1997) trabalharam com rede probabilística auto-organizada, e utilizaram dez dados para treinamento e dez para validação da rede, ou seja, 50% dos dados.

Pollard et al. (1992) comentam que o mapeamento, entre as entradas e saídas, obtido pelas redes neurais, pode ser determinado sem qualquer entendimento de suas relações funcionais. A desvantagem dessa abordagem é a necessidade de especificar um conjunto de entradas que abranja a região de interesse, uma vez que, sem o conhecimento

da real estrutura do processo, os resultados de uma experiência não podem ser extrapolados para fora da região em que a experiência foi realizada. Apresentam, também, o conceito de “*cross validation*”, ou validação cruzada, onde, após o treinamento da rede neural, utilizando um conjunto de dados de amostras, a qualidade do mapeamento é avaliada usando-se um diferente conjunto de dados. O melhor mapeamento é aquele que minimiza o erro da predição num conjunto de dados para os quais não foi treinada. Recomendam o uso da validação cruzada para controlar a quantidade de ruído que é aprendida, limitando o número de vezes que é permitido à rede para examinar o conjunto de dados durante o ciclo de aprendizagem. O treinamento da rede neural é interrompido antes que o mínimo erro, pelo critério dos mínimos quadrados, seja atingido. A motivação dessa abordagem é a obtenção da melhor representação do modelo atual do sistema, a partir de dados com ruído.

Em nosso trabalho, utilizamos nos vários testes diferentes tamanhos de conjuntos de dados para treinamento e para validação. Uma vez escolhido o tamanho do conjunto de dados, os dados para validação foram obtidos aleatoriamente. Um tamanho de dados para validação, que apresentou bons resultados, devido ao pequeno tamanho do conjunto de dados experimentais, e que foi bastante utilizado, era com cinco dados, o que significa cerca de 17%, e, conseqüentemente, 83% dos dados foram utilizados para treinamento.

Normalização das Variáveis de Entrada

Murkherjee & Biswas (1997) perceberam que, com a rede treinada embora houvesse boa concordância com os dados experimentais, as predições divergiram levemente dos dados experimentais nas extremidades das curvas. Essa divergência foi atribuída aos valores objetivos próximos a zero, ou a um, uma vez que os dados foram normalizados entre zero e um, e nessa região das extremidades da faixa, a inclinação da função sigmoideal é muito baixa, sendo desejável manter uma amplitude dos valores objetivos entre 0,2 e 0,8, uma vez que, com esse procedimento a rede apresentou significativa melhora.

Em nosso trabalho, nos testes iniciais, os valores das entradas foram normalizados entre -0,9 e +0,9, enquanto que os valores das saídas foram entre -0,8 e +0,8,

principalmente para utilizar a função de transferência tangente hiperbólica, necessária, como vimos, para o uso do OBS, cujos limites estão entre -1 e $+1$. Desta maneira, todos esses valores médios situam-se em torno de zero, e, conseqüentemente, a saída da função de transferência deve, também, ser em torno de zero, e a normalização deve afastar os valores das extremidades da curva.

Rede e treinamento iniciais

A rede inicial foi definida com vinte nós de entrada e dois de saída:

- O2 – Teor (%) de Oxigênio Residual dos gases na saída da torre de ciclone;
- OPAC – Emissão de particulados na saída da chaminé (Opacidade);
- TMPA, TMPB – Temperatura de saída dos gases da torre de ciclone nos ramais A e B, respectivamente;
- TORQ – Medida do torque do motor elétrico, para rotação do forno rotativo: quando aumenta indica que o forno está esquentando, e quando diminui indica que o forno está esfriando.
- NOX – Teor de NOx (ppm) nos gases de Combustão: quando aumenta indica que o forno está esquentando, e quando diminui indica que o forno está esfriando.
- FARIN – Vazão de farinha (t/h), alimentando o forno;
- CARV – Vazão de carvão (t/h) sendo alimentada nos maçaricos do forno (combustível);
- OLEO – Vazão de óleo combustível (t/h) para os maçaricos a óleo do forno;
- RPM – Rotação do forno rotativo em rpm;
- FANA, FANB – Rotação dos exaustores (ventiladores de tiragem), em %, nos ramais A e B, respectivamente;
- PRESA, PRESB – Perda de pressão (mbar) na saída da torre de ciclone nos ramais A e B, respectivamente;
- PCXF – Perda de pressão (mbar) na caixa de fumaça, situada na saída do forno e entrada da torre de ciclone;
- CLINK – Temperatura do clínquer na saída do resfriador;

- PH-CO – Teor de Monóxido de Carbono na saída da torre de ciclone: indica ocorrência de combustão incompleta;
- FSC, MS, MA – Dados de qualidade da farinha: Fator de Saturação, Módulo de Sílica e Módulo de Alumínio, respectivamente.

Forno 07 - Rede Totalmente Conectada ⇒ Nós hidden: 25

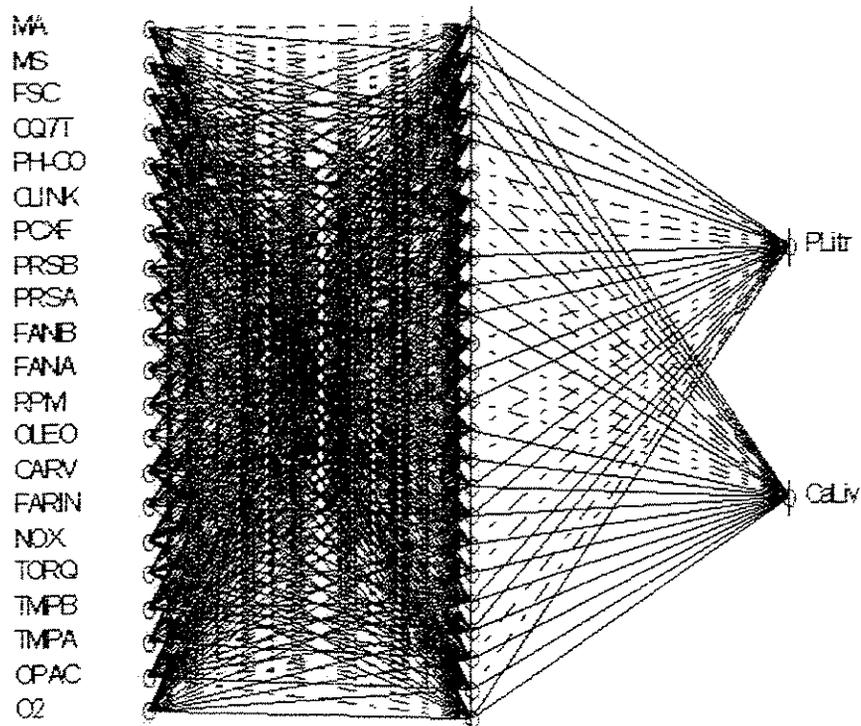


Figura 14 - Rede inicialmente proposta, com 25 nós intermediários, predizendo Cal Livre e Peso Litro

Inicialmente procuramos treinar uma única rede, com vários nós na camada intermediária, para calcular, tanto a Cal Livre como o Peso Litro. Para isso escolhemos um número de nós, da camada intermediária, maior que o número de nós de entrada, considerando que a aplicação do OBS pudesse otimizar o número de pesos da rede.

A função de transferência usada, como dito anteriormente, foi a tangente hiperbólica. Para o treinamento da rede foi escolhido o método do gradiente descendente

(*backpropagation*). Para o treinamento e teste da rede foi desenvolvido um programa em MATLAB.

Análise de Sensibilidade da Rede Treinada

Na análise de sensibilidade, após o treinamento da rede, e para que a função gere uma saída um pouco mais significativa, uma vez que o valor médio de cada entrada deve situar-se na parte central da função hiperbólica, o que pode não gerar uma excitação na saída da função, alteramos a média para um valor 20% acima, para calcular a influência de cada entrada nos parâmetros desejados: Cal Livre, ou Peso Litro. Assim, como caso base, para comparação da influência de cada entrada, a saída é calculada a partir de um conjunto de dados de entrada composto pela média de cada variável de entrada. Depois, cada variável de entrada tem sua média acrescida de 20%, e novamente será calculada a nova saída gerada, que, comparada com o valor gerado pelas médias, indicará a alteração real sofrida pela variável calculada, pelo aumento em cada entrada. Como as variáveis de entrada foram normalizadas, os percentuais calculados, para cada alteração na entrada, são comparáveis entre si, podendo ser utilizados para identificação da maior, ou menor influência de cada entrada na saída gerada. Procuramos um valor que promovesse um impacto significativo na saída, maior que os 10% propostos por Gontarski (2000), evitamos, porém, os dados nos limites, inferior ou superior, como utilizado por Dias (2000).

Capítulo 4

RESULTADOS E DISCUSSÕES

Dados de Entrada

Os dados de entrada, depois de normalizados entre $-0,9$ e $+0,9$, foram comparados com os dados de saída, também normalizados para a mesma faixa, cujos gráficos estão mostrados no Apêndice. A grande maioria dos dados apresentou uma boa distribuição, embora alguns deles apresentassem concentração em determinados pontos. A título de ilustração mostraremos apenas dois casos, apresentados na Figura 15 e na Figura 16, mas a análise pode ser estendida similarmente a todos os demais dados de entrada.

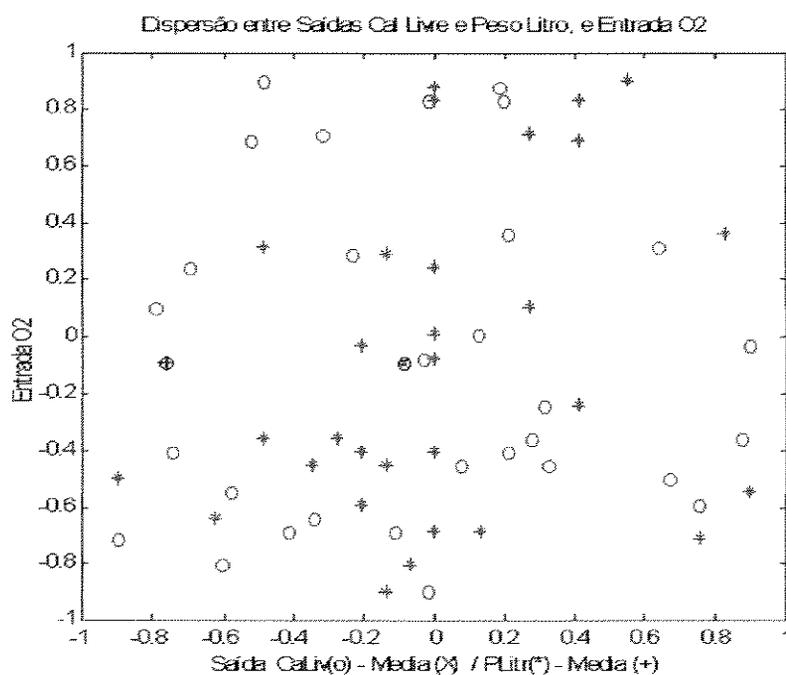


Figura 15 – Os dados do teor de oxigênio se distribuem bem em relação à Cal Livre (o – azul), e também e em relação ao Peso Litro (* - vermelho). O valor médio dos dados dessa entrada está relacionado com a média da Cal Livre, marcada com (x) e com a média do Peso Litro, marcada com (+).

A Figura 15 mostra a boa distribuição entre os dados do teor de oxigênio residual dos gases na saída da torre de ciclone (O₂), em relação aos respectivos dados de saída da Cal Livre, pelos pontos marcados em azul (o), e em relação aos respectivos dados de Peso Litro, marcados em vermelho (*). Foi efetuada a média dos dados de entrada, que também está relacionada com a média dos valores da Cal Livre, marcada com (x) e com a média dos valores do Peso Litro, marcada com (+).

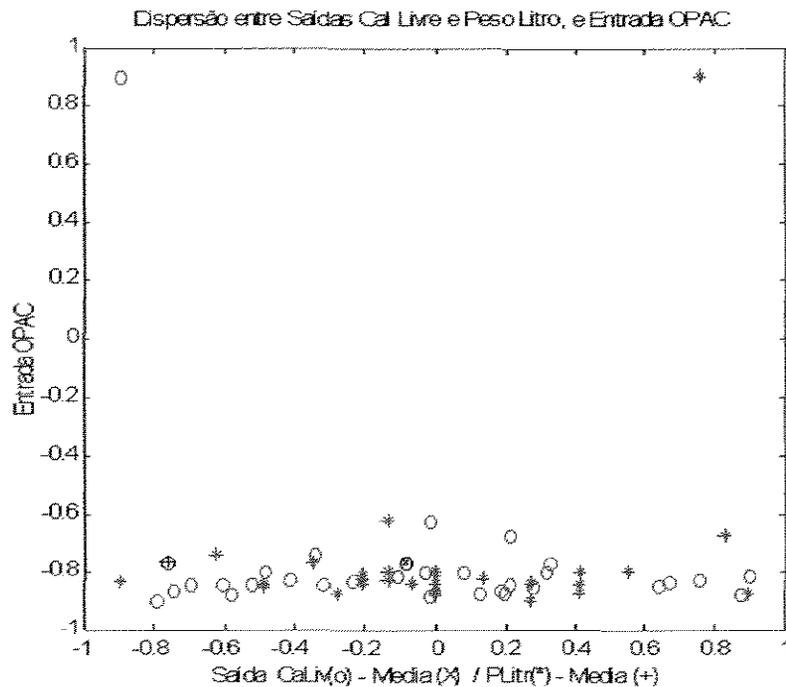


Figura 16 - Os dados de emissão de particulados se distribuem mal em relação às respectivas saídas de Cal Livre, marcadas em azul (o) e às respectivas saídas de Peso Litro, marcadas em vermelho (*). O valor médio dos dados dessa entrada está relacionado com a média da Cal Livre, marcada com (x) e com a média do Peso Litro, marcada com (+).

A Figura 16 também mostra a relação, porém mal distribuída, entre os dados de entrada referentes à emissão de particulados na saída da chaminé (OPAC), e as respectivas saídas de Cal Livre, marcadas em azul (o) e as respectivas saídas de Peso Litro, marcadas em vermelho (*). A média dos dados de entrada também está relacionada com a média dos valores da Cal Livre, marcada com (x) e, com a média dos valores do Peso Litro, marcada com (+). Alguns outros dados de entrada também apresentam esse tipo de má distribuição, que pode ser causada por deficiência na precisão dos instrumentos de medição, que

aparentemente operam em algumas faixas, e/ou por situações operacionais anormais, que podem ter ocorrido durante a coleta de dados.

A eliminação de vários conjuntos de dados, conforme explicado no item: *Levantamento dos Dados de Entrada* do capítulo anterior, tornou-os esparsos em relação ao tempo, impedindo uma análise de modelo temporal.

Treinamento da Rede: 25 nós intermediários, 2 saídas

A rede neural inicialmente proposta, com vinte entradas e vinte e cinco nós na camada intermediária, predizendo tanto a Cal Livre quanto o Peso Litro, apresentou, depois de treinada, os seguintes resultados para a Cal Livre:

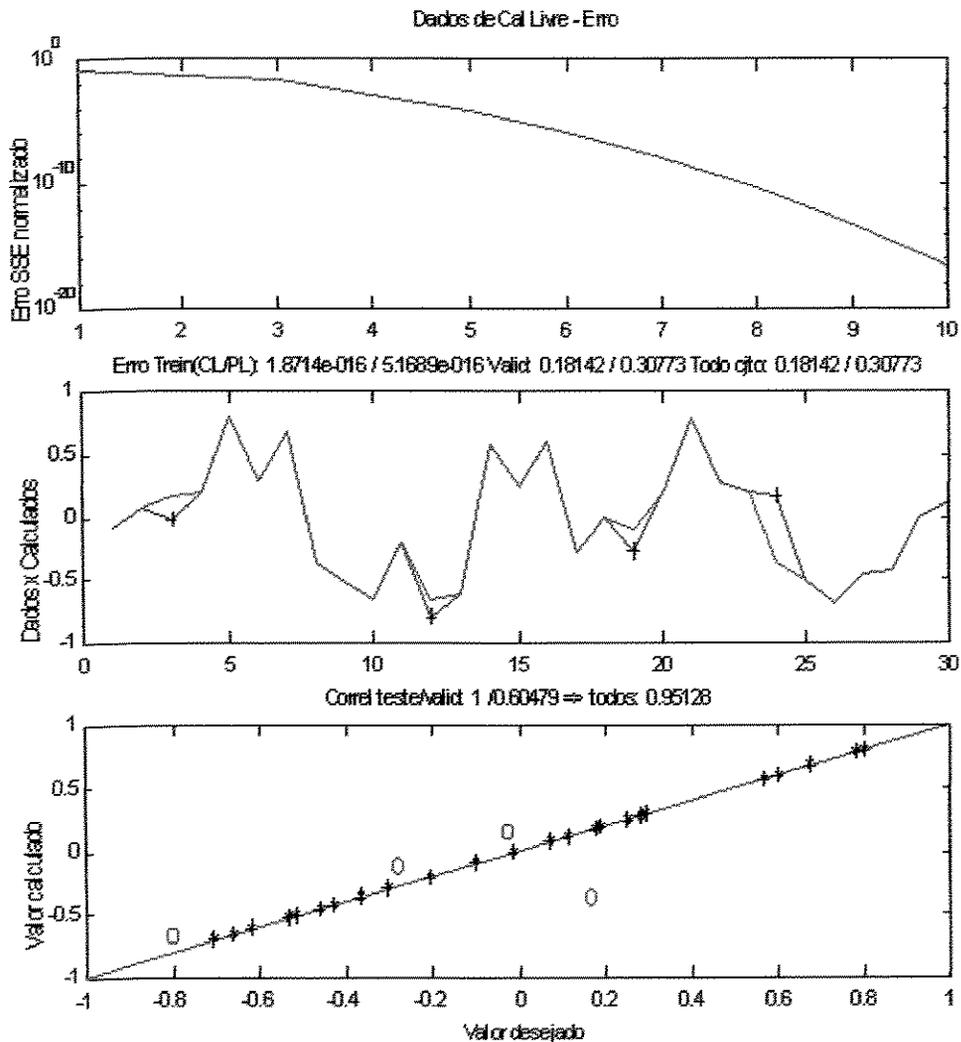


Figura 17 - Saída da rede com 25 nós intermediários, predizendo Cal Livre. No segundo gráfico, mostrando os valores calculados e os dados reais, os pontos (*) indicam os dados utilizados para validação. No terceiro gráfico, relacionando os valores calculados e os valores desejados, os pontos marcados com (*) são os dados utilizados para treinamento, enquanto que os dados marcados com (o) são os dados de validação.

Para a predição do Peso Litro a rede treinada apresentou a seguinte saída:

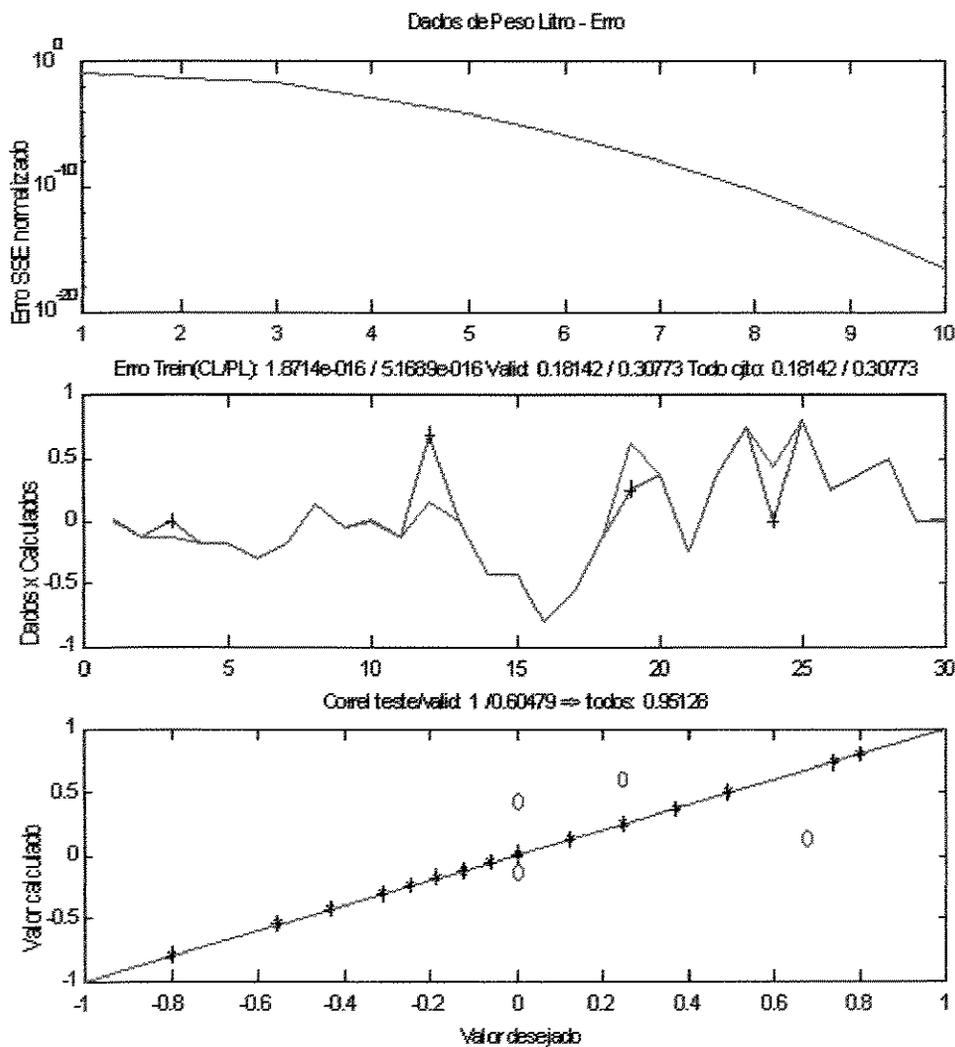


Figura 18 - Saída da rede com 25 nós intermediários, predizendo Peso Litro. A representação dos pontos é a mesma da Figura 17.

Pela análise desses resultados, vemos que a rede foi capaz de aprender bem, pelo seu índice de correlação $R^2=0,95$, considerando, na correlação, todos os pontos, incluindo os utilizados para treinamento e aqueles usados somente para validação da rede. Essa rede teve razoável habilidade de generalização, pelo índice de correlação dos quatro pontos

sorteados para validação: $R^2=0,60$. Disso deduz-se que a rede deve ter sido sobre treinada, e, muito provavelmente, que há poucos pontos para treinamento da rede.

Para reduzir um pouco a influência dos ruídos dos dados de entrada, o erro aceitável para treinamento foi elevado de 0,0005 para 0,002.

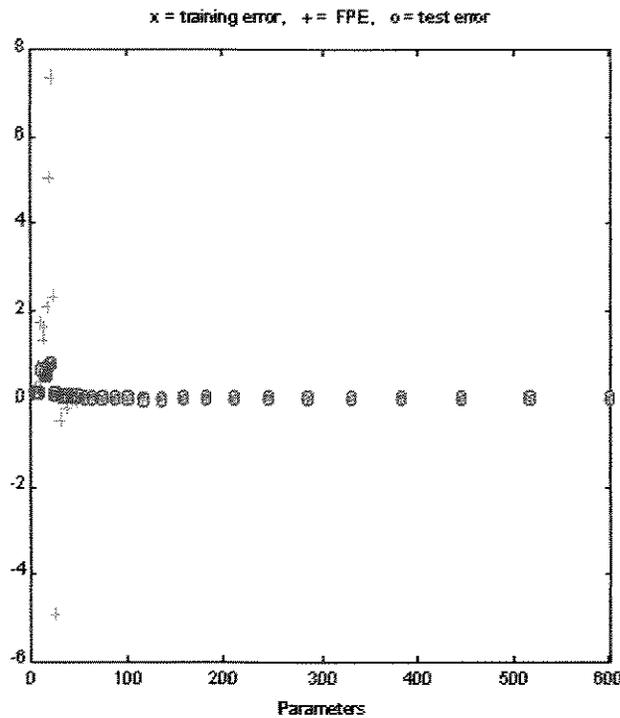


Figura 19 - Curva dos erros após retreinamento da rede, depois da redução de pesos com a metodologia OBS

Essa rede possui $(20*25 + 25*2)$ 550 pesos, sem considerar os pesos conectados aos *bias*. Aplicando o método OBS, para redução e otimização do número de pesos da rede, obtivemos gráfico de erro da Figura 19, após cada procedimento de remoção de pesos, pelo método OBS.

Aplicando a rede a todos os dados, após cada processo de remoção de pesos (14 % dos pesos de cada vez), obtivemos, na curva de erros da Figura 19, que está representada pelo gráfico em vermelho (o). Aplicando a rede somente aos dados de treinamento, os erros gerados estão representados pelos pontos em (+), e os pontos (*) representam as estimativas dos erros de generalização, para a rede com cada novo conjunto de pesos.

Analisando essas curvas de erros, obtivemos um ponto de mínima em 117 pesos, sem considerar os pesos relativos aos *bias*. Desse modo, a rede com os pesos otimizados ficou como mostrado na Figura 20:

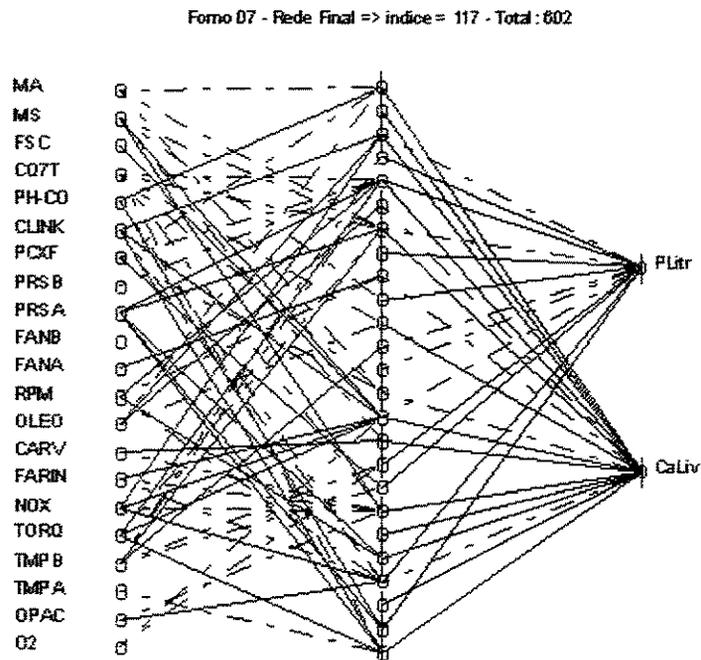


Figura 20 – Rede com 25 nós intermediários, após a redução dos pesos para 117

Notamos, também, que alguns nós de entrada ficaram sem conexões (pesos), assim como alguns da camada intermediária, indicando que não foram importantes para a modelagem.

As saídas desta rede, como podemos ver na Figura 21 e na Figura 22, para a predição da Cal Livre e do Peso Litro, respectivamente, apresentaram melhora no índice de correlação ($R^2=0,97$), e, também melhoraram sua capacidade de generalização, pois o índice de correlação para os dados para validação subiu para 0,79.

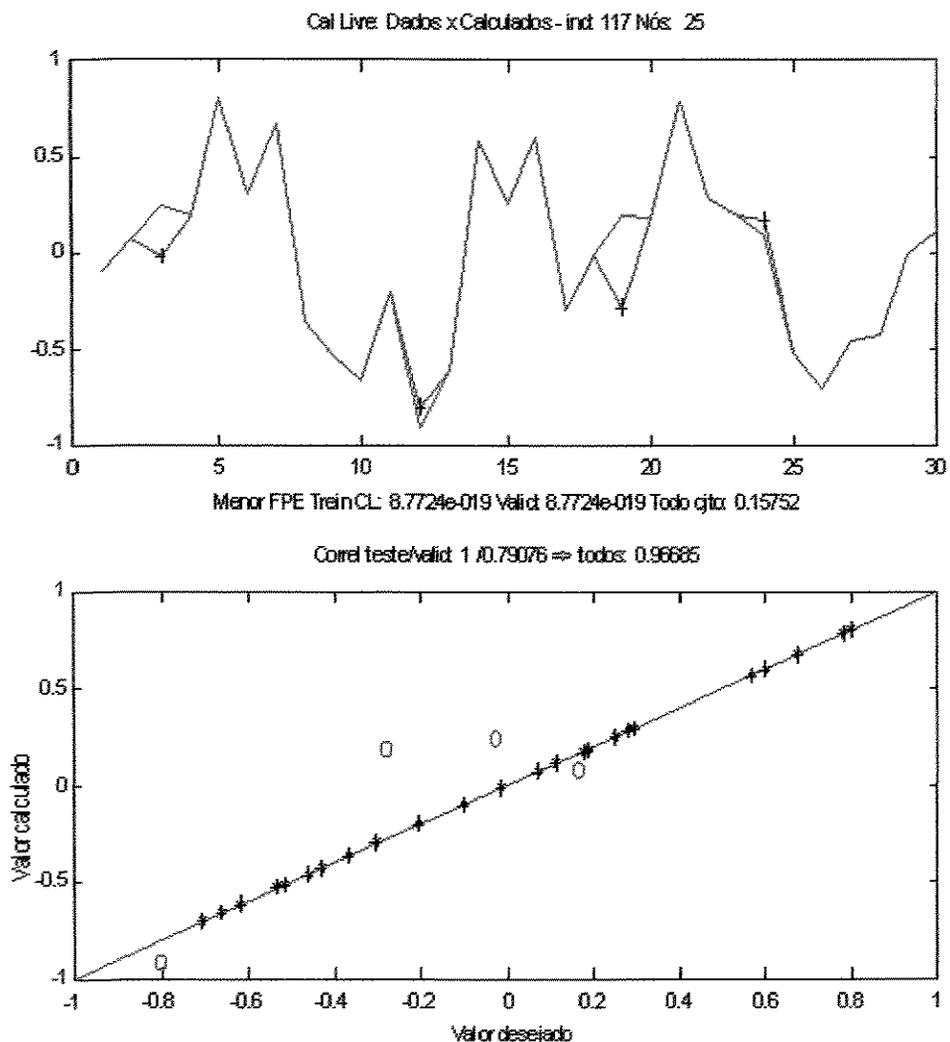


Figura 21 - Saída da rede reduzida para 117 pesos, predizendo Cal Livre

Na Figura 21 e na Figura 22, nos gráficos das respectivas partes superiores, as curvas em azul representam os pontos preditos pela rede para os pontos experimentais, marcados com (+) sobre a curva de dados esperados (experimentais), em vermelho.

A parte inferior dessas figuras apresenta a correlação entre os dados experimentais e os dados preditos pelo modelo. Se os valores forem coincidentes, todos os pontos devem situar-se sobre a diagonal marcada no gráfico, como ocorre com todos os dados utilizados

para treinamento da rede. Os pontos marcados em vermelho (o) representam os pontos sorteados para a validação da rede.

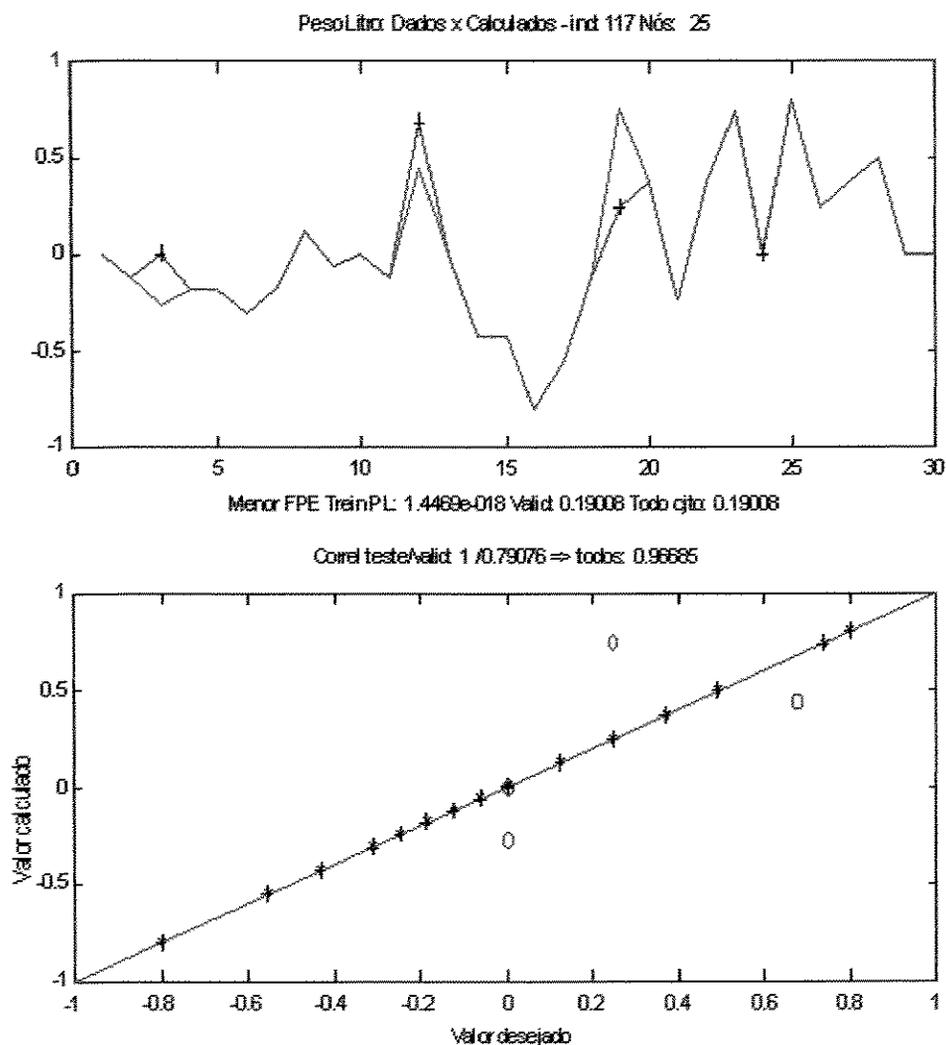


Figura 22 - Saída da rede reduzida para 117 pesos, predizendo Peso Litro

Comparando a saída dessa rede reduzida para 117 pesos com a saída da rede original (550 pesos), observamos que a generalização da rede, para os pontos sorteados para validação, não foi muito boa, nem para o Peso Litro, e nem para a Cal Livre, como podemos ver na Figura 23.

O pequeno número de dados, para treinamento e validação da rede, é o maior responsável por esse comportamento.

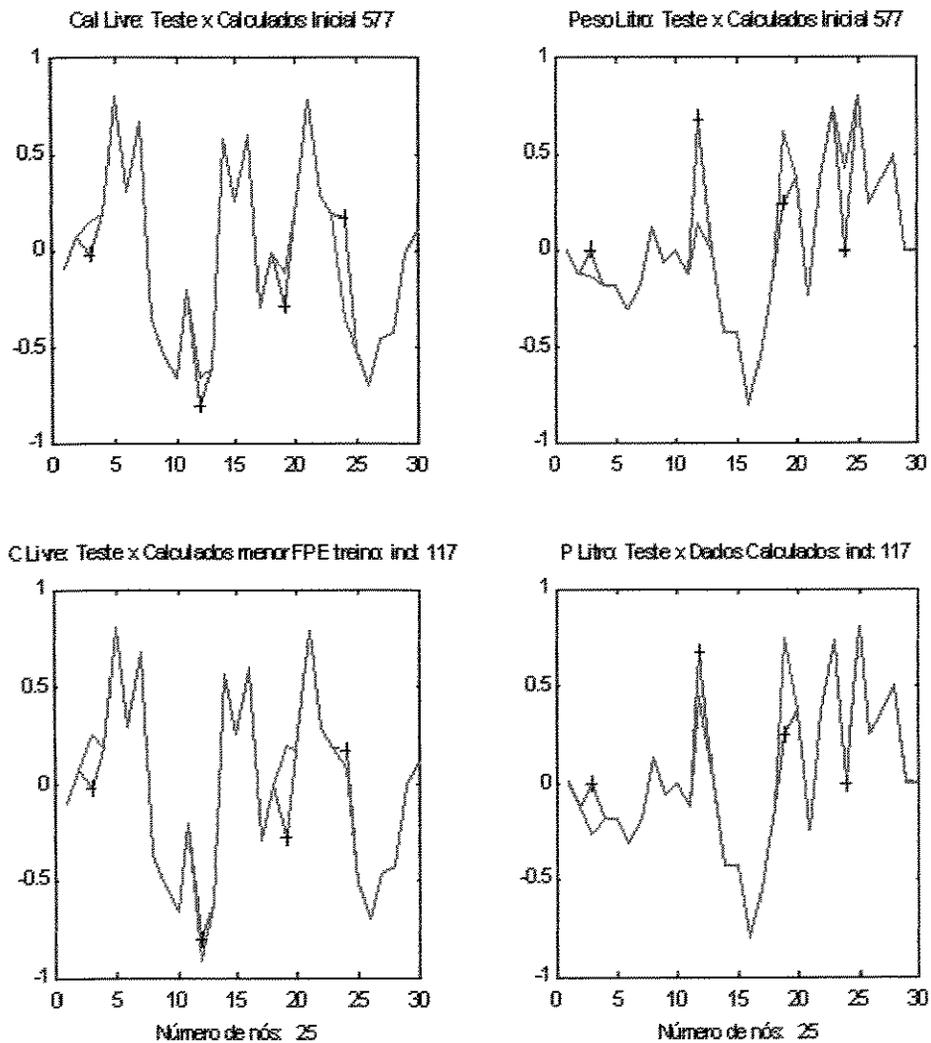


Figura 23 - Comparação entre as saídas da rede inicial e da rede com o número de pesos reduzido para 117.

Na Figura 23 a curva de dados experimentais está em vermelho, e o ponto (+) sobre ela representa cada resultado separado para a validação do modelo, cujo resultado, a curva em azul, apresenta um desvio em relação ao valor esperado.

Analisamos, também, uma rede com menor número de pesos, que apresentou baixo erro após o retreinamento, e que foi reduzida para 28 pesos.

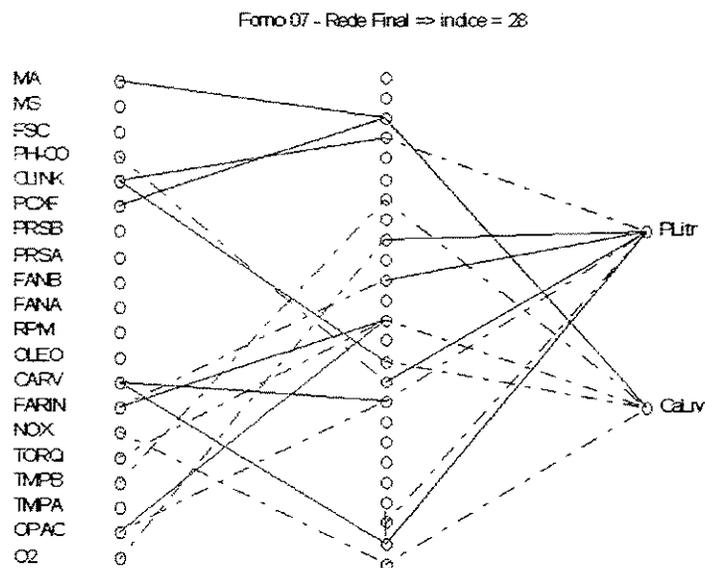


Figura 24 - Rede com número de pesos reduzido para 28

Essa drástica redução de pesos eliminou a conexão de nós de entrada à camada intermediária, indicando que essas entradas não devem ser importantes para a predição das propriedades estudadas. Por outro lado, ao eliminar as conexões entre os nós da camada intermediária até a camada de saída, há forte indicação do excesso de nós na camada intermediária, que também não devem ser necessários.

As saídas desta rede pioraram significativamente, como pode ser visto na Figura 25 e na Figura 26, uma vez que o índice de correlação calculado foi ($R^2=0,74$). Sua capacidade de generalização também piorou, como pode ser visto na parte inferior de ambas figuras, pois o índice de correlação para os dados de validação ficou em 0,50. Provavelmente houve a eliminação de pesos importantes, por isso a má predição obtida, ou seja, ou algum nó de entrada ou algum nó da camada intermediária, importante para o modelo foi eliminado.

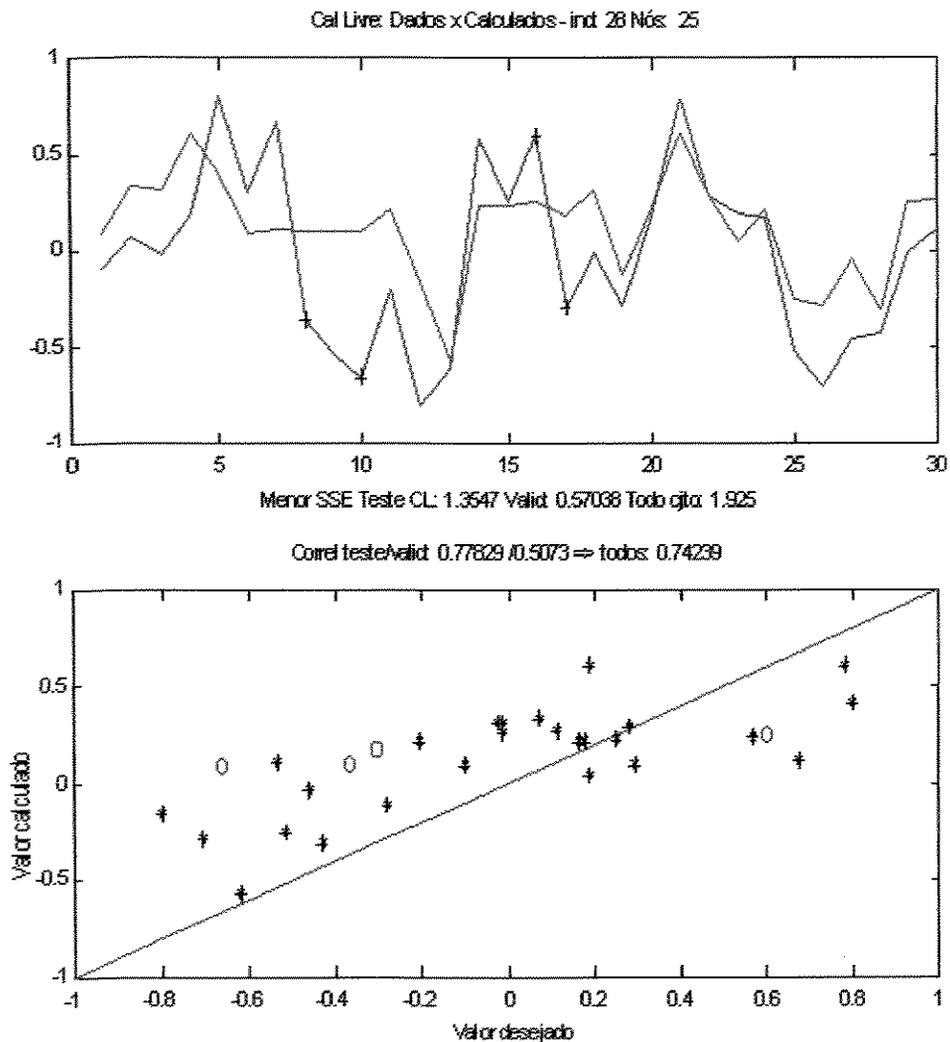


Figura 25 - Saída da rede reduzida para 28 pesos, predizendo Cal Livre

Na Figura 25, a curva em vermelho representa os pontos experimentais para predição da Cal Livre, sobre a qual está marcado o ponto experimental (+) utilizado para validação da rede, e a curva em azul mostra o desvio dos dados preditos em relação aos dados experimentais. Na parte inferior o gráfico, representando a correspondência entre os dados experimentais e os dados preditos, mostra que os dados utilizados para treinamento (*) não apresentaram boa concordância, assim como os dados sorteados, utilizados para validação da rede (o), também não.

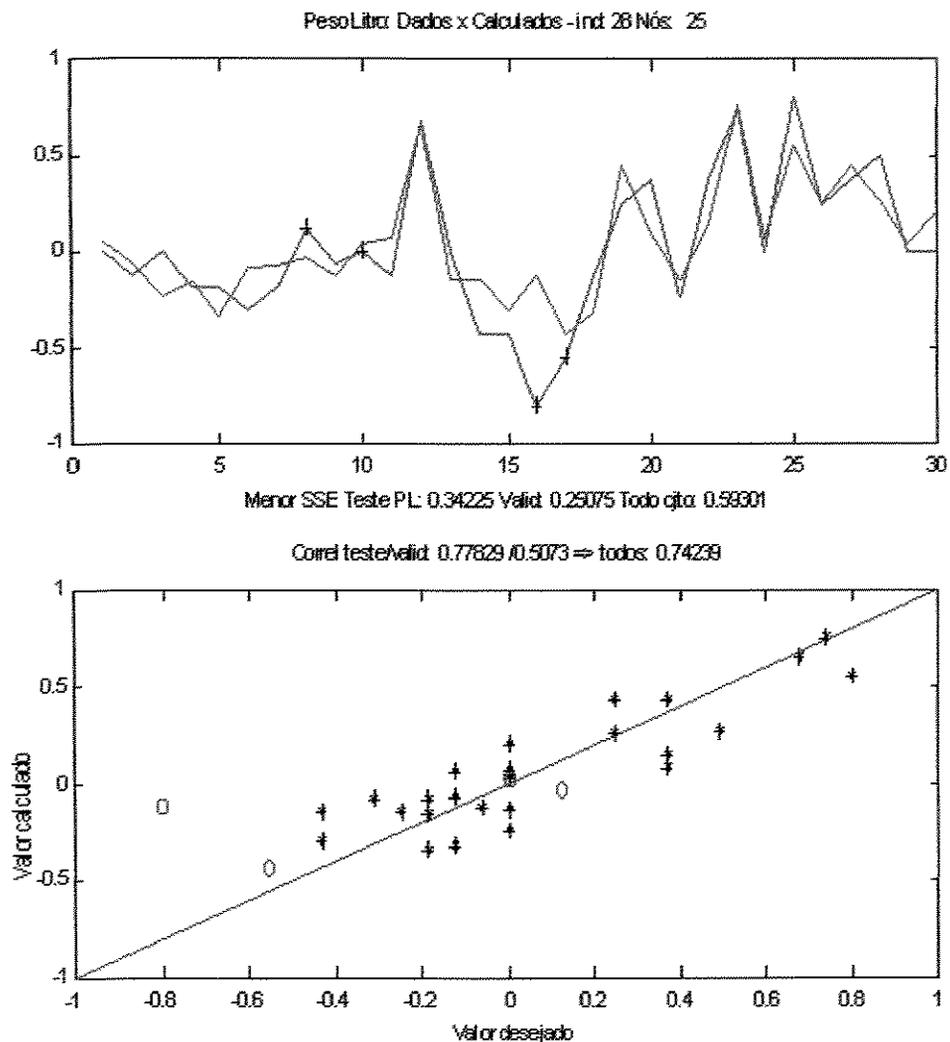


Figura 26 - Saída da rede reduzida para 28 pesos, predizendo Peso Litro

Da mesma forma, na Figura 26 a curva em vermelho representa os pontos experimentais de Peso Litro, sobre a qual está marcado o ponto experimental (+) utilizado para validação da rede, e a curva em azul mostra o desvio dos dados preditos em relação aos dados experimentais. Na parte inferior o gráfico, representando a correspondência entre os dados experimentais e os dados preditos, mostra que os dados utilizados para treinamento (*) não apresentaram boa concordância, assim como os dados sorteados, utilizados para validação da rede (o), também não.

Observando na Figura 24 que a rede reduzida para 28 pesos só utiliza onze nós da camada intermediária no modelo, decidimos trabalhar com uma nova rede, com onze nós intermediários, mas mantendo a saída com dois nós. Quanto à importância das entradas, deixamos para analisá-las somente depois da definição da melhor estrutura da rede.

Treinamento da Rede: 11 nós intermediários, 2 saídas

Depois da aplicação da mesma metodologia para treinamento e realização dos mesmos testes que foram feitos para a rede com 25 nós na camada intermediária e 2 nós na camada de saída, observamos que, na Figura 27, após a execução da técnica OBS para redução de pesos, ainda restaram 2 pesos sem conexão na camada intermediária. Apesar dos resultados de predição terem sido razoáveis ($R^2=0,87$), mas não da validação ($R^2=0,07$), efetuada com 4 pontos aleatórios não utilizados no treinamento, decidimos reduzir o número de nós da camada intermediária para 9 nós, por causa dos dois nós sem conexão, mantendo os dois nós da camada de saída, e então retreinar a rede.

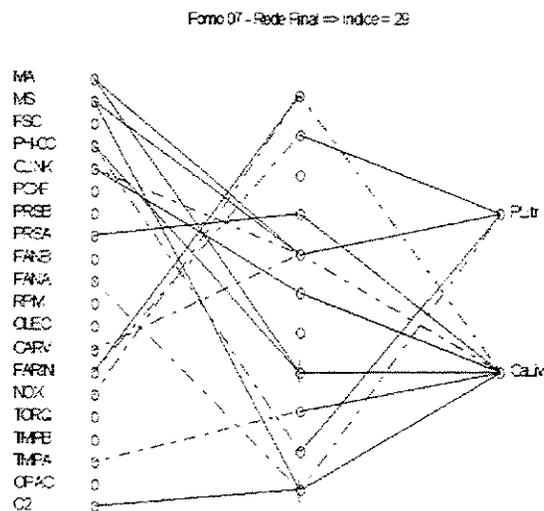


Figura 27 - Rede com 11 nós intermediários e 2 nós de saída

Treinamento da Rede: 9 nós intermediários, 2 saídas

Esta rede ficou com a seguinte estrutura inicial, com os nós totalmente conectados:

Forno 07 - Rede Totalmente Conectada => Nós hidden: 9

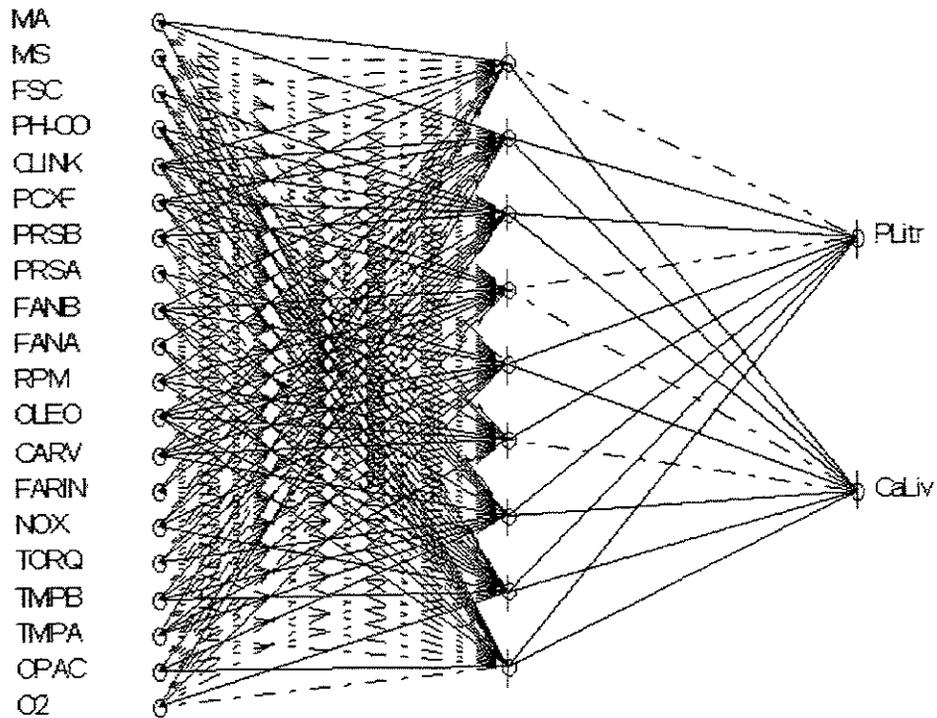


Figura 28 - Rede inicial com 9 nós na camada intermediária e 2 na camada de saída

Os procedimentos utilizados para o treinamento e teste desta rede foram os mesmos, mantendo-se também o número de dados aleatórios utilizados para a validação da rede em quatro, ou seja, cerca de 13 % dos dados.

Essa a rede apresentou os seguintes resultados para a predição da Cal Livre:

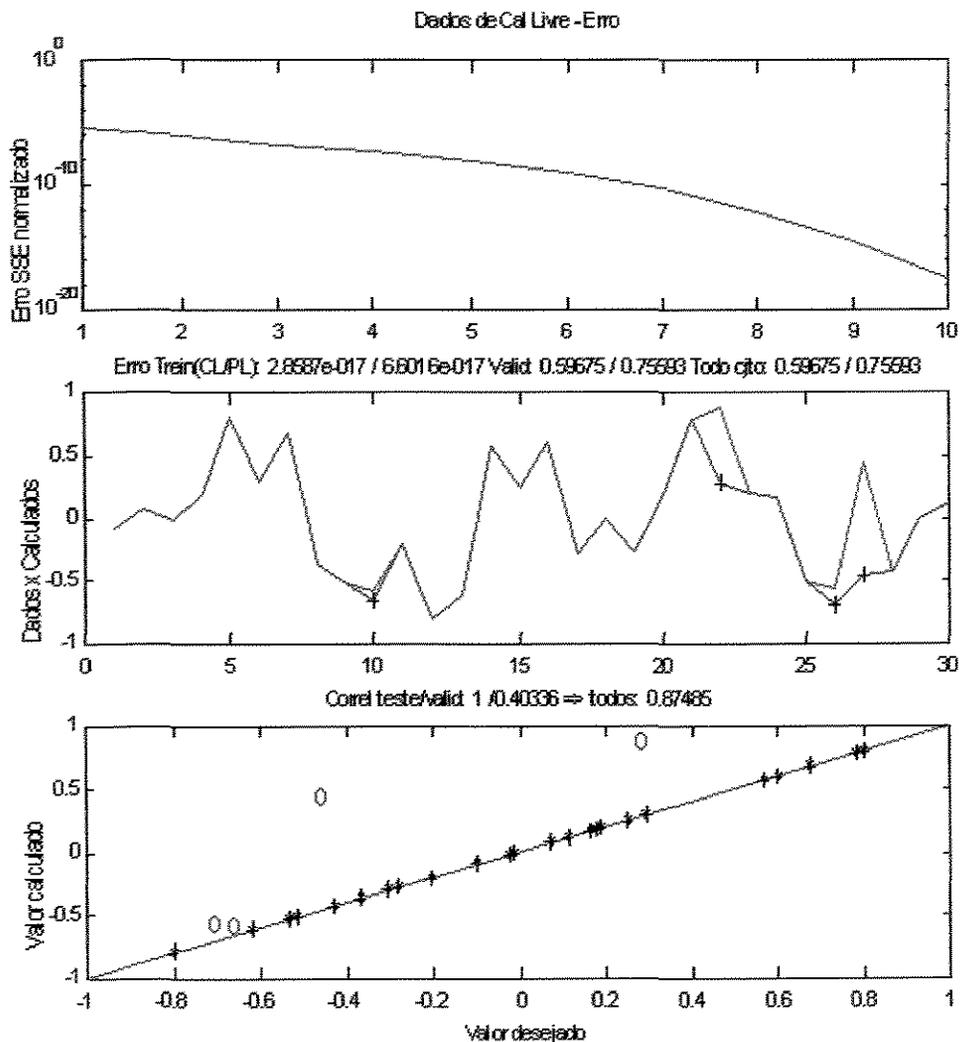


Figura 29 - Saída da rede com 9 nós intermediários predizendo Cal Livre

Na Figura 29 a curva em vermelho representa os pontos experimentais de Cal Livre, sobre a qual está marcado o ponto experimental (+) utilizado para validação da rede, e a curva em azul mostra o desvio dos dados preditos em relação aos dados experimentais. Na parte inferior o gráfico, representando a correspondência entre os dados experimentais e os dados preditos, mostra os dados utilizados para treinamento (*) que apresentaram boa concordância, embora os dados sorteados, utilizados para validação da rede (o) não ($R^2=0,4$).

Os resultados para a predição do Peso Litro foram:

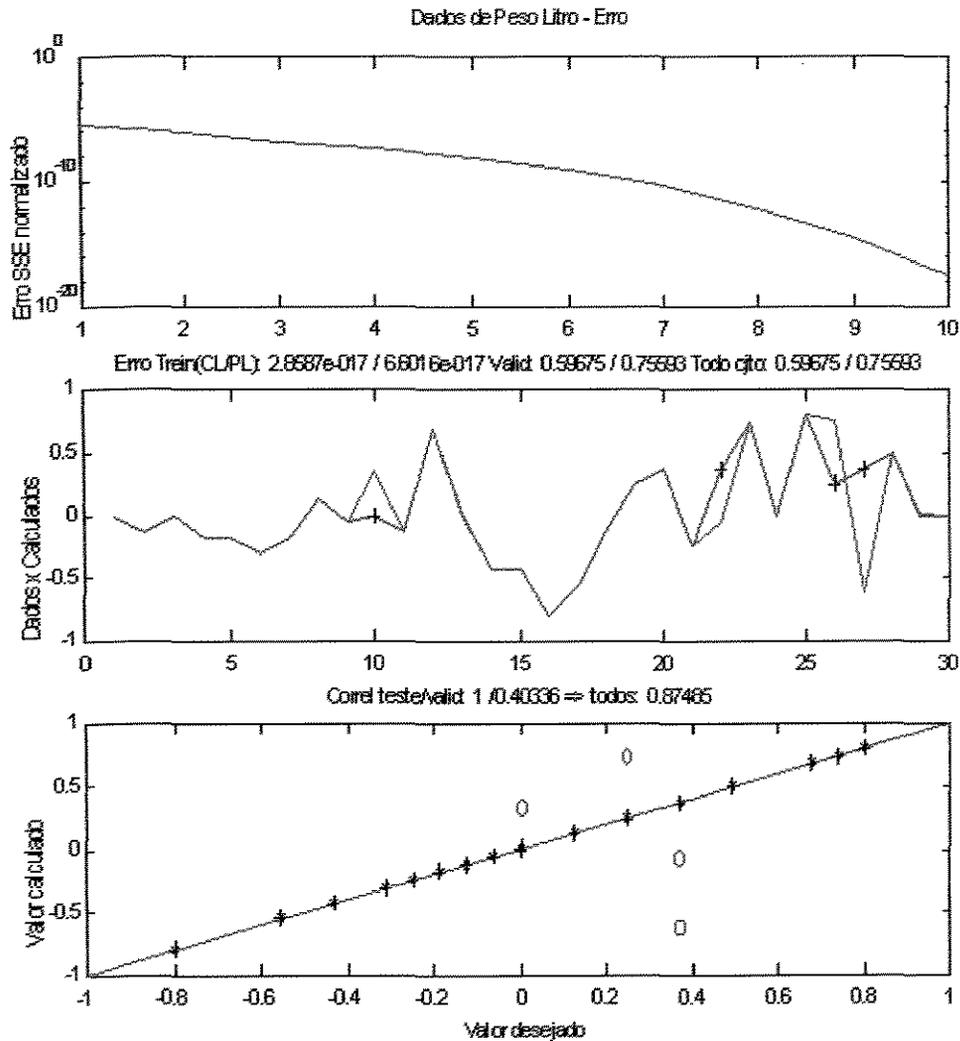


Figura 30 - Saída da rede com 9 nós intermediários predizendo Peso Litro

Da mesma forma, na Figura 30 a curva em vermelho representa os pontos experimentais de Peso Litro, sobre a qual está marcado o ponto experimental (+) utilizado para validação da rede, e a curva em azul mostra o desvio dos dados preditos em relação aos dados experimentais, e na parte inferior o gráfico, representando a correspondência entre os dados experimentais e os dados preditos, mostra que os dados utilizados para

treinamento (*) apresentaram boa concordância, embora os dados sorteados, utilizados para validação da rede (o) não ($R^2=0,4$).

O índice de correlação, para todos os dados, ficou em $R^2=0,87$, e, considerando somente os dados de validação em $R^2=0,40$.

Aplicando a metodologia do OBS, reduzimos os pesos ($20*9 + 9*2$) de 198 para 81, ficando com a configuração apresentada na Figura 31, a seguir.

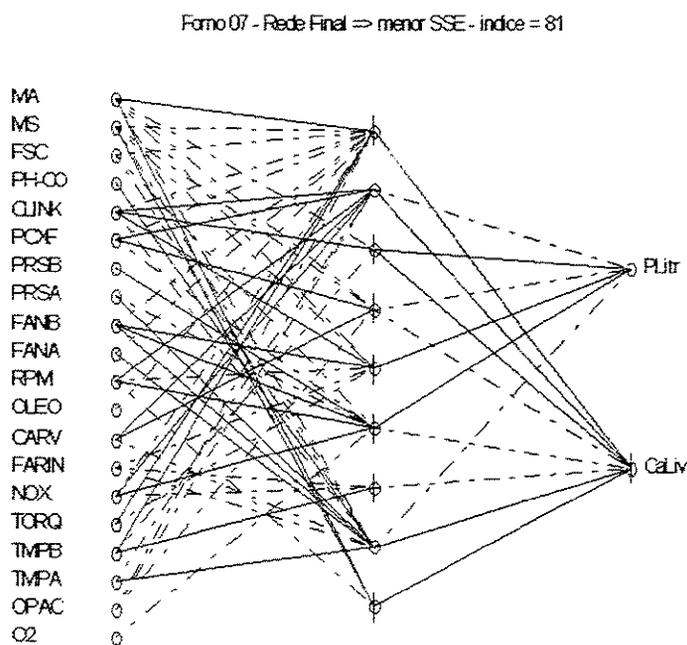


Figura 31 - Rede com 9 nós intermediários e pesos reduzidos para 81

Nessa rede otimizada percebemos que todos os nós, tanto da camada de entrada, como da camada intermediária, foram utilizados, indicando que, nesse nível de estrutura, todas os nós foram considerados necessários para a modelagem da Cal Livre e do Peso Litro.

E os resultados preditos por ela, usando quatro dados para validação, para a Cal Livre foram:

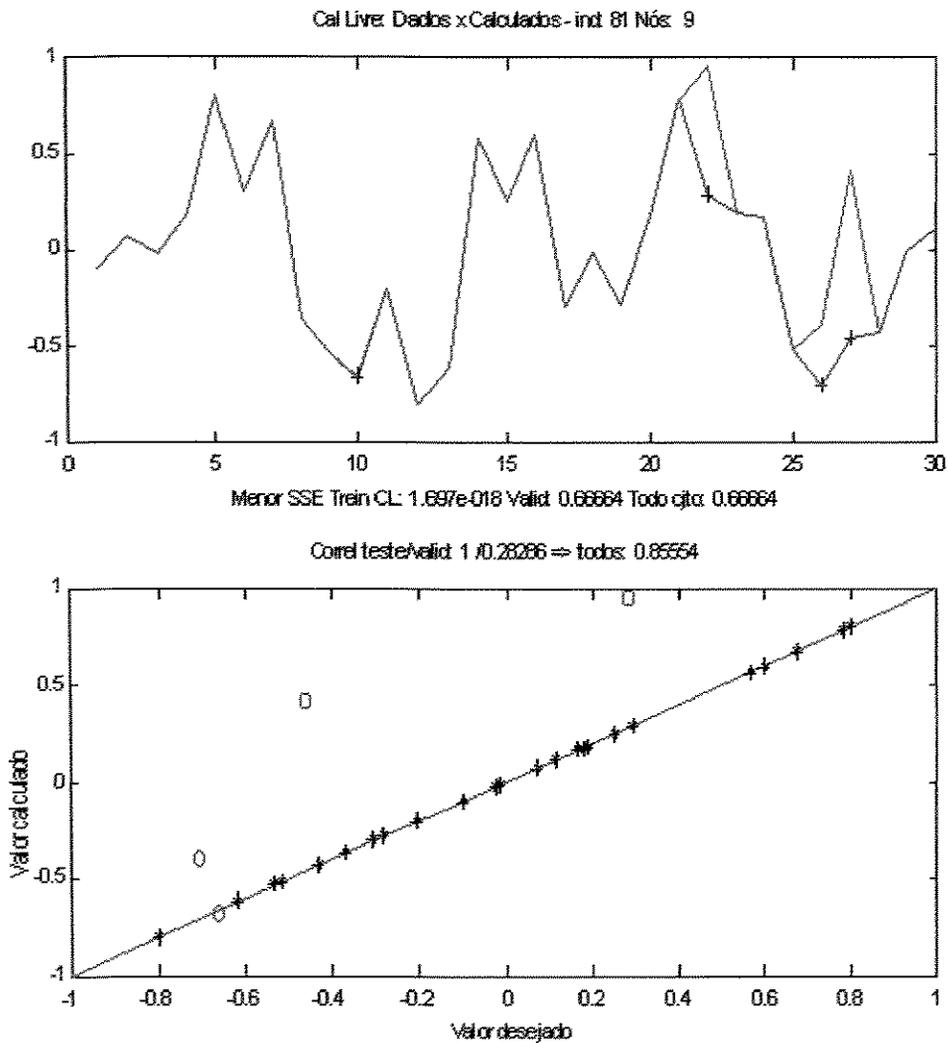


Figura 32 - Saída da rede com 9 nós intermediários e 81 pesos, predizendo Cal Livre

Na Figura 32 a curva em vermelho representa os pontos experimentais de Cal Livre, sobre a qual está marcado o ponto experimental (+) utilizado para validação da rede, e a curva em azul mostra o desvio dos dados preditos em relação aos dados experimentais. Na parte inferior o gráfico, representando a correspondência entre os dados experimentais e os dados preditos, mostra os dados utilizados para treinamento (*) que apresentaram boa concordância, embora os dados sorteados, utilizados para validação da rede (o) não ($R^2=0,26$).

E, para o Peso Litro os resultados foram:

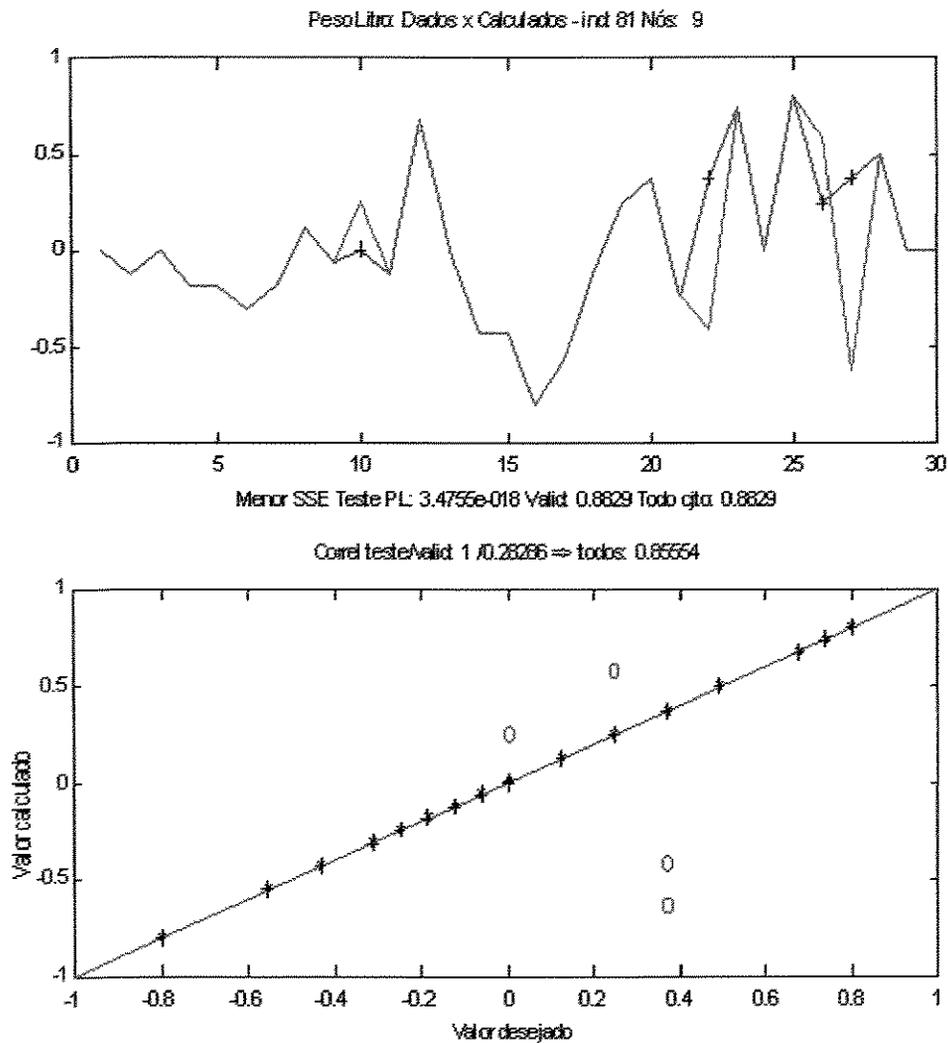


Figura 33 - Saída da rede com 9 nós intermediários e 81 pesos, predizendo Peso Litro

Na Figura 33, da mesma forma, a curva em vermelho representa os pontos experimentais de Cal Livre, e a curva em azul mostra o desvio dos dados preditos em relação aos dados experimentais, e, na parte inferior, o gráfico representando a correspondência entre os dados experimentais e os dados preditos.

O índice de correlação para todos os dados foi 0,86, e para o conjunto de validação 0,28.

Comparativamente com as redes iniciais, apresentadas na parte superior da Figura 34, praticamente não houve diferença. Alguns pontos melhoraram, outros pioraram, como pode ser visto a seguir.

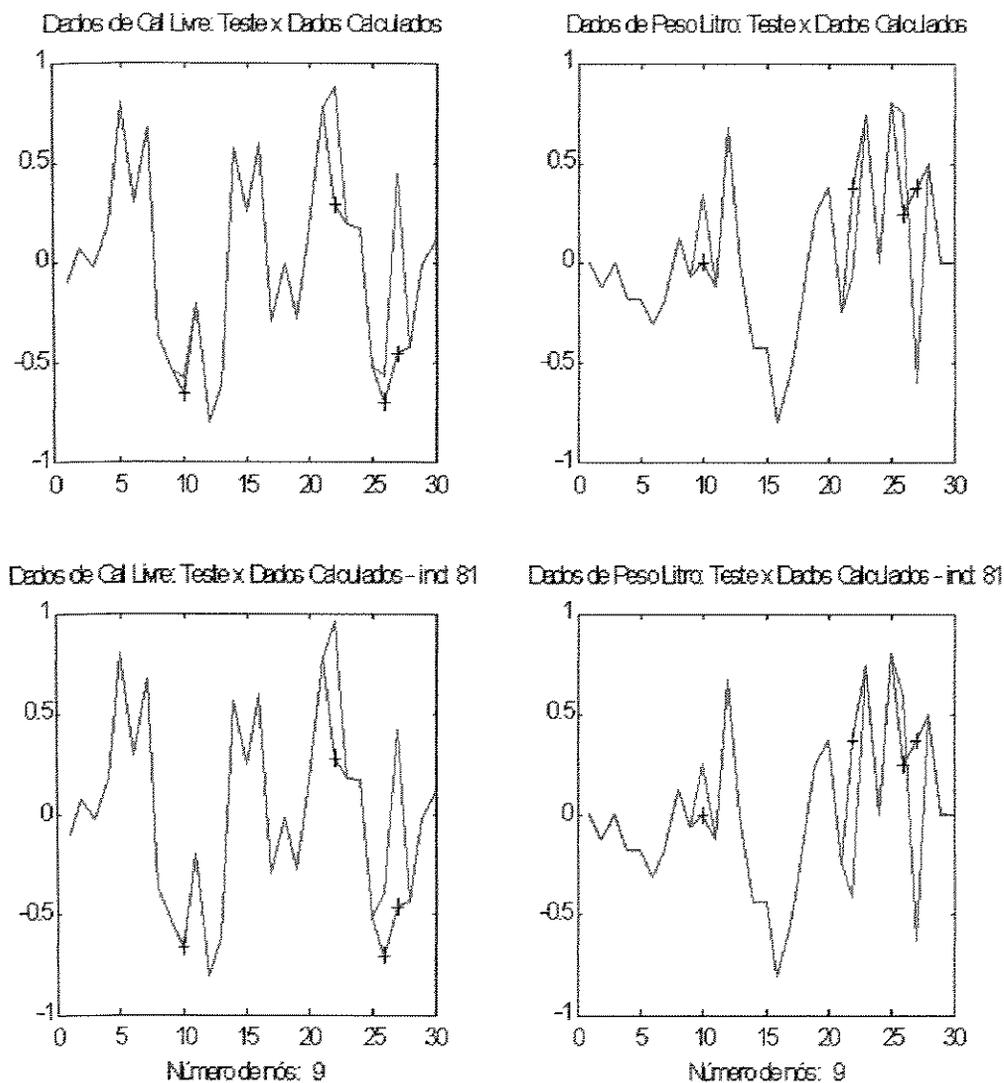


Figura 34 - Comparação entre saídas da rede de 9 nós intermediários, inicial e com pesos reduzidos para 81

Foram analisadas outras estruturas dessa rede, também com pesos reduzidos, como por exemplo, com 105 pesos, mas o resultado não foi significativamente diferente.

Número ideal de pesos da rede

Como vimos no capítulo anterior, um número excessivo de pesos pode guardar informações redundantes, levando a rede a ficar sobre treinada. Dessa consideração podemos concluir que o número de pesos não pode exceder muito o número de casos a serem apresentados para treinamento, sob o risco da rede decorá-los.

Considerando que temos vinte nós de entrada, uma maneira de diminuirmos os pesos é reduzindo o número de nós de saída. Assim, a rede, ao invés de predizer Cal Livre e Peso Litro concomitantemente, será dividida em duas, com as mesmas entradas para predizerem ou a Cal Livre, ou o Peso Litro.

Fizemos vários testes considerando apenas um nó na camada intermediária, que seria a condição com o menor número de nós na rede, mas, a rede mostrou-se sub treinada, não conseguindo mapear adequadamente os dados de saída.

Treinamento da Rede: 2 nós intermediários, 1 saída

Ao colocarmos 2 nós na camada intermediária teremos quarenta e dois pesos (sem os *bias*), que é maior que o número de casos disponíveis para treinamento (trinta casos). Considerando que poderemos reduzi-los através da técnica de OBS, definimos o número de casos para validação em cinco dados aleatorizados, restando vinte e cinco casos para treinamento da rede. Modificamos, também os limites de normalização, deixando, tanto os dados de entrada, como os de saída, normalizados entre $-0,9$ e $+0,9$.

A rede ficou com a configuração apresentada a seguir:

Forno 07 - Rede Totalmente Conectada => Nós hidden: 2 - Total: 45

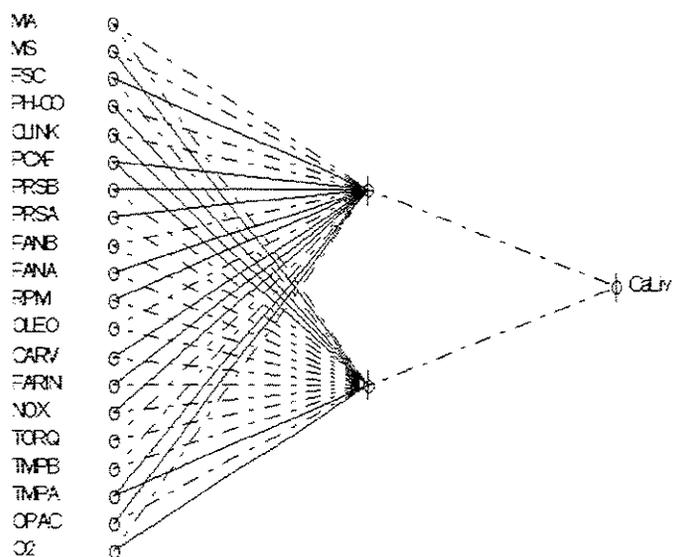


Figura 35 – Estrutura da rede com 2 nós intermediários, predizendo somente a Cal Livre

Forno 07 - Rede Totalmente Conectada => Nós hidden: 2 - Total: 45

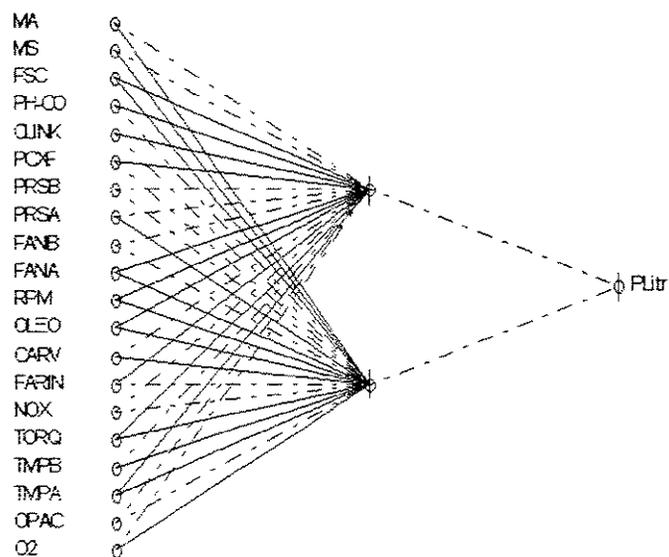


Figura 36 – Estrutura da rede com 2 nós intermediários, predizendo somente o Peso Litro

Estas estruturas de rede também foram treinadas e testadas conforme os procedimentos anteriores.

A rede predizendo a Cal Livre, conforme pode ser visto na Figura 37, apresentou um ótimo índice de correlação ($R^2=0,94$) considerando todos os dados experimentais, porém, ao testarmos somente com os dados sorteados para validação esse índice caiu bastante ($R^2=0,47$), ou seja, não teve boa capacidade de generalização. A curva em vermelho representa os pontos experimentais. Sobre ela estão marcados os pontos escolhidos para validação (+), e a curva azul representa o desvio dos valores preditos em relação aos valores experimentais. Na parte inferior da figura está a correlação entre os dados experimentais e os dados preditos (*), onde a concordância está na diagonal, e, os pontos selecionados para validação são aqueles marcados em vermelho (o).

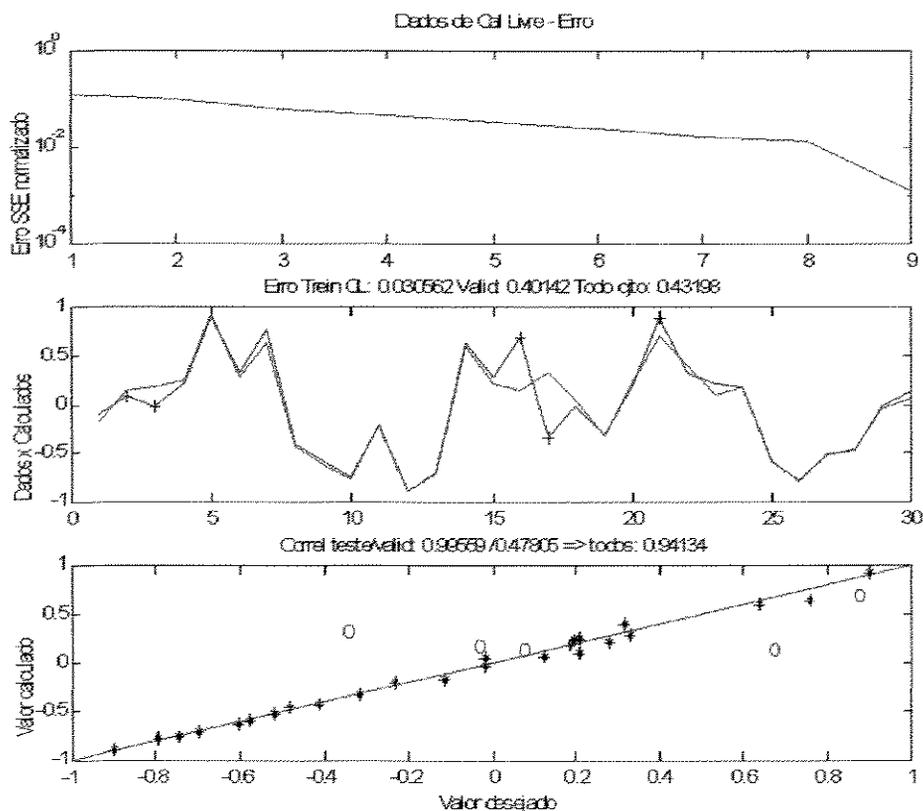


Figura 37 - Resultado inicial da rede com 2 nós intermediários e uma saída, predizendo a Cal Livre

A rede predizendo o Peso Litro, conforme pode ser visto na Figura 38, apresentou um índice de correlação ($R^2=0,86$) razoável, considerando todos os dados experimentais, e, ao testarmos somente com os dados sorteados para validação esse índice ficou em ($R^2=0,66$). O modelo não representou tão bem todos os dados experimentais quanto a rede para Cal Livre, porém generalizou melhor. A curva em vermelho representa os pontos experimentais. Sobre ela estão marcados os pontos escolhidos para validação (+), e a curva azul representa o desvio dos valores preditos em relação aos valores experimentais. Na parte inferior da figura está a correlação entre os dados experimentais e os dados preditos (*), onde a concordância está na diagonal, e, os pontos selecionados para validação são aqueles marcados em vermelho (o).

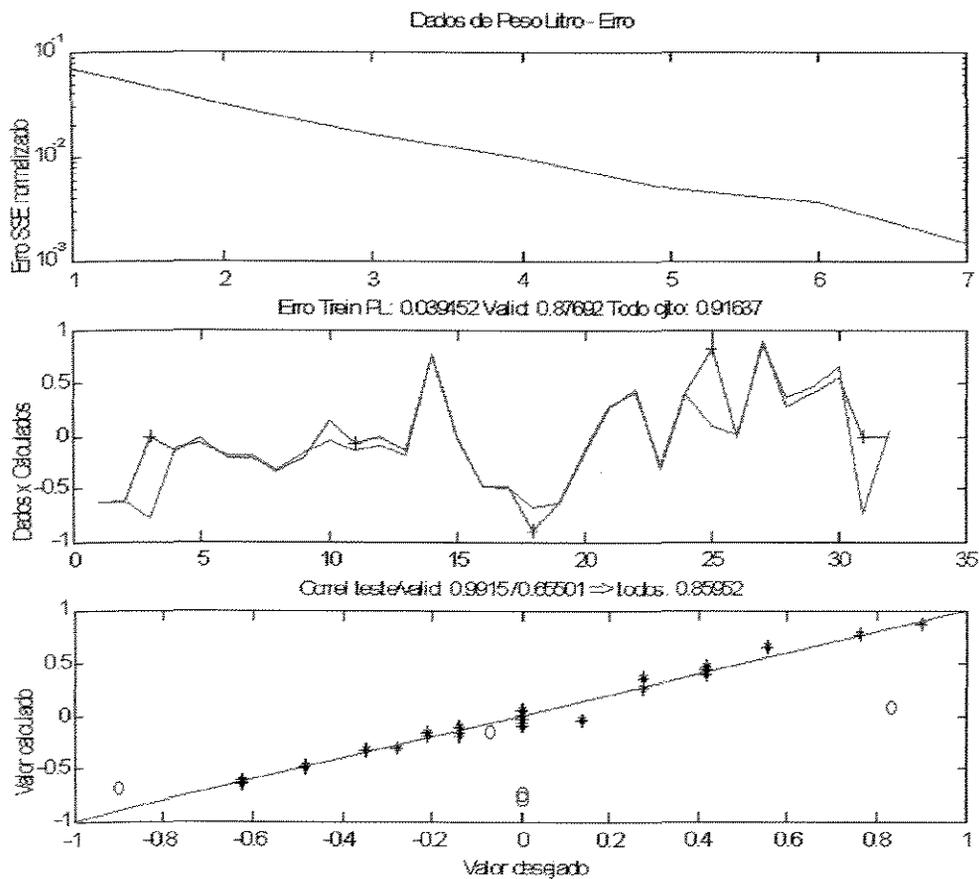


Figura 38 – Resultado inicial da rede com 2 nós intermediários e uma saída, predizendo o Peso Litro

Esses resultados são melhores do que aqueles obtidos para a rede predizendo simultaneamente as duas propriedades do clínquer.

Aplicando a técnica de OBS, obtivemos uma rede para predição de Cal Livre reduzida para 38 pesos.

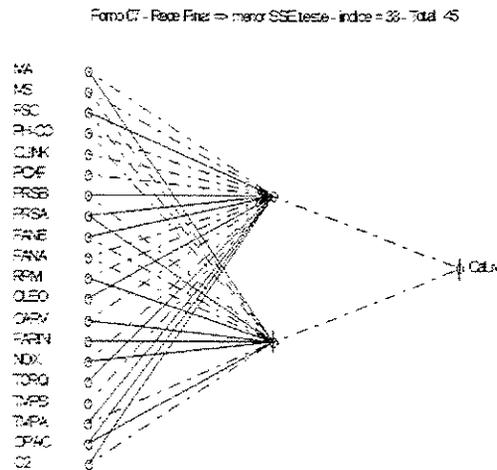


Figura 39 - Rede com 2 nós intermediários e uma saída, após redução de pesos, predizendo Cal Livre

E, aplicando a técnica de OBS para a rede predizendo Peso Litro, obtivemos uma rede com os pesos reduzidos para 31.

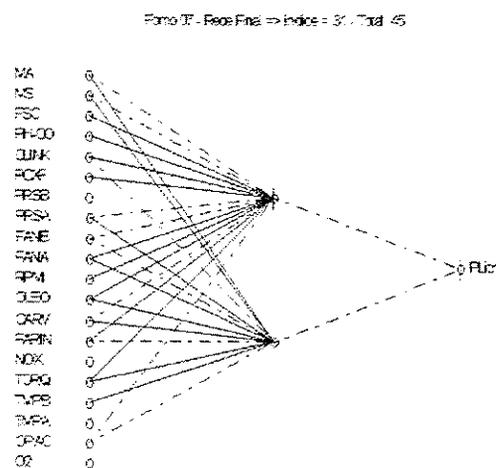


Figura 40 - Rede com 2 nós intermediários e uma saída, após redução de pesos, predizendo Peso Litro

Observamos que a rede para predição de Peso Litro desconsiderou quatro nós de entrada: teor de oxigênio residual dos gases na saída da torre de ciclone (O2), temperatura de saída dos gases da torre de ciclone no ramal A (TMPA), teor de NOx nos gases de combustão (NOX) e Queda de pressão no Ramal B (PRSB).

Os resultados de Cal Livre, preditos pela rede com os pesos reduzidos, foram:

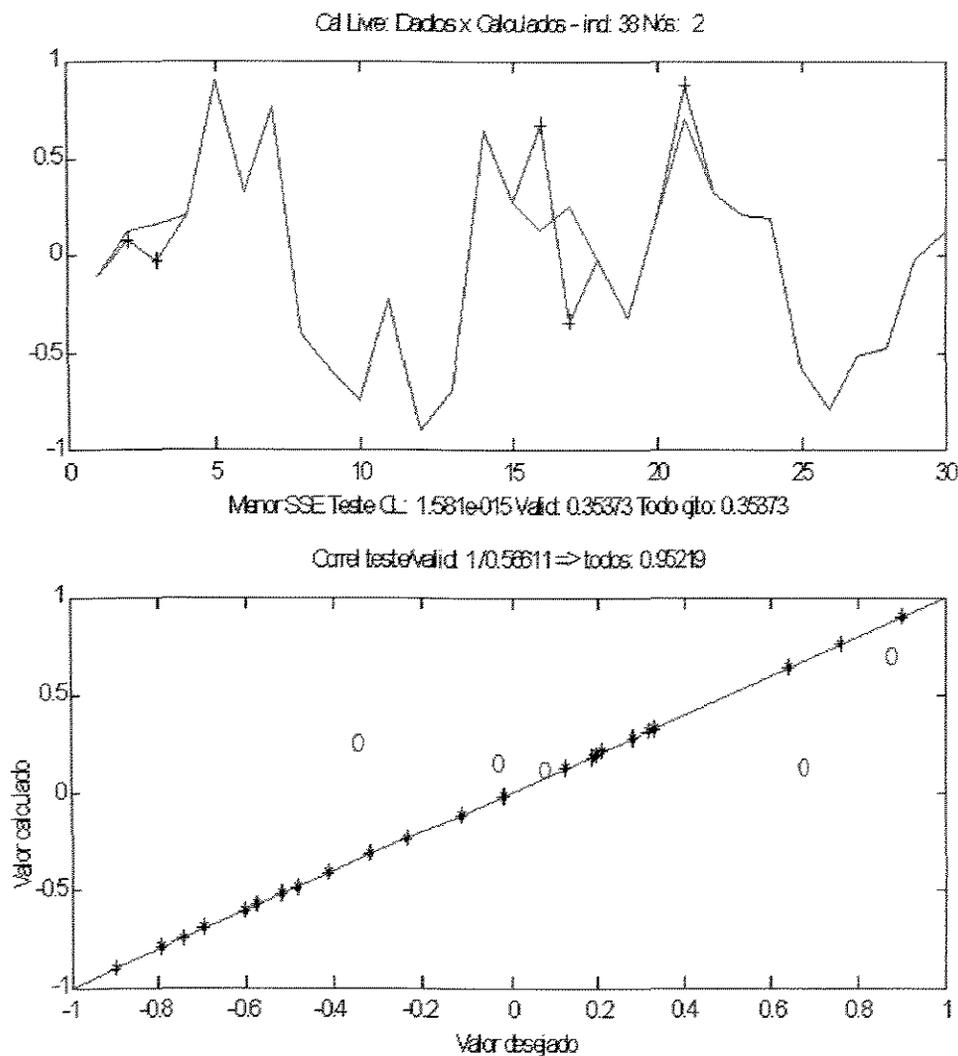


Figura 41 - Saídas da rede com 2 nós intermediários e 1 saída, predizendo Cal Livre, após redução de pesos

Com a redução de pesos, o resultado das redes melhorou. A rede predizendo a Cal Livre, conforme pode ser visto na Figura 41, apresentou um ótimo índice de correlação

($R^2=0,95$) considerando todos os dados experimentais, melhor que a rede totalmente conectada. Da mesma forma, ao testarmos somente com os dados sorteados para validação esse índice também melhorou ($R^2=0,57$), embora ainda sua capacidade de generalização não seja muito boa. A curva em vermelho representa os pontos experimentais. Sobre ela estão marcados os pontos escolhidos para validação (+), e a curva azul representa o desvio dos valores preditos em relação aos valores experimentais. Na parte inferior da figura está a correlação entre os dados experimentais e os dados preditos (*), onde a concordância está na diagonal, e, os pontos selecionados para validação são aqueles marcados em vermelho (o).

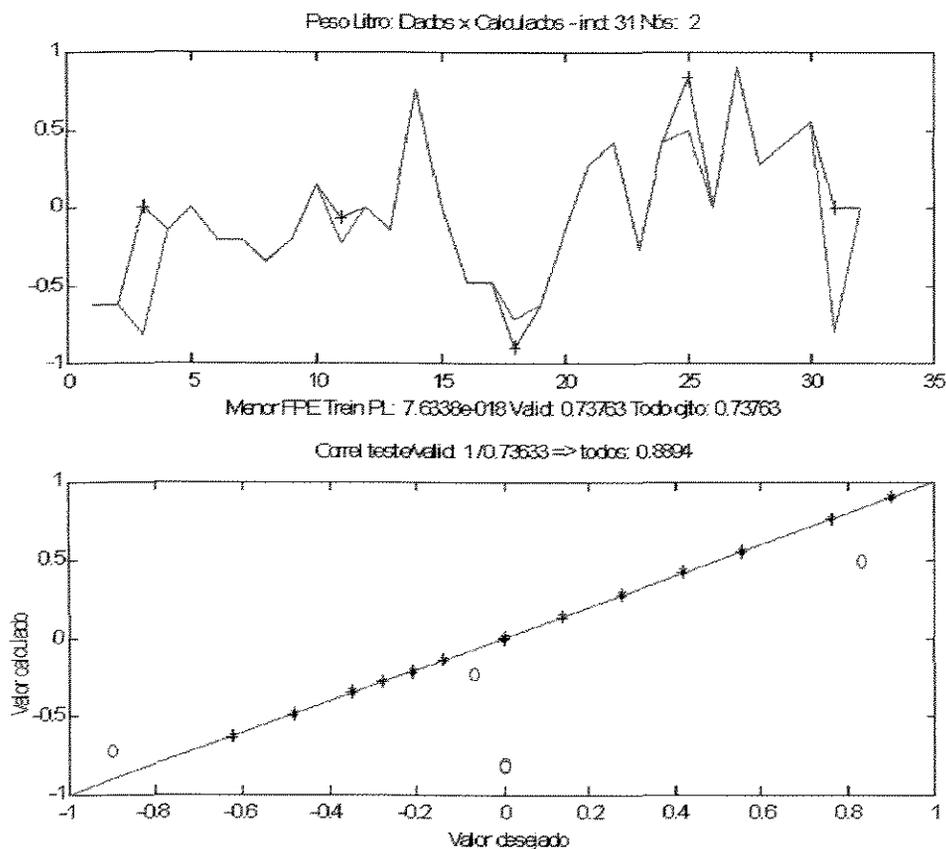


Figura 42 - Saídas da rede com 2 nós intermediários e 1 saída, predizendo Peso Litro, após redução de pesos

A melhoria pode ser vista, também, na rede predizendo o Peso Litro, Figura 42, pois o índice de correlação ($R^2=0,89$) melhorou com a redução de pesos, para todos os dados experimentais, assim como para os dados de validação ($R^2=0,74$), mostrando uma

boa generalização. A curva em vermelho representa os pontos experimentais. Sobre ela estão marcados os pontos escolhidos para validação (+), e a curva azul representa o desvio dos valores preditos em relação aos valores experimentais. Na parte inferior da figura está a correlação entre os dados experimentais e os dados preditos (*), onde a concordância está na diagonal, e, os pontos selecionados para validação são aqueles marcados em vermelho (o).

Comparando esses resultados obtidos das redes com redução de pesos, em relação às redes iniciais com dois nós na camada intermediária totalmente interligada, mostrada na parte superior da Figura 43, temos para a Cal Livre:

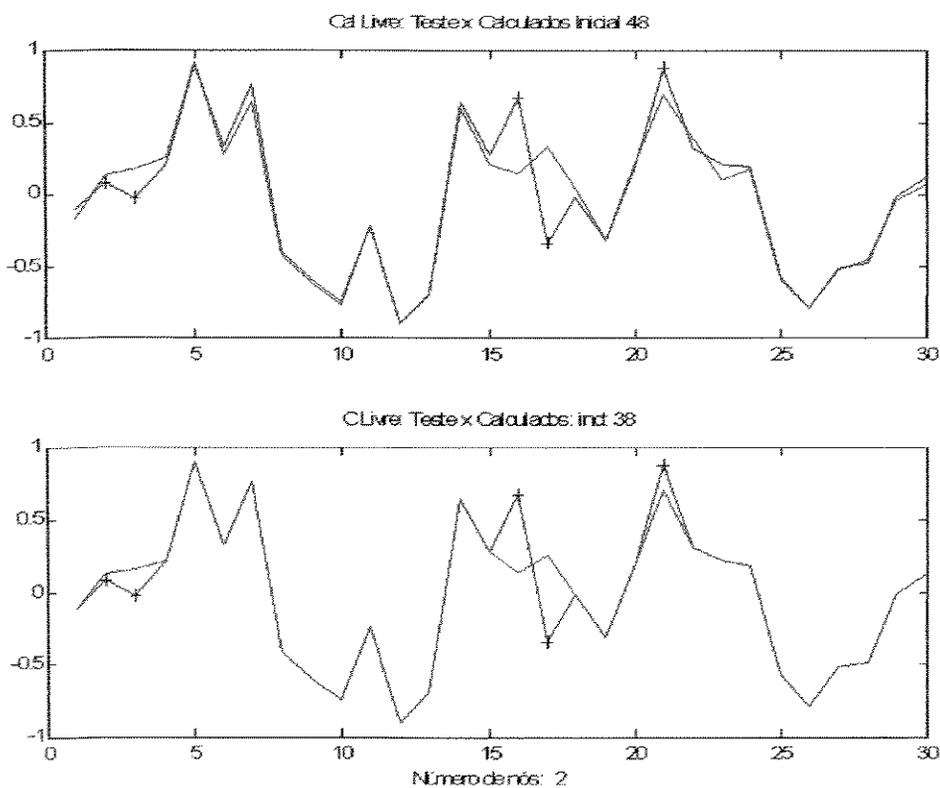


Figura 43 - - Comparação dos resultados da rede inicial, com dois nós na camada intermediária, e a rede depois da redução de pesos, predizendo Cal Livre.

Para o Peso Litro, comparamos os resultados obtidos das redes com redução de pesos, em relação às redes iniciais, com dois nós na camada intermediária, mostrada na parte superior da Figura 44:

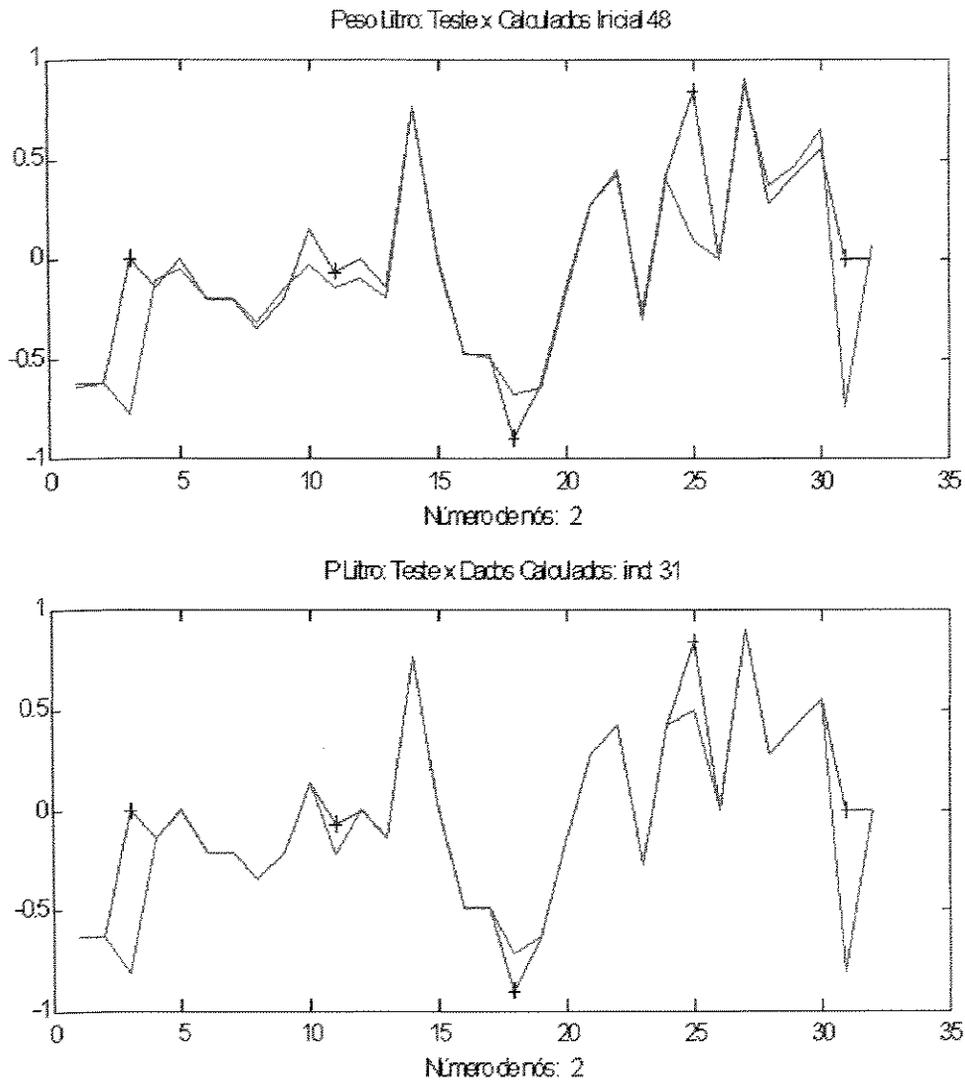


Figura 44 - Comparação dos resultados da rede inicial, com dois nós na camada intermediária, e a rede depois da redução de pesos, predizendo Peso Litro.

As redes com redução de pesos apresentaram melhor predição dos dados experimentais que as redes totalmente conectadas.

Análise de Sensibilidade dos dados de entrada

A análise de sensibilidade dos dados de entrada, para a rede totalmente conectada, foi efetuada através da aplicação da equação proposta por Garson (1991), que também foi utilizada por Elgibaly & Elkamel (1998), e está representada na Equação 20. Esta metodologia considera o valor absoluto dos pesos. Fizemos, também, a aplicação dessa mesma equação, considerando, entretanto, o sinal dos pesos, conforme nossas considerações no item “Análise de Significância das Variáveis de Entrada” do capítulo anterior.

Para a rede, após a otimização dos pesos, analisamos adicionalmente a saída da rede, relacionando o resultado obtido a partir dos valores médios de todas as entradas, que seria o caso base, com as saídas geradas a partir da modificação de cada entrada, acrescida de 20%. O valor percentual de cada alteração foi tabelado.

Análise de Sensibilidade: rede inicial, com dois nós intermediários, predizendo Cal Livre

Devido ao grande número de entradas, e, para visualizarmos melhor os resultados da análise da rede inicial, eles foram divididos e apresentados em duas tabelas. A primeira parte refere-se aos resultados das entradas mais importantes, como pode ser visto na Tabela 2 e a segunda parte das entradas menos importantes, mostrada na Tabela 3.

Tabela 2 - Análise de Sensibilidade da Rede Inicial: Entradas mais importantes para Cal Livre

Rede Inicial					
<i>Fórmula de Garson</i>			<i>Fórmula Modificada</i>		
Entrada		%	Entrada		%
Número	Nome		Número	Nome	
2	OPAC	14,0	2	OPAC	14,6
16	CLINK	11,1	17	PH-CO	12,1
17	PH-CO	9,7	16	CLINK	10,8
10	RPM	9,3	7	FARIN	8,5
20	MA	8,6	9	OLEO	7,6
1	O2	7,3	11	FANA	5,9
9	OLEO	7,2	10	RPM	5,4
18	FSC	5,2	1	O2	4,7
5	TORQ	4,3	13	PRSA	4,7
4	TMPB	4,2	5	TORQ	4,5
13	PRSA	3,3	18	FSC	4,2
11	FANA	3,1	20	MA	4,2
7	FARIN	3,0	4	TMPB	3,5

Observamos, na Tabela 2, que as três entradas consideradas mais importantes coincidiram para os dois métodos, com pequena variação no valor percentual calculado. A partir daí, os valores mudaram significativamente, impedindo uma conclusão sobre esses resultados.

Tabela 3 - Análise de Sensibilidade da Rede Inicial: Entradas menos importantes para Cal Livre

Rede Inicial					
<i>Fórmula de Garson</i>			<i>Fórmula Modificada</i>		
Entrada		%	Entrada		%
Número	Nome		Número	Nome	
12	FANB	2,4	19	MS	2,3
15	PCXF	1,7	6	NOX	2,2
8	CARV	1,6	15	PCXF	1,9
19	MS	1,4	3	TMPA	1,0
6	NOX	1,2	8	CARV	0,8
14	PRSB	0,7	12	FANB	0,8
3	TMPA	0,6	14	PRSB	0,2

Podemos observar na Tabela 3 que os resultados obtidos são similares, ou seja, todas elas são entradas consideradas menos importantes, para predição da Cal Livre, por ambas metodologias.

Análise de Sensibilidade: rede inicial, com dois nós intermediários, predizendo Peso Litro

Assim como para a Cal Livre, os resultados da análise da rede inicial foram divididos em duas partes. As entradas mais significativas são apresentadas na Tabela 4, e as menos significativas na Tabela 5.

Tabela 4 - Análise de Sensibilidade da Rede Inicial: Entradas mais importantes para Peso Litro

Rede Inicial					
Fórmula de Garson			Fórmula Modificada		
Entrada		%	Entrada		%
Número	Nome		Número	Nome	
12	FANB	13,5	12	FANB	16,0
2	OPAC	12,1	2	OPAC	12,1
15	PCXF	11,6	11	FANA	8,6
11	FANA	8,3	5	TORQ	8,4
5	TORQ	7,9	15	PCXF	8,1
20	MA	6,9	7	FARIN	5,5
19	MS	5,8	13	PRSA	4,7
9	OLEO	5,0	18	FSC	4,3
17	PH-CO	4,8	9	OLEO	4,1
18	FSC	4,7	4	TMPB	3,8
3	TMPA	4,4	20	MA	3,8
10	RPM	3,3	3	TMPA	3,6
16	CLINK	2,7	8	CARV	3,3
6	NOX	1,7	10	RPM	2,7
7	FARIN	1,6	17	PH-CO	2,7

Observamos, na Tabela 4, que as cinco entradas consideradas mais importantes, para a predição do Peso Litro, coincidiram para os dois métodos, com pequena variação no valor percentual calculado. A partir daí, os valores mudaram significativamente, impedindo uma conclusão sobre esses resultados.

Tabela 5 - Análise de Sensibilidade da Rede Inicial: Entradas menos importantes para Peso Litro

Rede Inicial					
<i>Fórmula de Garson</i>			<i>Fórmula Modificada</i>		
Entrada		%	Entrada		%
Número	Nome		Número	Nome	
14	PRSB	1,5	6	NOX	2,4
8	CARV	1,4	19	MS	2,1
1	O2	1,2	14	PRSB	2,0
13	PRSA	0,9	16	CLINK	1,5
4	TMPB	0,6	1	O2	0,3

Podemos observar na Tabela 5 que, exceto para o teor de oxigênio residual dos gases na saída da torre de ciclone (O2) e para a Queda de pressão no Ramal B (PRSB), onde ambas metodologias geraram resultados similares, ou seja, ambas são entradas pouco significativas para a predição do Peso Litro. Os demais valores obtidos não permitem uma conclusão.

Análise de Sensibilidade: rede com redução dos pesos, predizendo Cal Livre

Nesta análise consideramos também os resultados obtidos comparando a predição calculada com os valores médios das entradas, e a variação da saída para alterações de 20% em cada entrada. Para melhor visualizarmos os resultados dessa análise, eles também foram divididos em duas partes. A primeira parte, apresentada na Tabela 6, refere-se aos dados das entradas mais importantes, e a segunda parte, apresentada na Tabela 7, aos resultados das entradas menos importantes.

Tabela 6 - Análise de Sensibilidade da Rede com pesos otimizados: Entradas mais importantes para Cal Livre

Rede Otimizada - 38 pesos								
<i>Fórmula de Garson</i>			<i>Fórmula Modificada</i>			<i>Média com 20% acréscimo</i>		
Entrada		%	Entrada		%	Entrada		%
Número	Nome		Número	Nome		Número	Nome	
2	OPAC	13,9	2	OPAC	13,9	2	OPAC	9,4
16	CLINK	11,2	17	PH-CO	10,6	7	FARIN	6,8
10	RPM	9,7	16	CLINK	10,4	11	FANA	5,9
17	PH-CO	9,3	7	FARIN	10,1	17	PH-CO	5,6
20	MA	8,5	11	FANA	6,8	16	CLINK	1,7
1	O2	7,5	9	OLEO	6,8	4	TMPB	1,7
9	OLEO	7,3	13	PRSA	5,9	18	FSC	1,6
18	FSC	5,3	10	RPM	4,6	5	TORQ	0,8
4	TMPB	4,5	4	TMPB	4,2	13	PRSA	0,7
5	TORQ	4,1	18	FSC	4,1	14	PRSB	0,5
13	PRSA	4,0	6	NOX	4,0	12	FANB	0,5
11	FANA	3,4	1	O2	3,9	9	OLEO	0,5
7	FARIN	2,9	5	TORQ	3,8	10	RPM	0,3

Observamos, na Tabela 6, que só uma entrada, a emissão de particulados na saída da chaminé (OPAC) foi escolhida como a mais importante, coincidentemente nos três métodos utilizados. Dois métodos, embora apresentando variação nos valores percentuais calculados, consideraram os mesmos itens como as próximas quatro mais importantes entradas: vazão de farinha (FARIN), teor de monóxido de carbono na saída da torre de ciclone (PH-CO), rotação do exaustor no ramal A (FANA), e temperatura do clínquer na saída do resfriador (CLINK), enquanto que a equação proposta por Garson (1992) exceto para o teor de monóxido de carbono na saída da torre de ciclone (PH-CO), divergiu dessas outras duas metodologias. Para os demais valores não é possível uma conclusão clara.

Tabela 7 - Análise de Sensibilidade da Rede com pesos otimizados: Entradas menos importantes para Cal Livre

Rede Otimizada - 38 pesos								
<i>Fórmula de Garson</i>			<i>Fórmula Modificada</i>			<i>Média com 20% acréscimo</i>		
Entrada		%	Entrada		%	Entrada		%
Número	Nome		Número	Nome		Número	Nome	
12	FANB	2,4	20	MA	2,8	8	CARV	0,3
8	CARV	1,5	19	MS	2,4	6	NOX	0,3
15	PCXF	1,3	8	CARV	2,0	20	MA	0,3
19	MS	1,1	3	TMPA	1,9	1	O2	0,2
6	NOX	0,8	15	PCXF	1,2	3	TMPA	0,1
3	TMPA	0,7	14	PRSB	0,4	15	PCXF	0,0
14	PRSB	0,5	12	FANB	0,2	19	MS	0,0

Todos os três métodos, mostrados na Tabela 7, definiram as quatro entradas, listadas a seguir, como pouco (ou nada) importantes: módulo de sílica (MS), perda de pressão na caixa de fumaça (PCXF), temperatura de saída dos gases da torre de ciclone no ramal A (TMPA) e vazão de carvão combustível (CARV). Quanto às demais, não se pode concluir nada.

Análise de Sensibilidade: rede com redução dos pesos, predizendo Peso Litro

Da mesma forma, para a análise de sensibilidade das entradas da rede treinada para predição de Peso Litro, também foram considerados os resultados obtidos comparando a predição calculada com os valores médios das entradas, e a variação da saída para alterações de 20% em cada entrada. Os resultados dessa análise foram divididos em duas partes. A primeira parte, apresentada na Tabela 8, refere-se às entradas mais importantes, e a segunda parte, na Tabela 9, às entradas menos importantes.

Tabela 8 - Análise de Sensibilidade da Rede com pesos otimizados: Entradas mais importantes para Peso Litro

Rede Otimizada - 31 pesos								
<i>Fórmula de Garson</i>			<i>Fórmula Modificada</i>			<i>Média com 20% acréscimo</i>		
Entrada		%	Entrada		%	Entrada		%
Número	Nome		Número	Nome		Número	Nome	
15	PCXF	15,4	15	PCXF	15,5	2	OPAC	37,4
2	OPAC	15,3	2	OPAC	14,9	12	FANB	22,3
11	FANA	12,3	11	FANA	12,4	11	FANA	10,4
12	FANB	9,9	12	FANB	9,6	7	FARIN	9,6
20	MA	6,1	20	MA	6,2	16	CLINK	4,4
17	PH-CO	5,8	17	PH-CO	5,8	4	TMPB	4,2
19	MS	5,6	19	MS	5,8	5	TORQ	3,7
5	TORQ	5,3	13	PRSA	5,4	17	PH-CO	3,2
13	PRSA	5,1	5	TORQ	5,1	20	MA	1,8
9	OLEO	4,9	9	OLEO	4,9	8	CARV	1,2
8	CARV	4,2	8	CARV	4,6	9	OLEO	1,2
4	TMPB	3,6	4	TMPB	3,3	10	RPM	0,8
10	RPM	3,2	10	RPM	3,2	13	PRSA	0,8
18	FSC	1,9	18	FSC	2,0	15	PCXF	0,8
16	CLINK	0,9	16	CLINK	1,4	18	FSC	0,7
7	FARIN	0,4	7	FARIN	0,1	19	MS	0,0

Três entradas, apontadas pelos três métodos, na Tabela 8, são consideradas importantes, na análise de sensibilidade da rede obtida: emissão de particulados na saída da chaminé (OPAC) e rotação dos exaustores, nos ramais A e B (FANA e FANB). Quanto aos demais valores, nada se pode concluir.

Tabela 9 - Análise de Sensibilidade da Rede com pesos otimizados: Entradas menos importantes para Peso Litro

Rede Otimizada - 31 pesos								
<i>Fórmula de Garson</i>			<i>Fórmula Modificada</i>			<i>Média com 20% acréscimo</i>		
Entrada		%	Entrada		%	Entrada		%
Número	Nome		Número	Nome		Número	Nome	
1	O2	0,0	1	O2	0,0	1	O2	0,0
3	TMPA	0,0	3	TMPA	0,0	3	TMPA	0,0
6	NOX	0,0	6	NOX	0,0	6	NOX	0,0
14	PRSB	0,0	14	PRSB	0,0	14	PRSB	0,0

Podemos observar na Tabela 9 que, para estes resultados, os valores são iguais, ou seja, todas essas entradas não foram consideradas importantes para a predição do Peso Litro, nos três métodos utilizados.

Capítulo 5

CONCLUSÕES E RECOMENDAÇÕES

Neste trabalho efetuamos vários testes, comparações e análises, quanto à estrutura da rede, quanto ao procedimento de otimização dos pesos das redes, quanto ao modelo obtido, quanto à análise de sensibilidade das entradas das redes predizendo a Cal Livre ou o Peso Litro, cujas conclusões são apresentadas a seguir.

Quanto à estrutura de rede

Um eventual questionamento a este trabalho seria sobre a não utilização de modelos de rede temporais, tais como redes recorrentes, mais adequadas à identificação de um modelo temporal, cujo resultado depende do resultado imediatamente anterior. Estes modelos são bastante comuns em processos contínuos, como é o presente caso, porém, podem existir restrições quanto a imprecisões que ocorrem em mudanças de campanha, ou com alterações da qualidade de matéria-prima, ou em mudanças das especificações dos produtos. Para que a aplicação de modelos temporais fosse possível neste equipamento, teríamos que modificar procedimentos operacionais da planta, em termos de amostragem e análise de produtos, que, hoje, acontecem a cada quatro horas, num intervalo muito grande em relação à coleta dos dados operacionais, que ocorrem de hora em hora, pois o tempo morto, estimado para o forno, é cerca de 50 minutos, e os dados operacionais perdem significância em relação ao produto amostrado, para a análise temporal.

A rede utilizando algoritmo MLP – “Multi Layer Perceptron” - treinada em *backpropagation*, conseguiu modelar bem o forno, mesmo com poucos dados, e sendo estes esparsos, mostrando não ser necessária a modelagem de séries temporais, nem o uso de modelos recorrentes de rede.

Quanto ao procedimento de otimização dos pesos da rede

O método OBS – “Optimal Brain Surgeon” otimizou os pesos da rede treinada, melhorando, tanto o erro global, quanto a sua capacidade de generalização.

O número total de pesos deve ser similar, ou menor, que o número de casos apresentados para treinamento da rede, sem o que a rede pode tornar-se sobre treinada. A divisão da rede, para a predição de uma qualidade por rede, contribuiu para minimizar o número de pesos da rede, minimizando seu erro global e melhorando sua capacidade de generalização.

Quanto ao modelo obtido

Pelos resultados alcançados pelas redes desenvolvidas, podemos concluir que elas conseguiram prever bem as qualidades do clínquer, como pode ser observado na Figura 41, ressaltando a dificuldade em generalização devido aos poucos dados experimentais disponíveis. Nessa figura observamos que o índice de correlação, aplicado a todos os dados experimentais, comparados aos resultados preditos para a Cal Livre, ficou em $R^2=0,95$. Na Figura 42, observamos que o índice para a predição do Peso Litro ficou em $R^2=0,89$.

Quanto à análise de sensibilidade das entradas para a rede predizendo Cal Livre

Para a predição da Cal Livre com a rede treinada usando dois nós intermediários, a análise de sensibilidade das entradas apontou, em todos os métodos aplicados, como sendo a mais importante entrada:

- emissão de particulados na saída da chaminé (OPAC);

Esta entrada está particularmente associada ao correto processo de combustão do forno, indicando a influência direta do mesmo no valor da Cal Livre do clínquer.

Quanto a outras entradas importantes, os dois métodos utilizados para a análise da rede inicial, indicaram que as duas entradas seguintes também têm importância significativa no valor da Cal Livre do clínquer:

- o teor de monóxido de carbono na saída da torre de ciclone (PH-CO), e

- a temperatura do clínquer na saída do resfriador (CLINK),

Isto é, de certa forma, confirmado pela análise da rede com os pesos otimizados, pois, nos três métodos aplicados, elas também aparecem entre as quatro entradas seguintes mais importantes.

Já as entradas:

- vazão de farinha (FARIN), e
- rotação dos exaustores no ramal A (FANA)

que aparecem listadas por duas, das três metodologias utilizadas, entre as quatro entradas seguintes mais importantes, também devem ser consideradas como entradas importantes, e quanto às demais, os resultados não são conclusivos, se são muito importantes para o controle do valor da Cal Livre no clínquer.

Da mesma forma, analisando a rede inicial, as duas metodologias utilizadas indicaram as seguintes entradas como as menos importantes:

- temperatura de saída dos gases da torre de ciclone no ramal A (TMPA);
- perda de pressão na saída da torre de ciclone, no ramal B (PRESB);
- teor de NO_x nos gases de Combustão (NOX);
- Módulo de Sílica (MS);
- vazão de carvão combustível (CARV);
- perda de pressão na caixa de fumaça (PCXF);
- rotação dos exaustores no ramal B (FANB).

Todas essas entradas, a partir da análise de sensibilidade da rede otimizada, também foram consideradas sem importância para o controle da Cal Livre, exceto para:

- teor de NO_x nos gases de Combustão (NOX);
- rotação dos exaustores no ramal B (FANB), e
- perda de pressão na saída da torre de ciclone, no ramal B (PRESB)

que são apontadas em dois dos três métodos de análise utilizados.

Em princípio, todas essas entradas, poderiam ser descartadas, ou seja, poderíamos elaborar uma nova rede sem essas entradas, para a predição da Cal Livre. Cabe, porém a ressalva para uma dessas entradas:

- rotação dos exaustores no ramal B (FANB),

uma vez que ela é uma medição efetuada em um ramal, e a medição correspondente ao outro ramal foi considerada importante. Uma opção seria a simples eliminação de uma das entradas, e outra seria utilizar a média entre as duas, para uma nova rede.

Quanto à análise de sensibilidade das entradas para a rede predizendo Peso Litro

A análise de sensibilidade das entradas da rede treinada para predição do Peso Litro, usando dois nós intermediários, apontou, em todos os métodos aplicados, as entradas:

- emissão de particulados na saída da chaminé (OPAC),
- rotação dos exaustores no ramal A (FANA), e
- rotação dos exaustores no ramal B (FANB),

como sendo as mais importantes, no controle do valor do Peso Litro do clínquer produzido. A entrada:

- perda de pressão na caixa de fumaça (PCXF),

também parece ser importante, pois está entre as mais importantes entradas, indicadas pelos dois métodos utilizados na análise da rede inicial e por dois dos três métodos aplicados na análise da rede otimizada.

Na rede inicial, nos dois métodos utilizados, há a indicação da entrada:

- torque do motor elétrico, para rotação do forno rotativo (TORQ),

como sendo uma entrada importante, porém, na análise de sensibilidade da rede otimizada, a importância dessa entrada cai, nas três metodologias empregadas.

Do ponto de vista das entradas menos importantes, em todos métodos de análise de sensibilidade das entradas, seja para a rede inicial, como para a rede otimizada, as entradas:

- teor de oxigênio residual dos gases na saída da torre de ciclone (O₂), e
 - perda de pressão na saída da torre de ciclone, no ramal B (PRESB);
- são indicadas como sendo sem importância para a predição do Peso Litro.

Além dessa, a entrada:

- teor de NO_x nos gases de combustão (NO_x),
aparece em quatro das metodologias utilizadas, e a entrada:
- temperatura de saída dos gases da torre de ciclone no ramal A (TMPA),
aparece em três, e naquela onde não é indicada como não prioritária, é indicada a medição correspondente ao outro ramal:
- temperatura de saída dos gases da torre de ciclone no ramal B (TMPB).

Recomendações

Recomendamos que as redes propostas, para predição de Cal Livre e de Peso Litro sejam implantadas e testadas durante tempo suficiente para verificar sua concordância com os dados de processo, e que sejam efetuados testes na fábrica, para verificar a importância das entradas apontadas pelos métodos de análise de sensibilidade.

No caso das redes apresentarem dificuldade em predizer bem as propriedades desejadas no clínquer, os valores de Cal Livre e os valores de Peso Litro, recomendamos uma das duas possibilidades:

- Efetuar nova coleta de dados e retreinar as redes, ou
- Eliminar as entradas indicadas como de baixa importância e re-estruturar as redes atuais sem as mesmas e retreiná-las.

No caso do teste de campo apresentar diferença significativa com a importância das entradas, apontadas pelos testes de sensibilidade, sugerimos a aplicação de outras técnicas estatísticas para identificar a importância dessas entradas, assim como o uso do método PCA, utilizado para a ortogonalização das variáveis de entrada, auxiliando no estudo da importância das variáveis de entrada.

REFERÊNCIAS BIBLIOGRÁFICAS

BHAT, N. & McAVOY, T.J. Determining Model Structure for Neural Models by Network Stripping. *Computers chem. Engng.* Vol. 16, n. 4, p. 271-281, 1992.

BUFO, M. J. *Aplicação de Redes Neural Artificial como auxiliar na Predição do Desempenho de um Landfarming.* 2000. 94f. Dissertação (Mestrado em Engenharia Química) - Faculdade de Engenharia Química, Universidade Estadual de Campinas, Campinas.

DI MASSIMO, C.; MONTAGUE, G.A.; WILLIS, M.J.; THAM, M.T. & MORRIS, A.J. Towards Improved Penicillin Fermentation via Artificial Neural Networks. *Computers chem. Engng.* Vol. 16, n. 4, p. 283-291, 1992.

DIAS, W.P.S. Reduction of concrete Sorptivity with Age through Carbonation. *Cement and Concrete Research.* V.30, p.1255-1261, 2000.

ELGIBALY, A.A. & ELKAMEL, A.M. A New Correlation for Predicting Hydrate Formation Conditions for Various Gas Mixtures and Inhibitors. *Fluid Phase Equilibria.* V. 152, p. 23-42, 1998.

GAGNÉ, F & BLAISE, C. Predicting the Toxicity of Complex Mixtures using Artificial Neural Networks. *Chemosphere.* V.35, n°. 6, p. 1343-1363, 1997.

GARSON, G.D. Interpreting Neural-Network Connection Weights. *AI Expert.* Apr, p. 47-51, 1991.

GONTARSKI, C. A. U. – *Avaliação da utilização de Redes Neurais aplicadas a processos químicos.* 2000. 158 f. Tese (Doutorado em Engenharia Química) - Faculdade de Engenharia Química, Universidade Estadual de Campinas, Campinas.

GONTARSKI, C. A.; RODRIGUES, P.R.; MORI, M. & PRENEM, L.F. Simulation of an Industrial Wastewater Treatment Plant using Artificial Neural Networks. *Computers and Chemical Engineering.* V. 24, p.1719-1723, 2000.

HASSIB, B. & STORK, D.G. *Second Order Derivatives for Network Pruning: Optimal Brain Surgeon.* NIPS 5, Ed. S.J.Hanson et al., 164, San Mateo, Morgan Kaufmann, 1993.

KIM, Y.I.; MOON, K.C.; KANG, B.S.; HAN, C. & CHANG, K.S. Application of Neural Network to the Supervisory Control of a Reheating Furnace in the Steel Industry. *Control Engineering Practice.* V.6, p.1009-1014, 1998.

KOVÁCS, Z. L. *Redes Neurais Artificiais: fundamentos e aplicações: um texto básico*. 2ª ed. São Paulo: Edição Acadêmica, 1996.

LAI, S. & SERRA, M. Concrete Strength Prediction by means of Neural Network. *Construction and Building Materials*. V. 11, n.º 2, p. 93-98, 1997.

MURKHERJEE, A. & BISWAS, S.N. Artificial Neural Networks in Prediction of Mechanical Behavior of Concrete at High Temperature. *Nuclear Engineering and Design*. V.178, p. 1-11, 1997.

NEURALWARE. - *Neural Computing – A Technology Handbook for Professional II/Plus and NeuralWorks Explorer*. Pittsburgh, 1995.

NØRGAARD, M. *Neural Network Based System Identification Toolbox: for use with MATLAB*. Institute of Automation. Tech Report 95-E-773. Tech. University of Denmark. 1995, obtido em 1995 de pmn@iau.dtu.dk.

OLIVEIRA, K. P. S. - *Aplicação das Técnicas de Redes Neurais e de Análise de Componentes Principais na Modelagem de uma Lagoa Aerada da RIPASA S/A*. 2000. 64 f. Dissertação (Mestrado em Engenharia Química) - Faculdade de Engenharia Química, Universidade Estadual de Campinas, Campinas.

PHAM, D.T. – *An introduction to Artificial Neural Networks*. In: *Neural Networks for Chemical Engineering*. Great Britain. Elsevier Science B. V., 1995. Cap I

POLLARD, J.F; BROUSSARD, M.R.; GARRISON, D.B. & SAN, K.Y. Process Identification using Neural Networks. *Computers chem. Engng*. Vol. 16, n. 4, p. 253-270, 1992.

VOTORANTIM. *Apostila de laboratório*. Rio Branco-PR, 2000.

YEH, I-C. Modeling of Strength of High-performance Concrete using Artificial Neural Networks. *Cement and Concrete Research*. V. 28, n.º 12, p.1797-1808, 1998.

Anexos

DISTRIBUIÇÃO DOS DADOS DE ENTRADA VS SAÍDA

