

UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE QUÍMICA

*REDES NEURAI E TÉCNICAS DE PRÉ-PROCESSAMENTO
DE SINAIS EM ESPECTROSCOPIA DE REFLECTÂNCIA NO
INFRAVERMELHO PRÓXIMO*

TESE DE DOUTORADO

Candidato: Cesar Alexandre de Mello

Orientador: Prof. Dr. Ronei Jesus Poppi

Co-Orientador: Prof. Dr. João Carlos de Andrade

Campinas, novembro de 1998.



UNIDADE	IA
N.º CHAMADA:	UNICAMP
	M489r
V. Ex.	
TOMBO BC/	37578
PROC.	229199
C	<input type="checkbox"/>
D	<input checked="" type="checkbox"/>
PREÇO	R\$ 11,00
DATA	30/04/99
N.º CPD	400.102.808.9

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO INSTITUTO DE QUÍMICA
UNICAMP**

M489r Mello, Cesar Alexandre de
Redes neurais e técnicas de pré-processamento de sinais em espectroscopia de reflectância difusa no infravermelho próximo (NIRR) / Cesar Alexandre de Mello - - Campinas, [SP : s. n] , 1998.

Orientador: Ronei Jesus Poppi
Co-orientador: João Carlos de Andrade

Tese (doutorado) - Universidade Estadual de Campinas. Instituto de Química.

1. Calibração multivariada. I. Poppi, Ronei Jesus. II. Andrade, João Carlos de Andrade. III. Universidade Estadual de Campinas. Instituto de Química. IV. Título.

“ In this section a mathematical model of the growing embryo will be described. This model will be a simplification and an idealization, and consequently a falsification. It is to be hoped that the features retained for discussion are those of greatest importance in the present state of knowledge. ”

A. Turing, “The Chemical Basis of Morfogenesis”, Philos. Trans. R. Soc., vol. 237, pg. 5, London, 1952

À Lúgia, minha querida esposa e companheira de todos os momentos, aos nossos filhos Thomás e Vitor que tantas alegrias nos tem trazido

e

aos meus pais, Neto e Arlete, por terem me dado a vida e educado bem.

dedico.

AGRADECIMENTOS

ao Prof. Dr. Ronei Jesus Poppi, pela orientação segura, incentivo, amizade e exemplo de conduta científica;

ao Prof. Dr. João Carlos de Andrade, pela co-orientação e amizade;

aos meus Professores;

aos amigos, pelos bons momentos, Ronei, Paulo Filho, Carlos Leite, Bell Bueno, Zezzi, Bilão e Alex Magalhães.

aos colegas de laboratório, pela boa convivência e cooperação, Edenir, Eduardo, Cleidiane, Rosângela, Marcelo, Paulo Fidêncio, Silvio Lima, Iara, Ferrão, Renatão, Flávia, Élvio, Carlos, Clésia e Aline.

ao Instituto de Química da Unicamp;

aos funcionários do Instituto de Química;

ao CNPq, pela bolsa concedida.

e a todas as outras pessoas envolvidas direta ou indiretamente neste trabalho.

Resumo

Autor: Cesar Alexandre de Mello

Orientador: Prof. Dr. Ronei Jesus Poppi

Co-orientador: Prof. Dr. João Carlos de Andrade

Neste trabalho foi avaliado o desempenho de técnicas de pré-processamento de sinais e de modelamento de dados, para estimar o teor de nitrogênio em folhas de trigo a partir de espectros de reflectância difusa no infravermelho próximo (NIRR), obtidos em um espectrofotômetro de infravermelho próximo, com monocromador baseado em filtro óptico acústico sintonizável (AOTF).

As técnicas de pré-processamento de sinais utilizadas foram: filtro de Savitsky-Golay, filtro de Fourier, filtro de Wavelet e Correção multiplicativa de espalhamento de Luz (MSC). Os métodos de modelamento de dados utilizados foram: Redes Neurais ordinárias, Redes Neurais com apodização (RNP), Mínimos Quadrados Parciais (PLS), Mínimos Quadrados Parciais-Polinomial (POLIPLS) e Mínimos Quadrados Parciais-Redes Neurais (NNPLS).

Os resultados obtidos indicam que se desejar-se utilizar, com eficiência, a espectroscopia de reflectância difusa no infravermelho próximo para determinar o teor de nitrogênio em folhas deve-se, invariavelmente, corrigir efeitos relativos ao espalhamento de luz utilizando o MSC e tratar o espectro obtido de modo a remover ruídos, utilizando filtro de Fourier ou filtro de Wavelets; caso contrário, os erros de previsão serão elevados.

Utilizando como referência o método Kjeldahl para determinação de nitrogênio em plantas, observou-se que as redes neurais com apodização apresentaram desempenhos similares aos das redes neurais sem apodização e ao PLS redes neurais. Entretanto, ambos os tipos de redes e o PLS redes neurais, forneceram melhores resultados que o PLS e o PLS polinomial.

Abstract

Author : Cesar Mello

Adviser : Prof. Dr. Ronei Jesus Poppi

Co-adviser : Prof. Dr. João Carlos de Andrade

In this work it was evaluated the performance of signal preprocessing and data modeling-techniques, for nitrogen content estimation in wheat leaves, from near infrared diffuse reflectance spectra (NIRR), using a monochromator based on an acoustic optic tunable filter (AOTF).

The preprocessing techniques employed were : Savitsky-Golay filter, Fourier filter, Wavelet filter, and multiplicative scattering correction (MSC). The modeling methods used were: ordinary neural networks, pruning neural networks (RNP), partial least squares (PLS), polynomial/partial least squares (POLIPLS), and neural networks/partial least squares (NNPLS).

The results obtained show that to use the near infrared diffuse reflectance spectroscopy for nitrogen content estimation in leaves is necessary to correct the light scattering effect using the MSC and to treat the spectra obtained to noise cancellation using the Fourier or Wavelet filter.

It was observed, using as reference the Kjeldahl method for nitrogen content determination in the wheat leaves, that neural networks with pruning, ordinary neural networks and neural networks / PLS have similar performance. However, all neural networks based models provide better results than PLS and POLIPLS.

Conteúdo	
Resumo	<i>i</i>
Abstract	<i>ii</i>
Conteúdo	<i>iii</i>
1. Introdução Geral.....	1
2. A Espectroscopia de reflectância difusa no infravermelho próximo (NIRR).....	4
- A origem do espectro no infravermelho próximo.....	5
- A reflectância difusa.....	6
- A teoria de Kubelka-Munk: <i>O método das equações diferenciais</i>	9
- Adaptação da lei de Beer à reflectância difusa: <i>O método empírico</i>	18
3. Técnicas de pré-processamento de sinais.....	21
3.1 A suavização e a remoção de ruído instrumental.....	22

3.1.1 A suavização de ruído instrumental.....	24
- O filtro de média móvel.....	24
- O filtro de Savitzky-Golay.....	26
3.1.2 A remoção de ruído instrumental.....	29
- O filtro de transformada de Fourier.....	29
- O filtro de transformada de Wavelet.....	42
3.2 A correção do espalhamento de luz (MSC).....	48
3.3 A compressão de dados.....	51
- Compressão de dados através da técnica de Fourier.....	51
- Compressão de dados através da análise de componentes principais (PCA)	52
4. Técnicas quimiométricas de modelamento de dados.....	58
4.1 Redes Neurais.....	59
4.1.1 Neurônios artificiais.....	61
4.1.2 Sinapses e Pesos: A transmissão de sinais pelos neurônios.....	62
4.1.3 Funções de transferência.....	64
- A função sigmoidal ou logística.....	64
- A função degrau de Heaviside.....	65
- A função linear.....	66

4.1.4 A arquitetura de redes neurais.....	66
4.1.4.1 Cálculo do Net em uma rede neural.....	68
4.1.5 Treinamento das redes neurais.....	71
4.1.6 A otimização da Arquitetura.....	75
4.2 Redes neurais com apodização (Pruning).....	76
4.3 O método dos mínimos quadrados parciais (PLS).....	79
4.4 Outras técnicas de modelamento de dados não-lineares.....	80
- O PLS polinomial (POLIPLS).....	80
- O PLS redes neurais (NNPLS).....	81
5. Objetivos.....	82
6. Parte Experimental.....	84
6.1 Determinação de nitrogênio em folhas: <i>O método Kjeldahl</i>.....	85
6.2 Obtenção dos espectros de reflectância difusa no infravermelho próximo.....	86
6.3 Programas computacionais para o tratamento dos dados.....	93

7. Resultados e Discussão.....	96
7.1 Os espectros obtidos.....	97
7.1.1 A Homogeneidade das amostras.....	98
7.2 Filtros para minimização de ruído em espectros NIR.....	99
7.2.1 Filtro de transformada de Fourier.....	100
7.2.2 Filtro de transformada de Wavelet.....	102
7.2.3 Filtro de Savitzky-Golay.....	103
7.2.4 Comparação dos filtros utilizados.....	104
7.3 Resultados da correção de espalhamento de luz.....	105
7.4 A umidade das amostras.....	107
7.5 Seleção das amostras.....	109
7.6 Modelamentos.....	111
7.6.1 Avaliação do desempenho dos modelos.....	111
7.7 Redes neurais e o reconhecimento de padrões.....	131
7.8 Unindo classificação e modelamento.....	133
8. Conclusões.....	134
9.Referências Bibliográficas.....	137

CAPÍTULO 1

INTRODUÇÃO GERAL

1. Introdução geral

A química analítica encontra-se em uma fase muito peculiar de seu desenvolvimento. Existem métodos bem estabelecidos, precisos e eficientes para a determinação qualitativa ou quantitativa de quase todos os compostos inorgânicos e para boa parte dos compostos orgânicos. Infelizmente, a grande maioria destes métodos são dispendiosos, envolvem várias etapas e, principalmente, geram quantidades razoáveis de resíduos extremamente poluentes e agressivos ao meio ambiente. Além disso, tais métodos incentivam uma ampla gama de indústrias de reagentes químicos que, em geral, ainda utilizam processos altamente poluentes, fechando um ciclo vicioso que, a longo prazo, pode causar danos irreversíveis ao meio ambiente. Assim sendo, um dos grandes desafios da pesquisa em química analítica é o desenvolvimento de métodos de análises *in situ* que não necessitem de pré-tratamento de amostras, que sejam rápidos, não destrutivos, robustos e que não poluam o meio ambiente.

Uma técnica espectroscópica que preenche todos os pré-requisitos anteriores para um grande número de substâncias, em diferentes tipos de matrizes (por exemplo: alimentos, fármacos e plásticos), é a espectroscopia de reflectância difusa no infravermelho próximo¹.

A espectroscopia de reflectância difusa no infravermelho próximo permite analisar quantitativamente um grande número de substâncias em diferentes tipos de matrizes⁽²⁻⁹⁾, sem a necessidade da “abertura química” da amostra, reduzindo o consumo de reagentes, o tempo de análise e possibilitando análises *in situ*¹⁰. Estas vantagens conferem à técnica, características particularmente úteis à análises de rotina e controle *on line* de processos industriais⁽¹¹⁻¹²⁾.

Aplicações analíticas⁽¹³⁻¹⁴⁾ do infravermelho próximo vem sendo exploradas desde o início da década de cinquenta mas, devido às dificuldades instrumentais, a sua aplicação como técnica quantitativa só teve início trinta anos mais tarde, com o desenvolvimento e uso generalizado de microcomputadores. Graças ao interfaceamento entre equipamentos e microcomputadores, tornou-se simples obter e armazenar grandes quantidades de dados, de modo rápido e eficiente.

Uma vez superadas as dificuldades para a obtenção dos espectros de reflectância difusa no infravermelho próximo, o passo seguinte foi obter informações quantitativas destes espectros. Nesta fase observou-se que medidas de reflectância difusa no infravermelho próximo podem ser relacionadas à concentração de uma ou mais substâncias presentes na amostra, mas não de uma maneira direta, dificultando ou impossibilitando a aplicação de procedimentos clássicos de calibração univariada.

Atualmente, com o desenvolvimento da quimiometria, o que se faz é medir a reflectância em um grande número de comprimentos de onda, e utilizar um método de calibração multivariada¹⁵ para estabelecer a relação entre a reflectância difusa e concentração, ou outra propriedade de interesse. Dentre tais métodos destacam-se os não-lineares de calibração multivariada¹⁶ e mais recentemente as redes neurais¹⁷ e as redes neurais com apodização¹⁸.

CAPÍTULO 2

A ESPECTROSCOPIA DE REFLECTÂNCIA DIFUSA NO INFRAVERMELHO PRÓXIMO (NIRR).

2. Espectroscopia de reflectância difusa no infravermelho próximo (NIRR).

*A origem do espectro no infravermelho próximo*¹⁹

O infravermelho próximo²⁰ (NIR) é a região do espectro eletromagnético situada entre 800 nm e 2500 nm. Quando uma molécula absorve energia nesta região, observam-se transições do estado fundamental para o segundo e terceiro estados excitados, gerando as bandas de sobreton²¹ (overtone). Uma vez que poucas moléculas atingem os estados excitados, relativos ao primeiro e segundo sobreton, a intensidade de tais bandas são muito menores do que a intensidade da transição do estado fundamental para o primeiro estado excitado, observadas no infravermelho médio²².

Transições superiores são possíveis mas, como o número de moléculas que executam esta transição é muito pequeno, as bandas relativas a estas transições são muito pouco intensas e, portanto, possuem pouco significado prático.

A absorção de radiação no infravermelho próximo por substâncias de origem animal e vegetal²³ pode ser atribuída, fundamentalmente, à pequena massa do átomo de hidrogênio. Átomos de hidrogênio ligados covalentemente a átomos mais pesados e eletronegativos (O, N, C, etc) formam dipolos oscilantes anarmônicos que absorvem energia em regiões de alta frequência, isto é, na região do infravermelho próximo. Devido à anarmonicidade²⁴ existente na vibração da ligação entre átomos de hidrogênio e átomos mais pesados, as bandas de sobreton não ocorrerão em submúltiplos inteiros (1/2, 1/3) do comprimento de onda da

transição²⁴ fundamental, mas próximas aos submúltiplos (não inteiros) desta transição. A maioria das oscilações fundamentais das ligações dos átomos de hidrogênio situam-se na faixa de 2700 a 3600 *nm*. A maioria das bandas de sobreton são encontradas entre 400 e 2200 *nm*. As vibrações fundamentais podem combinar-se e gerar as bandas de combinação. A região entre 1800 e 2500 *nm* é ocupada pelas bandas de combinação²⁴.

Toda matéria de origem animal e vegetal é composta, basicamente, de proteínas, amido, açúcares, lipídios, celulose etc. Uma vez que todos estes constituintes possuem grupos funcionais (osciladores) em que os átomos de hidrogênio encontram-se ligados a átomos mais pesados (C, N, O), todos são candidatos à apresentarem absorção no infravermelho próximo. Os osciladores mais comuns²⁴ (dentre outros) são: -OH, -NH-, etc.

Como veremos no próximo item desta introdução, a espectroscopia de reflectância difusa no infravermelho próximo é um método capaz de fornecer informações quantitativas a respeito da composição da amostra e portanto, torna-se uma ferramenta analítica muito útil para análise de proteínas, amido, nitrogênio, celulose, nitrogênio, etc.

A Reflectância Difusa^{1, 25}

Quando uma superfície rugosa ou particulada, como um pó por exemplo, é irradiada com radiação NIR, uma parte da radiação incidente é refletida diretamente pela superfície. A este tipo de reflexão dá-se o nome de *reflectância especular*. À definição resumida, ora tecida sobre a reflectância especular, devemos acrescentar que a reflexão não impede que

as partículas da superfície, simultaneamente, absorvam a radiação NIR. Em verdade, além da reflexão e absorção, também é possível que ocorra, simultaneamente, dependendo do índice de refração da partícula, a refração e a difração da radiação NIR incidente.

A natureza da radiação refletida depende da natureza do feixe incidente e das propriedades ópticas da superfície em questão. Tal dependência, em relação as propriedades ópticas da superfície pode ser utilizada na caracterização de certos materiais entretanto, a aplicação da reflectância especular para fins quantitativos, na maioria dos casos, é ainda pouco explorada, dado que no processo de reflectância especular a radiação incidente interage com um pequeno número de moléculas da superfície e, portanto, o feixe refletido praticamente não é atenuado, possuindo quase a mesma intensidade da radiação incidente e para todos efeitos práticos, não fornece informações quantitativas sobre a amostra.

Uma outra parte da radiação incidente penetra na superfície da amostra e a reflexão ocorre logo abaixo da superfície (alguns milímetros), isto é, dentro da amostra. Uma vez que as partículas do interior da amostra apresentam-se com suas faces, estatisticamente distribuídas sob todos os possíveis ângulos, a radiação refletida no interior da amostra seguirá a mesma distribuição. Dado que as partículas em questão encontram-se no interior da amostra, a radiação refletida por uma partícula poderá ser novamente refletida por uma partícula vizinha e assim por diante, até que a radiação, por algum caminho estatisticamente provável, consiga retornar através da superfície e deixar a amostra. Este tipo de reflexão, que ocorre abaixo da superfície, é chamada de *reflectância difusa*. A associação entre reflectância especular e reflectância difusa, dá origem ao que se chama, genericamente, de espalhamento de luz. Um exemplo familiar deste

fenômeno é o espalhamento da luz solar, pelas moléculas da atmosfera terrestre. Se esta não existisse, o céu diurno pareceria negro, exceto na direção do Sol.

Na reflectância difusa, a radiação incidente entra em contacto milhares de vezes com as partículas da amostra e portanto, é severamente atenuada (perde energia), logo a radiação que deixa a amostra poderá fornecer informações qualitativas e quantitativas, sobre a amostra. O processo de reflectância difusa pode ser visualizado na Figura 2.1.

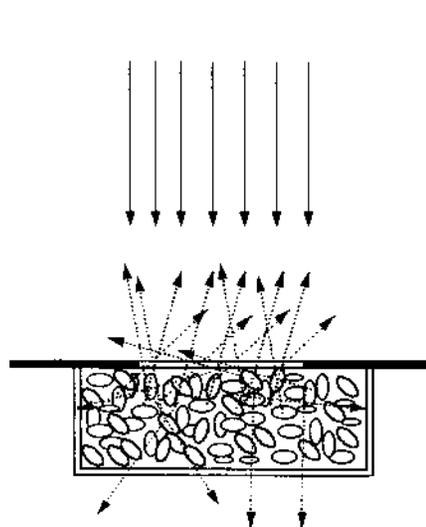


Figura 2.1. Representação esquemática da reflectância difusa

Naturalmente, existirão alguns caminhos, também estatisticamente prováveis, pelos quais a radiação não retornará pela superfície mas, será absorvida ou até mesmo atravessará as paredes do porta amostra, conforme representado na Figura 2.1.

A espectroscopia de reflectância difusa no infravermelho próximo (NIRR) permite analisar qualitativa e quantitativamente vários tipos de amostras sólidas, neste caso folhas moídas de trigo, milho e algodão, sem necessidade de “abertura química” da amostra. Devido a esta

característica, o consumo de reagentes e tempo de análise são bastante reduzidos. Entretanto, para que a espectroscopia NIRR possa adquirir finalidades quantitativas é necessário estabelecer uma relação matemática entre a intensidade da radiação espalhada ou que sofreu reflectância difusa, e a concentração da substância de interesse presente na amostra. Existem, basicamente, duas maneiras de estabelecer-se tal relação: a primeira delas é a teoria de Kubelka-Munk e a segunda é uma adaptação da lei Beer para a reflectância difusa. Começemos com a primeira.

A Teoria de Kubelka - Munk: Método das equações diferenciais^{1, 25}

A teoria fenomenológica desenvolvida por Kubelka e Munk em 1931, assume que a radiação que incide sobre uma superfície rugosa ou particulada (um pó por exemplo), pode ser dividida em dois fluxos. Um fluxo I que representa a radiação que penetra e é absorvida pela amostra, na direção do feixe incidente, e um fluxo J que retorna da amostra, devido à reflectância difusa, conforme representado na Figura 2.2.

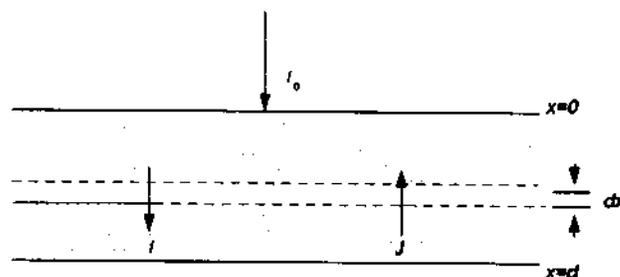


Figura 2.2. Representação gráfica dos fluxos envolvidos no modelo de Kubelka - Munk. I_0 é a intensidade da radiação incidente, I a fração absorvida e J fração que sofreu reflectância difusa.

Uma equação diferencial pode ser facilmente escrita para a variação (dI) da intensidade do feixe incidente à medida que este atravessa um elemento diferencial dx , da referida superfície (Fig. 2.2). A intensidade da radiação absorvida ou atenuação do feixe incidente pode ser escrita como sendo linearmente proporcional à intensidade do feixe incidente e à espessura do elemento dx , como segue:

$$dI = -K I dx \quad \text{eq.2.1}$$

O feixe incidente I , também será atenuado por perdas devidas à refração, difração e reflexão, isto é, perdas devidas à reflectância difusa. A atenuação do feixe incidente devido à reflectância difusa pode ser descrita pela equação 2.2, como segue:

$$dI = -S I dx \quad \text{eq.2.2}$$

em que S é a constante relativa às perdas por espalhamento. Combinando as equações 2.1 e 2.2, temos a atenuação total do feixe incidente, como segue:

$$dI = -K I dx - S I dx \quad \text{eq.2.3}$$

Para encerrarmos a dedução das equações da teoria de Kubelka e Munk, devemos ainda levar em consideração que a radiação do feixe I que foi espalhada, aumenta a intensidade do feixe J e vice versa. Assim sendo, a equação 2.3, que descreve a atenuação total no feixe I , torna-se:

$$dI = -(K + S)I dx + S J dx \quad \text{eq2.4}$$

Aplicando o mesmo raciocínio anterior podemos obter a equação que descreve o fluxo de reflectância difusa (J), como segue:

$$dJ = (K + S)J dx - S I dx \quad \text{eq2.5}$$

As equações 2.4 e 2.5, formam um conjunto de equações diferenciais ordinárias acopladas, as quais podem ser facilmente resolvidas através do método da transformada de Laplace. A transformada de Laplace direta, transformará as equações diferenciais em duas equações algébricas lineares simultâneas. Resolve-se estas equações algébricas e através da transformada de Laplace inversa, obtém-se a solução do sistema de equações diferenciais ordinárias. Obtida a solução, basta aplicar as seguintes condições de contorno:

$$\text{quando: } x = 0 \quad \Rightarrow \quad I = I_0$$

esta condição reflete o fato de que, na superfície não há atenuação do feixe, o que naturalmente é uma aproximação questionável mas, necessária à solução do sistema.

A segunda condição é a seguinte:

$$\text{quando: } x = d \quad \Rightarrow \quad j = 0$$

Esta condição representa o fato, ou melhor, a aproximação de que nas superfícies internas do porta amostra não há fluxos relativos a espalhamento.

Aplicando as condições de contorno à solução previamente obtida para o sistema, finalmente obtemos as equações que descrevem os fluxos I e J , como segue:

$$T = \frac{I_{x=d}}{I_0} = \frac{4\beta}{(1+\beta)^2 e^{\kappa d} - (1-\beta)^2 e^{-\kappa d}} \quad \text{eq2.6}$$

$$R_D = \frac{J_{x=d}}{I_0} = \frac{(1-\beta^2)(e^{\kappa d} - e^{-\kappa d})}{(1+\beta^2)e^{\kappa d} - (1-\beta^2)e^{-\kappa d}} \quad \text{eq2.7}$$

em que:

$$\kappa = \sqrt{K(K+2S)} \quad \text{eq.2.8}$$

$$\beta = \sqrt{\left(\frac{K}{K+2S}\right)} \quad \text{eq.2.9}$$

A interpretação físico-química para os fluxos I e J , pode ser feita se considerarmos os limites do modelo desenvolvido por Kubelka e Munk.

① *Interpretação do fluxo I (transmitância) :*

Analiseemos o que ocorre com a transmitância, quando as perdas por espalhamento tendem a zero:

$$\lim_{s \rightarrow 0} T = e^{-Kd}$$

portanto, no limite quando S tende a zero, isto é, quando não há espalhamento de luz, a expressão para T , torna-se :

$$T = e^{-Kd} \quad \text{eq.2.10}$$

A semelhança da eq2.10 com a lei de Beer é óbvia, o que nos leva à escrever:

$$K = \varepsilon C \quad \text{eq.2.11}$$

Assim, K pode ser interpretado, à princípio, como o produto εC , da lei de Beer. Entretanto, as coisas não são tão simples assim, pois na lei de Beer as absorções ocorrem, exclusivamente, na mesma direção do feixe incidente e na teoria de Kubelka-Munk, podem existir absorções de feixes refletidos por partículas vizinhas, aumentando o caminho óptico, conforme descrevemos no início deste item. Tais fatos permitem absorções em direções diferentes àquela do feixe incidente e portanto a analogia feita com lei de Beer, não é absolutamente válida. Algumas correções podem ser feitas para incorporar tal diferença. É possível demonstrar que se existe uma distribuição isotrópica dos feixes incidentes, o caminho médio percorrido pela luz é, aproximadamente, duas vezes o caminho percorrido na direção do feixe incidente e podemos escrever:

$$K = 2 \varepsilon C \quad \text{eq.2.12}$$

A análise do efeito do espalhamento de luz na transmitância (eq2.6), derivada na teoria de Kubelka e Munk, pode ser simplificada, se observarmos a Figura 2.3.

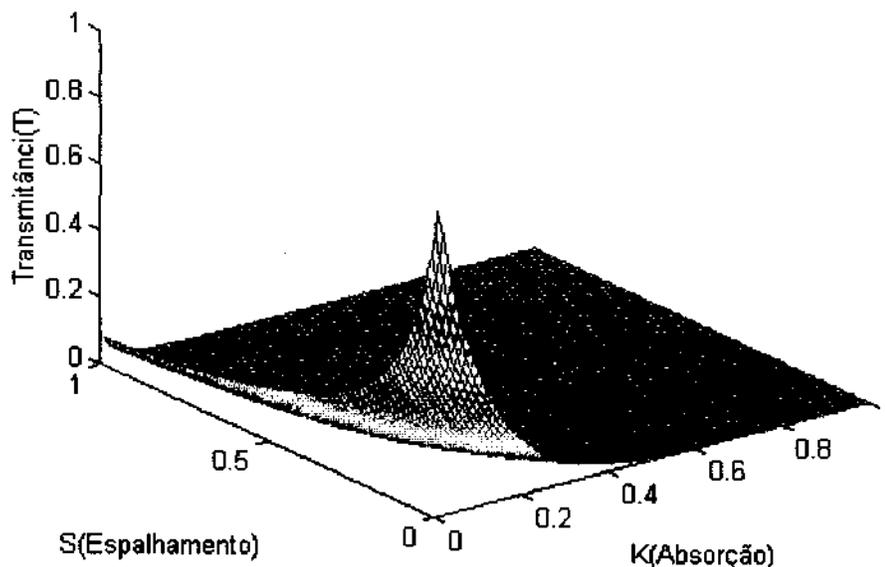


Figura 2.3. Superfície relativa ao comportamento da transmitância em função do espalhamento e da absorção, derivada da teoria de Kubelka e Munk.

Na superfície representada na Figura 2.3, podemos observar que à medida que o espalhamento de luz tende a zero, a variação na transmitância se aproxima à variação exponencial, prevista pela lei de Beer. Além disso, também é possível observar que, a espalhamentos elevados, a transmitância é sempre baixa, indicando que pouca, ou nenhuma radiação, atravessa a amostra.

② Interpretação do fluxo J (Reflectância) :

A interpretação direta do fluxo J é complexa e polêmica, pois este depende da interação simultânea de vários fatores relativos ao

espalhamento da radiação. Entretanto, uma análise menos ambiciosa indica que a intensidade deste fluxo é inversamente relacionada, fundamentalmente, ao tamanho das partículas pois, à medida que diminuimos a granulometria de um material, aumentamos a intensidade da luz espalhada. Isto ocorre pois, ao diminuirmos a granulometria aumentamos o número de superfícies capazes de refletir, absorver, difratar e refratar a luz incidente e, portanto, aumentamos a intensidade do espalhamento de luz. Um exemplo familiar deste fato é a cor branca do sal de cozinha, que na forma de grandes cristais é transparente e torna-se branco devido ao processo de moagem, em que se reduz a granulometria.

A interpretação da reflectância difusa (R_D) descrita na teoria de Kubelka e Munk (eq.2.7), pode ser simplificada se analisarmos a Figura 2.4.

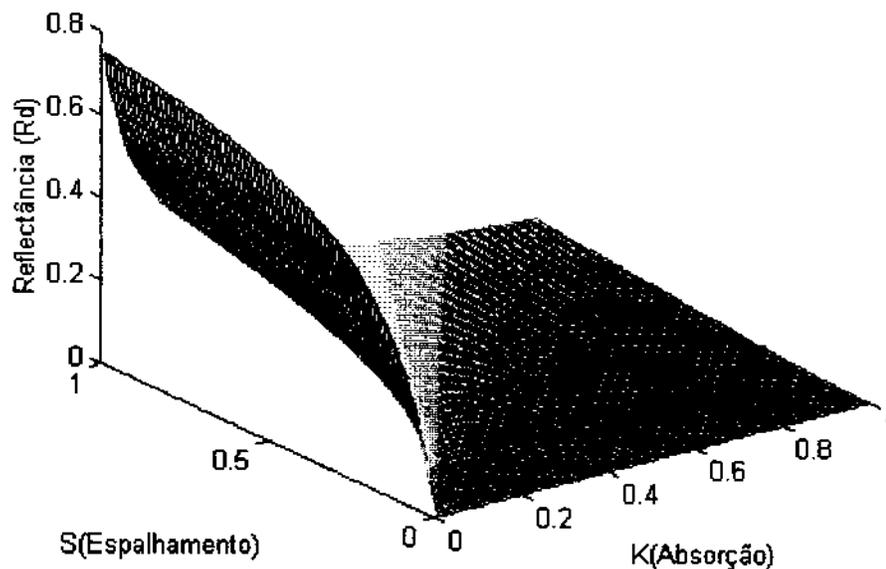


Figura 2.4. Superfície relativa ao comportamento da Reflectância difusa (R_D) em função do espalhamento e da absorção, derivada da teoria de Kubelka e Munk.

Na Figura 2.4 podemos observar que a intensidade da reflectância difusa (R_D) aumenta à medida que aumenta o espalhamento e diminui à medida que aumenta a absorção. Uma vez que o espalhamento de luz é inversamente proporcional à granulometria, se desejarmos aumentar a intensidade de medidas de reflectância, devemos reduzir a granulometria.

O fato de medidas de reflectância difusa serem mais intensas para amostras com menor granulometria representa, em algumas situações, uma séria limitação ao seu uso para finalidades analíticas quantitativas, pois amostras com a mesma concentração da substância de interesse, podem possuir espectros de reflectância difusa no infravermelho próximo, com intensidades diferentes, impossibilitando procedimentos de calibração. Uma maneira experimental de contornar este problema é a padronização da granulometria das amostras, o que representa um procedimento experimental de razoável complexidade, pois envolveria diversas operações de peneiração e subsequentes medidas da distribuição do tamanho de partícula. Entretanto, existem atualmente alguns métodos matemáticos que nos permitem eliminar ou reduzir significativamente, os efeitos relativos ao espalhamento luz devido à granulometria, conforme veremos mais adiante no texto.

A teoria de Kubelka e Munk pode apresentar uma solução extremamente simples para o fluxo J , isto é, para a reflectância difusa, nos casos em que a amostra é espessa o suficiente para evitar que a radiação incidente a atravesse completamente. Nestes casos a equação que descreve a reflectância pode ser escrita como:

$$R = \frac{1 - \beta}{1 + \beta} \quad \text{eq.2.13}$$

em que β é dado pela eq.2.9. Com algumas manipulações matemáticas, tediosas, é possível rescrever a eq.2.13, como segue:

$$f(R) = \frac{(1-R)^2}{2R} = \frac{K}{S} \quad \text{eq.2.14}$$

uma vez que (eq.2.11):

$$K \approx \varepsilon C$$

obtemos:

$$f(R) = \frac{\varepsilon}{S} \cdot C \quad \text{eq.2.15}$$

em que, a função $f(R)$ é geralmente chamada de função K-M.

Experimentalmente é observado que o fator S é dependente do comprimento de onda mas, felizmente a dependência observada é muito pequena ou praticamente nula para a maioria dos casos. Assim sendo, a variação da reflectância com o comprimento de onda deve possuir as mesmas características de um espectro de transmitância, dado que ambos são relacionados por uma constante multiplicativa: o fator S . À luz da discussão anterior, podemos deduzir que, analogamente ao que ocorre com espectros de absorção, a concentração de espécies absorventes pode ser relacionada à função K-M, como segue:

$$f(R) \propto C$$

Em última análise, é a relação anterior que, segundo a teoria de Kubelka e Munk, permite utilizar a espectroscopia de reflectância difusa no infravermelho próximo para finalidades analíticas quantitativas.

A grande dificuldade encontrada na aplicação da teoria de Kubelka e Munk em análises quantitativas é a obtenção de padrões de referência com 100% de reflectância, isto é, um material de referência essencialmente não absorvente. A utilização de padrões com reflectância inferior resultará em funções K-M com valores negativos para regiões espectrais onde a absorção for baixa.

Durante muitos anos o padrão primário para a reflectância difusa foi o óxido de magnésio finamente particulado. Superfícies preparadas com MgO tendem a ser frágeis e a deteriorar-se com tempo. Assim, a utilização de padrões secundários tornou-se comum, ainda que estes apresentem reflectância inferior à do MgO. Os materiais utilizados como padrões secundários, mais comuns, são: *NaF*, *NaCl*, *MgSO₄*, *Li₂CO₃*, *SiO₂*, *Al₂O₃*, *BaSO₄* e Teflon finamente particulado.

Conforme pudemos observar, a teoria de Kubelka e Munk é um excelente exercício para compreensão dos fenômenos envolvidos em medidas de reflectância difusa e de fundamental importância no projeto de novos métodos de análises quantitativas fundamentados na espectroscopia de reflectância no infravermelho próximo. Entretanto, a aplicação prática da função K-M é limitada pela dificuldade de obtenção de padrões com 100% de reflectância, estáveis no tempo.

② Adaptação da lei de Beer à reflectância difusa: *O método empírico*^{1,25}

Neste método o que se faz é associar diretamente a absorvância descrita na lei de Beer, com a reflectância difusa para os casos em que a espessura da amostra é suficientemente grande para evitar que a radiação incidente atravesse-a completamente. A justificativa para tal liberdade científica, é a seguinte: a intensidade da radiação incidente convertida em reflectância difusa, atenuada nos comprimentos de onda absorvidos pela amostra, é análoga à intensidade da radiação atenuada pela amostra em espectroscopia de transmitância de soluções. Sendo a analogia anterior verdadeira, podemos escrever:

$$A = -\text{Log}_{10} \left(\frac{R}{R_0} \right) \quad \text{eq.2.15}$$

em que R é a porcentagem de reflectância difusa da amostra e R_0 a porcentagem de reflectância difusa do padrão.

Neste método não há a necessidade de que o padrão apresente 100% de reflectância difusa pois, todas as medidas são feitas em relação ao mesmo padrão e cada aplicação envolve o uso de uma equação única, empiricamente derivada. O procedimento para a construção de uma curva de calibração usando espectros de reflectância é exatamente igual ao utilizado para espectros de transmitância (ou absorvância).

Naturalmente, a analogia utilizada para a obtenção da equação 2.15 é uma simplificação da situação física, pois associada à reflectância difusa encontra-se, invariavelmente, a reflectância especular. Uma consequência

desta associação é a perda da uniformidade do caminho óptico entre as amostras. O caminho nestes casos passa a ser função dos seguintes parâmetros: tamanho, forma, orientação das partículas, empacotamento das partículas na amostra, distribuição de granulometria, etc.

Assim sendo, o caminho óptico torna-se uma variável extra na equação 2.15 e, portanto, mesmo nos casos em que só um componente da amostra absorve radiação em um único comprimento de onda, serão necessárias medidas de absorbância em outros comprimentos de onda para que se possa montar um sistema de equações algébricas relacionando o caminho óptico e concentração, com a absorbância. Este fato foi durante alguns anos, um fator limitante ao uso da espectroscopia de reflectância difusa no infravermelho próximo em análises quantitativas.

Atualmente, o que se faz é medir a intensidade da reflectância em um grande número de comprimentos de onda (no espectro de reflectância difusa no infravermelho próximo) e utilizar um método de calibração multivariada para estabelecer a relação entre reflectância e concentração.

Neste ponto encerramos a introdução sobre espectroscopia de reflectância difusa no infravermelho próximo. Não houve a pretensão de descrever todos os aspectos teóricos mas somente o necessário à compreensão de como esta técnica foi aplicada, neste trabalho, à determinação de nitrogênio em folhas de trigo, milho e algodão.

CAPÍTULO 3

TÉCNICAS DE PRÉ-PROCESSAMENTO DE SINAIS

3. Técnicas de Pré-Processamento de Sinais

3.1 A suavização e a remoção de ruído instrumental^{26, 27}

Sinais analíticos são os dados originais obtidos de processos analíticos. Geralmente os sinais analíticos representam a variação de uma propriedade físico-química, em função do tempo e/ou espaço. Em química analítica, os sinais mais frequentemente encontrados são: espectros, voltamogramas, registros de um pH-metro em função do tempo, etc.

Infelizmente, a maioria dos sinais provenientes de instrumentos analíticos apresentam-se contaminados (digamos, *misturados*) com ruído, dificultando a interpretação e modelamento das propriedades associadas à estes sinais. Tipicamente, em química analítica, o ruído instrumental afeta os limites de detecção, de decisão e de determinação, os quais são figuras de mérito extremamente importantes na avaliação do desempenho de instrumentos e métodos analíticos²⁸.

Uma classificação, simplificada, dos tipos de ruídos instrumentais^{29,30} pode ser estabelecida se levarmos em consideração as frequências que os caracterizam. A classificação mais simples é a seguinte:

⇒ Ruído de baixa frequência.

⇒ Ruído de frequência intermediária.

⇒ Ruído de alta frequência.

O ruído de baixa frequência caracteriza-se, na prática, por deslocamentos, “drifts” e “trends” na linha base, de sinais analíticos. Este tipo de ruído pode ser minimizado instrumentalmente, acoplando-se a frequência da fonte de sinais com a frequência usada na detecção, através de um amplificador ““lock-in””. Também é possível minimizar ruídos de baixa frequência através de “software”, para tanto basta eliminar do sinal suas componentes de baixa frequência. Entretanto esta possibilidade é menos viável que a alternativa anterior pois a grande maioria dos sinais analíticos possuem informações úteis na região de frequências baixas, as quais seriam eliminadas conjuntamente com o ruído de baixa frequência.

Na classe de ruídos de frequência intermediária encontramos basicamente o ruído de interferência. O ruído de interferência mais comum é o ruído de interferência da linha de energia elétrica, com frequência fundamental da ordem de 60 Hz e suas harmônicas. Este tipo de ruído pode ser removido instrumentalmente, modulando-se a frequência da fonte de sinais e a frequência usada na detecção, longe da faixa de frequências associada ao ruído de interferência. Este tipo de ruído pode também, ser minimizado, com alguma facilidade, através de “softwares” projetados para esta finalidade. Veremos, claramente, como se faz isto, quando apresentarmos a teoria do filtro de transformada de Fourier.

Na última classe (classe dos ruídos de alta frequência), encontramos o tipo de ruído mais difícil de ser tratado: o ruído branco ou ruído gaussiano. A característica marcante deste tipo de ruído é a inexistência de uma frequência característica, mas uma distribuição de frequências características aleatoriamente espalhadas, sobre todas as frequências que compõem o sinal analítico. Este tipo de ruído origina-se no movimento térmico ou movimento Browniano dos elétrons em componentes de estado

sólido, presentes em instrumentos eletrônicos. Este tipo de ruído torna-se particularmente crítico em detectores de radiação eletromagnética, tanto que em alguns casos resfria-se o detector, na tentativa de reduzir este tipo de ruído. Entretanto, *gerar “frio”* é, geralmente, algo caro (proximadamente, cinco vezes mais caro que *gerar calor*) e pouco prático, especialmente para o caso de detectores. Além disso, como todos sabemos, é impossível atingir o zero absoluto (0 K) e, portanto, é impossível eliminar este tipo de ruído ou mesmo, reduzi-lo a níveis toleráveis, em boa parte das situações práticas. Entretanto, pode-se minimizar, em alguns casos, este tipo de ruído através de “softwares” que apliquem filtros de transformada de Wavelet ou filtros de transformada de Fourier, como descreveremos mais adiante, nesta tese.

3.1.1 A suavização de ruído instrumental: *O filtro de Savitzky-Golay e o filtro de média móvel.*

① *Filtro de média móvel*¹

O filtro de média móvel tornou-se, nos últimos anos, uma ferramenta de uso comum para a suavização de ruído em sinais analíticos. O motivo para tal popularidade deve-se à sua simplicidade teórica e a sua incorporação a diversos “softwares” comerciais, como por exemplo: Microcal Origin, Excel, Lotus, etc.

A idéia básica do método é dividir o sinal analítico em uma série de intervalos com a mesma largura, sequencialmente e ponto à ponto, e

substituir o valor do primeiro ponto do intervalo pelo valor médio do intervalo. O processo é executado na seguinte sequência:

① Define-se a origem do sinal ($x_{i=1}$).

② Define-se o número de pontos do intervalo (n).

③ Obtém-se a média dos pontos do intervalo: $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$

④ \bar{x} é primeiro ponto do novo sinal: x_1^{novo}

⑤ repete-se o processo (1 - 4) para $x_{i=2}$

⑥ Atualiza-se o valor de i e repete-se a etapa 5 até todo sinal ter sido suavizado.

O filtro de média móvel é simples de entender, fácil de implementar e de execução rápida. Entretanto, como nada é perfeito, os resultados que este tipo de filtro produz para sinais com picos estreitos é extremamente pobre. Tipicamente, este tipo de filtro distorce sinais analíticos, sendo esta distorção diretamente proporcional à largura do intervalo, como ilustrado para o diagrama da Figura 3.1.

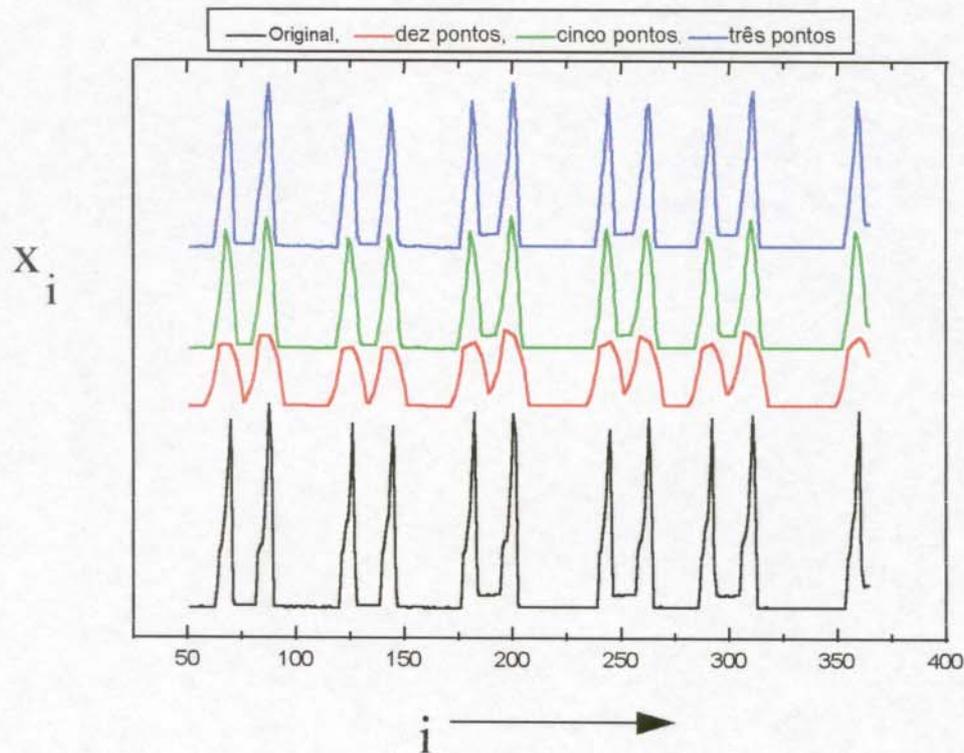


Figura 3.1. Representação do efeito da largura do intervalo usado no filtro de média móvel para a suavização de ruído em um diagrama.

O filtro de média móvel reduz ruído proporcionalmente à largura do intervalo mas, simultaneamente e proporcionalmente, distorce o sinal analítico. Assim sendo, a aplicação do filtro de média móvel fica redizida a sinais de banda extremamente larga, em que a largura do intervalo necessário a redução do ruído, não é crítica.

② O filtro de Savitzky-Golay^{32, 33}

O Filtro de Savitzky-Golay é um clássico entre os filtros de suavização de sinais analíticos e também encontra-se incorporado na maioria dos “softwares” comerciais para tratamento de dados. A idéia básica do método de Savitzky-Golay é a seguinte:

- ① Defini-se a origem do sinal
- ② Define-se a largura do intervalo
- ③ define-se o ponto central do intervalo
- ④ Ajusta-se através do método de mínimos quadrados, um polinômio de grau variável aos pontos do intervalo.
- ⑤ Utiliza-se o polinômio para estimar o valor do ponto em questão.
- ⑥ Desloca-se o intervalo para o ponto seguinte do sinal original e repete-se o processo anterior.

O filtro de Savitzky-Golay é menos sensível à largura do intervalo pois o novo ponto, isto é, o ponto do sinal já suavizado, é obtido através do ajuste dos pontos do intervalo por um polinômio e não pela simples média dos pontos vizinhos, como ocorre com o filtro de média móvel. Esta característica possibilita que o filtro de Savitzky-Golay seja aplicado a sinais analíticos com picos estreitos, apresentando resultados superiores aos obtidos com o filtro de média móvel. Entretanto, ainda se observa distorção do sinal à medida que o número de pontos utilizado no intervalo aumenta, conforme pode ser observado na Figura 3.2.

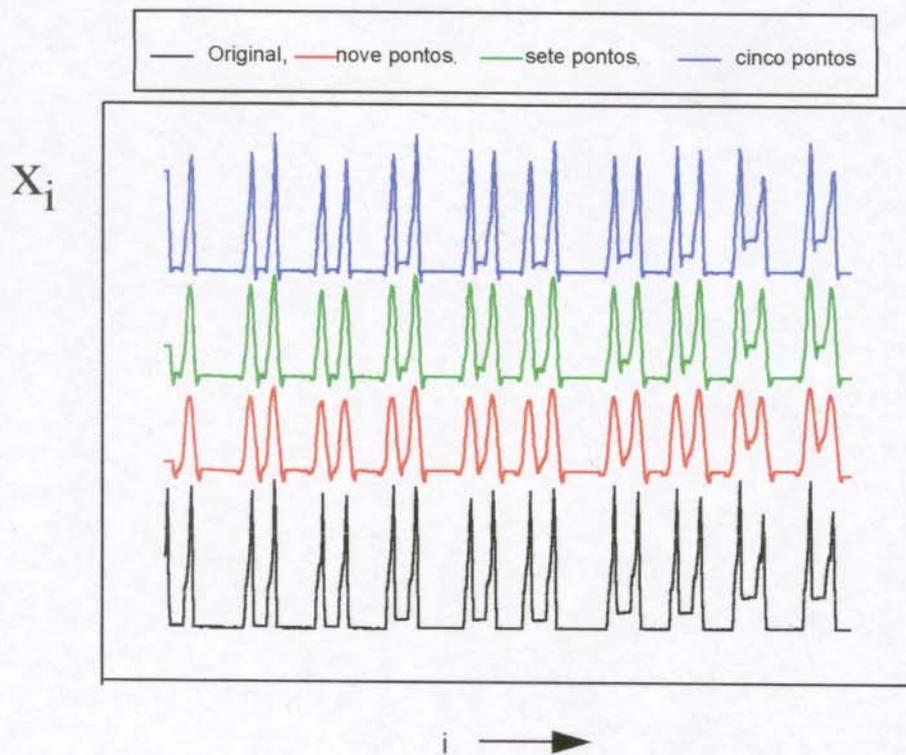


Figura 3.2. Representação do efeito da largura do intervalo usado na suavização de ruído com o filtro de Savitzky-Golay, em um fiagrama.

Conforme descrevemos, os filtros de suavização de ruído instrumental atuam diretamente sobre o sinal analítico e não sobre as frequências que compõem tal sinal. Esta característica torna os filtros de suavização pouco seletivos, isto é, incapazes de diferenciar o que é sinal do que é ruído e, portanto, no processo de suavização de ruído uma parte da informação útil contida no sinal é simultaneamente removida.

Felizmente, existem filtros que atuam diretamente sobre as frequências que compõem o sinal, possibilitando a remoção seletiva de ruído. Estes filtros são conhecidos como filtros de remoção de ruído. Os mais populares são os filtros baseados em transformada de Fourier e em transformada de Wavelet, conforme descreveremos no próximo item desta introdução.

3.1.2 A remoção de ruído instrumental: *O filtro de transformada de Fourier e o filtro de transformada de Wavelet.*

*O filtro de transformada de Fourier*³⁴⁻³⁷

A idéia básica deste tipo de filtro é aplicar-se a transformada de Fourier direta³⁸⁻⁴⁰, dada pela equação 3.1, para que se obtenha o sinal analítico representado no domínio das frequências (w), ou melhor, aplicamos a transformada de Fourier direta para obter o espectro de frequências $F(w)$, do sinal analítico.

$$F(w) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(\lambda) e^{i\lambda w} d\lambda \quad \text{Eq. 3.1}$$

Na equação 3.1, λ representa o domínio original do sinal analítico e $f(\lambda)$ o sinal analítico. Para espectros, λ representa os comprimentos de onda, para cromatogramas ou fiagramas, λ representa a variável tempo e assim por diante.

Uma vez obtido o espectro de frequências do sinal analítico, devemos cortar as frequências altas, visto serem estas frequências, na grande maioria dos casos, relacionadas ao ruído instrumental.

Finalmente aplicamos a transformada de Fourier inversa, dada pela equação 3.2 e recuperamos o sinal analítico inicial, livre de ruído.

$$f(\lambda) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} F(w) e^{-i\lambda w} dw \quad \text{Eq. 3.2}$$

A sequência de operações utilizada no processo de remoção de ruído, através da transformada de Fourier, podem ser facilmente entendidas se observarmos a Figura 3.3.

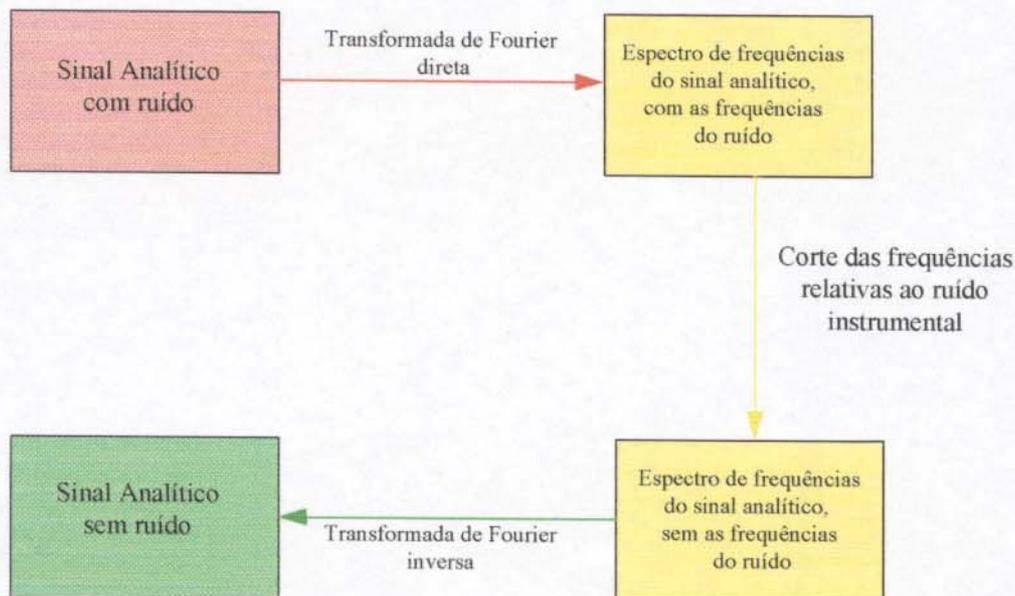


Figura 3.3. Representação esquemática da sequência de operações utilizada na aplicação do filtro de transformada de Fourier.

Para exemplificarmos a aplicação do filtro de transformada de Fourier escolhemos um sinal complexo, em que muitas frequências estão envolvidas. Este sinal não é proveniente de nenhum processo analítico,

mas permitirá ilustrar, de modo didático, alguns aspectos importantes do filtro de transformada de Fourier. Vejamos o exemplo.

① Sinal inicial

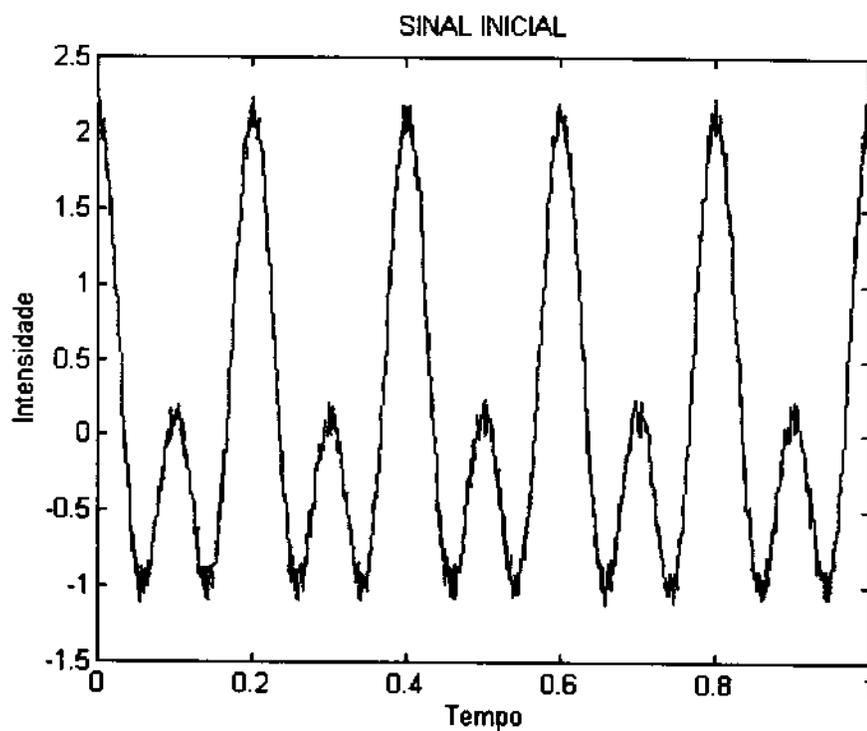


Figura 3.4. Sinal original, isto é, sinal + ruído.

② Espectro de frequências do sinal inicial

No espectro de frequências pode-se observar claramente que o sinal é formado pela seguintes frequências características:

- 5 Hz

-10 Hz

- Frequências altas, relativas a ruído

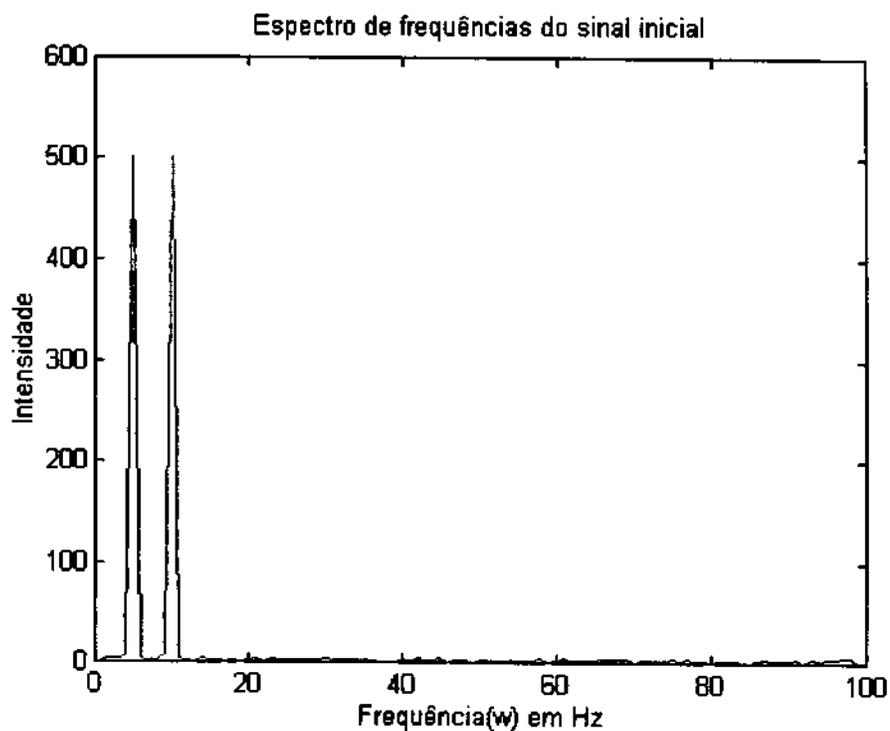


Figura 3.5. Espectro de frequências do sinal original.

③ Corte de frequências superiores a 10 Hz e retorno ao domínio do tempo.

Nesta etapa devemos cortar frequências superiores a 10 Hz. Para tanto, multiplicamos o espectro de frequências pelo valor 1, até a frequência 10 Hz, acima desta frequência multiplicamos por zero.

Efetuada o corte de frequências, aplicamos a transformada de Fourier inversa e obtemos o sinal livre de ruído.

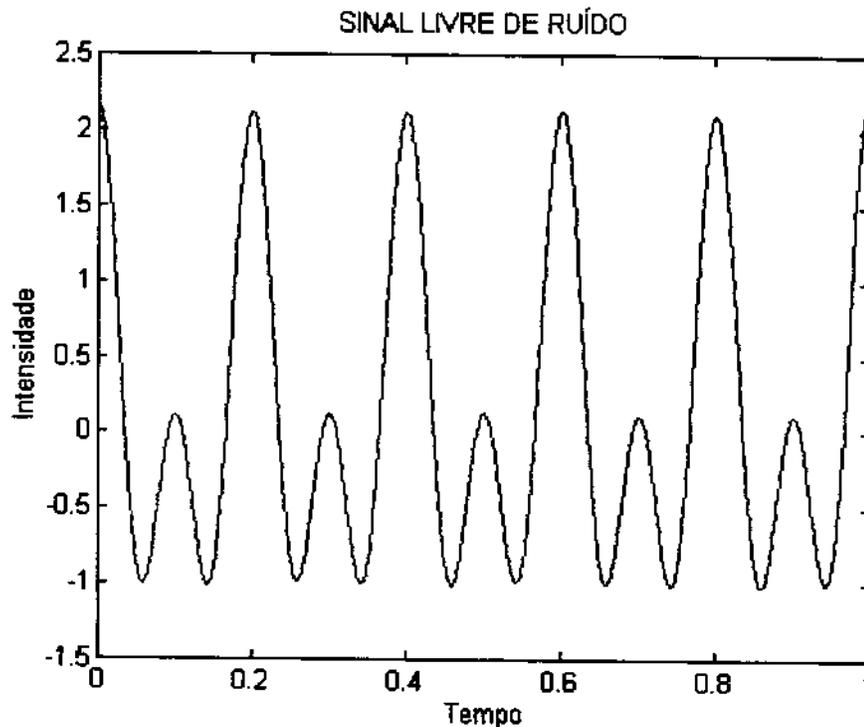


Figura 3.6. Sinal livre do ruído.

Uma vez apresentado os conceitos físicos e o método de funcionamento do filtro de transformada de Fourier, podemos apresentar a sua elegante definição matemática.

O filtro de transformada de Fourier pode ser entendido como a transformada de Fourier inversa (\mathcal{F}^{-1}), da transformada de Fourier direta (\mathcal{F}) de um sinal $f(\lambda)$ convolvido com uma função de apodização $h(\lambda)$. A transformada de Fourier direta de $f(\lambda)$ convolvido com $h(\lambda)$ é dada por:

$$F(S) = \mathfrak{F}(f(\lambda) * h(\lambda)) \quad \text{Eq.3.3}$$

Uma vez que a transformada de Fourier da convolução de duas funções é o produto das transformadas, podemos escrever:

$$F(S) = \mathfrak{F}(f(\lambda)) * \mathfrak{F}(h(\lambda)) \quad \text{Eq.3.4}$$

No domínio das frequências a convolução anterior (Eq. 3.4) se reduz à simples multiplicação das correspondentes transformadas de Fourier, como segue:

$$F(S) = F(w) H(w) \quad \text{Eq.3.5}$$

em que:

$$H(w) = \begin{cases} 1 & \text{se } w \leq w_{\text{critica}} \text{ (freq. de corte)} \\ 0 & \text{se } w > w_{\text{critica}} \end{cases}$$

aplicando a transformada de Fourier inversa à função $F(S)$

$$F(\lambda) = \mathfrak{F}^{-1}(F(S)) \quad \text{Eq.3.6}$$

obtemos o sinal $f(\lambda)$ livre de ruído.

Vale lembrar que o elegante desenvolvimento matemático é absolutamente idêntico ao “desenvolvimento didático”, utilizado no início da descrição do filtro de transformada de Fourier. A preferência por um dos dois desenvolvimentos fica reduzida à questões estéticas ou a idiossincrasias do leitor.

Enquanto nos filtros de suavização o fator crítico é a largura do intervalo utilizado na suavização, nos filtros de transformada de Fourier, o fator crítico é a frequência de corte. Se o corte de frequências for efetuado em uma posição errada, corre-se o risco de eliminar informação útil do sinal. Na Figura 3.7 ilustramos o que ocorreria se, por algum motivo, fosse cortado do espectro de frequências da Figura 3.5, todas frequências acima de 5 Hz e em seguida fosse aplicada a transformada de Fourier inversa.

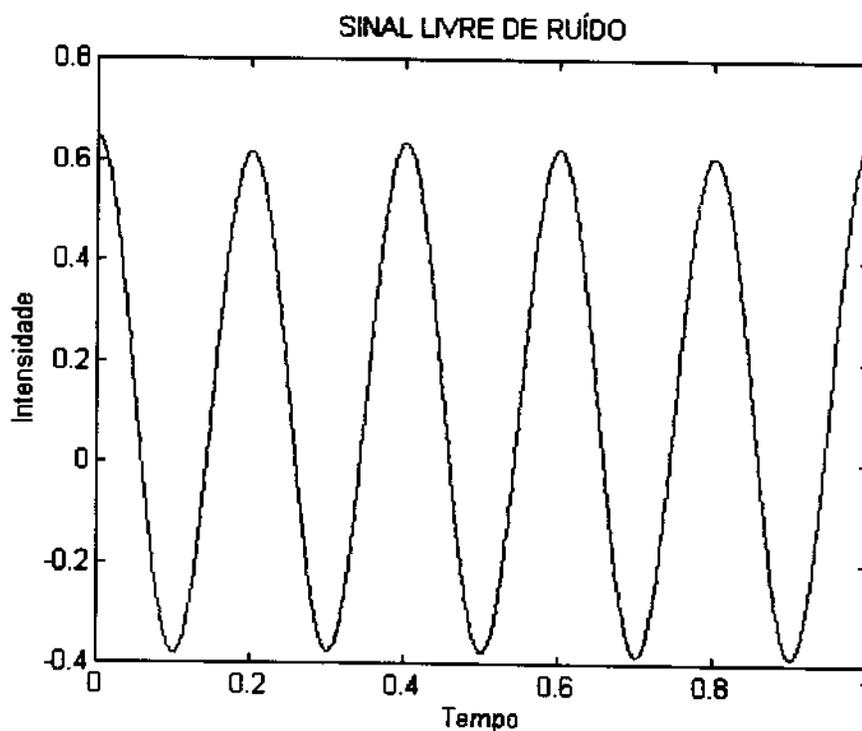


Figura 3.7. Sinal distorcido em função de um corte errado nas frequências.

Conforme pode ser observado na Figura 3.7, uma escolha errada da frequência crítica de corte, produz uma distorção catastrófica no sinal. O sinal tratado com o filtro de transformada de Fourier, neste caso, não apresenta nenhuma relação com o sinal original.

Felizmente, a escolha da frequência crítica é geralmente uma tarefa muito simples pois o ruído geralmente apresenta frequências altas e o sinal útil, frequências baixas. Na dúvida, devemos efetuar o corte bem longe das frequências baixas. Neste caso, talvez não estejamos cortando todo o ruído, mas certamente uma parcela será eliminada e não corremos o risco de distorcer o sinal.

O grande problema do filtro de transformada de Fourier certamente não é a escolha da frequência crítica, mas a sua aplicação a sinais não estacionários, isto é, sinais cujas frequências características variam com tempo (comprimento de onda para espectros, potencial para voltamogramas, etc).

Imagine dois sinais completamente diferentes, mas que apresentem espectros de frequências idênticos. Por exemplo, um sinal pode ser composto por duas frequências combinadas durante todo o intervalo de tempo, um segundo sinal pode possuir as mesmas duas frequências entretanto, estas podem aparecer, no sinal, em tempos diferentes. Os sinais seriam totalmente diferentes mas seus espectros de frequência, idênticos. Na Figura 3.8 podemos observar um sinal (sinal 1) com duas frequências distribuídas sobre todo o sinal. Na Figura 3.9 representamos um segundo sinal (sinal 2) formado pelas mesmas frequências do sinal anterior mas divididas em dois intervalos: no primeira metade do sinal temos somente a frequência 10 Hz e na segunda metade a frequência 20 Hz.

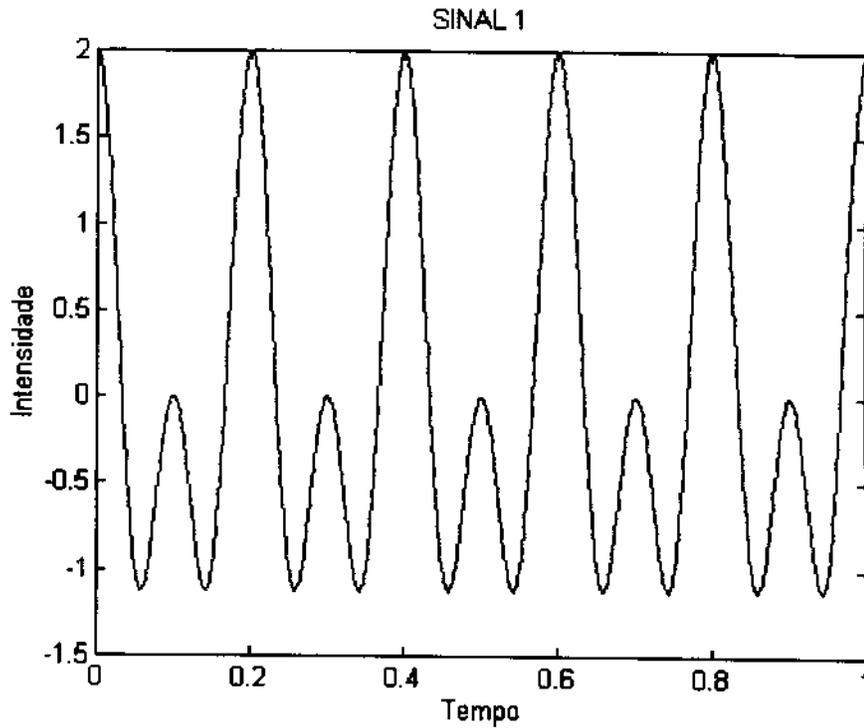


Figura 3.8. Sinal com duas frequências (10Hz e 20Hz), distribuídas sobre todo intervalo de tempo.

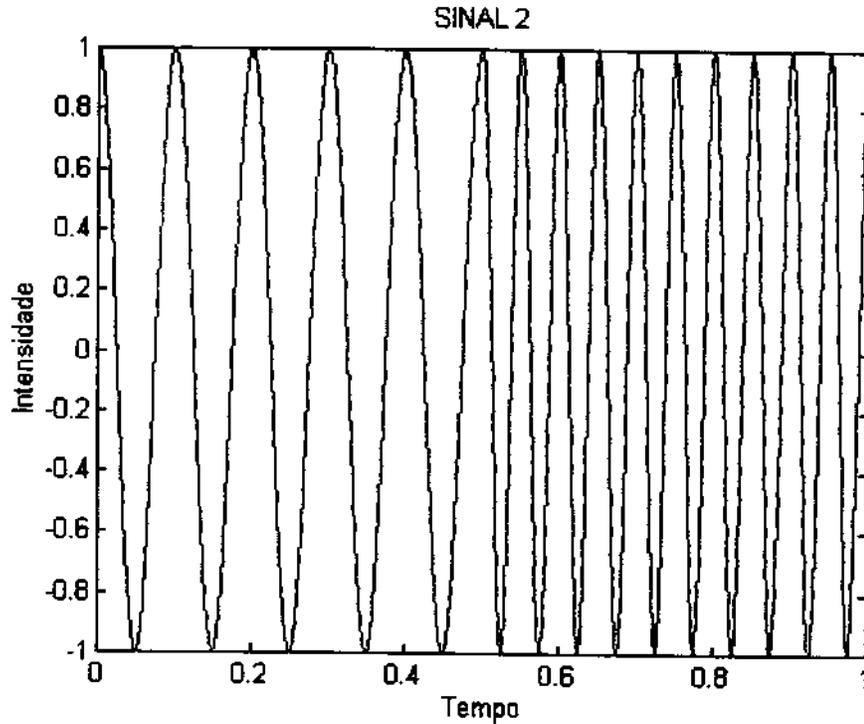


Figura 3.9. Neste sinal temos uma frequência (10 Hz) na primeira metade do sinal e outra frequência (20 Hz) na segunda metade .

Nas Figuras 3.10 e 3.11, representamos os espectros de frequência dos sinais 1 e 2, respectivamente. Observe que os espectros de frequência dos dois sinais apresentam, basicamente, as mesmas frequências características. Este fato nos levaria à uma interpretação errada dos sinais pois seríamos tentados a pensar que os sinais são, no mínimo, muito parecidos. Entretanto, os sinais são absolutamente diferentes.

O resultado prático do fato descrito no parágrafo anterior é que se aplicarmos o filtro de transformada de Fourier a um sinal não estacionário, no momento em que aplicarmos a transformada inversa podemos recuperar um sinal com características diferentes do sinal inicial.

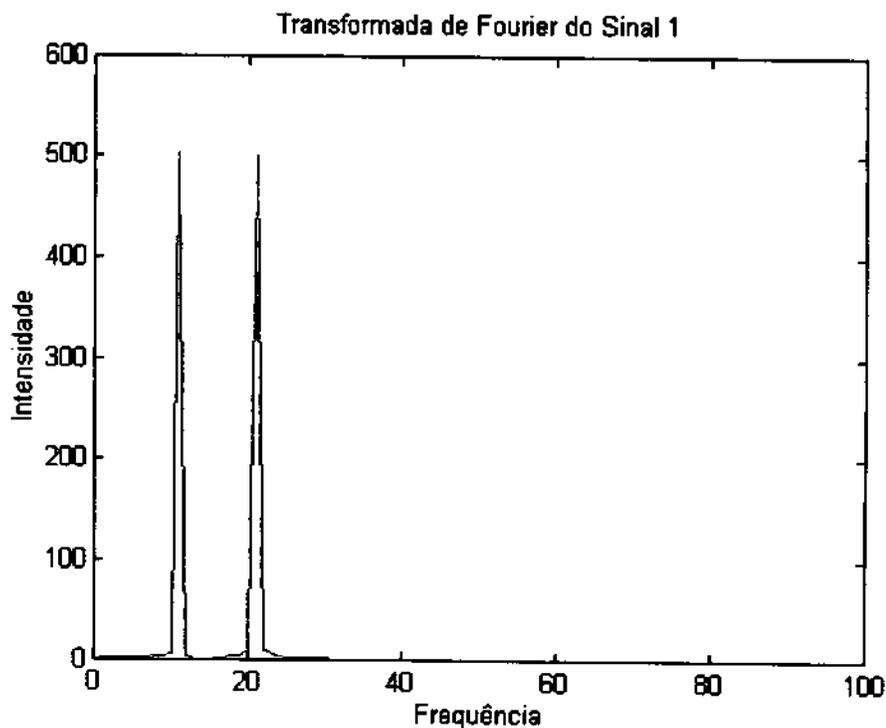


Figura 3.10. Transformada de Fourier do sinal 1.

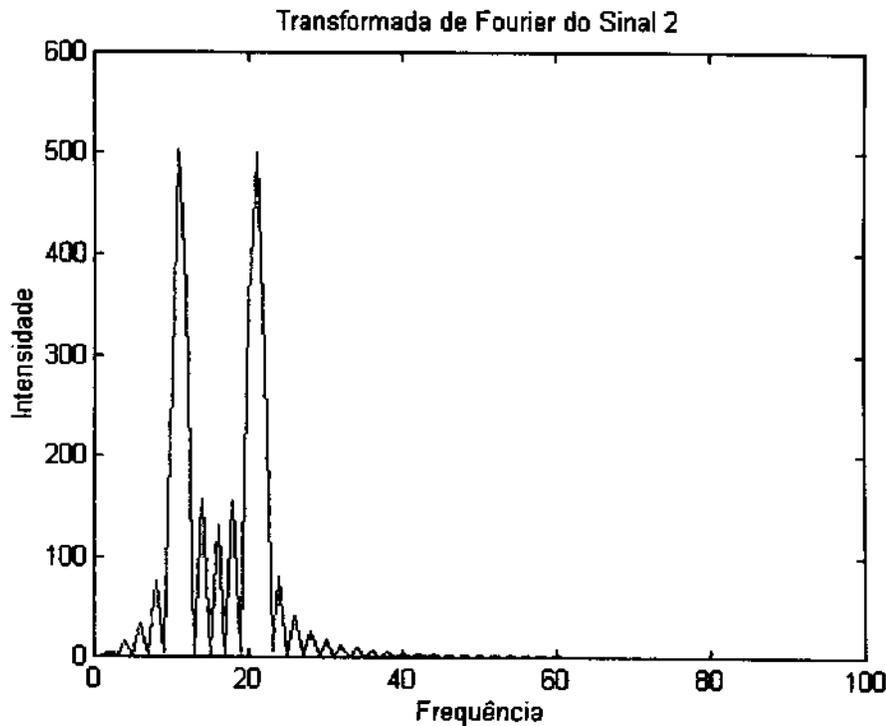


Figura 3.11. Transformada de Fourier do sinal.

Este fato ocorre porque a transformada de Fourier atua no sinal como um todo e não sobre segmentos, de modo a discriminar os tempos em que as frequências aparecem. A transformada de Fourier atua no sinal todo porque as funções utilizadas na decomposição do sinal para a obtenção do espectro de frequências, são funções seno e cosseno que à rigor, estendem-se de menos infinito até mais infinito, isto é, a transformada de Fourier utiliza para a decomposição do sinal, funções que não possuem suporte compacto no domínio do tempo. Observe novamente a equação utilizada na transformada de Fourier direta, reescrita abaixo, e o fato da falta de suporte compacto, tornar-se-á óbvio.

$$F(w) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(\lambda) e^{-i\lambda w} d\lambda \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(\lambda) (\cos(\lambda w) + i \operatorname{sen}(\lambda w)) d\lambda$$

Eq. 3.7

A solução para o problema da falta de suporte compacto das funções seno e cosseno, é aplicar a transformada de Fourier a segmentos do sinal e não diretamente sobre o sinal todo, num procedimento conhecido como transformada de Fourier de pequenos intervalos³¹ (STFT, *short time Fourier transform*). Quando bem aplicada, a STFT fornece informações ‘exatas’ sobre os tempos em que as frequências ocorrem. Assim sendo, devemos utilizar a STFT quando desejarmos aplicar o filtro de transformada de Fourier a sinais não estacionários. A única desvantagem prática deste procedimento é o aumento do tempo computacional. Matematicamente a STFT pode ser escrita como segue:

$$STFT \Rightarrow F(w) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (f(\lambda) l(\lambda_i, \lambda_f)) e^{i\lambda w} d\lambda \quad \text{Eq.3.8}$$

Na equação 3.8, o termo $l(\lambda_i, \lambda_f)$ representa a largura do intervalo utilizado na STFT.

Quando aplicamos a STFT para sinais não estacionários que mudam suas frequências em intervalos de tempo muito rápidos, a largura do intervalo deve ser inversamente proporcional à mudança de frequências do sinal, isto é, quanto mais rapidamente o sinal muda de frequência, menor deve ser a largura do intervalo usado na STFT. Neste ponto, nos deparamos com um problema teórico muito interessante, qual seja: até quando é possível reduzir a largura do intervalo utilizado na STFT, ou

melhor, com que exatidão, no domínio do tempo, as frequências podem ser determinadas em um sinal não estacionário ?

A resposta para questão anterior é dada pelo princípio da incerteza de Heisenberg^{41, 42}. O princípio da incerteza de Heisenberg mostra que não é possível conhecer exatamente e simultaneamente, tempo (t) e frequência (w), mas somente a seguinte relação entre estas grandezas:

$$\Delta w \Delta t \geq \frac{1}{4\pi}$$

Este fato, nos leva ao seguinte problema de resolução: quanto menor o intervalo de tempo utilizado na STFT, pior a resolução nas frequências e vice-versa.

Assim sendo, a transformada de Fourier nos fornece boa informação sobre as frequências de um sinal, mas não fornece informações exatas sobre o tempo em que estas frequências ocorrem. Já a STFT nos fornece boa informação sobre o tempo em que as frequências ocorrem mas não fornece informações exatas sobre as frequências. As características dos dois tipos de transformada de Fourier estão resumidas na Tabela 1.

Tabela 1. Desempenhos da STFT e da transformada de Fourier.

	Resolução	
	Tempo	Frequência
Transformada de Fourier	Pobre	Boa
STFT	Boa	Pobre

Apesar das dificuldades anteriormente descritas, os filtros de ruído fundamentados em transformada de Fourier (FT ou STFT) apresentam bons resultados para a maioria dos sinais encontrados na prática e são

considerados como técnica padrão, para a remoção de ruído, em diversas áreas da ciência e da engenharia.

O filtro de transformada de Wavelet⁴³⁻⁵²

Com o objetivo de superar os problemas de resolução apresentados pela transformada de Fourier e pela transformada de Fourier de pequenos intervalos (STFT), foi desenvolvida no início da década de setenta, por Ingrid Daubechie⁵³ no AT&T Bell Laboratórios, a transformada de Wavelet.

A transformada de Wavelet é uma transformação linear que analogamente à transformada de Fourier, nos permite representar um sinal no domínio das frequências que o compõem. No domínio das frequências também é possível cortar frequências e através da transformada de Wavelet inversa, recuperar o sinal livre de ruído.

A única diferença entre as transformadas de Fourier e de Wavelet são as funções de base utilizadas na decomposição do sinal. Nas transformadas de Fourier as funções de base para a transformação de domínio são as familiares funções seno e cosseno, as quais apresentam os problemas de falta de suporte compacto, anteriormente mencionados. Para a transformada de Wavelet, as funções de base são muito menos familiares e chamadas de *wavelets* (ondinhas). Estas funções de base, as wavelets, apresentam como principal característica, suporte altamente compacto, isto é, as wavelets não se estendem de $-\infty$ até $+\infty$ no domínio do tempo, mas possuem um ponto de início e um ponto de término muito bem definidos.

Ao contrário da transformada de Fourier, em que temos um único tipo de funções de base (senos e cossenos), na transformada de wavelet

temos vários tipos diferentes de funções de base, isto é, vários tipos diferentes de wavelets, divididas em famílias. As famílias de wavelets mais comuns são as seguintes: *Daubechies*, *Morlets*, *Coiflets*, *Haar* e as *Symmlets*. Um exemplo clássico de função de base utilizada em transformadas de Wavelet é a *wavelet* da família das *Morlets*, dada pela equação 3.9 e representada na Figura 3.12.

$$g = e^{i\omega t} e^{-t^2/2} \quad \text{Eq. 3.9}$$

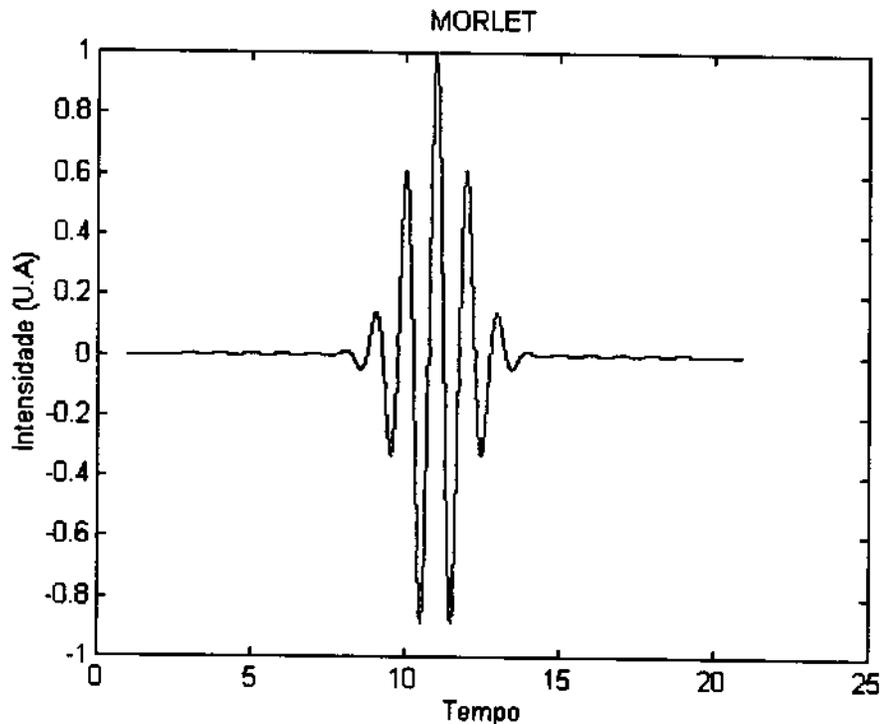


Figura 3.12. Função de base (wavelet) de MORLET.

Na Figura 3.12 podemos observar claramente que as wavelets possuem suporte altamente compacto, isto é, alta localização no eixo do tempo. Observe que a wavelet da Figura 3.12, oscila em torno do valor zero e devido a este fato estas funções de base foram denominadas wavelets, isto é, *ondinhas*.

A idéia básica do filtro de transformada de Wavelet é fazer a convolução do sinal inicial (sinal + ruído), $x(\lambda)$, com uma wavelet (uma morlet, por exemplo), $g(\lambda)$, cuja transformada de Fourier $\hat{g}(\omega)$ apresente valor zero acima de uma certa frequência crítica ($w_{critica}$), como segue:

$$W(\nu) = \int_{-\infty}^{\infty} x(\lambda) * g(\lambda) d\lambda \quad \text{Eq. 3.10}$$

Uma vez que as *wavelets* utilizadas em filtros de ruído instrumental, são compostas exclusivamente por frequências baixas, isto é, apresentam frequência crítica baixa, ao fazermos a convolução da equação 3.10 automaticamente cortaremos frequências altas, eliminando simultaneamente o ruído instrumental.

Uma função de base pode ser comprimida (por exemplo, período reduzido a metade, frequência crítica dobrada), no domínio tempo (comprimento de onda para espectros, tempo para cromatogramas, diferença de potencial para voltamogramas, etc.), de modo que podemos representar um mesmo sinal em escalas cada vez menores, removendo ruído sequencialmente em cada uma destas escalas. Além disso a wavelet é transladada através do sinal, permitindo remover ruído de sinais estacionários ou não estacionários. A transformada de Wavelet seria análoga a uma STFT em que se pudesse reduzir indefinidamente a largura do intervalo de análise do sinal (janela), sem violar o princípio da incerteza de Heisenberg. Entretanto, a relação entre a resolução no domínio do tempo ou das frequências é um jogo em que o *Croupier* é o princípio da

incerteza de Heisenberg, que independentemente dos jogadores, isto é, independente do tipo de transformada utilizada, fará valer suas regras. Assim sendo, se nas transformadas de Wavelet, devido à possibilidade de compressão das *wavelets*, é possível obter alta resolução no domínio do tempo, o princípio da incerteza nos diz que 'algo' deve ser perdido. De fato, à medida que comprimimos uma *wavelet* ganhamos resolução no tempo, mas simultaneamente, aumentamos sua frequência de corte. Portanto, só é possível cortar frequências altas nos sinais em que estas frequências, estejam limitadas a pequenos intervalos de tempo. Resumindo, se uma wavelet for comprimida, ganhamos resolução no tempo, mas devido a alteração na frequência da *wavelet*, perdemos precisão na frequência crítica de corte utilizada no corte de frequências altas. Portanto, o uso da transformada de Wavelet apresentará melhores resultados quando o sinal em questão possuir componentes de alta frequência de curta duração e componentes de baixa frequência de longa duração. Felizmente, a maioria dos sinais encontrados na prática possuem as características anteriores.

As *wavelets* geradas por compressão e translação de uma *wavelet* inicial (*wavelet* mãe), podem então, ser escritas como⁵²:

$$g_{a,b}(\lambda) = \frac{1}{\sqrt{a}} g\left(\frac{\lambda - b}{a}\right) \quad \text{Eq. 3.11}$$

Na Figura 3.13, representámos uma sequência de wavelets de Morlet, com diferentes parâmetros de escala (a), isto é, com diferentes frequências características.

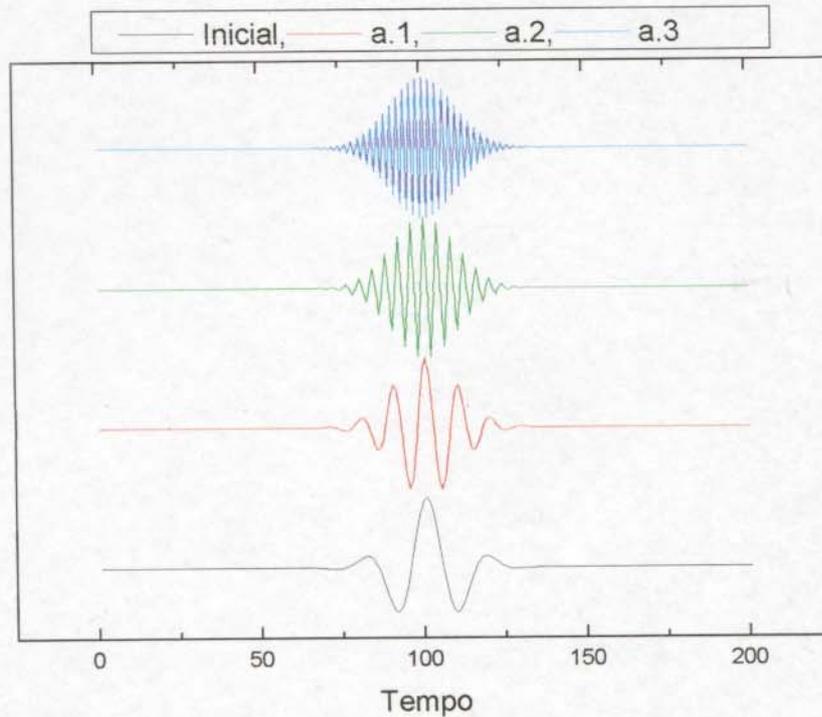


Figura 3.13. wavelets de Morlet com diferentes valores de escala (a).

Naturalmente, a transformada de wavelet torna-se função dos parâmetros de escala (a e b) e deve ser reescrita como segue:

$$W_f(a, b) = \int_{-\infty}^{\infty} x(\lambda) g_{ab}(\lambda) d\lambda \quad \text{Eq. 3.12}$$

em que a representa a escala (compressão) e b o deslocamento.

Assim sendo, cada valor de a define um tipo de filtro, isto é, um tipo de corte nas frequências. Na Figura 3.14, representamos a transformada de Wavelet como um banco de filtros^{31, 54}, sendo cada tipo de filtro responsável pela extração de um tipo de informação, do sinal original.

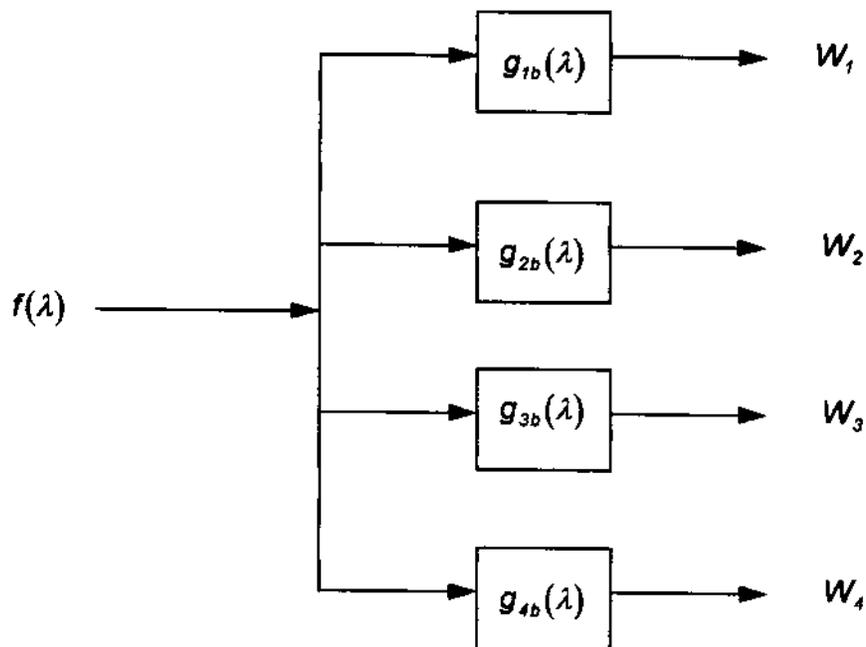


Figura 3.14. Representação esquemática de um banco de filtros de *wavelets*.

Por exemplo, a *wavelet* mais comprimida, deve remover o ruído do sinal. A segunda *wavelet* deve extrair uma segunda frequência que compõe o sinal. Frequência esta que pode ser associada a alguma propriedade de interesse, embutida no sinal e assim por diante.

O sinal inicial, livre de ruído, pode ser recuperado aplicando-se a transformada de Wavelet inversa relativa à *wavelet* mais comprimida, dada pela equação 3.13.

$$x(\lambda) = \int_0^{\infty} \int_{-\infty}^{\infty} W_r(a, b) g_{ab}(\lambda) db \frac{da}{a^2} \quad \text{Eq. 3.13}$$

Através da transformada inversa, além do sinal livre de ruído, também é possível obter a representação, no domínio inicial, das características

capturadas por outras *wavelet*, do banco de filtro. Para tanto, basta utilizar na transformada inversa a *wavelet* correspondente à característica desejada.

O processo de obter informações sobre um sistema, em diferentes escalas é uma nova alternativa para análise de sinais, conhecida como *análise em múltiplas escalas*⁵⁵ (MRA). A *análise de sinais em múltiplas escalas* é análoga a observação de um mapa em diferentes escalas, grandes escalas correspondem a uma visão não detalhada (continentes, por exemplo) , pequenas escalas correspondem á uma visão mais detalhada de uma determinada região (limites de uma cidade, por exemplo). Para sistemas fractais, por exemplo, a MRA apresentará sempre as mesmas características, independentemente da escala utilizada.

O campo das Transformadas de Wavelet é relativamente novo (cerca de 10 anos) e ainda não existem textos com boas características didáticas sobre o assunto. Além disso, a maioria dos textos é dirigido a matemáticos e talvez (ou melhor), certamente ainda não dominei, nem descrevi, todos os aspectos teóricos do assunto. Entretanto, para a maioria das aplicações em química analítica, os detalhes teóricos não são necessários e a introdução anterior é suficiente.

3.2 Correção de espalhamento de luz^{56,57} (MSC)

Uma das principais fontes de erro encontrada em determinações quantitativas, através da espectroscopia de reflectância difusa no infravermelho próximo, é o espalhamento de luz. O espalhamento de luz altera a relação funcional entre a intensidade das medidas de reflectância e

concentração das espécies absorventes presentes na amostra. A reflectância torna-se função não somente da concentração das espécies absorventes, mas também da não homogeneidade óptica da amostra, devida basicamente a diferenças de granulometria, geometria, empacotamento e orientação das partículas nas amostras, conforme descrito no capítulo 2 desta tese. Uma vez que, em determinações quantitativas, a intensidade da reflectância é a variável dependente, que deve ser modelada em função da concentração das espécies absorventes, temos duas opções: modelar simultaneamente reflectância e espalhamento de luz, ou minimizar o espalhamento de luz e em seguida construir um modelo para a relação entre a reflectância e a concentração das espécies absorventes, presentes na amostra.

Modelar, simultaneamente espalhamento e reflectância é uma tarefa extremamente difícil (impossível em alguns casos), pois a geometria e a orientação das partículas variam aleatoriamente, de amostra para amostra. Assim, se desejarmos construir modelos precisos e robustos, só nos resta a segunda opção, qual seja: minimizar o espalhamento de luz.

O método MSC (*Multiplicative Scattering Correction*) é uma técnica para corrigir o efeito do espalhamento de luz em espectros de reflectância difusa. O fundamento do método, está no fato de que a dependência funcional, em relação ao comprimento de onda, do espalhamento de luz é diferente daquela observada para a reflectância difusa. Assim, é possível utilizar um grande número de comprimentos de onda para diferenciar o espalhamento de luz da reflectância difusa.

No método MSC a interferência relativa ao espalhamento de luz é minimizada corrigindo-se os espectros de reflectância difusa de modo que todos os espectros possuam o espalhamento ‘ ideal ‘ e por ideal entenda-

se médio, visto ser este o único espalhamento possível de ser calculado. A técnica é muito simples e pode ser facilmente entendida se acompanharmos a sequência de operações abaixo.

Matematicamente (e resumindo), a correção é feita da seguinte forma:

1- A partir do conjunto total de espectros de reflectância difusa, calcula-se o espectro médio \bar{x} .

2- Faz-se a regressão linear para cada um dos k espectros (x_{ik}), do conjunto total de espectros, contra o espectro médio, sobre todos os i comprimentos de onda:

$$x_{ik} = u_k + v_k \bar{x}_i \quad \text{eq.3.14}$$

3- Correção final:

$$x_{ik}^{(\text{corrigido})} = \frac{(x_{ik}^{(\text{nao corrigido})} - u_k)}{v_k} \quad \text{eq.3.15}$$

O método MSC, quando bem aplicado, é capaz de reduzir significativamente os efeitos relativos ao espalhamento de luz. Tipicamente, obtém-se uma melhora expressiva no desempenho dos modelos contruídos para a relação existente, entre a intensidade de reflectância e a concentração das espécies absorventes.

3.3 Compressão de dados: *A análise de componentes principais (PCA) e Transformada de Fourier.*

Os instrumentos analíticos modernos nos permitem medir simultaneamente, de modo rápido e eficiente um grande número de parâmetros de um determinado sistema químico. Esta característica, extremamente útil e desejável, pode tornar-se um sério problema pois um aumento no número de dados, não representa necessariamente um maior conhecimento sobre o sistema, uma vez que nem todos os dados possuem informações relevantes. Além disso, o excesso de dados gera dificuldades computacionais relacionadas ao armazenamento e processamento destes dados. Assim, torna-se necessário a utilização de métodos matemáticos que nos permitam comprimir ou reduzir o conjunto de dados obtidos em determinado experimento, à uma coleção muito menor, mas que ainda retenha as informações realmente úteis, para a análise e modelamento do sistema. Existe uma série de métodos destinados à compressão de dados, entretanto, os mais utilizados são métodos fundamentados na transformada de Fourier e o método da análise de componentes principais (PCA).

① *Compressão de sinais através da transformada de Fourier*

Na compressão de dados pela técnica da transformada de Fourier, a idéia básica é utilizar o espectro de frequência ao invés do sinal analítico propriamente dito. No espectro de frequência é possível cortar frequências relativas ao ruído instrumental e frequências relativas a informações sem importância, reduzindo sua dimensão. O espectro de frequência assim tratado, pode ser armazenado ocupando menos espaço. O sinal analítico

pode ser resgatado com grande facilidade, através da transformada de Fourier inversa. O método da transformada de Fourier encontra grande aplicação na área de compressão e tratamento de imagens. O famoso algoritmo JPEG (Joint Photographics Experts Group), utilizado para compressão de imagens, aplica o método da transformada de Fourier (em duas dimensões), em uma de suas principais etapas. Com o algoritmo JPEG é possível alcançar taxas de compressão da ordem de 45%, sem distorção significativa da imagem.

② *Compressão de sinais através da análise de componentes principais (PCA)* ^{15, 26, 38, 59}.

“ A idéia básica da análise de componentes principais é encontrar combinações lineares das variáveis independentes, de modo a reduzir sua dimensão a uma coleção muito menor de parâmetros, que ainda descreva suas principais características. Na análise de componentes principais, a matriz das variáveis independentes é decomposta em uma soma de matrizes menores, que não podem mais ser reduzidas, mais uma matriz de erros, como segue:

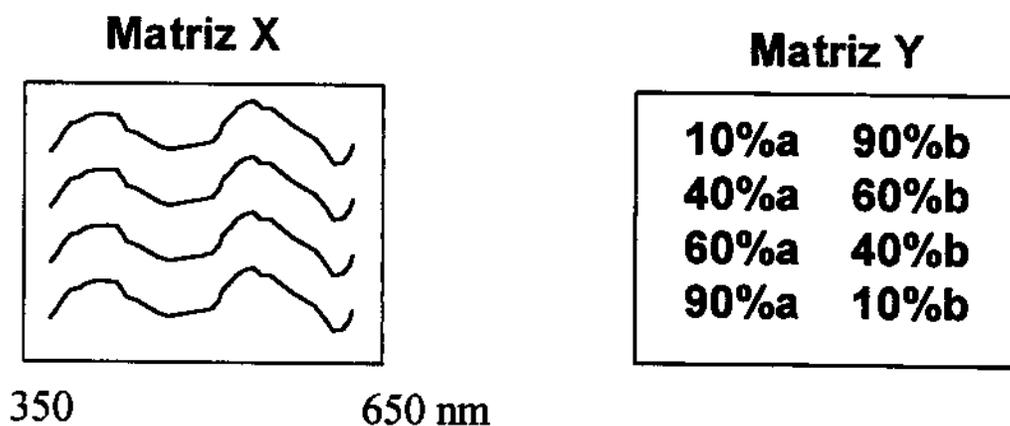
$$X' = t_1 p_1^T + t_2 p_2^T + \dots + t_k p_k^T + E \quad \text{eq. 3.16}$$

em que E é uma matriz de erros, t_i e p_i são os escores e loadings, respectivamente, da matriz das variáveis independentes”.

Esta é a definição que se encontra em boa parte dos livros e artigos sobre calibração multivariada. A definição anterior está totalmente correta, mas é quase impossível de entender, pelo menos para a grande maioria dos

mortais. Acompanhe o exemplo abaixo, e tudo se tornará claro (*pelo menos teoricamente!!!*).

Imagine que você deseja determinar simultaneamente dois íons metálicos, utilizando um complexante cromogênico inespecífico, através de espectroscopia de absorção, na região do visível (400 - 700 nm). Para tanto, você prepara diversas misturas, com diferentes proporções dos íons metálicos e adiciona o complexante cromogênico inespecífico. Em seguida, você obtém para cada uma das misturas, o respectivo espectro na região do visível. Executado o procedimento experimental anterior, você poderá montar a seguinte matriz de dados:



Na matriz *X* representamos os espectros na região do visível, cada uma das linhas (*m*) corresponde ao espectro de uma mistura e as colunas (*n*) são as absorbâncias lidas em diferentes comprimentos de onda. Na matriz *Y*, temos as respectivas proporções dos íons metálicos, presentes nas misturas.

No passo seguinte devemos centrar as colunas da matriz X na média para obter \bar{X} (basta subtrair de cada elemento da coluna, o valor médio da coluna toda) e calcular a covariância de \bar{X} , através da equação 3.17.

$$Cov(\bar{X}) = \frac{\bar{X}^T \bar{X}}{m-1} \quad \text{Eq.3.17}$$

Agora devemos obter a direção de maior variação da matriz $Cov(\bar{X})$. Isto é feito resolvendo-se a seguinte equação de autovalores e autovetores:

$$Cov(\bar{X}) p_i = \lambda_i p_i \quad \text{Eq.3.18}$$

em que λ_i é o autovalor associado ao autovetor p_i .

A esta altura, o leitor deve estar se perguntando: porque se utiliza a matriz covariância e não a matriz original? Qual o significado da covariância de uma variável com ela mesma? A resposta para estas questões é a seguinte: a matriz de covariância fornece informações de quanto uma absorvância lida em um determinado comprimento de onda se relaciona à uma absorvância lida em outro comprimento de onda. Se uma absorvância lida em certo comprimento de onda se relaciona muito com a absorvância lida em outro comprimento de onda, temos uma indicação de que estas duas absorvâncias podem ser substituídas por um valor correspondente à combinação linear delas. Naturalmente se observamos que um grupo de comprimentos de onda possuem alta covariância, podemos substituir o grupo todo pela combinação linear de suas absorvâncias. A maneira matemática de se obter estes grupos de alta

covariância é resolver a equação 3.18, obtendo seus autovalores e em seguida seus autovetores

Tipicamente, a equação 3.18 apresenta como solução uma série de autovalores, a partir dos quais obtém-se os respectivos autovetores. O primeiro autovetor, correspondente ao maior autovalor, fornece a direção no espaço das variáveis (colunas de X) que descreve a maior porcentagem da variância das amostras. Nos casos em que toda a variância do sistema não puder ser descrita usando um único autovetor, um segundo autovetor, ortogonal ao primeiro, pode ser obtido e assim por diante até que toda a variância das amostras seja capturada. Nesta etapa é possível utilizar um número de autovetores (ou autovalores) muito menor que o número de colunas da matriz X (número de comprimentos de onda) e ainda assim capturar a maior parte da variância dos dados, resultando na compressão dos dados.

Uma vez obtidas as direções de maior variância das amostras, o próximo passo é projetar a matriz X nesta nova base, como segue:

$$X p_i = t_i \quad \text{Eq. 3.19}$$

Na análise de componentes principais, os termos t_i são chamados de escores.

Resolvendo a equação 3.19, através do método dos mínimos quadrados, obtemos:

$$X = t_i p_i^T (p_i p_i^T) \quad \text{Eq. 3.20}$$

em que os elementos de:

$$P_i^T (P_i P_i^T)$$

são chamados, na análise de componentes principais, de loadings.

Assim, demonstramos que a matriz inicial pode ser representada através de combinações lineares de escores (t_i) e loadings (p_i), conforme descrevemos no início. Além disso, é possível utilizar um número menor de loadings e escores para representar a matriz X , possibilitando a compressão dos dados. O número de escores necessário para a descrição da matriz X é escolhido em função da variância capturada pelo autovalor e geralmente reduz severamente a dimensão do sistema.

Os escores são muito úteis, pois podemos relacioná-los com a concentração das espécies (**a** e **b**) nas misturas do nosso exemplo, em um típico procedimento de calibração, com a vantagem de que tal relação é feita para um número reduzido de escores e não para todas as absorvâncias lidas em todos os comprimentos de onda do espectro. Trocando em miúdos: “Para todos os efeitos, os escores podem ser usados como *absorvâncias* em procedimentos de calibração”. A maioria dos métodos de calibração multivariada utilizam os escores ao invés das absorvância, por exemplo: no método de calibração multivariada PCR^{58,59} (regressão dos componentes principais) uma relação linear entre os escores da matriz das variáveis independentes e as concentrações da matriz das concentrações, é estabelecida. No método dos mínimos quadrados parciais, obtém-se os escores da matriz X e da matriz Y , e a calibração é estabelecida entre os

escores de X e os escores de Y . Em nosso exemplo, poderíamos utilizar o PCR, para estabelecer a relação entre os escores dos espectros e a concentração das espécies **a** e **b**, presentes nas misturas.

Resumindo, os escores são extremamente versáteis e podem ser usados, com vantagens, em substituição às absorvâncias (ou outra variável) em qualquer procedimento de calibração multivariada.

CAPÍTULO 4

TÉCNICAS QUIMIOMÉTRICAS DE MODELAMENTO DE DADOS

4. Técnicas quimiométricas de modelamento de dados

4.1 Redes Neurais ⁶⁰

Recentes avanços em neurofisiologia têm desvendado vários mecanismos sobre o fluxo e o processamento de informações que ocorrem no cérebro humano. Alguns destes mecanismos foram modelados matematicamente permitindo a elaboração de algoritmos computacionais que simulam, ainda que de modo simplificado, a mais básica das estruturas cerebrais: *o neurônio* ⁶¹.

A capacidade de implementar computacionalmente versões simplificadas de neurônios biológicos deu origem a uma subespecialidade da inteligência artificial, conhecida como redes neurais. Existem várias definições para redes neurais⁶², portanto adotaremos aquela que nos parece mais geral e aplicável a qualquer área da ciência.

“Redes neurais é o nome dado a um conjunto de métodos matemáticos e algoritmos computacionais especialmente projetados para simular o processamento de informações e aquisição de conhecimento do cérebro humano.”

Operacionalmente podemos considerar uma rede neural como uma “caixa de processamento” que pode ser treinada para que, a partir de um conjunto de dados de entrada (inputs), possa gerar uma ou mais saídas (outputs), conforme representado na Figura 4.1.

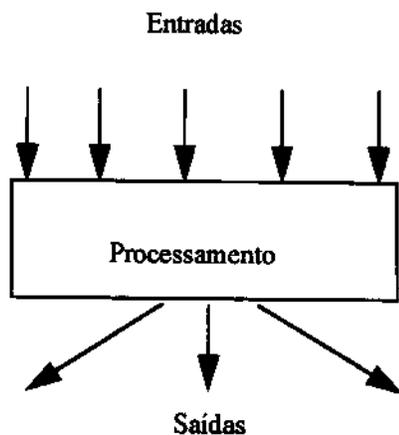


Figura 4.1. Representação operacional de uma rede neural.

Por exemplo, pode-se treinar uma rede neural para que a partir do espectro de RMN-H (input) de um composto orgânico ela apresente como saída (output) a estrutura molecular do mesmo, ou ainda os inputs podem ser dados clínicos de um paciente e os outputs o diagnóstico e tipo de evolução. Enfim, as aplicações são extremamente variadas e estão presentes em quase todas as áreas do conhecimento humano.

Neste ponto do texto, já temos uma visão geral sobre redes neurais e, portanto, estamos aptos a aprofundar nossos conhecimentos sobre o assunto. Como veremos, esta é uma tarefa fácil visto que as redes neurais são elaboradas a partir de cinco conceitos básicos, extremamente simples. Estes elementos são:

- ① *Neurônios artificiais*
- ② *Sinapses e pesos*
- ③ *Funções de transferência*
- ④ *Arquitetura de redes neurais*
- ⑤ *Treinamento de redes neurais.*

4.1.1 Neurônios artificiais^{17, 60 - 64}

Neurônios artificiais são as unidades básicas de processamento da informação, projetadas para simular o comportamento de neurônios biológicos. Simplificadamente, um neurônio biológico pode ser descrito como um corpo celular contendo dois conjuntos de ramificações: *dendritos* e *axônio* (Figura 4.2A). Analogamente aos neurônios biológicos, os neurônios artificiais possuem um corpo de processamento de informação com duas ramificações: entradas (dendritos) e saída (axônio), conforme representado na Figura 4.2B.

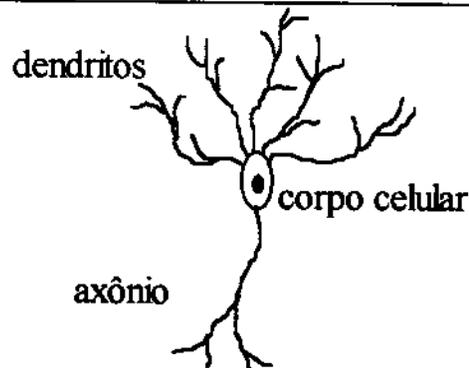
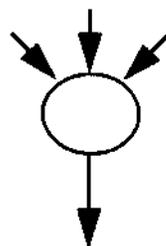


Figura 4.2A. Neurônio biológico

Entradas



Saída

Figura 4.2B. Neurônio artificial

4.1.2 Sinapses e pesos: *A transmissão de sinais pelos neurônios*^{65.66}

Sem entrar na físico-química do processo, a transmissão de sinais entre neurônios pode ser entendida do seguinte modo: O sinal entra no neurônio pelos dendritos, passa pelo corpo celular e em seguida é transmitido para outros neurônios, da rede neural, através do axônio. A passagem do sinal de um neurônio, para os dendritos de um outro neurônio é chamado de sinapse.

Sinapses representam barreiras que modulam o sinal que é trocado através delas e a quantidade de sinal trocado em uma sinapse, depende de um parâmetro chamado de *intensidade da sinapse*. Em um neurônio artificial a intensidade da sinapse é simulada por um fator de ponderação chamado *peso da sinapse* ou simplesmente peso, conforme ilustrado a Figura 4.3.

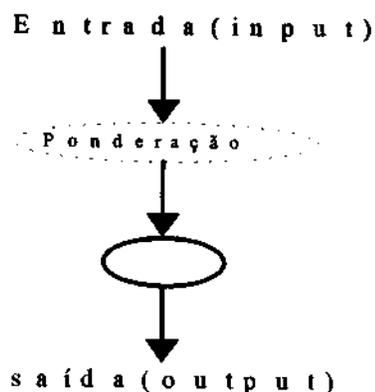


Figura 4.3. Representação da sinapse artificial

O sinal total que entra no corpo de processamento de um neurônio artificial é costumeiramente chamado de *Net*, cujo valor é calculado

através da simples multiplicação do sinal que chega ao neurônio pelo peso da sinapse em questão.

$$Net = \text{Sinal de entrada (input)} \times \text{Peso}$$

Neurônios possuem um grande número de dendritos e portanto, podem estabelecer várias sinapses com outros neurônios da rede neural, simultaneamente (Figura 4.4). Assim, o cálculo do *net* deve ser generalizado para *n* entradas (inputs).

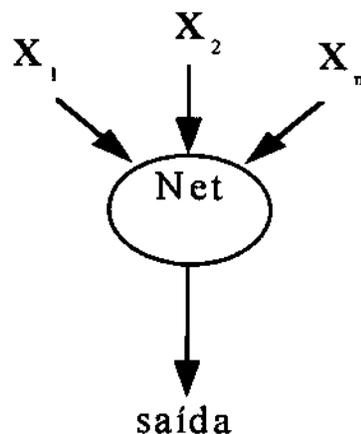


Figura 4.4 Representação generalizada da sinapse

Chamando os sinais de entrada de x_i e os pesos de w_i , o valor de *net* será dado por:

$$Net = \sum_{i=1}^n X_i W_i \quad \text{eq.4.1}$$

Até o momento vimos como calcular o sinal que entra em um neurônio, mas qual será o sinal de saída de um neurônio?

4.1.3 Funções de transferência^{17, 60, 67}

Como já poderíamos imaginar, a saída de um neurônio (output) será obtida em função do sinal de entrada (input), isto é, a saída será função do *net*.

$$\text{saída} = f(\text{net}) \quad \text{eq.4.2}$$

A função $f(\text{net})$ deve possuir as propriedades:

1) Não negativa

A função de transferência deve ser não negativa, pois ou o neurônio troca sinal com outro neurônio ou não troca, não há como trocar sinal negativo.

2) Contínua

A função de transferência deve ser contínua, pois um neurônio não pode trocar sinal com outro neurônio, indefinidamente.

Dentre as várias funções que preenchem os pré-requisitos anteriores, somente três são rotineiramente usadas, as quais sejam:

1) Função sigmoidal ou logística

Com certeza esta é a função de transferência mais utilizada em redes neurais. Esta função também é a que mais se aproxima da saída de um

neurônio biológico. Matematicamente, escrevemos a função sigmoidal como segue:

$$saida = f(net) = \frac{1}{1 + e^{-(net)}} \quad eq.4.3$$

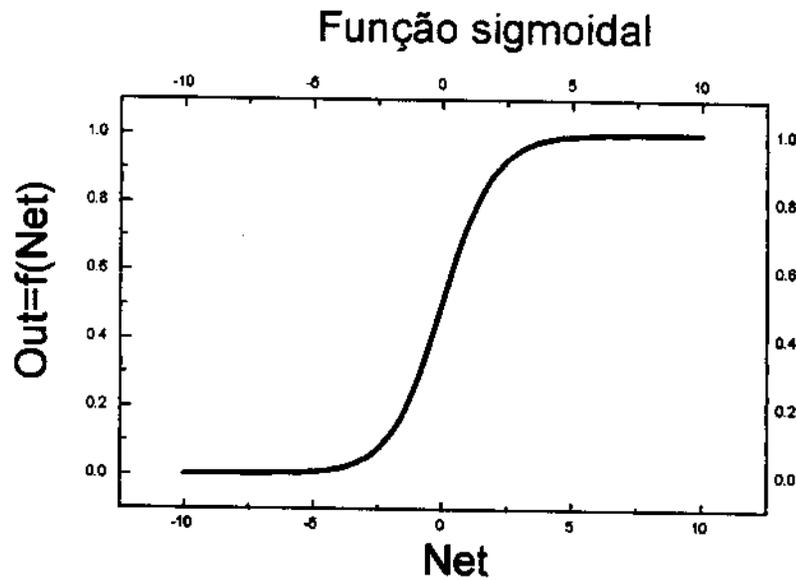


Figura 4.5. Função sigmoidal

2) Função degrau de Heaviside

Como podemos observar na Figura 4.6, a função de heavside é muito simples e só pode assumir os valores 0 ou 1.

$$out = f(net) = \begin{cases} 1 & \text{se } net > \text{valor limite} \\ 0 & \text{se } net \leq \text{valor limite(arbitrario)} \end{cases}$$

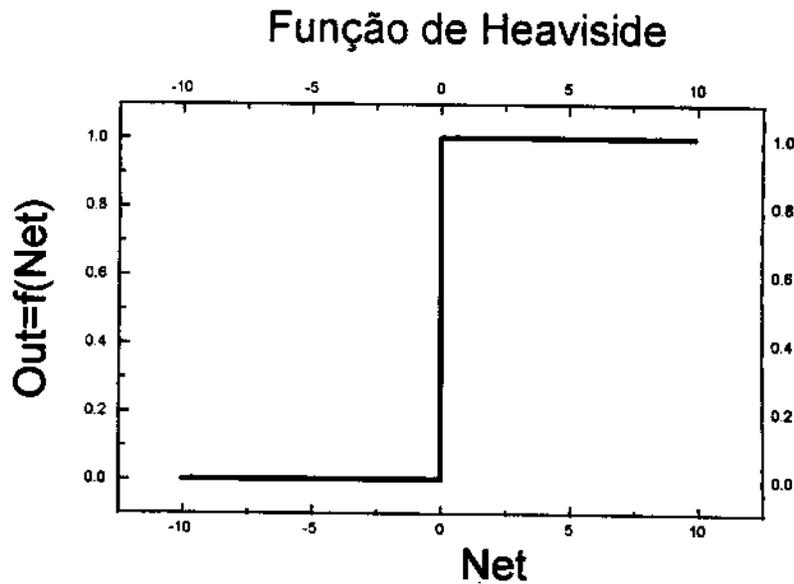


Figura 4.6. Função degrau de Heaviside

3) Função Linear

Esta função, naturalmente, dispensa comentários. A saída do neurônio será dada por:

$$saída = a (net) + b \quad a, b \in R_+^* \quad eq.4.4$$

4.1.4 Arquitetura de redes neurais

Uma rede neural artificial possui sempre uma *camada de entrada* e uma *camada de saída*. Entre a camada de entrada e a de saída, existe um número variável de *camadas intermediárias*. A esta disposição das camadas e número de neurônios por camada, dá-se o nome de *arquitetura* da rede neural. Na Figura 4.7 podemos observar a arquitetura típica de uma rede neural utilizada neste trabalho.

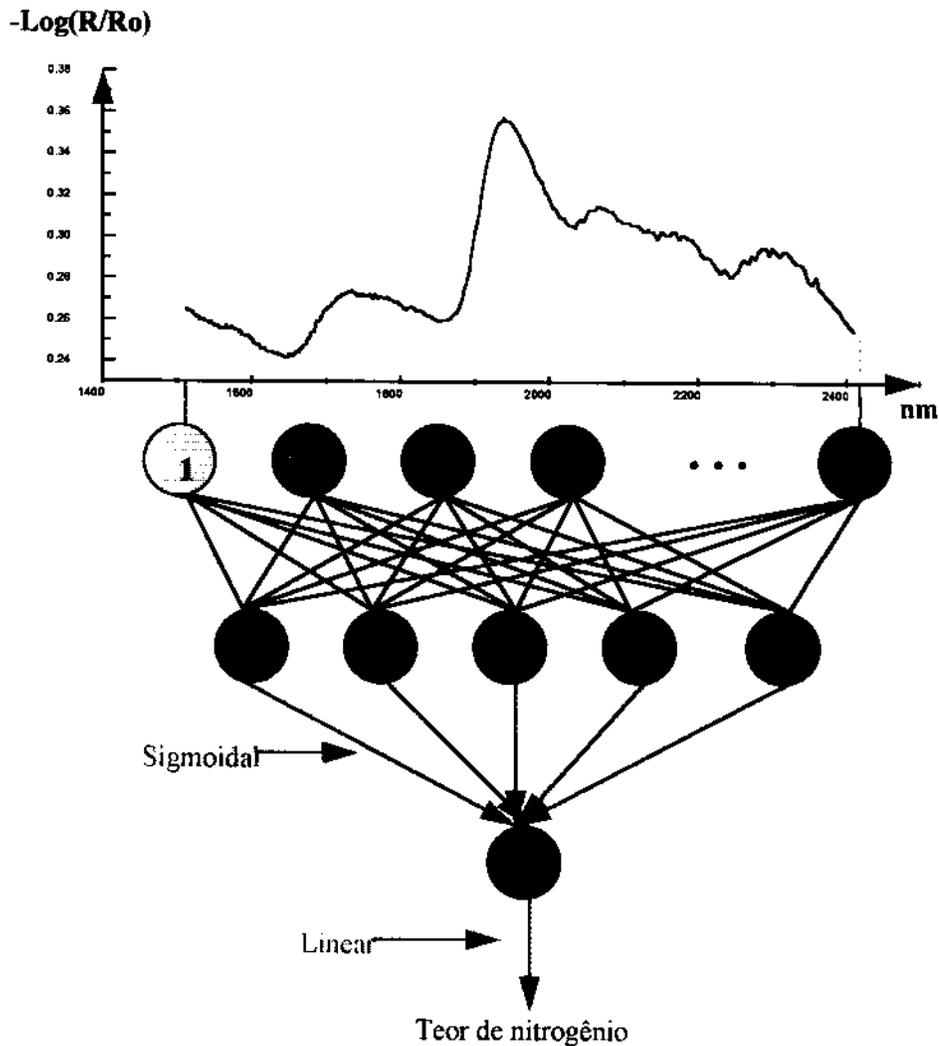


Figura 4.7. Representação de uma rede neural típica, utilizada neste trabalho.

Esta rede neural possui a seguinte arquitetura: 1 camada de entrada com 277 neurônios, uma camada intermediária com 5 neurônios e uma camada de saída com 1 neurônio. As funções de transferência utilizadas estão representadas na própria Figura 4.7. As entradas(inputs) foram as intensidades, isto é, valores de $(-\log(R/R_o))$ lidos em diferentes

comprimentos de onda, em espectros de reflectância difusa na região do infravermelho próximo (NIRR) de folhas de trigo, milho ou algodão. A saída (output) foi a concentração de nitrogênio nestas folhas.

Vale ressaltar que o número de camadas intermediárias e número de neurônios nestas, é um parâmetro ajustável conforme veremos adiante.

4.1.4.1 Cálculo do *Net* em uma rede neural

Já demonstramos (itens 4.1.2 e 4.1.3) como calcular o *Net* e a saída (output) para um neurônio. Vejamos agora como generalizar o cálculo do *Net* para uma rede neural.

É fácil imaginar que em uma rede neural com várias camadas, a saída de uma camada será o input da camada seguinte. Portanto, para calcular a saída final de uma rede neural basta calcular sequencialmente as entradas e saídas por toda a rede neural. Este cálculo que conceitualmente é muito simples, pode se tornar tedioso e complexo se a rede possuir várias camadas e muitos neurônios nestas camadas. Uma maneira mais simples e computacionalmente mais eficiente de se fazer isto, é “escrever” a rede neural de modo matricial. Assim, uma camada com n neurônios, cada um com m pesos, é representada por uma matriz W com n linhas e m colunas ($W_{n,m}$). Para uma rede neural com várias camadas, cada uma das camadas será representada por uma matriz W^l (em que l indica a camada). Para a rede neural representada na Figura 4.7, temos: $W_{277 \times 277}^1$, $W_{5 \times 1385}^2$ e $W_{1 \times 5}^3$.

Camada de entrada:

$$W_{277 \times 277}^1 = \begin{bmatrix} w_{1,1} & w_{1,2} & w_{1,3} & \dots & w_{1,277} \\ w_{2,1} & w_{2,2} & w_{2,3} & \dots & w_{2,277} \\ w_{3,1} & w_{3,2} & w_{3,3} & \dots & w_{3,277} \\ \vdots & \vdots & \vdots & & \vdots \\ w_{277,1} & w_{277,2} & w_{277,3} & \dots & w_{277,277} \end{bmatrix}$$

Camada intermediária:

$$W_{5 \times 1385}^2 = \begin{bmatrix} w_{1,1} & w_{1,2} & w_{1,3} & w_{1,4} & w_{1,5} & \dots & w_{1,1385} \\ w_{2,1} & w_{2,2} & w_{2,3} & w_{2,4} & w_{2,5} & \dots & w_{2,1385} \\ w_{3,1} & w_{3,2} & w_{3,3} & w_{3,4} & w_{3,5} & \dots & w_{3,1385} \\ w_{4,1} & w_{4,2} & w_{4,3} & w_{4,4} & w_{4,5} & \dots & w_{4,1385} \\ w_{5,1} & w_{5,2} & w_{5,3} & w_{5,4} & w_{5,5} & \dots & w_{5,1385} \end{bmatrix}$$

Camada de saída:

$$W_{1 \times 5}^3 = [w_{1,1} \quad w_{1,2} \quad w_{1,3} \quad w_{1,4} \quad w_{1,5}]$$

As entradas, isto é, valores de $(-\log(R/R_0))$ lidos em diferentes comprimentos de onda, no espectro de reflectância difusa no infravermelho próximo podem ser escritas vetorialmente do seguinte modo:

$$x_i^l = [x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5 \quad \dots \quad x_{277}]$$

O cálculo do *net* para cada um dos neurônios, da camada de entrada, por exemplo, pode ser escrito como:

$$Net = [Net_1, Net_2, \dots, Net_{277}] = \begin{bmatrix} w_{1,1} & w_{1,2} & w_{1,3} & \dots & w_{1,277} \\ w_{2,1} & w_{2,2} & w_{2,3} & \dots & w_{2,277} \\ w_{3,1} & w_{3,2} & w_{3,3} & \dots & w_{3,277} \\ \vdots & \vdots & \vdots & & \vdots \\ w_{277,1} & w_{277,2} & w_{277,3} & \dots & w_{277,277} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{277} \end{bmatrix}$$

O cálculo do *Net* pode ser generalizado, para qualquer camada, através da seguinte equação:

$$Net_j^l = \sum_{i=1}^m W_{ji}^l \cdot x_i^l \quad \text{eq.4.5}$$

Uma vez que toda entrada de uma camada pode ser relacionada com a saída da camada anterior, podemos reescrever a equação 4.5 como segue:

$$Net_j^l = \sum_{i=1}^m W_{ji}^l \cdot out_i^{l-1} \quad \text{eq.4.6}$$

Conhecidos o *Net* e as funções de transferência, podemos “propagar” o sinal pela rede, desde a entrada (input) até a saída (output).

4.1.5 Treinamento de redes neurais ⁶⁸

Para que uma rede neural execute tarefas como reconhecimento de padrões ou modelamento de dados (calibração), devemos treiná-la. Vejamos como se faz isto.

Na etapa de treinamento, uma série de entradas são apresentadas e propagadas pela rede neural. Neste trabalho, os inputs foram intensidades lidas em diferentes comprimentos de onda, para vários espectros de reflectância difusa na região do infravermelho próximo de folhas de trigo, milho ou algodão. Uma vez propagados, podemos calcular a saída da rede neural, isto é, o valor previsto da concentração de nitrogênio nestas folhas, conforme representamos na Figura 4.7.

Uma vez obtidos os valores previstos pela rede neural, podemos calcular um erro de previsão (E), o qual pode ser definido, em princípio, como o somatório do quadrado da diferença entre o valor previsto pela rede (saída) e o valor real (target), como segue:

$$E = \sum_{k=1}^n (y_k - \hat{y}_k)^2 \quad \text{eq.4.7}$$

onde y é o valor real (target) e \hat{y} o valor previsto (output) da concentração de nitrogênio para cada um dos n espectros (NIRR) utilizados no treinamento da rede neural. Conhecido o erro de previsão, quais parâmetros da rede neural devemos corrigir para minimizá-lo ? A resposta para esta pergunta é muito simples, pois uma vez fixada a arquitetura da rede neural, os únicos parâmetros passíveis de correções são os pesos, dado que não há outros parâmetros. Existem alguns métodos para se fazer tal correção, entretanto somente dois são usados com

frequência, os quais sejam: a **retropropagação de erros**⁶⁹ e o método de **Marquardt-Levenberg**⁷⁰.

No método da retropropagação de erros, o processo de correção dos pesos é iniciado na última camada e prossegue em direção à primeira camada, daí o nome **retropropagação**. Este processo pode ser visualizado na Figura 4.8.

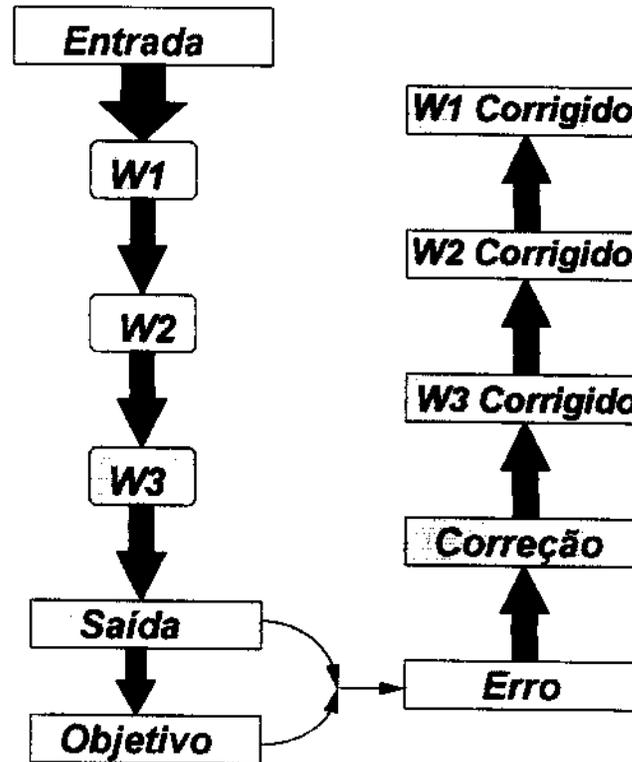


Figura 4.8. Representação da retropropagação para uma rede com uma camada de entrada(W1), uma intermediária(W2) e uma camada de saída(W3)

Sem entrar em deduções matemáticas, a equação para a correção dos pesos pode ser escrita como:

$$\Delta W_{ji}^l = \eta \cdot \delta_j^l \cdot out^{l-1} + \mu \cdot \Delta W_{ji}^{previo} \quad \text{eq.4.8}$$

De acordo com a equação anterior, a correção dos pesos é composta pela soma de dois termos que atuam conjuntamente para minimizar o erro de previsão. Vejamos o que cada um representa:

- Primeiro termo:

$$\eta \cdot \delta_j^l \cdot out^{l-1}$$

O fator de correção δ_j^l é obtido utilizando-se o método do gradiente descendente, cuja essência é que o erro (E) deve apresentar um mínimo em função do parâmetro que o causa, ou melhor, δ_j^l fornecerá os pesos que minimizam o erro (E). O parâmetro η chamado de velocidade de treinamento é introduzido para ponderar as correções, de modo a evitar grandes correções principalmente no final do processo iterativo, quando boa parte dos pesos já foi corrigido e uma correção muito acentuada pode comprometer todas as correções efetuadas.

- Segundo termo:

$$\mu \cdot \Delta W_{ji}^{previo}$$

A constante μ desloca o peso “ótimo” obtido pelo primeiro termo, forçando que o erro seja reavaliado neste novo peso. Este processo evita convergências para mínimos locais, pois um erro maior indica que o mínimo obtido pelo primeiro termo era realmente global; caso contrário, o mínimo era local e a busca pelo mínimo prossegue. Resumindo, o segundo

termo atua como dispositivo de segurança do primeiro termo, evitando mínimos locais.

No segundo método de treinamento as correções dos pesos são feitas de acordo com uma variante do método de Gauss-Newton⁷¹⁻⁷³, conhecida como método de Marquardt-Levenberg. Neste método as correções dos pesos são calculadas de acordo com a equação:

$$\Delta W = (J^T J + \lambda I)^{-1} J^T E \quad \text{eq.4.9}$$

onde J é a matriz jacobiana do erro para cada um dos pesos, λ é um escalar positivo, I é a matriz identidade e E o vetor de erros dado pela equação 4.7.

Em ambos os casos, a etapa de treinamento se encerra quando o erro atingir o critério de convergência previamente estabelecido. Neste ponto, a rede está treinada e podemos utilizar um conjunto de espectros de NIR, diferente daquele utilizado no treinamento, para avaliar as propriedades de generalização da rede neural artificial, as quais são geralmente boas, especialmente em situações em que os dados apresentem não linearidades. Entretanto, em alguns casos, as redes neurais artificiais, mesmo depois de terem sido treinadas, apresentam baixos erros de calibração e elevados erros de previsão, isto é, sobreajuste (overfitting), devido ao número excessivo de neurônios utilizados na camada intermediária. O número de neurônios na camada intermediária atua de modo análogo ao número de componentes principais utilizado na regressão de componentes principais⁷⁴⁻⁷⁵ (PCR), ou à ordem do polinômio utilizado em regressão polinomial⁷⁶, por exemplo. Portanto, um procedimento fundamental para evitar-se sobreajuste (overfitting), isto é,

pequenos erros de calibração e elevados erros de previsão, é a otimização do número de neurônios da camada intermediária.

4.1.6 A otimização da Arquitetura

A seleção do número ótimo de neurônios na camada intermediária ou otimização da arquitetura é feita variando-se o número destes neurônios e repetindo-se o processo de correção dos pesos. Escolhe-se então a configuração que apresentar os menores erros de calibração (%SEC) e previsão (%SEP), dados pelas eqs. 4.10 e 4.11

$$\%SEC = \frac{1}{\bar{c}_c} \sqrt{\frac{\sum_{i=1}^{n_c} (y_i - \hat{y}_i)^2}{n_c - k - 1}} \times 100 \quad \text{eq. 4.10}$$

$$\%SEP = \frac{1}{\bar{c}_p} \sqrt{\frac{\sum_{i=1}^{n_p} (y_i - \hat{y}_i)^2}{n_p}} \times 100 \quad \text{eq.4.11}$$

em que \bar{c}_c e \bar{c}_p são as concentrações médias da amostras utilizadas no conjunto de calibração e previsão respectivamente, n_c e n_p representam, respectivamente, o número de amostras usadas na calibração e na previsão, y_i é o valor real e \hat{y}_i o valor previsto pelo modelo e k é o número de parâmetros usados no modelo. Para redes neurais artificiais o

número de parâmetros utilizado no modelo não é conhecido e n_p é utilizado como uma aproximação, para o denominador da equação 4.10.

O leitor atento, a esta altura já deve ter observado que para construir modelos com redes neurais o conjunto de previsão também é utilizado na escolha da arquitetura ótima. Portanto, devemos ter um terceiro conjunto de espectros para realmente verificar as capacidades de previsão da rede neural otimizada. Afinado com este tipo de problema sempre utilizamos neste trabalho, três conjuntos de dados (calibração, validação e previsão) como veremos adiante no texto.

Conforme descrevemos, a otimização da arquitetura é basicamente uma tarefa de tentativa e erro e um tanto tediosa mas existe um modo de evitá-la, como veremos no próximo item.

4.2 Redes Neurais com Apodização^{77, 78}

A idéia básica deste método é inicializar a rede neural com um número razoável de neurônios na camada intermediária e, durante a etapa de treinamento cortar as conexões (ou pesos) dos neurônios que possuam pouca influência no erro E (eq.4.7). Neurônios que tiverem todas as conexões cortadas serão eliminados e, portanto, ao final dos “cortes”, sobrarão somente os neurônios realmente necessários ao modelamento. A técnica de apodização (pruning) reduz a complexidade da rede neural, melhorando suas capacidades de previsão, pois evita modelos sobre parametrizados (muitos neurônios) em que a possibilidade de sobreajuste (overfitting) é grande.

Existem basicamente dois métodos para a apodização de redes neurais: Optimal Brain Damage⁷⁹ (OBD) and Optimal Brain Surgeon^{80,81,82}

(OBS). Em ambos os métodos as conexões (ou pesos) são cortadas e a correspondente variação no erro E , chamada de saliência, é avaliada.

No método OBD as conexões são cortadas durante a etapa de treinamento e a rede neural não é retreinada após os cortes. No método OBS, as conexões são cortadas e, após o corte de uma conexão, a rede é retreinada, permitindo que um número maior de cortes seja efetuado. Além disso, no método OBS a rede neural é retreinada, aproximando-se os erros de treinamento por uma função quadrática, de modo a garantir a existência de um mínimo.

Resumidamente, o que se faz no método OBS é adicionar um termo, chamado termo de regularização^{83,84}, à função erro (E) dada pela equação 4.7, para obter a função “custo”:

$$C(W) = E(W) + \frac{1}{2} W^T D W \quad \text{eq.4.12}$$

O termo de regularização simplesmente penaliza pesos com valores elevados, sendo esta penalização, ponderada pelo fator D . O fator de regularização auxilia no processo de pruning e assegura a estabilidade numérica do método. Em seguida, expande-se a função custo em série de Taylor até termos de segunda ordem, em torno de um possível mínimo W_0 , como segue:

$$C(W) = C(W_0) + \frac{1}{2} (\nabla^2 E + D) (dW)^2 \quad \text{eq.4.13}$$

em que $dW = W - W_0$.

Em seguida, devemos cortar conexões e minimizar a função custo simultaneamente, o que pode ser conseguido aplicando-se o método dos multiplicadores de Lagrange, como segue:

$$\nabla C(W) = \lambda \nabla W \quad \text{eq.4.14}$$

em que λ é o multiplicador de Lagrange. Resolvendo a equação anterior obtemos o extremo vinculado:

$$dW = -\lambda H^{-1} \quad \text{eq.4.15}$$

em que H é a Hessiana da função custo com termo de regularização, dada por :

$$H = (\nabla^2 E + D) \quad \text{eq.4.16}$$

Neste ponto pode-se voltar a equação 4.12, explicitar a função $E(W)$, substituir o valor de dW e obter a seguinte equação para a saliência:

$$dE(W) = \lambda W_0 D H^{-1} + \frac{1}{2} \lambda^2 H^{-1} (\nabla^2 E + D) H^{-1} \quad \text{eq.4.17}$$

A equação da saliência (eq.4.17) nos fornece um critério matemático para o corte das conexões.

As técnicas de apodização de redes simplifica enormemente o processo de otimização da arquitetura e nos permite modelos com pequena possibilidade de sobreajuste (overfitting).

4.3 O método dos mínimos quadrados parciais^{85,86} (PLS)

A base do PLS é decompor a matriz (X) das variáveis independentes e a matriz (Y) das variáveis dependentes, em um produto de duas matrizes menores mais uma matriz de erro, como segue:

$$X = TP + E \quad \text{eq.4.18}$$

$$Y = UQ + F \quad \text{eq.4.19}$$

onde as matrizes T e U são chamadas de matrizes dos escores; P e Q matrizes dos loadings; E e F as matrizes de erro de X e Y respectivamente. Esta decomposição é muito útil nos casos em que a matriz X é mal condicionada, ou ainda, quando o número de amostras é menor que o número de variáveis independentes visto que podemos utilizar uma matriz T de dimensão inferior a da matriz X sem perda de informação útil, eliminando ruído e colinearidade dos dados.

Efetuada a decomposição anterior, o próximo passo é ajustar uma relação linear, quando possível, entre U e T , como segue:

$$U = b T + \varepsilon \quad \text{eq.4.20}$$

onde b é o coeficiente do ajuste, usualmente obtido com o algoritmo NIPALS⁸⁶. Finalmente, podemos substituir o valor de U na equação 4.19, e obter:

$$Y = b T Q + G \quad \text{eq.4.21}$$

e portanto, podemos obter os escores da matriz Y a partir dos escores da matriz X e vice-versa. Terminada a etapa de calibração pode-se fazer previsões⁸⁷ para amostras desconhecidas. Para tanto basta obter os escores da matriz X , o qual pode ser facilmente transformado em concentração, através da eq.4.21.

4.4 Outras técnicas de modelamento de dados não lineares⁸⁸: O PLS polinomial (POLIPLS) e o PLS redes Neurais (NNPLS).

Algumas extensões do método do PLS podem ser feitas nos casos em que os escores da matriz das variáveis independentes (matriz X) e os escores da matriz das variáveis dependentes (matriz Y), não possuam uma relação linear. Dentre estas, destacamos o PLS polinomial (POLIPLS) e o PLS rede neural (NNPLS).

O PLS polinomial é basicamente idêntico ao PLS linear, com a única diferença de que a relação entre os escores da matriz das variáveis independentes (u) e os escores da matriz das variáveis dependentes (t), é feita por um polinômio de grau variável, o que nos permite modelar relações não lineares entre t e u , como segue:

$$t = a_0 + a_1 u + a_2 u^2 + a_3 u^3 + \dots + a_n u^n \quad \text{eq.4.22}$$

Uma combinação muito eficiente entre as redes neurais artificiais e o PLS, é o PLS rede neural (NNPLS). No PLS rede neural a relação entre os escores da matriz das variáveis independentes (u) e os escores da

matriz das variáveis independentes (t), é feita por uma rede neural artificial, de arquitetura variável, conforme representado na Figura 4.9. Assim sendo, é possível modelar os casos lineares, bem como aqueles não-lineares. Para modelar relações lineares, basta utilizar na rede neural, funções de transferência lineares e para as relações não lineares deve-se utilizar funções de transferência não lineares, como por exemplo, a função sigmoideal. As outras etapas do PLS rede neural são idênticas as do PLS linear. Deve-se ter em mente que a diferença entre os modelos feitos com redes neurais artificiais e com o PLS rede neural, é que este é baseado em fatores (escores) e as redes neurais, geralmente, não.

Todavia, é possível elaborar modelos com redes neurais artificiais utilizando escores como entradas (inputs), conforme descrito no item 3.3 desta tese. Aliás, optamos por esta estratégia em nosso trabalho e obtivemos uma redução extremamente significativa no tempo computacional e na qualidade dos resultados.

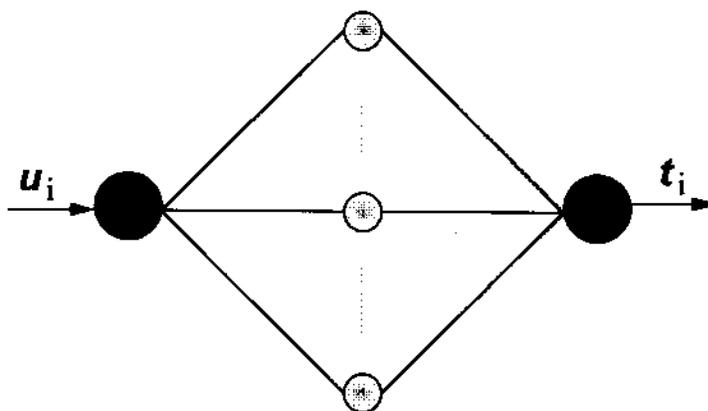


Figura 3.9. Representação esquemática do NNPLS. As linhas tracejadas indicam que a arquitetura da rede neural é ajustável.

CAPÍTULO 5

OBJETIVOS

5. OBJETIVOS

O objetivo deste trabalho é implementar, desenvolver e avaliar o desempenho relativo de métodos matemáticos e estatísticos, ainda pouco explorados em química analítica, aplicados ao pré-processamento de sinais e modelamento de dados que possibilitem a análise *in situ* do teor de nitrogênio em folhas de plantas, com pequenos erros, através da espectroscopia de reflectância difusa no infravermelho próximo (NIRR). O teor de nitrogênio em plantas é um parâmetro agrônômico extremamente importante pois permite prever a produtividade de um determinado cultivar, possibilitando a aplicação, racional, de medidas (adubação, por exemplo) que melhorem-na.

CAPÍTULO 6

PARTE EXPERIMENTAL

6. Parte Experimental

6.1 Determinação de nitrogênio em plantas: O método Kjeldahl

Obtivemos, junto ao INSTITUTO AGRONÔMICO DE CAMPINAS (IAC), amostras de folhas de trigo, milho e algodão, provenientes de diversos cultivares, localizados em diferentes cidades do estado de São Paulo. Foram obtidas 80 amostras, assim distribuídas: 60 amostras de folhas de trigo, 10 de milho e 10 de algodão.

O teor de nitrogênio de cada uma das amostras foi determinado através do método Kjeldahl⁸⁹. O fundamento do método Kjeldahl, é oxidar o nitrogênio das folhas a amônio, com uma mistura de ácido sulfúrico e peróxido de hidrogênio. Em seguida adiciona-se NaOH ao meio reacional, pois o íon amônio, em meio básico gera amônia, a qual é destilada por arraste de vapor e recebida em uma solução de ácido bórico, gerando borato. Finalmente a solução contendo o borato, recém formado, é titulada com ácido sulfúrico, utilizando uma mistura de indicadores (vermelho de metila e verde de bromo cresol) para detecção do ponto final.

Na etapa anterior à determinação de nitrogênio pelo método Kjeldahl, as folhas foram lavadas com uma solução de água e detergente (1%) e enxaguadas várias vezes com água destilada e uma vez com água deionizada. Em seguida, as folhas foram colocadas em sacos de papel (para identificação), secas em estufa a 65°C (mínimo de 24 h), moídas em um moinho do tipo Wiley com abertura de 2 mm e peneiradas em peneira de 80 mesh. Esta etapa do procedimento experimental está representada na Figura 6.1.

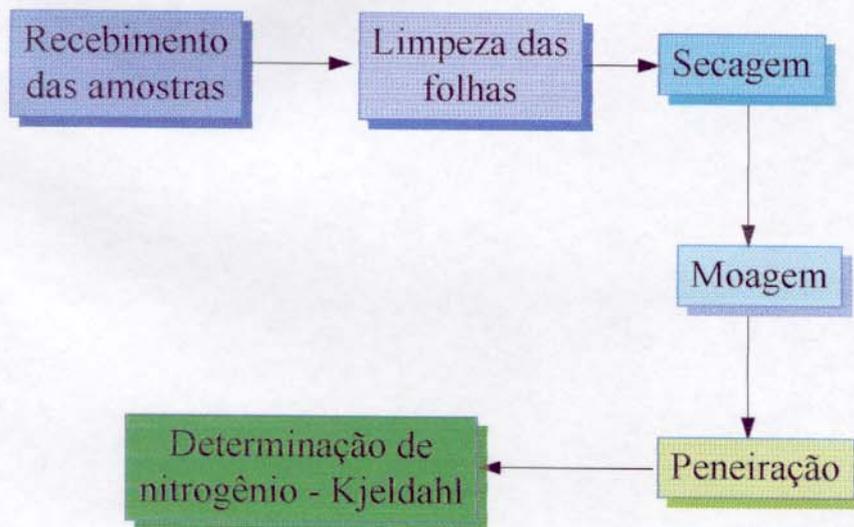


Figura 6.1. Representação da sequência de operações experimentais executadas.

6.2 Obtenção dos espectros de reflectância difusa no infravermelho próximo.

As amostras de folhas moídas foram peneiradas e a fração que passou através da peneira de 80 mesh foi dividida em três alíquotas. Para cada uma destas alíquotas, obtivemos um espectro de reflectância na região do infravermelho próximo entre 1500 e 2400 nm. O espectro final para cada uma das amostras foi a média dos espectros das três alíquotas. Antes da aquisição do espectro de cada alíquota (e para todas amostras), obteve-se o espectro do padrão de reflectância utilizado como referência em cálculos posteriores (padrão nº 99 da empresa LabSphere, Inc.). Assim, foram obtidos 6 espectros para cada amostra, resultando em um total de aproximadamente, 500 espectros de reflectância difusa no infravermelho próximo.

Os espectros NIR foram obtidos em um espectrofotômetro de infravermelho próximo com monocromador baseado em filtro óptico acústico sintonizável⁹⁰⁻⁹³ (AOTF), desenvolvido e gentilmente cedido pelo Prof. Dr. Célio Pasquini (IQ-UNICAMP) e pelo Sr. Renato Guchardi. Na montagem utilizada para as medidas de reflectância, adaptou-se ao espectrofotômetro uma fibra óptica bifurcada (um feixe conduz a radiação NIR até a amostra e o outro conduz a radiação NIR atenuada, devido às absorções, até o detector). Esta adaptação permitiu que as medidas de reflectância fossem feitas diretamente sobre as folhas moídas (ou sobre o padrão de reflectância, Teflon[®] finamente particulado), como representado nas Figuras 6.2.A, 6.2.B e 6.2.C.

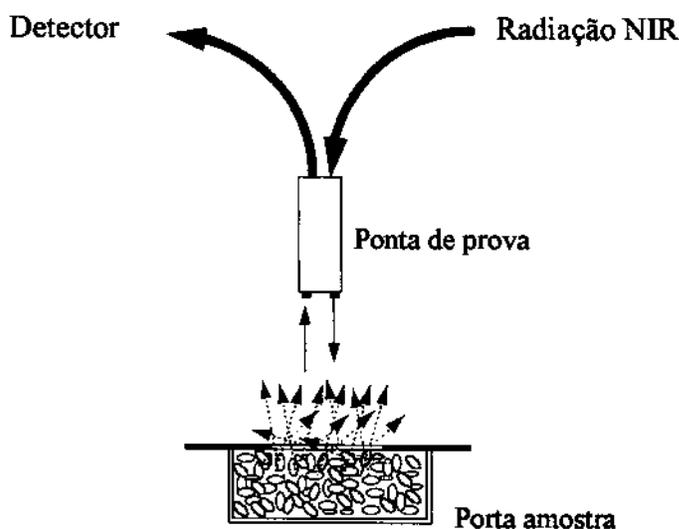


Figura 6.2.A. Representação da configuração utilizada para as medidas de reflectância difusa.



Figura 6.2.B A. Nesta foto pode-se observar o porta amostra (carregado com folhas de trigo moídas) e a ponta de prova, da fibra óptica.



Figura 6.2.C. Nesta foto pode-se observar o padrão de reflectância e a ponta de prova da fibra óptica.

Na montagem utilizada, apresentada na Figura 6.2.A, optou-se pelo uso de uma fibra óptica com baixo teor de hidroxilas para evitar-se a absorção da radiação NIR, pela própria fibra óptica.

Na fotografia apresentada na Figura 6.3, podemos observar, em detalhe, o porta amostra, o padrão de referência e alguns frascos com amostras de folhas de trigo, moídas.



Figura 6.3. Foto do porta amostra, padrão de referência e frascos com amostras de folhas de trigo moídas.

O espectrofotômetro de infravermelho próximo utilizado para a obtenção dos espectros de reflectância apresentava a configuração representada na Figura 6.4.

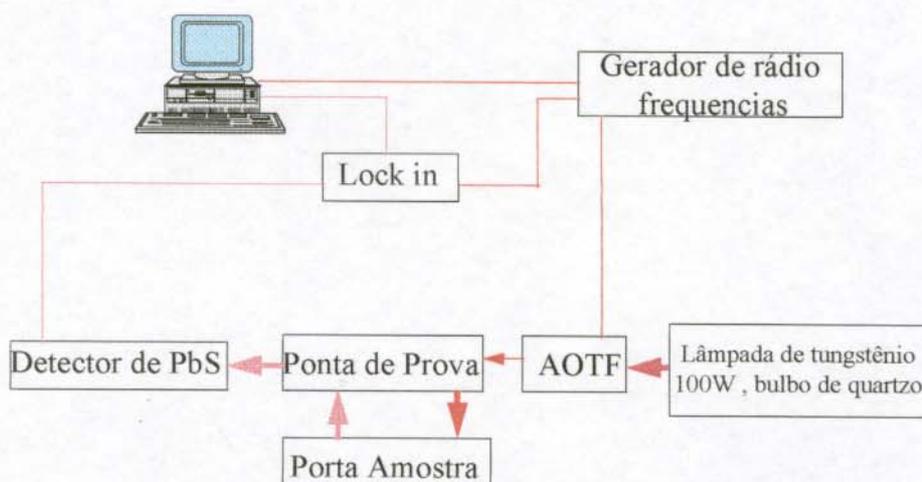


Figura 6.4. Representação, esquemática e simplificada, do espectrofotômetro NIRR.

O espectrofotômetro utilizado para a varredura, nos comprimentos de onda, apresentava o seguinte princípio de funcionamento:

O gerador de rádio frequências, controlado pelo microcomputador, aplica sobre um material piezoelétrico (LiNbO_3), acoplado à AOTF, uma diferença de potencial à uma determinada frequência, conforme representado na Figura 6.5.

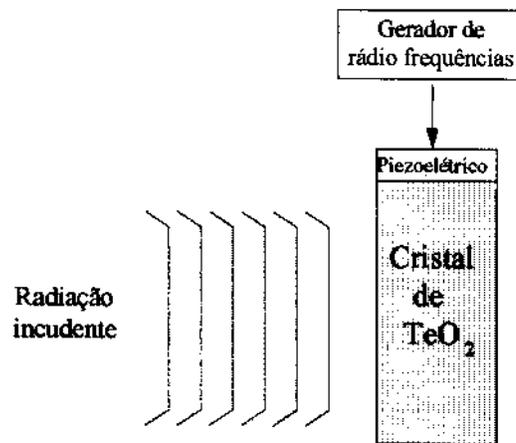


Figura 6.5. Representação simplificada, de uma AOTF.

Materiais piezoelétricos, quando submetidos à diferenças de potencial respondem expandindo-se ou contraindo-se; se a diferença de potencial é aplicada com uma determinada frequência, o piezoelétrico executa um movimento vibratório. Uma vez que o material piezoelétrico está acoplado a um cristal de TeO_2 , a vibração mecânica se propagará através deste cristal. Vibrações mecânicas (o som, por exemplo) quando se propagam em meios materiais, como o cristal de TeO_2 , geram regiões, internas ao cristal, de compressão e rarefação, como representado na Figura 6.6.

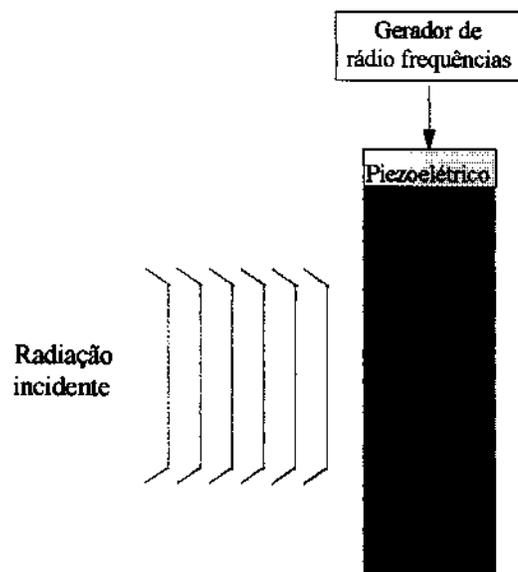


Figura 6.6. Representação simplificada do princípio de funcionamento da AOTF. As regiões claras indicam rarefação e as escuras, compressão.

A compressão e a rarefação alteram significativamente a densidade do meio (cristal de TeO_2), em diferentes posições, à medida que a onda se propaga. Consequentemente, teremos nestas posições, uma correspondente alteração do índice de refração e devido a estas variações periódicas (no espaço) no índice de refração, o cristal irá difratar a luz seletivamente. Assim, controlando-se a frequência aplicada sobre o cristal de TeO_2 , é possível controlar o processo de difração, o que, em última análise, torna a AOTF um *tipo* de grade de difração^{94,95}, isto é, um tipo de filtro que permite selecionar determinados comprimentos de onda.

Uma vez efetuada a seleção dos comprimentos de onda, através da AOTF, a radiação é direcionada para um dos feixes da fibra óptica, sendo ‘*conduzida*’ até a amostra. A radiação incide sobre a amostra, sofre o processo de reflectância difusa, sendo em seguida *capturada* pela outra extremidade da fibra e *conduzida* até o detector. Uma vez detectada a intensidade luminosa (atenuada nos comprimentos de onda absorvidos pelos componentes da amostra), o sinal segue para o amplificador “lock-

in". Neste dá-se o acoplamento das frequências, do sinal proveniente do detector com a frequência aplicada, pelo gerador de rádio frequências, na AOTF. Esta modulação permite que as medidas de reflectância difusa sejam feitas com o espectrofotômetro *aberto*, sem sofrer interferência da luz ambiente ou de outras fontes, eventualmente presentes no local. Na Figura 6.7 apresentamos a foto de uma vista geral do espectrofotômetro utilizado.

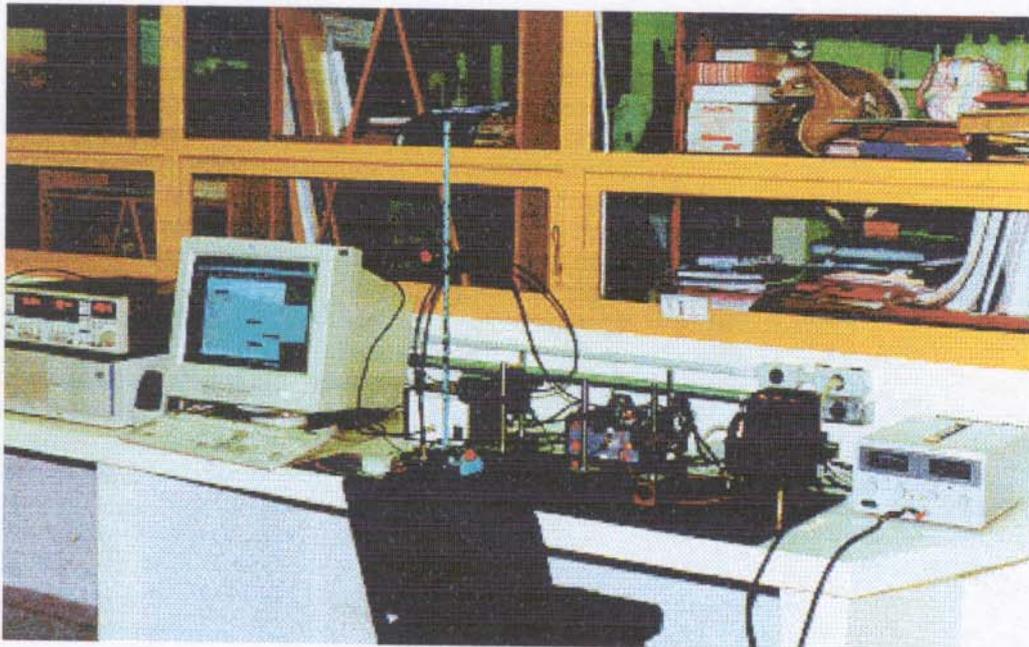


Figura 6.7. Vista geral do espectrofotômetro utilizado para a obtenção dos espectros de reflectância difusa no infravermelho próximo.

6.3 Programas computacionais para o tratamento dos espectros ⁹⁶⁻⁹⁸

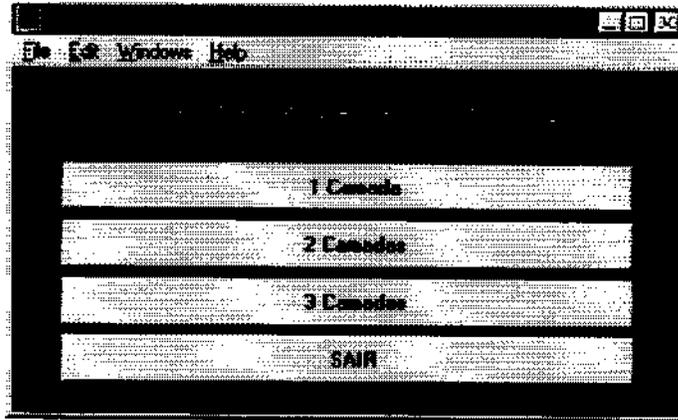
Desenvolvemos os seguintes programas para o tratamento dos dados espectroscópicos:

- O programa para a correção de espalhamento de luz (MSC) foi desenvolvido e implementado em Fortran Power Station[®] (MICROSOFT), e posteriormente, adaptado para MATLAB[®] (The MathWorks, Inc.).

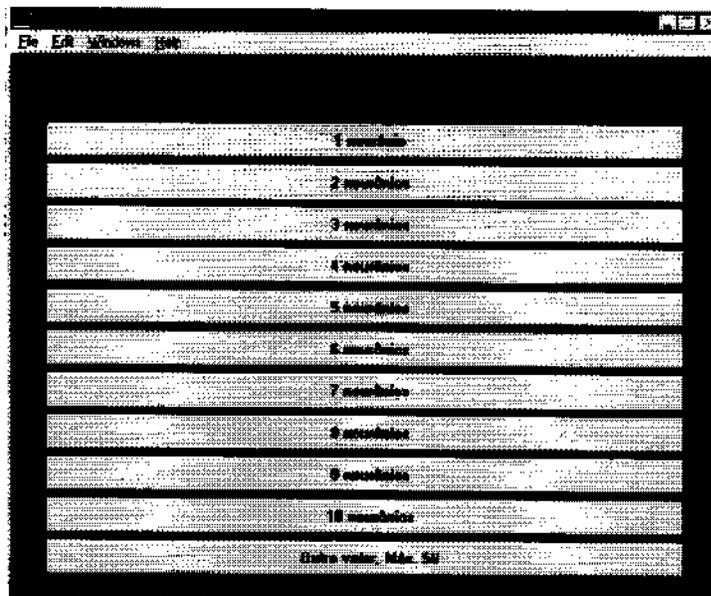
- O programa para o filtro de Wavelets foi desenvolvido e implementado em Fortran Power Station[®] (MICROSOFT) e, posteriormente, adaptado para MATLAB[®] (The MathWorks, Inc.).

- Desenvolvemos interfaces gráficas amigáveis (GUI) utilizando a linguagem e subrotinas do MATLAB[®] (The MathWorks, Inc.) para as seguintes técnicas de modelamento e pré-processamento de espectros (NIR): filtro de transformada de Fourier, Redes Neurais, PLS rede neural e Redes neurais com apodização. Por exemplo, o programa para modelamento dos dados através de redes neurais, apresentava a seguinte sequência de telas:

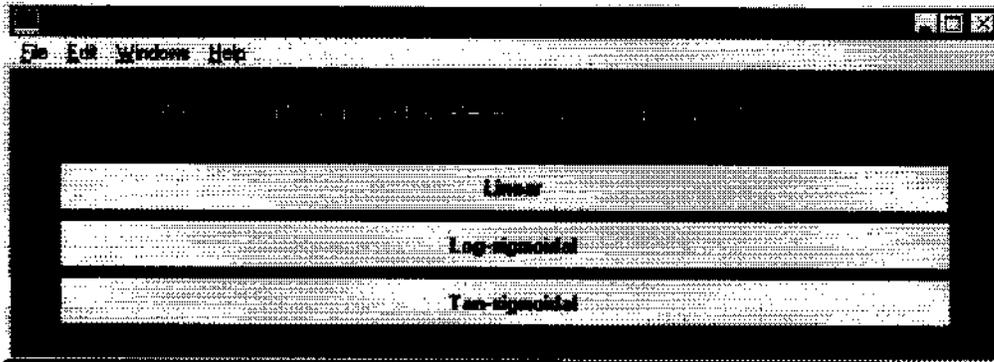
Escolhendo a arquitetura da rede neural



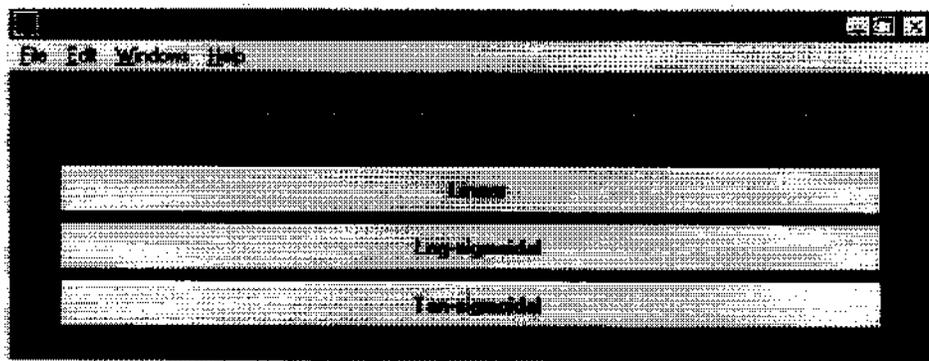
Selecionando o número de neurônios na camada intermediária (no caso de só uma camada intermediária ter sido escolhida)



Selecionando a função de transferência para a camada intermediária.....



Selecionando a função de transferência para a camada de saída



Conforme pode ser observado na sequência de *telas* anteriores, a utilização de redes neurais (e das outras técnicas de modelamento, também), não é uma tarefa complicada.

CAPÍTULO 7

RESULTADOS E DISCUSSÃO

7 Resultados e Discussão

7.1 Os espectros obtidos

Os espectros foram traçados lançando-se nas abscissas os comprimentos de onda (λ em nm) e nas ordenadas o logaritmo na base 10, da razão entre a reflectância da amostra (R) e a reflectância do padrão (R_0), isto é, utilizamos o método empírico descrito no item 2 desta tese. Na Figura 7.1 representamos três espectros típicos, um para cada tipo de amostra.

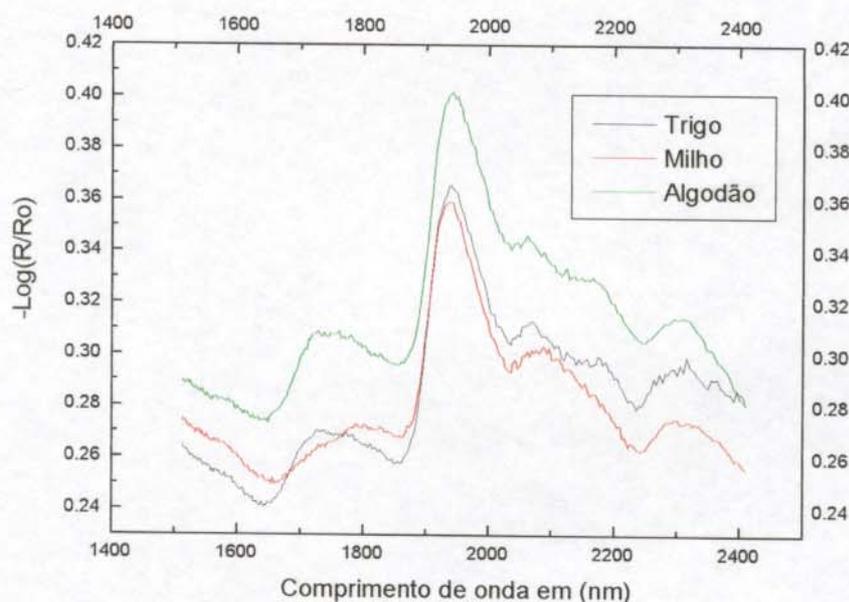


Figura 7.1. Espectros NIRR de folhas de milho, trigo e algodão.

Este tipo de transformação pode ser utilizada para os espectros obtidos neste trabalho pois tomamos o cuidado de construir o porta amostra em Teflon[®] grafitado, impedindo que a radiação incidente o

atravessa-se completamente eliminando os problemas teóricos, relativos à esta transformação, já discutidos no item 2 da introdução desta tese. Conforme descrito na parte experimental, não utilizamos um padrão com 100% de reflectância, o que nos impede, pelo menos teoricamente, de utilizar a transformação de Kubelka-Munk. Além disso, a transformação empírica é muito mais simples de se implementar computacionalmente.

7.1.1 Homogeneidade das amostras⁹⁹

As amostras foram separadas em três alíquotas (ou em três frações) pois devemos garantir que o espectro NIRr realmente represente a composição global da amostra. Esta precaução é necessária pois não é de se esperar que uma folha apresente a mesma concentração de nitrogênio em toda sua extensão. Na verdade, esta precaução deve ser tomada em todos os tipos de análises em que não é possível, ou pouco viável, obter a partir da amostra, uma solução verdadeira. No caso das soluções verdadeiras temos uma situação realmente confortável pois a difusão ou uma simples agitação resolve o problema.

Constatamos que os espectros para as três alíquotas, em todas as amostras, apresentavam variações iguais àquelas observadas em três espectros obtidos de uma mesma alíquota. Na Figura 7.2, podemos observar os espectros de três alíquotas de uma certa amostra (observe como é difícil diferenciá-los visualmente). O desvio médio calculado é da ordem de 0,005 unidades de reflectância. Portanto, os espectros NIRr obtidos refletem de modo fiel a composição global das amostras de folhas de trigo, milho e algodão.

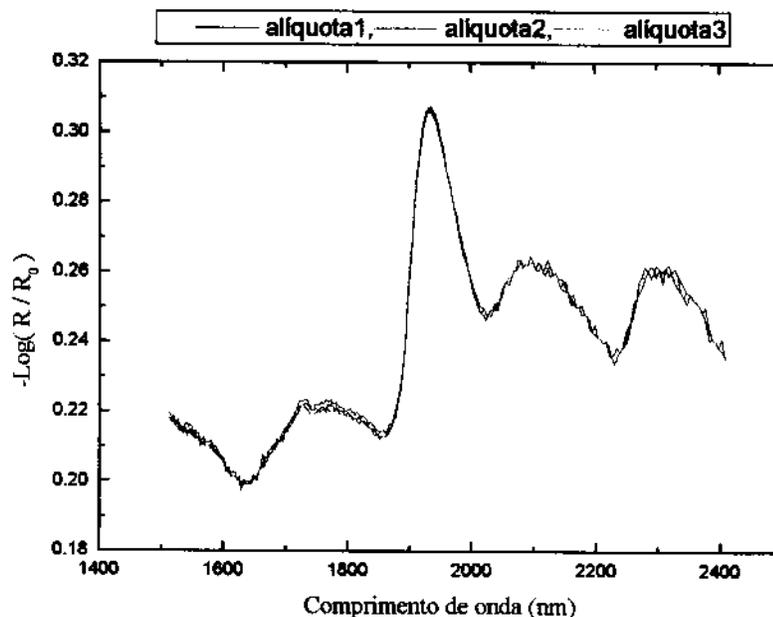


Figura 7.2. Espectros sobrepostos, de três alíquotas de uma mesma amostra.

Além do fato anterior, as medidas feitas com três alíquotas nos ajudariam a identificar possíveis problemas instrumentais durante a obtenção dos espectros de reflectância.

7.2 Filtros para a minimização do ruído em espectros NIRR

Os espectros obtidos apresentaram quantidades razoáveis de ruído. Esta característica pode inviabilizar determinações quantitativas com NIRR, dado que nesta região do espectro eletromagnético as intensidades das absorções são extremamente baixas (10 a 100 vezes menor que no infravermelho médio) e problemas de resolução tornam-se particularmente importantes. Portanto, para viabilizar a determinação quantitativa de nitrogênio nas folhas utilizando espectroscopia NIRR tivemos que minimizar o nível de ruído dos espectros.

7.2.1 Filtro de transformada de Fourier

Como foi descrito na introdução sobre pré-processamento de sinais (item 3, desta tese), devemos representar os espectros no domínio das frequências, cortar as frequências visto serem estas associadas ao ruído e através da transformada inversa recuperar o espectro, sem ruído. Entretanto, qual é o limite inferior para este corte, isto é, qual é a menor frequência que pode ser cortada do espectro de frequências sem que, ao aplicarmos a transformada de Fourier inversa, alteremos o espectro ?

A resposta para a questão anterior não é trivial, e este ainda é (e possivelmente sempre será) um problema não resolvido. Existem algumas sugestões na literatura; entretanto, todas elas são relacionadas a tipos específicos de sinais e não são diretamente extrapoláveis para espectroscopia NIRR. Na falta de melhor guia, resolvemos adotar a seguinte estratégia para o corte de frequências: cortávamos as frequências e avaliávamos a distribuição dos resíduos (espectro original menos espectro ‘filtrado’). Enquanto esta distribuição permanecesse gaussiana, prosseguíamos no corte; caso a forma da distribuição se alterasse, encerrávamos o processo de corte. O suporte teórico para o procedimento adotado é o seguinte: uma vez que o ruído que desejamos minimizar possui características aleatórias, a sua função de distribuição deve ser gaussiana. Logo a distribuição dos resíduos da aplicação do filtro de transformada de Fourier (STFT ou FT simples), isto é, espectro tratado menos o espectro original, também deve ser gaussiana. Na Figura 7.3 temos os espectros tratados com o filtro de FFT sobrepostos aos não tratados. Apresentamos somente os resultados da aplicação do filtro de transformada de Fourier para os espectros das folhas de trigo mas o

mesmo comportamento foi observado para as amostras de folhas de milho e algodão. Na Figura 7.4 representamos a distribuição dos resíduos após o corte das frequências altas, para os espectros das folhas de trigo.

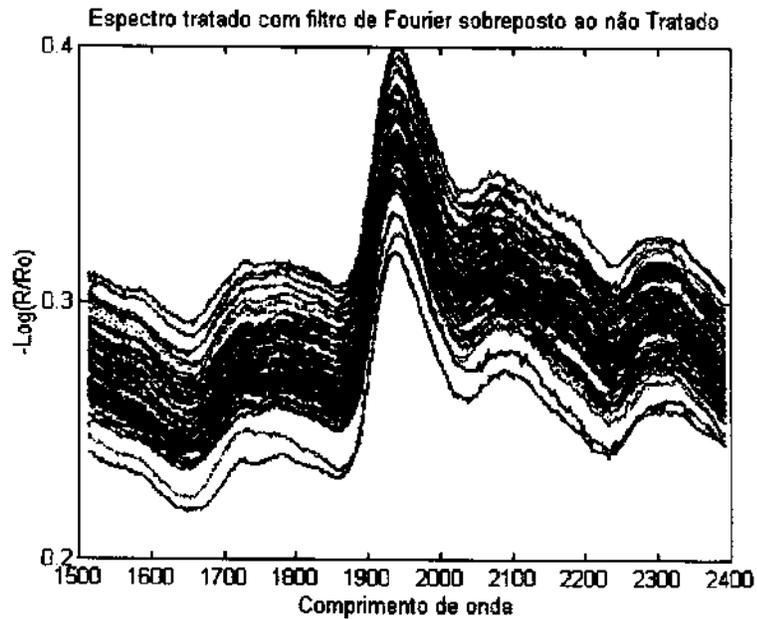


Figura 7.3. Efeito do filtro de transformada de fourier.

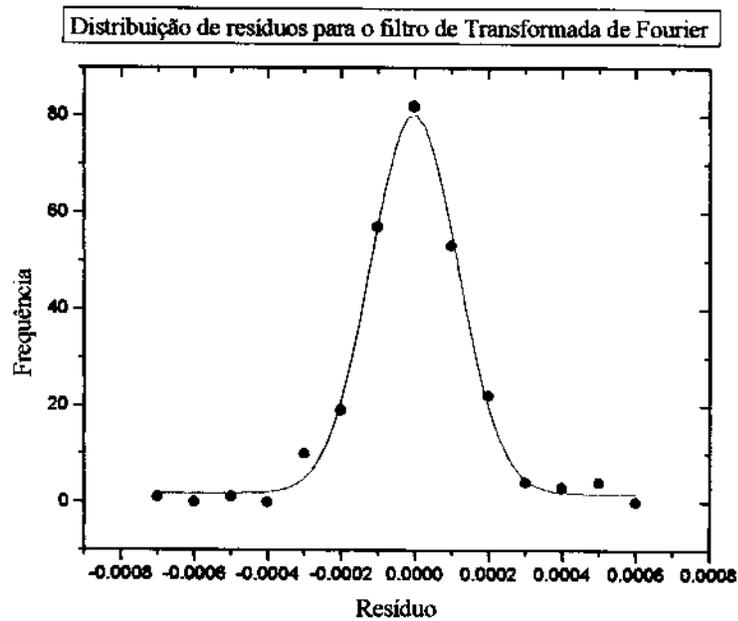


Figura 7.4. Distribuição dos resíduos do filtro de transformada de Fourier (STFT).

O processo adotado para o corte das frequências relativas ao ruído, nos garante que apenas ruído será excluído do espectro original, mas não nos garante, exatamente, que o ruído excluído é todo o ruído, presente no espectro (ou em outro sinal qualquer).

7.2.2 Filtro de transformada de Wavelet

A estratégia utilizada para a redução de ruído foi muito semelhante àquela utilizada no filtro de transformada de Fourier. A única diferença é que no caso das wavelets avaliávamos a distribuição dos resíduos em função do grau de compressão da função de base (Daubechies 4) usada na convolução com o espectro NIRR. Obtivemos que uma única compressão (redução do período da função de base à metade) foi suficiente para a redução do ruído, compressões superiores distorcem o espectro NIRR. Na Figura 7.5 temos os espectro tratados com o filtro de wavelets sobrepostos aos não tratados e na Figura 7.6 representamos a distribuição dos resíduos após a primeira compressão.

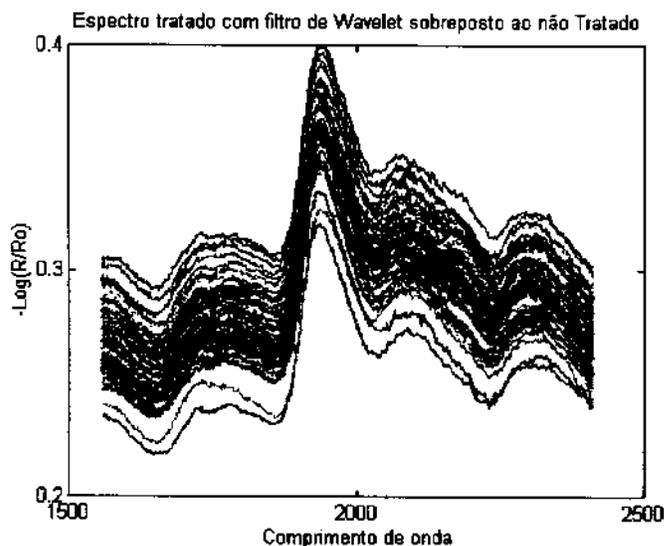


Figura 7.5. Efeito do filtro de wavelet.

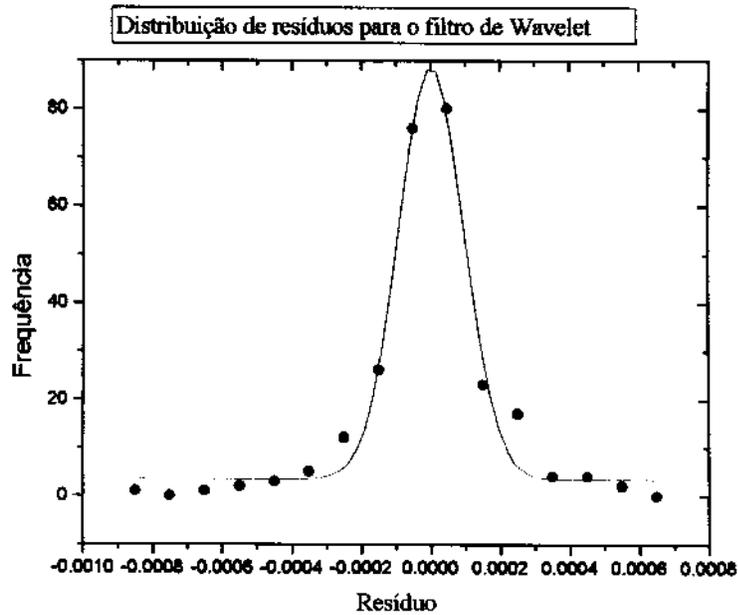


Figura 7.6. Distribuição dos resíduos do filtro de wavelet.

7.2.3 Filtro de Savitzky-Golay

Antes de apresentarmos os resultados vale lembrar que este tipo de filtro ‘suaviza’ os espectros mas não remove ruído, como é o caso dos filtros de transformada de Fourier e Wavelets. Neste filtro, descrito na introdução (secção 3), a interpolação foi feita a cada onze pontos usando a função x^2 . Outras interpolações (número de pontos e função) foram testadas mas descartadas pois causavam distorções nos espectros. Na Figura 7.7 temos os espectro tratados com o filtro de Savitzky-Golay sobrepostos aos não tratados e na Figura 7.8 representamos a distribuição dos resíduos após a aplicação do filtro de Savitzky-Golay.

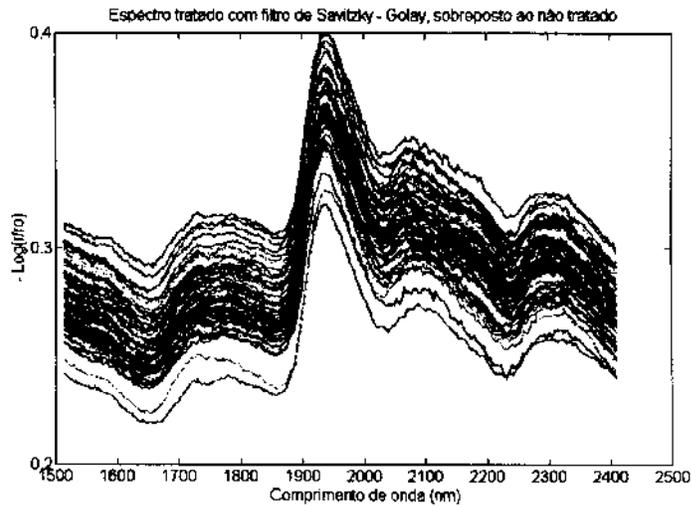


Figura 7.7. Espectros tratados com o filtro de Savitzky-Golay, sobrepostos aos não tratados.

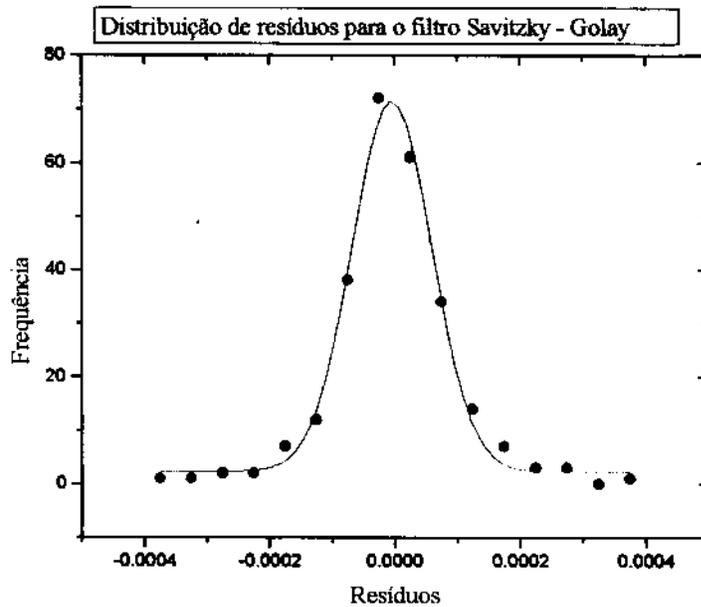


Figura 7.9. Resíduos do filtro de Savitzky-Golay

7.2.4 Comparação dos filtros utilizados

Um modo de comparar os filtros utilizados é avaliar a área da gaussiana ajustada à distribuição dos resíduos. Assim sendo, o filtro cuja

gaussiana apresentar a maior largura à meia altura (ou a maior área) será o mais eficiente. Gaussianas fortemente centradas em torno do valor zero (pequena largura), indicam que, possivelmente, pouco ruído foi removido. Resumindo, os resíduos devem apresentar distribuição gaussiana, centrada em zero e quanto mais larga a gaussiana, melhor o filtro.

Os filtros de Fourier e Wavelets apresentaram desempenho equivalente, visto que a área da gaussiana relativa ao filtro de Fourier (área= 0,025) foi, para todos efeitos práticos, igual à área da gaussiana relativa ao filtro de Wavelet (área = 0,023). Todavia, tanto o filtro de transformada de Fourier quanto o de transformada de Wavelet foram significativamente melhores que o filtro de Savitzky-Golay (área = 0,01).

Além do método anterior avaliamos a desempenho dos filtros de pré-processamento em função dos resultados que produziram nos modelos utilizados (PLS, PLS polinomial, PLS rede neural, redes neurais e redes neurais com pruning), isto é, avaliamos o filtro que permite construir modelos de melhor qualidade, isto é, menores erros. Discutiremos este fato após termos apresentado os resultados obtidos com os modelamentos.

7.3 Correção do espalhamento de luz (MSC, *Multiplicative Scattering Correction*)

Conforme podemos observar nas Figuras 7.10 e 7.11, a correção do espalhamento de luz (MSC) nos permitiu normalizar os espectros, isto é, removemos a inclinação da linha base (“drift”) e o deslocamento dos espectros em relação às ordenadas. Estes efeitos, conforme discutido na introdução desta tese, são relativos, basicamente, ao espalhamento de luz,

o qual é causado por diferenças de granulometria entre as amostras e não por diferenças de concentração, das espécies absorventes.

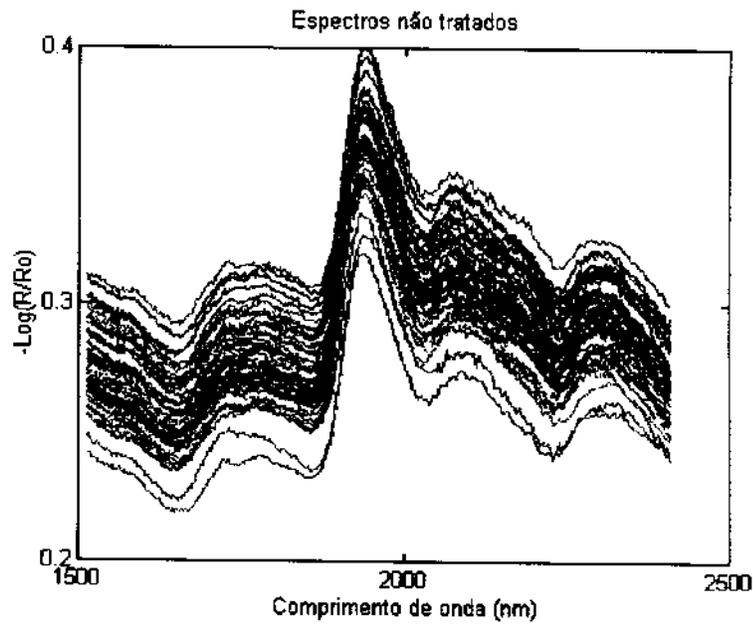


Figura 7.10. Espectros de folhas de trigo, antes do MSC.

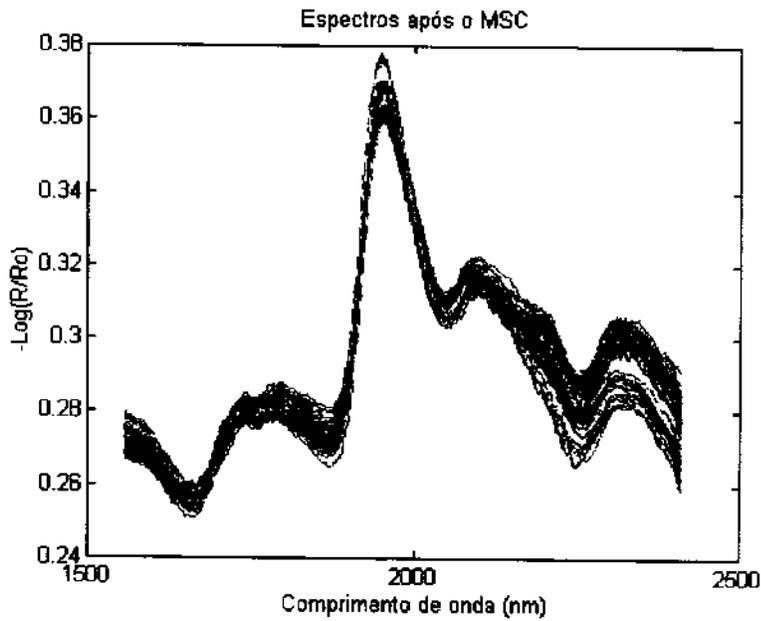


Figura 7.11. Espectros de folhas de trigo após o MSC.

7.4 A umidade das amostras

A água possui bandas de absorção muito intensas no infravermelho próximo, sendo que as duas principais absorções ocorrem em 1450 e 1940 nm. Como 98% do nitrogênio presente nas folhas se apresenta ligado em moléculas de proteína, as quais absorvem em regiões próximas as absorções da água, teremos alguma sobreposição entre as bandas de água e as bandas utilizadas para a determinação de nitrogênio. Assim, se a umidade variar aleatoriamente de amostra para amostra, teremos mais uma variável para modelar, sobre a qual não temos nenhuma informação ou controle. Entretanto seria muito fácil quantificar tal variável, basta fazer uma análise termogravimétrica¹⁰⁰ (TGA) para cada amostra e incluir o valor da umidade nos modelos. Todavia, seria este procedimento realmente necessário?

Além de desnecessário, este procedimento é indesejável visto que, conforme descrevemos na introdução, desejamos métodos rápidos para análise *in situ*. Para evitar este procedimento resolvemos utilizar em todos os modelos (PLS, POLIPLS, NNPLS, Redes neurais e Redes neurais com apodização) não os espectros NIRR, mas os escores destes espectros, pois conforme descrevemos na introdução desta tese (secção 3.3), podemos utilizar somente os escores relativos aos primeiros componentes principais, os quais não devem capturar variações aleatórias, como por exemplo, a umidade das amostras. Utilizando os escores relativos aos componentes que capturam a maior parte da variância dos dados, '*filtramos*' (reduzimos) o efeito da umidade. Naturalmente, os escores foram obtidos após o pré-processamento completo dos espectros.

Fizemos a análise termogravimétrica das amostras e observamos que todas apresentavam, praticamente, a mesma umidade, cerca de 5%. Na sequência de Figuras abaixo, apresentamos as curvas de perda de massa em função da temperatura, para a amostra que apresentou o maior teor de umidade (Figura 7.12), para a amostra com menor teor de umidade (Figura 7.13) e para uma amostra com teor de umidade intermediário (Figura 7.14).

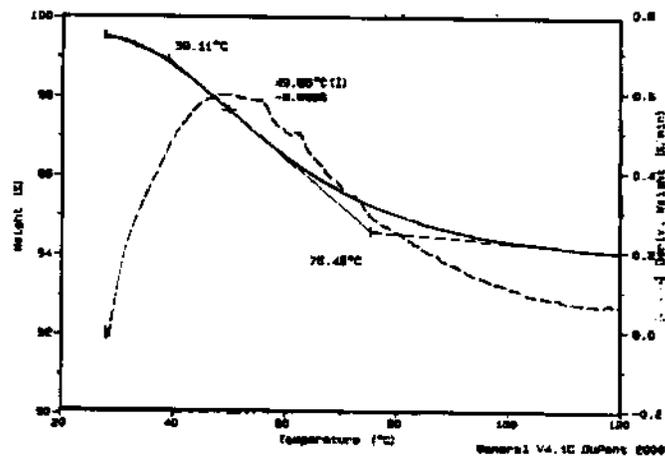


Figura 7.12. Curva representativa da variação da massa em função da temperatura para a amostra com maior teor de umidade (5,5 %).

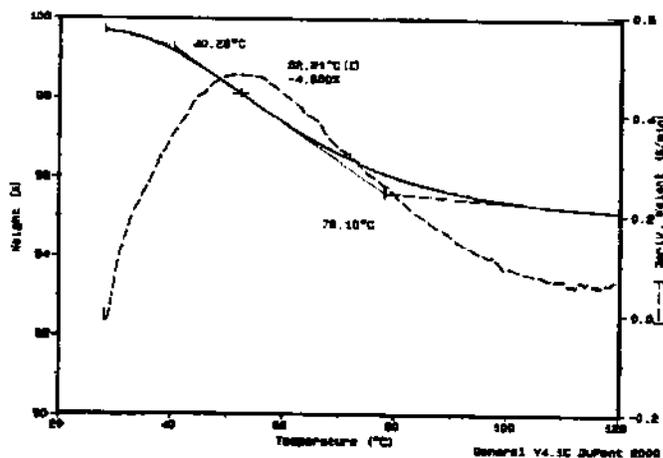


Figura 7.13. Curva representativa da variação da massa em função da temperatura para a amostra com menor teor de umidade (4,65 %).

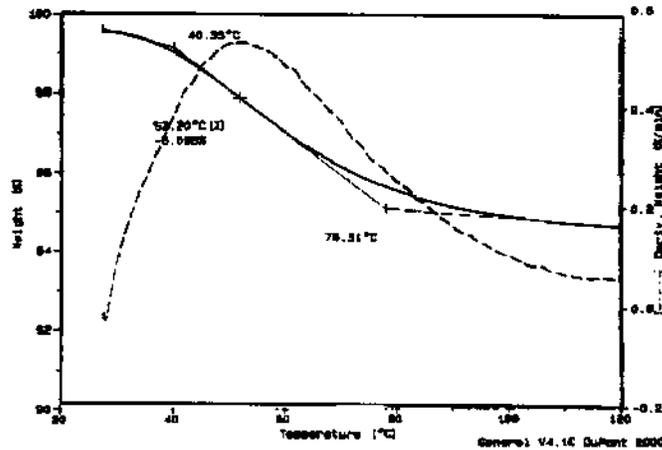


Figura 7.14. Curva representativa da variação da massa em função da temperatura para a amostra com teor de umidade intermediário (5,09 %).

O fato de todas as amostras apresentarem, praticamente a mesma umidade, indica que o processo de secagem foi igualmente eficiente (todas amostram com a mesma umidade) e que possivelmente o uso dos escores dos espectros para a construção dos modelos, possivelmente eliminará o efeito da umidade.

7.5 Seleção das amostras¹⁰¹

As amostras foram separadas em três conjuntos distintos. Um conjunto para calibração, outro para validação e outro conjunto para previsão. As amostras de folhas trigo foram distribuídas da seguinte maneira: 36 para calibração, 12 para validação e 12 para previsão.

Para executar a seleção das amostras de modo eficiente fizemos a análise de componentes principais (PCA) dos espectros (já tratados com filtro de Fourier e MSC) obtidos para as folhas de trigo. O gráfico dos escores do primeiro componente principal PC1 versus os escores do

segundo componente principal (PC2), nos permite reduzir a dimensão do sistema e visualizar possíveis agrupamentos (“clusters”) dos espectros, possibilitando a seleção homogênea das amostras usadas na calibração, validação e previsão (Figura 7.15).

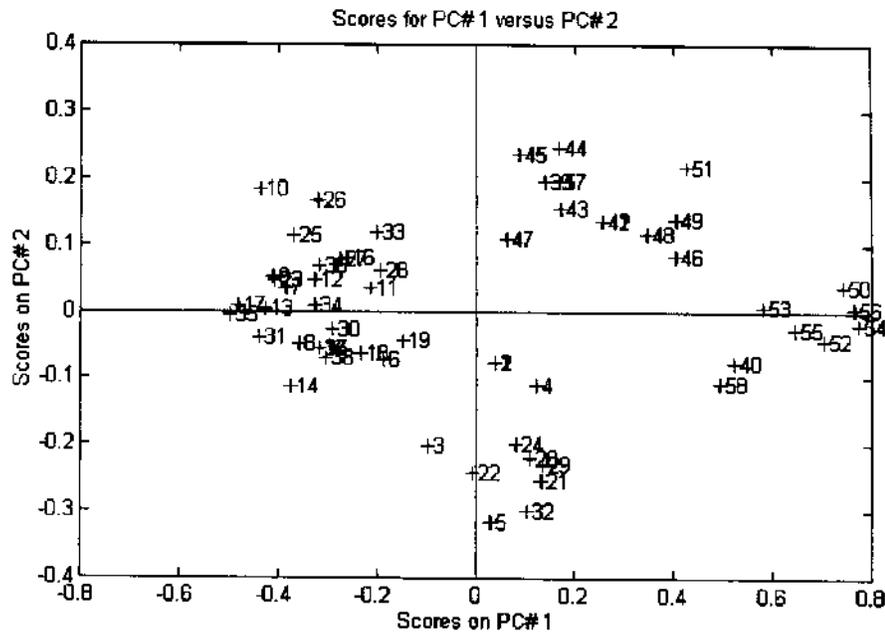


Figura 7.15. Diagrama utilizado para a seleção das amostras.

No diagrama apresentado na Figura 7.15, as cruces vermelhas representam uma determinada amostra e quanto mais próximas as cruces, mais semelhantes serão os espectros das amostras.

Além disso, o diagrama da Figura 7.15 nos auxilia a garantir que os espectros referentes ao menor e ao maior teor de nitrogênio, estejam no conjunto de calibração, visto que estamos construindo modelos locais, para os quais não devemos, à princípio, fazer extrapolações.

7.6 Modelamentos

Desenvolvemos modelos de calibração exclusivamente para as amostras de folhas de **trigo**. As amostras de milho e algodão foram utilizadas com outra finalidade que se tornará clara, mais adiante no texto.

Os modelos foram construídos combinando-se todas as técnicas de pré-processamento de sinais com todas as técnicas de modelamento, resultando em um total de oitenta modelos.

A construção dos modelos foi feita utilizando-se, não o espectro todo (todas intensidades de reflectância, lidas em todos os comprimentos de onda) mas, os cinco primeiros componentes principais (PC's) deste espectros, os quais capturavam aproximadamente, 95% da variância dos dados. A opção por cinco PC's deve-se ao fato de que se fosse utilizado um número superior ou inferior a este, aumentávamos os erros de validação e previsão em todos os modelos. Entretanto, nos modelos em que empregamos a derivada segunda dos espectros utilizamos vinte componentes principais (PC's), visto que neste caso estes PC's forneceram modelos com menores erros de calibração, validação e previsão.

7.6.1 Avaliação do desempenho dos modelos

Com o objetivo de avaliar a desempenho das técnicas de pré-processamento e de modelamento, executamos todas as combinações possíveis entre as técnicas de pré-processamento e os métodos de modelamento e avaliamos a porcentagem de erro padrão, para os conjuntos de calibração (%SEC), validação (%SEV) e previsão (%SEP).

As equações utilizadas para o cálculo destes erros foram as equações 7.1 e 7.12, respectivamente.

$$\%SEC = \frac{1}{\bar{c}_c} \sqrt{\frac{\sum_{i=1}^{n_c} (y_i - \hat{y}_i)^2}{n_c - k - 1}} \times 100 \quad \text{eq.7.1}$$

$$\%SEP = \frac{1}{\bar{c}_p} \sqrt{\frac{\sum_{i=1}^{n_p} (y_i - \hat{y}_i)^2}{n_p}} \times 100 \quad \text{eq.7.2}$$

onde \bar{c}_c e \bar{c}_p são as concentrações médias do conjunto de calibração e previsão respectivamente, n_c e n_p representam o número de amostras usadas na calibração e na previsão, k é o número de componentes principais (PC's) utilizados nos modelos, y_i é valor real e \hat{y}_i o valor do teor de nitrogênio previsto pelo modelo. A equação usada para o cálculo da %SEV é idêntica à equação da %SEP. Além dos parâmetros anteriores, lançamos em gráfico y_i contra \hat{y}_i , ajustamos uma relação linear entre eles e calculamos o coeficiente de correlação (r^2), o coeficiente angular (a) e o linear (b), para os três conjuntos (calibração, validação e previsão).

Uma vez calculados os parâmetros anteriores, podemos avaliar a desempenho individual e relativa dos modelos. Esta comparação foi feita utilizando-se basicamente os valores de %SEC, %SEV e %SEP sendo que os demais parâmetros (coef. de correlação, coef. angular e coef. linear) serviram de modo auxiliar. Utilizamos a seguinte estratégia:

1) Desempenho individual

Modelos que apresentavam %SEP e/ou %SEV duas vezes maior que o %SEC, foram foram imediatamente descartados, visto ser esta uma indicação de sobreajuste (overfitting), isto é, modelos com baixos erros de calibração e altos de previsão e validação. Além disso, também foram descartados modelos com erros (% SEC, % SEP ou % SEV) muito acima do aceitável (digamos 15 % de erro). Esta estratégia reduziu consideravelmente o número de modelos a serem analisados, isto é, comparados.

Assim sendo, a próxima tarefa é comparar todos os modelos restantes e escolher aqueles que apresentem melhor desempenho. Uma maneira de estabelecer esta comparação é construir uma tabela (Tabela 1) com os valores de % SEC, % SEV, % SEP, coeficiente de correlação (r^2), coeficiente angular e coeficiente linear, para cada um dos modelos e em seguida, compará-los. Esta tarefa é extremamente tediosa pois muitos modelos apresentam bons valores de % SEP, % SEC e % SEV mas, valores ruins para os outros parâmetros, dificultando a escolha dos melhores modelos. Entretanto, existe um modo de facilitar as coisas, como veremos à seguir.

Tabela 1. Resultados obtidos com os métodos de modelamento

Nº / modelo	Filt.	Der	SEC	SEP	SEV	r_c^2	r_p^2	r_v^2	a_c	a_p	a_v	b_c	b_p	b_v
1 RN	FFT	não	4.78	6.98	6.13	0.99	0.99	0.99	1.00	0.99	0.99	0.01	0.00	0.03
2 RNP	FFT	não	4.80	7.00	5.80	0.99	0.99	0.99	0.99	0.99	0.99	0.00	0.01	0.02
3 PLSRN	FFT	não	6.10	8.80	6.24	0.99	0.99	0.99	0.98	0.99	0.99	0.05	0.01	0.03
4 PLS	FFT	não	6.75	6.81	10.5	0.99	0.99	0.98	0.98	0.98	0.99	0.06	0.06	0.01

5	POLIPLS	FFT	não	6.42	7.83	11.1	0.99	0.98	0.98	0.96	0.92	0.97	0.15	0.26	0.09
6	RN	FFT	sim	5.03	6.33	10.5	0.99	0.99	0.97	0.98	1.03	0.99	0.06	0.04	0.11
7	RNP	FFT	sim	5.63	7.40	9.99	0.99	0.99	0.98	0.99	1.00	1.02	0.03	0.04	0.03
8	PLSRN	FFT	sim	6.13	6.19	8.60	0.99	0.99	0.99	0.98	0.99	0.99	0.05	0.00	0.04
9	PLS	FFT	sim	5.46	7.61	10.7	0.99	0.98	0.97	0.98	0.97	0.93	0.04	0.19	0.19
10	POLIPLS	FFT	sim	3.94	5.00	11.3	0.99	0.99	0.97	0.99	1.00	1.00	0.04	0.05	0.03
11	RN	SG	não	5.09	8.84	10.9	0.99	0.98	0.96	0.98	0.97	0.95	0.06	0.06	0.09
12	RNP	SG	não	6.43	6.55	8.65	0.98	0.99	0.97	0.98	0.96	0.97	0.04	0.06	0.04
13	PLSRN	SG	não	6.59	8.30	9.40	0.99	0.98	0.98	0.98	0.92	0.97	0.11	0.20	0.09
14	PLS	SG	não	5.91	6.20	8.00	0.99	0.99	0.98	0.98	0.95	0.96	0.06	0.2	0.1
15	POLIPLS	SG	não	5.42	7.83	11.1	0.99	0.98	0.97	0.95	0.92	0.96	0.15	0.26	0.01
16	RN	SG	sim	4.98	16.5	16.4	0.99	0.95	0.93	0.99	1.08	1.03	0.01	0.34	0.27
17	RNP	SG	sim	6.43	6.55	8.61	0.99	0.99	0.98	0.98	0.96	0.97	0.05	0.13	0.13
18	PLSRN	SG	sim	6.60	8.20	9.40	0.99	0.98	0.98	0.97	0.92	0.97	0.11	0.20	0.10
19	PLS	SG	sim	4.90	15.7	11.5	0.99	0.93	0.96	0.99	1.19	1.04	0.02	0.71	0.01
20	RN	não	não	5.49	5.99	11.9	0.99	0.99	0.95	0.98	0.96	0.94	0.06	0.17	0.12
21	RNP	não	não	6.51	6.9	8.1	0.99	0.99	0.98	0.98	0.95	0.96	0.03	0.14	0.15
22	PLS	não	não	4.60	6.85	9.42	0.99	0.98	0.98	0.99	0.95	0.97	0.03	0.15	0.10
23	POLIPLS	não	não	4.98	8.56	12.0	0.99	0.98	0.96	0.96	0.92	0.97	0.13	0.19	0.02

* Na construção da Tabela 1 utilizamos as seguintes abreviações: Redes neurais com apodização (RNP), Redes neurais (RN), Mínimos Quadrados Parciais - redes neurais (NNPLS), Mínimos Quadrados Parciais (PLS), Mínimos Quadrados Parciais-ajuste polinomial (POLIPLS), Filtro de transformada de Fourier (FFT), Filtro de Savitzky-Golay (SG) e derivada (Der.). Os sub índices c, v e p, indicam calibração, validação e previsão, respectivamente.

Um modo de transpor a dificuldade anterior é aplicar a análise de componentes principais (PCA) aos resultados obtidos com os modelos. A aplicação foi feita do seguinte modo: elaboramos uma matriz em que as linhas, representavam um método de modelamento e as colunas os parâmetros utilizados na avaliação do modelo (% SEP, % SEC , etc). Em seguida aplicamos a análise de componentes principais (PCA) sobre esta matriz e traçamos, em três dimensões, o gráfico paramétrico dos escores, para os três primeiros componentes principais (um escore em cada coordenada) os quais capturavam 99,5% da variância dos dados. O objetivo deste procedimento é absolutamente idêntico ao utilizado para a seleção das amostras (item 7.5 desta tese), isto é, observar possíveis agrupamentos dos resultados. O gráfico obtido, está apresentado na Figura 7.16.

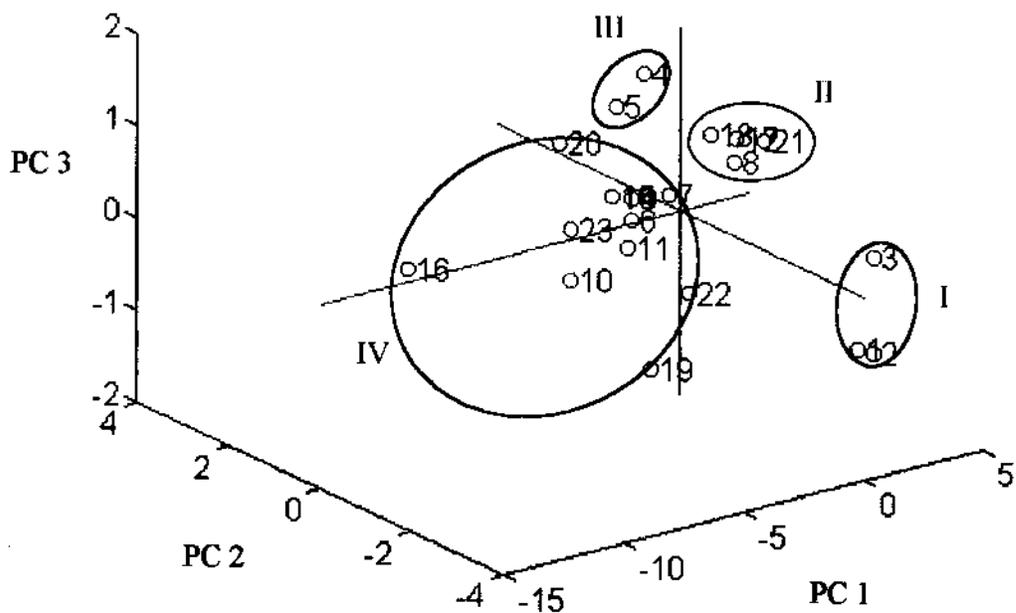


Figura 7.16. Gráfico utilizado para a seleção dos modelos.

No gráfico apresentado na Figura 7.16, podemos observar que existem quatro grupos básicos de resultados. Naturalmente, as curvas que delimitam os grupos são secções retas de superfícies envoltórias destes grupos, dado que estamos trabalhando em um espaço tridimensional.

Na análise anterior, não incluímos os modelos em que se utilizou o filtro de transformada de Wavelet pois, os resultados obtidos com os modelos em que se utilizou este tipo de filtro, são praticamente idênticos aos construídos com filtro de transformada de Fourier. Portanto, a inclusão destes resultados não acrescentaria nada à existência ou não de possíveis agrupamentos entre os dados, mas prejudicaria muito a visualização destes agrupamentos, devido ao aumento de pontos no gráfico da Figura 7.16.

A comparação dos componentes de um determinado grupo com a Tabela 1, nos leva à seguinte classificação:

➤ Grupo 1 : modelos com o melhor desempenho relativa

Neste grupo encontramos os modelos 1, 2 e 3. Como pode ser observado, na Tabela 1, todos estes modelos são baseados em redes neurais.

➤ Grupo 2 : modelos com desempenho intermediário

Neste grupo encontramos os modelos 8, 17, 18 e 21, todos baseados em redes neurais.

➤ Grupo 3 : modelos com desempenho satisfatório

Neste grupo encontramos os modelos 4 e 5. O modelo 4 é um modelo construído com o PLS e o modelo 5, com o PLS polinomial.

➤ Grupo 4 : modelos com desempenho insatisfatório

Neste grupo (obviamente) encontramos os modelos restantes, os quais não serão discutidos dado à sua baixa desempenho. Neste grupo, dentre outros, encontramos: modelos em que não se empregou filtro para ruído, modelos em que se utilizou a derivada primeira dos espectros e filtro de Savitzky-golay.

Terminada a análise anterior, fica simples obter os modelos que apresentaram melhores resultados, para cada um dos métodos de modelamento de dados. Escolhemos um modelo para cada método de modelamento, já que um dos nossos objetivos é avaliar a desempenho relativa dos seguintes métodos de modelamento: Redes Neurais com Apodização (RNP), Redes Neurais (RN), PLS redes neurais (NNPLS), PLS e PLS polinomial (POLIPLS).

Escolhemos os modelos do grupo I e do grupo III. Não escolhemos os modelos do grupo II, pois este grupo é formado exclusivamente por modelos baseados em redes neurais, os quais já foram escolhidos, quando selecionamos os modelos do grupo I. Assim sendo, os modelos selecionados foram:

① Redes neurais com apodização, correção de espalhamento de luz (MSC) e filtro de transformada de Fourier.

② Redes neurais, correção de espalhamento de luz (MSC) e filtro de transformada de Fourier.

③ PLS Redes Neurais, correção de espalhamento de luz (MSC) e filtro de transformada de Fourier.

④ PLS, correção de espalhamento de luz (MSC) e filtro de transformada de Fourier.

⑤ PLS Polinomial, correção de espalhamento de luz (MSC) e filtro de transformada de Fourier.

2) Desempenho relativa¹⁰²:

Selecionados os modelos que desejamos comparar (modelos 1 até 5 da página anterior), centralizaremos toda a análise subsequente sobre estes modelos, para obtermos uma classificação dos métodos de modelamento, isto é, na etapa anterior fizemos uma classificação grosseira, que será refinada nesta etapa.

Uma vez que os modelos selecionados anteriormente não apresentam sobreajuste (overfitting), podemos fundir os conjuntos de validação e previsão em um único conjunto, aqui chamado de conjunto de previsão combinado, e calcular para este conjunto o erro padrão de previsão combinado (% SEPC). Lembre-se, o objetivo de separar o conjunto de amostras de teste (previsão + validação) em dois conjuntos de 12 amostras (12 para validação e 12 para previsão), foi somente o de identificar modelos com sobreajuste.

A seguir apresentaremos em detalhe a desempenho de cada um dos cinco métodos de modelamento selecionados.

① PLS

Utilizamos correção de espalhamento de luz (MSC), 5 PC's e filtro de Fourier. Os resultados dos modelos utilizando o filtro de Wavelets ou Fourier foram estatisticamente (teste F) iguais, portanto, a opção por um deles pode ser feita em função da facilidade de implementação e no caso específico destes dados, optamos pelo filtro de transformada de Fourier.

% SEC	% SEPC
6,75	8,91

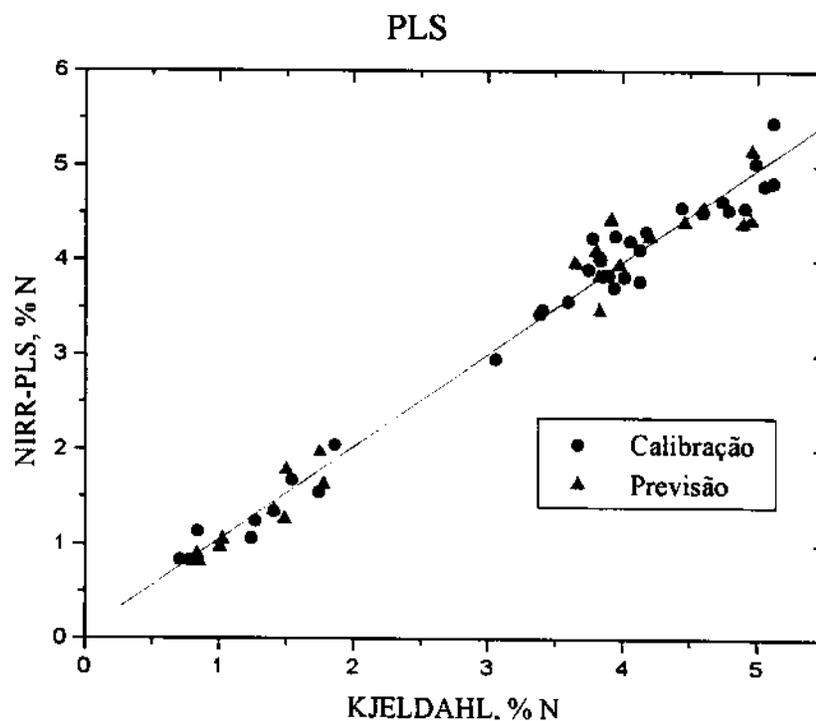


Figura 7.17. Valor previsto, para o teor de nitrogênio, pelo modelo de calibração NIRR-PLS versus o valor obtido com o método Kjeldahl.

② PLS Polinomial (POLIPLS)

Utilizamos correção de espalhamento de luz(MSC), 5 PC's e filtro de transformada de Fourier. A função utilizada para modelar a relação entre os escores da matriz dos espectros e os escores da matriz das concentrações foi a quadrática (x^2).

% SEC	% SEPC
6,42	9,65

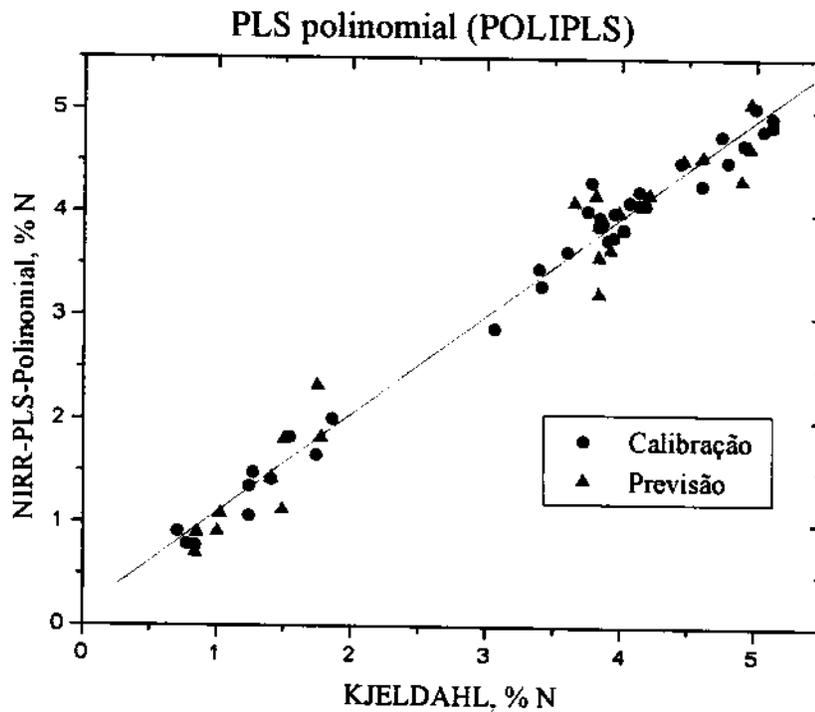


Figura 7.18. Valor previsto, para o teor de nitrogênio, pelo modelo de calibração NIR-PLS-POLIPLS versus o valor obtido com o método Kjeldahl.

③ Redes Neurais(RN)

Utilizamos como técnicas de pré-processamento: correção de espalhamento de luz(MSC), 5 PC's e filtro de Fourier.

% SEC	% SEPC
4,78	6,53

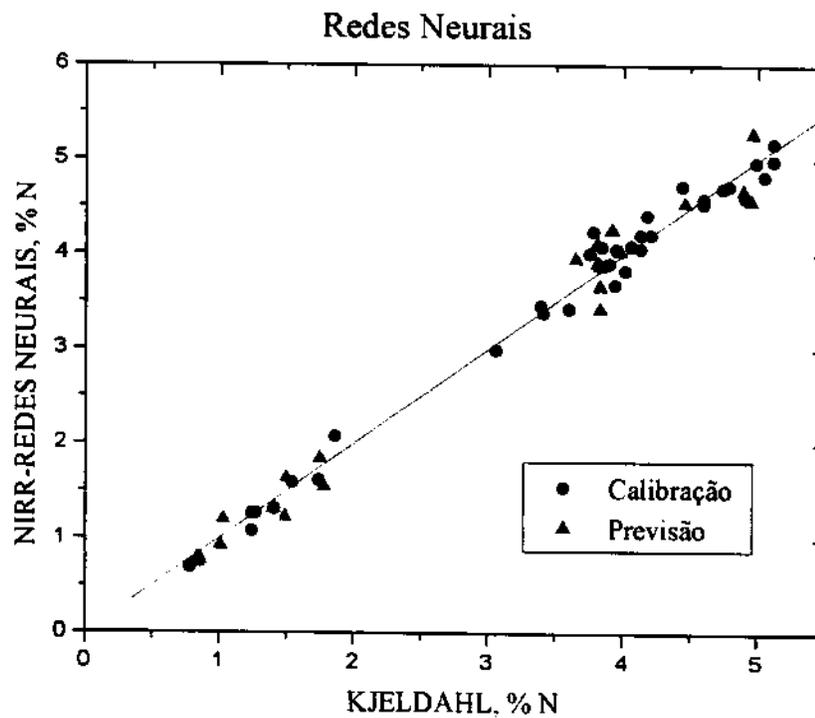


Figura 7.19. Valor previsto, para o teor de nitrogênio, pelo modelo de calibração NIRR-RN versus o valor obtido com o método Kjeldahl.

A arquitetura da rede neural que forneceu os menores erros de calibração (a arquitetura foi otimizada da maneira descrita no item 4.1.6), validação e previsão está representada na Figura 7.20 .

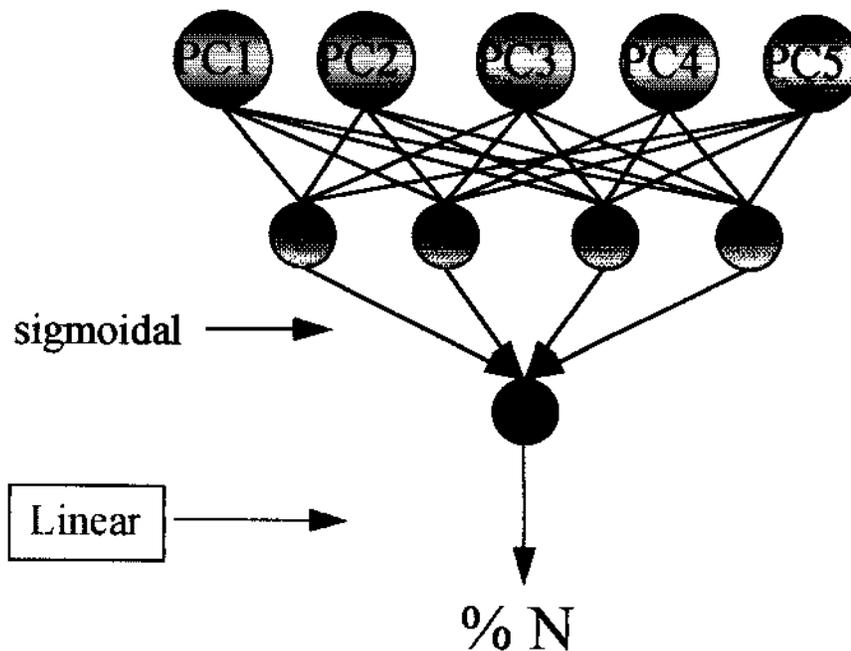


Figura 7.20. Representação da rede neural utilizada no nodelamento. Esta rede possui uma camada de entrada com cinco neurônios, uma camada intermediária com quatro neurônios e uma camada de saída com um único neurônio. A função de transferência utilizada entre a camada intermediária e a camada de saída foi a função sigmoide. Na camada de saída utilizou-se a função linear.

④ PLS Rede Neural (NNPLS)

Utilizamos correção de espalhamento de luz (MSC), 5 PC's e Filtro de transformada de Fourier.

% SEC	% SEPC
6,10	7,64

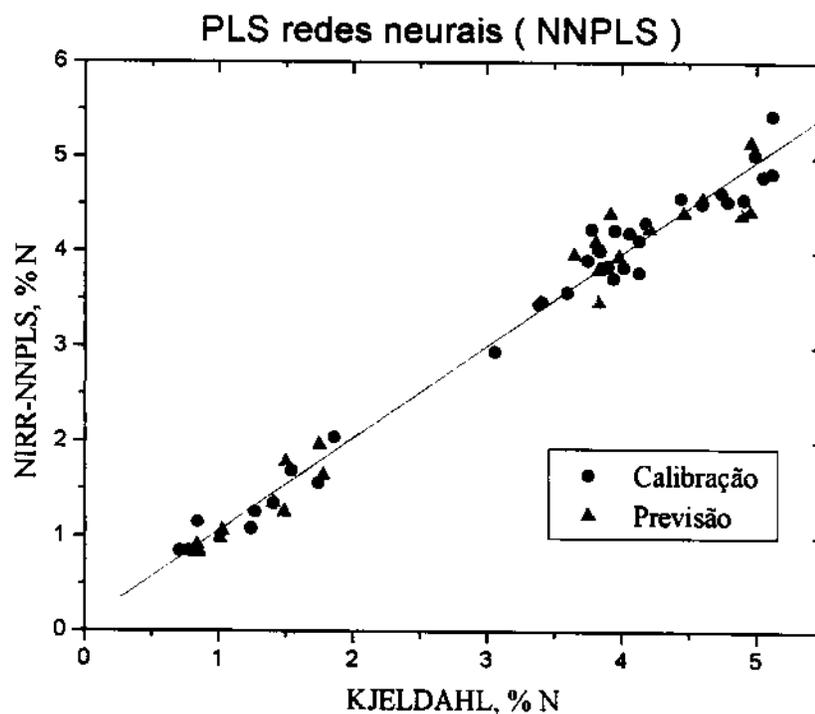


Figura 7.21. Valor previsto, para o teor de nitrogênio, pelo modelo de calibração NIRR-NNPLS versus o valor obtido com o método Kjeldahl.

Nesta versão não-linear do PLS a relação entre os escores da matriz dos espectros e os escores da matriz das concentrações foi modelada por uma rede neural com a seguinte arquitetura: uma camada de entrada com 5 neurônios (um para cada PC), três na camada intermediária e um na camada de saída. Entre a camada intermediária e a camada de saída utilizamos a função sigmoideal e na camada de saída a função linear.

⑤ Redes Neurais com Apodização (RNP)

Utilizamos como técnicas de pré-processamento: correção de espalhamento de luz (MSC), 5 PC's e filtro de transformada de Fourier.

% SEC	% SEPC
4,8	6,29

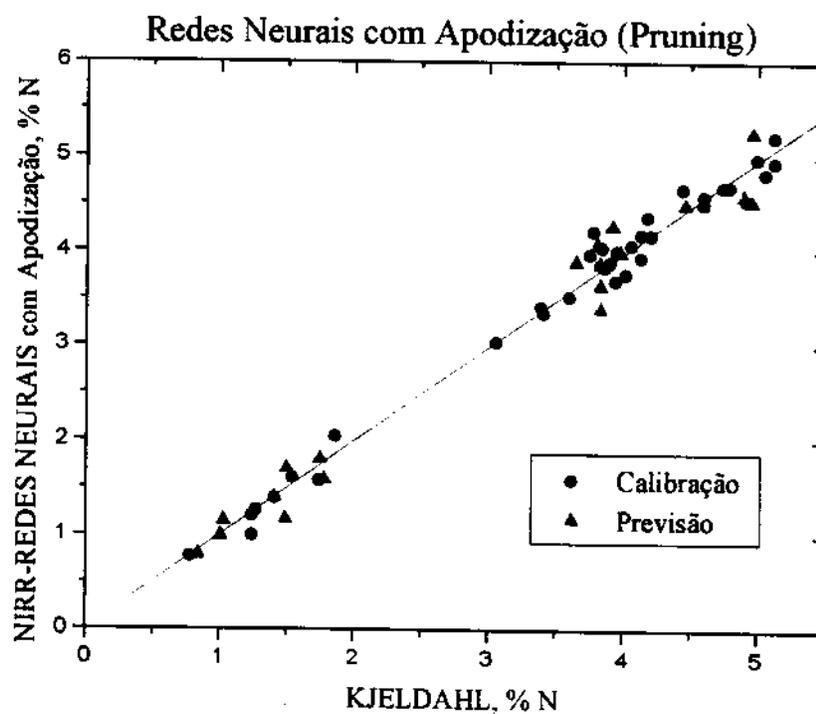


Figura 7.22. Valor previsto, para o teor de nitrogênio, pelo modelo de calibração NIRR-RNP versus o valor obtido com o método Kjeldahl.

A rede neural foi inicializada com a arquitetura apresentada na Figura 7.23:

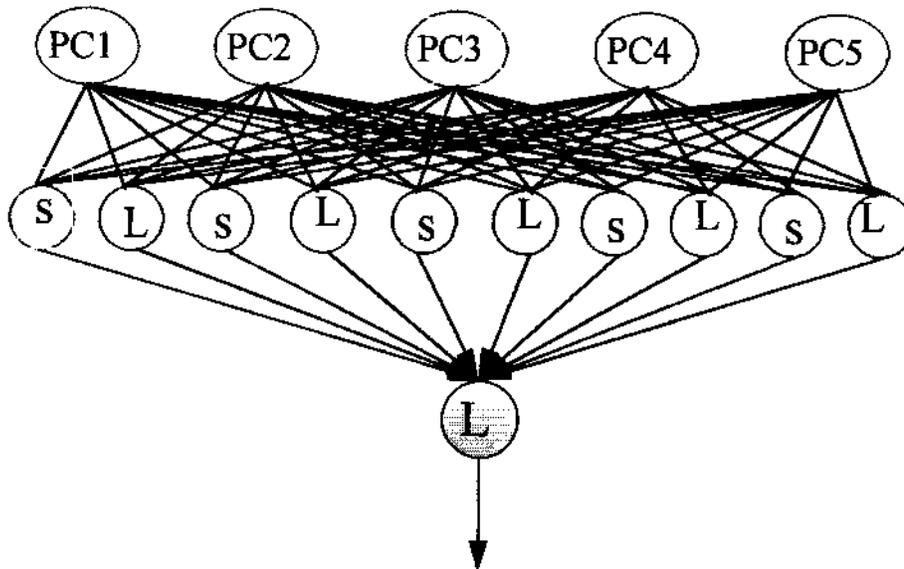


Figura 7.23. Rede neural de inicialização. L representa a função de transferência linear e S a função de transferência sigmoideal. Nesta rede neural temos uma camada de entrada com cinco neurônios, uma camada intermediária com dez neurônios e uma camada de saída com um neurônio.

Observe na Figura 7.23 que os neurônios da camada intermediária foram inicializados alternando-se as funções de transferência linear (L) e sigmoideal (S). Assim sendo, esta rede pode modelar igualmente bem relação lineares e não-lineares. Além disso, ao final do processo de apodização podemos verificar quais neurônios foram eliminados e, portanto, se somente os lineares tiverem sido eliminados, temos uma indicação de que os sistema em questão apresentava características realmente não-lineares; caso contrário o sistema seria linear. Naturalmente, se restassem neurônios lineares e não-lineares, teríamos ainda um sistema não-linear.

Após o treinamento da rede neural, restaram somente dois neurônios. Foram eliminados oito neurônios e com eles a possibilidade de sobreajuste (overfitting). A arquitetura da rede neural após a apodização dos neurônios dispensáveis, está representada na Figura 7.24:

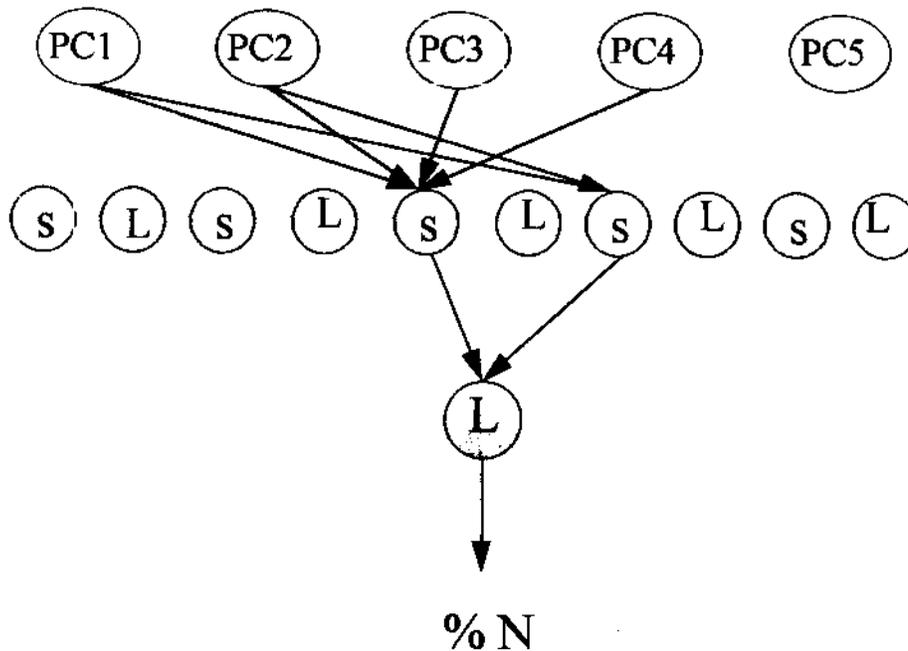


Figura 7.24. Rede neural após o corte dos neurônios desnecessários, apenas dois neurônios são suficientes para o modelamento.

conforme podemos observar na Figura 7.24, restaram somente neurônios não lineares e, portanto, nosso sistema possui características não-lineares.

Uma grande vantagem de utilizar-se redes neurais com pruning é que evitamos a característica mais perigosa das redes neurais: o sobreajuste (overfitting). Outra vantagem é que não necessitamos testar várias arquiteturas, isto é, não necessitamos variar o número de neurônios

na camada intermediária para em seguida escolhermos aquele que apresenta os menores erros de calibração, validação e previsão. Neste método “matamos dois coelhos com uma cajadada só”.

Para confirmarmos que os modelos não apresentam sobreajuste (overfitting), apresentamos em um gráfico de barras (Figura 7.25), os valores de % SEC e % SEPC para cada um dos modelos selecionados.

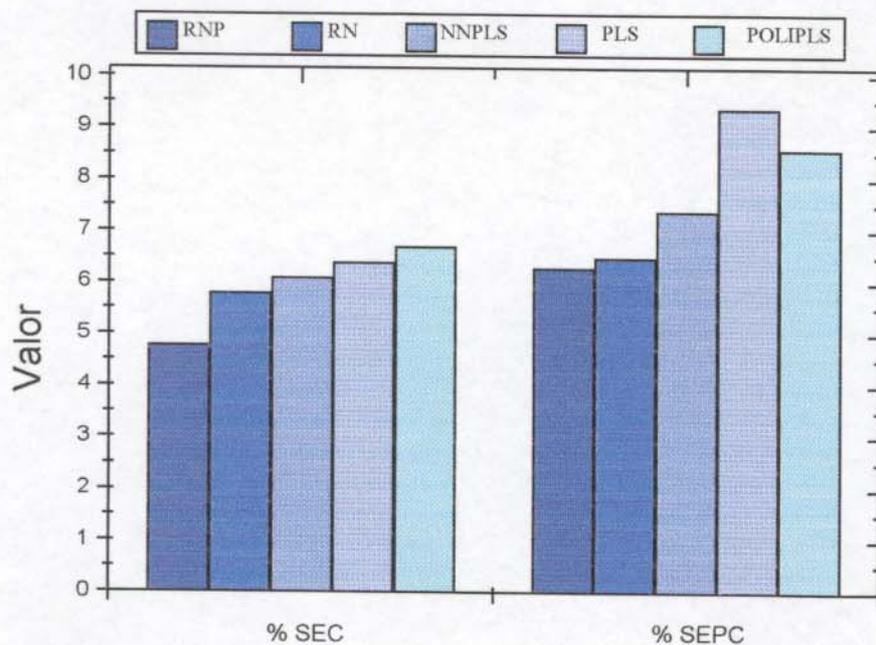


Figura 7.25. Gráfico de barras dos valores de % SEC e % SEP, para todos modelos testados.

Conforme pode ser facilmente observado na Figura 7.25, os modelos realmente não apresentam sobreajuste (overfitting) pois os valores de %SEC e %SEPC não são significativamente diferentes.

A análise, superficial, dos valores de % SEPC, apresentados para cada um dos modelos nos indica que a rede neural com apodização

(pruning) foi o método de modelamento mais eficiente, seguido pelas redes neurais ordinárias, PLS redes neurais, PLS e PLS polinomial. A desempenho das redes neurais com pruning foi 5 % melhor que as redes neurais ordinárias, 18 % melhor que o PLS redes neurais, 30 % melhor que o PLS e 35 % melhor que o PLS polinomial. Entretanto, seriam estas diferenças entre os valores de %SEPC, estatisticamente significativas, de modo a justificar a classificação anterior ?

Para comparar dois valores de %SEPC, podemos utilizar o teste F, dado pela equação 7.3, em que p_i e p_j são o número de amostras utilizadas no conjunto de previsão^{102. 103}.

$$F(p_i, p_j) = \left(\frac{SEP_i}{SEP_j} \right)^2 \quad \text{eq.7.3}$$

O valor crítico do teste $F_{critical}$ anterior, no nível de 95 % de confiança, para o conjunto de 24 amostras utilizadas na previsão, é:

$$F_{critical} = 1,98$$

valores calculados de F , maiores que o valor de $F_{critical}$ indicam que o desempenho dos modelos são significativamente diferentes, caso contrário os desempenhos dos modelos, serão iguais.

Quando aplicamos o teste F aos valores de % SEPC, obtidos com os diferentes métodos de modelamento selecionados, observamos que as redes neurais com pruning são significativamente melhores que o PLS ($F = 2.1$) e o PLS polinomial ($F = 2.4$). Entretanto, quando comparamos, utilizando o teste F , as redes neurais com pruning com o PLS Rede neural ($F = 1.5$) e com as redes neurais ordinárias ($F = 1.1$), observamos que as diferenças verificadas nos valores de %SEPC, não são significativas no nível de 95 % de confiança.

Os resultados obtidos com o teste F , no nível de 95 % de confiança, nos indicam que a desempenho dos cinco algoritmos de modelamento, pode ser finalmente classificada, basicamente, em duas classes. Na classe com menores erros, temos os modelos construídos com as redes neurais com pruning (RNP), redes neurais (RN) e PLS rede neural (NNPLS). Na classe com maiores erros, temos os modelos construídos com o PLS e PLS polinomial (POLIPLS). Modelos pertencentes a uma determinada classe, não apresentam diferenças nos valores de %SEPC, significativas ao nível de 95 % de confiança. Na Figura 7.26, apresentamos o gráfico de barras

dos valores do teste F , para cada um dos métodos de modelamento, onde se pode observar as duas classes de modelos, anteriormente citadas.

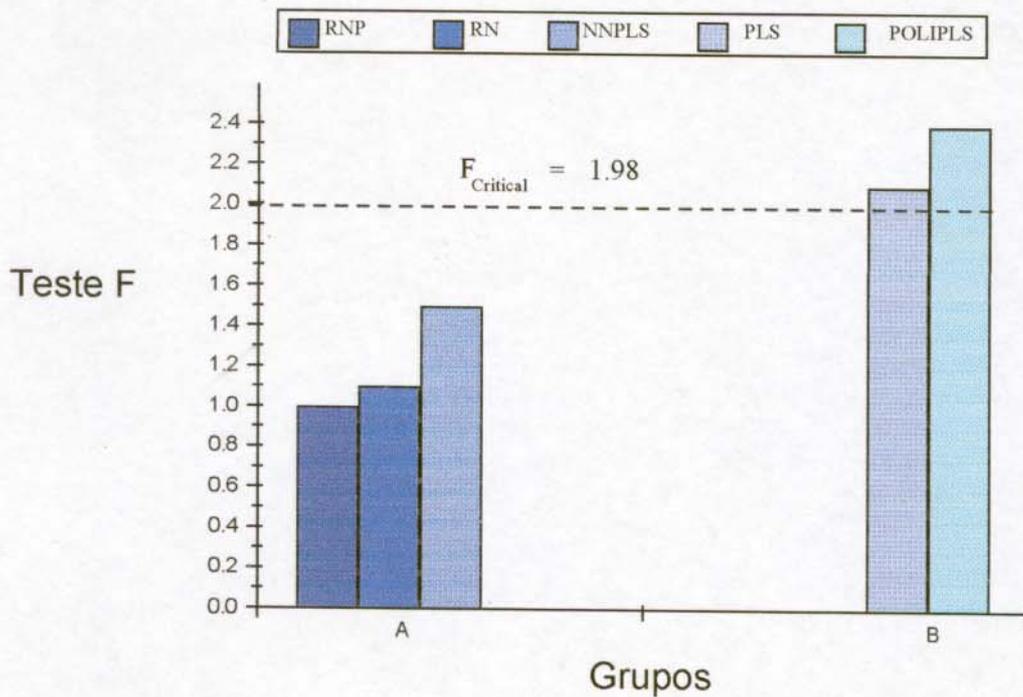


Figura 7.26. Gráfico de barras dos valores de teste F , utilizados para a comparação da desempenho dos modelos. (A) grupo de menores erros, (B) grupo de maiores erros.

Finalmente, a classificação dos modelos, segundo o teste F , é a seguinte:

1) Redes Neurais com apodização (prunning), Redes neurais e PLS redes neurais.

2) PLS e PLS polinomial

Conforme observamos não há diferenças significativas entres os modelos da primeira classe e a opção por um deles pode ser feita em função da velocidade de processamento, disponibilidade de “softwares”, etc. Neste caso específico faço a opção pelas redes neurais com apodização, visto que dentre outras vantagens, não necessitamos otimizar a arquitetura da rede, tarefa esta que consome *algum* tempo (paciência também). Além disso, podemos combinar, facilmente, o modelo obtido com as redes neurais com apodização com reconhecimento de padrões, como veremos a seguir.

7.7 Redes Neurais e o Reconhecimento de Padrões

Neste etapa do nosso trabalho desejávamos construir uma rede neural que a partir dos espectros de reflectância no infravermelho próximo (NIRR) fosse capaz de identificar (reconhecer padrões) entre amostras de folhas de trigo, milho e algodão, aquela que deu origem ao espectro. Para tanto, utilizamos os espectros de dez amostras de folhas de milho (cinco para calibrar e cinco para prever), dez de algodão (cinco para calibrar e cinco para prever) e cinquenta e quatro amostras de folhas de trigo (cinco para calibrar e 49 para prever). As entradas(inputs) da rede neural foram os espectros de trigo, milho e algodão e as saídas foram codificadas da seguinte maneira:

$$\text{trigo} = [1 0 0]$$

$$\text{milho} = [0 1 0]$$

$$\text{algodão} = [0 0 1]$$

A rede neural utilizada para tal tarefa, foi treinada com o método de Marquardt-Levenberg e apresentava a arquitetura representada na Figura 7.27:

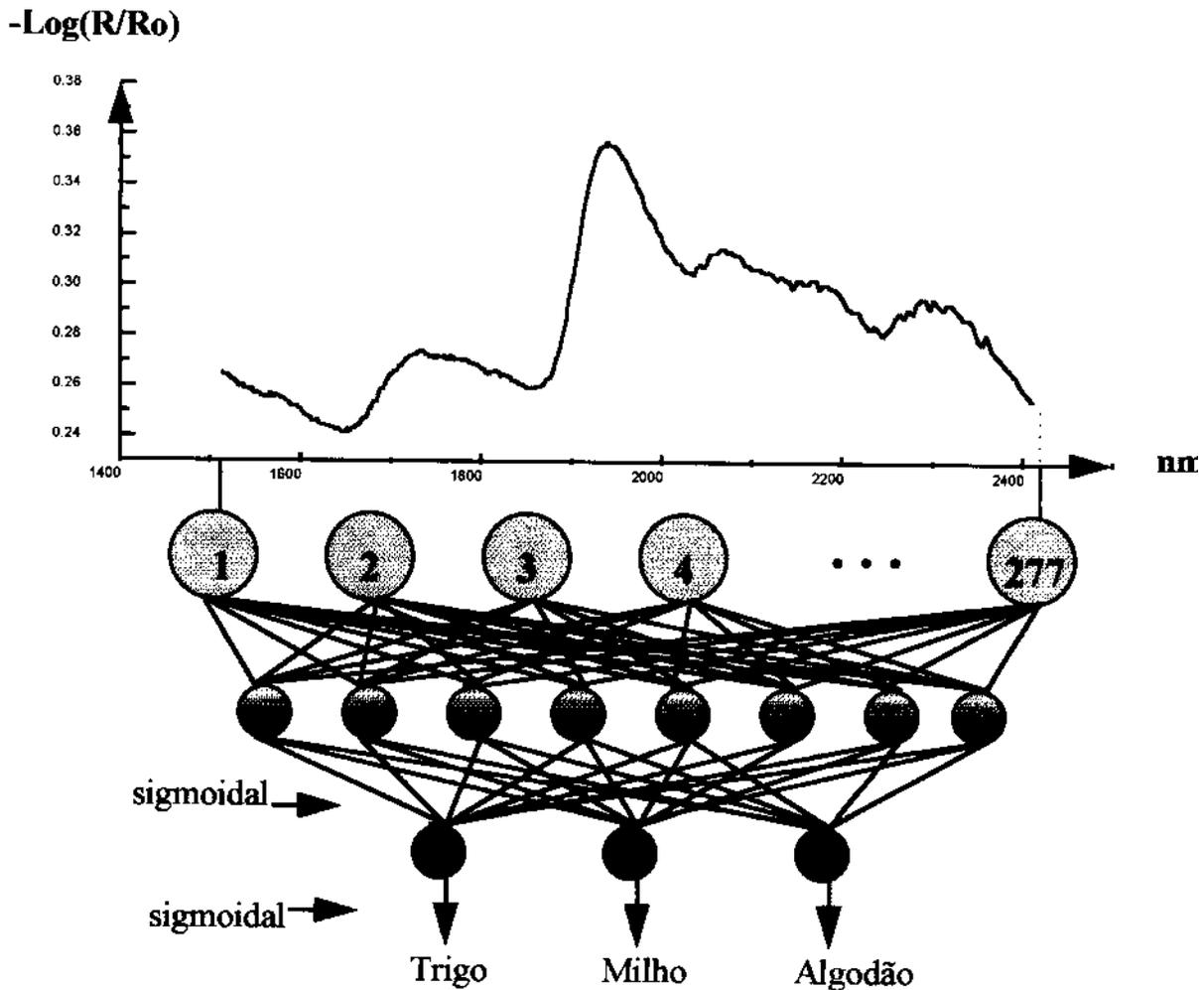


Figura 7.27. Representação da rede neural utilizada na classificação dos espectros. Esta rede neural possui uma camada de entrada com 277 neurônios, uma camada intermediária com oito neurônios e uma camada de saída com três neurônios. A função de transferência utilizada entre a camada de intermediária e a camada de saída foi a função sigmoide. Na camada de saída, utilizou-se a função sigmoide.

Utilizando a rede neural anterior, obtivemos 100% de classificações corretas, isto é, apresentávamos um espectro NIRR à rede neural e “*ela informava*” o tipo de amostra (folhas de trigo, milho ou algodão) que deu origem ao espectro. Enfim, a rede neural foi capaz de separar os espectros em três classes distintas.

7.8 Unindo classificação e modelamento

Uma vez que foi possível classificar os espectros utilizando redes neurais elaboramos um programa em MATLAB[®] em que inicialmente uma rede neural já treinada (a rede do item anterior) classifica os espectros em uma das três classes (trigo, milho ou algodão). Após a classificação segue-se a quantificação do teor de nitrogênio via uma rede neural com pruning previamente treinada com as amostras da calibração. Executamos o trabalho de modelamento (calibração) para o trigo conforme apresentado anteriormente, mas naturalmente o mesmo trabalho pode ser feito para o milho, o algodão ou outro tipo qualquer de cultura.

A vantagem de combinar classificação e modelamento em um único programa computacional, isto é, em um sistema especialista é a sua aplicação em laboratórios de rotina. O operador do espectrofotômetro NIR não necessita conhecer quimiometria, basta que o mesmo obtenha o espectro NIRR das amostras de folhas e o programa se encarregará do resto do trabalho, qual seja: a classificação e a previsão do teor de nitrogênio.

CAPÍTULO 8

CONCLUSÕES

8. Conclusões

Neste trabalho podemos obter várias conclusões, umas ligadas às técnicas de pré-processamento e modelamento e outras ligadas à aplicações práticas. Começemos com aquelas ligadas às técnicas de pré-processamento e modelamento.

A partir dos resultados obtidos podemos concluir que se desejarmos utilizar, com eficiência, a espectroscopia de reflectância difusa no infravermelho próximo (NIRR) para determinar nitrogênio em folhas devemos invariavelmente tratar o espectro obtido de modo a remover ruído e espalhamento de luz; caso contrário os erros de previsão serão elevados. Para remover o espalhamento de luz, devido às diferentes granulometrias das amostras, devemos usar o método MSC e, para remover o ruído, podemos optar entre o filtro de Fourier e o filtro de Wavelets, visto que ambos apresentaram desempenhos equivalentes.

A avaliação dos métodos de modelamento fundamentada simplesmente na comparação dos erros de previsão, pode induzir a conclusões incorretas. Portanto, para avaliar o desempenho de métodos de modelamento, deve-se invariavelmente aplicar testes de significância estatística, como por exemplo, o teste F.

Os métodos de modelamento testados foram classificados em dois grupos. Um grupo, com menores erros de previsão, formado pelas redes neurais com apodização, redes neurais e PLS redes neurais e um segundo grupo, com maiores erros de previsão, formado pelo PLS e PLS polinomial. Internamente a estes grupos não há diferenças significativas, no nível de 95 % de confiança, entre os valores de % SEC e % SEPC.

Conforme descrevemos, os modelos baseados em redes neurais apresentaram desempenhos significativamente superiores aos outros métodos de calibração multivariada. Uma vez que, os três métodos baseados em redes neurais não apresentaram diferenças significativas nos valores de % SEPC, a opção por um deles deve ser feita considerando fatores como o tempo computacional, a facilidade de implementação do método, a disponibilidade de “softwares”, etc. Entretanto podemos considerar que as redes neurais com apodização oferecem um maior potencial de aplicação em calibração multivariada que os demais métodos, pois o seu uso não requer a otimização da arquitetura da rede neural e reduz significativamente a possibilidade de overfitting. Além disso, somente as redes neurais com apodização oferecem a possibilidade de diagnosticar e modelar dados, independentemente das relações de linearidade ou de não-linearidade entre as variáveis dependente e independentes.

Os bons resultados obtidos e a possibilidade de unificação entre modelamento e classificação em um único software tem incentivado a substituição do tradicional método Kjeldahl para análise de nitrogênio em plantas pela metodologia desenvolvida neste trabalho, nos laboratórios de rotina do INSTITUTO AGRÔNOMICO DE CAMPINAS (IAC).

CAPÍTULO 9

Referências bibliográficas

9. Bibliografia

- [1] P. William e K. Norris, *Near-infrared Technology: in agricultural e foods industries*, American Association of Cereal Chemists, Minnesota, 1990.
- [2] D. L. Wetzel, *Anal. Chem.* **55**(12), 1165A (1983).
- [3] M. Twoney, G. Downey e P. B. McNulty, *J. Sci. Food Agric.* **67**, 77 (1984).
- [4] D. H. McQueen, R. Wilson e A. Kinnunen, *Talanta*, **42**, 2007 (1995).
- [5] E. Dreassi, G. Ceramelli, P. Corti, P. L. Perruccio e S. Lonard, *Analyst*, **121**, 219 (1996).
- [6] D. Jouan-Rimbaud, M. S. Khots, D. L. Massart, I. R. Last e K. A. Prebble, *Anal. Chim. Acta*, **315**, 243 (1995).
- [7] R. Tkachuk, *J. Am. Oil Chem. Soc.*, **58**, 819 (1981).
- [8] M. Hana, W. F. McClure e T. B. Whitaker, *J. Near Infrared Spectrosc.*, **3**, 133 (1995).
- [9] D. Wienker, W. Broek, W. Melssen, L. Buydens, R. Feldhoff, T. Kantimm, T. H. Fehre, L. Quick e K. Cammann, *Anal. Chim. Acta*, **317**, 1 (1995).
- [10] M. Martin, A. Garrison, M. Roberts, P. Hall e C. Moore, *Process Control Qual.*, **5**, 187 (1993).

- [11] D. Lambert, B. Descartes, J. R. Llinas, A. Espinosa, S. Osta, M. Sanchez e A. Martens, *Analisis*, **23**(4), M9-M13 (1995).
- [12] D. Lambert, B. Descartes, A. Espinosa, M. Sanchez, S. Osta, J. Gil, A. Martens e M. Valleur, *Analisis*, **23**(4), M20-M25 (1995).
- [13] W. Kaye, *Spectrochim. Acta*, **6**, 257 (1954).
- [14] W. Kaye, *Spectrochim. Acta*, **7**, 181 (1955).
- [15] H. Martens e T. Naes, 'Multivariate Calibration', John Wiley & Sons, Chichester, 1989.
- [16] S. Sekulic, M. B. Seasholtz, Z. Wang, S. E. Lee, B. R. Holt e B. R. Kowalski, *Anal. Chem.* **65**(19), 835 A (1993).
- [17] J. Zupan e J. Gasteiger, *Neural Networks for chemists: an introduction*, VCH, Weinheim, 1992.
- [18] J. Hertz, A. Krogh e R. G. Palmer, "Introduction to the Theory of Neural Computation", Addison Wesley, Menlo Park, 1991.
- [19] L. Bokobza, *J. Near. Infrared spectroscopy*, **6**, 3 (1998).
- [20] O. H. Wheeler, *J. Chem. Education*, **37**, 234 (1960).
- [21] S. E. Krikorian e M. Mahpour, *Spectrochim. Acta*, **29A**, 1233 (1973).
- [22] A. L. Smith, 'Applied infrared Spectroscopy', John Wiley & Sons, New York, 1979.

- [23] B. Osborne e T. Fearn, '*Near infrared Spectroscopy in Food Analysis*', John Wiley & Sons, New York, 1986.
- [24] J. J. Workman, *Appl. spectrosc.Reviews*, **31**(3), 251 (1996).
- [25] W. W. M. Wendlant e H. G. Hecht, *Reflectance Spectroscopy*, John Wiley & Sons, New York, 1966.
- [26] M. J. Adams, 'Chemometrics in Analytical Spectroscopy', The Royal Society of Chemistry, Wolverhampton(UK), 1995.
- [27] V. J. Barclay, R. F. Bonner e I. P. Hamilton, *Anal. Chem.* **69**(1), 78 (1997).
- [28] J. C. Miller e J. N. Miller, '*Statistics for Analytical Chemistry*', Ellis Horwood, Chichester, UK, 1993.
- [29] C. L. Erickson, M. J. Lysaght e J. B. Callis, *Anal. Chem.*, **64**, 1155 (1992).
- [30] S. E. Bialkowski, *Anal. Chem.*, **60**, 355 (1988).
- [31] W. H. Press, S. A. Teukolsky, W, T. Vetterling e B. P. Flannery, "Numerical Recipes in Fortram: *The Art of Scientific Computing*", 2nd ed., Cambridge University Press, New York, 1992.
- [32] A. Savitzky e M. J. E. Golay, *Anal. Chem.*, **36**, 1627 (1964).
- [33] J. Steiner, Y. Tertonia e J. Deltour, *Anal. Chem.*, **44**, (1972).

- [34] R. Bracewell, '*The Fast Fourier Transform e its Application*', McGraw-Hill, New York, 1965.
- [35] H. J. Nussbaumer, '*Fast Fourier Transform e Convolutions algorithms*', Springer-Verlag, New York, 1982.
- [36] P. Bloomfield, '*Fourier Analysis of time series*', John Wiley, New York, 1976.
- [37] C. Canuto, M. Y. Hussaini, A. Quarteroni e T. A. Zang, '*Spectral methods in fluid dynamics*', Springer-Verlag, New York, 1988
- [38] J. P. Gollub e G. L. Baker, "*Chaotic Dynamics*", Cambridge University Press, 1990.
- [39] Eugene Butkov, "*Física Matemática*", Guanabara Koogan, Rio de Janeiro, 1982.
- [40] G. B. Arfken e H. J. Weber, "*Mathematical Methods for Physicists*", 4th Edition, Academic Press, New York, 1995.
- [41] H. Eisberg e R. Resnick, "*Quantun Physics*", John Wiley & Sons, New York, 1986.
- [42] D. Halliday, R. Resnick e J. Walker, "*Fundamentals of Physics*", 5th Edition, John Wiley & Sons, New York, 1990.
- [43] Y. Meyer e R. D. Ryan, '*Wavelets: Algorithms e Applications*', SIAM, Philadelphia, 1993.

- [44] U. Depczynski, K. Jetter, K. Molt, A. Niemoller, *Chemom. Intell. Lab.Sys.* **39**, 19 (1997).
- [45] B. Walczak, B. van der Bogaert e D. L. Massart, *Anal. Chem.* **68**, 1742 (1996).
- [46] D. J. Rimbaud, B. Walczak, R. J. Poppi e D. L. Massart. *Anal. Chem.*, **69**, 4317 (1997)
- [47] V. J. Barclay, R. F. Bonner e I. P. Hamilton, *Anal. Chem.*, **69**(1), 78 (1997).
- [48] M. Wolkenstein, H. Hutter, S. G. Nikolov, I. Schmitz e M. Grasserbauer, *J. Trace e Micropobe Techniques*, **15**(1), 33 (1997).
- [49] M. Wolkenstein, H. Hutter, S. G. Nikolov, I. Schmitz e M. Grasserbauer, *Fresenius J. Anal. Chem.*, **357**, 783 (1997).
- [50] R. Castleman, “*Digital Image Processing*”, Prentice Hall, New Jersey, 1996.
- [51] G. Strang, *American Scientist*, **82**, 250 (1994).
- [52] M. Bos e E. Hoogendam, *Anal. Chim. Acta*, **267**, 73 (1992).
- [53] I. Daubechies, “*Ten Lectures on Wavelets*”, SIAM, Philadelphia, 1992.
- [54] P. P. Vaidyanatham, “*Multirate Systems e Filter Banks*”, Englewood Cliffs, New Jersey, 1993.

- [55] B. Walczak e D. L. Massart, *Trends in anal. chem.*, **16**(8), 451 (1997).
- [56] T. Isaksson e B. R. Kowalski, *Appl. Spectrosc.*, **47** (6), 702 (1993).
- [57] P. Geladi, D. MacDougall e H. Martens, *Appl. Spectrosc.*, **39**, 491 (1985).
- [58] B. M. Wise e N. B. Gallagher, “ *PLS Toolbox, for use with MATLAB™* “, Eigenvector Technologies, P.O. Box 483, 196 Hyacinth Avenue, Manson, WA 98831. (E-mail: 73633.2451@compuserve.com) , (FAX (509) 964-7-3973).
- [59] R. Kellner, J. M. Mermet, M. Otto e H. M. Widmer (Eds.), “Analytical Chemistry”, Wiley-VCH, New York, 1998.
- [60] J. Zupan e J. Gasteiger, *Anal. Chim. Acta*, **248**, 1 (1991).
- [61]] W. MacCulloch e W. Pitts, *Bull. Math. Biophys.*, **5**, 115 (1943).
- [62] Y. A. Liu e D. R. Baughman, ‘*Neural networks in bioprocessing e chemical engineering*’, Academic Press, San Diego (1995).
- [63] B. Katz, ‘Nerve Muscle e Synapse’, MacGraw Hill, New York (1996).
- [64] H. M. Cartwright, “Applications of Artificial Intelligence in Chemistry”, Oxford University Press, New York, 1995.
- [65] G. Katerman e J. R. Smits, *Anal. Chim. Acta*, **277**, 179 (1993).

- [66] P. J. Gemperline, *Chemom. e Intell. Lab. Sys.***39**, 29 (1997).
- [67] P. De B. Harrington, *Anal. Chem.*, **65**(15), 2167 (1993).
- [68] J. R. Long, V. G. Gregoriou e P. J. Gemperline, *Anal. Chem.*, **62**, 1791 (1990).
- [69] B. Wythoff, *Chemom.Intell. Lab. Sys.* **18**, 115 (1993).
- [70] J. R. S. Jang, C. T. Sun e E. Mizutani, 'Neuro Fuzzy e soft computing', Prentice Hall, Upper Saddle River (1997).
- [71] C. Cunha, "Métodos Numéricos para Engenharias e Ciências Aplicadas", Editora da UNICAMP, 1993.
- [72] S. C. Chapra e R. P. Canale, 'Numerical methods for engineers', MacGraw-Hill, New York (1990).
- [73] F. S. Acton, "Numerical Methods that Works", Harper e Row, New York, 1970.
- [74] E. V. Thomas e D. M. Haale, *Anal. Chem.*, **62**, 1091 (1990):
- [75] R. G. Brereton, "*Chemometrics*", Ellis Horwood, Chichester, UK, 1990.
- [76] A. F. Carley e P. H. Morgan, "*Computational Methods in the Chemical Sciences*", Ellis Horwood, Chichester, UK, 1989.
- [77] J.Heitz, A. Kragh, R. G. Palmer, 'Introduction to the theory of neural computation', Addison-Wesley, New York (1991).

- [78] B. Hassibi e D. G. Stork, 'Advances in neural information processing systems', Morgan Kaufmann, 1993.
- [79] Y. Le Cun, B. Boser, J. S. Denker and S. A. Solla, '*Optimal brain damage: Advances in Neural information processing systems*', vol.2, Morgan Kaufman, San Mateo, 1990, p.598
- [80] R. J. Poppi e D. L. Massart, *Anal. Chim. Acta.*, aceito para publicação em 1998.
- [81] C. Goutte, '*On use of a pruning prior neural networks*', in neural network for signal processing IV, Proceedings of IEEE International workshop, Piscataway (1996).
- [82] S. E. Fahlman e C. Libiere, in: D. S. Touretzky (Ed.), *Advances in Neural Information Processing Systems*, Vol. 2, Morgan Kaufmann, San Mateo, 1990, p.524.
- [83] J. Larsen e L. K. Hansen, "*Generalization desempenho of regularized neural network models*", Proc. of the IEEE workshop on neural networks for signal Proc. IV, Piscataway, New Jersey, 1994, p.164.
- [84] M. Norgaard, "Neural Network Based System Identification TOOLBOX for use with MATLABTM, version 1.1", Technical University of Denmark, 1997.
- [85] K. R. Beebe, B. R. Kowalski, *Anal. Chem.* **59**(17), 1007A (1987).

- [86] S. Wold, N. Kettaneh-Wold e B. Skagerberg, *Chemom. e Intell. Lab. Syst.*, **14**, 1989.
- [87] E. V. Thomas, *Anal. Chem.* **66**(15), 795 A (1994).
- [88] S. J. Quin e T. J. McAvoy, *Computers Chem. Engng.* **16**(4), 379 (1992).
- [89] O. C. Bataglia, J. P. F. Teixeira, A.M.C. Furlani e J. R. Gallo. 'Métodos de Análise Química de Plantas', Boletim Técnico Nº 78 do Instituto Agrônomo de Campinas, Campinas 1983.
- [90] D. C. Tran, *Anal. Chem.*, **64**(20), 917A (1992).
- [91] J. Xu, R. Stroud, "Acousto-Optic Devices: Principles, Design e Applications", Wiley-Interscience, New York, 1992.
- [92] R. Guchardi, P. A. Costa Filho, R. J. Poppi e C. Pasquini, *J. of Near Infrared Spectrosc.*, **6**, 333 (1998).
- [93] R. Guchardi, 'Construção e Avaliação de um Espectrofotômetro Infravermelho Próximo com Monocromador Baseado em Filtro Óptico Acústico Sintonizável (AOTF), Dissertação de Mestrado, Unicamp, Campinas 1997.
- [94] I. Kurtz, R. Dwelle e P. Katzka, *Rev. Sci. Instrum.*, **58**(11), 1996 (1987).
- [95] C. I. Chang, *Appl. Phys. Lett.*, **25**(7), 370 (1974).

- [96] H. Farrer, C. G. Becker, E. F. Chaves, H. F. Matos, M. A. Santos, M. L. Maia, 'Fortran Estruturado', Guanabara-Koogan, Rio de Janeiro, 1992.
- [97] Matlab, The Mathworks Inc., 24 primer Park Way, Natick, Mass., USA (1995)
- [98] H. Demuth e M. Beale, 'Neural Network Toolbox for use with Matlab', Manson, USA (1996).
- [99] B. Kratochvil, J. K. Taylor, *Anal. Chem.*, **53**(8), 925A (1981).
- [100] D. A. Skoog, F. J. Holler e T. A. Nieman, 'Principles of Instrumental Analysis', 5th ed., Saunders College Publishing, Philadelphia, 1998.
- [101] J. C. Davies, 'Statistics Data Analysis in Geology', John Wiley & Sons, New York, 1988.
- [102] P. J. Gemperline, J. R. Long e V. G. Gregoriou, *Anal. Chem.* **63**, 2313 (1991).
- [103] P. E. G. Box, G. W. Hunter e J. S. Hunter, "Statistics for Experimenters", John Wiley & Sons, New York, 1985.