

UNIVERSIDADE ESTADUAL DE CAMPINAS



INSTITUTO DE QUÍMICA - IQ

Estudo Comparativo entre Técnicas de Inteligência Artificial e Modelos Lineares em Determinações Quantitativas no Infravermelho Próximo.

Tese de Doutorado



1150054047



IQ

T/UNICAMP C823e

PAULO AUGUSTO DA COSTA FILHO

Orientador: Prof. Dr. RONEI J. POPPI

Campinas – São Paulo

Janeiro / 2003

UNIDADE	IQ
Nº CHAMADA	UNICAMP
	C823e
V	EX
TOMBO BCI	54047
PROC.	124103
C <input type="checkbox"/>	D <input checked="" type="checkbox"/>
PREÇO	R\$ 11,00
DATA	21/05/03
Nº CPD	

Biblid. 294032

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO INSTITUTO DE QUÍMICA
UNICAMP

CM00183367-5

C823e

Costa Filho, Paulo Augusto da
Estudo comparativo entre técnicas de
inteligência artificial e modelos lineares em
determinações quantitativas no infravermelho
próximo / Paulo Augusto da Costa Filho. --
Campinas, SP: [s.n.], 2003.

Orientador: Ronei Jesus Poppi.

Tese (doutorado) – Universidade
Estadual de Campinas, Instituto de Química.

1. Quimiometria. 2. Inteligência artificial.
3. Infravermelho próximo. I. Poppi, Ronei Jesus.
II. Universidade Estadual de Campinas. III. Título.

**Quando surgem os ventos das mudanças,
alguns constroem abrigos, outros moinhos.**

**Aos meus pais,
Paulo e Maria,
minhas irmãs
Cátia e Márcia**

Agradecimentos

Aos amigos do LAQQA, ao meu orientador Prof. Dr. Ronei J. Poppi, ao Prof. Dr. Celso U. Davanzo e aos funcionários do Instituto de Química que contribuíram para a execução deste trabalho.

Ao Laboratório de análises clínicas Canadá S.A, por ter cedido as amostras de soro humano utilizadas neste trabalho e à empresa Femto por ter cedido o espectrofotômetro FEMTO PLS plus. À Rhodia e à COOPERSUCAR por terem cedido parte dos conjuntos de dados utilizados neste trabalho.

Agradeço à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo financiamento do projeto (processo nº 98/5857-7) e pela bolsa concedida.

Resumo

Estudo Comparativo entre Técnicas de Inteligência Artificial e Modelos Lineares em Determinações Quantitativas no Infravermelho Próximo.

Autor: Paulo Augusto da Costa Filho

Orientador: Ronei Jesus Poppi

Palavras-Chave: Quimiometria, Inteligência artificial, infravermelho próximo.

Neste trabalho foi realizado um estudo comparativo entre o desempenho de métodos baseados em inteligência artificial e dois métodos de calibração multivariada linear (regressão linear múltipla (RLM) e mínimos quadrados parciais (PLS)), para determinações quantitativas na região do infravermelho próximo em três diferentes sistemas analíticos.

As técnicas de inteligência artificial pertencem à classe de mapeamento não linear que utiliza algoritmos inteligentes para modelagem de forma adaptativa de um determinado sistema. As técnicas de inteligência artificial estudada foram: algoritmo genético, rede neural, assim como sistemas híbridos como Neuro-Fuzzy e neurogenético.

Este trabalho reporta os resultados alcançados na modelagem de três sistemas bem diversos: 1-determinação do grau Brix e Pol em soluções de caldo de cana de açúcar; 2- determinação de um mineral contido num polímero; 3- determinação de glicose, colesterol total e triglicerídeos em amostras de plasma humano.

O estudo das amostras de cana de açúcar consistiu em estabelecer uma correlação entre os espectros na região no infravermelho próximo e o valor do grau Brix e Pol. O grau Brix e Pol são parâmetros utilizados para estimar a qualidade do açúcar. O melhor modelo de calibração multivariada para este sistema apresentou um erro padrão de previsão para o grau Brix e Pol de 0,35 e 0,55, respectivamente. Para o estudo quantitativo do mineral presente em um polímero foi empregada a espectroscopia de refletância difusa no infravermelho próximo. O objetivo desta análise foi determinar a percentagem do mineral contido no polímero, o qual confere propriedades especiais a este. Para esta análise, foi obtido um erro padrão de previsão de 0,009%. Para concluir este trabalho, foi realizada uma análise quantitativa de glicose, colesterol total e triglicerídeos em amostras de plasma humano utilizando a espectroscopia no infravermelho próximo. Para as amostras de glicose, colesterol total e triglicerídeos foram obtidos erros padrão de previsão de 6,54, 13,39 e 14,73 mg dL⁻¹, respectivamente.

Abstract

Comparative Study among Artificial Intelligence Techniques and Linear Models for Quantitative Analyses in the Near Infrared region

Author: Paulo Augusto da Costa Filho

Adviser: Ronei Jesus Poppi

Key words: Chemometric, artificial Intelligence, near infrared.

In this work it was done a comparative study among the performances of different artificial intelligence techniques and two multivariate linear calibration methods (Multiple Linear Regression and Partial Least Squares), in three different chemical systems for quantitative determination in the near infrared region.

The artificial intelligence techniques are a particular case of a non-linear models, which use intelligent algorithms to describe the behavior of one system in an iterative way. The artificial intelligence techniques studied were: genetic algorithm, neural network and Fuzzy logic, as well as hibrid systems as neurogenetic and Neuro-Fuzzy.

This thesis shows the results obtained from three different chemical systems: 1- solutions of sugar cane juice; 2- a mix between powders of a polymer and one special mineral; 3- determination of glucose, cholesterol and triglycerides in human plasma.

It was established a correlation between near infrared spectrum and Pol and Brix degrees, both parameters are used to evaluate the sugar quality. For this analysis, the best calibration model gave a standard error of prediction for Brix and Pol of 0.35 and 0.55, respectively. For the second system, it was used the diffuse reflectance in the near infrared range to analyse the mineral content in a polymer. The aim of this analysis was to determine the mineral concentration in this solid mix because it provides special chemical and physical properties to the polymer. The near infrared analyses provided very good results: it was obtained a standard error of prediction of 0.009%. To conclude this work, it was done quantitative analysis of glucose, cholesterol and triglycerides in human plasma. For glucose, total cholesterol and triglycerides it was obtained standard errors of prediction of 6.54, 13.39 and 14.73 mg dL⁻¹, respectively.

I. Formação Acadêmica

Doutorado em Ciências – Química Analítica

Estudo Comparativo entre Técnicas de Inteligência Artificial e Modelos Lineares em Determinações Quantitativas no Infravermelho Próximo

Universidade Estadual de Campinas.

Cidade: Campinas – SP.

Período: Março de 1998 – Janeiro de 2003.

Órgão financiador: FAPESP (proc. 98/5857-7).

Mestrado em Química

Algoritmo Genético na Seleção de Variáveis em Calibração Multivariada de Dados Espectroscópicos

Universidade Estadual de Campinas.

Cidade: Campinas – SP.

Período: Março de 1996 – Março de 1998.

Órgão financiador: CAPES.

Licenciatura em Química

Universidade Estadual de Campinas.

Cidade: Campinas – SP.

Período: Março de 1996 – Dezembro de 2000.

Bacharelado em Química

Universidade Estadual de Campinas.

Cidade: Campinas – SP.

Período: Março de 1992 – Dezembro de 1995.

II. Artigos Publicados

1. P. A. Costa Filho e R. J. Poppi, "Aplicação de Algoritmos Genéticos na seleção de variáveis em espectroscopia no infravermelho médio. Determinação simultânea de glicose, maltose e frutose", QUÍMICA NOVA 25, 46-52 (2002).
2. E. R. Pereira Filho, C. Mello, P. A. Costa Filho, M. A.Z. Arruda e R. J. Poppi, "Neuro-genetic approach for optimisation of the spectrophotometric catalytic determination of cobalt", ANALYTICA CHIMICA ACTA 433, 111-117 (2001).
3. P. A. Costa Filho e R. J. Poppi, "Determination of triglycerides in human plasma using near-infrared spectroscopy and multivariate calibration methods", ANALYTICA CHIMICA ACTA 446, 39-47 (2001).
4. P. A. Costa Filho e R. J. Poppi, "Use of near infrared spectroscopy for rapid estimation of sugar cane juice quality components", NEAR INFRARED SPECTROSCOPY: PROCEEDINGS OF THE 9TH INTERNATIONAL CONFERENCE ON NEAR INFRARED SPECTROSCOPY, 897-902, (2000).
5. P. A. Costa Filho e R. J. Poppi, "Algoritmo Genético em Química", QUÍMICA NOVA 22, 405-411 (1999).
6. R. Guchardi, P. A. Costa Filho, R. J. Poppi e C. Pasquini, "Determination of Ethanol and Methyl Tert-Butyl Ether (MTBE) in gasoline by NIR-AOTF based spectroscopy and multiple linear regression with variables selected by genetic algorithm", NEAR INFRARED SPECTROSCOPY 6, 333-339 (1998).

III. Trabalhos Apresentados em Congressos Internacionais

1. P. A. Costa Filho e R. J. Poppi, "Determination of Triglycerides in Human Plasma using Near Infrared Spectroscopy and Multivariate Calibration", 7th International Conference on Chemometrics in Analytical Chemistry - Antuérpia – Bélgica, Book of Abstracts, 64 (2000).
2. P. A. Costa Filho e R. J. Poppi, "Use of near infrared spectroscopy for rapid estimation of sugar cane juice quality components", 9th International Conference on Near-Infrared Spectroscopy - Verona – Itália, Abstract book, P4-27 (1999).
3. P. A. da Costa Filho, E. O. Cerqueira e R. J. Poppi, "Genetic algorithm for simultaneous selection of variables in generalized rank annihilation method applied to flow analysis", 7th International Conference on Flow Analysis - Piracicaba – Brasil, Book of Abstracts, 185 (1997)
4. C. Pasquini, R. J. Poppi, P. A. da Costa Filho e R. Guchardi, "Determination of ethanol and MTBE in gasoline by NIR AOTF based spectroscopy and multiple linear regression with variables selected by genetic algorithm", 8th International Conference on Near-Infrared Spectroscopy - Essen – Alemanha, Poster Abstracts, P2PETRO (1997)

IV. Trabalhos Apresentados em Congressos Nacionais

1. P. A. da Costa filho e R. J. Poppi, "Avaliação da possibilidade de determinação de glicose a níveis fisiológicos utilizando espectroscopia no infravermelho próximo", 23 a. Reunião Anual da SBQ - Poços de Caldas – MG, Livro de Resumos, QA-111 (2000).
2. P. A. da Costa Filho, C. Mello e R. J. Poppi, "Desenvolvimento e aplicação de um programa híbrido de inteligência artificial em espectroscopia de reflectância no infravermelho próximo", 10o. Encontro Nacional de Química Analítica - Santa Maria – RS, Livro de Resumos, QM-25 (1999).
3. C. Mello, E. O. de Cerqueira, R. J. Poppi, L. T. Kubota e P. A. da Costa Filho, "Aplicações do filtro de transformada de Fourier (FFT) na minimização de ruídos em sinais analíticos", 10o. Encontro Nacional de Química Analítica - Santa Maria – RS, Livro de Resumos, QM-26 (1999).
4. E. R. Pereira Filho, C. Mello, P. A. da Costa Filho, R. J. Poppi e M. A. Z. Arruda, "Determinação espectrofotométrica catalítica de Cobalto: emprego de redes neurais e algoritmos genéticos para otimização do método", 21 a. Reunião Anual da SBQ - Poços de Caldas – MG, Livro de Resumos, QA-132 (1998).
5. P. A. da Costa Filho e R. J. Poppi, "Algoritmo genético para seleção de comprimentos de onda em regressão linear múltipla na determinação simultânea de açúcares em solução com ATR/FT-IR", 20 a. Reunião Anual da SBQ - Poços de Caldas – MG, Livro de Resumos, QA-004 (1997).
6. C. Mello, R. J. Poppi, P. A. da Costa Filho, C. Reis e J. C. de Andrade, "Uso de técnicas de inteligência artificial para modelamento e otimização na determinação cinética catalítica de cromo(VI), 9o. Encontro Nacional de Química Analítica - São Carlos – SP (1997).

Abreviações

PCA	Análise de Componentes Principais
PLS	Mínimos Quadrados Parciais
MSC	Correção de Espalhamentos Múltiplos
SNV	Transformação Padrão Normal de Variação
DM	Distância de Mahalanobis
MLR	Regressão Linear Múltipla
DE	Distância Euclidiana
AG	Algoritmo Genético
NIR	Infravermelho Próximo
SEC	Erro Padrão de Calibração
SEP	Erro Padrão de Previsão
SVD	Decomposição em valores singulares

Capítulo 1 Métodos de Calibração Multivariada

1. Introdução a Calibração Multivariada.....	5
1.1 Arranjo de dados e notação Algébrica.....	5
1.2 Pré-processamento dos Dados.....	6
1.3 Distância Euclidiana e Mahalanobis.....	7
1.4 Relação entre Matrizes de dados.....	8
1.5 Regressão Linear Múltipla (RLM).....	9
1.6 Mínimos Quadrados Parciais (PLS).....	10
1.7 Detecção de Amostras Anômalas (outlier).....	11
1.8 Correção de Espalhamento Multiplicativo (MSC).....	12
1.9 Transformação padrão Normal de Variação (SNV).....	13

Capítulo 2 Algoritmo Genético

2. Introdução ao Algoritmo Genético.....	15
2.1 Teoria de Darwin.....	15
2.2 Algoritmo Genético.....	16
2.2.1 Algoritmo Genético Básico.....	17
2.3 Algoritmo Genético em Otimização.....	23
2.4 Operações e Técnicas Avançadas.....	24
2.5 Teoria dos Esquemas.....	26
2.6 Aplicação do Algoritmo Genético em Química.....	28
2.7 Algoritmo Genético na Seleção de Variáveis.....	29

Capítulo 3 Rede neural e Neurogenética

3. Introdução.....	31
3.1 Redes Neurais.....	31
3.2 Elementos Básicos de uma Rede neural.....	33
3.2.1 Tipos mais comuns de funções de transferência	34
3.3 Estrutura de uma Rede Neural	38
3.4 Princípio de processamento das Redes Neurais.....	40
3.5 Pré-processamento dos dados de Entrada da Rede Neural.....	42
3.5.1 Escalamento dos Dados de treinamento da Rede Neural.....	43
3.6 Treinamento de uma Rede Neural.....	46
3.6.1 Atualização dos valores dos pesos sinápticos.....	49
3.6.2 Cálculo do vetor gradiente do erro da camada Interna.....	50
3.6.2.1 Cálculo da atualização dos pesos	52
3.7 Critérios de Convergência.....	52
3.8 Rede Neurogenética.....	52
3.8.1 Implementação do Algoritmo Genético.....	53
3.8.2. Codificação de uma Rede Neural....	54
3.8.3 Etapa de treinamento e avaliação dos Cromossomos.....	55

Capítulo 4 Lógica Fuzzy e Neuro-Fuzzy

4	Introdução a Lógica Fuzzy	57
4.1	Teoria Fuzzy	57
4.2	Introdução aos Conceitos Matemáticos de Conjuntos Fuzzy	60
4.3	Funções de Pertinência	61
4.3.1	Operações entre funções de Pertinência	64
4.3.2	Classes de operadores dos Conjuntos Fuzzy	66
4.3.2.1	Operadores do Tipo Norma-T	67
4.3.2.2	Operadores do Tipo Norma-S ou Co-Norma-T	67
4.3.2.3	Complemento Fuzzy	68
4.3.2.4	Outros Operadores	69
4.3.3	O princípio de Extensão	69
4.4	Raciocínio Fuzzy	72
4.4.1	Mecanismos de Raciocínio Fuzzy	74
4.5	Métodos de Decodificação	76
4.6	Rede Neuro-Fuzzy	77
4.6.1	Sistema de Inferência Adaptativo Neuro-Fuzzy	78
4.7	Interpretação das Regras Fuzzy de Inferência	81

Capítulo 5 Introdução a Espectroscopia NIR

5. Introdução à Teoria do Infravermelho Próximo.....	85
5.1 Espectroscopia na Região do Infravermelho Próximo.....	85
5.2 Principais Fatores que influenciam nos trabalhos Quantitativos na região do infravermelho Próximo.....	88
5.3 Princípios Físicos Químicos envolvidos na espectroscopia do Infravermelho Próximo.....	89
5.3.1 Modelo Mecânico da Vibração Molecular.....	90
5.3.2 Princípio de Franck-Condon.....	92

Capítulo 6 Modelagem de Dados Simulados

6. Simulação Teórico.....	95
6.1 Modelagem de um sistema não Linear.....	95
6.2 Desenvolvimento dos Modelos de Inteligência Artificial.....	96
6.2.1 Rede Neural.....	96
6.2.2 Rede Neurogenética.....	97
6.2.3 Rede Neuro-Fuzzy.....	97
6.3 Resultados obtidos com os modelos desenvolvidos para os dados simulados.....	98
6.3.1 Modelagem aplicando a técnica de Redes Neurais.....	98
6.3.2 Modelagem aplicando a técnica de Rede Neuro-Fuzzy.....	100
6.3.3 Modelagem aplicando a técnica de Rede Neurogenética.....	102
6.4 Considerações Finais.....	104

Capítulo 7 Análise em Caldo de Cana de Açúcar

7. Introdução.....	105
7.1 Procedimento Experimental.....	106
7.1.1 Procedimento Experimental para Determinações do Grau Brix	107
7.1.2 Procedimento Experimental para Determinações do Grau Pol.....	107
7.1.2.1 Cuidados durante as Medidas Instrumentais.....	108
7.1.2.2 Parâmetros que podem afetar as medidas Polarimétricas.....	108
7.1.2.3 Definição da Escala Sacarimétrica.....	108
7.2 Desenvolvimento dos Modelos de Calibração Multivariada.....	109
7.2.1 Método dos Mínimos Quadrados Parciais (PLS).....	110
7.2.2 Algoritmo Genético com Regressão Linear Múltipla (AG/RLM).....	110
7.2.3 Rede Neural.....	111
7.2.4 Rede Neurogenética.....	111
7.2.5 Rede Neuro-Fuzzy.....	111
7.3 Resultados e Discussão.....	112
7.4 Determinação do grau Brix usando a técnica de infravermelho Próximo.....	112
7.5 Modelos de Calibração Multivariada para a Determinação do grau Brix.....	115
7.5.1 Modelo dos Mínimos Quadrados Parciais (PLS).....	115
7.5.2 Seleção de variáveis aplicando Algoritmo Genético (AG/RLM).....	116
7.5.3 Rede Neural.....	119
7.5.4 Rede Neurogenética.....	120
7.5.5 Rede Neuro-Fuzzy.....	121
7.6 Considerações Finais.....	122
7.7 Determinação do grau Pol usando a técnica de infravermelho Próximo.....	124
7.8 Modelos de Calibração Multivariada para a Determinação do Pol.....	124
7.8.1 Modelo dos Mínimos Quadrados Parciais (PLS).....	124

Capítulo 7 Análise em Caldo de Cana de Açúcar

7.8.2 Seleção de variáveis aplicando o Algoritmo Genético (AG/RLM).....	125
7.8.3 Rede Neural.....	128
7.8.4 Rede Neurogenética.....	130
7.8.5 Rede Neuro-Fuzzy.....	131
7.9 Considerações Finais.....	132

Capítulo 8 Reflectância Difusa em Sólidos

8. Introdução.....	133
8.1 Procedimento Experimental.....	134
8.2 Desenvolvimento dos Métodos de Calibração Multivariada.....	134
8.2.1 Métodos dos Mínimos Quadrados Parciais (PLS).....	134
8.2.2 Regressão Linear Múltipla com seleção de variáveis (AG/RLM).....	134
8.2.3 Rede Neural.....	135
8.2.4 Rede Neurogenética.....	135
8.2.5 Rede Neuro-Fuzzy.....	135
8.2.6 Pré-processamento dos Espectros.....	135
8.3 Resultados e Discussão.....	136
8.4 Modelos de Calibração Multivariada para determinar o Mineral.....	142
8.4.1 Modelos dos Mínimos Quadrados Parciais (PLS).....	142
8.4.2 Seleção de variáveis aplicando Algoritmo Genético (AG/RLM).....	142
8.4.3 Rede Neural.....	145
8.4.4 Rede Neurogenética.....	146
8.4.5 Rede Neuro-Fuzzy.....	147
8.5 Considerações Finais.....	149

Capítulo 9 Análise de Plasma Humano

9. Introdução.....	151
9.1 Procedimento Experimental.....	152
9.2 Desenvolvimento dos Métodos de Calibração Multivariada.....	153
9.2.1 Pré-processamento.....	153
9.2.2 Seleção das Amostras do conjunto de calibração e teste.....	153
9.2.2.1 Modelos de Calibração.....	153
9.2.2.2 Métodos dos Mínimos Quadrados Parciais (PLS).....	154
9.2.2.3 Regressão Linear Múltipla com Seleção de Variáveis (AG/RLM)	154
9.2.2.4 Rede Neural.....	154
9.2.2.5 Rede Neurogenética.....	155
9.2.2.6 Rede Neuro-Fuzzy.....	155

Triglicerídeos

9.3 Determinação de triglicerídeos em Plasma humano.....	156
9.4 Resultados e Discussão.....	157
9.4.1 Modelos dos Mínimos Quadrados Parciais (PLS).....	161
9.4.2 Seleção de variáveis aplicando Algoritmo Genético (AG/RLM).....	162
9.4.3 Rede Neural.....	164
9.4.3 Rede Neurogenética.....	166
9.4.5 Rede Neuro-Fuzzy.....	166
9.5 Considerações Finais.....	167

Capítulo 9 Análise de Plasma Humano

Colesterol

9.6 Determinação de Colesterol em Plasma humano.....	169
9.7 Resultados e Discussão.....	170
9.7.1 Modelos dos Mínimos Quadrados Parciais (PLS).....	171
9.7.2 Seleção de variáveis aplicando Algoritmo Genético (AG/RLM).....	172
9.7.3 Rede Neural.....	173
9.7.4 Rede Neurogenética.....	175
9.7.5 Rede Neuro-Fuzzy.....	175
9.8 Considerações Finais.....	177

Glicose

9.9 Determinação de Glicose em Plasma Humano.....	178
9.10 Resultados e Discussão.....	178
9.10.1 Modelos dos Mínimos Quadrados Parciais (PLS).....	179
9.10.2 Seleção de variáveis aplicando Algoritmo Genético (AG/RLM).....	181
9.10.3 Rede Neural.....	182
9.10.4 Rede Neurogenética.....	184
9.10.5 Rede Neuro-Fuzzy.....	184
9.11 Considerações Finais.....	185

Conclusões.....	187
Perspectivas de Trabalhos Futuros.....	189
Referências Bibliográficas.....	191

A partir de meados da década de 1980, o desenvolvimento da microeletrônica e a popularização dos computadores catalisaram o processo de mudanças e inovações em vários segmentos da sociedade.

Na química, o reflexo dos avanços tecnológicos incidiu pesadamente sobre a área das análises instrumentais, possibilitando a aquisição de maneira mais simples e rápida de um grande número de dados de uma mesma amostra e, conseqüentemente, o tratamento dos dados obtidos passou a exigir cálculos mais complexos. Tal fato, aliado ao crescente número de análises, à diversidade das amostras e à necessidade de resultados rápidos, abriram um leque de oportunidades para o desenvolvimento de uma nova área denominada quimiometria.

A quimiometria consiste essencialmente na aplicação de técnicas matemáticas e estatísticas, assim como de lógica matemática, em problemas químicos⁽¹⁾. O nome quimiometria originou-se da fusão entre a matemática e a química de maneira análoga a psicometria, biometria, econometria, etc.

As análises envolvendo espectroscopia foram as mais beneficiadas com a introdução da quimiometria na química. Nas últimas décadas, os estudos envolvendo dados espectroscópicos no infravermelho próximo em análises quantitativas e qualitativas apresentaram um significativo avanço⁽¹⁾. Isso se deve principalmente ao desenvolvimento e consolidação da quimiometria⁽²⁾, a qual tem contribuído com a aplicação e divulgação de uma série de ferramentas matemáticas que até então haviam sido largamente empregadas principalmente em áreas como: engenharia, estatística, matemática, psicologia, geologia, etc. Em engenharia ferramentas como: filtros digitais, análises das componentes principais, transformada de *wavelets*, algoritmos genéticos, redes neurais, lógica Fuzzy, etc, possuem várias publicações pioneiras em aplicações práticas sendo, portanto, a engenharia considerada uma das áreas precursoras da divulgação e popularização destas ferramentas matemáticas⁽³⁾.

Em química, o número de trabalhos publicados utilizando técnicas de inteligência artificial vem crescendo significativamente; porque estas ferramentas permitem a modelagem de dados com comportamento não linear de elevada complexidade⁽⁴⁾. Esses dados até então eram modelados pelas técnicas tradicionais como: mínimos quadrados parciais (PLS), regressão linear múltipla (RLM), regressão das componentes principais (PCR), etc⁽⁵⁾.

Na tentativa de tornar mais eficiente a solução de problemas extremamente complexos, a inteligência artificial tem evoluído para o que é chamado de *Soft computing*^(6,7). Este consiste na

associação de uma série de técnicas de inteligência artificial, que operam em sinergia para suprir ou minimizar deficiências individuais. Entre as ferramentas mais importantes destacam-se: redes neurais, teoria de conjuntos Fuzzy ou teoria dos conjuntos difusos, métodos de otimização não derivativos como os Algoritmos genéticos e a inteligência artificial convencional^(6,7). Cada uma destas técnicas destaca-se por apresentar um ou mais atributos que se assemelham a alguma característica associada à capacidade do cérebro humano de tomar decisões inteligentes.

Se tentarmos atribuir características individuais para cada uma das técnicas que integram a área de inteligência artificial, poderíamos dizer que: - As redes neurais apresentam a característica de aprendizagem e adaptação, a teoria de conjuntos Fuzzy, a capacidade de trabalhar com incerteza e imprecisão, os algoritmos genéticos, de realizar pesquisas randômicas e sistemáticas e, por último, mas não menos importante, a inteligência artificial convencional, que realiza manipulações simbólicas.

Tendo em mente o amplo leque de perspectivas que a inteligência artificial pode propiciar à área de química analítica, principalmente no desenvolvimento de modelos de calibração multivariada não lineares, nosso grupo de pesquisa (Laboratório de Quimiometria em Química Analítica - *LAQQA*) decidiu propor um trabalho que permitisse a união de medidas espectroscópicas no infravermelho próximo e técnicas que compõem a área de inteligência artificial. O especial interesse na região do infravermelho próximo provém da elevada potencialidade de aplicações e do fato de que, muitas vezes, esta região espectral está associada à não linearidade⁽⁸⁻¹²⁾.

Atualmente, há inúmeras aplicações da espectroscopia no infravermelho próximo em controle de qualidade de produtos farmacêuticos⁽¹³⁻¹⁷⁾, medida de hemoglobina em sangue⁽¹⁸⁾, caracterização de propriedades de polímeros em relação à configuração, conformação, cristalinidade e orientação⁽¹⁹⁾. Pela habilidade de analisar de forma rápida e não-destrutiva uma ampla faixa de propriedades físicas e químicas de várias amostras, a técnica da espectroscopia no infravermelho próximo parece ser capaz de realizar testes de controle de processo, particularmente para aplicações *in-line*, *on-line* e *at-line*^(20,21). Apesar de algumas áreas já empregarem esta ferramenta no controle de qualidade de produção, ela ainda se encontra em estágio experimental no campo de análise laboratorial médica.

Esta tese está dividida em 9 capítulos e uma conclusão final. No capítulo 1, será mostrada uma breve introdução da notação algébrica e teoria básica dos métodos de calibração multivariada, bem

como a interpretação geométrica da Regressão Linear Múltipla (RLM), mínimos quadrados parciais (PLS), detecção de amostras anômalas e a introdução de algumas ferramentas aplicadas para a correção de linha base. No capítulo 2, será apresentada a base da teoria geral do algoritmo genético, no caso especial para a seleção de variáveis. O capítulo 3 apresenta uma descrição detalhada do funcionamento das redes neurais utilizando métodos derivativos, além de uma estrutura híbrida entre redes neurais e algoritmo genético (Neurogenética), onde o algoritmo genético é utilizado para determinar o melhor conjunto de pesos da rede neural. O capítulo 4 é dedicado à introdução dos conceitos e elementos básicos de lógica Fuzzy e Neuro-Fuzzy. No capítulo 5, será fornecida uma breve introdução sobre a teoria de infravermelho próximo, bem como as vantagens e desvantagens desta técnica para amostras no estado sólido (pó, grão, etc) e líquido. O capítulo 6 mostra a avaliação das técnicas de inteligência em um sistema teórico bastante complexo. Neste exemplo, poderá se compreender melhor como funciona o processo de modelagem das técnicas de inteligência artificial. O capítulo 7 consiste em uma breve introdução sobre os parâmetros Brix e Pol e sua importância na indústria sucro-alcooleira. Posteriormente, são apresentados os resultados alcançados na estimativa destes parâmetros. O capítulo 8 apresenta uma introdução e os resultados alcançados na quantificação de um mineral que confere propriedades especiais a um particular polímero. O capítulo 9 apresenta os resultados alcançados na tentativa de determinar triglicerídeos, colesterol e glicose, a partir de amostras de plasma humano. Para finalizar este trabalho é apresentada uma conclusão geral.

Objetivos:

Os objetivos deste trabalho encontram-se dentro das perspectivas do Laboratório de Quimiometria em Química Analítica (LAQQA), do IQ-UNICAMP, que estão centrados basicamente no aprofundamento e domínio das técnicas quimiométricas com a finalidade principal de otimizar a quantificação em sistemas complexos. O objetivo específico deste trabalho foi a realização de um estudo comparativo entre o desempenho de metodologias que utilizam técnicas de inteligência artificial e modelos lineares baseados na Regressão Linear Múltipla e Método dos Mínimos Quadrados Parciais. Além do desenvolvimento de um programa híbrido, usando rede neural e algoritmo genético para a modelagem de espectros na região do infravermelho próximo em amostras de cana de açúcar, polímeros e plasma humano.

1. Introdução a Calibração Multivariada

1.1 Arranjo dos dados e notação Algébrica

Em problemas como os de calibração multivariada, onde o número de objetos e variáveis é elevado, torna-se absolutamente indispensável a disposição ordenada dos dados em forma de matriz, para facilitar sua manipulação. Em uma estrutura matricial, cada linha representa um objeto e as colunas, suas respectivas variáveis.

A Figura 1 ilustra a representação de uma matriz de dados, com n linhas e p colunas, onde o elemento da matriz x_{np} indica o valor do n -ésimo objeto e da p -ésima variável.

Em Quimiometria, os objetos geralmente são espécies químicas e as variáveis são, muitas vezes, derivadas das propriedades das espécies químicas.

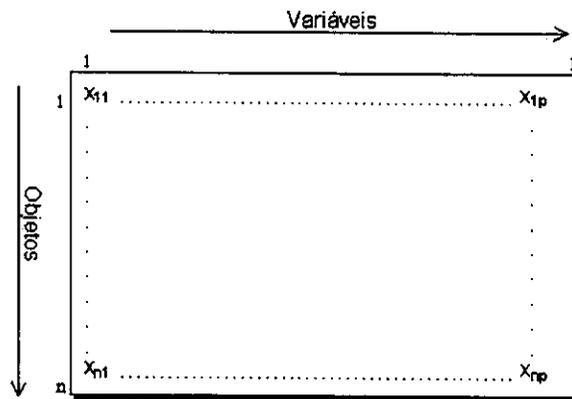


Figura 1. Representação matricial dos dados.

Neste trabalho, será usada a notação padrão de álgebra linear na descrição dos métodos matemático empregados. As matrizes serão denotadas por letras maiúsculas em negrito (\mathbf{X}), os vetores por letras minúsculas em negrito (\mathbf{y}). A transposta de uma matriz ou vetor será representada por um sobrescrito "T" (\mathbf{X}^T), a inversa pelo sobrescrito "-1" (\mathbf{y}^{-1}). A priori, todos os vetores serão vetores coluna, os correspondentes vetores linha serão representados por vetores transpostos. As matrizes diagonais serão denominadas por letras maiúsculas (S). As letras minusculas indicarão os escalares que podem ser elementos de matrizes (a_{ij}), de vetores (a_j) ou outras constantes como coeficientes de regressão (b)⁽²²⁾.

1.2 Pré-processamento dos dados

O pré-processamento é análogo às operações de pesagem, extração, secagem, centrifugação, precipitação de uma análise química e, freqüentemente, é a fase mais difícil, consome mais tempo e determina a qualidade e eficiência do método de calibração multivariada empregado.

Antes de desenvolver um modelo utilizando qualquer dos métodos de calibração multivariada, é aconselhável um pré-tratamento dos dados, para eliminar amostras anômalas (*outliers*), minimizar ruídos e informações superpostas de espécies de interesse, bem como de interferentes.

Durante a etapa de manipulação, preparação, condicionamento ou aquisição dos dados, pode-se ter variações nos procedimentos utilizados nas amostras, dando origem ao que chamamos de amostras anômalas. Os problemas mais típicos que podem causar o aparecimento de amostras anômalas são ⁽²³⁾: interferentes físicos (ex. espalhamento de luz na amostra); interferentes químicos (ex. instabilidade do analito); composição extrema (ex. níveis alto de resposta não linear); erro nos dados de referência (ex. erro na digitalização ou digitação dos dados); problemas nos dados instrumentais (ex. relação sinal/ruído muito baixa).

Um das maneiras de determinar estas amostras é através da análise das componentes principais⁽²³⁾, distância de Mahalanobis⁽²⁴⁻²⁵⁾, etc.

Entre os pré-processamentos, o mais utilizado é o escalamento dos dados, que consiste em atribuir a mesma influência relativa das diferentes variáveis independentes das suas unidades, uma vez que todas as variáveis passam a ser expressas em unidades de desvio padrão.

Para minimizar o ruído, pode-se utilizar o filtro com média móvel, transformada de Fourier, transformada de Wavelet, Savitsky-Golay, etc⁽²³⁾.

A derivada é freqüentemente usada para uma melhor definição de picos que se encontram sobrepostos em uma mesma região e para a correção de linha base.

Outro procedimento para o pré-tratamento dos dados, que será melhor abordado ao longo deste trabalho, é a seleção de variáveis que permite eliminar os termos que não são relevantes na modelagem, gerando uma submatriz com apenas as variáveis que possuem informações relevantes para o modelo.

1.3 Distância Euclidiana e Mahalanobis

Em Quimiometria, a medida de distância entre diferentes objetos é normalmente empregada para estabelecer o grau de similaridade entre eles possibilitando, desta maneira, determinar correlações entre grupos de amostras (*clusters*). Como consequência, esta ferramenta matemática é freqüentemente usada para detecção de amostras anômalas e seleção de amostras de calibração. Isto é realizado calculando a distância da amostra até o centro do grupo.

Entre as técnicas mais difundidas para a medida de distâncias estão a medida Euclidiana e de Mahalanobis.

A distância Euclidiana é representada pela seguinte equação matemática:

$$D_{iE} = \sqrt{(x_{i1} - \bar{x}_1)^2 + (x_{i2} - \bar{x}_2)^2} \quad \text{eq. (01)}$$

onde: x_{i1} e x_{i2} são os valores para o objeto i para as variáveis x_1 e x_2 , respectivamente. Os valores de \bar{x}_1 e \bar{x}_2 são as médias dos valores das variáveis x_1 e x_2 .

Realizando uma interpretação geométrica da equação Euclidiana, chegamos a conclusão que esta equação é equivalente à equação de uma circunferência. Portanto, a distância Euclidiana parte do princípio que os dados possuem uma dispersão simétrica, sendo o centro da equação determinado pelo centro geométrico espacial do conjunto dos dados. Entretanto, este tipo de medida de distância é pouco eficiente em sistemas onde as variáveis têm um alto grau de correlação, como por exemplo, nos espectros de infravermelho próximo. Neste caso, observa-se uma dispersão dos dados em uma direção preferencial do espaço, necessariamente não simétrica. Desta maneira, a distância Euclidiana não é a melhor técnica para afirmar se um ponto pertence ou não a um *cluster*. Para contornar este problema é necessário utilizar a distância de Mahalanobis.

A distância de Mahalanobis é representada pela seguinte equação:

$$D_{iM} = \sqrt{(x_i - \bar{x})C^{-1}(x_i - \bar{x})^T} \quad \text{eq. (02)}$$

onde: o C^{-1} indica a inversa da matriz de variância-covariância, x_i é o i -ésimo vetor da matriz de dados e \bar{x} é o vetor médio da matriz de dados.

O desenvolvimento da equação de Mahalanobis resulta em uma equação de uma elipse, conforme apresentado a seguir:

$$D_{iM} = \sqrt{\left(\frac{x_{i1} - \bar{x}_1}{\sigma_1}\right)^2 + \left[\left(\frac{x_{i2} - \bar{x}_2}{\sigma_2} - \rho_{12} \frac{x_{i1} - \bar{x}_1}{\sigma_1}\right) \frac{1}{\sqrt{1 - \rho_{12}^2}}\right]^2} \quad \text{eq. (03)}$$

onde: σ_1^2 e σ_2^2 indicam a variância da primeira e segunda variável, respectivamente. O parâmetro ρ_{12} indica o grau de correlação entre as duas variáveis.

Pela equação acima é possível observar que a informação explicada pela primeira variável é excluída da segunda variável. Ou seja, a distância de Mahalanobis exclui as informações redundantes dos dados. Além disso, quando as duas variáveis não possuem correlação, ou seja, $\rho_{12} = 0$, a equação se transforma na equação da distância Euclidiana. Portanto, podemos afirmar que a técnica de Mahalanobis é uma medida de distância mais robusta que a Euclidiana.

Entretanto, o uso da distância de Mahalanobis é apenas aconselhável quando o número de amostras ou objetos é maior que o número de variáveis. Pois a inversa da matriz de covariância pode mostrar-se instável para matriz de dados com alto nível de correlação. Para contornar este problema, geralmente é aconselhável utilizar a análise das componentes principais para reduzir a dimensionalidade do espaço das variáveis.

1.4 Relação entre Matrizes de dados

A calibração multivariada utiliza modelos matemáticos para estabelecer uma relação entre uma propriedade que possa ser monitorada, com alguma outra propriedade de interesse. A calibração multivariada pode ser dividida em duas fases: calibração e previsão. Na etapa de calibração é encontrada uma relação entre as medidas físicas (matriz X, variáveis independentes) e as propriedades de interesse (matriz Y, variáveis dependentes). Na etapa de validação, avalia-se a performance do modelo matemático aplicando a um segundo conjunto de amostras. Os dados usados nesta fase recebem o nome de conjunto de previsão.

A representação esquemática da etapa de calibração está ilustrada na Figura 2.

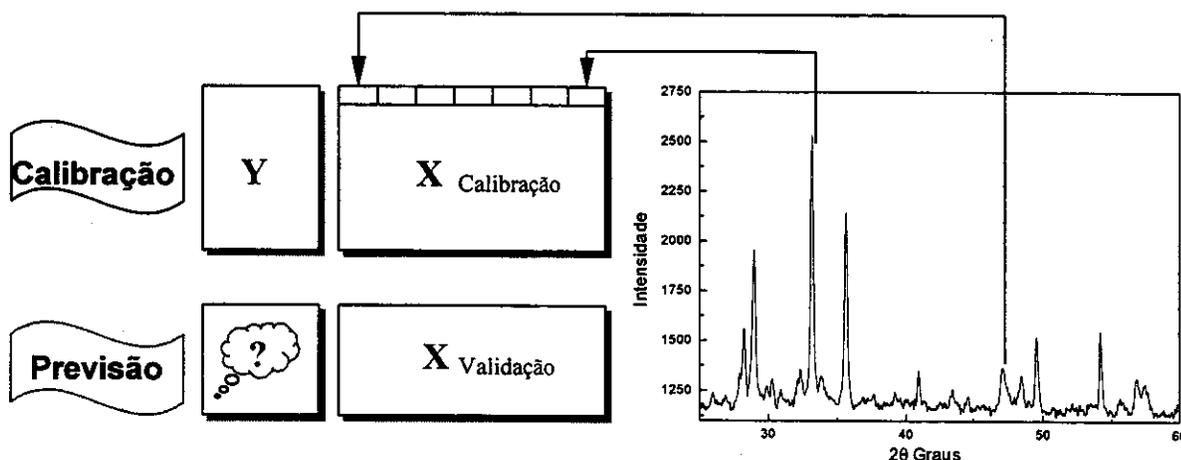


Figura 2. Organização dos dados para a calibração multivariada.

1.5 Regressão linear Múltipla (RLM)

O objetivo da regressão linear múltipla é encontrar uma combinação linear entre as variáveis independentes (medidas instrumentais) e as variáveis dependentes (propriedade de interesse) que minimiza o erro na estimativa dessas propriedades de interesse⁽²⁰⁾. Tal relação pode ser escrita como:

$$\hat{y} = x_1 b_1 + x_2 b_2 + x_3 b_3 + \dots + x_p b_p + e \quad \text{eq. (04)}$$

$$\hat{y} = \mathbf{x}^T \mathbf{b} + e \quad \text{eq. (05)}$$

onde: “ \hat{y} ” é propriedade estimada (variável dependente); “ \mathbf{x} ” é vetor com as medidas instrumentais (variáveis independentes); “ \mathbf{b} ” é o vetor dos coeficientes de regressão; “ e ” é o resíduo ou erro.

A regressão linear múltipla nada mais é que a resolução de sistemas lineares. Portanto, as seguintes condições são válidas (p é o número de variáveis e n é o número de amostras): se $p > n$, (onde se tem mais variáveis que amostras) não existe solução para \mathbf{b} ;
 se $p = n$, (onde se tem número de amostras e variáveis iguais) há uma única solução para \mathbf{b} ;
 se $p < n$, (onde se tem mais amostras que variáveis) não existe uma solução exata para \mathbf{b} .

A regressão linear múltipla pode ser entendida como a projeção da(s) coluna(s) das variáveis dependentes sobre o espaço definido pelas colunas da matriz de variáveis independentes.

Os coeficientes da regressão linear múltipla (equação 04) são calculados utilizando o critério dos mínimos quadrados, através da equação 06:

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad \text{eq. (06)}$$

As limitações da regressão linear múltipla concentram-se no número de amostras que não pode ser menor do que o número de variáveis, e em determinar a inversa de $(\mathbf{X}^T \mathbf{X})$, pois se as variáveis independentes possuírem uma elevada correlação, esta inversa pode não existir. Mesmo assim, se $(\mathbf{X}^T \mathbf{X})^{-1}$ existir com correlações altas, os valores previstos para “b” terão erros elevados.

1.6 Mínimos Quadrados Parciais (PLS)

O método dos mínimos quadrados parciais^(26,27) é um modelo baseado em variáveis latentes (fatores), onde cada fator é definido como uma combinação linear das variáveis originais da matriz \mathbf{X} (variáveis independentes) ou \mathbf{Y} (variáveis dependentes).

A primeira componente principal correspondente ao maior autovalor é, pela definição, a direção no espaço de \mathbf{X} que descreve a máxima quantidade de variância nas amostras. Quando toda a variância de um conjunto de amostras não puder ser explicada por apenas uma componente principal, uma segunda componente principal perpendicular ou ortogonal à primeira será usada. Após a modelagem, teoricamente a matriz dos quadrados dos resíduos deverá conter apenas a variância não explicada associada ao ruído, uma vez que os fatores são calculados de maneira que expliquem a máxima variância possível associada à propriedade de interesse que se deseja modelar. Além disso, esta técnica também pode modelar algumas não linearidades.

A importância da ortogonalidade das componentes principais é devido ao fato de que somente desta forma pode-se garantir que a nova base gerada resulte de uma combinação de vetores linearmente independentes, portanto constituindo um novo espaço vetorial.

O método dos mínimos quadrados parciais (PLS – Partial Least Squares) foi proposto inicialmente por H. Wold é calculado pelo algoritmo NIPALS⁽²⁶⁾. Este consiste em escrever a matriz \mathbf{X} de posto p como uma soma de matrizes de posto 1, conforme a equação 07.

$$\mathbf{X} = \mathbf{M}_1 + \mathbf{M}_2 + \mathbf{M}_3 + \dots + \mathbf{M}_p \quad \text{eq. (07)}$$

Uma matriz de posto 1 pode ser escrita como produto interno de dois vetores, os escores (\mathbf{t}_h ou \mathbf{u}_h) e os pesos (\mathbf{p}_h^T ou \mathbf{q}_h^T).

$$\mathbf{X} = t_1 \mathbf{p}_1^T + t_2 \mathbf{p}_2^T + \dots + t_p \mathbf{p}_p^T \quad \text{eq. (08)}$$

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad \text{eq. (09)}$$

Da mesma forma, a matriz \mathbf{Y} (matriz de variáveis dependentes) também é decomposta como:

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{F} \quad \text{eq. (10)}$$

Após obter-se as relações externas das componentes principais de \mathbf{X} ($\mathbf{X} = \mathbf{TP}^T + \mathbf{E}$) e \mathbf{Y} ($\mathbf{Y} = \mathbf{UQ}^T + \mathbf{F}$), realiza-se uma relação interna entre os escores de \mathbf{X} e \mathbf{Y} , a fim de se determinar uma relação linear da forma $u_h = b_h t_h$ onde:

$$b_h = \mathbf{u}_h \mathbf{t}_h / \mathbf{t}_h^T \mathbf{t}_h \quad \text{eq. (11)}$$

O parâmetro b representa o coeficiente angular do vetor que melhor explica a relação existente entre os escores de \mathbf{X} e \mathbf{Y} . Como as componentes principais de \mathbf{X} e \mathbf{Y} são calculadas independentemente, a relação entre elas não é necessariamente a melhor. Para minimizar este problema, faz-se uma leve rotação nos eixos das componentes principais de forma a não comprometer a ortogonalidade. Esta leve rotação das componentes principais provoca um desvio da direção de maior variância das matrizes originais.

1.7 Detecção de Amostras Anômalas (outliers)

Em calibração multivariada, normalmente faz-se a detecção de anomalias no bloco das variáveis dependentes (\mathbf{Y}) e independentes (\mathbf{X}). No bloco \mathbf{Y} é necessário apenas o uso de ferramentas tradicionais como o teste de Grubbs^(6,21) ou Dixon^(6,21), enquanto que, para o bloco \mathbf{X} , é necessário o uso de ferramentas mais sofisticadas, que operam em espaços multidimensionais.

Uma importante medida para determinar a presença de uma amostra anômala é a medida de influência (*Leverage*). Esta ferramenta matemática indica o grau de influência de uma amostra na estimativa dos parâmetros de regressão como: resposta, coeficientes de regressão e erro padrão.

A medida de influência (h_i) está relacionada à distância de Mahalanobis, conforme a seguinte equação ilustra:

$$h_i = \frac{1}{n} + \frac{(D_i M)^2}{n-1} \quad \text{eq. (12)}$$

onde: n é o número de objetos ou amostras e $D_i M$ é distância de Mahalanobis, calculada a partir dos valores dos escores normalizados, obtidos a partir da análise das componentes principais.

A equação acima pode ser convertida em uma equação que somente depende dos valores dos escores normalizados. Portanto, por questão de praticidade e velocidade de processamento, a medida de influência é calculada pela seguinte equação:

$$h_i = \text{diag}(TT^t) \quad \text{eq. (13)}$$

onde: a matriz T representa os escores normalizados e “diag”, o operador que fornece os elementos contidos na diagonal principal da matriz de dados.

Existem inúmeros critérios para avaliar se uma amostra deve ou não ser considerada como outlier. Entretanto, será apresentado apenas o critério mais comum, o qual leva em consideração o número de parâmetros de regressão (g) e o número de amostras (n), sendo representado pela seguinte expressão matemática: $\text{lim} = \frac{2g}{n}$.

Quando o valor do parâmetro h_i superar o limite (lim), a amostra é pode ser um possível outlier.

1.8 Correção de espalhamento Multiplicativo (MSC)

O método de correção de espalhamento multiplicativo (*multiple scattering correction*) é comumente aplicado para a correção de linha base, proveniente principalmente da não homogeneidade da distribuição de partículas em materiais sólidos²¹, em espectroscopia no infravermelho próximo.

Este método assume que os comprimentos de onda da luz espalhada possuem uma dependência distinta entre a luz espalhada e a absorvida pelos constituintes da amostra. Portanto, teoricamente, é possível separar estes dois sinais. Este método tenta remover o efeito do espalhamento pela linearização de cada espectro por um espectro ideal. Para efeito de cálculo, considera-se que o espectro ideal é o espectro médio do conjunto de dados para o qual deseja realizar a correção de linha base. Em seguida, utiliza-se uma regressão linear para calcular o coeficiente angular e linear do gráfico entre o espectro ideal e o espectro que vai ser corrigido. O espectro corrigido é calculado subtraindo cada ponto do espectro pelo valor do coeficiente linear e dividindo este valor pelo coeficiente angular.

Esta técnica garante que a informação química seja preservada enquanto, as diferenças entre os espectros sejam minimizadas. Entretanto, esta técnica somente pode ser empregada para espectros que possuem relação linear com a concentração.

1.9 Transformação Padrão Normal de Variação (SNV)

Normalmente, os espectros de reflectância no infravermelho próximo apresentam problemas de linha base, inclinações e algumas vezes curvaturas, devido principalmente ao espalhamento de luz^(23,28,29). O espalhamento é fortemente dependente do comprimento de onda da luz, do tamanho das partículas, do nível de compactação, do índice de refração da amostra, etc. Para minimizar este efeito, é necessário o uso de técnicas como a transformação padrão normal de variação (*Standard Normal Variate Transformation*). Esta técnica é aplicada para corrigir os efeitos do espalhamento multiplicativo e o tamanho de partícula, de maneira análoga à correção de espalhamento multiplicativo (MSC).

Apesar do MSC e SNV terem a mesma finalidade, ou seja corrigir a linha base, estas duas técnicas são distintas. O SNV não necessita de um espectro ideal para fazer a correção dos espectros. A correção é realizada pela normalização de cada espectro para o seu próprio desvio padrão, conforme ilustrado pelas equações a seguir:

$$\text{Média do espectro} \quad \bar{x}_i = \sum_{j=1}^p \frac{X_{i,j}}{p} \quad \text{eq. (14)}$$

$$\text{Espectro Corrigido} \quad X_{i(SNV)} = \frac{(X_{i,j} - \bar{x}_i)}{\sqrt{\frac{\sum_{j=1}^p (X_{i,j} - \bar{x}_i)^2}{p-1}}} \quad \text{eq. (15)}$$

onde: X representa uma matriz com n espectros e p comprimentos de onda, \bar{x}_i é a média do vetor contendo o espectro i da matriz X .

Como este método corrige cada espectro independentemente, ele pode ser aplicado em casos onde ocorra maior variabilidade entre um conjunto de espectros.

2. Introdução ao Algoritmo Genético

2.1. Teoria de Darwin

A teoria da evolução das espécies originou-se das observações realizadas por Charles R. Darwin, no século 19 durante uma expedição científica ao redor do mundo. Nesta ocasião Darwin assumia a tarefa de naturalista a bordo do navio Beagle.⁽³⁰⁾ Durante esta jornada de cinco anos, Darwin presenciou inúmeras situações que o levaram a conjecturar sobre a possível seleção natural das espécies. No arquipélago dos Galápagos, ele verificou que cada ilha possuía uma fauna característica, principalmente com relação a pássaros, tartarugas e lagartos. Em 1858, Darwin apresentou a teoria da seleção natural numa sessão da Academia Britânica de Ciências. No ano seguinte, publicou um resumo do livro “Natural Selection”, onde introduzia a teoria da Evolução das espécies. Segundo Darwin, no mecanismo de seleção natural, os indivíduos mais adaptados têm maior probabilidade de sobrevivência do que os menos adaptados. Como consequência disso, os mais adaptados teriam maior número de descendentes, o que provocaria, na geração seguinte, um aumento no seu número, em relação aos não adaptados. Tais mudanças remetem à evolução biológica, a qual ocasiona mudanças nas características das populações dos organismos que transcendem o período de vida de um único indivíduo. As mudanças nas populações que são consideradas evolutivas são aquelas herdáveis via material genético, de uma geração para a outra. A evolução biológica pode ser pequena ou substancial, sendo estes efeitos responsáveis pelas adaptações dos organismos a diferentes ambientes⁽³⁰⁾.

Os indivíduos podem ser vistos como uma dualidade entre seu código genético (*genótipo*) e suas características comportamentais, fisiológicas e morfológicas (*fenótipo*). Em sistemas evoluídos naturalmente, não existe uma relação biunívoca entre um gene (elemento do genótipo) e uma característica (elemento do fenótipo): um único gene pode afetar diversos traços fenotípicos simultaneamente (*pleiotropia*) e uma única característica fenotípica pode ser determinada pela interação de vários genes (*poligenia*). A pleiotropia e poligenia geralmente tornam os resultados de variações genéticas imprevisíveis. Sistemas naturais em evolução são fortemente pleiotrópicos e altamente poligênicos. O mesmo não ocorre em sistemas artificiais, onde uma das principais preocupações é com o custo computacional do sistema. Assim, em sistemas artificiais, existe uma relação de um-para-um entre genótipo e fenótipo^(30,31).

A seleção natural escolhe os melhores fenótipos, de acordo com as restrições ambientais e parâmetros evolutivos. Em sistemas biológicos reais, o ambiente está em constante mudança, fazendo com que populações estejam em constante evolução em direção a novos pontos de ótimo. Neste caso, assumindo que as mudanças ambientais são significativas, a taxa evolutiva deve ser suficientemente elevada para acompanhar as mudanças ambientais. Caso a taxa de evolução não seja significativa, isso pode implicar na extinção de uma determinada espécie.

2.2 Algoritmo Genético

A partir da teoria evolutiva das espécies de Darwin, John H. Holland pesquisador da Universidade de Michigan, propôs nos anos 60 a construção de um algoritmo matemático para otimização de sistemas complexos, batizado de Algoritmo Genético⁽³²⁾.

Foi nos anos oitenta que o algoritmo genético recebeu um grande impulso em diversas áreas de aplicação científica, devido principalmente à popularização dos computadores e conseqüentemente aparecimento de computadores cada vez mais rápidos e potentes. Suas áreas de maior aplicação são ⁽³³⁾: processamento de imagem; modelagem e identificação de sistemas; filtros adaptativos para o cancelamento de ruídos; robótica; parâmetros de redes neurais, etc.

Em química, a primeira aplicação de algoritmo genético ocorreu no início dos anos noventa, por Lucasius e Kateman⁽³⁴⁾. As aplicações deste algoritmo cresceram, principalmente devido às vantagens que este apresenta quando comparado a outros métodos de otimização. Como principais vantagens, pode-se destacar ⁽³²⁾: não requer informações sobre o gradiente da superfície de resposta; eventuais descontinuidades da superfície de resposta não afetam a performance da otimização; a presença de mínimos locais não reduz a eficiência do algoritmo; a performance do algoritmo tem apresentado excelentes resultados para problemas de otimização de grande escala.

Basicamente há quatro tipos principais de algoritmos genéticos: Algoritmo genético de Holland, programa genético de Koza, programa evolucionário de Fogel e estratégia evolucionária de Rechenberg. Eles diferem pelo mecanismo de codificação e procedimento de seleção para a reprodução. O mais popular é o de Holland, pois este codifica as informações na forma de código binário (0,1)⁽³⁵⁻³⁶⁾.

2.2.1 Algoritmo Genético Básico

O algoritmo genético foi estruturado de forma que as informações referentes a um determinado sistema pudessem ser codificadas de maneira análoga à dos cromossomos biológicos. Desta forma, o algoritmo proposto assemelha-se ao processo evolutivo biológico, facilitando a confecção do algoritmo matemático.

O algoritmo genético básico envolve cinco passos: codificação das variáveis, criação da população inicial, avaliação da resposta, cruzamento e mutação.

i. Codificação das variáveis

O cromossomo biológico é composto de genes (genótipo), os quais são responsáveis por uma determinada característica do indivíduo (fenótipo), como por exemplo, cor dos olhos, altura, cor da pele, etc. Através de uma analogia, é possível construir um cromossomo artificial, por exemplo, por um processo químico onde cada gene representa um determinado parâmetro que se deseja otimizar: temperatura, agitação, tempo, quantidade do catalisador (conforme Figura 3).

Os cromossomos artificiais são usualmente implementados na forma de vetores, onde cada componente do vetor é conhecido como *gene*. Os possíveis valores que um determinado gene pode assumir são denominados de *alelos*.

Temperatura	Catalisador	Agitação	Tempo
35	8	43	56
1 0 0 0 1 1	0 0 1 0 0 0	1 0 1 0 1 1	1 1 1 0 0 0

Figura 3. Codificação dos parâmetros do cromossomo em números binários de seis bits.

Neste tipo de sistema, a codificação em número binário é a forma mais simples de representar as informações relativas para cada parâmetro que será otimizado (como temperatura, catalisador, agitação e tipo de catalisador). Assim, cada cromossomo artificial será representado por

uma seqüência de valores de zero e um (forma binária para cada parâmetro). As vantagens deste tipo de código são ⁽³²⁾: torna as operações genéticas mais simples de serem executadas; torna-se possível o trabalho com variáveis que assumem valores contínuos; facilita a variação da faixa de trabalho dos parâmetros e apresenta relativa simplicidade na variação da precisão dos valores de cada parâmetro.

A Figura 3 mostra a codificação dos valores relativos a quatro parâmetros que serão otimizados. Neste caso, deve-se codificar para binário o valor de temperatura igual a 35 graus, o valor do catalisador igual a 8 gramas, o valor de agitação igual a 43 rpm e o valor do tempo igual a 56 minutos. Tomando-se como seis bits cada um desses parâmetros, tem-se assim o cromossomo formado por 24 unidades binárias. Nesse caso, os seis primeiros bits formam o gene correspondente à temperatura e, na seqüência, de seis bits os genes para catalisador, agitação e tempo.

ii - Criação da população inicial

Através de um gerador aleatório, cria-se uma população inicial de cromossomos, como mostrado na Figura 4, com uma população de sete indivíduos (onde cada indivíduo é representado por um cromossomo).

```
[1]100000000001001000101010  
[2]000100100100100000000010  
[3]100001000000000000010000  
[4]000000000100000000001000  
[5]010001000000010000000000  
[6]000000001000000001000000  
[7]000001000000001000000100
```

Figura 4. Representação de uma população aleatória de sete cromossomos.

Sendo a população inicial aleatória, garantimos que não haverá nenhum tipo de influência tendenciosa do meio externo. Desta forma tenta-se garantir a idoneidade da pesquisa. Entretanto, há casos onde existem informações adicionais que podem ser introduzidas na geração da população inicial, minimizando o tempo de computação.

Na população inicial, pode-se estipular qual será o número máximo e mínimo de unidades binárias que um determinado cromossomo poderá possuir. Esse número é algumas vezes um fator limitante, podendo estar associado ao método utilizado para calcular a resposta.

iii - Avaliação da resposta ou *Aptidão*

Deve-se encontrar o valor associado ao desempenho de cada cromossomo, relacionado ao sistema de interesse. A resposta de cada cromossomo é o resultado mais importante no procedimento do algoritmo genético.

A resposta (aptidão) é uma característica intrínseca do indivíduo. Esta indicará qual o nível de habilidade que um indivíduo possui para sobreviver a predadores, pestes e a outros obstáculos, até que possa se reproduzir. Transportando para o algoritmo matemático, seria sua habilidade de produzir a melhor resposta. Portanto, esta resposta será responsável pela vida ou morte dos indivíduos. Em alguns algoritmos, é o grau de aptidão que determina a probabilidade com que um determinado cromossomo irá cruzar. Portanto, desta forma, pode-se determinar quais são os indivíduos que poderão realizar mais de um cruzamento durante uma mesma geração.

iv - Cruzamento

O cruzamento representa a permuta entre os genes de diferentes indivíduos. A partir de uma população inicial, uma nova população é formada pelo cruzamento aleatório entre os cromossomos. Portanto, os filhos recebem informações de seus progenitores através do material genético proveniente deste cruzamento. Durante a permuta de material genético entre os indivíduos, haverá uma tendência de que os indivíduos que possuem características dominantes transmitam-nas para as gerações futuras. Após algumas gerações, pode-se observar que uma elevada taxa dos indivíduos possuem a presença de variáveis '*dominantes*'. O cruzamento é a etapa responsável pela convergência do processo de otimização. Um aspecto importante em um algoritmo genético diz respeito em como escolher os indivíduos que serão submetidos ao cruzamento. Aqui também temos diversas alternativas possíveis, sendo que entre as mais comuns está o cruzamento entre indivíduos aleatórios (onde são escolhidos aleatoriamente indivíduos da população atual) ou atribui-se um peso maior para os indivíduos mais aptos. O processo de cruzamento dá-se pela escolha aleatória dos indivíduos e da posição onde ocorrerá a ruptura seguida pela troca de material genético. O número de rupturas geralmente varia entre 1 e 2. Isto se deve ao fato de que quanto maior for o número de rupturas, menor será a semelhança entre os pais e os filhos, dificultando a convergência.

A Figura 5 mostra a etapa do cruzamento entre dois indivíduos, para o gene (pedaço do cromossomo) relacionado à variável catalisador, que está sendo otimizada.

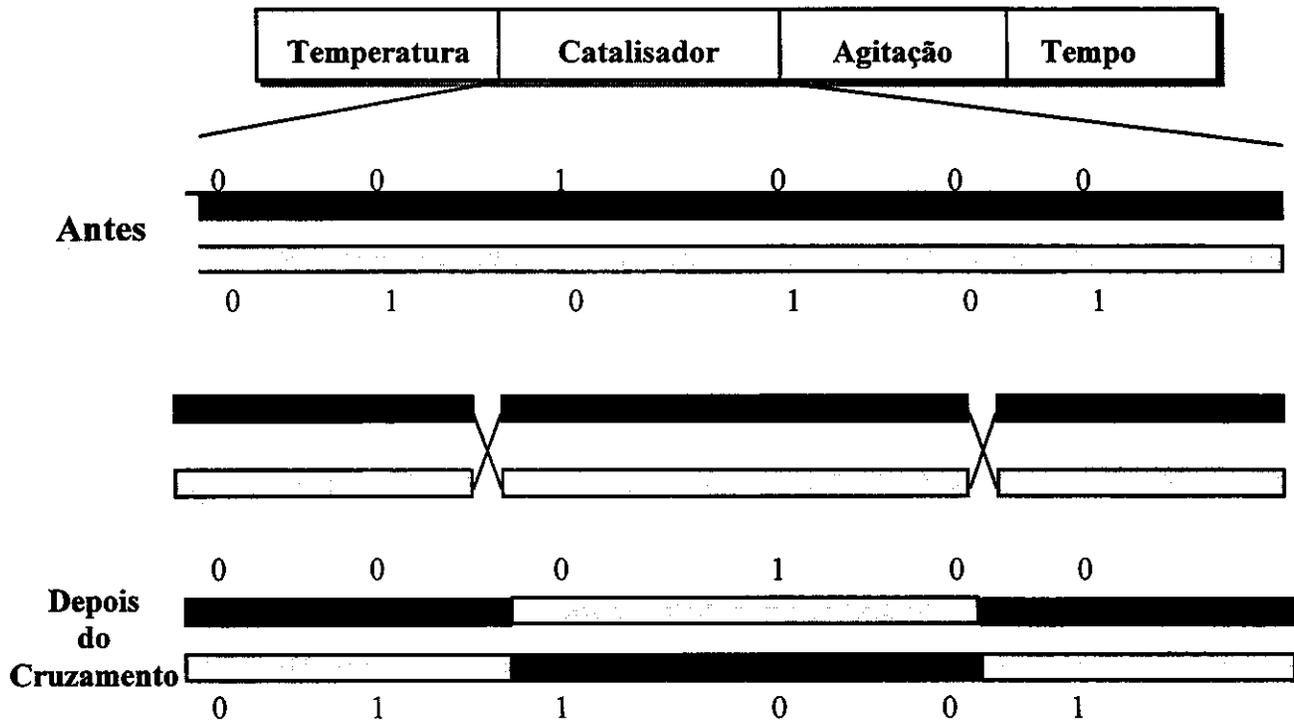


Figura 5. Operação de cruzamento entre cromossomos.

v - Mutação

Durante o processo de reprodução, pode ocorrer uma perturbação das informações contidas dentro dos genes, o que acarreta em uma mutação do material genético.

A mutação promove alterações no código genético de uma pequena parcela da população, onde muitas vezes pode atuar de forma benigna. Se a alteração gerar características melhores, estas serão transmitidas para os demais indivíduos ao longo das próximas gerações.

A mutação ajuda a solucionar o problema do confinamento em mínimos locais, pois promove alterações que direcionam a pesquisa para fora desta região. A Figura 6 representa dois cromossomos, onde ocorreu mutação no material genético de um deles. A mutação na representação binária é realizada pela troca de **um** por **zero** e vice-versa, conforme pode ser notado na Figura 6.

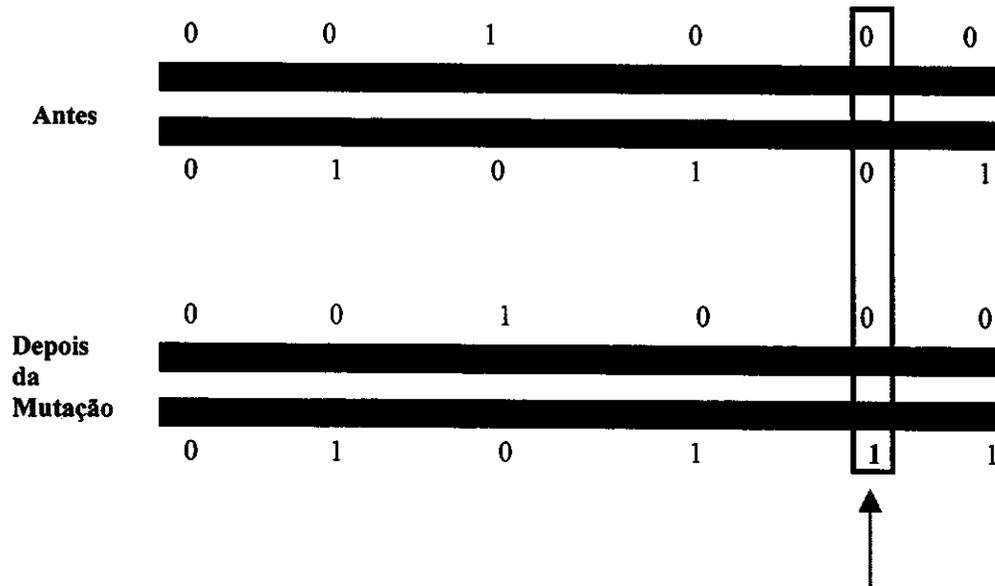


Figura 6. Operação de mutação nos cromossomos após o cruzamento.

vi - Principais Etapas do Algoritmo Genético básico

A Figura 7 mostra as principais etapas do algoritmo genético básico. O processo é continuamente repetido até que uma das condições de parada sejam satisfeitas. Como critério de parada para o algoritmo genético, normalmente adota-se o número máximo de gerações ou um erro mínimo desejado.

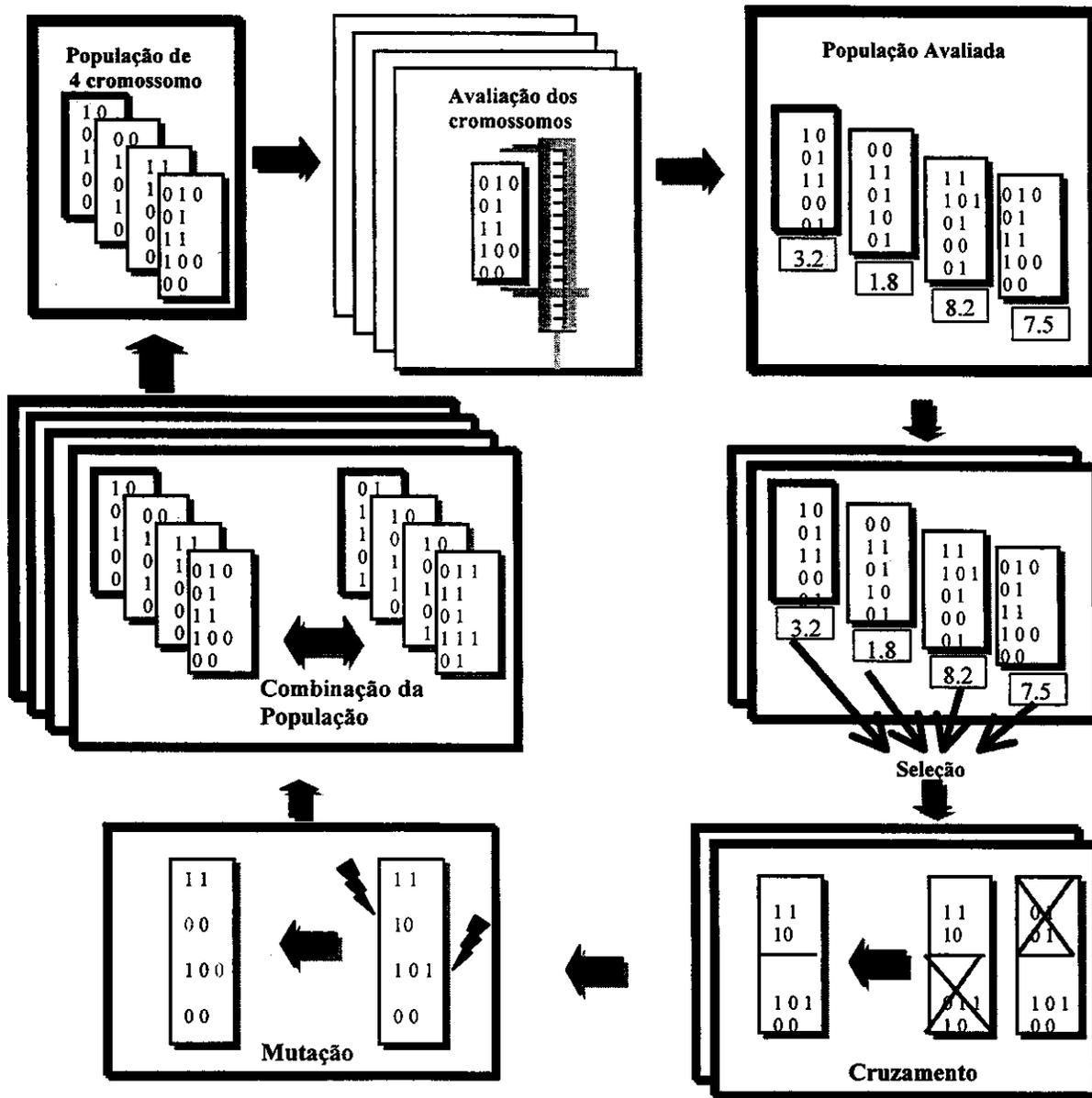


Figura 7. Principais etapas do algoritmo genético básico.

Para um bom desempenho do algoritmo, faz-se necessário otimizar parâmetros como probabilidade de mutação, probabilidade de cruzamento, tamanho da população inicial e número de gerações.⁽³²⁾ O tamanho da população inicial é escolhido entre 50-500 membros. Normalmente, utiliza-se probabilidade de mutação de aproximadamente 1%, probabilidade de cruzamento ao redor de 90%. Enquanto, o número de gerações irá depender do tamanho da população inicial.⁽³²⁾

2.3 Algoritmo Genético em Otimização

A otimização envolve a maximização ou minimização de uma função ou uma série de variáveis num processo. Dentre os métodos de otimização comumente utilizados podem ser destacados o método de otimização Simplex⁽²⁶⁾, recozimento simulado (Simulated Annealing)⁽³⁵⁾, Algoritmo Genético e os Métodos Clássicos de Gradiente⁽³⁷⁾.

O estudo desses métodos de otimização é importante, pois sistemas químicos geralmente podem ser complexos, quando se trabalha com um grande número de variáveis. Isto torna muitas vezes impossível ou impraticável sua otimização em processos de tentativa e erro.

Para aumentar a probabilidade de encontrar um ótimo global em sistemas complexos, o algoritmo deve ter alguns mecanismos inteligentes, para que possa caminhar em direção das regiões de melhor desempenho.

De acordo com o grau de informações que se possui de um determinado sistema, o método de otimização pode ser classificado em método fraco ou forte. Os métodos fracos são aqueles onde se fazem poucas aproximações. Este é caracterizado por exigir um alto tempo de processamento, contudo é um método mais robusto⁽³³⁾. Os métodos fortes, dedicados ou tailorizados assumem um grande número de aproximações. Estes métodos apresentam como grande vantagem o curto tempo de processamento; contudo, fornecem um modelo frágil, tendo uma estreita faixa de aplicação⁽³³⁾.

Nenhum dos métodos de otimização garante que a resposta encontrada seja a melhor, portanto esta resposta é chamada de solução estimada. Apesar de não ser a melhor solução, esta pode apresentar resultados suficientemente bons para muitos propósitos experimentais⁽³²⁾.

O algoritmo genético encontra-se em uma faixa intermediária entre os métodos de otimização fracos e fortes - sendo considerado moderado. Ele concilia a vantagem de ser um método robusto e gastar um tempo médio de processamento⁽³²⁾.

O diferencial do algoritmo genético encontra-se no fato deste realizar uma pesquisa paralela na superfície de resposta e não uma pesquisa ponto a ponto como os demais métodos. Isto confere ao algoritmo genético uma maior robustez⁽³²⁾. As regras de seleção usadas pelo algoritmo genético podem ser classificadas em probabilísticas e não determinísticas.⁽³²⁾ Há essencialmente três fatores que diferenciam o algoritmo genético das técnicas convencionais de otimização^(32,33): pesquisa a partir de uma população e não de um simples ponto; pesquisa via amostragem; pesquisa usando operadores estocásticos (regras não determinísticas).

2.4 Operações e Técnicas Avançadas

Após o desenvolvimento do primeiro algoritmo genético, inúmeras modificações foram propostas com o objetivo de aperfeiçoar o desempenho e robustez do programa.

Os primeiros algoritmos eram compostos de basicamente quatro etapas: geração da população inicial, cruzamento, mutação e avaliação. Atualmente em programas mais sofisticados pode-se encontrar inúmeros outros operadores genéticos, como:

- Dominância
- Distinção de sexo:
 - . Homossexuais
 - . Heterossexuais
 - . Assexuados
- Incestos

Dominância: Quando um determinado gene é responsável por uma característica que faz com que um indivíduo se destaque entre um determinado grupo. Desta forma, torna-se interessante a preservação desta informação e a propagação para gerações futuras, podendo acelerar o processo de convergência.

O operador de dominância possui a função de localizar e preservar os genes responsáveis pelo destaque em seu desempenho. Esse operador é implementado neste algoritmo matemático, fazendo com que após um determinado número de gerações, seja inicializada uma sub-rotina que emparelhe dois cromossomos selecionados aleatoriamente e crie um novo indivíduo que possua as características dominantes dos progenitores. Este tipo de operação está representado no esquema, da Figura 8, onde os *zeros* representam as informações dominantes e os valores *um* as recessivas. Portanto, além da população proveniente do cruzamento, há também o acréscimo de mais estes indivíduos à população final desta geração.

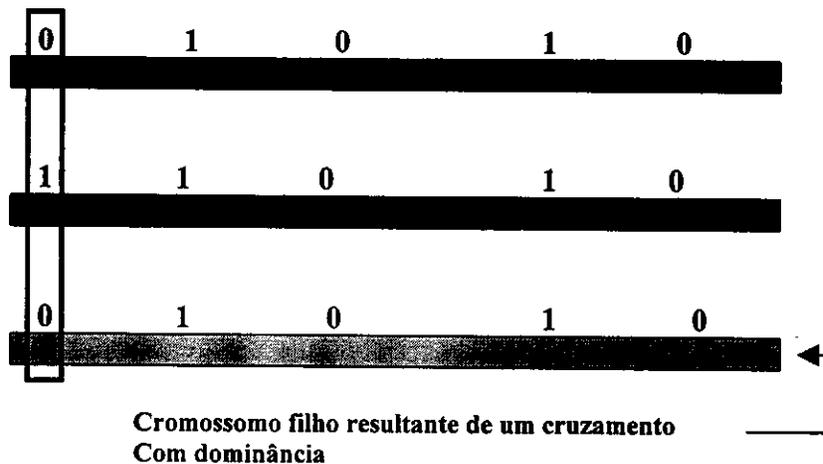


Figura 8. Esquema ilustrativo de um cruzamento com dominância.

Distinção de Sexos: O cruzamento é a operação onde são gerados novos indivíduos a partir de cromossomos pais. O cruzamento pode apresentar-se de três formas distintas⁽³²⁾: homossexuais; heterossexuais e assexuado.

O cruzamento por distinção de sexo é realizado através de um bit reservado para distinguir o sexo dos cromossomos; por exemplo, representando-se um indivíduo do sexo masculino como tendo o bit “1”, em uma determinada posição do cromossomo. A partir deste bit, é possível controlar o cruzamento entre os indivíduos⁽³²⁾.

Nos cruzamentos homossexuais ocorre o cruzamento entre cromossomos de mesmo sexo e nos heterossexuais ocorre o cruzamento entre cromossomos de sexo distintos. Já no cruzamento assexuado ocorre troca de material genético com o próprio indivíduo.

O controle pela distinção de sexo faz com que a diversificação seja maior, já que determinadas características somente um determinado sexo possui. Isso pode ser observado na natureza, uma vez que seres inferiores são hermafroditas.

Incestos: O cruzamento entre indivíduos muito semelhantes leva a geração de um cromossomo muito parecido com os dos pais. Desta forma não há nenhuma melhora significativa para a nova geração⁽³²⁾. Portanto, pode-se criar uma sub-rotina que evite o incesto, durante a etapa de cruzamento. O incesto é mais provável ocorrer após várias gerações.

2.5 Teoria dos Esquemas

A teoria dos esquemas foi proposta por Holland para tentar explicar por que os algoritmos genéticos conseguem convergir para um ponto de máximo ou mínimo. Para atingir este objetivo, Holland propôs as seguintes teorias: teorema de crescimento dos esquemas e a hipótese dos blocos construtivos⁽³²⁾. Um esquema é uma representação genérica capaz de descrever diversos cromossomos simultaneamente.

Um esquema é construído inserindo um caracter genérico (*) no alfabeto dos genes, indicando que esse alelo pode representar qualquer um dos possíveis valores. Por exemplo, o esquema [* 1 0 0 1] representa os cromossomos [0 1 0 0 1] e [1 1 0 0 1]. O esquema [0 * 0 * 1] representa quatro cromossomos: [0 0 0 0 1], [0 0 0 1 1], [0 1 0 0 1] e [0 1 0 1 1]. Portanto, o esquema [1 1 1 0 0] pode representar somente um determinado cromossomo, enquanto que o esquema [* * * * *] representa todos os cromossomos possíveis de comprimento 5. Cada cromossomo de comprimento m é representado por 2^m esquemas. Por exemplo, considere o cromossomo [* * * * *]. Este cromossomo é representado pelos seguintes 2^5 esquemas:

[0 1 0 0 1]	[* * 0 0 1]	[* * * 0 1]
[* 1 0 0 1]	[* 1 * 0 1]	[* * 0 * 1]
[0 * 0 0 1]	[* 1 0 * 1]	[* * 0 0 *]
[0 1 * 0 1]	:	:
[0 1 0 * 1]	:	:
[0 1 0 0 *]	[0 1 0 * *]	[0 1 * * *]	... [* * * * *]

A ordem de um esquema S , $o(S)$, é definida como o número de 0's e 1's presentes no esquema, isto é, o número de posições fixas (caracteres diferentes de “*”) presentes no esquema. A ordem de um esquema define sua especificidade, de modo que quanto maior a ordem, mais específico é o esquema.

O comprimento de um esquema S , denotado por $d(S)$, é a maior distância entre posições fixas de um cromossomo. O comprimento que define o nível de compactação da informação contida no esquema.

A aptidão de um esquema S na geração t , $eval(S, t)$, é definido como a média do grau de aptidão de cada um dos cromossomos membros de uma população que podem ser representados pelo esquema S . Assumindo que há p cromossomos $\{x'_{i_1}, \dots, x'_{i_p}\}$ representados pelo esquema S_i na geração t :

$$eval(s_i, t) = \frac{1}{p} \sum_{j=1}^p eval(x'_{i_j}) \quad \text{eq. (16)}$$

onde $eval(x'_{i_j})$ é o grau de aptidão do indivíduo x'_{i_j} .

Se Tam_Pop representa o tamanho da população, o grau de aptidão médio da população na geração t , $\bar{F}(x)$ é dado por:

$$\bar{F}(x) = \frac{1}{Tam_Pop} \sum_{i=1}^{Tam_Pop} eval(x'_i) \quad \text{eq. (17)}$$

O número de cromossomos do esquema S_i na geração t pode ser representado por $\xi(S_i, t)$. Holland demonstrou que é possível estabelecer uma relação entre estes parâmetros, chegando a seguinte relação matemática⁽²⁹⁾:

$$\xi(S_i, t+1) \geq \frac{\xi(S_i, t) eval(S_i, t)}{\bar{F}(t)} \left[1 - p_c \frac{\delta(S_i)}{m-1} - o(S_i) p_m \right] \quad \text{eq. (18)}$$

onde p_c e p_m são as probabilidades de cruzamento e mutação, respectivamente, e m o comprimento do cromossomo.

A inequação acima é conhecida como equação de crescimento reprodutivo do esquema. Esta equação é deduzida admitindo que a função de aptidão assume apenas valores positivos.

A equação de crescimento mostra que a seleção aumenta a amostragem de esquemas cujo grau de aptidão está acima da média da população, e este aumento é exponencial. A seleção, por si só, não introduz nenhum novo esquema (não representado na geração inicial “ t_0 ”). Esta é a razão da introdução do operador de cruzamento, o qual possibilita a troca de informação estruturada, ainda que aleatória. Além disso, o operador de mutação introduz uma variabilidade maior na população. O efeito (destrutivo) combinado destes operadores não é significativo se o esquema é curto e de ordem baixa.

2.6 Aplicação do Algoritmo Genético em Química

Na química tem-se observado uma acentuada tendência na aplicação de algoritmo genético em inúmeros problemas de otimização de sistemas complexos, essencialmente devido à robustez e versatilidade do algoritmo.

Na química analítica tem-se utilizado o algoritmo genético para otimização de parâmetros internos (pesos) de rede neural⁽³⁸⁾. Uma de suas aplicações é o treinamento da rede para análise quantitativa de espectros de fluorescência de raios-X.⁽³⁹⁾ Outra aplicação bastante difundida é a seleção de comprimentos de onda em determinações espectroscópicas, principalmente no infravermelho próximo, utilizando calibração multivariada⁽⁴⁰⁾. Geralmente nestes casos utiliza-se o método dos mínimos quadrados parciais⁽⁴⁰⁾ ou a regressão linear múltipla⁽⁴¹⁾ para a construção de um modelo que possa correlacionar os espectros com as espécies de interesse. Este tem como objetivo determinar o número mínimo de comprimentos de onda que possuem informação relevante para a análise de interesse.

O algoritmo genético também tem sido utilizado no mapeamento de átomos, a fim de se determinar a mínima distância interatômica⁽⁴²⁾. Outro campo que está sendo explorado é a pesquisa com aplicação em planejamentos experimentais, que consiste no planejamento de n experimentos, onde há variação de alguns fatores para otimizar o sistema⁽⁴³⁾.

Muitos sistemas químicos apresentam um alto nível de complexidade, sendo muitas vezes necessário o uso de uma função para explicá-los. Para solucionar este tipo de problema utiliza-se uma poliotimização. Um exemplo de poliotimização em química analítica é a otimização de distâncias entre picos cromatográficos, como função das condições cromatográficas (composição da fase móvel, temperatura do injetor, valor de pH, etc.)⁽⁴⁴⁾. Pode-se encontrar outro exemplo de poliotimização na determinação de glicose em sangue humano por fotometria, baseado em algoritmo genético. A otimização envolve 12 parâmetros químicos e tecnológicos, com o intuito de monitorar seis propriedades ópticas. O resultado obtido foi comparado com o do método clássico de regressão dos mínimos quadrados parciais não lineares⁽⁴⁴⁾. Encontram-se ainda exemplos da aplicação em determinação da constante de estabilidade de sistemas de cloreto de cádmio em meio contendo perclorado de chumbo e éter coroa diclorohexil-18-coroa-6, a partir de dados calorimétricos e polarográficos⁽⁴⁰⁾ e na seleção e detecção de amostras anômalas dentre um conjunto de amostras, para construção do modelo de calibração.⁽⁴⁵⁻⁴⁶⁾ Também encontram-se exemplos na otimização da separação de bases fracas ou soluções ácidas por cromatografia líquida de fase reversa, com

variação de pH e modificação orgânica na fase móvel⁽⁴⁷⁾, na otimização da estimativa dos parâmetros cinéticos de uma reação⁽⁴⁸⁾ e para a otimização da intensidade de linhas de emissão atômica de traços metálicos em pó de alumina como uma função da condição de excitação⁽⁴⁹⁾, e na análise conformacional de biopolímeros.⁽⁵⁰⁾

2.7 Algoritmo Genético na seleção de variáveis

Tem-se observado nos últimos anos uma tendência na implementação de técnicas espectroscópicas em análises químicas rotineiras. Isto se deve essencialmente à simplicidade, rapidez, baixo custo, versatilidade, bem como outros atributos de diferenciação proporcionados pela técnica.

A calibração multivariada contribuiu de forma significativa para a consolidação desta técnica, viabilizando a determinação simultânea de mais de uma espécie, mesmo em situações adversas, onde os métodos convencionais de análises seriam difíceis ou impossíveis de serem aplicados (por exemplo, determinações “on-line” em processos).

Em análises espectrofotométricas, muitas das aplicações de calibração multivariada envolvem um elevado número de variáveis. Muitas vezes, o uso de somente algumas das variáveis que carregam mais informações podem proporcionar maior segurança e facilidade na interpretação do modelo. Assim, as variáveis (como absorbância em diferentes comprimentos de onda) que possuem uma menor relação sinal/ruído e/ou não linearidades podem ser descartadas.

Atualmente, o número de algoritmos que se propõem a selecionarem variáveis é pouco expressivo. Dentre estes, se destacam a regressão linear múltipla por passos (*stepwise multiple linear regression*)⁽²³⁾ e os que empregam o uso do algoritmo genético.

A implementação do algoritmo genético na seleção de variáveis difere das demais aplicações no que tange à codificação do problema e à função de resposta, já que as outras etapas permanecem inalteradas.

Na codificação do problema em seleção de variáveis de espectros, considera-se que o cromossomo possui um número definido de alelos, onde cada alelo representa uma das variáveis do espectro (absorbância em cada comprimentos de onda). Assim sendo, o cromossomo terá o mesmo número de variáveis contidas no espectro, conforme representada na Figura 9.

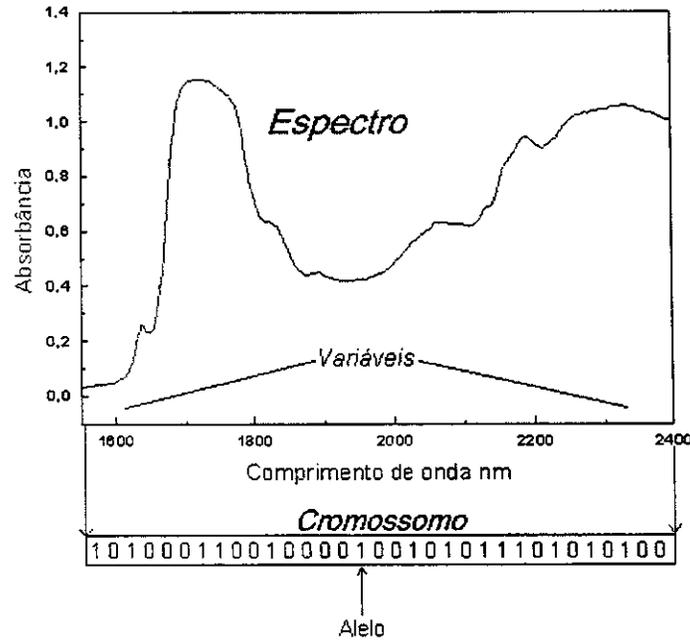


Figura 9. Codificação de um espectro na forma de um cromossomo.

Quando se utiliza o algoritmo genético para a seleção de variáveis, é necessário, dependendo do modelo de calibração, utilizar algumas “condições de contorno” como, por exemplo, pré-estabelecer o número máximo de variáveis selecionadas, caso da regressão linear múltipla.

Na seleção de variáveis utiliza-se também o auxílio do código binário (0,1) para codificar o problema. Cada alelo pode assumir o valor *um* ou *zero*, sendo que quando a posição referente a uma determinada variável for igual a *um*, isto implica na seleção da variável. Por outro lado, se a posição contiver o valor *zero*, isto indica que a variável não foi selecionada.

3. Introdução

3.1 Redes Neurais

As redes neurais são fruto da tentativa dos cientistas de compreenderem melhor o funcionamento do cérebro humano. O surgimento desta ferramenta matemática nasceu do interesse mútuo de pesquisadores de diversas áreas, como psicologia, neurofisiologia, computação e engenharia⁽⁵¹⁾. Entretanto, apesar de seu caráter multidisciplinar, cada uma das áreas que estuda o tema rede neural artificial, desenvolveu-se praticamente independentemente, devido ao fato de possuírem abordagens bem distintas. Por exemplo, os neurofisiologistas desejam desvendar os segredos dos estímulos nervosos dos neurônios, enquanto os psicólogos procuram criar modelos que explicam as estruturas cognitivas responsáveis pelo comportamento dos seres humanos.

Apesar do estudo do funcionamento do cérebro humano e da idéia de construção de uma máquina que realize tarefa de forma inteligente seja bem antiga, somente no início da década de quarenta, um grupo de cientistas deu início ao projeto de desenvolvimento de uma máquina que raciocinasse de forma semelhante aos seres humanos. O objetivo deste trabalho era basicamente viabilizar o desenvolvimento de dispositivos eletrônicos que pudessem essencialmente desempenhar o mesmo papel de seres humanos nas tomadas de decisões, processamento de informações, otimização, etc⁽⁵¹⁾. Entretanto, após mais de sessenta anos de pesquisa, é um consenso que este objetivo ainda está longe de ser atingido. Contudo, vários avanços significativos foram realizados neste campo.

As redes neurais basicamente podem ser classificadas em dois tipos distintos: as redes supervisionadas e não supervisionadas. As redes com treinamento supervisionado se diferenciam das redes não supervisionadas, pelo motivo das primeiras necessitarem de um conjunto inicial de padrões para a realização do treinamento e para estimar valores das propriedades de interesse, enquanto as redes neurais não supervisionadas são destinadas à área de classificação de padrões. Entre as redes neurais não supervisionadas existem inúmeros métodos de treinamento. Entre os quais podemos citar: teoria de ressonância adaptativa (ART1); memória associativa bidirecional (BAM); memória associativa temporal (TAM); mapas auto-organizáveis de Kohonen (SOM); aprendizado competitivo; memória associativa linear, etc⁽³⁸⁾. As redes neurais supervisionadas possuem um grande e variado número de métodos de treinamento; entre os principais, podemos citar: máquina de boltzmann (BM); quantização vetorial por aprendizagem (LVQ); rede lógica

adaptativa (ALN); rede neural probabilística (PNN); retropropagação (BP), rede de função de base radial (RBF), etc⁽³⁸⁾.

As redes neurais têm apresentado uma contribuição significativa na análise e desenvolvimento de modelos não lineares. Esta técnica tem um forte caráter multidisciplinar, com ênfase em áreas como a estatística, engenharia, física, química, biologia, marketing, etc.

Em engenharia, particularmente, há inúmeras aplicações. Dentre estas podemos destacar: análise e processamento de sinais; controle de processos, robótica, classificação de dados, reconhecimento de padrões em linha de montagem, filtros contra ruídos eletrônicos, - análise de imagens, análise de voz, diagnóstico médico, - previsões de séries temporais, etc^(38,51). Isto é atribuído ao fato de ser uma das primeiras áreas a utilizar esta ferramenta matemática.

Particularmente, na área de química analítica, o número de aplicações das redes neurais tem crescido expressivamente nos últimos anos⁽⁵²⁾. As principais aplicações encontram-se na classificação de padrões e na determinação de propriedades de interesse químico e/ou físico. Para tanto, são usadas as informações provenientes de medidas físicas, como espectros no infravermelho próximo, espectros de ressonância magnética nuclear, espectros na região do visível e ultravioleta, espectros de massa, espectros de fluorescência de raios-X, cromatogramas, voltamogramas, etc.

A Tabela 1 mostra uma série de aplicações onde as redes neurais têm apresentado sucesso, tendo destaque maior em casos onde os sistemas têm comportamento não linear.

Tabela 1. Exemplos de aplicações de Redes Neurais em química analítica.

Propriedade modelada	Técnica de análise	Referência
Nitrogênio em trigo	espectros no infravermelho próximo	53
Propriedades de polímeros	espectros no infravermelho próximo	54
Teor de metanol e etanol	espectros no UV/VIS	55
Componentes em misturas binárias e ternárias	espectros no UV/VIS	56
Ingredientes ativos em fármacos	espectros no UV/VIS	57
Determinação de [Cu] e [Zn]	voltametria	58
Determinação de [Fe], [Ni] e [Cr]	fluorescência de raios-X	59
Número de octanagem em gasolina	cromatografia gasosa	60

O princípio básico de funcionamento de uma rede neural pode ser considerado relativamente simples, quando comparada a outras ferramentas matemáticas, como: transformada de wavelets, métodos de ordem superior como *PLS multi-way*⁽⁶¹⁾, *GRAM*⁽⁶²⁾, etc. Entretanto, tem

sido criado um misticismo ao redor desta técnica, devido ao fato do seu nome estar diretamente associado ao funcionamento do cérebro humano, além da associação realizada entre as unidades básicas das redes neurais e os elementos que são responsáveis pela capacidade do homem de aprender. Contudo, a concepção de que as redes neurais são uma ferramenta que opera igual ao cérebro humano não é completamente verdadeira. A complexidade de uma rede neural é muito inferior ao processo de aprendizagem e manipulação de informações existentes no cérebro humano. Basta lembrar que o cérebro possui aproximadamente 10^{11} neurônios, com aproximadamente 10^{13} sinapses, enquanto uma rede neural possui em média 10 - 20 neurônios, ou seja, é um modelo muito simplificado da complexa estrutura existente no cérebro humano. Além disso, o cérebro humano é capaz de manipular informações imprecisas e incertas, variáveis lingüísticas, emoções, etc. Nenhuma das redes neurais proposta até hoje foi capaz de apresentar estas habilidades.

As expectativas geradas por um milagroso desempenho das redes neurais têm criado frustração e inúmeras críticas de alguns grupos de pesquisa. Talvez este motivo explique a resistência encontrada para sua utilização como ferramenta de análise⁽⁶³⁾. É importante salientar que as redes neurais bem como outras técnicas matemáticas de calibração e classificação não são capazes de realizar milagres. Portanto, a melhor forma de lidar com estas limitações é ter um profundo conhecimento sobre o sistema químico e/ou físico que se deseja modelar.

De forma simples, podemos ver as redes neurais como sendo uma função de mapeamento, a priori não linear, de um espaço de variáveis de entradas (medidas instrumentais) para um espaço de variáveis de saídas (propriedades físicas e/ou químicas que se deseja monitorar). A função de mapeamento é composta por uma série de parâmetros que são ajustados de forma iterativa.

Uma rede neural é uma estrutura altamente organizada composta por uma série de unidades processadoras de sinais denominados de neurônios, os quais são conectados entre si. Cada conexão de um neurônio está associada a um sinal de entrada a ser processado pelo neurônio. Depois de processados os sinais de entrada, o neurônio gera um sinal que é transmitido para a saída da unidade processadora.

3.2 Elementos Básicos de uma Rede Neural

Em uma rede neural há três elementos básicos que a caracterizam. O primeiro elemento é o conjunto de pesos sinápticos - que possuem a finalidade de ponderar as informações transmitidas entre os neurônios, atribuindo diferentes pesos às informações processadas. O segundo termo trata

do nível de ativação, que é responsável pela somatória do produto entre os pesos sinápticos e as informações transmitidas pelos os neurônios. O nível de ativação pode ser representado pela seguinte expressão: $v = \sum_{k=1}^l x_k w_k - \theta$, onde a variável x representa a informação a ser transmitida, ou seja, a informação que entra no neurônio; w é o peso sináptico, ou seja, uma espécie de peso que irá ponderar a importância da informação fornecida ao neurônio e θ é o termo de polarização ou bias. Este último parâmetro tem a finalidade de deslocar a função no espaço de excursão. O terceiro elemento é a função de transferência, a qual determina se o estado de ativação irá assumir valores discretos ou contínuos. Para o caso discreto, geralmente utiliza-se a função de ativação binária do tipo degrau. No caso contínuo, as funções normalmente não são lineares e devem ser monotonicamente crescentes, deriváveis e assumirem valores em um intervalo finito do domínio dos números reais.

3.2.1 Tipos mais comuns de funções de transferência

Como vimos, praticamente podemos utilizar qualquer tipo de equação matemática como função de transferência, desde que esta satisfaça as condições apresentadas anteriormente. Entretanto, existem algumas funções que são mais utilizadas, principalmente pelo fato de suas derivadas serem bem conhecidas e fáceis de serem calculadas. Além disso, possuem um segmento linear, o que permite a função a flexibilidade de modelar dados lineares e não lineares. Entre as funções mais comuns de modelamento, podemos citar as funções linear, logística, tangente hiperbólica e arco tangente.

Função Linear: As funções lineares são as mais simples e com certeza as mais conhecidas. Uma função linear é composta de um coeficiente angular α que determina a inclinação da reta e um coeficiente linear b que tem a finalidade de transladar a função ao longo do eixo y .

Uma função linear em um espaço bidimensional, pode ser escrita da seguinte forma: $y = \alpha x + b$, onde seus parâmetros podem assumir qualquer valor no domínio dos números reais.

A derivada primeira da função linear fornece a inclinação da reta, conforme ilustrado na Figura 10 b.

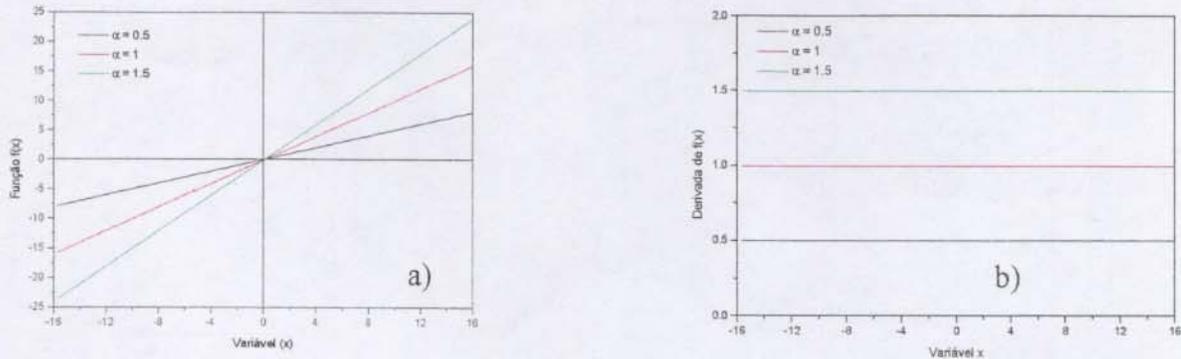


Figura 10. a) Função linear com três diferentes valores de coeficientes angulares. b) Primeira derivada das funções lineares em relação à x .

Função Logística ou sigmoidal: A função logística pode ser representada de forma genérica,

segundo a expressão,
$$y = \frac{1}{1 + e^{-ax}}$$

Esta função depende de dois parâmetros, o primeiro deles é o coeficiente a que determina a inclinação da função e a velocidade de convergência para o ponto de saturação da função. Como pode ser observado na Figura 11.a, quanto maior o valor do coeficiente a , mais rapidamente irá convergir para um valor constante. O segundo parâmetro é o valor que a variável x assume ao longo do eixo x . Estes dois parâmetros são suficientes para prever o perfil da equação em uma determinada região do espaço das variáveis. Além disso, esta função somente assume valores reais no intervalo entre 0 e 1 para a variável y .

A primeira derivada da função logística é expressa matematicamente por:
$$\frac{dy}{dx} = a \frac{1}{1 + e^{-ax}} \left(1 - \frac{1}{1 + e^{-ax}} \right)$$
. O gráfico desta função possui comportamento gaussiano, conforme pode ser observado na Figura 11.b. A variação no parâmetro a , proporciona uma variação na dispersão da função, ou seja, este parâmetro possui uma função análoga ao desvio padrão na função gaussiana.

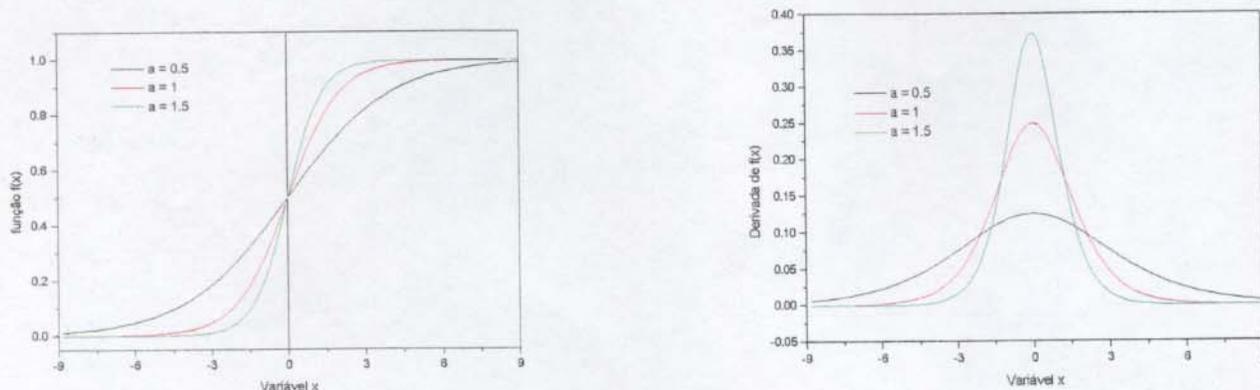


Figura 11. a) Função logística para três diferentes valores do parâmetro a . b) representa a primeira derivada da função logística em relação a x .

Tangente hiperbólica: A função tangente hiperbólica é uma composição de termos exponenciais que podem ser expressos conforme a equação: $y = \frac{e^{ax} - e^{-ax}}{e^{ax} + e^{-ax}}$.

Esta função possui um comportamento similar à função logística. Entretanto, difere-se da primeira devido aos valores que sua imagem pode assumir. No caso da função logística, a imagem assume valores no domínio dos reais positivos entre 0 e 1. Para a função tangente hiperbólica, a imagem pode assumir tanto valores negativos como positivos, conforme pode-se observar na Figura 12.a, mas entre $-1,0$ e $+1,0$.

Semelhante a função logística, o parâmetro a possui influência na convergência para a região de saturação da função e, conseqüentemente, na inclinação da curva, conforme pode ser verificado na Figura 12.a.

A primeira derivada da função tangente hiperbólica é dada pela seguinte equação: $\frac{dy}{dx} = a \left(1 - \left(\frac{e^{ax} - e^{-ax}}{e^{ax} + e^{-ax}} \right)^2 \right)$. O gráfico da função possui um comportamento similar ao da derivada da

função logística, diferenciando-se principalmente, na magnitude do valor que y pode assumir para o mesmo valor de x , conforme pode ser observado ao compararmos as Figuras 12.b e 11.b.

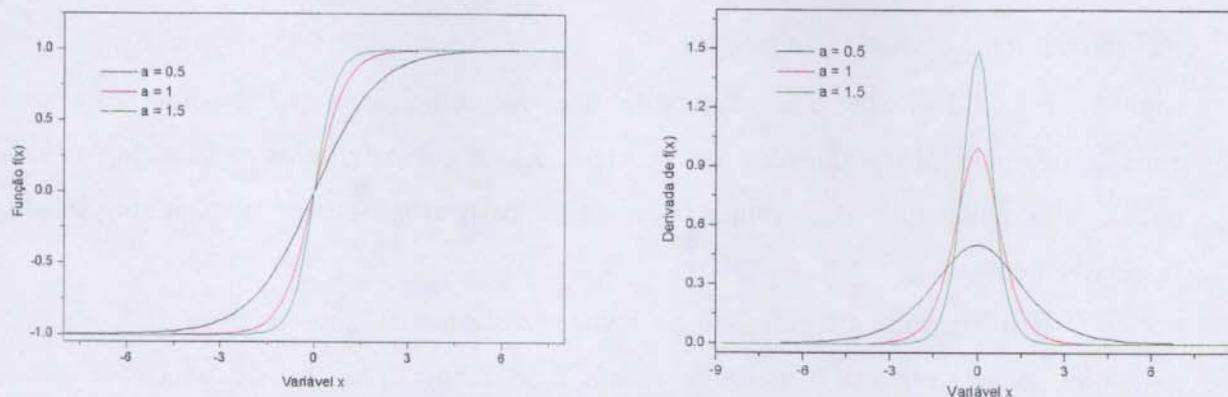


Figura 12. a) Função tangente hiperbólica com três diferentes valores do parâmetro a . b) Primeira derivada da função tangente hiperbólica em relação a x .

Função Arco Tangente: A função arco tangente que pode ser expressa como $y = \arctan(ax)$, tem um comportamento similar às funções anteriormente apresentadas. A faixa dinâmica da função de transferência encontra-se no intervalo compreendido entre $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. O perfil da função arco tangente pode ser observado na Figura 13.a. Esta função possui como característica principal uma taxa de convergência mais lenta para a faixa de saturação, quando comparada com as funções anteriores.

A derivada da função arco tangente é dada por: $\frac{dy}{dx} = a \frac{1}{1+a^2x^2}$. A derivada desta função

tem um comportamento similar ao inverso de uma função quadrática, assumindo o valor máximo para x igual a zero. O parâmetro a influencia diretamente na intensidade da função.

A grande vantagem desta função é a facilidade de sua implementação e de processamento dos cálculos.

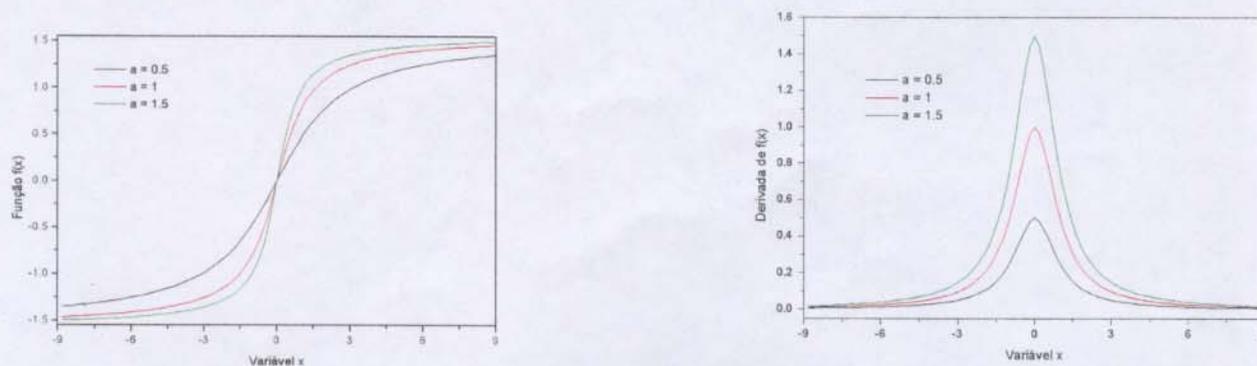


Figura 13. a) Função arco tangente para três diferentes valores do parâmetro a . b) Primeira derivada da função arco tangente com relação a x .

3.3 Estrutura de uma rede neural

A Figura 14 ilustra a arquitetura de uma rede neural, com quatro neurônios e um bias de entrada, três neurônios e um *bias* na camada interna e dois neurônios na camada de saída. Cada círculo e as linhas que os conectam representam os neurônios e as ligações sinápticas (pesos), respectivamente.

O termo arquitetura refere-se ao número de camadas que compõem a rede, número de neurônios de cada camada, os pesos de cada ligação sináptica, os tipos de funções de transferência usadas e o algoritmo de treinamento da rede neural.

Uma rede neural é composta de no mínimo três camadas, uma camada de entrada, uma camada interna e uma camada de saída. O número de camadas internas pode variar; entretanto, normalmente utiliza-se apenas uma camada interna. Quando se introduz novas camadas internas, a etapa de treinamento demanda um maior tempo de processamento. Isso ocorre porque o número de pesos de treinamento cresce significativamente, além de tornar o processo de ajustes dos pesos sinápticos mais complexo.

Segundo vários autores, o uso de redes neurais com mais de uma camada interna só se justifica para casos específicos onde a relação entre as variáveis independentes e dependentes tem um tipo de descontinuidade⁽⁶³⁾. Vários estudos demonstram que com apenas uma camada interna e com funções de transferências não lineares, uma rede neural já apresenta a propriedade de ser um aproximador universal.

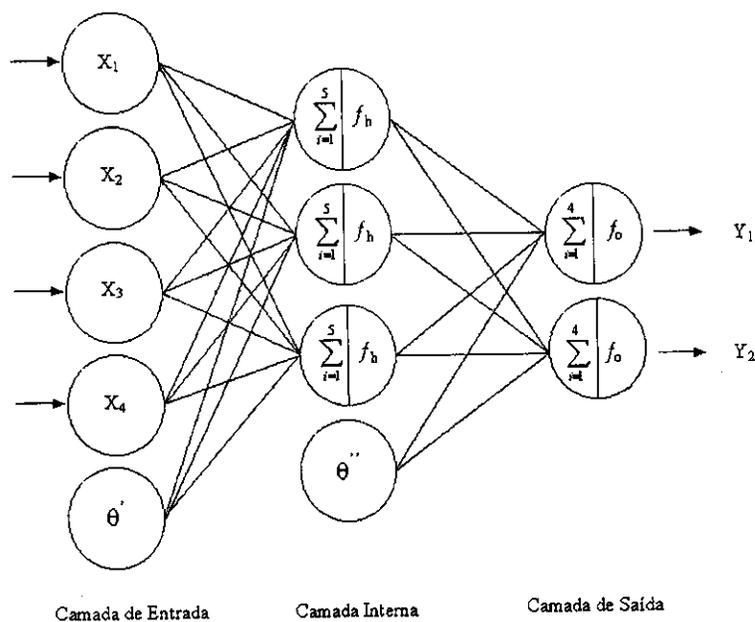


Figura 14. Estrutura de uma rede neural com uma camada interna.

Podemos imaginar que o número de camadas e neurônios de uma rede neural possui a mesma função do número de termos de um polinômio necessários para aproximar uma dada função $P(x)$. Ou seja, como nos casos dos polinômios, redes com um número pequeno de parâmetros podem não ser capazes de aproximar com precisão suficiente os dados de entrada e saída. Em contrapartida, uma rede neural com um elevado número de parâmetros pode não conseguir generalizar com precisão suficiente, devido a um sobre-ajuste da função para um conjunto específico de dados de entrada e saída. Este sobre-ajuste é mais crítico quando os dados possuem um alto grau de incerteza em suas variáveis de modelamento, devido a ruídos instrumentais ou limitações da precisão dos dados obtidos pelo método padrão de análise, ou seja, imprecisão na variável dependente y .

Uma rede neural com funções de transferência linear é equivalente à regressão linear múltipla. Se trabalharmos com dados com comportamento linear, é possível observar que os resultados obtidos convergem para os obtidos pelo método da regressão linear múltipla. A única diferença entre os dois métodos é a forma como os parâmetros são estimados. A regressão linear utiliza um método exato que consiste na inversão da matriz de variância-covariância, enquanto a rede neural aplica um método iterativo. Entretanto, a rede neural provavelmente irá gastar um maior tempo de processamento, pois o seu modelo parte do pressuposto que os dados possuem um forte caráter não linear e apenas uma sutil contribuição linear, o que dificulta o modelamento dos dados com comportamento linear.

Quando temos apenas uma camada interna e mais de uma saída de previsão, podemos utilizar a equação abaixo para estimar os valores de saída de uma rede neural:

$$\hat{y}_m = f_o \left[\theta'' + \sum_{j=1}^{nh} w_j^2 f_h \left(\sum_{i=1}^{nd} w_{ij}^1 x_i + \theta' \right) \right] \quad \text{eq. (19)}$$

onde \hat{y}_m indica a saída estimada pela rede, f_o e f_h representam as funções de transferência selecionadas, para a camada intermediária e de saída, respectivamente; θ'' e θ' são os valores dos bias ou termo de polarização; nd e nh são os números de variáveis de entrada e da camada interna respectivamente. Como já sabemos, os valores de w representam pesos e os valores de x as variáveis de entrada.

3.4 Princípio de processamento das Redes Neurais

Para facilitar a compreensão de cada uma das etapas que compreendem a fase de treinamento e previsão de uma rede neural, será apresentado um exemplo que permita ao leitor acompanhar todas as etapas envolvidas no treinamento de uma rede neural.

O primeiro passo é gerar valores aleatórios para cada um dos pesos da rede neural. Entretanto, porque devemos começar com uma estrutura com valores de pesos escolhidos de forma aleatória? Devido ao fato de não sabermos onde se encontra a melhor solução. O valor inicial dos pesos faz com que o processo de minimização dos erros inicie em uma região arbitrária do espaço da superfície de erro.

Apesar de desejarmos otimizar uma superfície de erros, não temos a informação do valor do ponto no mínimo global da função. Ou seja, não sabemos se durante o processo de otimização foi encontrado o ponto de mínimo global ou local. Para sistemas reais, é impossível afirmar que o valor obtido trata-se de um mínimo ou máximo global. Entretanto, se os pesos da rede neural forem calculados várias vezes, partindo de pontos diferentes da superfície de erro, é possível ter uma boa idéia se o ponto encontrado é um mínimo local. Contudo, vale ressaltar que esta técnica não pode ser usada para afirmar que realmente foi encontrado um ponto de mínimo ou máximo global. Na prática o processo de otimização é cessado quando a função de erro atingir um valor que satisfaça as necessidades do sistema estudado.

O próximo passo é efetuar as operações matemáticas necessárias para o treinamento, previsão das saídas e ajuste dos pesos de uma rede neural. Para a realização dos cálculos que serão apresentados a seguir, os pesos serão representados por W^c_{ij} , onde a letra W é a abreviatura da palavra *weight* (peso), c indica qual camada pertence o respectivo peso, ij representa que o peso é referente conexão do i -ésimo elemento da c -ésima camada com o j -ésimo elemento da $(c+1)$ -ésima camada. Ou seja, W^1_{11} representa um peso da primeira camada que está realizando uma conexão sináptica entre o primeiro neurônio da primeira camada e o primeiro neurônio da segunda camada. Este tipo de notação será de grande ajuda ao leitor, para que esse possa acompanhar o raciocínio inicial. Entretanto, trata-se de uma notação muito carregada e, por este motivo, alguns autores preferem omitir o parâmetro c . Inicialmente, iremos utilizar este parâmetro, posteriormente, ele será omitido, para simplificar a notação.

A Tabela 2 mostra o conjunto de pesos iniciais utilizados para o processo de treinamento de uma rede neural com uma arquitetura semelhante à da Figura 14.

Tabela 2. Conjunto de pesos obtidos por um gerador de números pseudo-aleatórios.

<i>Neurônios da Primeira Camada</i>	<i>Neurônios da Segunda Camada</i>
$W^1_{11} = 0,5$	$W^2_{11} = 0,37$
$W^1_{12} = -0,56$	$W^2_{12} = 0,65$
$W^1_{13} = 0,12$	$W^2_{21} = 0,34$
$W^1_{21} = 0,32$	$W^2_{22} = 0,25$
$W^1_{22} = -0,17$	$W^2_{31} = 0,71$
$W^1_{23} = -0,82$	$W^2_{32} = 0,46$
$W^1_{31} = -0,41$	$W^2_{\theta''1} = 1$
$W^1_{32} = 0,82$	$W^2_{\theta''2} = 1$
$W^1_{33} = 0,67$	
$W^1_{41} = 0,55$	
$W^1_{42} = 0,58$	
$W^1_{43} = 0,90$	
$W^1_{\theta'1} = 1$	
$W^1_{\theta'2} = 1$	
$W^1_{\theta'3} = 1$	

Para iniciar o processo de treinamento da rede neural, é necessário que haja um conjunto de dados de treinamento que consiste em uma tabela contendo valores de medidas físicas e/ou químicas e valores associados a estas medidas que se deseja prever. Para o treinamento da rede neural, será utilizado um conjunto de dados hipotéticos que estão apresentados na Tabela 3.

Tabela 3. Conjunto de dados para o treinamento de uma rede neural com quatro entradas e duas saídas.

Entrada 1	Entrada 2	Entrada 3	Entrada 4	Saída 1	Saída 2
0,03	0,09	0,36	0,01	0,86	0,35
0,18	0,26	0,06	0,32	0,55	0,14
0,85	0,65	0,01	0,24	0,47	0,81
0,08	0,95	0,62	0,67	0,36	0,77
0,23	0,57	0,05	0,11	0,92	0,09

3.5 Pré-processamento dos dados de entrada da Rede

Para que o processo de treinamento da rede inicie, deve-se primeiro realizar um pré-processamento dos dados, com o objetivo de reduzir a dimensionalidade destes.

Quando se trabalha com um extenso número de dados, é normalmente empregada a redução de sua dimensionalidade, para eliminar informações redundantes ou sem correlação com a propriedade que se deseja analisar. Uma técnica muito conhecida e utilizada em quimiometria é a análise das componentes principais (PCA). Entretanto, existem outras ferramentas que podem ser empregadas na redução da dimensionalidade dos dados de entradas de uma rede neural como: coeficientes de *wavelets* e coeficientes da transformada de Fourier.

O interesse da redução da dimensionalidade não é simplesmente para que possamos visualizar tendências do conjunto de dados, mas também aumentar a velocidade de treinamento da rede, para tornar viável a utilização de um processador mais simples, melhorar a habilidade de generalização e a robustez do modelo. Portanto, a aplicação de redes neurais em dados espectroscópicos geralmente está associada à redução de dimensionalidade, através do auxílio de umas das técnicas citadas anteriormente. Neste caso, as entradas utilizadas para o treinamento e previsão de novas amostras são os valores dos escores obtidos na análise de componentes principais. No caso de *wavelets* e Fourier, são empregados os seus respectivos coeficientes.

3.5.1 Escalamento dos Dados de Treinamento da Rede Neural

A Tabela 3 contém um conjunto de valores de entrada e saída que serão empregados no exemplo do treinamento da rede neural. Entretanto, antes que possamos realizar o treinamento da rede, os valores de entrada e saída devem ser escalados, de forma que estejam na faixa dinâmica das funções de transferência. A faixa dinâmica é a região onde uma variação em x proporciona uma variação no valor da função. Fora da faixa dinâmica, a função de transferência encontra-se na região de saturação, ou seja, o valor da função $f(x)$ é constante e independente do valor x . Portanto, nesta situação, o tempo de aprendizagem tende a infinito, pois a saída da rede tende a um valor constante, o qual não fornece informação para que a atualização dos pesos se processe de forma a minimizar o erro de previsão do conjunto de dados de treinamento.

O escalamento dos dados é realizado pela seguinte equação matemática:

$$x_i^{esc.} = \frac{(x_i - x_{min.})}{x_{max.} - x_{min.}} (l_{max.} - l_{min.}) + l_{min.} \quad \text{eq. (20)}$$

onde $x_i^{esc.}$ é o elemento i do vetor de dados x escalado; x_i é o elemento i do vetor x ; $x_{min.}$ e $x_{max.}$, representam o menor e maior valor do vetor de dados x , respectivamente; $l_{max.}$ e $l_{min.}$ representam o valor máximo e mínimo que o vetor de dados escalados pode assumir, respectivamente.

A Tabela 4 mostra os valores escalados do conjunto de dados de entrada e saída da rede neural da Tabela 3, entre o intervalo de 0,2 a 0,8 (unidades arbitrárias). O intervalo de 0,2 a 0,8 foi escolhido por ser a faixa dinâmica da função sigmoideal. Para outras funções, deve-se escolher outro intervalo, por exemplo, para a função tangente hiperbólica, é aconselhado utilizar a faixa entre -0,8 a +0,8.

Tabela 4. Conjunto de dados escalados para o treinamento de uma rede neural com quatro entradas e duas saídas.

Entrada 1	Entrada 2	Entrada 3	Entrada 4	Saída 1	Saída 2
0,20	0,20	0,54	0,20	0,74	0,42
0,31	0,32	0,25	0,48	0,40	0,24
0,80	0,59	0,20	0,41	0,32	0,80
0,24	0,80	0,80	0,80	0,20	0,77
0,35	0,53	0,24	0,29	0,80	0,20

Neste exemplo ilustrativo, iremos utilizar a função de transferência do tipo logarítmica sigmoideal ou logística na camada intermediária de neurônios, conforme apresentada na equação 21. A equação 22 fornece o nível de ativação do neurônio. Para os neurônios da camada de saída, usaremos a função linear identidade.

$$y = \frac{1}{1 + e^{-\left(\sum_{k=1}^l x_k w_k - \theta\right)}} \quad \text{eq. (21)}$$

onde y representa o valor de saída da função; x representa o vetor de dados a ser processado durante o treinamento, w o conjunto de pesos que conecta duas camadas adjacentes e θ o parâmetro de polarização da função, mais conhecidos como bias. Neste exemplo consideramos que o parâmetro θ tem valor fixo igual a zero.

$$v = \sum_{k=1}^l x_k w_k - \theta \quad \text{eq. (22)}$$

onde v indica o nível de ativação da saída do neurônio.

O primeiro passo consiste em calcular o valor de saída da primeira camada para o primeiro conjunto composto de quatro valores de entrada ($x_1=0,20$; $x_2=0,20$; $x_3=0,54$ e $x_4=0,20$). Para simplificar a notação matemática, será omitida a informação referente ao número da iteração. A seguinte notação, y_1^2 , deve ser lida como saída da estimativa de y do primeiro neurônio da segunda camada (camada intermediária).

$$y_1^2 = \frac{1}{1 + e^{-[(0,5 * 0,20) + (0,32 * 0,20) + (-0,41 * 0,54) + (0,55 * 0,20) - 0]}} = 0,5131$$

$$y_2^2 = \frac{1}{1 + e^{-[(-0,56 * 0,20) + (-0,17 * 0,20) + (0,82 * 0,54) + (0,58 * 0,20) - 0]}} = 0,6540$$

$$y_3^2 = \frac{1}{1 + e^{-[(0,12 * 0,20) + (-0,82 * 0,20) + (0,67 * 0,54) + (0,9 * 0,20) - 0]}} = 0,5991$$

Para calcular a saída da rede neural, utilizam-se os valores das saídas da camada anterior como entrada para os neurônios da camada de saída. A saída dos neurônios da camada de saída será dada por:

$$y_1^3 = [(0,37 * 0,5131) + (0,34 * 0,6540) + (0,71 * 0,5991) - 0] = 0,8376$$

$$y_2^3 = [(0,65 * 0,5131) + (0,25 * 0,6540) + (0,46 * 0,5991) - 0] = 0,7726$$

Uma vez que para os neurônios da saída, utiliza-se a função linear identidade.

O passo seguinte é determinar o erro de previsão da rede para cada uma das saídas, com o objetivo de realizar as devidas correções nas matrizes de pesos.

O erro na saída da rede neural é calculado subtraindo-se o valor previsto pela rede do valor real associado ao conjunto de dados de entrada. O erro associado às saídas y_1 e y_2 será calculado conforme apresentado a seguir:

$$e_{n_j} = s_j - y_{n_j} \tag{23}$$

onde s_j é o valor do padrão associada ao neurônio da saída j e y_{n_j} é o valor gerado pela rede no neurônio da saída j na iteração n .

Para o primeiro padrão apresentado à rede neural, o erro associado às saídas y_1 e y_2 são $-0,0976$ e $-0,3526$, respectivamente.

Se os erros obtidos forem considerados elevados é preciso iniciar o processo de correção dos pesos. A correção dos pesos pode ser realizada basicamente por dois diferentes critérios: o primeiro critério consiste na atualização dos pesos da rede após a apresentação de cada um dos padrões (chamado de padrão-a-padrão); o segundo é conhecido por batelada, pois a correção dos pesos somente é realizada após todos os padrões serem processados pela rede. No final deste

procedimento, é calculado um erro quadrático médio dos padrões de treinamento, o qual é utilizado para estabelecer a correção dos pesos sinápticos da rede.

Comparando-se os dois métodos de treinamento, padrão-a-padrão e por batelada ou lote, verifica-se que o método padrão-a-padrão apresenta vantagens como⁽⁵¹⁾:

1. Demanda menos memória do computador, pois este método processa cada conjunto de informação dos padrões de entrada isoladamente. Conseqüentemente, o número de variáveis temporárias para armazenar e processar as informações dos dados é menor que as equivalentes para o método em batelada. Isso implica diretamente na velocidade de processamento.
2. O método padrão-a-padrão confere ao algoritmo de treinamento maior flexibilidade para escapar de mínimos e máximos local, pois atribui uma característica estocástica ao processo de otimização. Se inicialmente a primeira amostra indica que o processo de otimização deve seguir em uma determinada direção, a segunda amostra pode indicar que a melhor direção é em outra região. Portanto, inicialmente o processo de otimização dos pesos, segue um comportamento pseudo-aleatório, que lentamente tende a uma direção preferencial.

O método padrão a padrão apresenta a desvantagem de demandar maior tempo de processamento até que convirja para um valor de ótimo local ou global.

3.6 Treinamento de uma rede neural

Para a atualização dos pesos de uma rede neural, é necessário usar um algoritmo de treinamento. Entre os principais algoritmos utilizados, podemos destacar⁽³⁷⁾: método do passo descendente, método do gradiente conjugado, retropropagação de erro, método de Gauss-Newton, método de Newton, método de Levenberg-Marquardt e algoritmo genético.

Para ilustrar o processo de atualização dos pesos, será utilizado o método de retropropagação dos erros, por se tratar do método de treinamento mais difundido e fácil de ser ilustrado.

Os pesos da rede neural são atualizados a partir da informação do erro da saída da rede. Para calcular o erro na saída da rede neural, é utilizado o erro quadrático médio, por ser uma medida mais sensível a variações de erro. Esta média atribui um peso maior às amostras com erro maior, minimizando desta forma a influência que o número de padrões de treinamento e saídas

possam exercer sobre o resultado final. O erro quadrático do j -ésimo neurônio na n -ésima iteração será definido como:

$$J_n = \frac{1}{2} e_{nj}^2 \quad \text{eq. (24)}$$

Como pode existir mais de neurônio na camada de saída, deve-se realizar a otimização levando em consideração o erro de todos os neurônios da camada de saída. Portanto, é necessário realizar uma soma dos erros quadrados médios dos k neurônios existentes na camada de saída, que é representado por: $J_n = \frac{1}{2} \sum_{j=1}^k e_{nj}^2$. É importante lembrar que a saída y da rede neural é função dos pesos sinápticos e, portanto, a soma do erro quadrático médio é uma função de todos os pesos da rede. Assim sendo, pode-se escrever a soma dos erros quadrados médio sobre todos os k neurônios da saída como :

$$J_n(w) = \frac{1}{2} \sum_{j=1}^k (s_{nj} - y_{nj}(w))^2 \quad \text{eq. (25)}$$

Desta maneira obtém-se a função de erro que deve ser otimizada com relação aos pesos para produzir o menor erro possível de previsão. Entretanto, não existe uma solução de forma fechada para este problema de otimização, o que torna necessário recorrer a um processo iterativo de ajuste dos pesos.

Como neste problema deseja-se otimizar o erro, então deve-se calcular o vetor gradiente de erro em relação aos pesos da rede. Isso é representado pela seguinte expressão: $\frac{\partial J_n(w)}{\partial w_{nj}}$. O vetor

de gradiente de erro indica a direção em que os pesos devem ser ajustados para otimizar a função de erro. Contudo, é importante lembrar que se deseja minimizar a função de erro. Portanto, neste caso, deve-se caminhar na direção oposta do gradiente. A nova direção do vetor dos pesos é calculada partindo dos pesos anteriores e do vetor de gradiente. Matematicamente, a regra de

ajuste dos pesos pode ser expressa como: $w_{n+1,ji} = w_{n,ji} - \eta \frac{\partial J_n(w)}{\partial w_{n,ji}}$ eq. (26)

onde: η_n é o passo do algoritmo da iteração n . No algoritmo de retropropagação original, é

utilizado um passo constante durante todas as iterações do algoritmo, passo este conhecido como taxa de aprendizagem.

Para calcular o vetor gradiente de erro, é preciso saber o erro associado a cada neurônio. Para os neurônios da camada de saída, o erro é associado individualmente a cada neurônio, enquanto na camada escondida, o erro está distribuído entre vários neurônios. Para contornar este problema, o gradiente da função erro é calculado em relação à saída de cada neurônio, como apresentado a seguir:

$$J_n = \frac{1}{2} \sum_{k=1}^l e_{nk}^2 \quad \text{eq. (27)}$$

$$\frac{\partial J_n(\mathbf{w})}{\partial y_{n_j}} = \sum_{k=1}^l e_{nk} \cdot \frac{\partial e_{nk}}{\partial y_{n_j}} \quad \text{eq. (28)}$$

Para calcular o termo $\frac{\partial e_{nk}}{\partial y_{n_j}}$, deve-se utilizar a regra da cadeia, obtendo a seguinte equação:

$$\frac{\partial J_n(w)}{\partial y_{n_j}} = \sum_{k=1}^l e_{nk} \cdot \frac{\partial e_{nk}}{\partial v_{n_k}} \cdot \frac{\partial v_{n_k}}{\partial y_{n_j}} \quad \text{eq. (29)}$$

Para continuar a demonstração, é necessário definir o erro na saída do k-ésimo neurônio como: $e_{n_k} = s_k - y_{n_k}$ ou $e_{n_k} = s_k - f_k(v_{n_k})$, onde v é o nível de ativação do neurônio k na n -ésima iteração.

Partindo da equação de erro da saída do k-ésimo neurônio, calcula-se a derivada parcial do erro em relação ao nível de ativação para o k-ésimo neurônio.

$$\frac{\partial e_{n_k}}{\partial v_{n_k}} = -f'_k(v_{n_k}) \quad \text{eq. (30)}$$

O próximo passo será calcular a derivada parcial do nível de ativação em relação à saída do j-ésimo neurônio, obtendo as seguintes equações:

$$v_{n_k} = \sum_{j=0}^q w_{n_{kj}} y_{n_j} \quad \text{eq. (31)}$$

$$\frac{\partial v_{n_k}}{\partial y_{n_j}} = w_{n_{kj}} \quad \text{eq. (32)}$$

Partindo das equações desenvolvidas até o momento, obtém-se uma nova equação do gradiente de erro, a qual é expressa matematicamente por:

$$\frac{\partial J_n(\mathbf{w})}{\partial y_{n_j}} = - \sum_{k=1}^l e_{nk} \cdot f'_k(v_{n_k}) \cdot w_{n_{kj}} \quad \text{eq. (33)}$$

3.6.1 Atualização dos valores dos pesos sinápticos

Para a atualização dos pesos, iremos adotar a seguinte regra $\mathbf{w}_{n+1} = \mathbf{w}_n - \eta \nabla J_n(\mathbf{w}_n)$. Como podemos observar esta expressão matemática é composta de três termos já apresentados ao leitor. Neste momento, vale ressaltar que não trabalhamos mais com escalares e sim com vetores. Portanto, neste momento temos um vetor de pesos representado por \mathbf{w} ; além disso, iremos utilizar um vetor de derivadas parciais do erro em relação aos pesos, o qual é representado por $\nabla J_n(\mathbf{w}_n)$. Quando temos uma matriz composta de vetores gradiente, esta recebe o nome especial de matriz Jacobiana.

Como sabemos, um vetor é caracterizado por possuir direção e magnitude, portanto, o vetor de pesos \mathbf{w}_n está direcionado no espaço. A direção espacial do vetor é que irá determinar se iremos obter um bom resultado ou não durante a previsão do valor da propriedade de interesse. O nosso objetivo neste momento é reorientar a direção do vetor de pesos \mathbf{w}_n , de forma que este apresente o melhor desempenho possível de previsão para a rede neural. Para que isso seja possível, iremos utilizar a informação do vetor gradiente ($\nabla J_n(\mathbf{w}_n)$) para redirecionar o vetor \mathbf{w} na direção que minimiza o vetor erro de previsão. Ou seja, neste momento estamos realizando uma mudança na direção do vetor \mathbf{w} .

Além disso, a variável η que é denominada taxa de aprendizagem, tem a finalidade de atribuir um maior ou menor peso ao vetor gradiente, orientando para que a otimização caminhe mais ou menos rapidamente em uma determinada direção. Para finalizar o processo, somamos o vetor gradiente ao vetor \mathbf{w}_n e, como consequência obtém a nova direção dos pesos \mathbf{w}_{n+1} na iteração $n+1$.

Neste momento, podemos dar continuidade em nosso exemplo prático. Para tanto, iremos primeiro calcular o vetor gradiente, usando a seguinte expressão:

$$\frac{\partial J_n(\mathbf{w})}{\partial y_{n_j}} = - \frac{1}{1 + e^{-x}} \left(1 - \frac{1}{1 + e^{-x}} \right) \sum_{k=1}^l e_{nk} \cdot w_{n_{kj}}, \text{ para o ajuste dos pesos.}$$

3.6.2 Cálculo do vetor gradiente de erro da camada interna

$$\frac{\partial J_1(w)}{\partial y_1} = -0,5131 * (1 - 0,5131) * [(0,37 * 0,8376) + (0,65 * 0,7726)] = -0,2029$$

$$\frac{\partial J_1(w)}{\partial y_2} = -0,6540 * (1 - 0,6540) * [(0,34 * 0,8376) + (0,25 * 0,7726)] = -0,1081$$

$$\frac{\partial J_1(w)}{\partial y_3} = -0,5991 * (1 - 0,5991) * [(0,71 * 0,8376) + (0,46 * 0,7726)] = -0,2282$$

3.6.2.1 Cálculo da atualização dos pesos

Para a atualização dos pesos, será aplicada a seguinte regra: $w_{n+1} = w_n - \eta \nabla J_n(w_n)$, onde a taxa de aprendizagem η assumirá um valor fixo de 0,1.

O vetor de gradiente de erro associado à camada de saída será igual ao erro da saída, que calculamos anteriormente.

A seguir, mostraremos a atualização do conjunto de pesos sinápticos que conecta a camada de saída à camada interna, após serem atualizados:

$$w_{11}^2 = 0,37 - [0,1 * (-0,0976)] = 0,3798$$

$$w_{12}^2 = 0,65 - [0,1 * (-0,3526)] = 0,6853$$

$$w_{21}^2 = 0,34 - [0,1 * (-0,0976)] = 0,3498$$

$$w_{22}^2 = 0,25 - [0,1 * (-0,3526)] = 0,2853$$

$$w_{31}^2 = 0,71 - [0,1 * (-0,0976)] = 0,7198$$

$$w_{32}^2 = 0,46 - [0,1 * (-0,3526)] = 0,4953$$

A seguir, mostraremos a atualização do conjunto de pesos sinápticos que conecta a camada de entrada a camada interna, após serem atualizados:

$$w_{11}^1 = 0,50 - [0,1 * (-0,2029)] = 0,5203$$

$$w_{12}^1 = -0,56 - [0,1 * (-0,1081)] = -0,5492$$

$$w_{13}^1 = 0,12 - [0,1 * (-0,2282)] = 0,1428$$

$$w_{21}^1 = 0,32 - [0,1 * (-0,2029)] = 0,3403$$

$$w_{22}^1 = -0,17 - [0,1 * (-0,1081)] = -0,1592$$

$$w_{23}^1 = -0,82 - [0,1 * (-0,2282)] = -0,7972$$

$$w_{31}^1 = -0,41 - [0,1 * (-0,2029)] = -0,3897$$

$$w_{32}^1 = 0,82 - [0,1 * (-0,1081)] = 0,8308$$

$$w_{33}^1 = 0,67 - [0,1 * (-0,2282)] = 0,6928$$

$$w_{41}^1 = 0,55 - [0,1 * (-0,2029)] = 0,5703$$

$$w_{42}^1 = 0,58 - [0,1 * (-0,1081)] = 0,5908$$

$$w_{43}^1 = 0,90 - [0,1 * (-0,2282)] = 0,9228$$

Após os pesos serem atualizados, é introduzido um novo conjunto de entradas na rede, dando início a uma nova iteração do processo de treinamento. O processo de treinamento da rede somente é finalizado quando o número máximo de iteração é atingindo, ou quando o critério de convergência é alcançado.

O número máximo de iterações e o critério de convergência “erro mínimo desejado” são dois parâmetros importantes que devem ser considerados durante a etapa de treinamentos. Geralmente, as pessoas que utilizam uma rede neural pela primeira vez imaginam que o ideal seria usar um número de iterações elevado e um erro tendendo a zero. Contudo, este procedimento pode levar a um sobre-ajuste dos dados, tornando o modelo pouco robusto e conseqüentemente sem aplicabilidade prática para o sistema estudado.

Isso normalmente acontece porque as redes neurais inicialmente tentam estabelecer correlações diretas entre as entradas e saídas, minimizando o erro de forma homogênea para o conjunto de treinamento. Desta maneira, no início do processo de treinamento, são consideradas principalmente as variações correlacionadas entre as entradas e saídas dos padrões de treinamento, que modelam o comportamento de todas as amostras. Em um segundo momento, o modelo começa a ser ajustado para explicar relações específicas entre os padrões de entrada e saída. Geralmente, estas relações estão relacionadas a ruídos e/ou incertezas, o que obviamente degrada a habilidade de generalização do modelo de calibração.

Desta maneira, é óbvio que um tempo de treinamento longo ou um valor do critério de convergência muito baixo não contribuem para a modelagem dos dados. O ideal neste caso é

atribuir uma maior importância ao valor do erro máximo desejado. Em alguns sistemas reais, o analista possui uma boa idéia do erro associado ao sistema, o que permite que este atribua um critério de convergência coerente ao erro embutido nos padrões utilizados para construção do modelo de calibração.

3.7 Critérios de Convergência

Existem vários critérios de convergência utilizados para cessar o processo de ajuste dos pesos da rede neural; entretanto, praticamente todos utilizam a informação do vetor gradiente de erro como um parâmetro para finalizar o processo de minimização do erro. Isso se deve ao fato do vetor gradiente de erro possui informações de quanto o processo de otimização encontra-se próximo do ponto de mínimo global ou local da superfície de erro.

Os critérios de convergência sugeridos para o algoritmo de retropropagação segundo Haykin são⁽⁵¹⁾:

1. O algoritmo converge quando a norma Euclidiana da estimativa do vetor gradiente $\|\nabla J(w)\|$ atinge um valor suficientemente pequeno, pré-definido.
2. Considera-se que o algoritmo convergiu quando a variação do erro quadrático de uma série de iterações não apresenta uma variação significativa, ou seja, isso implica que foi atingido um estado estacionário no processo de otimização.
3. Pode-se considerar que o algoritmo convergiu quando o valor do erro quadrático médio for igual ou menor a um valor pré-especificado.

O critério de parada do treinamento menos aconselhável é o que pré-estabelece um número máximo de iterações. Geralmente este critério pode levar a um modelamento inadequado dos dados devido à sobreajuste dos dados. Os outros critérios de parada levam em consideração a capacidade de generalização do modelo, ou seja, após cada iteração se avalia se o novo modelo perdeu capacidade de generalização ou robustez. Esta é uma maneira de se evitar a construção de um modelo que somente apresenta bom desempenho para um conjunto específico de dados.

3.8 Rede Neurogenética

Nesta parte iremos tratar das redes denominadas de neurogenéticas. Elas pertencem ao grupo de algoritmos híbridos, propostos para melhorar o desempenho dos algoritmos tradicionais de modelamento e classificação de padrões⁽⁶⁴⁾.

Como foi apresentado anteriormente, o método tradicional de treinamento das redes neurais utilizam informações do gradiente local da função objetivo, com a finalidade de minimizar a superfície de erro. Entretanto, estes métodos são sensíveis a mínimos locais, o que pode proporcionar dificuldades ao processo de treinamento da rede neural. Um dos grandes problemas do treinamento da rede neural é o fato que os métodos que utilizam o gradiente da função possuem apenas informações locais sobre a superfície de erro⁽³⁷⁾. Este tipo de informação local não é suficiente para afirmarmos que o mínimo global da superfície de erro foi alcançado. Entretanto, o algoritmo, após alcançar a situação de mínimo, realiza algumas pesquisas locais para verificar se existe outra direção que pode ser seguida para minimizar o erro.

Uma maneira proposta para superar este problema é a utilização do algoritmo genético para o treinamento da rede neural. Espera-se que o algoritmo genético apresente resultados iguais ou melhores que os métodos tradicionais de treinamento, isso porque o algoritmo genético é um método de otimização não derivativo, ou seja, menos sensível a presença de mínimos locais.

3.8.1 Implementação do Algoritmo Genético

Para a implementação de um algoritmo genético, é necessário inicialmente definir o problema que se deseja solucionar: *“Desejamos obter um conjunto de pesos, de maneira que estes ofereçam o menor erro de previsão para um conjunto de padrões de treinamento. Para tanto, será utilizada uma rede neural com uma arquitetura pré-definida e fixa”*.

Depois de definido o problema, nós devemos codificá-lo de maneira que seja possível a implementação do algoritmo genético. A codificação de cada parâmetro, peso de uma conexão sináptica, será realizada em um segmento de um vetor linha composto de zeros e uns. Cada parâmetro otimizado está em uma região fixa e bem determinado do vetor, conforme ilustrado na Figura 15.

Peso 1	Peso 2	Peso 3	Peso 4	Peso 5
--------	--------	--------	--------	--------

Figura 15. Codificação de um sistema na forma de um cromossomo.

3.8.2 Codificação de uma rede neural

Para codificar a rede, deve-se inicialmente imaginar como podemos reescrevê-la na forma de um vetor. Para tanto, pode-se criar um vetor contendo segmentos com os valores dos pesos dos neurônios e dos *bias* da primeira e segunda camada, respectivamente, conforme ilustrado na Figura 16.

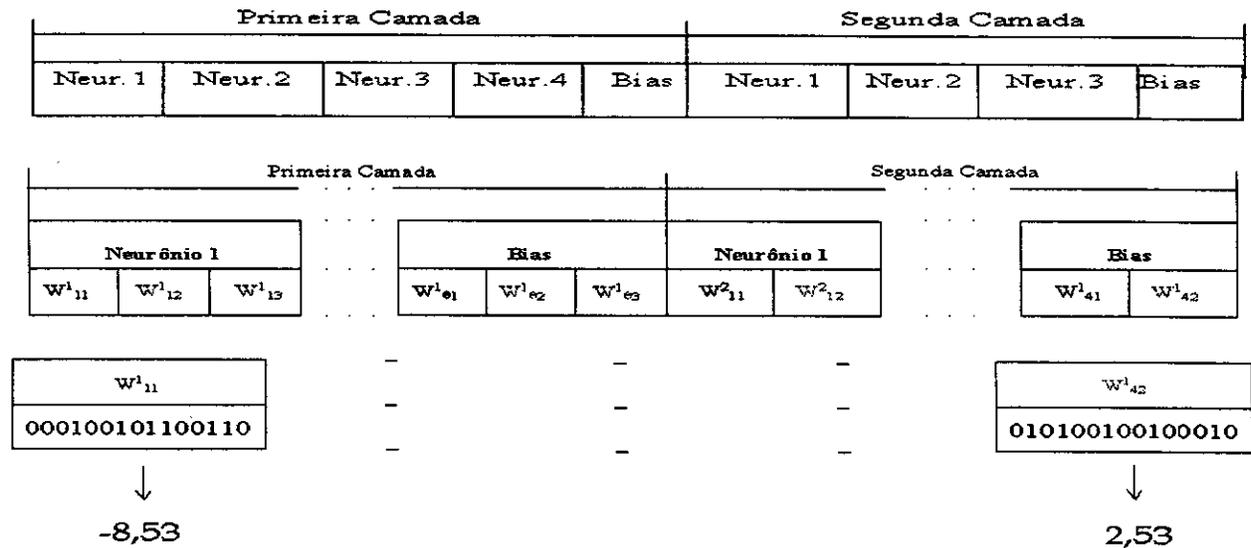


Figura 16. Codificação dos pesos de uma rede neural na forma de um vetor.

Na Figura 16, pode-se observar que cada “peso” está representado por um número fixo de elementos pertencentes ao vetor. Ou seja, neste caso, cada peso possui 15 elementos compostos de 0 e 1. Isso implica que esta sequência pode representar qualquer número entre 0 e 32767. Para que possamos utilizar esta informação, nós inicialmente devemos converter o respectivo número binário em um número decimal entre 0 e 32767. Após esse procedimento, utiliza-se uma outra relação matemática para converter o respectivo número decimal em um número que se encontre dentro da faixa de interesse. Por exemplo, se desejamos que os pesos variem entre valores compreendidos entre -10 e 10 unidades, devemos estabelecer uma relação que atribua o valor 0 ao número -10 e o valor 32767 ao número 10. Esse procedimento faz com que esta escala tenha uma precisão de 0,0006 unidades. Portanto, neste ponto, o leitor deve observar que existe um compromisso entre precisão, faixa em que os valores dos pesos devem assumir e, conseqüentemente, o número de elementos do vetor final. Este fator implica diretamente na qualidade dos dados obtidos e no tempo de processamento dos dados. Quanto maior o número de elementos de um vetor, mais memória será necessária para a realização dos cálculos.

3.8.3 Etapa de treinamento e avaliação dos cromossomos

Depois de estabelecido como será codificado o problema, o próximo passo é a realização de sua operacionalização. Para tanto, nós devemos construir um programa que seja capaz de gerar um número pré-definido de vetores compostos de uma sequência aleatória entre 0 e 1. A este conjunto de vetores iniciais é denominado de população inicial de cromossomos. Em seguida, cada sequência do cromossomo associado a um peso será decodificada, ou seja, esta será convertida em um número real compreendido, por exemplo, na faixa entre -10 a 10. Após este procedimento ter sido realizado, teremos o conjunto de pesos dos neurônios associados à primeira e segunda camada da rede neural.

Neste ponto, é calculada a ativação das entradas da segunda camada, proporcionada pelos estímulos proveniente da camada de entrada. Para que se possa calcular a ativação da segunda camada, nós devemos inicialmente calcular o nível de ativação (eq. 22) de cada um dos neurônios da segunda camada. Em seguida, para calcular a saída da rede neural, utilizam-se os valores das saídas da camada interna como entrada para os neurônios da camada de saída.

O passo seguinte é determinar o erro de previsão da rede neurogenética para cada uma das saídas, com o objetivo de avaliar o grau de aptidão do cromossomo associado à solução proposta.

O erro na saída da rede neural é calculado subtraindo-se o valor previsto pela rede, do valor real associado ao conjunto de dados de entrada. Depois de avaliado o grau de aptidão de cada cromossomo pertencente à população de uma geração, inicia-se o processo de seleção, cruzamento e mutação, com o objetivo de gerar uma nova população mais apta ao sistema. A adaptação de um indivíduo (cromossomo) é indicada pela diferença entre a saída da rede e o valor esperado para um determinado padrão de entrada.

Após cada nova geração, a diferença entre o valor estimado e o previsto para um determinado padrão de entrada deve ser minimizada, até que seja alcançado um valor aceitável, segundo os critérios do operador. Para o treinamento da rede neurogenética, nós optamos por utilizar a técnica de treinamento padrão a padrão, pois desta maneira o número de amostras de treinamento da rede é menor. Isso se deve ao fato de que não é necessário o uso de um conjunto interno de avaliação, semelhante ao requerido pelo método de seleção de variáveis com algoritmo genético.

4. Introdução à Lógica Fuzzy

4.1 Teoria Fuzzy

Historicamente, os modelos probabilísticos foram as primeiras ferramentas matemáticas desenvolvidas para representar as incertezas associadas a uma série de fenômenos físicos. Os modelos probabilísticos partem do princípio que o sistema de interesse apresenta incertezas de caráter aleatório; portanto, o seu comportamento não pode ser explicado por um modelo determinístico⁽⁶⁾.

Na natureza, existem inúmeros sistemas que não podem ser representados por um modelo determinístico como, por exemplo: - Suponhamos que se tenha um fragmento de material radioativo que emita partículas gama. Com o auxílio de um dispositivo de contagem, podemos registrar o número de partículas emitidas durante um intervalo de tempo específico. É evidente que não poderemos antecipar precisamente o número de partículas emitidas, ainda que se conheça de modo exato a forma, a dimensão, a composição química e a massa do objeto em estudo. Por isso, parece não existir um modelo determinístico razoável que forneça o número de partículas emitidas, como uma função das várias características pertinentes ao material da fonte. Entretanto, com o uso de um modelo probabilístico, é possível realizar uma descrição estatística completa deste sistema.

Um processo considerado probabilístico é caracterizado quando: 1- o espaço de amostragem de diferentes processos é constante ao longo do tempo; 2- a frequência de ocorrência, ou probabilidade, de um evento dentro do espaço de amostragem é constante e não muda de experimento para experimento; 3- uma seqüência de resultados deste processo não pode ser prevista. Contudo, existem alguns sistemas que a incerteza não apresenta um comportamento aleatório, tornando inviável o uso de modelos probabilísticos. Na tentativa de modelar esta classe de sistemas foi desenvolvida a teoria de conjuntos Fuzzy⁽⁶⁵⁻⁶⁸⁾.

A principal diferença entre a imprecisão nos sistemas Fuzzy e a incerteza probabilística é que a primeira trata de possibilidade (plausibilidade) determinística, ao passo que a segunda lida com a verossemelhança de eventos não-determinísticos. A incerteza que está por trás do conceito de probabilidade geralmente se relaciona com a ocorrência de um fenômeno simbolizado pelo conceito de aleatoriedade, como, por exemplo, quando se tenta determinar “a probabilidade de tirar uma determinada bola de uma urna”, ou “a probabilidade de chover na noite de natal”. Por outro lado, a imprecisão nos conjuntos Fuzzy está relacionada à ambigüidade encontrada na definição de conceitos lingüísticos como “pessoa inteligente” ou “carro velho”.

A teoria de conjuntos Fuzzy foi desenvolvida nos meados da década de sessenta pelo Prof. Dr. Lotfi A. Zadeh⁽⁶⁾. Após a apresentação deste trabalho pioneiro, uma série de trabalhos teóricos sobre lógica e conjuntos Fuzzy foram desenvolvidos nos Estados Unidos, Europa e Japão. Entretanto, apesar da teoria Fuzzy ter sido desenvolvido nos Estados Unidos, foram os japoneses os primeiros a aplicarem com sucesso os conceitos da teoria Fuzzy a um vasto número de produtos eletrônicos⁽⁶⁹⁻⁷¹⁾.

Os termos conjunto e lógica Fuzzy são normalmente empregados como sinônimos pela maioria dos autores, portanto, torna-se difícil estabelecer a diferença entre estes dois termos. Entretanto, a designação de conjuntos Fuzzy é usada para descrever toda a teoria de conjuntos Fuzzy que engloba, inclusive, alguns conceitos de lógica Fuzzy. A teoria de conjuntos Fuzzy é vista como uma extensão da teoria de conjuntos clássicos, a qual está associada aos conceitos básicos de funções de pertinência, operações com conjuntos Fuzzy, números Fuzzy, relações Fuzzy, regras Fuzzy, regra composicional de inferência, etc. Enquanto, a lógica Fuzzy pode ser vista como uma extensão da lógica bi-valores e multi-valores⁽³⁷⁾.

A teoria da lógica Fuzzy proposta pelo professor Zadeh é um caso generalizado da teoria clássica dos conjuntos. Segundo a teoria clássica dos conjuntos, um determinado elemento possui apenas dois estados, 0 ou 1. Enquanto, a teoria da lógica Fuzzy é mais flexível, permitindo a um dado elemento x possuir um grau de pertinência contínuo entre $[0,1]$ para um determinado grupo Y .

A Figura 19 ilustra um caso que será utilizado para mostrar a diferença entre a teoria clássica de conjuntos e Fuzzy. Suponha a seguinte situação hipotética: em um curso do segundo grau, o professor de química separou os alunos de uma sala de aula em vários grupos. A cada um dos grupos foi permitida a escolha de um tema para realizar um projeto.

Um dos temas propostos pelo professor consistiu em determinar uma maneira de classificar os elementos químicos da tabela periódica em grupos bem definidos. Os alunos responsáveis por este tema decidiram utilizar como critério de classificação as propriedades físicas de algumas substâncias químicas da tabela periódica.

Após terem conseguido as informações necessárias sobre as propriedades físicas, os alunos não sabiam como organizar os dados de maneira que pudessem distinguir as classes dos elementos químicos. Após refletirem muito sobre o tema, um dos alunos lembrou de uma aula de matemática, na qual o professor introduziu o conceito de conjuntos e o diagrama de Venn.

Baseando-se na teoria de conjuntos e no diagrama de Venn, os alunos tentaram classificar os elementos em grupos bem distintos. Após representar graficamente os dados no diagrama de

Venn, os alunos classificaram os elementos químicos em quatro conjuntos, conforme apresentado na Figura 19. Entretanto, alguns dos elementos apresentaram características intermediárias entre algumas das classes pré-definidas.



Figura 17. Representação do Diagrama de Venn 4 conjuntos Fuzzy.

Devido às limitações do modelo adotado, o grupo decidiu atribuir os elementos não classificados apropriadamente, ao conjunto vizinho mais próximo. Entretanto, a abordagem mais apropriada para a classificação dos elementos seria a utilização da teoria de conjuntos Fuzzy.

A teoria de conjuntos Fuzzy associa a cada conjunto uma região Fuzzy ou nebulosa. Esta região é caracterizada por associar um grau de pertinência distinto aos elementos que não pertencem totalmente a um dado conjunto. Desta maneira é possível capturar informações adicionais, que permitem realizar a classificação de uma maneira mais adequada.

Na Figura 17, cada conjunto é representado por um círculo colorido (vermelho, verde, azul e violeta) e envolta de cada um dos conjuntos existe uma região acinzentada, que representa uma zona de transição que associa um grau de pertinência ao elemento presente nesta região. Esta região de transição pode ser vista como um gradiente, que aumenta o grau de pertinência da periferia para o centro da circunferência. Portanto, os elementos no interior do círculo colorido possuem um grau de pertinência igual a 1 e os elementos pertencente a região acinzentada possuem graus distintos de pertinência que variam no intervalo $[0,1[$. Analisando a Figura 17 de forma qualitativa, segundo os princípios da teoria Fuzzy, pode-se dizer que o elemento representado pela letra A não pertence a nenhum dos conjuntos propostos. O grau de pertinência de cada um dos conjuntos (violeta, verde, azul e vermelho) é igual a zero nesta região do espaço. Portanto, é necessário criar um novo conjunto para este elemento. O elemento B pertence ao conjunto

vermelho com um grau de pertinência maior que 0 e menor que 1. Certamente, não podemos afirmar que este elemento tenha as mesmas características dos elementos contidos no interior do conjunto representado pelo círculo em vermelho. Entretanto, não é correto afirmar simplesmente que este elemento não tenha nenhuma característica similar às apresentadas pelos elementos do conjunto em vermelho.

O elemento C possui uma característica dual entre os conjuntos representados pelos círculos em azul e violeta. Neste caso, cada um dos conjuntos exerce um grau de pertinência distinto sobre este elemento. Portanto, podemos dizer que as propriedades físicas desse elemento são um híbrido das observadas para o conjunto azul e violeta. O elemento D apresenta um alto grau de complexidade, pois possui diferentes graus de pertinência dos conjuntos representados pelos círculos em vermelho, verde e azul. O grau de pertinência de cada um dos elementos é definido de acordo com a semelhança que este apresenta com relação a um conjunto específico.

O mesmo raciocínio pode ser aplicado para uma série de sistemas químicos, que apresentam inúmeras situações de incertezas e ambigüidades, semelhantes à classificação de um elemento como metal ou não metal, solução ácida ou básica, solução diluída ou concentrada. Em todos os casos citados, a classificação é complexa na região de interface dos conjuntos.

4.2 Introdução aos conceitos Matemáticos de Conjunto Fuzzy

A necessidade da noção de conjunto surge quando se necessita organizar, resumir ou generalizar informações a respeito de uma série de objetos.

Um conjunto é definido por uma coleção de pares ordenados, conforme apresentado pela equação abaixo.

$$A = \{(x, \mu_A(x)) \mid x \in X\} \quad \text{eq. (34)}$$

Esta equação diz que os elementos pertencentes ao universo X apresentam diferentes graus de pertinência para um dado conjunto A, onde o grau de pertinência de cada elemento é expresso pela função $\mu_A(x)$.

Na teoria clássica de conjuntos o grau de pertinência é bivalente, podendo assumir os valores 1 ou 0. Portanto, esta teoria segue a lógica clássica de Aristóteles onde uma determinada afirmação somente pode ser verdadeira ou falsa.

A Figura 18 mostra uma representação gráfica de uma função de pertinência bivalente e multivalente. Na representação bivalente, Figura 18.a, é possível observar que a mudança de

estados ocorre de maneira brusca, enquanto na representação multivalente, Figura 18.b, a transição de estados ocorre de maneira suave.

Uma transição brusca como mostra a função de pertinência bivalente leva a uma descontinuidade no processo de classificação. Os elementos próximos à região de fronteira são excluídos apesar de possuírem um comportamento muito similar aos elementos que pertencem a um determinado conjunto.

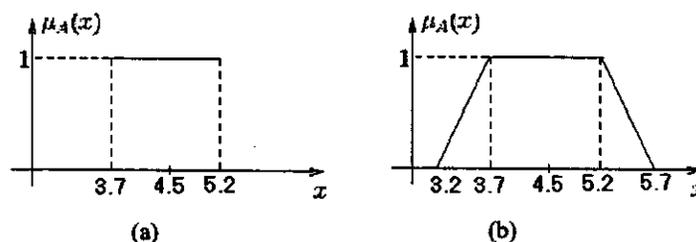


Figura 18. Funções de pertinência: (a) Conjunto clássico, (b) Conjunto Fuzzy.

Em geral, o formato da função de pertinência é restrito a uma certa classe de funções, representadas por parâmetros específicos. Os formatos mais comuns das funções de pertinência são: triangular, trapezoidal, sino generalizada, gaussiana, sigmoidal e direita-esquerda⁽⁷³⁾. Além dos formatos tradicionais existe uma forma bastante utilizada em aplicações práticas: o conjunto unitário (*singleton*).

4.3 Funções de Pertinência

As funções de pertinência são um dos mais importantes componentes dos conjuntos Fuzzy, pois têm a função de armazenar toda a informação dos conjuntos Fuzzy. Por este motivo, um grande número de pesquisadores têm destinado parte de seus trabalhos ao desenvolvimento de novas funções de pertinência. Basicamente, estes trabalhos consistem em tentar determinar a forma da função de pertinência que melhor represente o comportamento da incerteza contida em um sistema. Desta maneira, é possível desenvolver um modelo que explique de maneira mais eficiente o comportamento de um dado sistema.

As principais funções de pertinência são:

1. **Função triangular:** é caracterizada por três parâmetros $\{a, m, b\}$ como apresentado a seguir:

$$\text{triângulo}(x; a, m, b) = \begin{cases} 0, & x \leq a. \\ \frac{x-a}{m-a}, & a \leq x \leq m. \\ \frac{b-x}{b-m}, & m \leq x \leq b. \\ 0, & b \leq x. \end{cases} \quad \text{eq. (35)}$$

2. **Função trapezoidal** é caracterizada por quatro parâmetros $\{a, b, c, d\}$ como apresentado a seguir:

$$\text{trapezoidal}(x; a, b, c, d) = \begin{cases} 0, & x \leq a. \\ \frac{x-a}{b-a}, & a \leq x \leq b. \\ 1, & b \leq x \leq c. \\ \frac{d-x}{d-c}, & c \leq x \leq d. \\ 0, & d \leq x. \end{cases} \quad \text{eq. (36)}$$

A função triangular e trapezoidal são caracterizadas pela simplicidade da implementação computacional, o que as tornam atrativas para aplicações em tempo real. Além disso, estas funções são mais simples de serem utilizadas quando se deseja representar matematicamente o conhecimento de um especialista.

3. **Função gaussiana:** é caracterizada por dois parâmetros $\{c, \sigma\}$ como apresentado a seguir:

$$\text{Gaussiana}(x; c, \sigma) = e^{-\frac{1}{2} \left(\frac{x-c}{\sigma} \right)^2} \quad \text{eq. (37)}$$

4. **Função Sino generalizado ou Cauchy:** é caracterizada por três parâmetros $\{a, b, c\}$:

$$\text{Sino Generalizada}(x; a, b, c) = \frac{1}{1 + \left| \frac{x-c}{a} \right|^{2b}} \quad \text{eq. (38)}$$

A função de pertinência sino generalizado é uma generalização direta da distribuição de Cauchy, usada em teoria de probabilidade. Os parâmetros desta função permitem ajustar a função de acordo com as necessidades de sua aplicação. Os parâmetros a e c são capazes de ajustar o centro e a posição da função, enquanto o parâmetro b é responsável pelo ajuste da inclinação da função no ponto de cruzamento.

5. **Função Sigmoidal:** é caracterizada por dois parâmetros {a,c} como apresentado a seguir:

$$Sigmoidal(x; a, c) = \frac{1}{1 + e^{-a(x-c)}} \quad \text{eq. (39)}$$

6. **Função Direita-Esquerda:** é caracterizada por dois parâmetros {a,c}, como apresentado a seguir:

$$ED(x; c, \alpha, \beta) = \begin{cases} F_E\left(\frac{c-x}{\alpha}\right), & x \leq c \\ F_D\left(\frac{x-c}{\beta}\right), & x \geq c \end{cases} \quad \text{eq. (40)}$$

$$F_E(x) = \sqrt{\max(0, 1-x^2)}, \quad \text{eq. (41) e eq. (42)}$$

$$F_D(x) = e^{-|x|^2}$$

7. **Função Unitária (singleton):** é caracterizada por dois parâmetros {m,h} como apresentado a seguir:

$$Singleton(x; m, h) = \begin{cases} h = 1 & x = m \\ h = 0 & x \neq m \end{cases} \quad \text{eq. (43)}$$

A Figura 19 mostra os perfis das funções de pertinência apresentadas.

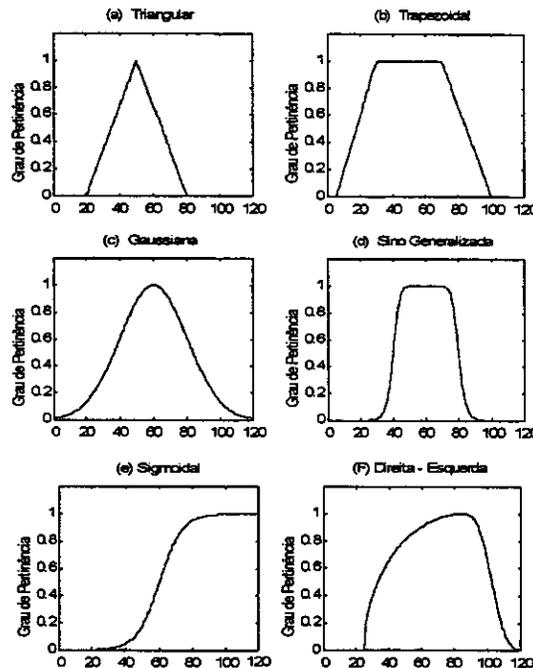


Figura 19. Exemplo das Funções de pertinência mais comumente empregadas em conjuntos Fuzzy.

Cuidados adicionais devem ser tomados quando se deseja utilizar métodos adaptativos que envolvam técnicas derivativas para o ajuste dos parâmetros das funções^(67,74-75). Nesse caso, a função deve ser derivável, e de preferência, possuir uma derivada bem conhecida. Algumas funções propostas podem possuir derivadas complexas de serem implementadas computacionalmente, podendo, portanto, comprometer o processo de otimização de modelagem do sistema. Assim, o melhor procedimento é utilizar funções que possuam derivadas bem conhecidas e simples de serem implementadas computacionalmente como, por exemplo: a função gaussiana e sino generalizado⁽³⁷⁾.

As funções de pertinência da lógica Fuzzy possuem propriedades similares às observadas na teoria de clássica de conjuntos. Portanto, as operações como união, intersecção e complemento são também válidas para os conjuntos Fuzzy. Além dessas operações, existem outros operadores que permitem realizar normalização, dilatação, concentração e intensificação de contraste das funções de pertinência.

4.3.1 Operações entre Funções de Pertinência

1. União

O operador de união tem a propriedade de agregar informações de dois ou mais conjuntos em um novo conjunto. A operação de união fornece o menor conjunto do universo de discurso que contém simultaneamente as informações de dois ou mais conjuntos.

A equação matemática que descreve a operação de união pode-se representada como:

$$\mu_{A \cup B}(x) = \max[\mu_A(x_1), \mu_B(x_2)] = \mu_A(x_1) \vee \mu_B(x_2) \quad \text{eq. (44)}$$

Na prática, existem duas maneiras de realizar a união entre os conjuntos. A primeira é utilizando o operador de máximo, que realiza uma comparação entre os elementos recíprocos dos conjuntos Fuzzy e retorna o maior valor. A segunda maneira é utilizando o operador lógico “ou”.

2. Intersecção

A intersecção é um operador que informa o que dois ou mais conjuntos possuem em comum. Portanto, podemos imaginar que este operador destaca a semelhança entre os conjuntos. O novo conjunto formado é o maior conjunto do universo de discurso que possui apenas as informações dos conjuntos de interesse.

A equação abaixo ilustra a operação de intersecção entre os conjuntos A e B.

$$\mu_{A \cap B}(x) = \min[\mu_A(x_1), \mu_B(x_2)] = \mu_A(x_1) \wedge \mu_B(x_2) \quad \text{eq. (45)}$$

A intersecção pode ser representada pelo operador de mínimo ou o operador lógico “e”.

3. Complemento

O complemento de um conjunto é definido como sendo todos os elementos do universo de discurso que não pertencem a este conjunto. Entretanto, ele é diferente da teoria clássica de conjuntos, pois um conjunto A e seu complemento (A') são sempre não disjuntos, ou seja, a intersecção entre ambos não é o conjunto vazio. Da mesma forma, a união de um conjunto Fuzzy e os eu complemento não é igual ao universo de discurso.

O diagrama de Venn é apresentado na Figura 20 para ilustrar as operações de complementariedade, intersecção e união entre dois conjuntos Fuzzy A e B . Isso permite uma melhor compreensão de como estas operações atuam entre as funções de pertinência.

A Figura 20 mostra dois conjuntos Fuzzy A e B e as operações de união, complemento e intersecção entre eles. A Figura 20.c ilustra que a união entre um conjunto Fuzzy e o seu complemento não resulta no universo de discurso. Enquanto, a Figura 20.d mostra que a intersecção entre um conjunto Fuzzy e o seu complemento é diferente do conjunto vazio. Estes dois exemplos foram apresentados graficamente, pois são os mais difíceis de serem aceitos, visto que toda a nossa cultura é baseada na teoria de conjuntos clássicos. A Figura 21 apresenta outro exemplo que ilustra as operações de união e intersecção com conjuntos Fuzzy.

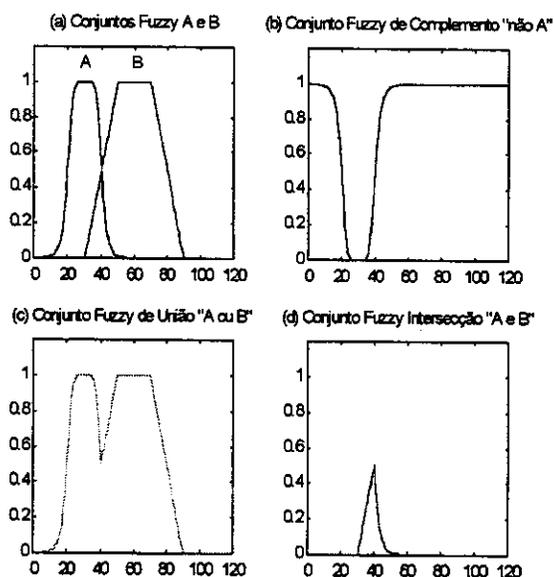


Figura 20. a) Funções de pertinência que representam os conjuntos Fuzzy A e B ; b) Complemento da função de pertinência do conjunto Fuzzy A ; c) Função de pertinência resultante da união entre os conjuntos Fuzzy A e B ; d) Função de pertinência resultante da intersecção entre os conjuntos Fuzzy A e B .

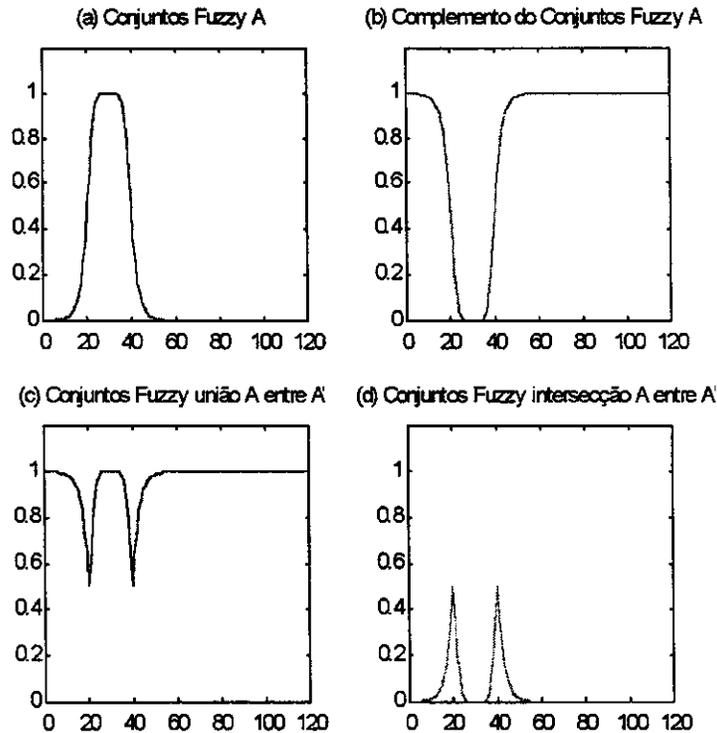


Figura 21. a) Função de pertinência do conjunto Fuzzy A; b) Complemento da função de pertinência do conjunto Fuzzy A; c) Função de pertinência resultante entre a união do conjunto Fuzzy A e o seu complemento; d) Função de pertinência resultante entre a intersecção do conjunto Fuzzy A e o seu complemento.

4.3.2 Classes de Operadores dos Conjuntos Fuzzy

A teoria de conjuntos Fuzzy possui duas classes gerais de operadores (normas triangulares e co-normas triangulares), que são uma extensão dos operadores tradicionais de união e intersecção da teoria dos conjuntos clássicos. Estes operadores foram desenvolvidos para atribuírem maior flexibilidade à modelagem de uma série de sistemas.

O operador norma triangular é uma classe de funções que permite realizar a operação de intersecção entre as funções de pertinência. Esta classe de funções é comumente chamada de norma-T, onde a letra T é a abreviatura de triangular.

Para definir as equações matemáticas envolvidas nas operações da norma-T, iremos necessitar de dois conjuntos Fuzzy A e B definidos no universo X. Para este caso, a norma-T será uma função de duas entradas definidas nos domínios $T : [0,1] \times [0,1] \rightarrow [0,1]$, onde “X” representa o operador de produto cartesiano. A função de pertinência resultante do produto cartesiano entre os conjuntos A e B, é definida como:

$$\mu_{AXB}(x) = \min(\mu_A(x_1), \mu_B(x_2)) \text{ ou } \mu_{A \cap B}(x) = T(\mu_A(x_1), \mu_B(x_2)) \quad \text{eq. (46)}$$

O produto cartesiano é representado por uma matriz relacional, que contém todas as combinações possíveis entre os pares ordenados formados pelo produto cartesiano dos elementos do universo de entrada e saída. Os elementos desta matriz podem assumir qualquer valor entre 0 e 1. Tal valor também é chamado de intensidade de mapeamento ou força de ligação.

4.3.2.1 Operadores do Tipo Norma-T

Os operadores mais populares da norma-T são: mínimo, produto algébrico, produto limitado e produto drástico.

- **Mínimo**

$$T_1(\mu_A(x_1), \mu_B(x_2)) = \min(\mu_A(x_1), \mu_B(x_2)) \quad \text{eq. (47)}$$

- **Produto Algébrico**

$$T_2(\mu_A(x_1), \mu_B(x_2)) = \mu_A(x_1)\mu_B(x_2) \quad \text{eq. (48)}$$

- **Produto Limitado**

$$T_3(\mu_A(x_1), \mu_B(x_2)) = 0 \vee (\mu_A(x_1) + \mu_B(x_2) - 1) \quad \text{eq. (49)}$$

- **Produto Drástico**

$$T_4(\mu_A(x_1), \mu_B(x_2)) = \begin{cases} \mu_A(x_1), & \text{se } \mu_B(x_2) = 1 \\ \mu_B(x_2), & \text{se } \mu_A(x_1) = 1 \\ 0, & \text{se } \mu_A(x_1) \text{ e } \mu_B(x_2) < 1. \end{cases} \quad \text{eq. (50)}$$

De maneira análoga, para a norma-T, existe uma classe de funções chamada de co-norma-T ou norma-S, a qual é definida genericamente pela expressão matemática:

$$\mu_{A \cup B}(x) = S(\mu_A(x_1), \mu_B(x_2)).$$

4.3.2.2 Operadores do Tipo Norma-S ou Co-Norma-T

O operador norma-S é uma classe de funções que permite realizar a operação de união entre funções de pertinências contidas no mesmo universo de discurso.

As normas-S mais comuns são máximo, soma algébrica, soma limitada, soma drástica, etc.

• **Máximo**

$$S_1(\mu_A(x_1), \mu_B(x_2)) = \max(\mu_A(x_1), \mu_B(x_2)) \quad \text{eq. (51)}$$

• **Soma Algébrica**

$$S_2(\mu_A(x_1), \mu_B(x_2)) = \mu_A(x_1) + \mu_B(x_2) - \mu_A(x_1)\mu_B(x_2) \quad \text{eq. (52)}$$

• **Soma Limitada**

$$S_3(\mu_A(x_1), \mu_B(x_2)) = 1 \wedge (\mu_A(x_1) + \mu_B(x_2)) \quad \text{eq. (53)}$$

• **Soma Drástica**

$$S_4(\mu_A(x_1), \mu_B(x_2)) = \begin{cases} \mu_A(x_1), & \text{se } \mu_B(x_2) = 0 \\ \mu_B(x_2), & \text{se } \mu_A(x_1) = 0 \\ 0, & \text{se } \mu_A(x_1) \text{ e } \mu_B(x_2) > 0. \end{cases} \quad \text{eq. (54)}$$

4.3.2.3 Complemento Fuzzy

O complemento é uma função contínua que opera sobre o espaço de discurso das funções de pertinência $N : [0,1] \rightarrow [0,1]$. Para que o operador de complemento possa ser aplicado sobre uma função de pertinência, os seguintes axiomas devem ser satisfeitos:

$$\text{Condições de contorno: } \begin{cases} N(0) = 1 \\ N(1) = 0 \end{cases}$$

$$\text{Condição de monotonicidade: } N(\mu_A(x_1)) \geq N(\mu_A(x_2)) \quad \text{se } \mu_A(x_1) \leq \mu_A(x_2)$$

$$\text{Condição de involução: } N(N(\mu_A(x))) = \mu_A(x)$$

Os operadores de complemento Fuzzy mais famosos são o complemento de Sugeno e o complemento de Yager.

$$\text{O complemento de Sugeno é definido como: } N_p(\mu_A(x)) = \frac{1 - \mu_A(x)}{1 + p * \mu_A(x)}$$

onde p é um parâmetro que pode assumir qualquer valor no intervalo $[-1, \infty]$. Quando p é igual a zero, o operador de Sugeno tem o mesmo comportamento do complemento definido para os conjuntos clássicos.

$$\text{O complemento de Yager é definido como: } N_r(\mu_A(x)) = (1 - \mu_A(x))^{\frac{1}{r}}$$

onde o parâmetro r é definido no domínio dos números reais positivos.

4.3.2.4 Outros Tipos de Operadores

1. **Concentração:** Diminui os valores da função de pertinência

$$Con(\mu_A(x)) = \mu_A(x)^2 \quad \text{eq. (55)}$$

2. **Dilatação:** Aumenta os valores da função de pertinência

$$Dil(\mu_A(x)) = \sqrt{\mu_A(x)} \quad \text{eq. (56)}$$

3. **Intensificação de contraste:** Dilata o conjunto para valores de pertinência acima de 0,5 e concentra o conjunto para valores abaixo de 0,5.

$$Int(\mu_A(x)) = \begin{cases} 2 * \mu_A(x)^2, & \text{se } 0 \leq \mu_A(x) \leq 0,5 \\ (1 - \mu_A(2 * (1 - \mu_A(x))^2)) & \text{se } 0,5 \leq \mu_A(x) \leq 1 \end{cases} \quad \text{eq. (57)}$$

4.3.3 O Princípio de Extensão

Em aplicações de escopo prático, é imprescindível que seja possível operar em universos de discurso distintos. Por exemplo, se desejarmos modelar um sistema que dependa das variáveis temperatura, pressão e volume, é necessário que as funções de pertinência que representem os conjuntos contidos no universo possam interagir entre si.

O princípio de extensão é um procedimento da teoria de conjuntos Fuzzy que permite realizar um mapeamento ponto a ponto entre conjuntos Fuzzy. Este princípio garante que existe uma função não necessariamente explícita que agregue todas as variáveis do sistema que se deseja modelar, de forma análoga à função $f(x,y,z)$.

Por exemplo, suponhamos que exista uma função explícita que mapeie o espaço no domínio de X para Y, sendo que no domínio de X existe um conjunto Fuzzy A definido como:

$$A = \frac{\mu_A(x_1)}{x_1} + \frac{\mu_A(x_2)}{x_2} + \dots + \frac{\mu_A(x_n)}{x_n}. \text{ O sinal de "+" indica que se trata de uma coleção de objetos}$$

discretos.

Suponha que o conjunto Fuzzy A seja definido entre o intervalo -3 e 3, e que conhecemos os valores dos graus de pertinência de cada um dos elementos entre o intervalo -3 e 3. Portanto, poderemos representar os elementos desse conjunto Fuzzy como:

$$A = \frac{0}{-3} + \frac{0,1}{-2} + \frac{0,3}{-1} + \frac{0,5}{0} + \frac{0,7}{1} + \frac{0,4}{2} + \frac{0,2}{3}.$$

O princípio de extensão garante que a imagem de um conjunto Fuzzy A sobre uma função de mapeamento $f(\cdot)$ fornece como resultado os elementos de um conjunto Fuzzy B, ou seja:

$$B = f(A) = \frac{\mu_A(x_1)}{y_1} + \frac{\mu_A(x_2)}{y_2} + \dots + \frac{\mu_A(x_n)}{y_n}. \text{ Portanto, se considerarmos que existe neste caso uma}$$

função de mapeamento explícita do tipo: $f(x) = x^2 + x - 5$, o conjunto Fuzzy B definido no domínio de Y pode ser representado por: $B = \frac{0}{1} + \frac{0,1}{-3} + \frac{0,3}{-5} + \frac{0,5}{-5} + \frac{0,7}{-3} + \frac{0,4}{1} + \frac{0,2}{7}$. Entretanto,

normalmente o mapeamento entre as variáveis não é unívoco, ou seja, para cada elemento definido no domínio de X pode existir mais de uma imagem definida no domínio de Y. Para contornar este problema, utiliza-se o grau de pertinência com maior valor, conforme ilustrado a seguir:

$$B = \frac{(0 \vee 0,4)}{1} + \frac{(0,1 \vee 0,2)}{-3} + \frac{(0,3 \vee 0,5)}{-5} + \frac{0,2}{7}, \quad B = \frac{0,4}{1} + \frac{0,2}{-3} + \frac{0,5}{-5} + \frac{0,2}{7}. \text{ Este mesmo}$$

procedimento pode ser estendido para um mapeamento n-dimensional. Entretanto, o operador de máximo deve ser interpretado como um operador de união, ou seja, segundo o que foi mostrado anteriormente, este pode ser substituído por qualquer um dos operadores de norma-S.

Normalmente, o princípio de extensão é expresso na forma de matrizes relacionais ou função de relação. Este tem a finalidade de mapear os elementos de um domínio ordinário para o domínio Fuzzy. A relação entre o mapeamento de dois espaços, utilizando uma matriz relacional, pode ser expresso como $B(y) = A(x) \bullet R(x, y)$. Onde “ \bullet ” é o operador composicional que indica uma operação generalizada, similar à uma norma-T ou norma-S. Essa expressão é chamada de regra composicional de inferência.

A regra composicional de inferência diz que, se a relação Fuzzy ($R(x, y)$) que representa um sistema é conhecida, então a resposta do sistema ($B(y)$) pode ser determinada a partir de uma excitação conhecida ($A(x)$).

Para que o leitor possa compreender melhor como funciona estas relações matemáticas, propomos o seguinte exemplo, suponha que desejamos fazer um mapeamento entre dois espaços (Pressão e Volume), onde o espaço pressão da matriz relacional assume os seguintes valores $P = \{1,4; 1,5; 1,6; 1,7; 1,8\}$, enquanto o espaço do volume assume os valores $V = \{48; 46; 44; 42; 40\}$. Neste problema, um determinado material apresenta variação em seu volume segundo a variação da pressão exercida sobre este. Entretanto, não é possível estabelecer com exatidão o volume do material a uma dada pressão. Para modelar este sistema serão aplicados

os princípios da lógica Fuzzy, partindo da experiência de um especialista da área. Depois da entrevista com o especialista, o projetista do sistema Fuzzy foi capaz de estabelecer a seguinte matriz relacional:

		Pressão →					
		1,4	1,5	1,6	1,7	1,8	
$R(v, p) =$	1	0,8	0,2	0,1	0	48	↑ Volume
	0,8	1	0,8	0,2	0,1	46	
	0,2	0,8	1	0,8	0,2	44	
	0,1	0,2	0,8	1	0,8	42	
	0	0,1	0,2	0,8	1	40	

A matriz relacional $R(v, p)$ estabelece o grau de pertinência entre os domínios definidos pela variável pressão e volume. Podemos verificar na primeira coluna que a pressão de 1,4 atmosferas tem o grau de pertinência igual a 1 para o volume de 48 cm³; para a mesma pressão, o grau de pertinência é igual a 0,8, para o volume de 46 cm³. Segundo o que podemos observar, o grau de pertinência diminuiu segundo a redução do volume do material. O comportamento inverso é verificado quando se aumenta a pressão do sistema.

Nosso objetivo é determinar o conjunto Fuzzy do domínio da variável volume induzido por um conjunto Fuzzy do domínio da variável pressão, a partir a matriz relacional conhecida. Portanto, teremos que utilizar a seguinte equação: $P(p) = V(v) \bullet R(v, p)$.

O vetor de pertinência de entrada $P = \{0,3; 0,5; 0,75; 0,9; 1\}$ irá induzir uma saída (um vetor de pertinência) no espaço volume. Para que o leitor possa acompanhar o raciocínio, serão apresentadas às operações necessárias para realizar a indução de um conjunto Fuzzy de um espaço, para apenas a primeira coluna da matriz relacional. Para as demais colunas, o processo é similar.

$$\begin{aligned}
 \min(1; 0,3) &= 0,3 \\
 \min(0,8; 0,5) &= 0,5 \\
 \min(0,2; 0,75) &= 0,2 & \max(0,3; 0,5; 0,2; 0,1; 0) &= 0,5 \\
 \min(0,1; 0,9) &= 0,1 \\
 \min(0, 1) &= 0
 \end{aligned}$$

O vetor de pertinência do espaço pressão, $P = \{0,3; 0,5; 0,75; 0,9; 1\}$, induz ao seguinte vetor de pertinência do espaço volume: $V = \{0,5; 0,75; 0,8; 0,9; 1\}$. Desta maneira, mostramos como é realizado o mapeamento entre conjuntos Fuzzy contidos em domínios distintos.

Neste ponto, já foram definidas as funções de pertinência e as operações que permitem a agregação das informações armazenadas nos conjuntos Fuzzy. O próximo passo será apresentar o sistema de inferência Fuzzy que aplica os princípios de raciocínio Fuzzy para armazenar e processar o conhecimento de um sistema.

4.4 Raciocínio Fuzzy

Nesta seção, será apresentado como os conceitos de variáveis lingüísticas permitem a ligação entre as regras Fuzzy e o raciocínio aproximado. Existem dois tipos básicos de inferência Fuzzy : o “modus ponens” (modo afirmativo) e o “modus tollens” (modo negativo), ambos operando de forma a se basear em premissas ou condições, as quais geram uma determinada consequência.

A regra básica de inferência tradicional opera com dois valores lógicos, segundo o qual nós podemos inferir a verdade de uma proposição B, a partir da verdade de A. Neste tipo de situação, pode-se usar o procedimento de inferência do raciocínio Fuzzy para derivar conclusões, a partir de um conjunto de regras Fuzzy e fatos conhecidos.

Premissa 1 (fato) :	x é A e y é B		
Premissa 2 (regra 1) :	Se x é A ₁ e y é B ₁	Então	z é C ₁
Premissa 3 (regra 2) :	Se x é A ₂ e y é B ₂	Então	z é C ₂
Premissa 4 (regra 3) :	Se x é A ₃ e y é B ₃	Então	z é C ₃
:	:	:	:
Premissa n (regra n-1) :	Se x é A _{n-1} e y é B _{n-1}	Então	z é C _{n-1}
Consequência (conclusão) :			z é C _i

Para estabelecer uma relação entre os termos antecedentes (A e B) e o termos consequentes (C), é necessário uma matriz relacional (R).

Matematicamente, o processo de inferência pode ser representado por:

$$C_1 = (A \times B) \bullet (A_1 \times B_1 \rightarrow C_1)$$

$$R_1 = A_1 \times B_1 \rightarrow C_1$$

$$C_1 = (A \times B) \bullet R_1$$

$$C_2 = (A \times B) \bullet (A_2 \times B_2 \rightarrow C_2)$$

$$R_2 = A_2 \times B_2 \rightarrow C_2$$

$$C_2 = (A \times B) \bullet R_2$$

$$C_3 = (A \times B) \bullet (A_3 \times B_3 \rightarrow C_3)$$

$$R_3 = A_3 \times B_3 \rightarrow C_3$$

$$C_3 = (A \times B) \bullet R_3$$

$$C_{n-1} = (A \times B) \bullet (A_{n-1} \times B_{n-1} \rightarrow C_{n-1})$$

$$R_{n-1} = A_{n-1} \times B_{n-1} \rightarrow C_{n-1}$$

$$C_{n-1} = (A \times B) \bullet R_{n-1}$$

$$C_i = (A \times B) \bullet (R_1 \cup R_2 \cup R_3 \dots \cup R_{n-1})$$

$$C_i = [(A \times B) \bullet R_1] \cup [(A \times B) \bullet R_2] \cup [(A \times B) \bullet R_3] \cup \dots \cup [(A \times B) \bullet R_{n-1}]$$

$$C_i = C_1 \cup C_2 \cup C_3 \cup \dots \cup C_{n-1}$$

onde os termos C_1, C_2, C_3 e C_{n-1} são os conjuntos Fuzzy inferidos pelas regras 1, 2, 3 até n.

O símbolo “•” representa o operador de composição que tem a função de estabelecer uma relação matemática entre a matriz relacional e as premissas iniciais. Desta maneira, é possível determinar o termo conseqüente de cada uma das regras de inferência.

O primeiro passo para construir uma estrutura de um sistema de inferência Fuzzy é estabelecer os seguintes três componentes básicos⁽³⁷⁾: 1- Uma base de dados, onde se define a função de pertinência usada nas regras Fuzzy; 2- Uma base de regras, contendo a seleção de regras Fuzzy; 3- Um mecanismo de raciocínio, o qual avalia o desempenho do procedimento de inferência sobre as regras e permite que seja possível derivar uma conclusão a partir das informações disponíveis.

Na base de dados, são definidos o número e o tipo de funções de pertinência utilizadas para cada uma das variáveis no mapeamento do sistema. Por exemplo, se estivermos modelando uma reação química que depende das variáveis: temperatura, velocidade de agitação, concentração de

um reagente específico e tempo de aquecimento, devemos inicialmente estabelecer quantas funções de pertinência são necessárias para cada variável e o formato da função de pertinência. Segundo a experiência de um químico, a variável temperatura para este sistema pode ser representada por três funções de pertinência do tipo gaussiana. A cada uma das funções de pertinência, associam-se os termos lingüísticos: baixa, média e alta. O mesmo procedimento é adotado para definir o número e a forma das funções de pertinência para as demais variáveis.

É importante salientar que a forma das funções de pertinência e o número de termos lingüísticos são critérios subjetivos. Portanto, o número e o formato das funções de pertinência na modelagem de um sistema pode variar significativamente, segundo os critérios da pessoa responsável em desenvolver o sistema de inferência Fuzzy⁽⁷⁶⁻⁷⁷⁾.

O passo seguinte é criar um conjunto de regras plausíveis que possam estabelecer uma relação entre as funções de pertinência de cada uma das variáveis como:

1. **Se** (temperatura é alta) **e** (velocidade é moderada) **e** (concentração é média) **e** (tempo é médio) **Então** (Rendimento é alto);
2. **Se** (temperatura é média) **e** (velocidade é alta) **e** (concentração é baixa) **e** (tempo é médio) **Então** (Rendimento é médio);
3. **Se** (temperatura é baixa) **e** (velocidade é alta) **e** (concentração é alta) **e** (tempo é alto) **Então** (Rendimento é baixo).

Para finalizar, é necessário um mecanismo de raciocínio que possa processar as informações inseridas no sistema.

4.4.1 Mecanismos de Raciocínio Fuzzy

Existem três sistemas de inferência Fuzzy que têm sido largamente empregado⁽³⁷⁾:

- a) Modelo Fuzzy Mandani: este modelo utiliza conjuntos Fuzzy para os termos antecedentes e consequentes das regras Fuzzy. A saída final é representada por um conjunto Fuzzy resultante da agregação da saída inferida de cada regra. Como consequência, é necessário empregar um método de decodificação que converta a saída para um número ordinário.

- b) Modelo Fuzzy Sugeno: o termo conseqüente é representado por uma função das variáveis de entrada. Neste modelo, a saída é obtida pela média ponderada das saídas inferidas de cada regra e os coeficientes de ponderação são dados pelos graus de ativação das respectivas regras.

- c) Modelo Fuzzy Tsukamoto: O termo conseqüente utiliza funções de pertinência monotônicas. A saída final é obtida pela média ponderada das saídas inferidas de cada regra.

A diferença entre estes três sistemas de inferência Fuzzy está no procedimento de agregação e decodificação dos dados.

A Figura 22 ilustra como as informações são processadas dentro de um sistema de inferência Fuzzy do tipo Mandani. Neste exemplo, foram utilizadas apenas três regras de inferência, onde cada regra está representada por uma linha da Figura 22. A cada uma das colunas, está associada uma variável específica (Temperatura, Velocidade, Concentração e Tempo).

Partindo do conjunto de regras apresentado anteriormente, podemos ler a primeira linha como:

Se (temperatura é baixa) e (velocidade é alta) e (concentração é alta) e (tempo é alto) **Então** (Rendimento é baixo).

Para determinar a contribuição de cada uma das regras na modelagem do sistema, é realizado o seguinte procedimento: para cada uma das variáveis que compõe o termo antecedente (Temperatura, Velocidade, Concentração, Tempo) é inserido um valor de entrada, o qual está associado à respectiva variável. Por exemplo, na Figura 22, existem quatro valores de entrada: temperatura, velocidade e concentração igual a 0,5 e tempo igual a 0,6. Para cada uma das três regras de inferência, estes valores de entrada possuem diferentes graus de pertinência, conforme pode ser observado pela área sombreada nas funções de pertinência da Figura 22.

Graficamente, o grau de pertinência de cada entrada é determinado pelo ponto de intersecção entre uma reta vertical, que parte do valor da entrada representado na base da função de pertinência e o seu perfil de contorno. Por exemplo, a primeira linha na Figura 22 que representa a regra de inferência 1 mostra que o valor de entrada 0,5 para as variáveis temperatura, velocidade e concentração implica em um grau de pertinência igual a 1, 0,1 e 0,75, respectivamente. Para uma

entrada igual a 0,6, a variável tempo assume um grau de pertinência de 0,8. Após determinar a contribuição de cada variável do termo antecedente para a primeira regra, o próximo passo é agregar a informação de cada uma das entradas com o objetivo de estabelecer o valor do termo conseqüente. Como descrevemos anteriormente a agregação das informações dos termos antecedentes é realizada pelos operadores da norma-T.

Para a modelagem deste sistema, foi aplicada a norma-T do tipo produto algébrico, que realiza a operação de multiplicação entre os graus de pertinência das variáveis do termo antecedente. Este mesmo procedimento é realizado para cada uma das regras definidas previamente.

Posteriormente, a informação de cada uma das regras contida no termo conseqüente é condensada por um dos operadores de composição do tipo norma-S. Este operador tem a função de unir todas as informações em apenas um conjunto, conforme ilustra o desenho representado na última linha da coluna rendimento, na Figura 22.

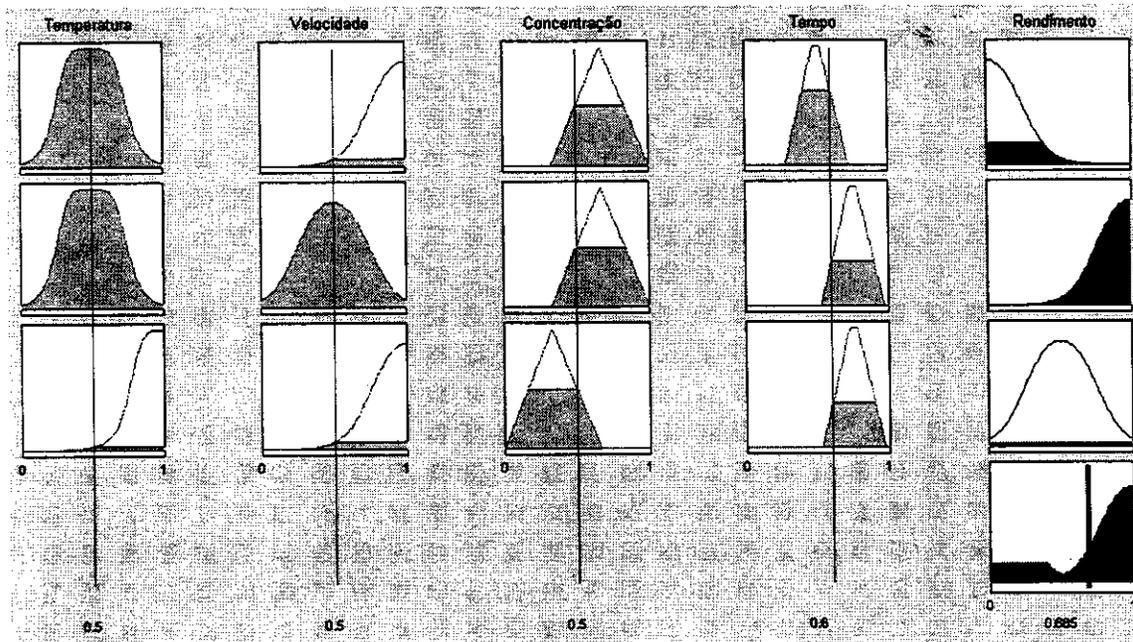


Figura 22. Processo de interação entre as funções de pertinência das variáveis de entrada e saída.

4.5 Métodos de Decodificação

Para converter um conjunto Fuzzy em um número ordinário, é necessário utilizar um método de decodificação. O objetivo de decodificar é obter um valor numérico que melhor represente os valores Fuzzy inferidos da variável lingüística de saída.

Em geral há cinco métodos para extrair a informação de um conjunto Fuzzy A para um universo U: Centróide da área (Z_{COA}); Bissetor da área (Z_{BOA}); Média dos máximos (Z_{MOM}); Maior dos máximos (Z_{LOM}) e Centro dos máximos (Z_{COM})⁽⁷⁾. Entretanto, os métodos mais utilizados são centróide da área, média dos máximos e centro dos máximos.

Centróide da Área: Este método também é conhecido como método do centro de gravidade, pois ele calcula o ponto que divide ao meio a área resultante da agregação entre os termos consequentes de cada regra de inferência.

Este método possui dois problemas. O primeiro ocorre quando as funções de pertinência não são sobreponíveis, o que pode implicar em um centro geométrico vazio. O segundo problema está associado à sobreposição de áreas que não são contabilizadas; além disso, este método necessita que seja realizada uma integração numérica da área.

Centro dos Máximos: Este método também é conhecido pelo método de decodificação pelas alturas, pois utiliza uma média ponderada dos máximos dos picos das funções de pertinência para calcular a saída ordinária.

Média dos Máximos: Nesta abordagem, a saída é calculada pela média de todos os máximos da função de pertinência de saída.

4.6 Rede Neuro-Fuzzy

Quando o sistema dos parâmetros de um sistema Fuzzy é ajustado segundo a experiência do operador, dizemos que trata-se de um sistema Fuzzy normal. Entretanto, em alguns casos, não é suficiente apenas a experiência do operador, pois a complexidade do sistema é muita elevada. Neste caso é, possível a utilização de um sistema híbrido entre redes neurais e lógica Fuzzy. Basicamente, a rede neural tem a finalidade de ajustar os parâmetros de um sistema Fuzzy, de maneira que este seja capaz de modelar um dado sistema. A rede neural pode ser aplicada para ajustar a posição e a abertura das funções de pertinência; entretanto, existem trabalhos que propõem o uso das redes neurais para selecionar o melhor número e formato das funções de pertinência para cada variável. Isso torna o sistema completamente automático, ou seja, neste caso, o operador apenas necessita informar o número de entradas e seus valores. Este tipo de abordagem recebe o nome de Sistema de Inferência Neuro-Fuzzy. Uma outra abordagem alternativa de ajuste

dos parâmetros de um sistema Fuzzy é o uso de algoritmos genéticos; entretanto, atualmente existem poucos trabalhos nesta direção.

4.6.1 Sistema de Inferência Adaptativo Neuro-Fuzzy

Neste ponto, desejamos introduzir a idéia de sistema de inferência adaptativo Neuro-Fuzzy, partindo do que já foi mencionado na seção de redes neurais e lógica Fuzzy. Para tanto, utilizaremos um exemplo que utiliza o sistema de inferência Fuzzy Sugeno de primeira ordem, com duas entradas (x e y) e uma saída z . O primeiro passo é determinar as regras de inferência Fuzzy “Se – Então”⁽⁷⁸⁾.

Regras:

- 1- se x é A_1 e y é B_1 então $f_1 = p_1x + q_1y + r_1$
- 2- se x é A_2 e y é B_2 então $f_2 = p_2x + q_2y + r_2$

Nesta estrutura, A_i e B_i são conjuntos Fuzzy que fazem parte do que chamamos de antecedente da regra, enquanto $f_i = p_i x + q_i y + r_i$ constitui o termo conseqüente da regra Fuzzy. Geralmente $f_{(x,y)}$ é um polinomial das variáveis de entrada x e y , mas nada impede de ser qualquer função, desde que descreva apropriadamente a saída de um modelo como a região Fuzzy específica da regra Fuzzy antecedente.

A Figura 23 ilustra a arquitetura de uma rede Neuro-Fuzzy, com duas entradas (X e Y) e uma saída Z .

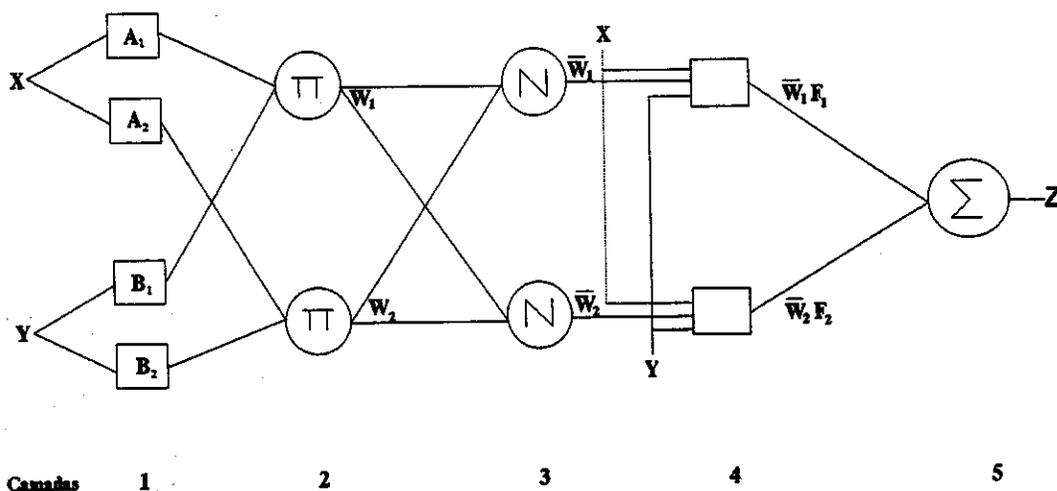


Figura 23. Arquitetura de uma rede Neuro-Fuzzy.

Na camada 1, as entradas são convertidas do domínio de números ordinários para o domínio dos conjuntos Fuzzy, utilizando uma função de pertinência apropriada.

Os valores contidos em cada um dos neurônios da primeira camada podem ser representados por:

$$O_{1,i} = m_{A_i(x)}, \text{ para } i = 1,2 \quad \text{eq. (58)}$$

$$O_{1,i} = m_{B_{i-2}(x)}, \text{ para } i = 3,4 \quad \text{eq. (59)}$$

onde $O_{1,i}$ é o grau de pertinência de um conjunto Fuzzy ϕ (A_1, A_2, B_1 ou B_2) e este especifica o grau que uma dada entrada x ou y satisfaz a quantificação de ϕ ; x e y são as entradas para a primeira camada; A_i e B_{i-2} são as variáveis lingüísticas semelhantes à “alta concentração” e “baixa concentração”.

A segunda camada utiliza neurônios que têm a função de fornecer o produto de todos os sinais de entrada como saída. Este termo é representado pela letra “w”, que, por analogia à rede neural, pode ser chamada de peso sináptico ou peso de excitação.

$$O_{2,i} = w_i = m_{A_i(x)} m_{B_i(y)}, \text{ } i = 1,2 \quad \text{eq. (60)}$$

Cada saída de um neurônio é o resultado da intersecção das funções de pertinência. A Figura 24 ilustra o processo realizado na segunda camada da rede Neuro-Fuzzy.

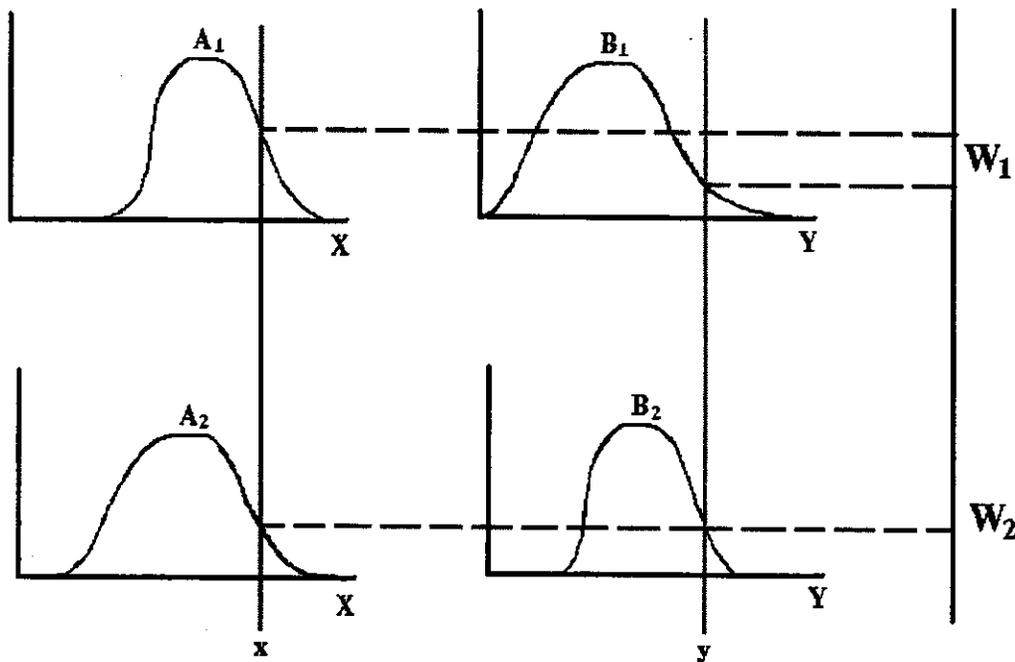


Figura 24. Processo de determinação dos pesos de excitação de uma rede Neuro-Fuzzy.

A terceira camada tem a finalidade de normalizar os w_i provenientes da camada anterior.

$$O_{3,i} = \bar{w}_i = \frac{w_i}{w_1 + w_2}, i = 1,2 \quad \text{eq. (61)}$$

Na quarta camada, cada termo da função $f_{(x,y)}$ é multiplicado por \bar{w}_i , ponderando desta maneira a contribuição de cada uma das equações para um determinado conjunto de entrada (x e y).

$$O_{4,i} = \bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i) \quad \text{eq. (62)}$$

Na camada de saída (quinta camada), temos a agregação de todas as informações na forma de um somatório de todos os neurônios da camada anterior.

$$O_{5,i} = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad \text{eq. (63)}$$

Para realizar o ajuste dos parâmetros da rede Neuro-Fuzzy, é utilizado um algoritmo híbrido que envolve o método dos mínimos quadrados e retropropagação dos erros. Para compreender como esse algoritmo funciona, iremos utilizar a arquitetura apresentada na Figura 23.

Inicialmente, devemos dividir nosso problema em duas partes. A primeira etapa é a otimização dos termos consequentes, onde os seus parâmetros possuem comportamento linear. A segunda etapa consiste em otimizar a premissa, que é composta de um conjunto de parâmetros não lineares (funções de pertinência). Portanto, para o treinamento híbrido, devemos inicialmente considerar que as premissas são fixas e otimizar os termos consequentes. Em um segundo momento, o processo inverso deve ser realizado, ou seja, os parâmetros das premissas são otimizados e os termos consequentes são mantidos fixos.

Quando os valores dos parâmetros das premissas são fixos, a saída geral da arquitetura apresentada na Figura 23, pode ser expressa como uma combinação linear dos parâmetros consequentes:

$$F = \frac{w_1}{w_1 + w_2} f_1 + \frac{w_2}{w_1 + w_2} f_2 \quad \text{eq. (64)}$$

$$F = \bar{w}_1 (p_1 x + q_1 y + r_1) + \bar{w}_2 (p_2 x + q_2 y + r_2) \quad \text{eq. (65)}$$

$$F = (\bar{w}_1 x)p_1 + (\bar{w}_1 y)q_1 + (\bar{w}_1)r_1 + (\bar{w}_2 x)p_2 + (\bar{w}_2 y)q_2 + (\bar{w}_2)r_2 \quad \text{eq. (66)}$$

Observando a equação 66, podemos verificar que os pesos de excitação e os valores de entrada possuem uma relação linear com os parâmetros dos termos conseqüentes p_1, q_1, r_1, p_2, q_2 e r_2 . Portanto, a função dos mínimos quadrados é otimizar estes parâmetros, com o objetivo de obter o menor erro na saída da rede.

Em um segundo momento, considera-se os parâmetros lineares otimizados e fixos e inicia-se a otimização dos parâmetros não lineares da premissa⁽⁷²⁾. Para a otimização dos parâmetros não lineares, é utilizado o método do gradiente descendente, de maneira análoga ao apresentado para a rede neural.

4.7 Interpretação das Regras Fuzzy de Inferência

Nesta seção, desejamos apresentar uma abordagem alternativa para interpretar como as regras Fuzzy de inferência agem durante o processo de modelagem de um determinado sistema. Para alcançarmos esse objetivo, iremos partir de uma rede Neuro-Fuzzy com arquitetura similar à apresentada na Figura 25.

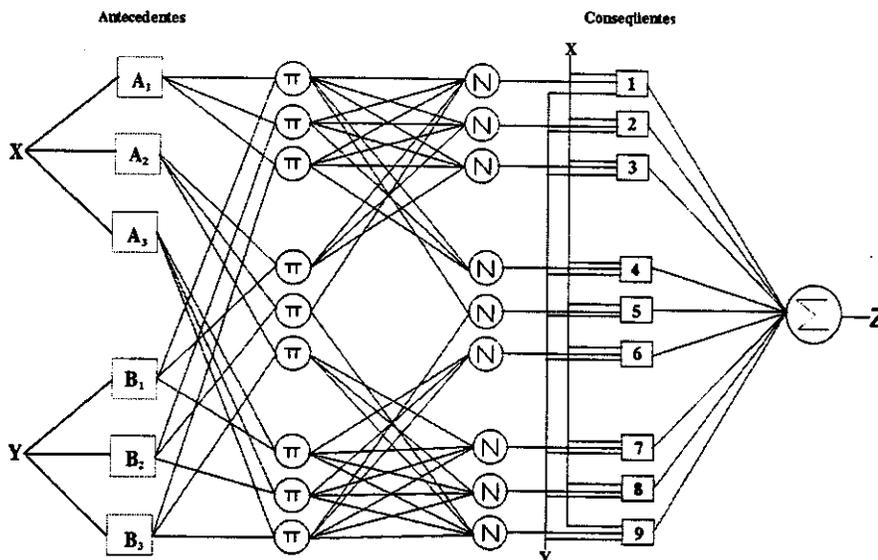


Figura 25. Arquitetura de uma rede Neuro-Fuzzy.

A Figura 25 ilustra uma arquitetura Neuro-Fuzzy com duas entradas (x e y) e uma saída (z). A cada uma das entradas são associadas três funções de pertinência, as quais são representadas pelas letras A e B e com os sub-índices 1, 2 e 3. A combinação entre as funções de pertinência gera um total de 9 regras, sendo responsáveis pelo mapeamento do espaço. A Figura 26 mostra como as funções de pertinência são combinadas para realizar o mapeamento do sistema.

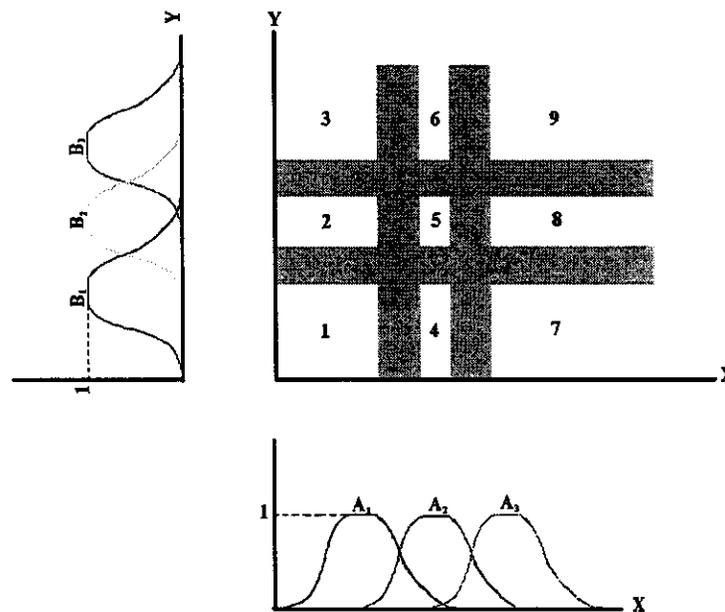


Figura 26. Esquema do mapeamento formado pelas funções de pertinência de cada uma das entradas.

O espaço de mapeamento sofre uma granularização, a qual é definida pelos termos lingüísticos associadas a cada um dos domínios das variáveis de entrada. A partição pode ser uniforme (com todos os termos lingüísticos iguais e equidistantes) ou não uniforme⁽³⁷⁾.

O número de termos lingüísticos (funções de pertinência) define a granularidade das partições de cada universo. Um número pequeno de termos lingüísticos define uma partição esparsa, enquanto um número maior resulta numa partição fina. Portanto, quando utilizamos um alto número de funções de pertinência, o mapeamento torna-se mais robusto e eficiente, pois este é responsável por modelar apenas uma restrita região do espaço.

Uma das grandes vantagens de utilizar as regras para a modelagem de um sistema está associada ao fato de que cada uma das regras que compõem o sistema de mapeamento deste sistema é independente entre si, ou seja, associa ao sistema uma maior flexibilidade de ajuste, pois não é necessário fazer o ajuste das demais regras.

A partição do universo também pode ser vista como uma forma de compressão de dados. Utilizando agrupamentos Fuzzy com informações similares, pode-se esconder parte da informação inútil, indesejada ou redundante. Assim, a granularização pode ser usada para focar a análise de acordo com o interesse do usuário.

Na Figura 26, pode-se observar a partição do espaço em nove regiões distintas, onde o limite de separação de cada região é delimitado por uma região sombreada. A largura da região sombreada está associada ao grau de sobreposição entre as funções de pertinência de um determinado domínio. Analisando a Figura 26, podemos notar que a região central da figura é a que possui a maior área sombreada. Esse comportamento é devido ao alto grau de sobreposição das funções de pertinência A_2 e B_2 . Para ambas as funções de pertinência, pode ser verificado que existe uma sobreposição à direita e à esquerda da função. Portanto, podemos concluir que esta região é a que possui a maior habilidade de modelar as incertezas contidas neste sistema. Isso implica que as regiões de maior sobreposição são responsáveis pelo ajuste mais fino do modelo⁽³⁷⁾.

5. Introdução à Teoria do Infravermelho Próximo

5.1 Espectroscopia na Região do Infravermelho Próximo

A origem da espectroscopia no infravermelho, data do início do século dezenove, com o trabalho pioneiro do músico e astrônomo alemão Frederick William Herschel. Durante sua adolescência, Herschel despertou um grande interesse pela astronomia, motivando-o a construir um telescópio para observar os corpos celestes. Inicialmente, a astronomia era apenas um passatempo que Herschel e sua irmã cultivavam⁽⁷⁹⁾. Entretanto, com o passar dos anos, Herschel começou a realizar estudos sérios que consistiam essencialmente no mapeamento dos corpos celestes. Estes estudos resultaram na descoberta de várias estrelas e nebulosas, mas sua grande descoberta no campo da astronomia foi o planeta Urâno em 1781. O interesse pela astronomia despertou a curiosidade de Herschel com relação às propriedades físicas da radiação eletromagnética na região do visível. Herschel acreditava que a compreensão das propriedades físicas da luz poderia ajudá-lo em seus estudos dos corpos celestes.

Em 1800, durante a execução de um experimento que consistia em utilizar um prisma para separar as faixas espectrais associadas à região espectral do visível, Herschel observou que cada uma das cores decompostas pelo prisma apresentava uma quantidade de energia distinta. Para monitorar a quantidade de energia associada a cada cor, ele teve a brilhante idéia de utilizar um termômetro. Para sua surpresa, ele verificou que abaixo do vermelho, onde não havia luz visível, era a região que apresentava a maior temperatura⁽⁷⁹⁾.

O experimento de Herschel foi importante não somente pela descoberta do infravermelho, mas também porque ele demonstrou que existem formas de luz que não podem ser observadas pelo olho humano.

A Figura 27 ilustra o equipamento proposto criado por Herschel. A partir deste estudo, outros trabalhos foram desenvolvidos utilizando outras faixas espectrais tais como: ultravioleta, infravermelho próximo, médio e longínquo.



Figura 27. Este desenho ilustra o experimento de Herschel para medir a energia associada à luz branca decomposta por um prisma.

O desenvolvimento da espectroscopia no infravermelho foi alavancado pela potencialidade que esta técnica apresentou na caracterização e quantificação de diferentes espécies químicas. Inicialmente, devido às limitações instrumentais, os trabalhos envolvendo espectroscopia em química restringiram-se basicamente à identificação e quantificação de algumas poucas espécies químicas em casos bem específicos⁽⁸⁰⁾. Posteriormente, com o desenvolvimento de equipamentos mecânicos e ópticos mais precisos, o número de aplicações multiplicou-se.

Em uma primeira fase, as aplicações qualitativas concentraram-se principalmente na faixa espectral conhecida como região do infravermelho médio, pois, nesta região, podem-se observar claramente as bandas de absorção de grupos orgânicos específicos como: N-H, C-H, O-H, C-C, etc. Desta maneira, esta técnica foi largamente empregada pelos químicos orgânicos para auxiliar na caracterização de diversas substâncias químicas⁽⁷⁹⁾. Por outro lado, as aplicações quantitativas concentraram-se basicamente na região do visível, devido ao fato de que uma série de sistemas químicos respeitam a lei de Beer-Lambert.

Somente, a partir de 1970 iniciou-se uma nova fase dos estudos espectroscópicos quantitativos, promovidos pelo desenvolvimento dos espectrofotômetros com transformada de

Fourier, da informática, do interfaceamento dos instrumentos eletrônicos com os computadores pessoais e da introdução de recursos matemáticos mais sofisticados.

A partir destas inovações tecnológicas, os estudos quantitativos expandiram suas fronteiras para as regiões espectrais no infravermelho próximo e médio. Isso desencadeou um surpreendente interesse de vários grupos de pesquisas acadêmicos, governamentais e industriais, na tentativa de desenvolver metodologias de análise. A consequência foi um impacto direto no crescimento da produção e produtividade industrial, redução de gastos, minimização de resíduos industriais, etc. A iniciativa da comissão de grãos canadenses de aprovar a substituição do método de Kjeldahl (método padrão para determinação de proteínas em grãos) por uma metodologia baseada na espectroscopia no infravermelho próximo (NIR – *Near Infrared*), permitiu que, no ano de 1995, o Canadá economizasse mais de \$ 2,5 milhões de dólares e evitasse a geração de 47 toneladas de resíduos cáusticos⁽⁷⁹⁾.

Os primeiros trabalhos que realmente proporcionaram o interesse pelo estudo da espectroscopia no infravermelho próximo como ferramenta de análise industrial foram desenvolvidos na década de 1970, pelo grupo de pesquisa do Professor Karl Norris, quando ele era responsável por um grupo de pesquisa do Departamento de Agricultura dos Estados Unidos da América. Entretanto, as limitações tecnológicas da época não permitiram o desenvolvimento vertiginoso do NIR como atualmente é observado em diversos setores industriais como agrícola, petroquímico, alimentício, farmacêutico, etc⁽⁸¹⁾.

A espectroscopia no infravermelho próximo é uma técnica particularmente útil para o estudo quantitativo de substâncias orgânicas, embora o espectro contenha essencialmente informações dos grupos C-H, N-H, O-H e S-H⁽⁷⁹⁾. Além disso, o espectro no infravermelho próximo é dependente de parâmetros físicos, como tamanho das partículas, polimorfismo, solventes residuais, umidade, temperatura, pressão, densidade, textura superficial, forma, compactação, etc. Por esta razão, uma comparação direta entre os espectros é inviável de ser realizada. Em contrapartida, esta característica do infravermelho próximo permite que seja possível estimar propriedades físicas e químicas de um vasto número de materiais.

Em materiais sólidos granulares, a interação da radiação eletromagnética na faixa do infravermelho próximo pode ser interpretada da seguinte forma: cada partícula no sistema granular age como um ponto de espalhamento, onde a interação partícula – luz é afetada pelas partículas vizinhas. Geralmente, quando o feixe de luz incide sobre a superfície rugosa da amostra, pode

ocorrer os fenômenos de reflexão, transmissão, refração e absorção da luz incidente. A proporção relativa de cada fenômeno depende somente das propriedades ópticas do material, mas alguns parâmetros como tamanho de partícula, forma, densidade de empacotamento e orientação da amostra, podem afetar os valores de reflectância e intensidade da banda de absorbância π . A luz espalhada pelo material é uma medida que contém uma componente especular e outra difusa. O comportamento óptico é uma função direta do coeficiente de espalhamento e absorção do material sólido, além disso, estes coeficientes são dependentes do comprimento de onda. O espalhamento múltiplo atenua parte da energia transmitida através da partícula. Como resultado, os efeitos de espalhamento e transmissão tendem a ser opostos entre si, principalmente no caso de partículas transparentes e semitransparentes. Os raios que emergem da superfície, após terem penetrado no material sólido, sofrem uma seletiva absorção e assim contêm informação sobre a amostra. Este método oferece um espectro de boa qualidade e permite que sejam realizadas determinações qualitativas e quantitativas.

5.2 Principais Fatores que Influenciam nos Trabalhos Quantitativos na Região do Infravermelho Próximo

A *temperatura* é uma variável importante para soluções aquosas. A variação de poucos graus na temperatura da solução pode resultar em mudanças substanciais no perfil do espectro. Este parâmetro é menos crítico para espectros de reflectância NIR de sólidos. Entretanto, em trabalhos de alta precisão este parâmetro deve ser considerado.

Umidade e solventes residuais são parâmetros que devem ser avaliados durante a execução de um trabalho utilizando a espectroscopia NIR.

Em trabalhos de determinação de água, a umidade da sala de aquisição dos espectros deve ser constante. Além disso, é necessário um cuidado adicional com relação à contaminação de solventes residuais na amostra, pois estes podem contribuir para mudanças significativas no perfil do espectro.

Em trabalhos que envolvem o uso de amostras sólidas, a *espessura* é outro fator que influencia diretamente o espectro. A variação na espessura das amostras implica na variação do caminho óptico. Isso ocorre porque a radiação NIR é absorvida muito menos do que a radiação no infravermelho médio e por este motivo a luz pode penetrar vários milímetros no interior da

amostra. Este fator é mais sensível quando a amostra possui uma espessura na ordem de milímetros.

Em amostras sólidas o tamanho, a compactação e a superfície das amostras devem ser semelhantes às das amostras padrões. Caso contrário, isso pode implicar em diferença no espalhamento da luz, o que provoca mudanças significativas do espectro.

A compactação da amostra pode mudar sua superfície, adicionando uma componente de reflectância especular no espectro de reflectância NIR. A reflectância especular pode distorcer os picos de reflectância difusa, interferindo na realização de trabalhos quantitativos.

A **estrutura polimórfica** de algumas espécies químicas pode ser observada, porque a reflectância NIR mede diretamente substâncias sólidas cristalinas; portanto, podem ser observadas variações na estrutura cristalina.

5.3 Princípios Físico-Químicos envolvidos na Espectroscopia do Infravermelho Próximo

Para compreendermos as leis que regem a espectroscopia vibracional é necessário utilizarmos a lei de Hooke e o princípio de Franck-Condon⁽⁸²⁾.

A lei de Hooke utiliza o modelo mecânico clássico para explicar os movimentos dos corpos. Esta mesma abordagem pode ser aplicada para calcular as vibrações fundamentais de uma molécula diatômica na região do infravermelho médio. A Figura 28 ilustra uma molécula diatômica hipotética, a qual podemos utilizar para fazer uma analogia com um oscilador harmônico.

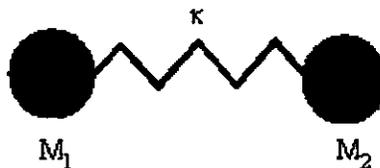


Figura 28. Representação de um oscilador diatômico.

A energia potencial do oscilador harmônico pode ser expressa como:

$$E_p = \frac{1}{2}k(r - r_e)^2 \quad \text{eq. (67)}$$

onde r é a distância internuclear entre os dois átomos, r_e é a distância internuclear de equilíbrio entre os dois átomos e k é a constante de força.

Para a análise da região do infravermelho próximo, a lei de Hooke não se aplica devido ao fato de que essa região é composta de bandas de combinação e sobretons. Desta forma, torna-se imprescindível aplicar o princípio de Franck-Condon, o qual considera a anarmonicidade do sistema.

As bandas de combinação surgem quando uma molécula poliatômica possui dois ou mais modos vibracionais que podem interagir entre si. A banda de combinação resultante é posicionada na frequência que possui valor igual à soma de cada uma das frequências dos módulos vibracionais envolvidos na interação. Por exemplo, a banda de combinação da água ocorre em 1940 nm, a qual é formada pela interação entre os módulos vibracionais de um estiramento assimétrico a 3500 cm^{-1} e pela deformação angular a 1645 cm^{-1} da ligação O-H da molécula de água. Portanto, ao realizarmos a soma desses dois módulos vibracionais ($3500 + 1645 = 5145\text{ cm}^{-1}$) teremos que a banda de combinação da água deve surgir a 1943 nm.

Os sobretons são transições entre estados vibracionais não consecutivos. As bandas provenientes desse tipo de transição são caracterizadas por apresentarem uma intensidade de 10–100 vezes mais fracas do que a transição entre o estado fundamental e o primeiro estado excitado.

5.3.1 Modelo Mecânico da Vibração Molecular

Por analogia à lei de Hooke, pode-se considerar que uma molécula diatômica é composta de dois corpos sólidos (átomos) de massas (M) distintas e um cabo (ligação química) que os conectam com uma constante de força (κ).

No modelo mecânico, os átomos vibram com uma frequência (ν_0), podendo ser representada pela equação 68:

$$\nu_0 = \frac{1}{2\pi} \sqrt{\frac{\kappa}{\mu}} \quad \text{eq. (68)}$$

$$\mu = \frac{M_1 M_2}{M_1 + M_2} \quad \text{eq. (69)}$$

onde μ é a massa reduzida.

Na espectroscopia vibracional, o movimento dos átomos de uma molécula está relacionado à energia da molécula. Portanto, para que ocorra variação da energia da molécula, é necessário que a frequência de vibração entre os átomos varie.

A mudança da frequência da vibração de uma ligação está relacionada o comprimento de onda da radiação que incide sobre a molécula.

As vibrações das moléculas podem ser descritas pelos estiramentos ou deformações angulares entre os átomos de uma ligação química. O estiramento é promovido durante a mudança da distância interatômica dos átomos de uma ligação química, enquanto a deformação angular envolve a mudança no ângulo da ligação entre átomos.

Como havíamos comentado anteriormente, o modelo mecânico proposto segue a lei de vibração harmônica. Entretanto, em sistemas reais as moléculas não são osciladores harmônicos perfeitos, devido à repulsão Coulômbica entre os átomos e/ou dissociação das ligações químicas. Estes fenômenos físicos são responsáveis pela anarmonicidade mecânica da molécula. Por este motivo, é possível observar as bandas de combinações e sobretons na região do infravermelho próximo. Além disso, a intensidade dos sobretons e bandas de combinação dependem do grau de anarmonicidade da molécula⁽⁸³⁻⁸⁴⁾.

Existem dois tipos de anarmonicidade, a mecânica e a elétrica. A primeira está relacionada com o fato de que o espaçamento entre os níveis eletrônicos não é constante. Além disso, existe uma distância entre dois átomos que levará à dissociação da ligação química entre eles. A energia potencial de dissociação de uma ligação química é descrita pela equação de Morse.

$$E_p = D_e (1 - e^{-\beta(r-r_e)})^2 \quad \text{eq. (70)}$$

onde D_e é a energia de dissociação da ligação; β é uma constante; r_e a distância internuclear de equilíbrio; r a distância internuclear

Para que ocorra a dissociação da ligação, é necessário que a energia potencial da molécula seja igual à energia de dissociação. Esta condição é satisfeita quando a diferença entre a distância internuclear de equilíbrio e a internuclear é suficientemente grande para que o termo $e^{-\beta(r-r_e)}$ tenda a zero.

A anarmonicidade elétrica está associada às transições vibracionais que se caracterizam por apresentarem uma transição entre dois níveis vibracionais não consecutivos (sobretons). Este comportamento tem influência direta no momento de dipolo elétrico da molécula.

5.3.2 Princípio de Franck–Condon

O princípio de Franck-Condon diz: “Sabe-se que quando uma molécula vibra, a probabilidade de encontrarmos um determinado átomo em um certo ponto é inversamente proporcional à sua velocidade no ponto. Portanto, o átomo em uma vibração molecular permanece um tempo maior nas configurações com energia cinética menor. As configurações de menor energia são aquelas nas quais a energia potencial é aproximadamente idêntica à energia total do sistema ou da intersecção do nível da energia vibracional com a superfície de energia potencial da molécula. Assim, os fótons são mais prováveis de serem absorvidos quando os núcleos estão estacionários ou com um movimento lento. Além disso, a excitação resultante da absorção do fóton não pode ser transferida imediatamente para o núcleo. Portanto, o núcleo irá continuar a movimentar-se lentamente, imediatamente após o processo de absorção. Assim, no estado excitado, a configuração nuclear também tenderá a estar próxima da intersecção do nível de energia vibracional com a superfície de energia potencial excitada. Portanto, a transição tende a tomar lugar entre o nível vibracional na qual a configuração nuclear é a mesma em ambos os estados, e elas tendem a ocorrer quando a energia cinética nuclear é pequena”.

O modelo da lei de Hooke anteriormente descrito considera que a energia da ligação entre os dois corpos segue um comportamento contínuo. Entretanto, a níveis atômicos, o sistema tem um comportamento discreto, o qual é melhor representado pela mecânica quântica.

Segundo a mecânica quântica, a variação de energia de uma ligação química possui comportamento discreto e não contínuo, como o observado para sistemas macroscópicos. No modelo quântico, a energia de vibração é representada, conforme mostrado na equação 71:

$$E_v = \left(v + \frac{1}{2} \right) \frac{h}{2\pi} \sqrt{\frac{k}{\mu}} \quad \text{eq.(71)}$$

onde v é o número quântico de vibração (0,1 ,2 ,3 , ...) e h é a constante de Plack.

Para calcular a energia necessária para a transição entre os níveis vibracionais, devemos utilizar a equação 72:

$$\Delta E_{vib} = h\nu = h\left(\frac{c}{\lambda}\right) = hc\tilde{\nu} \quad \text{eq. (72)}$$

onde

$$\frac{1}{\lambda} = \tilde{\nu}(cm^{-1}) = \frac{1}{2\pi c} \sqrt{\frac{k(M_1 + M_2)}{M_1 * M_2}} \quad \text{eq. (73)}$$

Devido ao fato do modelo que descreve o comportamento das moléculas não ser um oscilador harmônico perfeito, torna-se possível observar as bandas na região do infravermelho próximo.

Quando a anarmonicidade é considerada, a equação 72 é expressa da seguinte forma:

$$\Delta E_{vib} = h\nu[1 - (2\nu + \Delta\nu + 1)y] \quad \text{eq. (74)}$$

onde y é um fator de anarmonicidade.

A anarmonicidade mecânica no infravermelho próximo está relacionada ao fato de que as ligações envolvendo átomos de hidrogênio vibram com maior amplitude devido à baixa massa do átomo de hidrogênio. Portanto, as bandas de absorção observadas na região do infravermelho próximo surgem dos sobretons das vibrações de estiramento envolvendo grupos funcionais com átomos de hidrogênio ou combinação de módulos vibracionais de estiramento e deformação angular.

A Tabela 5 apresenta as principais regiões espectrais do infravermelho próximo e a natureza da transição vibracional associada a elas⁽⁸⁰⁾.

As bandas observadas na região do infravermelho próximo são sobretons ($\Delta\nu = \pm 2, \pm 3, \dots$) e combinações do nível fundamental de vibrações ($\Delta\nu = \pm 1$) com os sobretons. As probabilidades das transições envolvidas nos sobretons são menores que as das transições do nível fundamental, e por este motivo, a intensidade do sinal é menor que o observado na região do infravermelho médio. Devido a este fato, esta região espectral pode ser utilizada na análise de amostras biológicas, pois a intensidade da banda de água é menor, permitindo desta maneira que importantes informações sobre a amostra possam ser extraídas.

Tabela 5. Atribuição das regiões espectrais às possíveis transições vibracionais.

Região Espectral (nm)	Natureza da transição vibracional
2200 – 2450	Estiramento da combinação C-H
2000 – 2200	Estiramento da combinação N-H, O-H
1650 – 1800	Primeiro sobretons do estiramento C-H
1400 – 1500	Primeiro sobretons do estiramento N-H, O-H
1100 – 1225	Segundo sobretons do estiramento C-H
950 – 1100	Segundo sobretons do estiramento N-H, O-H
850 – 950	Terceiro sobretons do estiramento C-H
775 - 850	Terceiro sobretons do estiramento N-H, O-H

6. Simulação Teórica

6.1 Modelagem de um Sistema Não-Linear

Para comparar o desempenho das técnicas de inteligência artificial em casos extremos de complexidade, foi proposta uma simulação teórica de um sistema com comportamento não linear.

O objetivo desta simulação foi criar um modelo matemático que fosse capaz de estabelecer uma relação entre dois parâmetros X e Y (variáveis independentes) e um outro parâmetro Z (variável dependente).

Neste exemplo, optamos por trabalhar com somente duas variáveis, para que o leitor possa visualizar em um gráfico tridimensional a complexidade da superfície a ser modelada e a habilidade das técnicas de inteligência artificial em aproximar funções de alta complexidade. Entretanto, vale salientar que o mesmo processo pode ser aplicado para sistemas com muitas variáveis.

Para a simulação, utilizou-se a função matemática que gerou a superfície de resposta da Figura 29. A função matemática é a seguinte:

$$Z = (3 \times (1 - X)^2 \times \exp(-X^2 - (Y + 1)^2)) - (10 \times (\frac{X}{5} - X^3 - Y^5) \times \exp(-X^2 - Y^2)) - (\frac{1}{3} \times \exp(-(X + 1)^2 - Y^2))$$

Os valores dos parâmetros X e Y variaram entre -3 e +3, conforme pode ser observado na superfície de resposta.

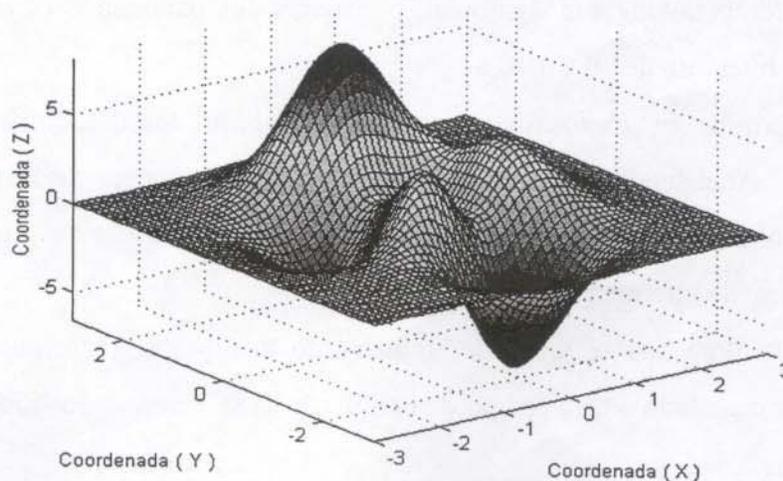


Figura 29. Superfície de resposta utilizada na simulação.

Para a simulação respeitar condições similares às obtidas nos experimentos reais, acrescentou-se ruídos aleatórios aos valores dos parâmetros X, Y e Z da ordem de $-0,04$ a $+0,04$.

6.2 Desenvolvimento dos Modelos de Inteligência Artificial

Para o treinamento e teste da rede neural, Neuro-Fuzzy e neurogenética, foi gerada uma superfície com 2000 pontos, semelhante à apresentada na Figura 29. Posteriormente, os pontos que compõem a superfície foram ordenados em uma matriz de variáveis independentes, onde cada elemento da matriz está associada aos valores das coordenadas (x,y) da superfície gerada. Os valores da coordenada z foram utilizados para criar um vetor de variáveis dependentes.

Para o treinamento, foram selecionados 1600 pontos igualmente distribuídos ao longo de toda a superfície. Os restantes 400 pontos foram empregados para a avaliação do modelo.

6.2.1 Rede Neural

Para o ajuste dos pesos que conectam os neurônios, utilizaram-se dois diferentes métodos de treinamento, Levenberg-Marquardt e retropropagação de erros. Para ambos modelos foi utilizada a mesma arquitetura e idênticos critérios de finalização. Os critérios de convergência adotados foram um erro inferior a 10^{-5} ou um número máximo de iterações igual a 1000.

Para o treinamento da rede neural, foram utilizados dois parâmetros de entrada (as coordenadas X e Y) e uma saída (coordenada Z). Devido ao fato de termos decidido utilizar a função de transferência tangente sigmoideal, os valores dos parâmetros de entradas e saída foram escalonados no intervalo de $-0,8$ a $+0,8$.

O número de neurônios da camada intermediária foi determinado pelo processo de tentativa e erro. Analisando o valor do erro no final do processo de treinamento e a taxa de aprendizagem, concluímos que a arquitetura mais eficiente é obtida quando empregamos 25 neurônios na camada intermediária.

Para cada uma das técnicas de treinamento dos pesos, foi cronometrado o tempo de processamento e calculado o erro padrão de calibração (SEC) e previsão (SEP). O parâmetro SEP é definido como:

$$SEP = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad \text{eq. (75)}$$

onde n é o número de amostras do conjunto teste; y_i é o valor conhecido da i -ésima amostra; \hat{y}_i é o valor previsto para a i -ésima amostra.

A equação para calcular o valor do SEC é similar à apresentada para calcular o SEP. A única diferença é que o SEC é calculado utilizando as amostras do conjunto de calibração, enquanto que no SEP, o erro é calculado a partir das amostras do conjunto de previsão.

6.2.2 Rede Neurogenética

A arquitetura da rede neurogenética foi inicialmente otimizada por tentativa e erro, obtendo-se a seguinte configuração: 2 neurônios de entrada, 25 neurônios na camada intermediária com função de transferência tangente sigmoideal, um neurônio na camada de saída com função de transferência linear e ligações direta entre as entradas e o neurônio de saída. Os parâmetros do programa para a seleção do melhor conjunto de pesos foram inicializados para uma população inicial de 100 cromossomos, probabilidade de mutação de 1% e probabilidade de cruzamento de 90%. Como critérios para finalização do programa, estabeleceu-se um número máximo de 1000 iterações ou um erro mínimo de 10^{-5} .

6.2.3 Rede Neuro-Fuzzy

Semelhante ao processo adotado para as outras duas técnicas, o processo de otimização da arquitetura da rede Neuro-Fuzzy foi realizada por tentativa e erro. O melhor desempenho foi alcançado com dois neurônios na camada inicial. A cada neurônio, foram associadas 8 funções de pertinência do tipo sino generalizada. Para o ajuste das regras foi utilizado o método do gradiente descendente.

O programa foi configurado para terminar o processo de otimização após atingir 1000 iterações ou alcançar um erro mínimo de 10^{-5} .

Os programas de rede neural e Neuro-Fuzzy utilizados nesta tese fazem parte do toolbox de rede neural e lógica Fuzzy do Matlab 5.2. O programa de rede neurogenética foi desenvolvido pelo

autor deste trabalho, utilizando o ambiente Matlab 5.2. Todos os programas foram executados em um Pentium III 800 Mhz, com 128 Mb de memória RAM.

6.3 Resultados obtidos com os Modelos Desenvolvidos para os dados Simulados

A seguir, serão apresentados os resultados obtidos para a modelagem dos dados simulados.

6.3.1 Modelagem Aplicando a Técnica de Redes Neurais

Na modelagem com rede neural, optamos por avaliar dois métodos de treinamento distintos: Levenberg-Marquart e retropropagação dos erros. Segundo a literatura, estas duas técnicas possuem desempenhos similares; entretanto, o método de Levenberg-Marquart converge mais rápido para um ponto de máximo ou mínimo.

Este estudo comparativo servirá de base para escolher a técnica de treinamento que será aplicada para a modelagem dos sistemas reais, propostos neste trabalho.

Os resultados obtidos na modelagem dos dados simulados estão apresentados na Tabela 6. Essa tabela mostra o resultado do melhor modelo de calibração e o valor do erro médio dos resultados, após executar o programa de treinamento 10 vezes.

Tabela 6. Resultados do desempenho dos modelos obtidos com o método de Levenberg-Marquardt e Retropropagação dos erros.

	Levenberg-Marquardt	Retropropagação dos erros
SEC ¹	0,0691	0,0892
SEC ² médio	0,0712	0,0913
SEP	0,1071	0,1234
SEP médio	0,1106	0,1301
Tempo de convergência	≈ 2 minutos	≈ 4 minutos
Tempo de convergência médio	≈ 2 min. e 30seg.	≈ 5 minutos
Tempo de processamento	≈ 1 hora e 15 minutos	≈ 1 hora e 23 minutos

1. SEC do melhor modelo de calibração.
2. SEC médio dos 10 modelos de calibração.

Partindo do modelo de calibração desenvolvido com rede neural, realizou-se a previsão dos padrões de treinamentos. A Tabela 6 mostra que os resultados dos erros obtidos utilizando o método de Levenberg-Marquart e retropropagação são similares e parece não haver vantagem da utilização de uma em relação à outra.

Para avaliar o desempenho do modelo gerado, empregou-se um conjunto externo de 400 novas amostras. Estas amostras foram geradas para estarem distribuídas de maneira uniforme ao longo do espaço estudado.

As novas amostras foram utilizadas para prever o valor do parâmetro Z , para os respectivos valores de X e Y de entrada. Com os valores previstos do parâmetro Z , gerou-se uma superfície de resposta, conforme ilustrada na Figura 30.

Como podemos observar, apesar da nova superfície de resposta possuir um número menor de pontos, essa ainda conservou as mesmas características da superfície de resposta de treinamento apresentada na Figura 29. É possível observar a presença de três regiões de máximos e mínimos, indicando que o mapeamento da função para a região de interesse procedeu de maneira apropriada aos objetivos deste trabalho.

A rede neural mostrou ser capaz de estabelecer uma função de mapeamento que pode ser aplicada para prever novas amostras contidas na região do mapeamento das amostras de calibração. Neste caso, o erro de previsão da rede é consideravelmente baixo, conforme podemos observar na Figura 30.

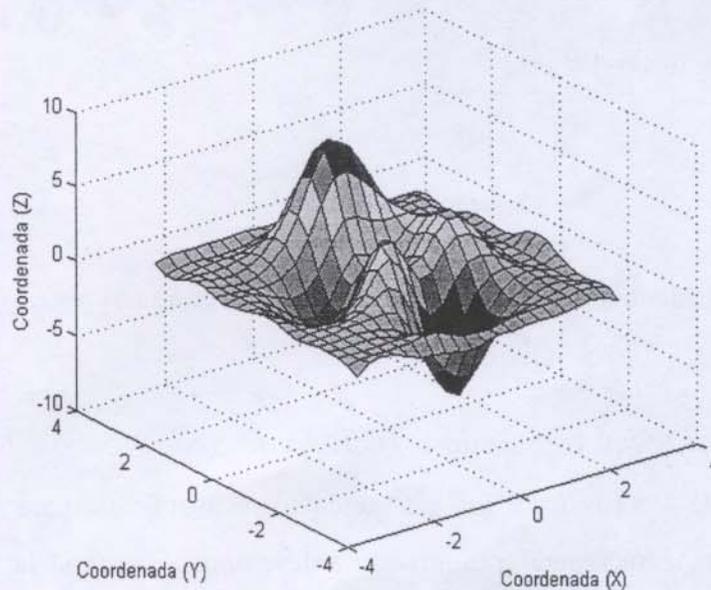


Figura 30. Superfície de resposta do conjunto de 400 amostras de validação do modelo da rede neural.

A Figura 31 mostra o resíduo não modelado pela rede, que é composto de informações do sistema e de ruído aleatório. Neste caso, o resíduo ao redor das regiões de fronteira da superfície, contém uma quantidade significativa de informações, quando comparado com o ruído aleatório inserido aos dados.

A região central da superfície apresenta um resíduo menor que as regiões de fronteira. Além disso, se levarmos em consideração que o valor absoluto dos pontos da superfície na região central são significativamente superiores ao valor do resíduo, chegaremos a conclusão que o resíduo não modelado pode ser desprezado. Isso indica que o modelo construído apresenta um bom desempenho para prever amostras que não estão ao redor das regiões de fronteira da superfície.

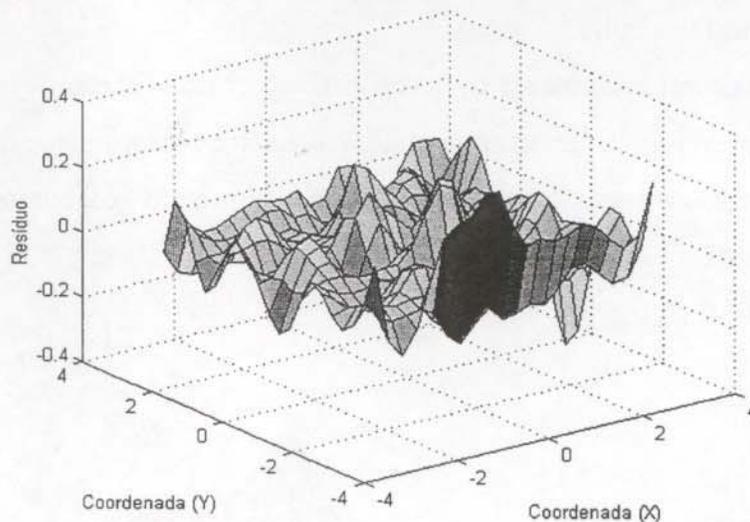


Figura 31. Resíduo não modelado das amostras do conjunto teste.

6.3.2 Modelagem Aplicando a Técnica de Rede Neuro-Fuzzy

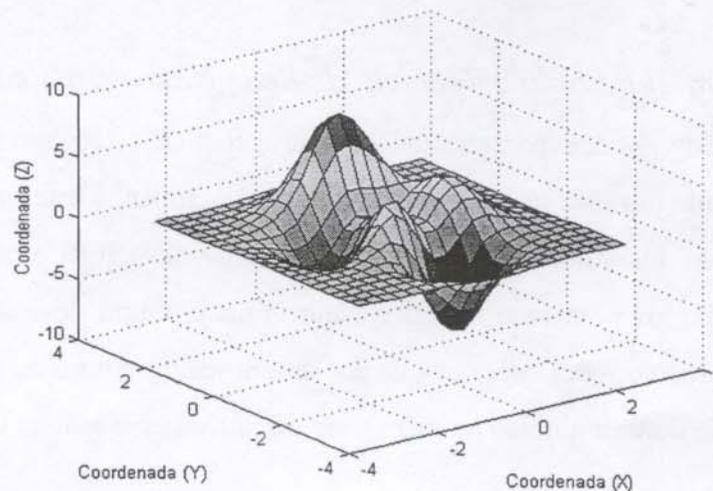
A rede Neuro-Fuzzy foi a segunda técnica aplicada para modelar os dados. Segundo a literatura, esta técnica apresenta um ótimo desempenho na modelagem de dados com alto grau de não linearidade. A Tabela 7 apresenta os resultados obtidos para os mesmos conjuntos de dados empregados no modelo da rede neural.

Tabela 7. Resultados do desempenho do modelo obtido com o programa Neuro-Fuzzy.

	Neuro-Fuzzy
SEC	0,0066
SEC Médio	0,0075
SEP	0,0099
SEP Médio	0,0115
Tempo de convergência	≈ 3 minutos
Tempo de convergência médio	≈ 4 minutos
Tempo de processamento	≈ 1 horas 40 minutos

Ao comparar os resultados das Tabelas 6 e 7, verificamos que o modelo Neuro-Fuzzy apresentou um desempenho superior ao obtido pela rede neural.

A Figura 32 mostra as amostras do conjunto teste, onde pode-se verificar o ótimo modelamento, sem problemas de bordas⁽⁵¹⁾. Podemos concluir, a partir desses resultados, que o modelo Neuro-Fuzzy realmente é um modelo mais eficiente para dados altamente complexos e possui uma taxa de convergência mais rápida que a rede neural. Estas conclusões são válidas se considerarmos que foi empregada a melhor arquitetura para cada um dos modelos.

**Figura 32.** Superfície de resposta do conjunto de 400 amostras de validação do modelo da rede Neuro-Fuzzy.

A Figura 33 ilustra o resíduo não modelado pela rede Neuro-Fuzzy. Como pode ser observado, o resíduo está na mesma ordem de grandeza dos valores do ruído aleatório inserido nos dados empregados neste estudo. Como a quantidade de informação não modelada não é

significativa, observamos que o modelo Neuro-Fuzzy foi capaz de explicar quase 100% do comportamento deste sistema. A Figura 33 mostra ainda que o resíduo não modelado é similar para toda a superfície, reforçada a afirmação que este modelo é menos sensível ao problemas de modelagem das bordas.

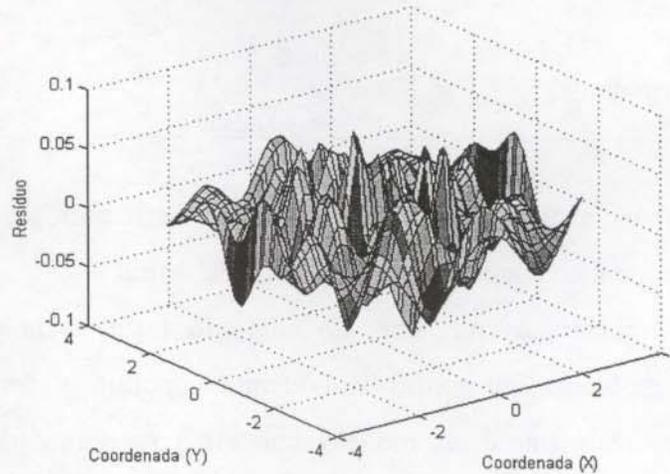


Figura 33. Resíduo das amostras do conjunto teste.

6.3.3 Modelagem Aplicando a Técnica de Rede Neurogenética

Os resultados obtidos para o modelo neurogenético estão apresentados na Tabela 8. O modelo desenvolvido por esta técnica apresentou um desempenho inferior, quando comparado aos modelos anteriores. Os valores de erro padrão de previsão (SEP) e calibração (SEC) são bem maiores que os obtidos pela rede neural e Neuro-Fuzzy. Outra desvantagem deste modelo é o tempo de processamento, bem superior que aos anteriores. O tempo de processamento necessário para a convergência do treinamento também é significativamente maior.

Tabela 8. Resultados do desempenho dos modelos obtidos com o programa Neurogenético.

	Rede Neurogenético
SEC	0,195
SEC Médio	0,201
SEP	0,231
SEP Médio	0,243
Tempo de convergência	≈ 11 minutos
Tempo de convergência médio	≈ 15 minutos
Tempo de processamento	≈ 43 horas 30 minutos
Tempo de processamento Médio	≈ 44 horas 25 minutos

Apesar do baixo desempenho, podemos verificar que o modelo é bastante robusto, pois é capaz de estimar valores do parâmetro Z para novas amostras com precisão similar às das amostras de calibração, conforme apresentado na Tabela 8 e ilustrado na Figura 34.

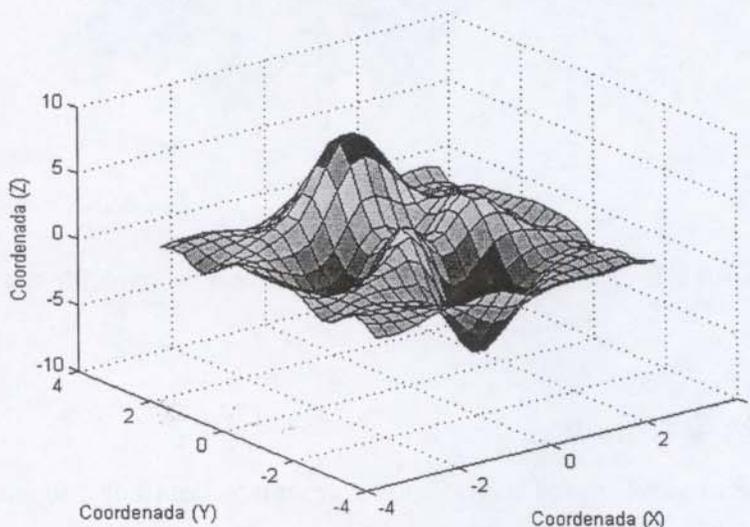


Figura 34. Superfície de resposta do conjunto de 400 amostras de validação do modelo da rede neurogenética.

A Figura 35 mostra o resíduo para as amostras do conjunto teste. O resíduo apresenta-se distribuído de maneira homogênea ao redor do zero para todas as amostras. Pelo gráfico da Figura 35, não é possível observar nenhum tipo de erro sistemático do modelo.

Para o modelo neurogenético, é difícil afirmar a partir das informações obtidas, que o modelo apresenta o problema do modelamento de bordas⁽⁴⁸⁾, uma vez que a superfície de resíduo é mais ou menos homogênea ao longo de todo o espaço mapeado. Entretanto, acreditamos que esta técnica apresenta o mesmo problema, visto que as funções de mapeamento são as mesmas empregadas pela rede neural treinada com o método de Levenberg-Marquart e retropropagação dos erros.

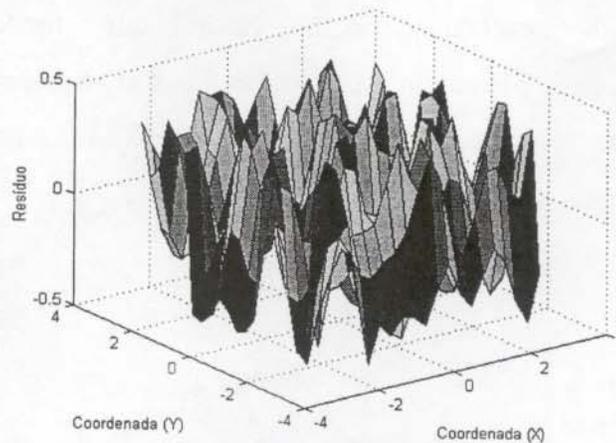


Figura 35. Resíduo das amostras do conjunto teste para o treinamento com rede neurogenética.

6.4 Considerações Finais

De maneira geral, todos os modelos apresentam uma boa robustez, que é constatada pela similaridade entre os erros das amostras de previsão e calibração. Entretanto, verificamos que a técnica Neuro-Fuzzy é capaz de desenvolver um modelo de calibração não linear mais eficiente que a rede neural e neurogenética, pois o erro de previsão é significativamente menor e não apresenta problemas para modelar a superfície perto das regiões de fronteira.

Os resultados do treinamento da rede neural indicam que o método de Levenberg-Marquart apresenta desempenho melhor com relação à taxa de convergência e à capacidade de estimar o valor do parâmetro Z . Portanto, a partir destas observações iniciais, decidimos usar o método de Levenberg-Marquart como padrão para o treinamento da rede neural nas aplicações das amostras reais.

Este estudo inicial deixa claro que o tempo de processamento para o treinamento de uma rede neurogenética é a principal desvantagem que esta técnica apresenta, quando comparada a outras técnicas. Isso indica que para esta técnica ser aplicada como uma forma alternativa de treinamento, é necessário aperfeiçoar o algoritmo de treinamento de maneira que esse possa ser processado mais rapidamente. Caso contrário, este método de treinamento possui pouco interesse prático, apesar de teoricamente apresentar um grande potencial.

7. Introdução

Na indústria açucareira, o termo açúcar é empregado para se referir à sacarose, enquanto, os demais açúcares como frutose, galactose e glicose são chamados de açúcares redutores⁽⁸⁵⁾.

A produção dos cristais de açúcar é realizada a partir da solução concentrada de caldo de cana de açúcar, que é basicamente constituída de uma mistura de sacarose, açúcares redutores, sais minerais, ácidos orgânicos e água. Ao final das etapas de processamento, o caldo de cana é convertido em cristais de açúcares, com composição específica segundo as exigências do cliente. Quando maior a percentagem de sacarose existente nos cristais, maior a pureza do açúcar⁽⁸⁵⁾.

A composição média do açúcar pode variar entre: sacarose (97,70% - 99,93%), açúcares redutores (0,22% - 0,01%), matéria orgânica (0,37% - 0,01%), cinzas (0,30% - 0,01%) e umidade (0,31%- 0,04%)⁽⁸⁵⁾.

A presença dos açúcares redutores (como por exemplo, frutose, galactose e glicose) é considerada impureza, diminuindo consideravelmente o valor agregado do produto final para a venda. Portanto, o controle do nível de sacarose, e indiretamente, dos açúcares redutores, é de extrema importância econômica para este setor industrial.

Segundo os critérios da indústria açucareira, os açúcares podem ser classificados em: açúcar refinado (99,90% de sacarose); açúcar bruto (99,30% - 98,00% de sacarose); açúcar escuro (um açúcar bruto que pode ser utilizado para o consumo direto) e açúcar “Tel quel” (um produto sem referencia de qualidade). Normalmente, a qualidade do caldo da cana é avaliado pelo valor do grau Brix e Pol.

O grau Brix indica a quantidade de sólidos solúveis em 100 gramas de solução, sendo o material sólido constituído principalmente de açúcares (glicose, frutose e sacarose) e sais minerais⁽⁸¹⁾. O grau Brix com valor igual a 10 indica que a solução contém 10 gramas de material sólido dissolvido em 100 gramas de solução. Em soluções concentradas do xarope do caldo de cana, o grau Brix pode atingir valores entre 60 e 70%.

A medida do grau Brix baseia-se no índice de refração das substâncias químicas presentes no meio. O refratômetro normalmente é calibrado com uma solução aquosa de sacarose pura a 20°C. Portanto, é possível estabelecer uma relação entre o índice de refração a 20°C e a percentagem de massa total de sólidos solúveis na solução aquosa pura de sacarose. Um incremento no índice de refração indica um aumento na concentração de sólido solúvel em solução⁽⁸⁵⁾.

A análise polarimétrica (Pol) é o método tradicionalmente empregado para determinar a quantidade de sacarose no caldo de cana de açúcar industrial. Esta técnica baseia-se na propriedade

que algumas substâncias químicas possuem de desviar o plano da luz polarizada em uma dada direção do plano da luz. Por exemplo, a frutose em solução desvia a luz polarizada para a esquerda. Uma solução com valor do grau Pol de 98% indica que esta amostra possui aproximadamente 98% de sacarose⁽⁸¹⁾.

Embora a análise polarimétrica forneça bons resultados sob condições ideais, esta técnica pode apresentar sérias limitações, principalmente quando existem na solução outras substâncias que também possuem atividade óptica. Portanto, em certos casos, faz-se necessário a utilização de métodos mais precisos como os métodos enzimáticos ou a cromatografia líquida e gasosa⁽⁸⁶⁾. Entretanto, ao considerarmos fatores como custo, tempo de análise e complexidade da execução destes tipos de análise, chegamos à conclusão que estes métodos ainda não satisfazem totalmente a necessidade da indústria açucareira.

Atualmente, já existem alguns trabalhos que acenam com a possibilidade de utilizar a espectroscopia no infravermelho próximo como uma técnica alternativa para a realização de análises quantitativas.⁽⁸⁷⁾ Estes trabalhos mostram que esta técnica apresenta um grande potencial em análises quantitativas *in-line* ou *on-line* do grau Brix e Pol.

O objetivo principal deste estudo é avaliar a possibilidade do desenvolvimento de uma metodologia de análise *off-line* dos parâmetros Brix e Pol, utilizando espectros na região do infravermelho próximo (NIR) de amostras de caldo de cana de açúcar e verificar qual o modelo de calibração multivariada mais adequado para ser utilizado.

7.1 Procedimento Experimental

A aquisição dos espectros foi feita em um espectrofotômetro FEMTO PLS PLUS, dispersivo, no Centro de Pesquisa da COOPERSUCAR em Piracicaba, utilizando uma cubeta de quartzo com caminho óptico de 1 mm e um sistema automático de sucção de amostra. O sistema de sucção é composto de uma bomba e tubos de polietileno com diâmetro de 1 mm. Antes de iniciar a sucção da amostra, os tubos de polietileno são lavados internamente passando um fluxo contínuo de água destilada por um tempo que varia entre um a dois minutos.

Antes da solução de caldo de cana de açúcar ser analisada, esta é previamente filtrada para eliminar partículas sólidas dispersas em solução, que poderiam afetar as medidas espectroscópicas ou ocasionar um bloqueio no sistema durante a sucção ou limpeza.

Durante a aquisição dos espectros, utilizou-se a própria solução de caldo de cana de açúcar da amostra subsequente para a limpeza do sistema. Desta maneira, o tempo de aquisição dos espectros foi otimizado e, simultaneamente, a geração de resíduos minimizada.

A aquisição dos espectros foi realizada entre 1000 nm e 2400 nm, com uma resolução de 2nm. O processo de aquisição, limpeza e substituição da amostra necessitam de, aproximadamente, 2 minutos.

A COOPERSUCAR adota como metodologia padrão para determinar o valor do grau Brix e Pol a refratometria e polarimetria, respectivamente.

7.1.1 Procedimento Experimental para Determinação do Grau Brix

O caldo de cana de açúcar deve ser filtrado em um papel de filtro; em seguida algumas gotas do caldo filtrado devem ser transferidas para a superfície do prisma do refratômetro, com um bastonete de borracha. A superfície do prisma deve ser previamente limpa e seca, antes de introduzir o material.

Depois de adicionado o material, deve-se esperar alguns segundos antes de iniciar a medida. Este procedimento visa garantir que a amostra e a superfície do prisma estejam em equilíbrio térmico. Em seguida é realizada a leitura, anotando o valor do grau Brix. Posteriormente, a leitura é corrigida para a temperatura padrão de 20°C.

7.1.2 Procedimento Experimental para Determinação do Grau Pol

Deve-se transferir 200 mL do caldo coletado para um béquer. Em seguida, são adicionadas entre 2 a 3 gramas de subacetato de chumbo. A solução formada deve ser agitada com um bastão de vidro por, aproximadamente, 2 minutos; em seguida, o material é filtrado sobre um papel de filtro em pregas colocado sobre um béquer. Os primeiros 20 mL do material filtrado devem ser descartados.

Com o filtrado absolutamente límpido, lavar 3 vezes sucessivas o tubo de polarização, enchendo-o em seguida com a solução.

7.1.2.1 Cuidados durante as medidas instrumentais

1. A solução deverá proporcionar um percurso ótico livre de interferências.
2. A clarificação, com subacetato de chumbo, deverá proporcionar soluções límpidas e de coloração atenuada.
3. O tubo polarimétrico deverá estar totalmente preenchido e livre de bolhas.
4. As janelas das extremidades dos tubos deverão estar limpas.
5. A aferição do aparelho deve ser periódica, utilizando os padrões que o acompanham.

7.1.2.2 Parâmetros que podem afetar as medidas polarimétricas

As medidas polarimétricas têm sua magnitude é afetada pela temperatura, comprimento de onda da luz e pelas características do material opticamente ativo. Mantendo constantes estas condições o, ângulo de rotação é diretamente proporcional à concentração do componente ativo e ao caminho óptico da luz no interior da solução.

7.1.2.3 Definição da Escala Sacarimétrica

Uma solução pura de sacarose contendo ± 26 gramas em 100 mL de solução, colocada em um tubo de 20 cm de comprimento, a 20 °C, proporciona um desvio do plano de polarização ao qual se atribui o valor de 100. O valor zero corresponde à água destilada.

100 ° S ----- 26 g / 100 mL

L ° S ----- X g / 100 mL

Como o Pol é definido como percentagem em massa, o valor acima obtido deverá ser corrigido pela densidade da solução:

$$Pol = \frac{0,26 * leitura}{Densidade a 20^{\circ} C} \frac{g}{100 mL} \quad \text{eq. (76)}$$

7.2 Desenvolvimento dos modelos de calibração multivariada

A COOPERSUCAR nos cedeu dois conjuntos de dados com 307 espectros para a análise do grau Brix e 262 espectros para a determinação do grau Pol.

Os espectros foram submetidos a um pré-processamento, que envolveu a eliminação das regiões compreendida entre 1806 nm e 1918 nm e acima de 2385 nm. Em seguida, foi realizada uma correção de linha base através da correção de sinal multiplicativo (MSC) e alisamento com o filtro da transformada de Fourier. Posteriormente, foi utilizada a análise das componentes principais para separar as amostras em dois conjuntos (calibração e teste).

Para a construção do modelo de calibração do grau Brix, foram empregadas 263 amostras e para o teste do modelo, 44 amostras. O modelo de calibração do grau Pol foi construído com 218 amostras. Para o conjunto teste do modelo desenvolvido foram usadas as 44 amostras restantes.

A Figura 36 ilustra a distribuição do valor do grau Brix para as 307 amostras de caldo de cana de açúcar. Para o estudo deste parâmetro, foi utilizada a faixa de concentração entre 13 e 24 unidades do grau Brix.

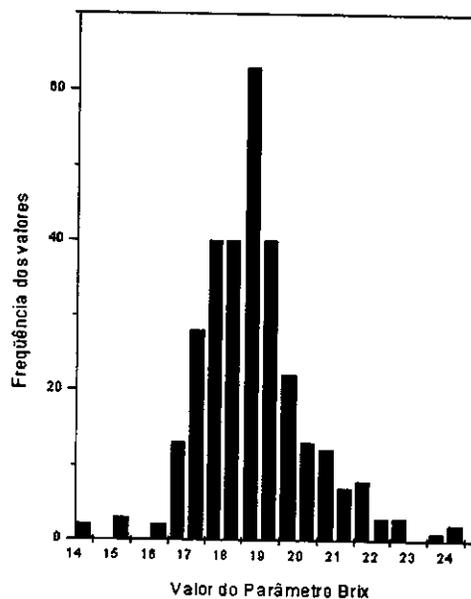


Figura 36. Distribuição do valor de grau Brix para as 307 amostras utilizadas.

A Figura 37 ilustra a distribuição do valor do grau Pol para as 262 amostras de caldo de cana de açúcar. Para o grau Pol, a faixa de concentração das amostras analisadas variou entre 8 a 21 unidades do grau Pol.

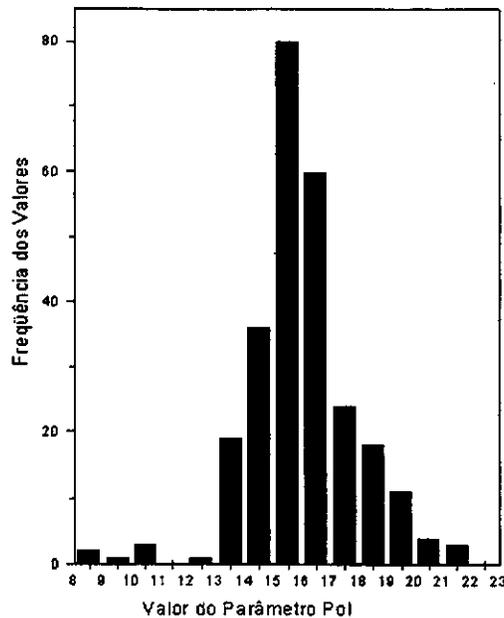


Figura 37. Distribuição do valor do grau Pol para as 262 amostras utilizadas.

Nos modelos de rede neural, neurogenética e Neuro-Fuzzy, utilizou-se, como entrada, os valores escalados entre +0,8 e -0,8 dos escores obtidos pela análise de componentes principais, sendo a saída da rede neural relacionada aos valores escalados do grau Brix e Pol, respectivamente.

Para ter uma mesma base de comparação entre os resultados obtidos, utilizaram os mesmos critérios de finalização entre os métodos de inteligência artificial, ou seja, número máximo de 600 iterações ou um erro mínimo de 10^{-5} .

7.2.1 Método dos Mínimos Quadrados Parciais (PLS)

O modelo dos mínimos quadrados parciais foi construído utilizando oito componentes principais, para as análises do grau Brix e Pol. Esse número ideal de componentes principais foi determinado por validação cruzada. Para cada um dos parâmetros, foi construído um modelo de calibração.

7.2.2 Algoritmo genético com Regressão Linear Múltipla (AG/RLM)

Para o modelo de Regressão Linear Múltipla com seleção das variáveis, o conjunto de amostras de calibração do parâmetro Brix com 263 amostras foi subdividido em dois grupos. O primeiro subconjunto com 193 amostras foi utilizado para o desenvolvimento do modelo de calibração. O segundo subconjunto com 70 amostras foi utilizado para a validação interna do modelo de

calibração. Para testar o modelo de calibração, foi utilizado o mesmo grupo de 44 amostras do modelo dos mínimos quadrados parciais.

Um procedimento análogo foi empregado durante a etapa do desenvolvimento e avaliação do modelo de calibração multivariada para o parâmetro Pol. Para este parâmetro, foram utilizadas 148 amostras para a construção do modelo, 70 amostras para a validação interna e 44 amostras para o conjunto teste.

O programa do algoritmo genético foi configurado de maneira que a população inicial possuísse 50 cromossomos, um número máximo de 600 iterações, e probabilidade de cruzamento e mutação de 90% e 1%, respectivamente.

7.2.3 Rede Neural

A arquitetura da rede neural criada para a determinação do grau Brix e Pol foi inicialmente otimizada por tentativa e erro, obtendo-se a configuração: 8 neurônios de entrada, 2 neurônios na camada intermediária com função de transferência tangente sigmoideal e um neurônio na camada de saída com função de transferência linear. Utilizou-se o método de Levenberg – Marquardt para treinar a rede (correção dos pesos da rede).

7.2.4 Rede Neurogenética

A arquitetura da rede neurogenética otimizada apresentou a configuração: 8 neurônios de entrada, 5 neurônios na camada intermediária com função de transferência tangente sigmoideal e, entre as entradas e a saída, foi introduzido um neurônio com função de transferência linear, para a realização de uma ligação direta entre a camada de saída e a entrada.

O algoritmo genético para treinamento dos pesos foi configurado de maneira que a porcentagem de cruzamento e mutação fossem de 90% e 1%, respectivamente.

7.2.5 Rede Neuro-Fuzzy

A arquitetura otimizada para os modelos do grau Brix e Pol apresentaram a mesma configuração, três entradas, com duas funções de participação do tipo gaussiana cada uma. Os parâmetros da rede Neuro-Fuzzy foram ajustados pelo método do gradiente descendente.

7.3 Resultados e Discussão

7.4 Determinação do grau Brix usando a espectroscopia no Infravermelho Próximo

Os estiramentos das ligações químicas O-H da molécula de água apresentam uma intensa absorção na região espectral do infravermelho próximo. Como consequência, o espectro de soluções aquosas possui uma contribuição majoritária do espectro da água em relação aos de outras espécies em solução. Isso dificulta a análise de uma série de espécies químicas presentes em matrizes aquosas. Esse problema é mais sério quando a intensidade do sinal de absorbância associado ao analito de interesse é significativamente inferior ao da ligação O-H da banda da água. Neste caso, as variações no perfil do espectro provenientes da concentração do analito de interesse são ocultadas pelo intenso sinal da banda da água.

Um dos recursos utilizados para minimizar a influência da banda de água é selecionar as regiões espectrais onde as ligações das moléculas de açúcares possuem as bandas de absorção mais intensas. Desta maneira, é possível tentar reduzir a influência da absorção da ligação O-H da banda da água e simultaneamente maximizar o sinal proveniente dos grupos orgânicos dos açúcares. Entretanto, na região do infravermelho próximo esta tarefa não é muito simples, pois as bandas de absorção possuem um elevado grau de sobreposição. O procedimento mais comum para realizar esta tarefa é o uso de ferramentas matemáticas específicas (seleção de variáveis ou faixas espectrais, subtração do espectro da água, etc). Outro procedimento alternativo é a redução do caminho óptico que a luz percorre pela amostra. Alguns trabalhos descritos na literatura reportam que análises quantitativas de amostras em matrizes aquosas são mais eficientes quando se utiliza uma cubeta com caminho óptico inferior a 1 cm.

Uma cubeta com caminho óptico menor implica em uma redução na sensibilidade da análise; entretanto, causa também uma forte redução na banda de absorção da água. Este procedimento também possui a finalidade de evitar a saturação da banda de água. Isso resultaria na perda de informações contidas na região onde a banda de água absorve. Por este motivo, foi utilizada uma cubeta de quartzo com caminho óptico de 1 mm.

Existem outros procedimentos instrumentais que podem ser adotados para melhorar a qualidade do sinal como, por exemplo, aumentar o número de varreduras por amostra. Estes procedimentos são adotados para aumentar a relação sinal/ruído do espectro, tendo como objetivo evitar que sinais de baixa intensidade provenientes do analito de interesse se misturem ao ruído. Como consequência, podem ser perdidas informações importantes correlacionadas com o analito de interesse.

A Figura 38 ilustra o perfil da coleção de espectros empregados para a construção e avaliação dos modelos de calibração multivariada.

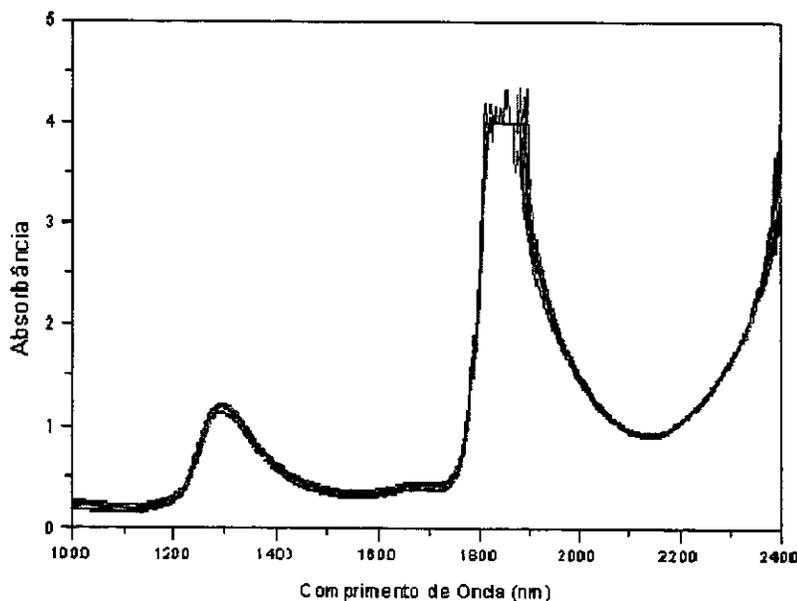


Figura 38. Espectros de absorvância das amostras de caldo de cana de açúcar na região do infravermelho próximo.

Segundo a literatura, o espectro da água na região do infravermelho próximo apresenta quatro bandas de absorção intensas em 970, 1190, 1450 e 1950 nm. Entretanto, é importante ressaltar que a presença de outras espécies pode ocasionar um deslocamento na posição e alteração na intensidade das bandas de absorção do espectro da água⁽⁸⁶⁾.

Na Figura 38, podemos observar as bandas de absorção ao redor de 1300 nm e 1850 nm, que se apresentam deslocadas pela provável interação das moléculas de água com as inúmeras substâncias químicas contidas na solução do caldo de cana de açúcar.

A Figura 38 mostra claramente a forte absorção da banda de água na faixa compreendida entre 1806 nm a 1918 nm. Nesta região, o sinal apresenta-se saturado e com uma baixa relação sinal/ruído, observado devido à elevada variação aleatória do sinal próximo da região de máximo da banda. Por estas razões, a região compreendida entre 1806 nm a 1918 nm foi excluída deste estudo. A região acima de 2385 nm também foi excluída por apresentar uma baixa relação sinal/ruído.

A Figura 39 ilustra as amostras selecionadas para o conjunto de calibração, validação e teste. Tomamos o cuidado de selecionar amostras para cada um dos conjuntos, de modo que sejam as mais representativas possíveis do espaço amostral estudado. As amostras foram escolhidas de modo que o conjunto de calibração possuísse a faixa mais ampla do espaço amostral e o conjunto teste, a

mais estreita. Este cuidado é necessário, pois normalmente os modelos de calibração multivariada não podem ser usados para estimar o valor da propriedade de interesse fora do domínio que o modelo foi desenvolvido.

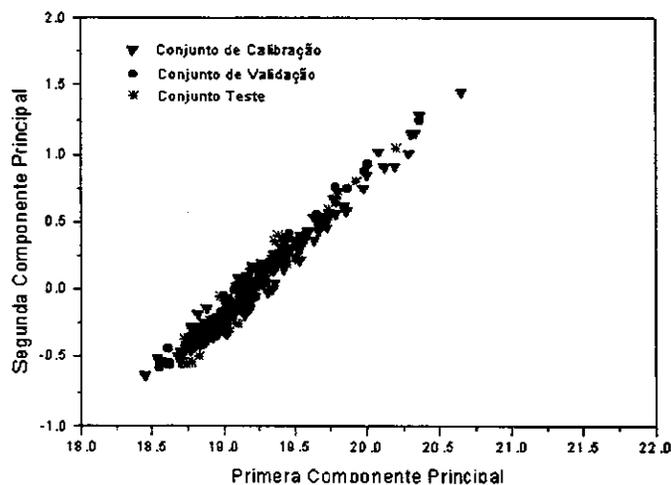


Figura 39. Gráfico dos escores da primeira pela segunda componente principal.

A Figura 40 ilustra o gráfico que mostra a influência de cada uma das amostras empregadas do modelo de calibração e validação.

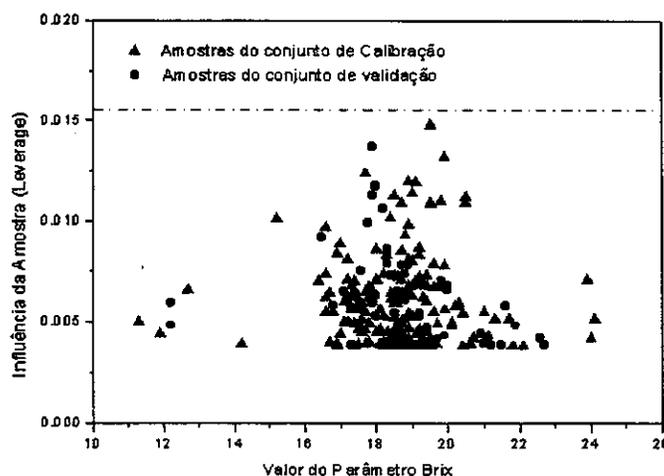


Figura 40. Gráfico de influência das amostras no modelo de calibração.

Como podemos observar, Figura 40, as amostras empregadas neste estudo possuem uma distância inferior ao limite máximo de influência tolerável. Isso indica que este conjunto de amostras não possui anomalias que possam distorcer o modelo de calibração.

Optamos por não apresentar o *leverage* das amostras do conjunto teste, porque se o modelo de calibração não apresenta nenhuma anomalia, ele deverá ser capaz de prever de maneira eficiente

novas amostras. Caso isso não ocorra, provavelmente o problema deve estar associado às novas amostras e não ao modelo de calibração. Como o objetivo desta etapa é desenvolver um modelo de calibração que seja o mais robusto possível, não nos interessa saber se o conjunto teste contém uma ou mais amostras com comportamento distinto das utilizadas no modelo de calibração.

Se posteriormente for observado que alguma(s) da(s) amostra(s) pertencente(s) ao conjunto teste apresentam o valor de previsão muito diferente do valor padrão, poderemos realizar um estudo utilizando o *leverage*. Isso seria feito para verificar se existe alguma diferença significativa entre esta amostra e as amostras empregadas no modelo de calibração. Caso isso seja constatado, provavelmente esta amostra possui o valor do parâmetro de interesse fora dos limites da curva de calibração e/ou porque ocorreu alguma variação durante a manipulação da amostra e/ou aquisição do espectro. Portanto, poderemos simplesmente descartar esta amostra do conjunto teste. Caso contrário, se não for verificado que esta amostra seja anômala, deveremos reavaliar o modelo de calibração.

A seguir, serão apresentados os resultados obtidos para cada um dos cinco modelos de calibração desenvolvidos para a previsão do grau Brix nas amostras de caldo de cana de açúcar.

7.5 Modelos de Calibração Multivariada para a Determinação do grau Brix

7.5.1 Modelo dos Mínimos Quadrados Parciais (PLS)

O modelo dos mínimos quadrados parciais apresentou um bom desempenho para a previsão das amostras do conjunto teste. Para este método, foi obtido um coeficiente de correlação para as amostras de calibração e teste de 0,972 e 0,968, respectivamente, e um valor de erro padrão de previsão (SEP) de 0,39 grau Brix.

Neste trabalho, também foi calculado o erro padrão de calibração (SEC), cujo valor também foi de 0,39 graus Brix. A similaridade entre os valores de SEC e SEP indica que o modelo desenvolvido apresenta uma boa robustez. Estes resultados estão apresentados na Tabela 9 no final da seção de resultados e discussão.

A Figura 41 mostra a boa concordância entre os valores previstos e os valores de referência.

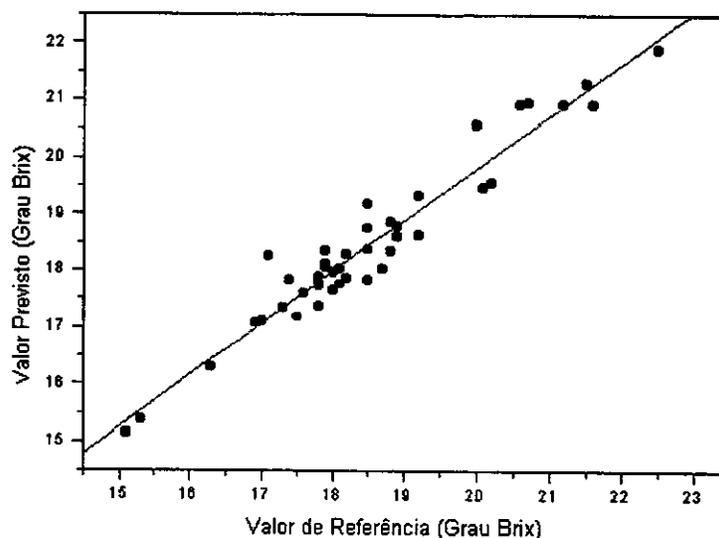


Figura 41. Valores do método referência versus os valores obtidos por NIR para a determinação do grau Brix, utilizando PLS.

7.5.2 Seleção de Variáveis Aplicando Algoritmo Genético (AG/RLM)

O programa do algoritmo genético foi empregado em conjunto com a técnica de regressão linear múltipla, com o objetivo de selecionar o melhor conjunto de variáveis que possa ser aplicado para previsão do grau Brix. A Figura 42 ilustra o perfil do espectro obtido para a solução de caldo de cana e as variáveis selecionadas pelo algoritmo genético.

Observamos que as variáveis que apresentam maior correlação com o grau Brix são aquelas que estão localizadas mais distantes do máximo da banda de absorção da água, com exceção do comprimento de onda em 1335 nm.

Os comprimentos de ondas selecionados foram atribuídos como⁽⁸⁰⁾: 1045 nm e 1047 nm ao estiramento e deformação da ligação $-CH_2$; 1141 nm, 1181 nm e 1219 nm ao segundo sobretom do estiramento $-CH$; 1335 nm e 1440 nm ao estiramento e deformação da ligação $-CH$; 1660 nm ao primeiro sobretom do estiramento da ligação $-CH$; 2116 nm ao estiramento da ligação $-C=O$; 2208 nm ao estiramento $-CH$ e $-C=O$ e 2261 nm ao estiramento e deformação da ligação $-OH$.

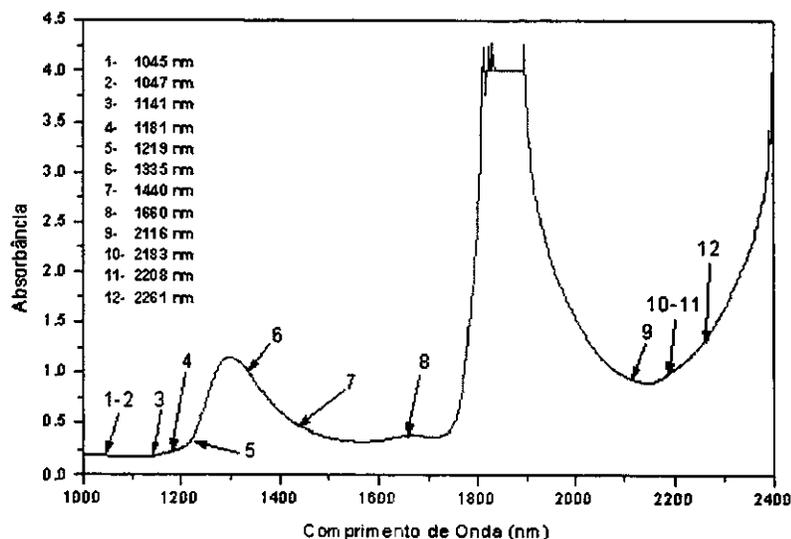


Figura 42. Perfil do espectro da solução de cana de açúcar com os comprimentos de ondas selecionados pelo algoritmo genético.

A Figura 43 apresenta o gráfico dos valores previstos pelo modelo de regressão linear múltipla versus os valores fornecidos pelo método de referência. Pode-se observar uma boa concordância entre os valores previstos e os valores de referência. Aparentemente, este modelo apresentou um desempenho levemente superior ao obtido pelo método dos mínimos quadrados parciais. Isso indica que a seleção de variáveis pode contribuir para a construção de um modelo mais simples e tão eficiente como os obtidos por outras técnicas de calibração multivariada. Isso é possível pelo fato desta técnica empregar somente os comprimentos de onda que apresentam menor influência da banda de água e uma maior relação com os estiramentos dos grupos orgânicos dos açúcares e outros sólidos solúveis.

Os valores previstos pelo modelo para as amostras do conjunto teste apresentam um coeficiente de correlação de 0,966 e erro padrão de previsão (SEP) de 0,35 graus Brix, enquanto as amostras do conjunto de validação apresentam resultados um pouco superiores. Entretanto, este comportamento já era esperado, devido ao fato das variáveis terem sido selecionadas para minimizar principalmente o erro de previsão das amostras do conjunto de validação. Este comportamento é compreensível, já que na maioria dos casos as amostras utilizadas para avaliar o modelo de calibração não representam completamente o comportamento do sistema que se deseja modelar. Desta forma, o modelo de calibração sempre apresenta uma flexibilidade na habilidade de previsão de novas amostras.

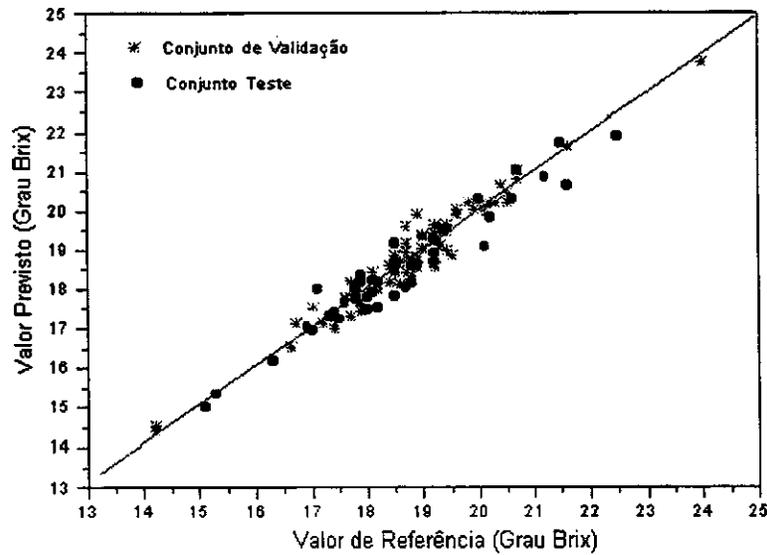


Figura 43. Valores do método referência versus os valores obtidos por NIR. As amostras do conjunto teste estão representadas pelo símbolo (●) e as do conjunto de validação por (*).

A Figura 44 ilustra o erro relativo de cada uma das amostras empregadas nos conjuntos de validação e teste. O gráfico do erro relativo será apresentado apenas para o melhor modelo de calibração multivariada.

Figura 44 apresenta o erro relativo versus o valor do grau Brix. Esta maneira de apresentar os dados permite ao leitor observar se existe alguma correlação entre o erro relativo e o valor de Brix. Como podemos verificar, aparentemente não existe nenhuma correlação entre eles.

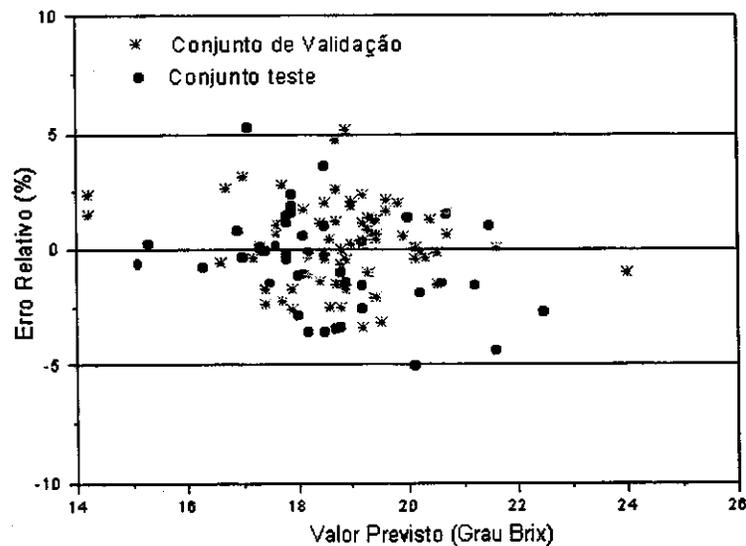


Figura 44. Erro relativo obtido na previsão do grau Brix por NIR, utilizando a regressão linear múltipla e algoritmo genético.

Este modelo apresentou um erro relativo máximo de aproximadamente $\pm 5\%$, sendo que a maioria das amostras apresentou erro inferior a este. Este erro está na mesma ordem do erro do método padrão; portanto, dificilmente obteremos resultados melhores que os apresentados para este modelo de calibração multivariada.

7.5.3 Rede Neural

O modelo não linear baseado na técnica de rede neural mostrou-se menos eficiente que os modelos lineares apresentados anteriormente, com um erro padrão de previsão (SEP) de 0,46 graus Brix. Entretanto, o modelo utilizando rede neural mostrou ser bastante robusto, pois apresenta o erro padrão de calibração e previsão similares entre eles. Estes resultados estão apresentados na Tabela 9, juntamente com os resultados dos demais modelos desenvolvidos para a previsão do grau Brix.

A Figura 45 mostra o gráfico dos valores previstos pelo modelo com rede neural versus os valores do método de referência. Para este modelo, foi obtido um coeficiente de correlação de 0,958. Podemos observar que as amostras estão mais dispersas ao redor da curva ajustada, do que as dos modelos lineares. Isso indica que estas amostras possuem um erro relativo superior ao observado no modelo de regressão linear múltipla e no PLS.

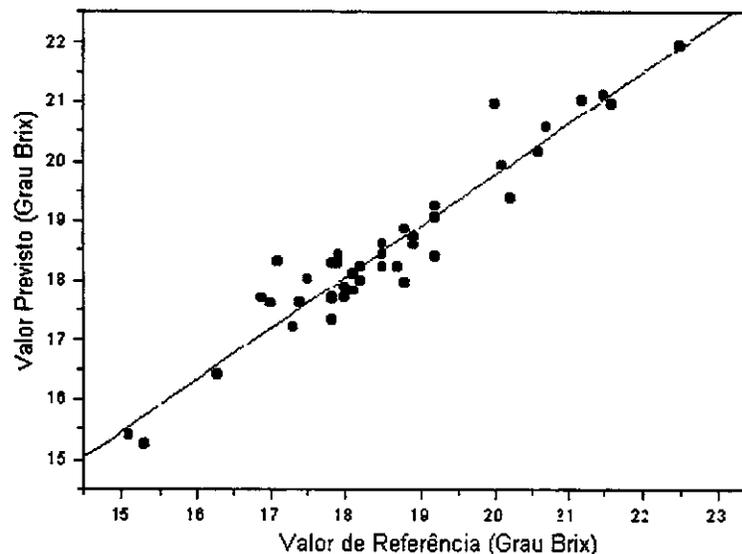


Figura 45. Valores do método referência versus os valores obtidos por NIR utilizando a rede neural.

7.5.4 Rede Neurogenética

O modelo neurogenético forneceu resultados inferiores aos obtidos pela técnica de rede neural. O coeficiente de correlação obtido para o conjunto teste foi de 0,743. Além disso, o valor de SEP (0,77 graus Brix) e SEC (0,72 graus Brix) são, aproximadamente, duas vezes maiores que os obtidos para os modelos lineares.

A Figura 46 apresenta os valores previstos versus os valores fornecidos pelo método de referência. Pode-se observar uma grande dispersão dos dados ao redor da curva ajustada. Este comportamento indica que o modelo apresenta uma baixa correlação entre os valores previstos e os valores fornecidos pelo método de referência. A dispersão dos pontos indica que esse conjunto de amostras possui um erro relativo superior aos obtidos pelas técnicas lineares e rede neural.

Os resultados alcançados por esse modelo mostram que esta técnica não foi capaz de capturar completamente as informações correlacionadas com a variação da quantidade de sólido solúvel em solução. Acreditamos que o pobre desempenho deste modelo, quando comparado ao modelo de rede neural, deve estar relacionado ao fato da rede neurogenética não ter tido tempo suficiente para convergir para a melhor solução. Seria de esperar que apresentassem resultados semelhantes, visto que esta técnica utiliza as mesmas funções de mapeamento que a da rede neural, diferenciando apenas no algoritmo de treinamento.

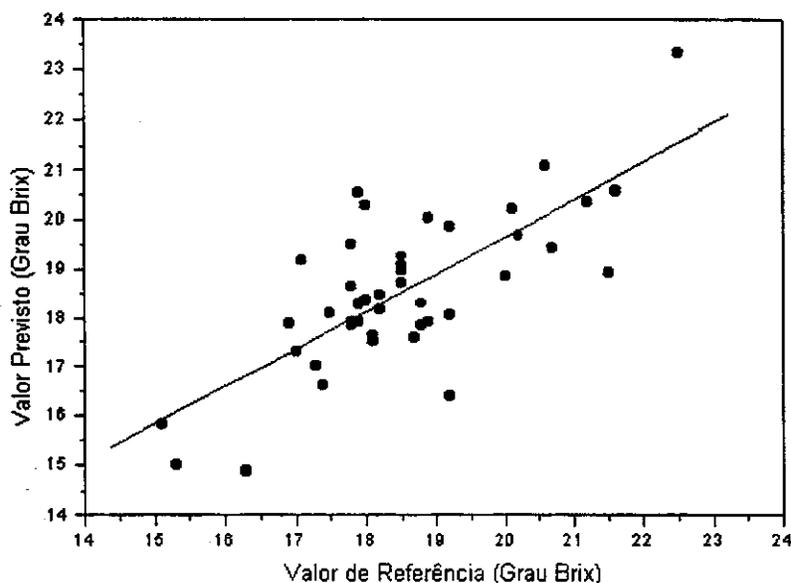


Figura 46. Valores do método referência versus os valores obtidos por NIR, utilizando a rede neurogenética.

7.5.5 Rede Neuro-Fuzzy

O modelo Neuro-Fuzzy apresentou um desempenho um pouco melhor que o modelo neurogenético. Como podemos observar na Figura 47, os pontos apresentam-se menos dispersos ao redor da curva ajustada do que os apresentados na Figura 46.

Os valores previstos versus os valores do método padrão apresentaram um coeficiente de correlação de 0,865. A correlação entre os dados é significativamente superior aos obtidos pela rede neurogenética, entretanto, inferior ao alcançado pela rede neural e modelos lineares.

Os valores de SEC (0,78 graus Brix) e SEP (0,62 graus Brix) confirmam o baixo desempenho de previsão deste modelo de calibração. O pobre desempenho deste modelo está relacionado ao fato desta técnica ter sido desenvolvida para a modelagem de sistemas onde existe a presença de não linearidade dos dados. Para sistemas lineares, pensamos que esta técnica não seja recomendada, pois dificilmente apresentará resultados com qualidade similares as obtidas pelos modelos lineares.

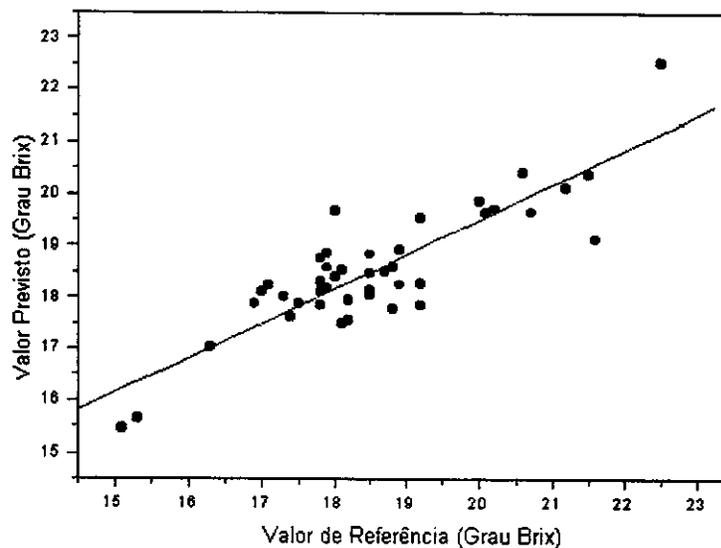


Figura 47. Valores do método referência versus os valores obtidos por NIR, utilizando a rede Neuro-Fuzzy.

Teoricamente, os modelos não lineares adaptativos, semelhantes aos utilizados pelas técnicas de inteligência artificial, são capazes de convergirem para um modelo linear. Entretanto, este processo é lento e, geralmente, a solução proposta é menos eficiente que a oferecida pelos métodos lineares. Por esse motivo, acreditamos que esta técnica tenha apresentado resultados inferiores que os métodos lineares.

Para explicar a diferença entre os resultados oferecidos pela rede neural e Neuro-Fuzzy, pensamos que o problema esteja associado a uma série de fatores como velocidade de treinamento, taxa de convergência, complexidade da arquitetura, etc. Estas variáveis possuem um impacto direto no desempenho do algoritmo de treinamento dos pesos da rede. Isso pode implicar em resultados bem distintos com relação à habilidade de previsão dos modelos propostos por estas duas técnicas de inteligência artificial.

7.6 Considerações Finais

A comparação entre os resultados obtidos para cada método foi realizada utilizando o teste F com 95% de confiança que apresenta um valor crítico de 1,67. Neste caso, para realizar a comparação entre os modelos de calibração, foi utilizado como referência o modelo que apresentou o menor valor de SEP. Para este sistema, o modelo com menor erro padrão de previsão foi o modelo desenvolvido a partir da regressão linear múltipla.

Os resultados da Tabela 9 mostram que o método dos mínimos quadrados parciais e rede neural não apresentam diferenças significativas na habilidade de previsão, quando comparado ao modelo obtido pela regressão linear múltipla. Entretanto, os modelos neurogenético e Neuro-Fuzzy apresentam a habilidade de previsão significativamente inferior.

Tabela 9. Valores de erro e do teste F obtidos para os 5 modelos de calibração multivariada na estimativa do grau Brix.

	PLS	RLM	Rede Neural	Neurogenética	Neuro-Fuzzy
SEC (graus Brix)	0,39	0,32	0,46	0,77	0,78
SEP (graus Brix)	0,39	0,35	0,45	0,72	0,62
Teste - F	1,24	-	1,29	4,23	3,13

$F_{\text{Crítico } 95\%} = 1,67$

Os resultados apresentados neste trabalho mostram que os modelos não lineares não devem ser vistos como uma solução absoluta para a modelagem de dados com comportamentos ainda não determinados. Geralmente, as pessoas são levadas a acreditar que os modelos não lineares são capazes de modelar sistemas lineares e não lineares com a mesma habilidade. Esse tipo de raciocínio é facilmente aceito pelo fato de que podemos pensar que o mapeamento linear é um caso particular de um mapeamento não linear.

Para que um mapeamento não linear convirja para um mapeamento linear é necessário ter controle sobre fatores como quantidade e qualidade das informações disponíveis sobre o sistema (número de experimentos), erro do método padrão, modelo matemático não linear utilizado, tempo de treinamento, valor dos parâmetros iniciais, etc.

Como podemos observar, existem vários parâmetros a que deve ser levados em consideração. Portanto, a melhor filosofia é iniciar com um modelo linear e, caso este não apresente resultados satisfatórios, então realizar um estudo comparativo empregando algumas técnicas não lineares.

7.7 Determinação do grau Pol usando a Técnica de Infravermelho Próximo

No estudo que envolve a modelagem do grau Pol, optamos por não apresentar a figura que ilustra o perfil do espectro de cana de açúcar, pois os espectros das soluções de cana de açúcar utilizados neste trabalho são similares aos apresentados na modelagem do grau Brix.

O procedimento da seleção das amostras para os conjuntos de calibração e teste foi realizada de maneira similar a conduzida para o grau Brix. Foram empregados os gráficos dos escores e *leverages* para selecionar as melhores amostras para cada um dos conjuntos. Novamente decidimos por não apresentar estes gráficos devido a fato de já termos apresentado gráficos similares na seção que trata os resultados do grau Brix.

7.8 Modelos de Calibração Multivariada para a Determinação do Pol

7.8.1 Modelo dos Mínimos Quadrados Parciais (PLS)

O modelo dos mínimos quadrados parciais mostrou-se bastante eficiente para estimar o valor do grau Pol. Esta observação é baseada no fato do modelo construído com PLS ter apresentado uma alta correlação entre o valor previsto pelo modelo e os correspondentes valores de referência das amostras empregadas no conjunto de calibração e teste.

Segundo os valores do erro padrão de calibração (SEC igual 0,57 graus Pol) e previsão (SEP igual a 0,61 graus Pol) apresentados na Tabela 10, pode-se verificar que o modelo desenvolvido possui um bom desempenho e uma boa robustez.

A Figura 48 mostra os resultados previstos versus os valores de referência que possui um coeficiente de correlação de 0,964. Graficamente, é possível constatar a boa correlação entre os resultados previstos e os valores fornecidos pelo método padrão. Este gráfico indica um comportamento homoscedástico dos dados ao longo de toda a curva de calibração, ou seja, a variância na habilidade de previsão do sistema parece ser constante e igual ao longo de toda a faixa estudada. Isso indica que não existe nenhum tipo de comportamento sistemático do erro ou seja apresenta heteroscedasticidade associada ao modelo.

Para a seleção das amostras, foram utilizadas as informações da análise da primeira versus a segunda componente principal, a distância de Mahalanobis e o valor do grau Pol.

A análise visual dos gráficos das componentes principais e a distância de Mahalanobis não permitem distinguir as amostras com baixo valor do grau Pol das demais amostras. Por este motivo optamos por utilizar a informação do valor do grau Pol para selecionar as amostras com valores inferiores a 13 unidades.

O valor do grau Pol mostra que existem 6 amostras com valor inferior a 13 unidades. Por este motivo, optamos por utilizar 5 amostras para o modelo de calibração e deixar apenas 1 amostra para ser usada na fase de avaliação do modelo. Este procedimento tende a aumentar as chances de desenvolvermos um modelo de calibração mais robusto e eficiente com relação à sua habilidade de previsão. A capacidade de mapeamento do modelo está diretamente relacionada ao número e a faixa de concentração que estas amostras cobrem. Apesar do recurso adotado neste trabalho, o ideal seria a utilização de um maior número de amostras com valores abaixo de 13 unidades, com a finalidade principal de avaliar melhor o desempenho de previsão do modelo para amostras que se encontram perto do limite inferior da faixa de calibração.

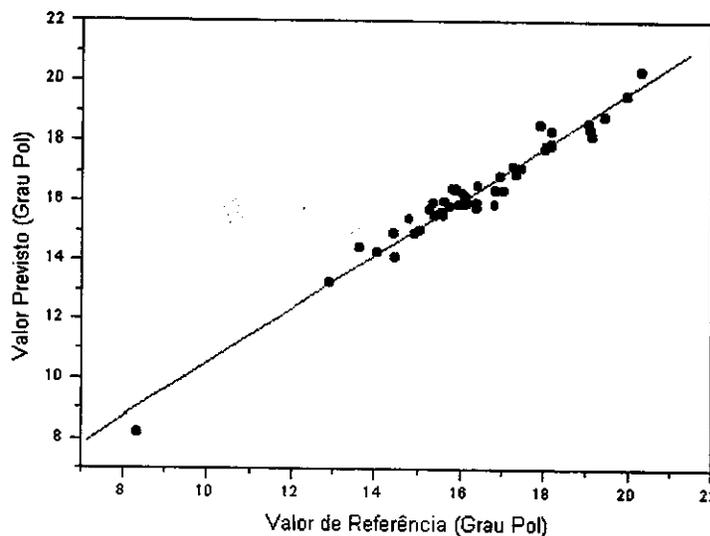


Figura 48. Valores do método referência versus os valores obtidos por NIR, utilizando o PLS

Apesar do bom desempenho de previsão obtido, conforme apresentado na Figura 48, pensamos que seria pouco prudente afirmar que este modelo seja o ideal para a modelagem deste sistema. Não possuímos informações suficientes para avaliar a habilidade do modelo para prever amostras com baixos valores de Pol. Entretanto, adotando uma postura mais conservadora, poderíamos afirmar que este modelo, no pior dos casos, pode ser utilizado para análises de caráter quantitativo na região entre 13 e 21 e, semi-quantitativo, para valores inferiores a 13 unidades do grau Pol.

7.8.2 Seleção de Variáveis Aplicando Algoritmo Genético (AG/RLM)

Após o processo iterativo de seleção das variáveis, o programa do algoritmo genético selecionou 14 variáveis como sendo as mais importantes para descrever o comportamento deste

sistema. O perfil de um dos espectros com os comprimentos de onda selecionados pode ser observado na Figura 49.

A Figura 49 mostra que os comprimentos de ondas selecionados encontram-se em regiões fora da banda de água, com exceção do comprimento de onda em 1280 nm. Além disso, podemos observar que a maioria dos comprimentos de onda selecionados está na faixa entre 1480 nm - 1750 nm. A seleção dos comprimentos de onda nesta região espectral está relacionada ao fato de que as principais vibrações moleculares dos açúcares ocorrem nesta faixa.

Os comprimentos de onda selecionados foram atribuídos às seguintes vibrações moleculares⁽⁸⁰⁾: 1045 nm estiramento -CH e deformação do -CH₂; 1280 nm segundo sobretom do estiramento -CH; 1485 nm primeiro sobretom do estiramento -OH; 1590 nm primeiro sobretom do estiramento -OH; 1655 nm, 1660 nm, 1708 nm, 1711 nm e 1749 nm primeiro sobretom do estiramento -CH; 2183 nm estiramento assimétrico do -CH₂ e estiramento C=C; 2230 nm e 2241 nm estiramento e deformação -OH; 2273 nm estiramento -OH e C-C; 2286 nm estiramento e deformação C-H.

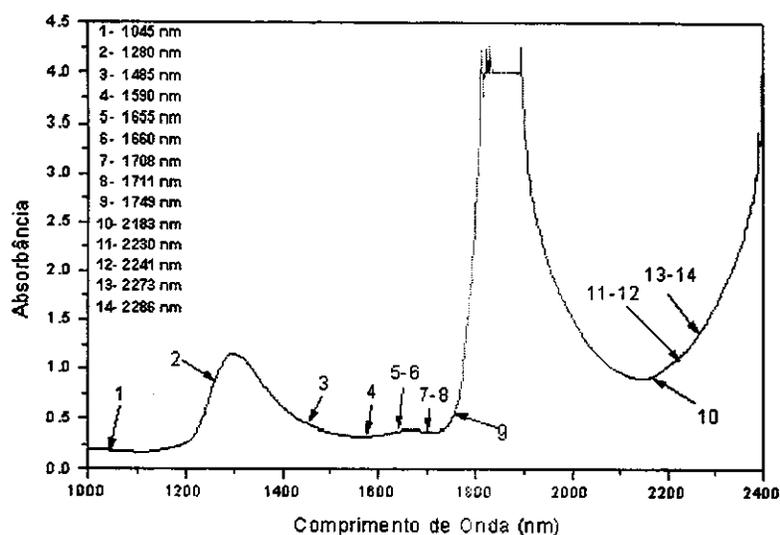


Figura 49. Perfil do espectro da solução de cana de açúcar com os comprimentos de onda selecionados pelo algoritmo genético.

A Figura 50 apresenta o gráfico dos valores previstos versus os valores fornecidos pelo método padrão de análise, para as amostras empregadas no conjunto de validação e teste. A análise visual da Figura 50 mostra uma alta correlação entre os valores previstos pelo modelo e os valores obtidos pelo método de referência. Para as amostras do modelo de calibração, validação e teste, foi obtido um coeficiente de correlação de 0,984, 0,964 e 0,968, respectivamente.

Para este modelo de calibração, foi obtido um erro padrão de calibração (SEC) e previsão (SEP) para as amostras do conjunto teste de 0,54 e 0,59 graus Pol, respectivamente. Por outro lado, o conjunto de validação apresentou um erro de previsão de 0,61 graus Pol.

Os valores de erro padrão de calibração e previsão mostram uma boa concordância entre si, indicando uma boa robutez do modelo de calibração.

Pode-se observar no gráfico da Figura 50, que as amostras do conjunto teste possuem um menor grau de dispersão ao longo da curva ajustada do que as amostras utilizadas para validar as variáveis selecionadas pelo algoritmo genético. Isso indica que o modelo de calibração não foi superajustado para as amostras de treinamento.

Outra informação importante fornecida por este sistema está relacionada com a capacidade de previsão do valor do grau Pol para valores abaixo de 13 unidades do grau Pol.

Como pode ser observado na Figura 50, o modelo é capaz de prever suficientemente bem a amostra do conjunto teste com valor de 8,2 graus Pol. O interessante desta amostra é que ela encontra-se abaixo do limite inferior das amostras do conjunto de validação, mas dentro da faixa de calibração das amostras do conjunto de calibração. Isso indica que durante a etapa de seleção de variáveis, o conjunto de validação não possuía informações sobre esta região de mapeamento e apesar disso, o modelo foi capaz de estimar o valor do parâmetro Pol de maneira muito eficiente. Isso mostra que o sistema possui um comportamento linear e o modelo de calibração não foi superajustado para o conjunto de variáveis selecionadas.

Estes resultados também fornecem mais evidências de que é possível desenvolver um modelo de calibração que seja capaz de prever o valor de Pol para a faixa entre 8 a 21 unidades do grau Pol, mantendo a mesma qualidade de previsão que a observada na faixa entre 13 a 21. Para o conjunto de validação, foram utilizadas 2 amostras com valores próximos de 10,5 e estas possuem um erro de previsão similar às demais amostras empregadas neste estudo.

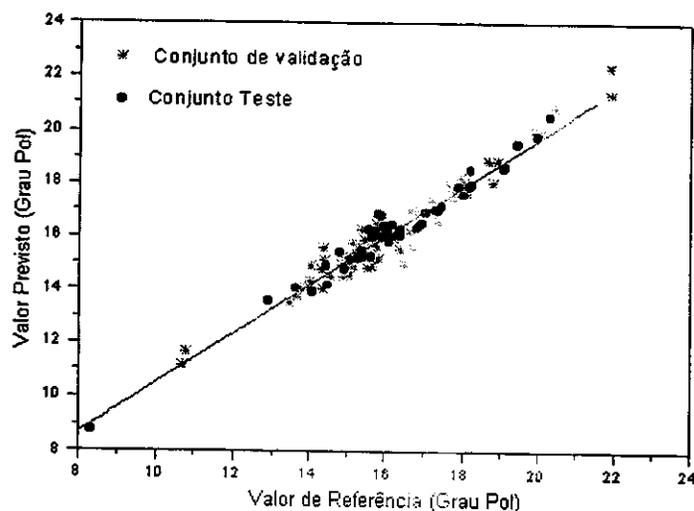


Figura 50. Valores do método referência versus os valores obtidos por NIR. As amostras do conjunto teste estão representadas pelo símbolo (●) e as do conjunto de validação, por (*).

7.8.3 Rede Neural

Analisando-se os resultados apresentados na Tabela 10, pode-se perceber que os resultados alcançados para o modelo desenvolvido mostram que as amostras do conjunto teste apresentaram um menor resíduo de previsão que as amostras do conjunto de calibração. Este tipo de comportamento é normal, pois todo modelo de calibração possui flexibilidade que permite pequenas variações na sua qualidade de previsão.

O desempenho superior do conjunto teste indica que o modelo da rede neural não foi superajustado para as amostras do conjunto de calibração. Quando isto acontece obtêm-se erros para as amostras de calibração bastante inferiores que as amostras do conjunto teste.

Os resultados obtidos para o modelo da rede neural mostraram-se numericamente superiores aos obtidos para os modelos lineares (mínimo quadrados parciais e regressão linear múltipla), com valores de SEP de 0,55 graus Brix. Contudo, acreditamos que esta diferença não possa ser considerada como sendo significativa. Portanto, para que se possa realizar uma análise mais precisa do desempenho dos modelos, optamos por utilizar o teste estatístico F para avaliar a similaridade entre as variáveis (nesse caso os valores de SEP) de duas populações. Desta forma pode-se verificar se existe alguma diferença significativa entre os resultados alcançados pelos diferentes modelos de calibração. Analisando-se os valores de F em relação ao valor de $F_{\text{Crítico}}$, Tabela 10, verifica-se que não existe diferença significativa entre os modelos lineares e rede neural.

A Figura 51 ilustra os valores previstos versus os valores do método de referência para as amostras do conjunto teste. Como pode ser observado existe uma boa concordância entre os valores ao longo de toda a curva, com um coeficiente de correlação de 0,981.

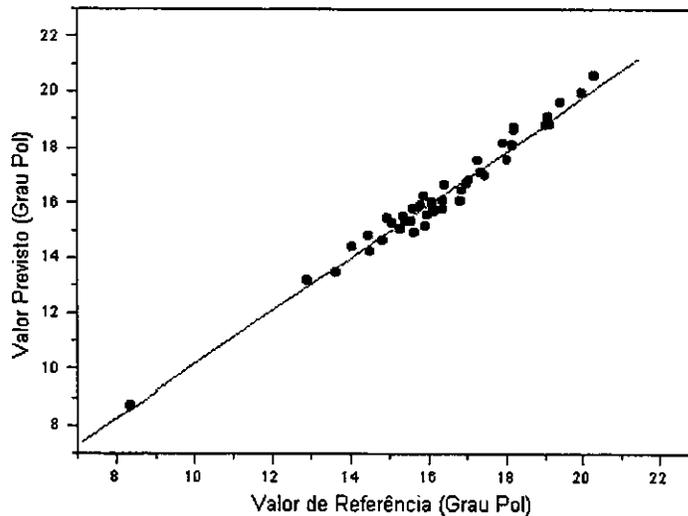


Figura 51. Valores do método referência versus os valores obtidos por NIR, utilizando a rede neural.

A Figura 52 apresenta o erro relativo de previsão das amostras do conjunto teste utilizando o modelo desenvolvido com a técnica de redes neurais. Podemos observar que o erro relativo não mostra nenhuma relação direta com o valor do grau Pol. Isso indica que a qualidade de previsão do modelo independe do valor do grau Pol.

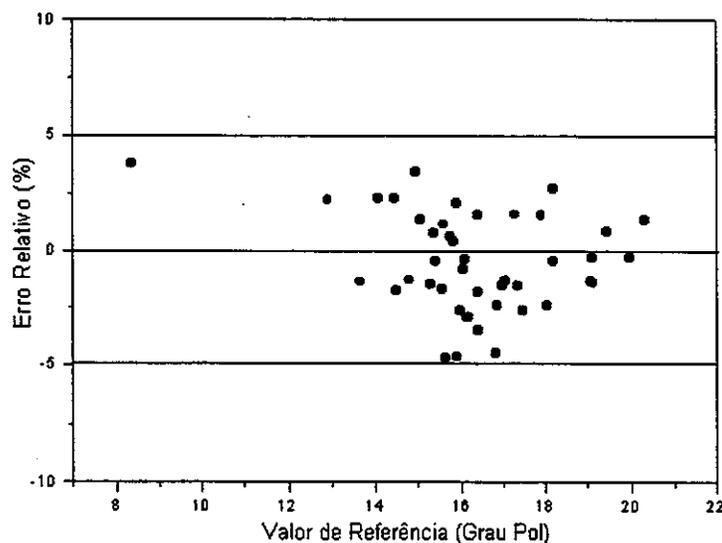


Figura 52. Erro relativo obtido na previsão do grau Pol por NIR, utilizando rede neural.

Como pode ser observado no gráfico da Figura 52, as amostras empregadas na validação do modelo de calibração apresentaram erro relativo inferior a $\pm 5\%$. Esses valores de previsão podem ser considerados de alta qualidade, visto que o método padrão de análise possui a mesma ordem de grandeza de incerteza.

Desta maneira acreditamos que este método pode ser utilizado como uma metodologia de análise alternativa, pois apresenta resultados similares ao método padrão, além de agregar atributos como velocidade de análise, redução de custos com reagentes e menor descarte de resíduos químicos. Também permite o desenvolvimento de análises *on-line*, o que possibilitará avaliar com antecedência a qualidade de um determinado lote do produto, podendo oferecer ao responsável pela produção um tempo maior para que esse possa tomar as decisões necessárias com relação a este produto.

7.8.4 Rede Neurogenética

O modelo neurogenético apresentou resultados de baixa qualidade, os quais são expressos pelo erro padrão das amostras de previsão e calibração e pelos seus respectivos coeficientes de correlação. Os valores dos erros obtidos podem ser observados na Tabela 10. Os coeficientes de correlação das amostras de calibração e teste foram de 0,781 e 0,752, respectivamente. Aparentemente, esta técnica de treinamento não foi capaz de convergir para um modelo similar ao obtido pela rede neural.

O gráfico da Figura 53 mostra a baixa correlação entre os valores previstos e os valores obtidos pelo método padrão. Podemos verificar uma alta dispersão dos pontos ao longo da curva ajustada o que indica que este modelo apresenta um alto erro relativo para as amostras do conjunto teste. O interessante deste modelo é que ele foi capaz de prever muito bem a amostra com valor próximo 8. Entretanto, para a região onde existe um número maior de amostras o desempenho é muito inferior aos dos modelos desenvolvidos para as técnicas anteriores.

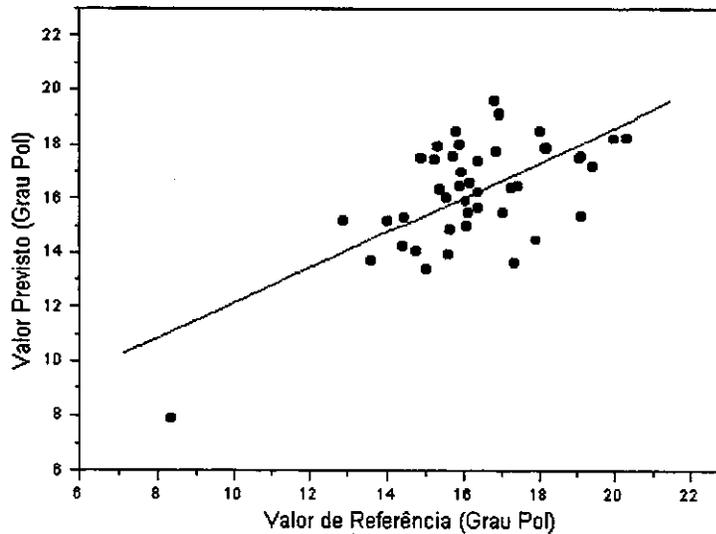


Figura 53. Valores do método de referência versus os valores obtidos por NIR, utilizando a rede neurogenética.

7.8.5 Rede Neuro-Fuzzy

O modelo Neuro-Fuzzy apresentou o pior desempenho de previsão entre as técnicas testadas para a modelagem deste sistema. O gráfico da Figura 54 ilustra a baixa correlação entre os resultados previstos e os valores obtidos pelo método de referência, com um coeficiente de correlação de 0,698.

Este modelo apresentou um comportamento muito similar ao observado no modelo neurogenético, ou seja, este foi capaz de estimar muito bem a amostra com baixo valor de grau Pol. Entretanto, para a faixa de grau Pol entre 13 e 21, a qualidade de previsão é muito baixa, com valor de SEP de 1,41 graus Pol.

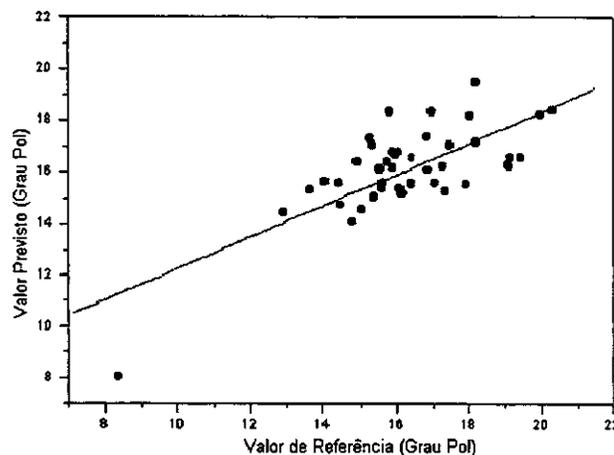


Figura 54. Valores do método de referência versus valores obtidos por NIR, utilizando a rede Neuro-Fuzzy.

7.9 Considerações Finais

Associamos o baixo desempenho dos modelos neurogenético e Neuro-Fuzzy ao fato deste sistema químico possuir predominantemente características lineares, tornando o uso destas técnicas pouco aconselhado.

O fato da técnica empregando rede neural ter apresentado resultados tão bons quanto os modelos lineares está provavelmente relacionado com a metodologia de treinamento da rede, que converge para um ponto de máximo ou mínimo de maneira mais eficiente que as empregadas pelas redes neurogenéticas e Neuro-Fuzzy.

A Tabela 10 apresenta os melhores resultados dos valores de erro padrão de previsão e calibração e o Teste F com valor crítico de 95% de confiança, para cada uma das técnicas estudadas. O Teste F mostra que não existe diferença estatisticamente significativa entre os erros de previsão dos modelos lineares e da rede neural.

Tabela 10. Valores de erro e do Teste F obtidos para os 5 modelos de calibração multivariada na estimativa do grau Pol.

	PLS	RLM	Rede Neural	Neurogenética	Neuro-Fuzzy
SEC (graus Pol)	0,57	0,52	0,57	0,95	1,04
SEP (graus Pol)	0,61	0,59	0,55	1,02	1,41
Teste - F	1,23	1,15	-	3,44	6,57

$F_{\text{Crítico } 95\%} = 1,67$

Este trabalho mostrou que é possível estimar diretamente a concentração de sacarose em amostras de cana de açúcar aplicando técnicas espectroscópicas, o que tem um impacto direto na redução do custo no controle de qualidade. Também pode permitir um controle de qualidade mais rigoroso, pois, teoricamente, é possível estimar a concentração de sacarose de todo o material produzido em diferentes fases do processo de fabricação. Isso implica em maior flexibilidade no processo de fabricação, pois o operador irá saber o momento exato que deve intervir para efetuar os ajustes necessários no processo de fabricação dos cristais de açúcar.

8. Introdução

A espectroscopia na região do infravermelho próximo é uma ferramenta interessante para o controle de qualidade. O uso desta técnica permite a realização de análises qualitativas e quantitativas em diversos tipos de matrizes, além de possibilitar a análises rápidas e, em alguns casos, sem a necessidade de pré-tratamento das amostras. Tais atributos têm despertado o interesse de diversos setores industriais como^(79,88-90) farmacêutico, alimentício, de plástico, de fertilizantes e cerâmicas. Estes setores têm demonstrado um interesse especial pela espectroscopia de reflectância difusa no infravermelho próximo em aplicações que envolvem o controle de qualidade “in-line” e “on-line”. Este interesse decorre do fato desta técnica permitir o controle de qualidade de misturas sólidas, possibilitando a caracterização, otimização e o controle de processos de misturas de materiais sólidos em tempo real.

O maior desafio na análise de misturas sólidas está no processo de amostragem, o qual deve ser representativo e não exercer qualquer tipo de influência na composição do produto final a ser analisado. O método tradicional de amostragem consiste em remover amostras de diferentes pontos na mistura. Entretanto, esta metodologia pode gerar uma série de problemas como⁽⁹¹⁻⁹⁴⁾: 1 a remoção das amostras por meio de uma sonda de amostragem causa distúrbios na composição da mistura sólida; 2 O material sólido presente na região onde a amostra foi coletada tende a segregar em direção à cavidade formada após a amostragem; 3 esta metodologia permite que seja realizada no máximo entre 10 e 20 amostragens em um curto intervalo de tempo, sendo este número insuficiente para realizar uma caracterização adequada da mistura sólida, especialmente em casos de um forte desvio da distribuição normal. Portanto, uma metodologia que permita examinar misturas sólidas “in situ” e que seja não invasiva seria uma alternativa muito interessante para este tipo de análise. Devido a uma série de vantagens, a técnica de reflectância difusa é o método não invasivo mais promissor para análise de misturas sólidas.

Além da possibilidade de estimar propriedades químicas, o espectro de reflectância difusa no infravermelho próximo fornece informações sobre as propriedades físicas do material como⁽⁹¹⁻⁹⁴⁾: tamanho de partículas, forma, densidade de empacotamento e orientação da amostra, ampliando significativamente o número de aplicações em que esta técnica pode ser utilizada.

O objetivo deste trabalho consistiu em desenvolver um modelo de calibração multivariada que permita estimar a concentração de um mineral adicionado a um polímero, portanto, após o desenvolvimento do modelo de calibração e validação, este modelo pode ser empregado

futuramente para prever a concentração do mineral introduzido ao polímero sem a necessidade de abertura de amostra.

8.1 Procedimento Experimental

A empresa Rhodia sediada em Lion (França) no forneceu um conjunto de 54 espectros de reflectância difusa de um determinado polímero, com a concentração de um mineral entre 0,00% e 1,63% em massa. A aquisição dos espectros foi realizada por um espectrofotômetro da marca NIR-System Foss 6500 acoplado a um acessório de reflectância difusa. Os espectros foram obtidos na região espectral entre 1100 - 2500 nm, com resolução de 2 nm e 32 leituras por espectro. As amostras cedidas foram previamente analisadas pela técnica de absorção atômica, onde o erro da análise é estimado em $\pm 0,008\%$.

8.2 Desenvolvimento dos Modelos de Calibração Multivariada

Para cada um dos modelos de calibração multivariada foram utilizados 42 espectros para a modelagem e 12 espectros para o conjunto teste. Nos modelos de rede neural, neurogenética e Neuro-Fuzzy, utilizaram-se como entrada, os valores escalados entre $-0,8$ e $+0,8$ dos escores obtidos pela análise de componentes principais. A saída da rede neural foi relacionada aos valores escalados da percentagem do mineral contido no polímero.

Para usar a mesma base de comparação entre os resultados obtidos pelo métodos de calibração multivariada, utilizou-se o mesmo critério de finalização para os diferentes métodos de inteligência artificial: número máximo de 600 iterações ou um erro mínimo de 10^{-5} .

8.2.1 Métodos dos Mínimos Quadrados Parciais (PLS)

O modelo dos mínimos quadrados parciais foi construído utilizando-se 8 componentes principais, sendo esse o número ideal determinado pelo método de validação cruzada.

8.2.2 Regressão Linear Múltipla com seleção de variáveis (AG/RLM)

Para o processo de seleção de variáveis, foram selecionadas 12 amostras para fazerem parte do conjunto de validação. Estas amostras foram escolhidas a partir das 42 amostras pertencentes ao conjunto de calibração utilizados pelos outros modelos de calibração multivariada. Para o conjunto teste, foram utilizadas as mesmas 12 amostras usadas nos demais modelos. Portanto, para a

regressão linear múltipla, foram utilizadas 30 amostras para o modelo, 12 amostras para a validação e 12 amostras para o conjunto teste.

O algoritmo genético foi executado utilizando a seguinte configuração: um número total de 600 iterações que representam o número de gerações, população inicial de 100 cromossomos, probabilidade de cruzamento e mutação de 90% e 1%, respectivamente.

8.2.3 Rede Neural

A arquitetura da rede neural criada foi otimizada por tentativa e erro, obtendo-se a seguinte arquitetura:- 6 neurônios de entrada, 2 neurônios na camada intermediária, com função de transferência tangente sigmoideal e 1 neurônio na camada de saída, com função de transferência linear. Utilizou-se o método de Levenberg - Marquardt para treinar a rede (correção dos pesos da rede).

8.2.4 Rede Neurogenética

A arquitetura neurogenética otimizada apresentou a configuração:- 6 neurônios de entrada, 2 neurônios na camada intermediária, com função de transferência tangente sigmoideal. Entre as entradas e a saída, foi introduzida uma conexão com função de transferência linear para a realização de uma ligação direta. O algoritmo genético para treinamento dos pesos foi configurado de maneira que as percentagens de cruzamento e mutação fossem 90% e 1% respectivamente.

8.2.5 Rede Neuro-Fuzzy

A arquitetura otimizada apresentou a configuração: duas entradas, contendo cada uma delas quatro funções de pertinência do tipo gaussiana. Os parâmetros da rede Neuro-Fuzzy foram ajustados pelo método do gradiente descendente.

8.2.6 Pré-processamento dos Espectros

Para minimizar problemas de linha base proveniente do espalhamento de luz das amostras, foi empregada a técnica conhecida como Transformação Padrão Normal de Variação (SNV). Detalhes sobre o princípio matemático empregado por esta ferramenta matemática podem ser obtidos no capítulo I desta Tese. Para a minimização do ruído, foi utilizado o filtro com transformada de Fourier.

8.3 Resultados e Discussão

Este estudo envolveu o uso de 54 espectros de reflectância difusa de um polímero específico produzido pela empresa Rhodia. As propriedades físicas e químicas deste polímero estão intimamente ligadas com a concentração de um determinado mineral que é adicionado a este material durante o seu processo de fabricação. Portanto, o controle da concentração desse mineral é extremamente importante para estimar as propriedades físicas e químicas do produto final.

Atualmente, o controle de qualidade deste material é realizado durante o processo de fabricação do polímero, onde a amostra é coletada e realizada a determinação da concentração média do mineral presente no polímero. Entretanto, segundo informações da empresa, a metodologia padrão empregada para este tipo de análise é cara e demanda um tempo relativamente grande. Esse motivo despertou o interesse da empresa em desenvolver uma metodologia alternativa de análise que forneça resultados similares e confiáveis aos alcançados pelo método de referência (absorção atômica).

Para realizar este trabalho, foi proposto um estudo empregando a técnica de espectroscopia de reflectância difusa na região espectral do infravermelho próximo. Esta decisão foi tomada com base no elevado número de trabalhos publicados que reportam o potencial desta técnica em análises quantitativas e/ou qualitativa de amostras sólidas. Além disto, esta ferramenta permite a realização de análises *in-line*, *on-line*, *at-line* e *off-line*, atendendo plenamente os requisitos traçados inicialmente neste estudo.

Por ser um trabalho em cooperação com uma empresa que exige de sigilo industrial na fabricação de seus produtos, não tivemos acesso a qualquer tipo de informação química ou física sobre a composição do polímero ou do mineral. Acreditamos que, para o objetivo deste trabalho estas informações não sejam de importância fundamental, pois o que queremos é avaliar os diferentes modelos de calibração multivariada estudados.

A Figura 55 ilustra o perfil da coleção de espectros utilizados neste estudo. Como podemos verificar os espectros possuem perfil semelhantes. Entretanto, é possível observar algumas diferenças provenientes da influência do mineral nas interações químicas de alguns grupos orgânicos do polímero.

A Figura 55 mostra a coleção de espectros separados em 5 conjuntos, os quais estão representados por diferentes cores. Em uma primeira etapa, cada um dos conjuntos foi classificado segundo o valor da concentração do mineral presente no polímero. A Tabela 11 indica a faixa de concentração em que cada conjunto foi classificado nesta primeira etapa.

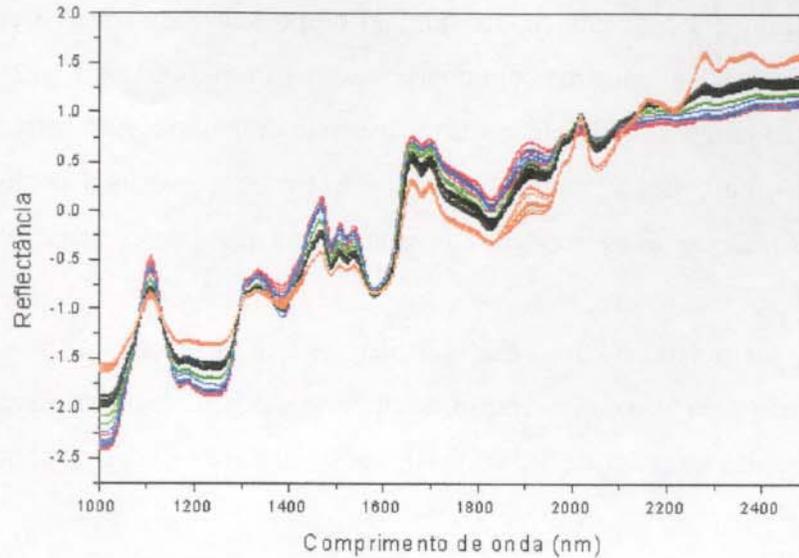


Figura 55. Espectros de reflectância das amostras do polímero com diferentes concentrações do mineral.

Tabela 11. Classificação dos conjuntos de espectros, segundo a faixa de concentração do mineral presente no polímero.

Conjunto	Concentração (%)	Cor (ver Figura 55)
1	0,00	Vermelho
2	0,010 – 0,020	Azul
3	0,110 – 0,128	Verde
4	0,300 – 0,310	Preto
5	1,630 – 1,640	Laranja

A coleção de espectros apresentados na Figura 55 estão centrados na média. Este comportamento é devido ao fato dos dados já terem sido submetidos a um pré-processamento para a correção de variação de linha base, utilizando a transformação padrão normal de variação (SNV). Optamos por utilizar esta técnica de correção de linha base por ser uma ferramenta que apresenta bom desempenho na eliminação do efeito do espalhamento de luz multiplicativo. Além disso, esta técnica apresenta a vantagem de não necessitar da informação de outros espectros para realizar a correção de linha base. Por esse motivo, esta técnica pode ser empregada em sistemas onde o perfil dos espectros mostra uma alta variabilidade.

A Figura 55 mostra que a concentração do mineral proporciona mudanças significativas no perfil do espectro do polímero. As amostras com concentração entre 0,000% e 1,640% apresentam variações de intensidade ao longo de todo o espectro, sugerindo uma correlação entre a intensidade dos espectros e a variação da concentração do mineral. Entretanto, apenas a análise visual dos espectros não permite afirmar se estas variações possuem uma correlação linear ou não com a concentração do mineral.

Comparando os espectros das 54 amostras, verificamos que as mudanças mais drásticas são observadas para as amostras com alta concentração do mineral. As principais mudanças podem ser vistas nas bandas posicionadas ao redor de 1000, 1100, 1900, 2050 nm e para os comprimentos de onda acima de 2200 nm.

As bandas que mais se destacam são as centradas em 1900, 2050 e 2250 nm. A primeira banda apresenta variações significativas com o incremento da concentração do mineral. Entretanto, o fato mais interessante é que esta banda é extremamente sensível a variação da concentração do mineral. A sensibilidade torna-se mais evidente para altas concentrações do mineral, onde uma pequena variação na concentração do mineral proporciona mudanças significativas na intensidade da banda. Essa diferença de sensibilidade surge um comportamento não linear dos dados nesta região.

A banda em 2050 nm mostra um aumento significativo na intensidade e torna-se melhor definida para as concentrações entre 1,630% - 1,640% do mineral. Entretanto, aparentemente não é possível estabelecer uma relação linear entre a variação da sua intensidade e a concentração do mineral.

No comprimento de onda ao redor de 2250 nm, podemos acompanhar a evolução do aumento da intensidade de uma banda de acordo com o aumento da concentração do mineral contido no polímero. Uma análise visual dos dados sugere que esta banda apresenta uma alta correlação linear com a concentração do mineral.

Ainda pela análise visual dos espectros da Figura 55, verifica-se que a banda em 2050 nm seja oriunda de uma interação química entre o mineral e o polímero, visto que na ausência do mineral não é possível observá-la.

A intensidade da banda em 2050 nm evolui segundo o aumento da concentração do mineral introduzido ao polímero. As amostras com concentração entre 1,630% e 1,640% do mineral apresentam uma banda intensa e bem definida.

Observamos, também, que a intensidade da banda em 2250 nm é mais sensível a pequenas variações da concentração do mineral ao redor de 0,300% e 0,310%, pois, para essa faixa de concentração, essa banda apresenta uma significativa variação na intensidade do pico. Esta variação surge que talvez haja alguma mudança no tipo de interação química existente entre o polímero e o mineral e que esta dependa diretamente da concentração do mineral.

Segundo a literatura⁽⁸⁰⁾, as bandas ao redor de 1100, 1500, 1900, 2050 e 2250 nm podem ser associadas às vibrações das ligações químicas N-H, C-H, O-H e C-O. Portanto, uma ou mais ligações destes grupos orgânicos poderiam estar interagindo diretamente com o mineral, e como resultado, alterando suas propriedades físicas e químicas.

Para selecionar as amostras do conjunto de calibração, teste e validação, optamos por utilizar a análise de componentes principais (PCA). Esta técnica auxilia na visualização da similaridade entre as amostras, tornando mais eficiente o processo de seleção das amostras que devam ser utilizadas no desenvolvimento e avaliação do modelo de calibração.

A Figura 56 apresenta o gráfico dos escores da primeira contra a segunda componente principal. Nesta Figura, podemos observar a presença de quadros grupos de amostras que estão intimamente correlacionados com a concentração do mineral no polímero. Além disso, podemos observar uma relação quadrática entre os escores da primeira e segunda componente principal, indicando uma correlação não linear entre os escores dos espectros das amostras.

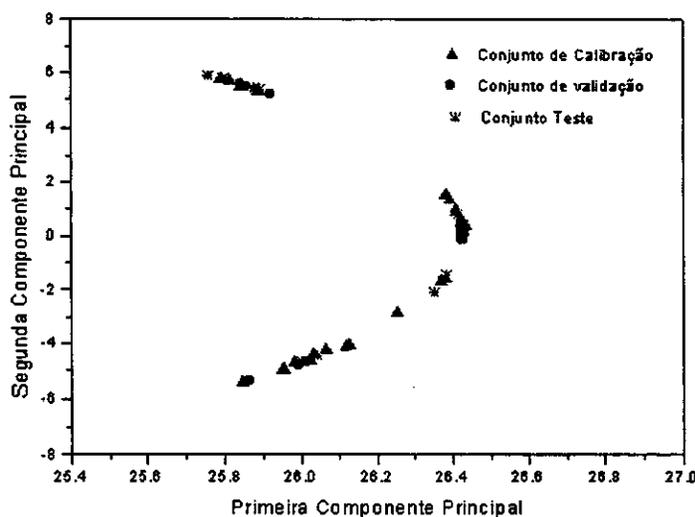


Figura 56. Gráfico dos escores da primeira contra a segunda componente principal.

A Figura 57 ilustra melhor a correlação entre os grupos de amostras e a concentração do mineral presente no polímero. Este gráfico nos permite classificar de maneira mais eficiente os grupos de amostras.

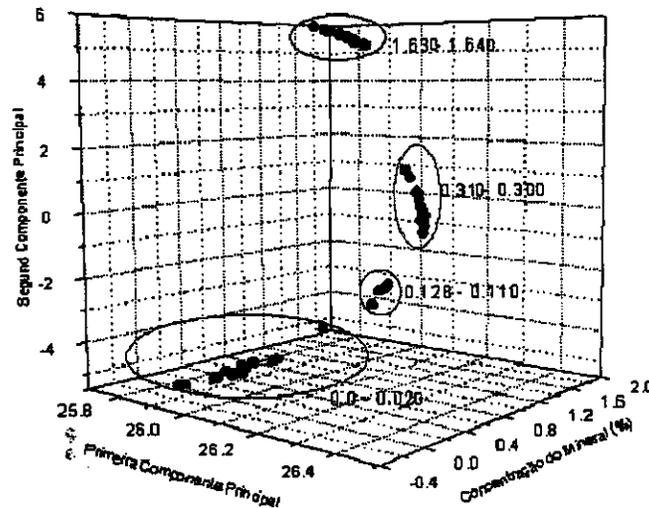


Figura 57. Gráfico da primeira e segunda componente principal contra a concentração do mineral no polímero.

A Figura 57 surge que os 5 conjuntos de amostras propostos inicialmente, Tabela 11, podem ser reduzidos para quatro grupos, conforme ilustrado pelos círculos em preto. Em cada grupo foram selecionadas algumas amostras para comporem o conjunto de calibração, validação e teste. O procedimento de selecionar as amostras a partir das informações contidas nos escores, garante que os conjuntos de amostras selecionadas sejam os mais representativos possíveis do espaço de mapeamento.

Apesar de na Figura 56 termos apresentado as amostras selecionadas para três conjuntos (calibração, validação e teste), somente iremos utilizar estes três conjuntos para o modelo de regressão linear múltipla (RLM). Os demais modelos de calibração irão usar apenas dois conjuntos (calibração e teste), onde o conjunto de calibração será composto pelas amostras do conjunto de validação e calibração do modelo de regressão linear múltipla (RLM).

A Figura 58 mostra o gráfico de influência. Este gráfico indica se existe alguma amostra que possa exercer maior influência sobre o modelo de calibração. Desta maneira, é possível detectar previamente a presença de amostras anômalas, as quais podem distorcer o modelo de calibração. O ideal para o desenvolvimento de um bom modelo de calibração é utilizar amostras que tenham níveis de influência similares.

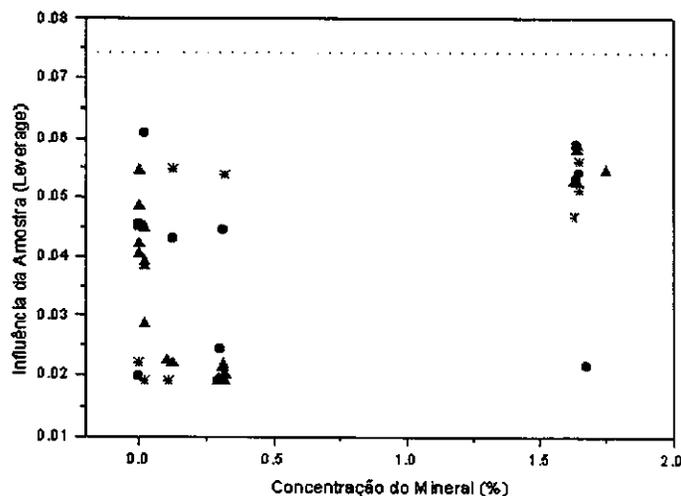


Figura 58. Gráfico de influência da amostras de polímero. As amostras do conjunto de calibração são representadas por (▲), conjunto de validação por (●) e conjunto teste por (*).

A Figura 58 indica que as amostras possuem níveis de influência distintos sobre o modelo de calibração. Entretanto, cada uma delas apresenta um nível de influência inferior ao valor máximo tolerável, mostrando desta maneira que, *a priori*, os conjuntos de calibração e validação não contêm nenhuma amostra anômala.

Apesar da Figura 58 mostrar a influência das amostras do conjunto teste, estas não possuem qualquer tipo de influência sobre o modelo de calibração. Somente as amostras do conjunto de calibração e validação apresentam influência direta ou indireta no modelo de calibração.

Neste estudo, foram desenvolvidos 5 modelos de calibração multivariada, dois modelos lineares (mínimos quadrados parciais e regressão linear múltipla) e três modelos não lineares (rede neural, neurogenética e Neuro-Fuzzy). Entretanto, existem alguns trabalhos que reportam o método dos mínimos quadrados parciais é capaz de modelar alguma não linearidade. Contudo, o método dos mínimos quadrados parciais é tradicionalmente considerado um modelo linear.

A seguir, iremos apresentar os resultados alcançados para cada um dos modelos de calibração multivariada desenvolvidos e finalizaremos com uma tabela apresentando todos os resultados obtidos para este sistema.

8.4 Modelos de Calibração Multivariada para a Determinação do Mineral

8.4.1 Modelo dos Mínimos Quadrados Parciais (PLS)

A Figura 59 apresenta o gráfico dos valores previstos pelo modelo dos mínimos quadrados parciais versus os valores da concentração do mineral presente no polímero, determinado pelo método de referência. Podemos verificar uma boa concordância entre os resultados previstos e os valores da porcentagem do mineral contido no polímero.

Para este modelo de calibração, foi obtido um coeficiente de correlação de 0,995 e um erro padrão de previsão de 0,049%, para as amostras empregadas no conjunto teste.

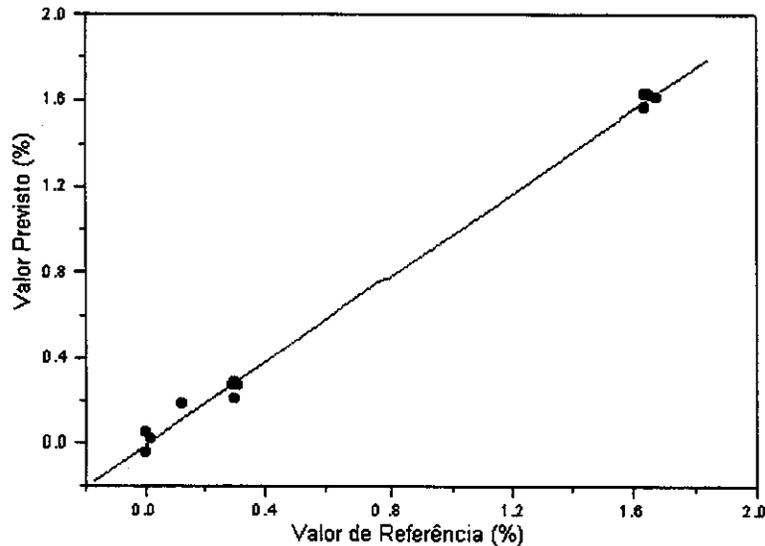


Figura 59. Valores do método referência versus os valores obtidos por NIR, utilizando o PLS.

Os resultados obtidos mostram que, apesar dos dados indicarem um comportamento não linear, o método dos mínimos quadrados parciais é capaz de produzir resultados satisfatórios.

8.4.2 Seleção de Variáveis Aplicando Algoritmo Genético (AG/RLM)

O modelo de regressão linear múltipla foi construído a partir dos comprimentos de onda selecionados pelo algoritmo genético. A Figura 60 ilustra o perfil do espectro de uma das amostras do polímero e os comprimentos de onda selecionados.

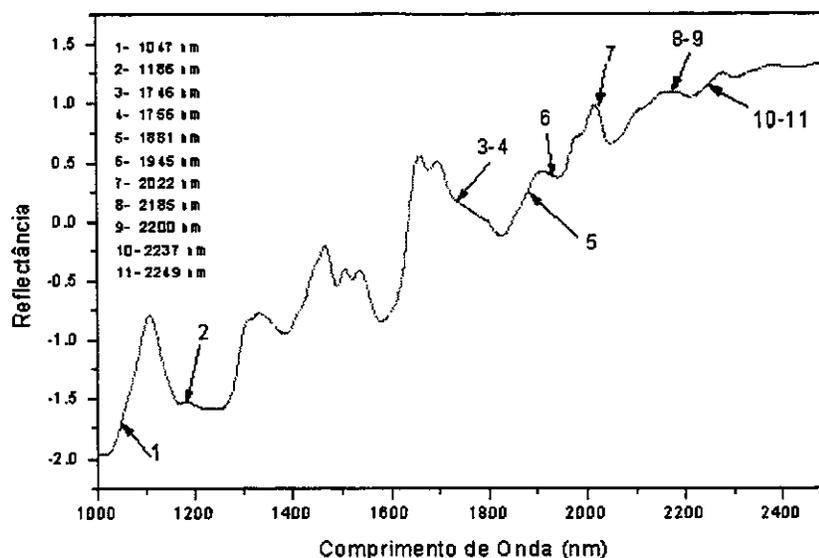


Figura 60. Comprimentos de onda selecionados pelo algoritmo genético.

Utilizando informações da literatura, foi realizada a atribuição dos comprimentos de onda selecionados pelo algoritmo genético com os possíveis estiramentos vibracionais associados a estes comprimentos de onda. Têm-se as seguintes possíveis atribuições⁽⁸⁰⁾: 1047 nm estiramento C-H e segundo sobretom da deformação -CH; 1186 nm estiramento C-H e sobretom do segundo estiramento C-C; 1746 e 1756 nm primeiro sobretom do estiramento -CH₂; 1881 nm sobretom do primeiro estiramento -OH; 1945 nm primeiro sobretom do estiramento e deformação -OH; 2022 nm segundo sobretom do estiramento C=O; 2185 nm primeiro sobretom do estiramento assimétrico -CH₂; 2237 nm primeiro sobretom do estiramento -NH e 2249 nm estiramento e deformação -OH.

Os resultados do modelo utilizando regressão linear múltipla mostram que as variáveis selecionadas apresentam um ótimo desempenho para as amostras do conjunto de validação; entretanto, o mesmo não é observado para o conjunto teste. A Figura 61 mostra que o modelo de calibração apresenta dificuldades para estimar novas amostras principalmente as com concentração próxima a 0,30%.

Os resultados obtidos pelo algoritmo genético confirmam as observações realizadas inicialmente com relação ao comportamento dos espectros. Podemos verificar que o algoritmo genético somente selecionou comprimentos de onda próximos das bandas que apresentam maior correlação com a variação da concentração do mineral (Figura 60). Por esse motivo, nenhum comprimento de onda entre 1200 nm - 1700 nm foi selecionado.

Além disso, o modelo mostrou ser menos eficiente para prever as amostras com concentração próxima a 0,300% do mineral. Esse comportamento provavelmente deve estar

relacionado ao fato do algoritmo genético ter selecionado os comprimentos de onda em 2237 nm e 2249 nm. Esses dois comprimentos de onda apresentam uma alta correlação linear (0,70) com a concentração do mineral. Contudo, o modelo de regressão linear múltipla também utiliza comprimentos de onda em 1881 nm e 1945 nm em seu modelo. Esses comprimentos de onda apresentam um comportamento bem distinto para a faixa de concentração do mineral compreendida entre 0,300% - 0,310%. Os espectros das amostras com este valor de concentração mostram uma alta sensibilidade para uma pequena variação da concentração do mineral na banda centrada em 1900 nm. Isso pode estar relacionado a alguma não linearidade do sistema para esta faixa específica de concentração, justificando, desta maneira, a deficiência do modelo para estimar a concentração do mineral nesta faixa de concentração.

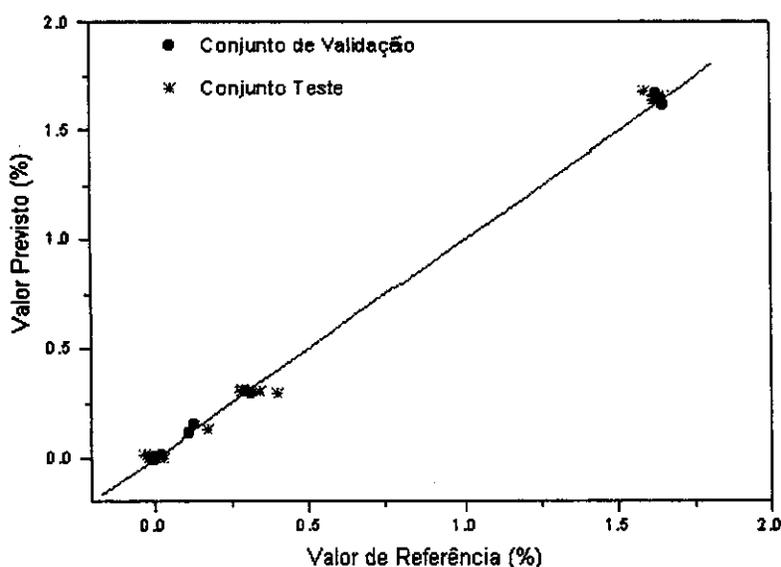


Figura 61. Valores do método referência versus os valores obtidos por NIR.

As amostras do conjunto de validação apresentaram um erro padrão de previsão de 0,021%, enquanto as amostras do conjunto de teste fornecem um erro de 0,049%, ou seja, o conjunto teste apresenta um erro de aproximadamente duas vezes maior que o obtido para as amostras do conjunto de validação.

A diferença entre a qualidade dos resultados obtidos para o conjunto de validação e teste está relacionado principalmente com o baixo desempenho de previsão do modelo para as amostras com concentração próxima de 0,300%.

Pensamos que uma maneira de minimizar este efeito seria a inclusão de um número maior de amostras ao conjunto de calibração e validação. Desta maneira, o algoritmo genético executaria o processo de seleção de variáveis de uma maneira mais eficiente, tentando selecionar outras regiões

espectrais que apresentam melhor correlação com a concentração do mineral. Entretanto, como não dispunhamos de um número maior de amostras, optamos por tentar realizar o processo de treinamento um número maior de vezes com diferentes critérios de parada. Quando trabalhamos com um número pequeno de amostras, o algoritmo genético tende a fornecer a melhor solução para este conjunto, independente se esta solução seja um caso particular do conjunto de amostras utilizada durante o processo de otimização. Uma maneira de tentar contornar este problema é alterar os parâmetros de otimização, na tentativa de tornar o algoritmo menos eficiente e, desta maneira, encontrar uma solução com desempenho inferior, mas que apresente um desempenho similar para o conjunto de validação e teste.

Inicialmente tentamos variar o número de iterações ou gerações, mantendo constante o erro mínimo desejado. Em seguida, mantivemos constante o número de iterações e variamos o valor do erro mínimo desejado.

Para ambos os casos, os resultados foram pouco satisfatórios. Quando o número de iterações é baixo, não há tempo suficiente para que o algoritmo convija para um erro similar ao obtido pelo PLS. Desta maneira, tanto o conjunto de validação como o teste apresenta erros de previsão superiores aos observados pelo PLS.

Quando utilizamos um número alto de iterações, o modelo sempre converge para um resultado similar ao apresentado na Figura 61. Por esses motivos, optamos por manter os critérios de parada similar para todos os modelos de calibração, para termos uma base de comparação similar entre eles.

8.4.3 Rede Neural

O modelo de rede neural mostrou melhor habilidade de previsão da concentração do mineral do que os métodos lineares anteriormente descritos. Foram obtidos para este modelo um coeficiente de correlação de 0,9999 e um SEP de 0,018%. Ao comparamos estes dois parâmetros com os obtidos para os modelos lineares, podemos verificar que existe uma quantidade significativa de informação que as técnicas anteriores não foram capazes de inserir ao modelo de calibração.

A Figura 62 ilustra os valores previstos pelo modelo da rede neural versus os valores de referência das amostras do conjunto teste.

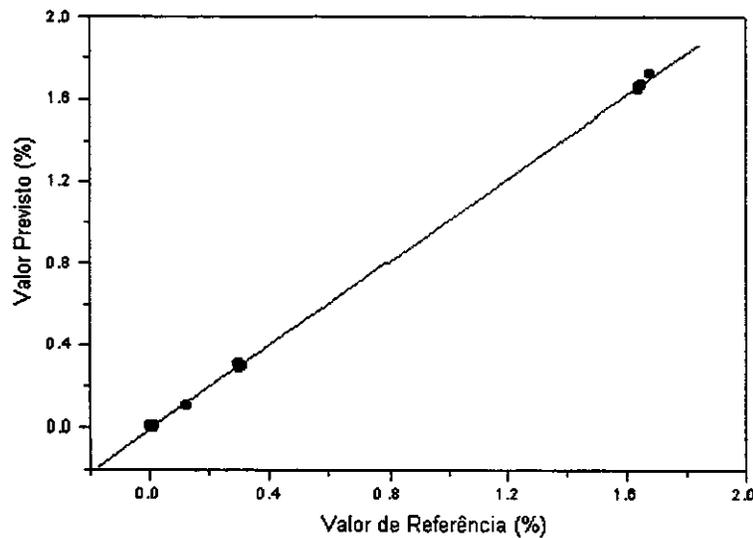


Figura 62. Valores do método referência versus os valores obtidos por NIR, utilizando a rede neural.

A Figura 62 corrobora os resultados numéricos apresentados anteriormente. Podemos observar um perfeito ajuste dos valores previstos versus os valores de referência ao longo de toda a curva, mostrando que o modelo é capaz de prever com a mesma qualidade toda a faixa de concentração utilizada neste estudo.

Os resultados obtidos pelo modelo de previsão com rede neural confirmam que existe uma parcela não linear de informações nos dados que os modelos lineares não são capazes de incluir no modelo de previsão. Desta maneira, mostra-se como os métodos não lineares podem contribuir no desenvolvimento de um modelo de calibração mais eficiente.

8.4.4 Rede Neurogenética

O modelo de rede neurogenética mostrou um desempenho melhor que os métodos lineares, contudo inferior aos observados no modelo desenvolvido pela rede neural. Este modelo apresentou um coeficiente de correlação de 0,9971 e um erro padrão de previsão de 0,021%. A Figura 63 mostra os valores previstos pelo modelo da rede neurogenética versus os valores de referência das amostras do conjunto teste.

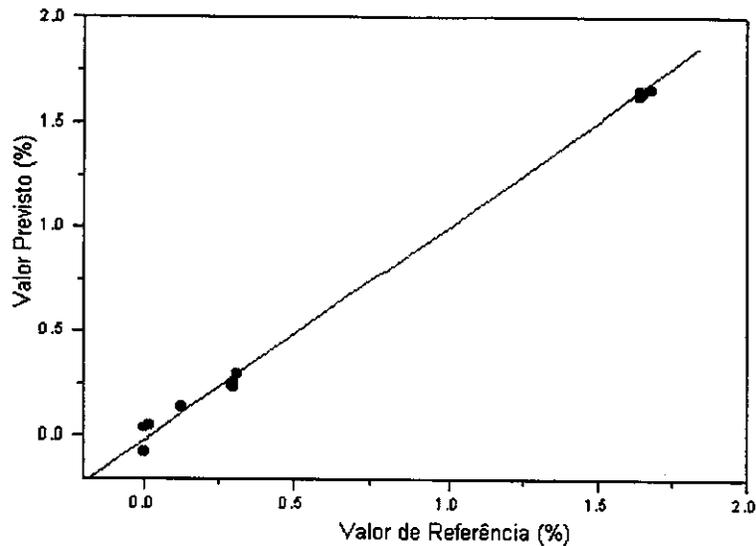


Figura 63. Valores do método referência versus os valores obtidos por NIR, utilizando a rede neurogenética.

A Figura 63 mostra que o modelo neurogenético é pouco eficiente para prever as amostras que não contêm o mineral. Este modelo não é capaz de discriminar uma amostra que possui uma concentração muito baixa do mineral de uma amostra que não contém o mineral. Entretanto, para esta aplicação, não acreditamos que este seja um problema sério, pois observamos que o modelo desenvolvido foi capaz de prever com boa exatidão uma amostra com a concentração de 0,128% do mineral. Assim, este modelo poderia ser utilizado para prever amostras na faixa de 0,12% a 1,6%.

8.4.5 Rede Neuro-Fuzzy

O modelo rede Neuro-Fuzzy mostrou um desempenho bem superior aos demais modelos de calibração multivariada. Observamos que para sistemas com comportamento não lineares, o método Neuro-Fuzzy apresenta uma diferença significativa na habilidade de previsão, quando comparado ao método de rede neural e neurogenética. Esses resultados reforçam as observações do estudo simulado desenvolvido na primeira parte desta tese.

A Figura 64 ilustra os valores previstos pelo modelo da rede Neuro-Fuzzy versus os valores de referência das amostras do conjunto teste.

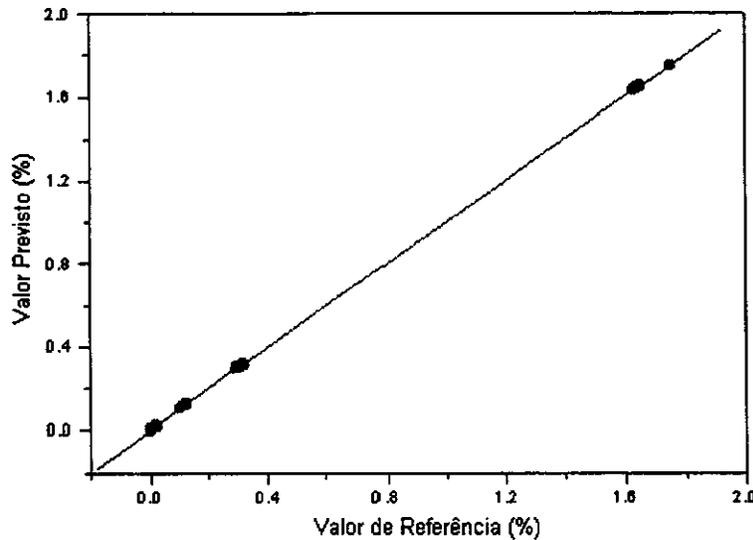


Figura 64. Valores do método referência versus os valores obtidos por NIR, utilizando a rede Neuro-Fuzzy.

A Figura 64 mostra um perfeito ajuste dos valores previstos versus os valores do método de referência. Foram obtidos um coeficiente de correlação que tende a 1 e um erro padrão de previsão de 0,009%. O erro padrão de previsão é da mesma ordem do erro do método de referência, mostrando que a união da técnica de infravermelho próximo e Neuro-Fuzzy pode oferecer resultados tão bons quanto os obtidos pelo método padrão (absorção atômica) para este sistema.

Os resultados exatos deste trabalho só foram possíveis de ser alcançados devido ao fato de que as amostras utilizadas como referência possuem um alto grau de homogeneidade. Também o método de referência é uma técnica que oferece resultados com um alto grau de precisão e exatidão. Podemos concluir que o desempenho do modelo de calibração está fortemente relacionado com a preparação das amostras e com a precisão e exatidão do método de referência.

A Figura 65 apresenta o resíduo do valor previsto pelo modelo Neuro-Fuzzy, ou seja, quanto o valor previsto é diferente do valor real. Como podemos observar, não existe nenhum indício de erro sistemático. Além disso, o valor do resíduo mostra um comportamento aleatório ao longo de toda a faixa da concentração do mineral.

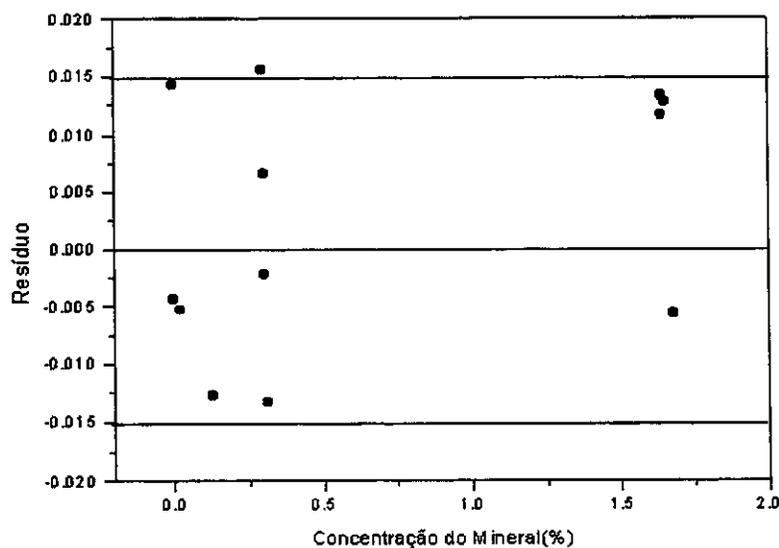


Figura 65. Resíduo obtido na previsão da concentração do mineral presente no polímero, utilizando a rede Neuro-Fuzzy.

8.5 Considerações Finais

Para concluir o estudo deste sistema, foi realizada uma análise comparativa entre os cinco modelos desenvolvidos, com o objetivo de verificar se há alguma diferença significativa entre os resultados fornecidos por cada um dos modelos de calibração.

Para comparar os cinco modelos de calibração multivariada foi aplicado o teste F com 95% de confiança, o qual permite avaliar com um grau de 95% de certeza se a variância de duas populações são similares ou não. Partindo deste princípio, pode-se avaliar se o erros de previsão de dois modelos de calibração distintos são semelhantes ou não. Entretanto, para que se possa comparar modelos de calibração diferentes é necessário usar amostras similares para os conjuntos de calibração e teste. Caso contrário, é muito difícil de avaliar se a diferença do desempenho na previsão provém devido ao conjunto de amostras utilizado ou da capacidade de previsão do modelo.

A Tabela 12 apresenta os valores do teste F calculado para os modelos de calibração empregados neste estudo. Para realizar o cálculo do teste F, utilizamos o modelo que apresentou o menor erro padrão de previsão (SEP) como referência.

O valor do $F_{\text{crítico}}$ para uma confiança de 95% é de 2,69. Os resultados mostram que os modelos do PLS, RLM, rede neural e rede neurogenética apresentam desempenho inferior ao modelo Neuro-Fuzzy. Como pode ser observado na Tabela 12, os valores do teste F são superiores ao valor do $F_{\text{crítico}}$ aceitável.

Tabela 12. Valores de erro e do Teste F obtidos para os modelos de calibração multivariados na estimativa do mineral presente no polímero.

	PLS	RLM	Rede Neural	Neurogenética	Neuro-Fuzzy
SEC (%)	0,047	0,030	0,008	0,015	0,005
SEP (%)	0,049	0,049	0,018	0,021	0,009
Teste - F	29,64	29,64	4,00	5,44	-
Teste - F	5,44	5,44	0,07	-	0,18

$F_{\text{Crítico } 95\%} = 2,69$

Os resultados também mostram que os modelos Neuro-Fuzzy e PLS são bastante robustos, já que não existe uma grande variação entre os resultados do erro padrão de calibração e previsão. Além disso, podemos afirmar que os modelos de rede neural e neurogenético apresentam valores similares entre si, indicando que não há diferenças significativas entre eles.

Portanto, podemos concluir que os modelos não lineares apresentaram desempenho superior aos lineares. Dentre os modelos não lineares, o método Neuro-Fuzzy é o mais adequado para este sistema.

9. Introdução

A espectroscopia no infravermelho tem apresentado um elevado potencial no campo de aplicações médicas, principalmente no controle preventivo e tratamento de doenças. Isso é possível devido a alterações do material biológico (proteínas, lipídios e ácidos nucleicos) que estão diretamente associadas a determinadas doenças⁽⁹⁵⁾. A alteração no material biológico provoca consideráveis mudanças no perfil do espectro do infravermelho próximo, as quais podem ser detectadas e correlacionadas com parâmetros ou espécies químicas de interesse⁽⁹⁶⁾.

Em análises clínicas ou aplicações médicas, a espectroscopia no infravermelho próximo tem mostrado um grande destaque. Essa tendência decorre do fato que o material biológico é opticamente transparente à radiação no infravermelho próximo, o que possibilita realizar análises não invasivas em seres vivos. Entretanto, a água presente em abundância nos tecidos biológicos proporciona uma série de problemas por apresentar algumas bandas intensas.

Embora existam alguns trabalhos mostrando a potencialidade da espectroscopia no infravermelho próximo em análises de matrizes biológicas, ainda não existem exemplos de aplicações de análises clínicas de rotina.

O objetivo deste trabalho foi realizar um estudo quantitativo da concentração de glicose, triglicerídeos e colesterol em plasma de seres humanos sem nenhum pré-tratamento químico e/ou físico das amostras. Este estudo empregou diretamente as medidas de absorção na região do infravermelho próximo com o objetivo de estabelecer uma relação entre a concentração do analito de interesse e os valores de absorbância medidas. Uma série de modelos de calibração multivariada lineares (PLS e RLM/AG) e não lineares (rede neural, Neuro-Fuzzy e neurogenética) foi empregada para avaliar qual o modelo mais adequado e verificar se modelos simples como a regressão linear múltipla podem ser utilizados neste tipo de análise. Este tipo de informação pode ser posteriormente empregado para propor um método alternativo de monitoramento de glicose, triglicerídeos e colesterol em plasma humano. Esta metodologia de análise poderia ser empregada em um primeiro momento como uma técnica de análise em medicina preventiva.

9.1 Procedimento Experimental

Durante um período de seis meses, o Laboratório de Análises Clínicas Canadá S. A. de Osasco – São Paulo, nos forneceu um total de 350 amostras de plasma humano, que foram transportadas para o Instituto de Química da Unicamp, após serem acondicionadas em tubos de ensaio de polietileno e congeladas abaixo de 0 °C.

As amostras foram previamente analisadas pelo instrumento TAGA 3000 – Winner que emprega metodologia enzimática. O controle de precisão e exatidão foi avaliado utilizando um padrão certificado de plasma CONTROL-LAB. Segundo o método de referência adotado pelo laboratório, as concentrações dos analitos (triglicerídeos, colesterol e glicose) possuem um desvio da ordem de ± 5 a $\pm 10\%$.

Para a aquisição dos espectros, as amostras foram descongeladas em um banho termostático até atingir $25 \pm 0,2$ °C. Posteriormente, foram transferidas para uma sala com temperatura controlada, onde foram coletados os espectros.

As medidas espectrais foram realizadas utilizando o espectrofotômetro com transformada de Fourier Bomem Hartmann & Braun DA-8, equipado com uma lâmpada de quartzo (5000 nm - 400 nm), separador de feixe (*beamsplitter*) de CaF₂ (2500 nm -400 nm) e um detector InGaAs IPH 6300L, resfriado com nitrogênio líquido.

A aquisição dos espectros foi realizada na faixa compreendida entre 714 e 2500nm, com uma resolução de 2,5 nm e uma janela óptica de 0,5 nm. O tempo de aquisição de cada espectro (64 leituras) demandou aproximadamente 2 minutos.

Os espectros de absorvância das amostras de plasma humano foram obtidos utilizando uma cubeta de quartzo de 1 mm de caminho óptico. As amostras foram introduzidas dentro da cubeta, utilizando uma seringa descartável de 1 mL. Após cada medida, a cubeta foi lavada utilizando uma solução de 1 mol L⁻¹ de ácido acético, água deionizada, uma solução de 1% de dextrana de grau clínico (Sigma), água deionizada e acetona (P.A.) para secar.

As seringas e as luvas usadas foram descartadas em uma caixa para descarte de resíduos hospitalares. Os resíduos químicos e os plasmas foram armazenados em bombonas plásticas para posterior descarte no Hospital das Clínicas da UNICAMP.

9.2 Desenvolvimento dos Modelos de Calibração Multivariada

9.2.1 Pré-processamento

A fase de pré-processamento dos dados consistiu em um processo por tentativa e erro da minimização dos erros de previsão, utilizando uma combinação de ferramentas matemáticas como: subtração do espectro da água, filtro digital baseado no algoritmo de Savitzky-Golay e transformada de Fourier para alisamento do espectro, correção de espalhamento múltiplo (MSC), análise de componentes principais (PCA), primeira e segunda derivadas e seleção de variáveis. Associado aos pré-processamentos foram testados diferentes métodos de calibração multivariada, com a finalidade de modelar um possível comportamento não linear associado às espécies químicas de interesse.

9.2.2 Seleção das Amostras do Conjunto de Calibração e Teste

As amostras do conjunto de calibração e teste foram selecionadas utilizando as informações da primeira e segunda componentes principais que explicam aproximadamente 95% da variância dos dados. Este critério foi utilizado devido ao fato destas duas componentes possuírem uma quantidade significativa de informação, permitindo a seleção do conjunto de amostras mais representativo para o modelo de calibração e teste.

9.2.2.1 Modelos de Calibração

Em cada um dos modelos de calibração multivariada foram utilizados 30 espectros para a avaliação do modelo (conjunto teste) e um número variável de amostras para a construção do modelo de calibração.

Nos modelos empregando redes neural, neurogenética e Neuro-Fuzzy, utilizou-se como entrada os valores escalados dos escores obtidos pela análise de componentes principais entre +0,8 e -0,8. A saída da rede neural foi relacionada aos valores escalados da concentração do analito de interesse.

Para realizar uma comparação entre os resultados obtidos e avaliar o desempenho de cada modelo, utilizou-se o mesmo critério de finalização entre os métodos de inteligência artificial: número máximo de 600 iterações ou um erro mínimo de 10^{-5} .

9.2.2.2 Método dos Mínimos Quadrados Parciais (PLS)

O modelo dos mínimos quadrados parciais foi construído usando um número variado de componentes principais para cada um das espécies estudadas (glicose = 3, colesterol = 3 e triglicerídeos = 10). Para determinar o número ideal de componentes principais, foi empregado o método de validação cruzada.

9.2.2.3 Regressão Linear Múltipla com Seleção de Variáveis (AG/RLM)

O algoritmo genético foi executado com uma configuração fixa para cada um dos analitos. Foi utilizado um número total de 600 iterações que representam o número de gerações, população inicial de 50 cromossomos, probabilidade de cruzamento e mutação de 90% e 1%, respectivamente.

Este método de seleção de variáveis necessita de três grupos de amostras, conjunto de calibração, conjunto de validação e conjunto teste. O conjunto de validação é empregado para avaliar as variáveis selecionadas pelo algoritmo genético durante o processo de otimização do modelo de calibração.

9.2.2.4 Rede Neural

A arquitetura da rede neural foi otimizada por tentativa e erro para cada um dos analitos estudados, obtendo-se as seguintes configurações: - **Glicose**: 3 neurônios de entrada, 3 neurônios na camada intermediária com função de transferência tangente sigmoide e 1 neurônio na camada de saída com função de transferência linear; - **Colesterol**: 3 neurônios de entrada, 2 neurônios na camada intermediária com função de transferência tangente sigmoide e 1 neurônio na camada de saída com função de transferência linear; - **Triglicerídeos**: 8 neurônios de entrada, 4 neurônios na camada intermediária com função de transferência tangente sigmoide e 1 neurônio na camada de saída com função de transferência linear.

O treinamento da rede foi realizado utilizando-se o método de Levenberg - Marquardt.

9.2.2.5 Rede Neurogenética

A arquitetura da rede neurogenética foi otimizada por tentativa e erro. Depois de várias tentativas de otimização, chegou-se à conclusão que a melhor arquitetura foi a mesma empregada para a rede neural. O algoritmo genético para treinamento dos pesos foi configurado de maneira que as porcentagens de cruzamento e mutação fossem de 90% e 1%, respectivamente.

9.2.2.6 Rede Neuro-Fuzzy

A arquitetura da rede Neuro-Fuzzy otimizada apresentou a seguinte configuração: - **Glicose**: 2 entradas com 2 funções de pertinência do tipo gaussiana cada uma; - **Colesterol**: 2 entradas com 3 funções de pertinência do tipo gaussiana cada uma; - **Triglicérides**: 3 entradas com 2 funções de pertinência do tipo gaussiana cada uma. Os parâmetros da rede Neuro-Fuzzy foram ajustados pelo método do gradiente descendente.

9.3 Determinação de Triglicerídeos em Plasma Humano

Os triglicerídeos são a forma química na qual uma série de gorduras apresentam-se nos alimentos que ingerimos diariamente. Após a ingestão dos alimentos, uma parte dos triglicerídeos contidos nos alimentos é transferida para a corrente sanguínea.

A análise do plasma humano mostra que os triglicerídeos presentes no plasma são derivados não apenas das gorduras ingeridas durante as refeições. Uma parcela dos triglicerídeos é também produzida em nosso corpo, a partir de outras fontes de energia como os carboidratos.

O excesso de triglicerídeos no plasma é chamado de hipertrigliceridemia e pode levar a formação de placas microscópicas (arterosclerótica), que causam a obstrução das artérias coronárias. Isso pode ocasionar graves problemas cardíacos ao homem⁽⁹⁷⁻⁹⁸⁾. De acordo com a Associação Americana de Medicina, a melhor maneira de prevenir os problemas cardíacos é adotando uma alimentação mais saudável composta de frutas, vegetais, cereais e carne com baixos níveis de gordura animal⁽⁹⁷⁾. Entretanto, mesmo adotando estas medidas preventivas, é necessário o controle periódico do teor de triglicerídeos no plasma humano.

Normalmente, as pessoas são levadas a acreditar que a elevada taxa de triglicerídeos no sangue esteja relacionada apenas a uma alimentação pouco saudável. Contudo, em alguns casos a elevada concentração de triglicerídeos pode estar relacionada a outras doenças, como por exemplo, a diabetes não tratadas⁽⁹⁷⁾.

Segundo a Associação Americana de Medicina, os níveis de triglicerídeos nos adultos são definidos como: níveis normais, com valores inferiores a 200 mg dL⁻¹; níveis intermediários quando estão entre 200 a 400 mg dL⁻¹; níveis altos quando estão entre 400 a 1000 mg dL⁻¹ e níveis muito altos quando a sua concentração supera 1000 mg dL⁻¹⁽⁹⁷⁾.

Acreditamos que seja possível desenvolver uma metodologia de análise do teor de triglicerídeos baseado em espectroscopia no infravermelho próximo, que possa contribuir significativamente ao avanço das análises de interesse biológico. Isso possibilitará a realização de análises a custos mais baixos e, simultaneamente, proporcionar um aumento significativo na velocidade da análise, permitindo diagnosticar e avaliar a evolução ou retrocesso de algumas doenças como, por exemplo, a hipertrigliceridemia e a diabetes.

A Figura 66 ilustra a fórmula estrutural química genérica dos triglicerídeos, onde cada grupo R representa a estrutura química de um hidrocarboneto.

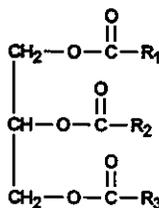


Figura 66. Fórmula estrutural química genérica dos triglicerídeos.

9.4. Resultados e Discussão

Nesta seção, serão apresentados os resultados obtidos na determinação da concentração de triglicerídeos em amostras de plasma humano. Este estudo consistiu em estabelecer um modelo de calibração multivariada entre os espectros coletados na região do infravermelho próximo e os valores de concentração de triglicerídeos. As Figuras 67.a e 67.b mostram o conjunto de espectros obtidos para cada amostra de plasma e a sua primeira derivada, respectivamente.

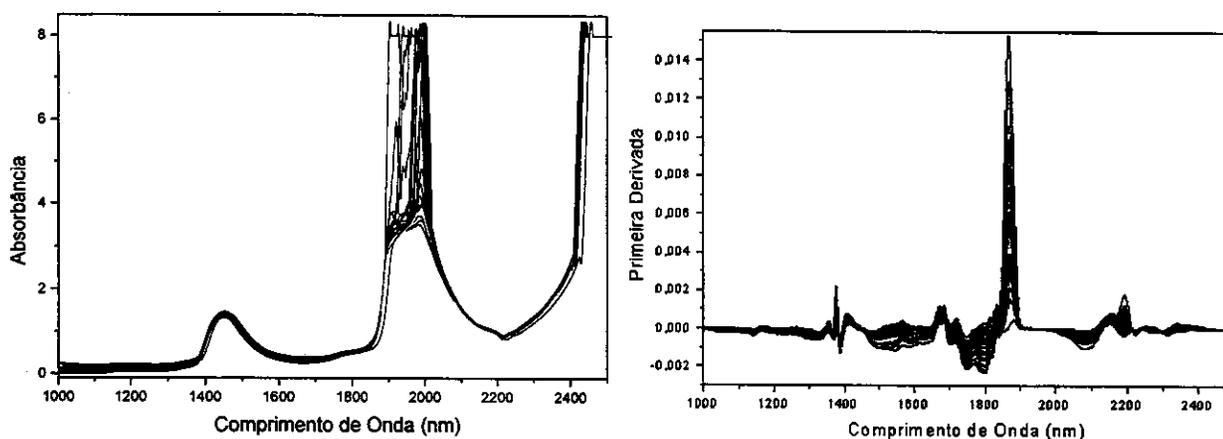


Figura 67. a. Coleção de espectros de NIR de plasma humano. b. Primeira derivada da coleção de espectro na região do infravermelho próximo.

Na Figura 67.a, é possível observar as bandas características de absorção de água. Trata-se de duas bandas que apresentam intensa absorção do grupo OH, localizadas na região ao redor de 2000 nm e de 1450 nm.

Devido ao intenso sinal da banda de água, a realização de estudos envolvendo analitos em matrizes biológicas é considerado um desafio fascinante. Espécies químicas como triglicerídeos, colesterol, proteínas e glicose apresentam absorção na mesma faixa espectral da banda de água. Além disso, as concentrações destas espécies são muito baixas, o que proporciona um sinal com ordem de grandeza muito inferior ao da banda de água. O estudo qualitativo e/ou quantitativo destas

espécies torna-se difícil de ser realizado sem a utilização de equipamentos de alta precisão e ferramentas matemáticas adequadas.

Para contornar ou minimizar estes problemas, é possível utilizar recursos que minimizem o sinal da banda de água. Um dos recursos mais comuns é o uso de uma cubeta com caminho óptico de 1 mm e pré-processamentos matemáticos dos espectros. Além disso, normalmente é necessário utilizar outras ferramentas matemáticas que minimizem variações sistemáticas dos dados como os problemas de linha base e ruídos instrumentais. Neste trabalho foram utilizados ambos os procedimentos citados anteriormente, com o objetivo de minimizar o sinal da banda da água e preservar as informações dos analitos de interesse neste estudo.

Antes da construção dos modelos de calibração multivariada, foi realizada a seleção do conjunto de amostras mais representativas do espaço amostral estudado. Para isto, foi usado os escores da primeira versus da segunda componente principal, conforme ilustrado na Figura 68. O gráfico da Figura 68 ilustra os três grupos de amostras utilizados neste estudo (conjunto de calibração, teste e validação).

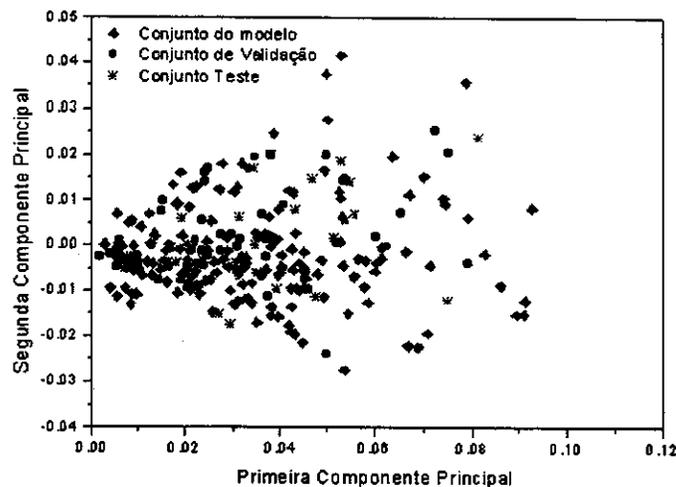


Figura 68. Gráfico dos escores da primeira versus a segunda componente principal.

A priori, a seleção de variáveis necessita de um conjunto de amostras para construir o modelo de calibração e um segundo conjunto de amostras para avaliar se as variáveis selecionadas apresentam um bom desempenho de previsão (conjunto de validação). Além desses dois conjuntos, a seleção de variáveis emprega um conjunto adicional de amostras externas (conjunto teste) para garantir que o conjunto de variáveis selecionadas não é uma solução específica para as amostras do conjunto de validação. Isso ocorre porque a seleção de variáveis é um processo iterativo que se baseia em informações de um conjunto restrito de amostras.

Em alguns casos, as amostras do conjunto de validação podem não representar suficientemente bem o comportamento global do sistema estudado, gerando um modelo de calibração que oferece bons resultados apenas para o caso particular das amostras do conjunto de validação.

Os demais métodos de calibração empregados nesta tese utilizam informações apenas do conjunto de calibração, necessitando apenas de um conjunto teste externo de amostras para avaliar a qualidade do modelo desenvolvido. No caso da seleção de variáveis, um subconjunto das amostras do conjunto de calibração foi selecionado para compor o conjunto de amostras de validação.

Para os modelos de rede neural, neurogenética e Neuro-Fuzzy, o número de amostras é um fator importante, pois determina a qualidade do mapeamento espacial. Observamos que mesmo em sistemas com comportamento linear, o uso de modelos não lineares necessita de um alto número de amostras pois, desta forma, a velocidade de convergência do modelo é mais rápida. Por este motivo, optamos por utilizar o maior número de amostras possíveis para construir o modelo de calibração. Contudo, o número de amostras do conjunto de calibração não deve crescer indefinidamente, pois existe um compromisso entre a habilidade de previsão e a imprecisão dos dados inseridos no modelo de calibração.

Depois de selecionadas as amostras de cada um dos grupos (calibração, teste e validação), utilizou-se a distância de Mahalanobis para avaliar o grau de influência de cada amostra. Esta técnica permite avaliar se alguma amostra apresenta um comportamento distinto com relação a um determinado grupo de amostras. Tal comportamento pode ser interpretado como um possível “outlier” (amostra anômala). Portanto, a exclusão desta amostra seria o procedimento mais adequado.

A Figura 69 ilustra o gráfico de influência das amostras utilizadas neste estudo. Como podemos verificar, todas as amostras apresentam valores inferiores ao do limite de tolerância aceitável, o qual é representado pela linha tracejada.

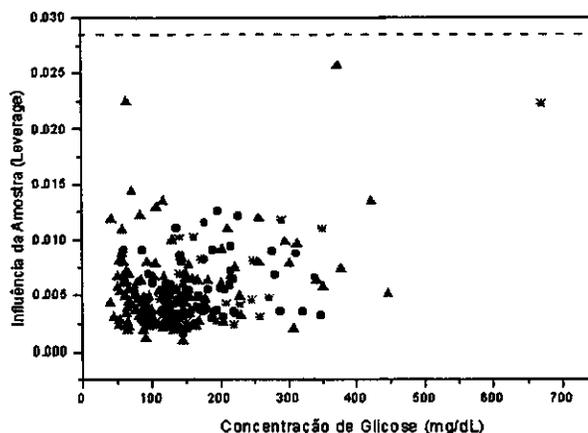


Figura 69. Gráfico da influência de cada amostra, onde as amostras do conjunto de calibração são representadas pelo símbolo (▲), conjunto teste por (●) e conjunto de validação por (*).

O estudo desenvolvido para a determinação de triglicerídeos mostrou que os melhores resultados foram alcançados após ser utilizado a transformada de Fourier para minimizar o ruído, seguida pela correção de linha base, subtração do espectro de água e o uso da primeira derivada. O uso da primeira derivada tem como objetivo a deconvolução das informações sobrepostas no espectro e minimizar o problema de linha base.

Segundo a literatura⁽⁹⁶⁾, a maioria das informações relacionadas aos triglicerídeos encontra-se nas faixas espectrais entre (1600 nm e 1850 nm) e (2222 nm e 2381 nm). Por este motivo, concentramos nossos esforços iniciais nos comprimentos de onda próximos a estas regiões. Após um exaustivo trabalho que envolveu a seleção de diferentes faixas espectrais próximas às regiões anteriormente relatadas, seguido por uma série de combinações de diferentes pré-processamentos dos dados, chegou-se à conclusão que a melhor faixa espectral encontra-se na região entre 1850 nm e 1450 nm.

A Figura 70 mostra o perfil do conjunto de espectros ao final da etapa de pré-processamento dos dados. Observou-se que o sinal dos espectros apresenta baixa intensidade. Este comportamento já era esperado devido à baixa concentração das substâncias de interesse biológico presentes nas amostras de plasma humano.

Este fato evidencia a necessidade de utilizar um espectrofotômetro com alta relação sinal/ruído para este tipo de estudo. Caso contrário após o final da etapa de pré-processamento o ruído teria sido amplificado pela primeira derivada para a mesma ordem de grandeza do sinal do analito. Isso provavelmente inviabilizaria o estudo na faixa de concentração proposta neste trabalho.

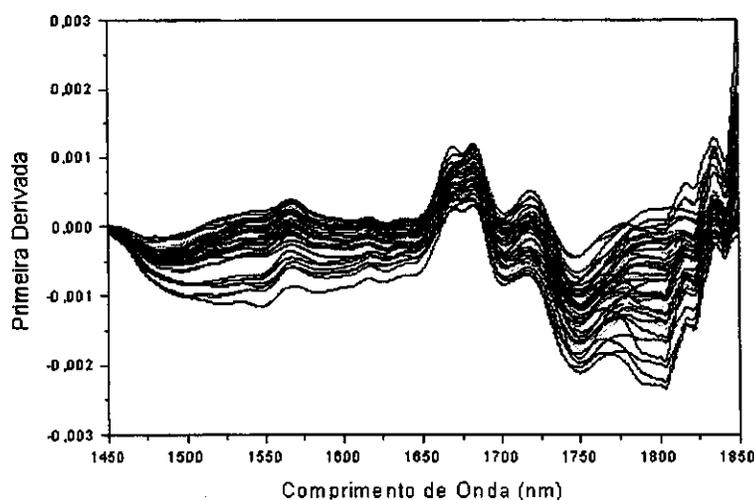


Figura 70. Primeira derivada do conjunto de amostras de plasma humano na faixa espectral entre 1450 nm e 1850 nm.

9.4.1 Modelo dos Mínimos Quadrados Parciais (PLS)

Os resultados obtidos usando o modelo dos mínimos quadrados parciais (PLS) mostram um bom desempenho na estimativa da concentração de triglicerídeos. Este estudo envolveu 230 amostras, sendo destinadas 200 amostras para construção do modelo e 30 amostras (conjunto teste) foram empregadas para avaliar a qualidade de previsão do modelo calibração.

A Figura 71 mostra uma boa correlação (coeficiente de correlação de 0,977) entre o valor de concentração previsto e o valor de concentração considerado como referência.

O modelo desenvolvido apresentou um erro ao redor de $\pm 10\%$ e um erro padrão de previsão (SEP) de $17,89 \text{ mg dL}^{-1}$. Entretanto, é importante ressaltar que o erro do método de referência tem desvio na faixa entre $\pm 5\text{--}\pm 10\%$. Portanto, considerando a imprecisão do método de referência, podemos concluir que os resultados de previsão para este tipo de análise mostraram um bom desempenho. Além disso, observou-se que o modelo de calibração multivariada possui um comportamento linear para uma extensa faixa de concentração, conforme ilustrado na Figura 71 onde existe uma amostra com concentração de 628 mg dL^{-1} . Entretanto, acreditamos que seja pouco prudente afirmar que este modelo possa ser utilizado para estimar concentrações tão altas, pois não existem amostras suficientes para avaliar se o modelo realmente apresenta um bom desempenho de previsão para concentrações acima de 300 mg dL^{-1} . Contudo, acreditamos que este seja um forte indício que este modelo seja capaz de prever amostras com concentração de triglicerídeos entre 80 a 600 mg dL^{-1} .

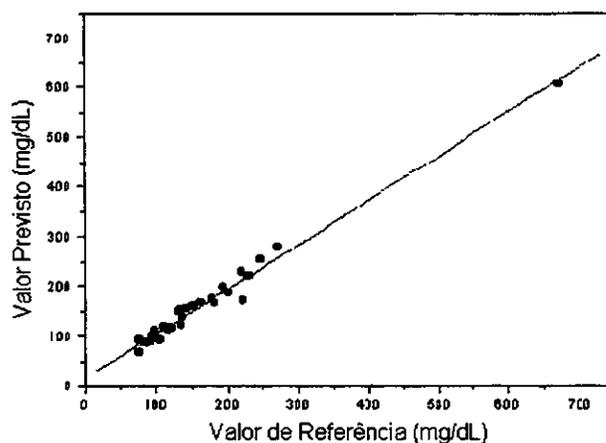


Figura 71. Valores do método referência versus os valores obtidos por NIR.

9.4.2 Seleção de Variáveis Aplicando Algoritmo Genético (AG/RLM)

O algoritmo genético foi empregado para selecionar o melhor subconjunto de variáveis que estabeleça uma relação linear entre concentração e absorvância. Desta maneira, é possível a utilização de um modelo matemático mais simples, como a regressão linear múltipla (RLM). Além disso, a seleção de variáveis elimina as variáveis que possuem pouca ou nenhuma informação associada à espécie de interesse. Para o desenvolvimento do modelo de calibração, foram usadas 140 amostras, para o conjunto de validação 60 amostras e para o conjunto teste 30 amostras.

A Figura 72 apresenta o conjunto de variáveis (comprimentos de onda) selecionadas pelo algoritmo genético. Foram selecionados 17 comprimentos de onda (1801, 1784, 1756, 1746, 1734, 1726, 1715, 1714, 1699, 1684, 1623, 1590, 1585, 1514, 1568, 1522 e 1485 nm), os quais estão relacionados com sobretons e bandas de combinação das ligações⁽⁸⁰⁾ -CH, -CH₂, -CH₃ e -OH.

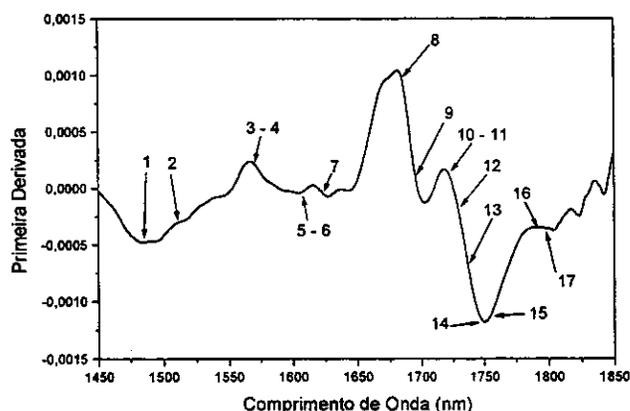


Figura 72. Primeira derivada do espectro de plasma humano e os comprimentos de onda selecionados pelo algoritmo genético.

Os resultados obtidos mostram um desempenho similar aos alcançados pelo método dos mínimos quadrados parciais, com valores de SEP de 13,76 e 16,93 mg dL⁻¹, para os conjuntos de validação e teste.

A diferença do valor de SEP entre os dois conjuntos está associada ao fato de que as amostras do conjunto teste possuem maior variabilidade no valor da concentração que as do conjunto de validação.

Como podemos observar na Figura 73, que mostra os valores previstos pela metodologia proposta contra os valores de referência, a variabilidade do conjunto teste é maior, devido à presença de uma amostra com valor de concentração de triglicerídeo de 628 mg dL⁻¹. Esta amostra é a principal responsável pela diferença observada entre os valores do SEP dos conjuntos de validação e teste. Isso ocorre porque o SEP é calculado sobre o resíduo obtido da diferença entre o valor previsto e o fornecido pelo método padrão. Isso implica que o resíduo possui uma influência direta no valor do SEP. Além disso, devemos lembrar que o resíduo é elevado ao quadrado, portanto, a contribuição das amostras com resíduos maiores é potencializada em relação às amostras com resíduos menores. Entretanto, é importante ressaltar que o erro relativo para esta amostra é similar ao obtido para as outras amostras, indicando que o modelo possui a mesma qualidade de previsão ao longo de toda a faixa de concentração estudada.

Concluindo, o SEP e SEC somente podem ser comparados diretamente quando operamos sobre faixas de concentrações similares. Em sistemas que operam em faixas de concentrações distintas, é necessário tomar cuidados adicionais para que não sejam realizadas interpretações equivocadas. Por este motivo, neste estudo, o SEP e o erro relativo foram avaliados em conjunto durante a análise da robustez do modelo.

A forma mais simples de contornar este problema seria excluir a amostra com alto valor de triglicerídeos do conjunto teste. Entretanto, nós acreditamos que esta amostra indica que seja possível desenvolver um modelo de calibração capaz de prever amostras entre 80 a 600 mg dL⁻¹ de triglicerídeos, fornecendo desta maneira uma importante informação adicional a este estudo.

A Figura 73 mostra os resultados de previsão obtidos pela regressão linear múltipla versus os valores do método de referência. A reta obtida a partir dos dados do conjunto de validação mostra um coeficiente de correlação de 0,979 e, para os de teste, de 0,987.

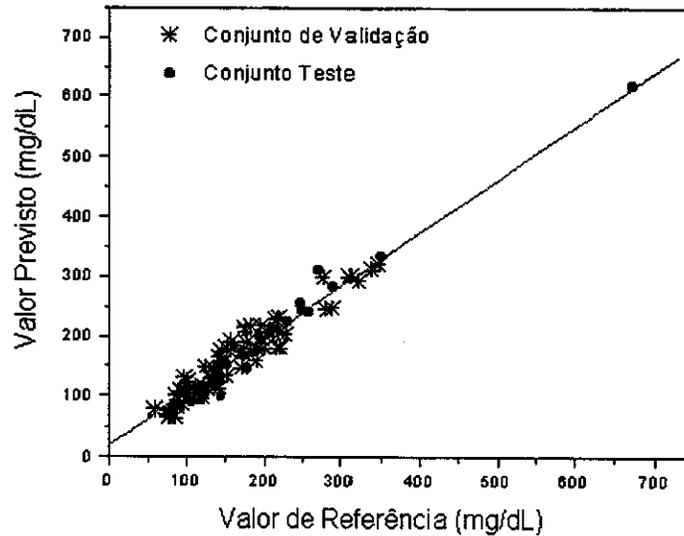


Figura 73. Valores do método referência versus os valores obtidos por NIR, utilizando a regressão linear múltipla. As amostras do conjunto teste estão representadas pelo símbolo (•) e as do conjunto de validação, por (*).

9.4.3 Rede Neural

O modelo de rede neural foi empregado na tentativa de modelar alguma não linearidade do sistema.

Os resultados obtidos mostram que o modelo com rede neural tem um desempenho similar ao alcançado pelos modelos lineares (RLM e PLS), indicando que os espectros no infravermelho próximo das amostras de plasma humano apresentam predominantemente uma relação linear com concentração de triglicerídeos.

A Figura 74 mostra os resultados de previsão obtidos pela rede neural versus os valores do método de referência. A reta ajustada aos dados tem um coeficiente de correlação de 0,985, sendo obtido um SEP de $14,73 \text{ mg dL}^{-1}$.

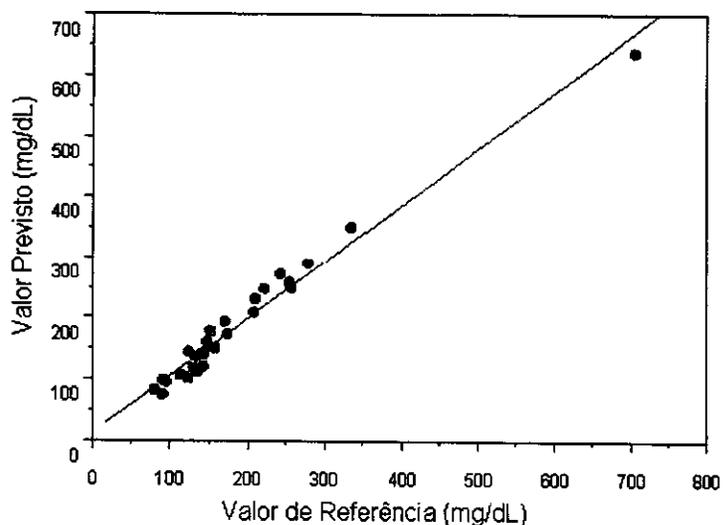


Figura 74. Valores do método referência versus os valores obtidos por NIR, utilizando a rede neural.

A Figura 75 ilustra o erro relativo associado às amostras do conjunto teste ao longo da faixa de concentração estudada. A Figura 76 mostra que o erro relativo possui um comportamento aleatório ao redor do zero (tomando como base todas as amostras), indicando que não há qualquer tipo de erro sistemático associado ao modelo desenvolvido.

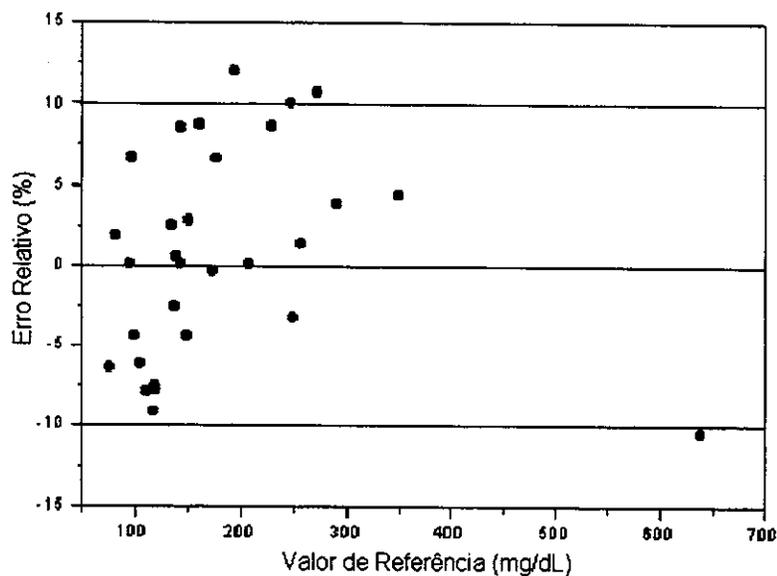


Figura 75. Erro relativo obtido na previsão da concentração de triglicerídeos por NIR.

Como poder ser constatado na Figura 75, a amostra com concentração de 628 mg dL^{-1} de triglicerídeos apresenta um erro relativo dentro da faixa de variação do método de referência ($\pm 10\%$).

9.4.4 Rede Neurogenética

O modelo de rede Neurogenético apresentou um desempenho de previsão inferior, quando comparado aos demais métodos de calibração multivariada.

A Figura 76 mostra os resultados de previsão obtidos pela rede neurogenética versus os valores do método de referência. A reta ajustada aos dados apresentaram um coeficiente de correlação de 0,869, sendo obtido um SEP igual a $21,32 \text{ mg dL}^{-1}$.

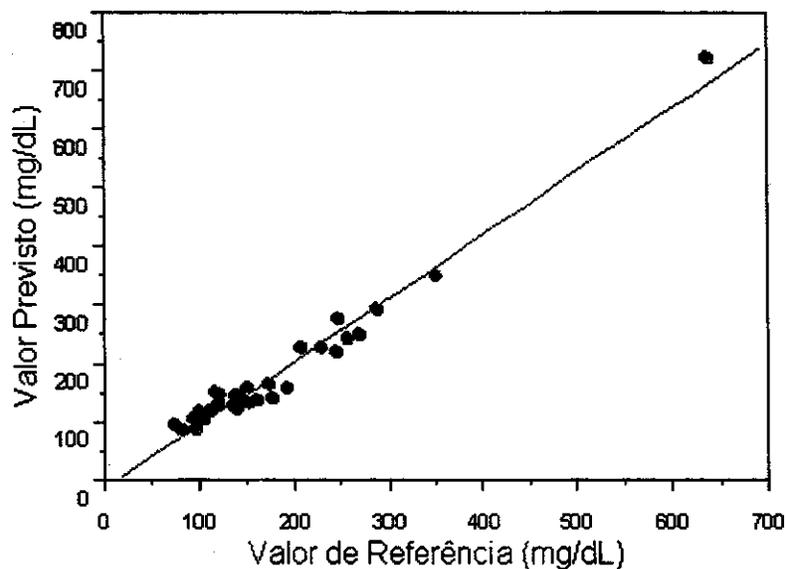


Figura 76. Valores do método referência versus os valores obtidos por NIR, utilizando a rede neurogenética.

9.4.5 Rede Neuro-Fuzzy

O modelo Neuro-Fuzzy mostrou um desempenho similar ao apresentado pelo modelo neurogenético. Como havíamos observado anteriormente, esta técnica apesar de muito poderosa, apresenta baixo desempenho na modelagem de dados com comportamento linear. Entretanto, acreditávamos que este modelo pudesse contribuir de alguma maneira para minimizar o problema na imprecisão no valor da concentração do analito de interesse. Contudo, não observamos nenhuma contribuição deste modelo nesta direção e que justifique sua aplicação na modelagem deste sistema em particular.

A Figura 77 mostra os resultados de previsão obtidos pela rede neural versus os valores do método padrão. A reta ajustada aos dados apresentou um coeficiente de correlação de 0,911, sendo obtido um SEP igual a 20,75 mg dL⁻¹.

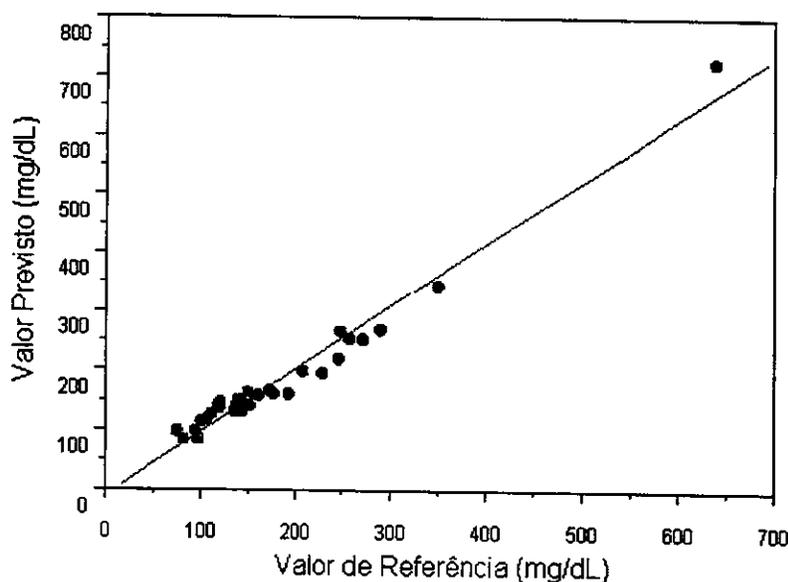


Figura 77. Valores do método referência versus os valores obtidos por NIR, utilizando a rede Neuro-Fuzzy.

9.5 Considerações Finais

Para concluir o estudo que envolveu a determinação da concentração de triglicérides em amostras de plasma humano, realizamos um estudo comparativo entre os modelos para verificar se existe alguma diferença significativa entre eles, além de tentar observar se os modelos não lineares apresentam alguma contribuição na modelagem deste tipo de dados.

Este trabalho consistiu em utilizar o teste F com valor crítico de 95%, que permite avaliar com um grau de 95 % de certeza se a variância de duas populações são similares ou não. Partindo deste princípio pode-se avaliar se o erro de previsão de dois modelos de calibração distintos usando o mesmo grupo de amostras são semelhantes ou não.

Segundo o teste $F_{95\%}$ os modelos do PLS, RLM e rede neural apresentam desempenho similares, pois o valor do $F_{\text{crítico}}$ é de 1,84. Entretanto, para fins práticos, estamos fortemente inclinados a recomendar o método dos mínimos quadrados parciais ou regressão linear múltipla. Apesar de termos observado que o modelo com rede neural apresentou o melhor modelo de

previsão, a complexidade do modelo não justifica a ganho no desempenho de previsão, pois em ambos os casos, os resultados encontram-se dentro do limite tolerável do erro de previsão para este tipo de análise.

Este estudo também mostrou que os modelos de redes neurogenético e Neuro-Fuzzy não são aconselháveis de serem aplicados, pois além da maior complexidade, apresentam um desempenho de previsão estatisticamente inferior aos demais modelos de calibração, conforme pode ser observado na Tabela 13.

Tabela 13. Valores de erro e do Teste F obtidos para os 5 modelos de calibração multivariados na estimativa de triglicérides em plasma.

	PLS	RLM	Rede Neural	Neurogenética	Neuro-Fuzzy
SEC (mg dL ⁻¹)	17,35	14,95	12,50	18,65	19,57
SEP (mg dL ⁻¹)	17,89	16,93	14,73	21,32	20,75
Teste-F	1,47	1,15	-	2,09	1,98

F_{Crítico 95} = 1,84

9.6 Determinação de Colesterol em Plasma humano

No corpo humano o colesterol é usado como matéria prima para a produção de membranas celulares, de alguns hormônios e de tecidos biológicos. A Figura 79 ilustra a fórmula estrutural química da molécula de colesterol. O transporte do colesterol entre as células do corpo humano é realizado pelos lipídios. Entretanto, quando dissolvido no sangue, o colesterol é transportado por lipoproteínas. Estas lipoproteínas são classificadas em lipoproteínas de baixa densidade (LDL) e lipoproteínas de alta densidade (HDL). O elevado nível de LDL colesterol (superior a 130 mg dL^{-1}) no sangue é chamado de hipercolesterolemia^(97,99).

Aproximadamente de um terço a um quarto do colesterol transportado no sangue é realizado por lipoproteínas de alta densidade. Acredita-se que o HDL remove o excesso de colesterol das artérias para o fígado. Portanto, o baixo nível de HDL (inferior a 35 mg dL^{-1}) pode ser um alto risco a saúde humana. Por este motivo, nas análises de saúde de rotina os médicos solicitam os níveis de LDL e HDL de seus pacientes⁽⁹⁷⁾.

As lipoproteínas de baixa densidade juntamente com outras substâncias formam placas finas e duras que se depositam lentamente nas paredes das artérias que irrigam o coração e o cérebro. O bloqueamento das artérias pode ocasionar o infarto levando a morte ou a perda total ou parcial do movimento de algumas partes do corpo⁽⁹⁷⁾.

Essencialmente, o colesterol pode ser obtido de duas maneiras distintas pelo organismo humano: a) O corpo humano, principalmente o fígado, produz quantidades da ordem de 1000 mg ao dia; b) através da ingestão de alimentos animais (ovos, carne, peixe, leite, etc.) e vegetais (frutas, grãos, cereais, sementes, etc.).

A Associação Americana do Coração⁽⁹⁷⁾ recomenda que a ingestão média diária de colesterol seja menor que 300 mg ; portanto, deve-se evitar a ingestão do excesso de alimentos ricos em gorduras saturadas, buscando sua substituição por vegetais.

Segundo a Associação Americana do Coração, as pessoas propensas a disfunções cardíacas coronárias possuem níveis de colesterol superior a 240 mg dL^{-1} (hipercolesterolemia); o risco moderado está associado a pessoas com níveis compreendidos entre 200 a 239 mg dL^{-1} , os valores recomendáveis para uma pessoa saudável deve ser inferior a 200 mg dL^{-1} ⁽⁹⁷⁾.

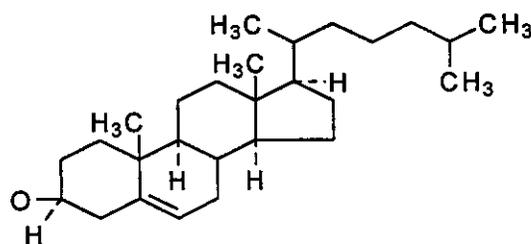


Figura 78. Fórmula estrutural química da molécula de colesterol.

9.7 Resultados e Discussão

O colesterol em plasma humano foi uma das três espécies químicas analisadas neste estudo. A análise quantitativa desta espécie foi realizada utilizando a faixa espectral entre 1480 nm - 1900 nm. Esta região foi selecionada após um exaustivo trabalho que compreende o emprego de duas faixas espectrais compreendidas entre as regiões sugeridas pela literatura⁽⁹⁶⁾ (2374 nm e 2212 nm) e (1480 nm e 1900 nm). O perfil dos espectros das amostras empregadas neste trabalho está ilustrado na Figura 79.

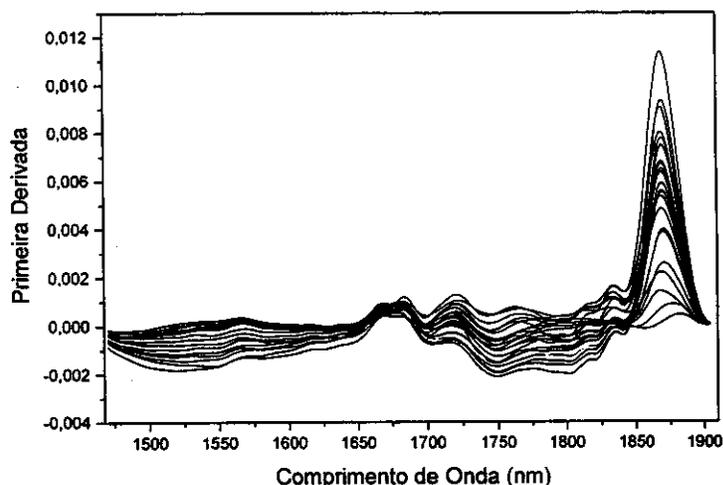


Figura 79. Perfil da primeira derivada dos espectros de plasma humano utilizado na determinação de colesterol, após o pré-processamento dos dados.

O estudo deste analito envolveu 237 amostras de plasma humano, que foram inicialmente separadas em dois conjuntos para o desenvolvimento do modelo com o método dos mínimos quadrados parciais (PLS). Para o grupo de calibração, foram utilizadas 207 amostras e as 30 amostras restantes foram usadas no grupo teste.

Para a técnica de regressão linear múltipla com seleção das variáveis, o grupo com 207 amostras (conjunto de calibração) foi dividido em dois subconjuntos com 130 e 77 amostras.

Foram utilizadas 130 amostras para o modelo de calibração e 77 amostras para a validação, durante o processo de seleção de variáveis, enquanto o conjunto com as 30 amostras restantes foi empregado para testar a robustez do modelo de calibração multivariada.

Os modelos que empregam técnicas de inteligência artificial (redes neural, neurogenética e Neuro-Fuzzy) utilizaram o mesmo conjunto de amostras que o modelo dos mínimos quadrados parciais (PLS).

A seleção de cada um dos grupos foi baseada no gráfico da primeira versus a segunda componente principal, conforme ilustrado na Figura 80.

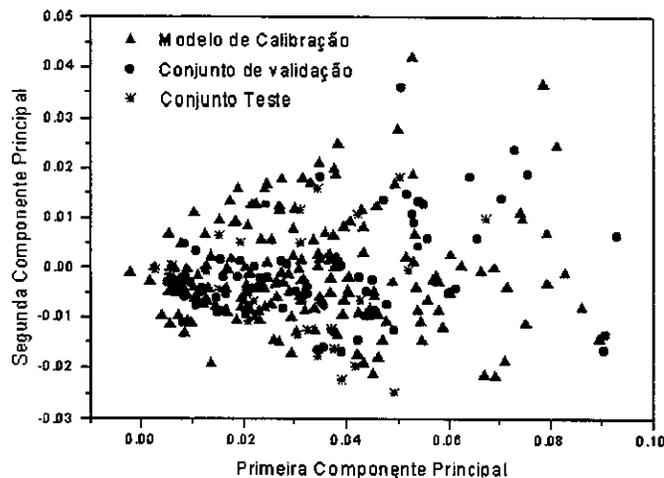


Figura 80. Gráfico dos escores da primeira versus a segunda componente principal.

A seleção das amostras foi realizada de maneira que cada grupo possuisse o conjunto mais representativo de amostras do espaço amostral estudado. Além disso, foi utilizada a influência para verificar o grau de influência de cada uma das amostras utilizadas no conjunto de calibração e validação interna.

9.7.1 Modelo dos Mínimos Quadrados Parciais (PLS)

O modelo desenvolvido a partir do PLS mostrou um bom desempenho de previsão para a concentração de colesterol em amostras de plasma humano. A Figura 81 apresenta os resultados previstos pelo modelo versus os valores fornecidos pelo método de referência.

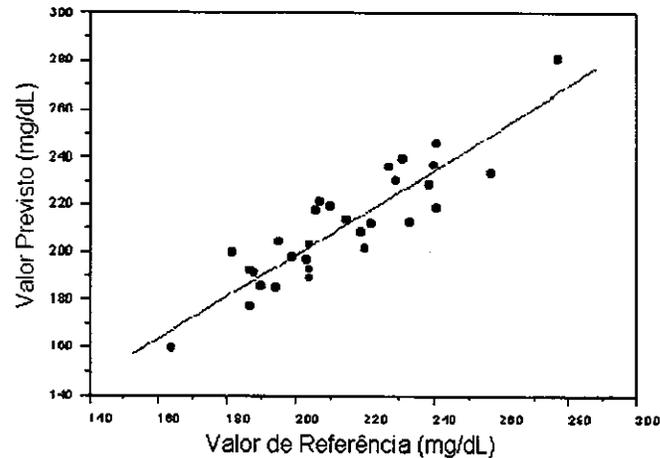


Figura 81. Valores do método referência versus os valores obtidos por NIR utilizando o PLS.

A partir do conjunto de amostras empregadas foi possível obter um valor satisfatório do coeficiente de correlação (0,893) da reta obtida entre os valores previstos e de referência. O erro padrão de previsão (SEP) foi de 14,12 mg dL⁻¹.

9.7.2 Seleção de Variáveis Aplicando Algoritmo Genético (AG/RLM)

A Figura 82 mostra o conjunto de variáveis selecionadas pelo algoritmo genético. Foram selecionados 3 comprimentos de onda (1728, 1684 e 1496 nm), os quais podem estar assim relacionados⁽⁸⁰⁾: 1728 nm ao primeiro sobretom do estiramento -CH das ligações -CH₂, 1684 nm ao primeiro sobretom do estiramento -CH das ligações -CH₃ e 1496 nm ao primeiro sobretom do estiramento -OH. O comprimento de onda em 1684 nm corresponde a um ponto de máximo de uma banda de absorção.

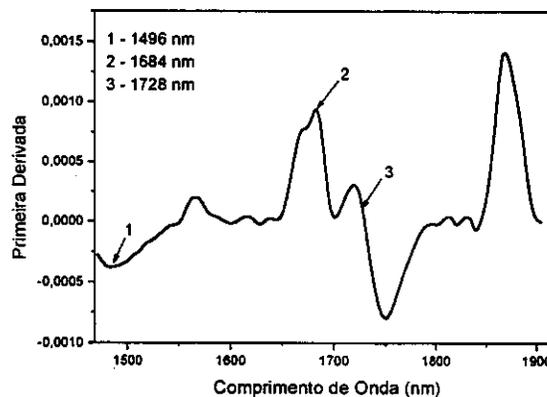


Figura 82. Primeira derivada do espectro de plasma humano e os comprimentos de onda selecionados pelo algoritmo genético.

Com o uso do algoritmo genético para a seleção de variáveis foi possível reduzir o número de variáveis para apenas três, permitindo desta maneira simplificar o modelo de calibração e estabelecer relações entre o modelo matemático e interpretações químicas. O procedimento de interpretação química é geralmente muito difícil de ser realizado com outros modelos de calibração multivariada.

A Figura 83 mostra os valores do método de referência versus os valores obtidos pelo NIR.

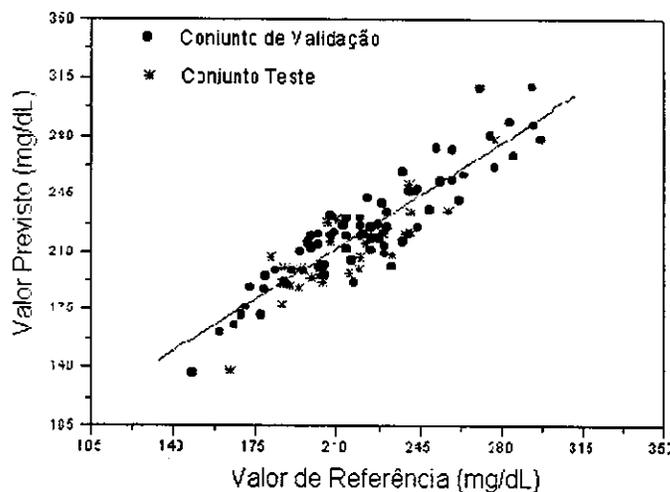


Figura 83. Valores do método referência versus os valores obtidos por NIR; o conjunto de validação e teste estão representados pelo símbolos (●) e (*), respectivamente.

Os resultados obtidos com o modelo de regressão linear múltipla são um pouco melhores que os alcançados com o modelo dos mínimos quadrados parciais. Isto pode ser constatado quando comparamos os valores do coeficiente de correlação (0,921) da reta obtida mostrada na Figura 83, SEP (13,97 mg dL⁻¹). Conforme pode ser observado, o modelo de regressão linear múltipla apresenta valores levemente superiores aos apresentados pelo modelo dos mínimos quadrados parciais.

9.7.3 Rede Neural

A rede neural mostrou um desempenho muito similar aos métodos lineares, conforme pode ser observado na Figura 84. A maior discrepância é verificada para a amostra com o menor valor de concentração (164 mg dL⁻¹). Esta amostra apresenta um resíduo significativamente superior quando comparada com o resíduo das demais amostras. O procedimento normal seria eliminar

esta amostra. Entretanto, optamos por mantê-la, pois a distância de Mahalanobis não mostra diferença significativa entre esta amostra e as demais amostras utilizadas neste estudo. Portanto, este comportamento pode ser um indício de que o modelo de calibração não tem o mesmo desempenho de previsão para toda a faixa de concentração, devido ao fato do qual não dispomos de um número suficiente de amostras para modelar concentrações inferiores a 175 mg dL^{-1} ou o limite de detecção do infravermelho próximo para este analito encontra-se nesta faixa.

O modelo de regressão linear múltipla acena com o mesmo comportamento, mas de maneira mais sutil, conforme pode ser verificado na Figura 83.

Em uma primeira análise, acreditamos que a dificuldade de modelagem esteja mais fortemente relacionada com a dificuldade da rede neural de modelar as regiões de fronteira. Entretanto, para termos certeza, teríamos que dispor de um número maior de amostras para realizar um trabalho mais minucioso deste sistema na faixa de concentração próxima a 164 mg dL^{-1} .

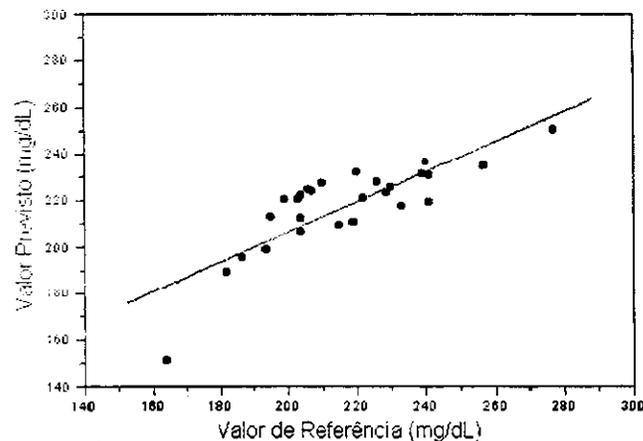


Figura 84. Valores do método referência versus os valores obtidos por NIR, utilizando a rede neural.

O modelo construído utilizando a técnica de redes neurais apresentou, para a reta ajustada entre o valor de referência e o do NIR um coeficiente de correlação de 0,839, sendo obtido um SEP de $14,52 \text{ mg dL}^{-1}$. A eliminação da amostra com maior valor de resíduo resultaria em uma melhora significativa principalmente dos valores do coeficiente de correlação e SEP, visto que estes dois parâmetros estão relacionados com o quadrado do resíduo.

9.7.4 Rede Neurogenética

A rede neurogenética forneceu-nos um modelo de calibração multivariada pouco eficiente na habilidade de prever a concentração de colesterol em amostras de plasma humano. Na Figura 85, observamos uma alta dispersão das amostras ao redor da curva ajustada. Este alto grau de dispersão reflete um baixo valor do coeficiente correlação (0,721) entre o valor de referência e o valor previsto. O baixo desempenho do modelo é também assinalado pelo valor do SEP (18,34 mg dL⁻¹).

O modelo mostrou o mesmo baixo nível de desempenho ao longo de toda a curva, indicando que talvez fosse necessário aumentar o número de iterações do processo de treinamento da rede. Esta conclusão é reforçada pelos resultados obtidos com a rede neural, já que esta apresentou resultados bem melhores. Teoricamente, esperávamos que o desempenho da rede neural e neurogenética fossem similares.

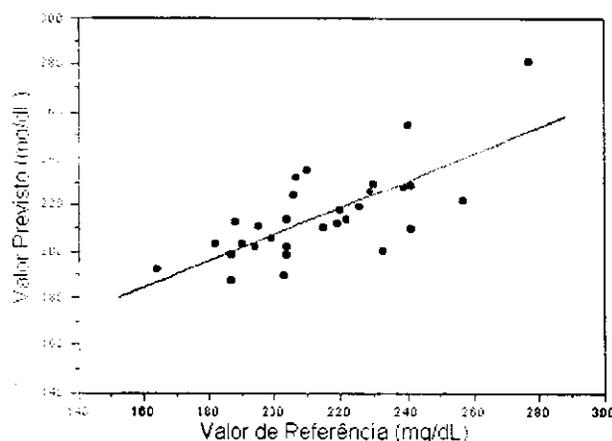


Figura 85. Valores do método referência versus os valores obtidos por NIR, utilizando a rede neurogenética.

9.7.5 Rede Neuro-Fuzzy

Com o uso desta técnica, foi possível desenvolver um modelo de calibração melhor em relação aos outros métodos não lineares, conforme apresentado na Figura 86. A Figura 87 mostra o erro relativo ao longo da faixa de concentração estudada.

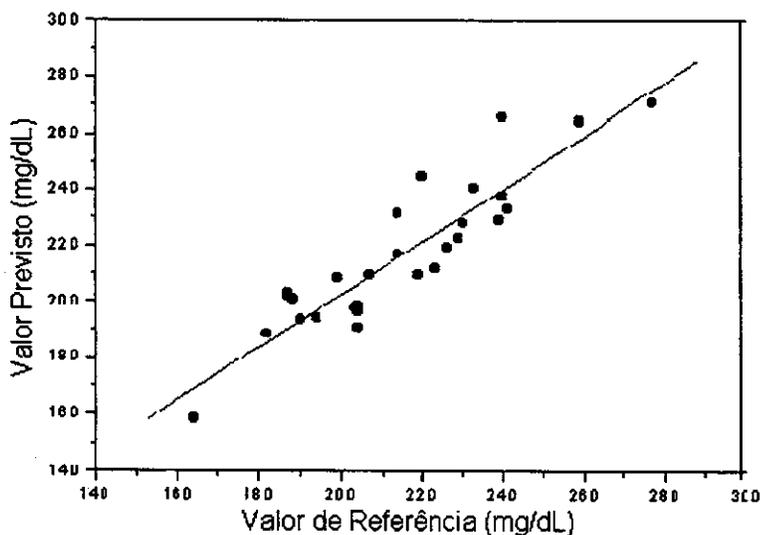


Figura 86. Valores do método referência versus os valores obtidos por NIR.

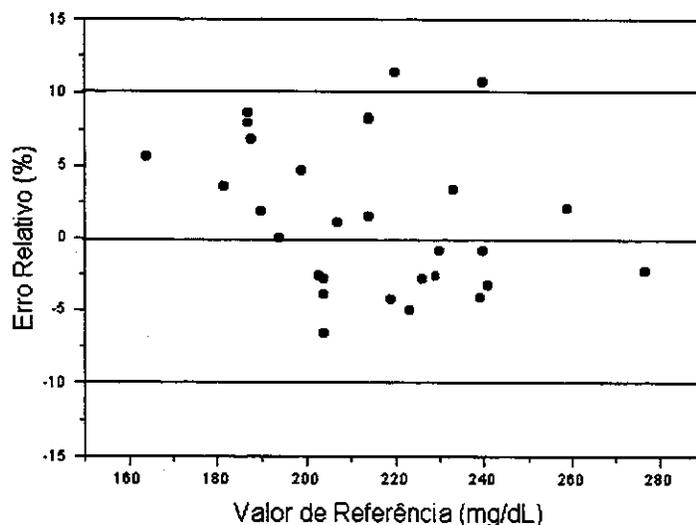


Figura 87. Erro relativo obtido na previsão da concentração de colesterol por NIR.

O modelo de rede Neuro-Fuzzy mostrou-se um pouco mais robusto, principalmente com relação às amostra com concentração próxima do limite inferior. Este comportamento pode indicar que o modelo da rede neural realmente apresenta o problema de modelagem de bordas. Uma maneira para contornar este comportamento seria a inserção de mais amostras com concentrações próximas ao limite inferior ou tentar realizar o processo de treinamento da rede neural um número maior de vezes.

9.8 Considerações Finais

Apesar do melhor desempenho do modelo Neuro-Fuzzy, os valores obtidos para o coeficiente de correlação (0,911) da reta apresentada na Figura 87, e SEP (13,39 mg dL⁻¹) não são significativamente melhores que os obtidos pelo PLS e RLM. Esta conclusão é baseada nos valores do teste F com 95% de confiança, conforme apresentado na Tabela 14.

Tabela 14. Valores de erro e do teste F obtidos para os 5 modelos de calibração multivariada na estimativa colesterol em plasma.

	PLS	RLM	Rede Neural	Neurogenética	Neuro-Fuzzy
SEC (mg dL ⁻¹)	13,20	14,04	10,37	17,58	12,47
SEP (mg dL ⁻¹)	14,12	13,97	14,52	18,34	13,39
Teste - F	1,11	1,09	1,18	1,87	-

F_{Crítico 95%} = 1,84

Os resultados da Tabela 14 mostram que apenas o modelo da rede neurogenética apresenta um desempenho significativamente inferior aos demais modelos de calibração. Portanto, partindo destes resultados, concluímos que o melhor modelo de calibração para ser empregado seria a regressão linear múltipla ou o PLS, visto que são as técnicas mais simples e difundidas, além de que atualmente já existem inúmeros softwares comerciais que apresentam estas ferramentas à disposição.

9.9 Determinação de Glicose em Plasma humano

O interesse pelo desenvolvimento de uma metodologia alternativa para a determinação da concentração de glicose em fluidos biológicos está associado ao elevado número de análises deste analito. Normalmente, em um laboratório de Análises Clínicas, mais de 50% das determinações realizadas mensalmente se refere a determinação da concentração de glicose. Este número expressa claramente a importância que esta análise possui quando comparada com outros analitos como: uréia, albumina, creatinina, triglicérides, etc. A fórmula estrutural química da glicose está apresentada na Figura 88.

Os métodos tradicionais de análise de glicose em plasma envolvem o uso de ensaios enzimáticos, seguidos pela detecção fotométrica ou eletroquímica. Apesar da elevada demanda do número de análises, os kits para os ensaios enzimáticos ainda possuem preços abusivos. Esta política de preços é decorrente do fato de não existir uma análise alternativa que concorra neste segmento do mercado. Portanto, o desenvolvimento de uma metodologia de análise aplicando medidas espectroscópicas pode proporcionar uma redução significativa dos gastos e conseqüentemente, uma diminuição nos custos das análises.

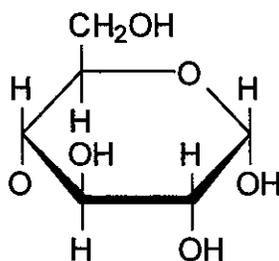


Figura 88. Fórmula estrutural química da molécula de Glicose.

9.10 Resultados e Discussão

A Figura 89 ilustra o perfil do conjunto de espectros utilizados após a etapa de pré-processamento dos dados e a faixa espectral selecionada (1485 nm e 1900 nm). A escolha desta região baseou-se em informações reportadas na literatura⁽⁹⁶⁻¹⁰¹⁾, seguida por uma etapa de otimização da melhor região espectral. A otimização da região espectral foi realizada por tentativa e erro, que consistiu em desenvolver vários modelos de calibração e observar qual apresentava um aumento significativo ou diminuição do erro quando comparado a faixa espectral proposta na literatura (2374 nm e 2111 nm) e (2000 nm e 1450 nm).

Segundo o que observamos, a região entre 2374 nm e 2111 nm não apresenta nenhuma contribuição significativa ao desempenho de previsão do modelo de calibração multivariada. Associamos este fato ao motivo de que o artigo que reporta esta região espectral trabalha com amostras de pacientes com diabetes. Nesse caso, a concentração de glicose é entre duas e três vezes superior à proposta neste trabalho. Heise e colaboradores⁽⁹⁶⁾ utilizaram amostras com concentrações entre 50 e 510 mg dL⁻¹ e concentração média de 208 mg dL⁻¹. Enquanto, isso neste trabalho opera com uma faixa de concentração inferior (entre 78 e 160 mg dL⁻¹ e concentração média de 104 mg dL⁻¹).

De acordo com as observações constatadas neste trabalho, a região entre 2374 nm e 2111 nm apresenta uma baixa relação sinal/ruído, o que dificulta a extração de qualquer informação que possa ser aplicada para modelar este sistema. Entretanto, acreditamos que no caso particular de Heise e colaboradores⁽⁹⁶⁾, esta região possa ter contribuído para modelar melhor o conjunto de dados analisado, visto que provavelmente o sinal relacionado à glicose possa ser significativamente mais intenso nesta região, para altas concentrações. Portanto, provavelmente a relação sinal/ruído deve ser maior para altas concentrações de glicose, o que permite extrair informações necessárias para a modelagem de amostras com altos valores de concentração de glicose.

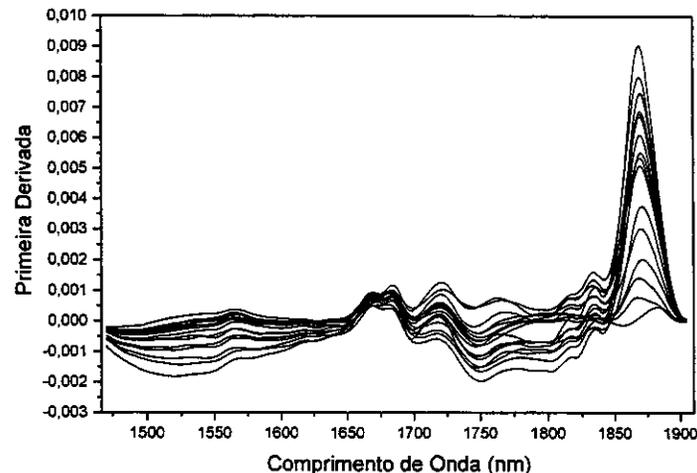


Figura 89. O perfil da primeira derivada dos espectros de plasma humano utilizado na determinação de glicose, após o pré-processamento dos dados.

9.10.1 Modelo dos Mínimos Quadrados Parciais (PLS)

O resultado obtido usando o modelo dos mínimos quadrados parciais (PLS) mostrou um ótimo desempenho para a estimativa da concentração de glicose. Este estudo envolveu 266 amostras, sendo 236 amostras destinadas para a construção do modelo de calibração multivariada e 30 amostras (conjunto teste) foram empregadas para avaliar a qualidade de previsão do modelo de calibração

multivariada. A seleção dos conjuntos de amostras foi realizada utilizando o gráfico da primeira versus a segunda componente principal, de maneira análoga à apresentada para a análise de triglicerídeos.

A Figura 90 mostra uma boa correlação (coeficiente de correlação de 0,921) entre a reta ajustada dos valores das concentrações previstas e os valores das concentrações consideradas como referência.

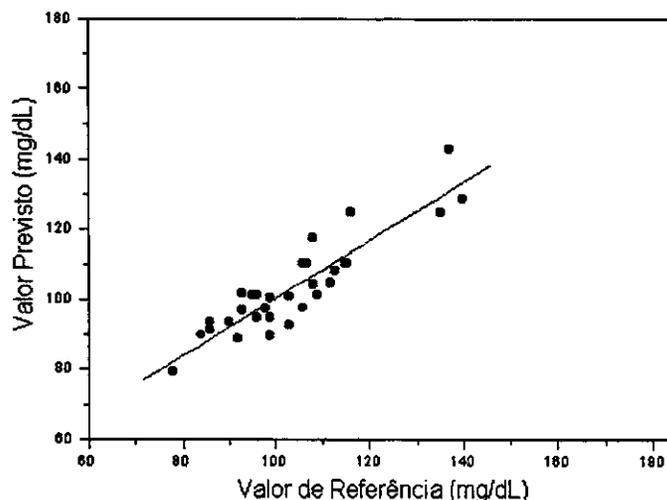


Figura 90. Valores do método referência versus os valores obtidos por NIR.

O modelo desenvolvido apresentou um erro padrão de previsão (SEP) de $6,54 \text{ mg dL}^{-1}$. Considerando o desvio do método de referência, podemos concluir que os resultados de previsão para este tipo de análise são muito bons. Além disso, observou-se que o modelo de calibração apresentou um erro relativo com distribuição aleatória e sempre menor que $\pm 10\%$ ao longo de toda a faixa de concentração estudada, conforme pode ser observado na Figura 91.

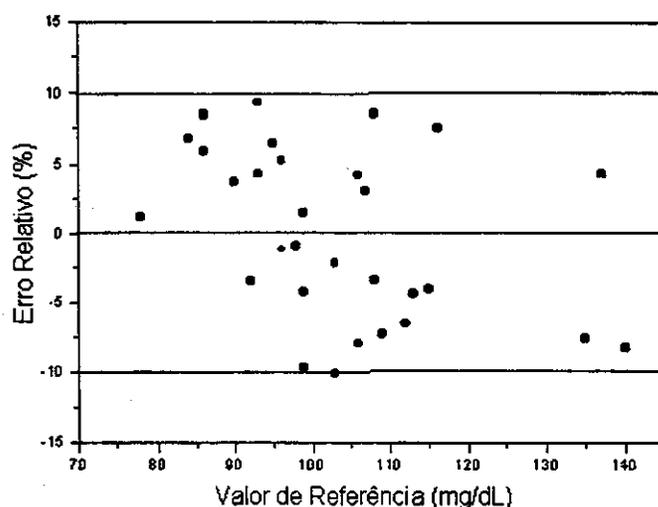


Figura 91. Erro relativo obtido na previsão da concentração de glicose, por NIR.

9.10.2 Seleção de Variáveis Aplicando Algoritmo Genético (AG/RLM)

O algoritmo genético foi empregado para tentar encontrar um subconjunto de variáveis que estabeleçam a melhor relação linear entre concentração e os dados espectrais, viabilizando a utilização de um modelo matemático mais simples como a regressão linear múltipla (RLM). Além disso, a implementação deste algoritmo é mais simples de ser realizado e possibilita interpretações químicas, o que normalmente é difícil de ser realizado em modelos matemáticos mais complexos. O desenvolvimento deste modelo de calibração envolveu 160 amostras para o modelo de calibração, 76 amostras para o conjunto de validação e 30 amostras para o conjunto teste.

A Figura 92 mostra o conjunto de variáveis selecionadas pelo algoritmo genético. Foram selecionados 3 comprimentos de onda (1872, 1672 e 1548 nm), os quais podem estar relacionados com⁽⁸⁰⁾: 1872 nm é uma banda de combinação entre o estiramento -OH e o estiramento -C-O, 1672 é um estiramento -CH do primeiro sobretom da grupo -CH₃ e 1548 é um estiramento -OH do primeiro sobretom de ligação intramolecular da ponte de hidrogênio.

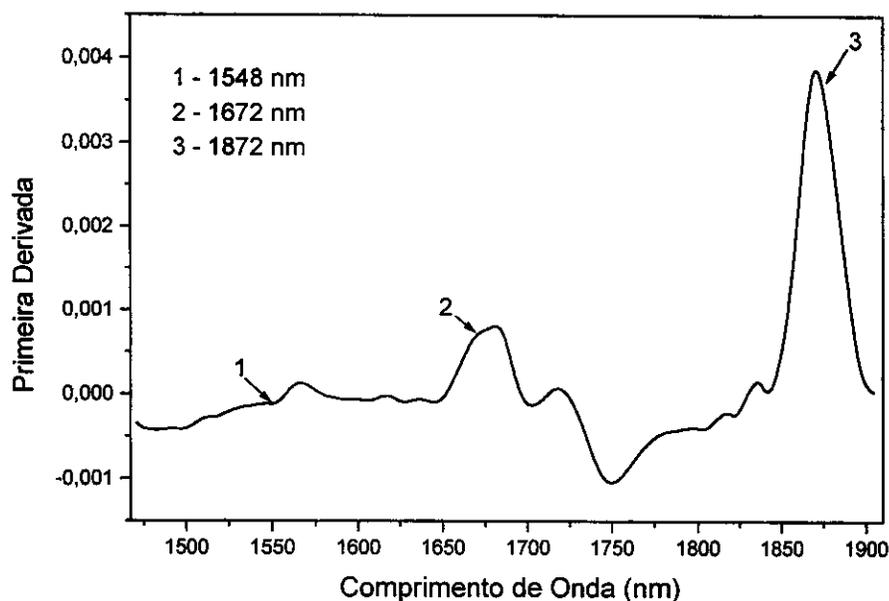


Figura 92. Primeira derivada do espectro de plasma humano e os comprimentos de onda selecionados pelo algoritmo genético.

A seleção de variáveis mostrou que, para este sistema, os comprimentos de onda próximos das bandas de absorção de maior intensidade são os mais importantes, indicando que a modelagem da glicose é simples de ser implementada com o auxílio do método de regressão linear múltipla.

Os resultados obtidos mostram um desempenho similar aos alcançados pelo método dos mínimos quadrados parciais. A reta ajustada entre os valores obtidos pelo método proposto e o método de referência apresentou um coeficiente de correlação de 0,915 para o conjunto teste e 0,894 para o conjunto de validação, SEP de 6,49 (teste) e 7,13 mg dL⁻¹ (validação).

A Figura 93 mostra os resultados da RLM/GA para os conjuntos de validação e teste. Conforme podemos verificar, os conjuntos teste e de validação apresentam desempenho similares. Entretanto, algumas amostras apresentaram um erro relativo maior que os observados sem a seleção de variáveis, indicando que provavelmente este não seja o melhor conjunto de variáveis que minimizam o erro de previsão. Uma opção para solucionar este problema seria a execução do programa um número maior de vezes.

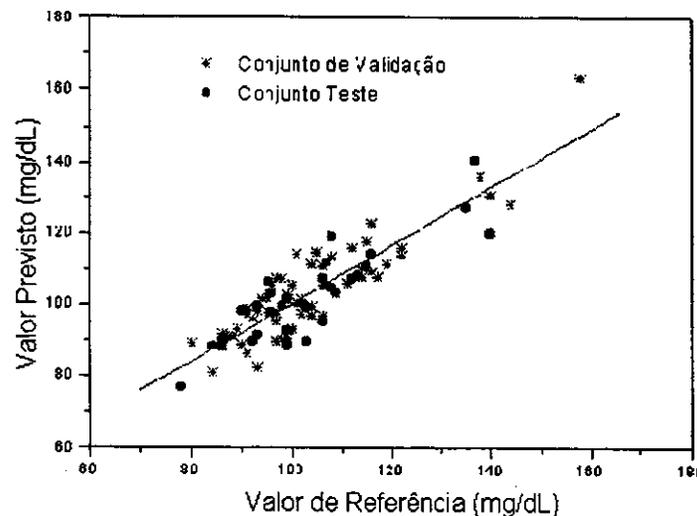


Figura 93. Valores do método referência versus os valores obtidos por NIR, para o conjunto teste (●) e validação (*).

9.10.3 Rede Neural

O modelo de rede neural foi empregado na tentativa de modelar alguma não linearidade do conjunto de dados permitindo, desta maneira, construir um modelo mais eficiente e proporcionar uma qualidade de previsão superior à observada pelos modelos lineares.

A Figura 94 mostra os resultados de previsão da rede neural versus os valores do método de referência. Os resultados obtidos pelo modelo de rede neural mostram aparentemente um desempenho similar aos métodos lineares, conforme pode ser verificado ao comparar os valores do coeficiente de

correlação (0,879) da reta obtida entre os valores previstos e os dados pelo método de referência, SEP (10,05 mg dL⁻¹).

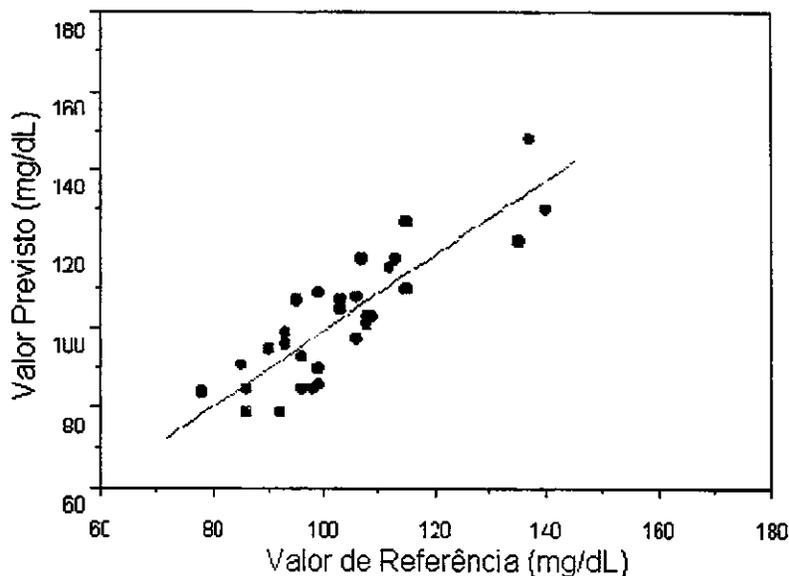


Figura 94. Valores do método de referência versus os valores obtidos por NIR.

Como já havíamos observado anteriormente para os outros modelos, os erros de previsão encontram-se dentro da faixa do desvio do método de referência. É importante ressaltar que não encontramos uma correlação direta entre o valor de concentração e o erro relativo, ou seja, aparentemente o erro relativo independe da concentração. Contudo, não podemos afirmar o mesmo com relação ao resíduo (diferença absoluta entre o valor previsto e o valor calculado com o método de referência). Observamos que conforme o valor de concentração aumenta, maior é o resíduo. Entretanto, este é um problema do método de referência, que não pode ser contornado utilizando os métodos de calibração multivariada.

Os resultados obtidos pelo modelo de rede neural indicam que os dados possuem um comportamento linear. Em sistemas lineares, o emprego de métodos não lineares geralmente não apresenta nenhum tipo de vantagem. Normalmente, os métodos não lineares demandam um tempo maior de processamento para obterem resultados similares aos alcançados pelos métodos lineares. Entretanto, em vários casos, mesmo um tempo maior de processamento pode resultar em um desempenho inferior, quando comparado aos obtidos pelos métodos lineares ou, em alguns casos, na obtenção de um modelo de calibração multivariada pouco robusto.

9.10.4 Rede Neurogenética

O modelo de rede neurogenético resultou em um modelo aparentemente menos eficiente que o desenvolvido para os métodos anteriores. A Figura 95 mostra o valor previsto pelo modelo de calibração multivariada versus o valor obtido pelo método de referência. Para este sistema a reta ajustada tem um coeficiente de correlação de 0,769 e SEP de 15,54 mg dL⁻¹.

No caso particular do modelo de rede neurogenética, é mais evidente a diferença entre os modelos lineares e não lineares, devido principalmente ao fato do treinamento ser mais lento que o da rede neural. Certamente, quando o tempo de treinamento tende ao infinito, a rede neurogenética e rede neural irão convergir para o modelo linear.

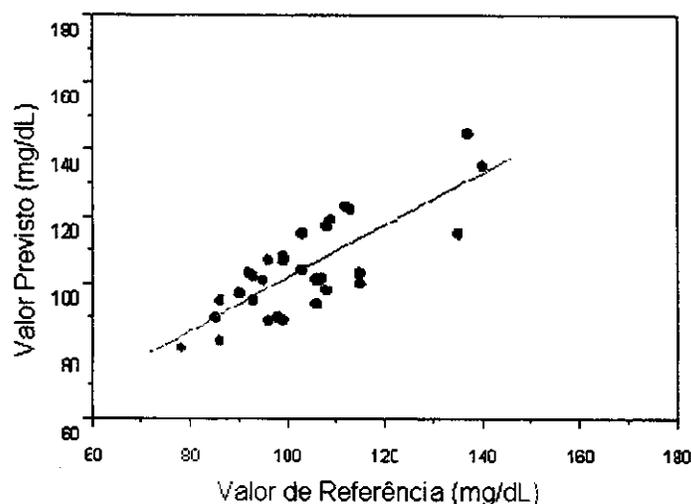


Figura 95. Valores do método de referência versus os valores obtidos por NIR.

9.10.5 Rede Neuro-Fuzzy

O modelo de rede Neuro-Fuzzy apresentou um tempo de processamento e habilidade de previsão similar ao da rede neural, sendo obtido um SEP de 11,45 mg dL⁻¹. A Figura 96 apresenta os valores de referência versus valores previstos, cuja reta ajustada tem um coeficiente de correlação de 0,865.

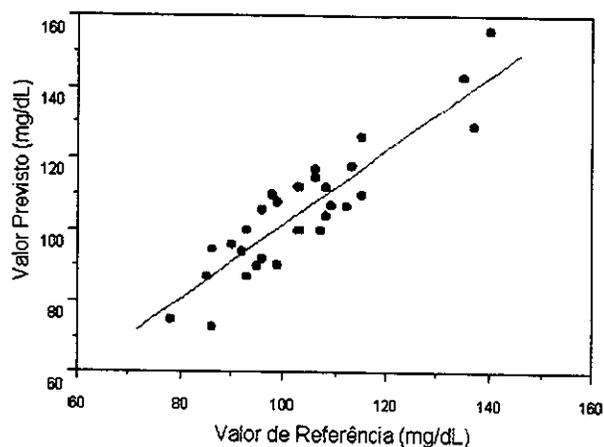


Figura 96. Valores do método referência versus os valores obtidos por NIR.

9.11 Considerações Finais

Neste trabalho foi empregado o teste F, para comparar o desempenho dos modelos de calibração multivariada e avaliar se existe uma diferença significativa entre eles. A Tabela 15 apresenta os resultados dos valores de SEC e SEP para os diversos modelos de calibração empregados.

O teste F com 95 % de confiança foi realizado utilizando o valor do SEP do modelo com o menor erro de previsão (Mínimos Quadrados Parciais) como modelo de referência. A tabela do teste F com 95 % de confiança, para dois conjuntos com 30 amostras, o valor crítico é de 1,84. Portanto, segundo os valores apresentados na Tabela 16, concluímos que os modelos lineares são similares, enquanto os modelos não lineares apresentam um desempenho inferior. Contudo, os modelos de rede neural e Neuro-Fuzzy poderiam também ser aceitos se considerarmos que o método padrão apresenta um desvio ao redor de $\pm 10\%$. Entretanto, como não observamos nenhuma vantagem na utilização dos modelos de inteligência artificial perante os modelos tradicionais, estamos fortemente inclinados a rejeitar o uso destas técnicas de inteligência artificial para a determinação de glicose em plasma humano.

Tabela 15. Valores de erro e do Teste F obtidos para os 5 modelos de calibração multivariada na estimativa de glicose em plasma.

	PLS	RLM	Rede Neural	Neurogenética	Neuro-Fuzzy
SEC (mg dL ⁻¹)	7,32	7,85	7,05	13,42	8,23
SEP (mg dL ⁻¹)	6,54	7,13	10,05	15,54	11,45
Teste-F	-	1,18	2,36	5,64	3,07

F_{Crítico 95%} = 1,84.

Conclusões Gerais

As conclusões desta tese podem ser abordadas de duas maneiras : 1- avaliação das técnicas de inteligência artificial na modelagem de dados químicos e simulados ; 2- Aplicação da espectroscopia no infravermelho próximo para o controle de qualidade em amostras biológicas, um polímero e cana de açúcar.

1- Avaliação das técnicas de inteligência artificial na modelagem de dados químicos e simulados.

Foi observado que os modelos de inteligência artificial (redes neural, neurogenética e Neuro-Fuzzy) quando aplicados a dados com comportamento não linear, agregam um benefício significativo à modelagem do sistema, sendo, portanto, uma ferramenta que deve ser avaliada quando modelos lineares (PLS e RLM) apresentam resultados pouco satisfatórios na modelagem do sistema de interesse.

As redes neural e Neuro-Fuzzy mostraram ser capazes de modelar sistemas com comportamento linear tão bem quanto os modelos lineares. Entretanto, devido ao grau de complexidade associado a estas técnicas, não é interessante seu uso em sistemas onde modelos lineares apresentam um desempenho satisfatório. Entretanto, se o sistema apresentar um real comportamento não linear como o apresentado nos capítulos 6 e 8, certamente o uso destas ferramentas ofereceriam resultados significativamente melhores que justificam a sua aplicação.

O sistema simulado apresentado no capítulo 6 desta Tese agregou informações significativamente relevantes para a compreensão dos comportamentos das redes neurais e rede Neuro-Fuzzy. Observamos que a rede neural treinada apresenta o que se chama de problema de borda, devido à insuficiência de informações na região de fronteira, apresentando um baixo desempenho na capacidade de previsão nestas regiões. Além disso, esta aplicação deixa bem claro que não é possível realizar extrapolações a partir dos modelos desenvolvidos. Em contrapartida, a capacidade da rede Neuro-Fuzzy de manipular com a imprecisão dos dados permite que o modelo gerado seja capaz de modelar todo o espaço amostral com a mesma habilidade. Mesmo as regiões de bordas onde existem poucas informações são bem modeladas. Portanto, acreditamos que talvez seja possível aplicar o modelo de rede Neuro-Fuzzy para fazer pequenas extrapolações, estimando propriedades que estejam fora da região de mapeamento das amostras de calibração.

2- Aplicação da técnica de infravermelho no controle de qualidade em amostras biológicas, de polímero e de cana de açúcar.

Outro aspecto importante, avaliado neste trabalho foi a possibilidade de aplicar a técnica de infravermelho como uma ferramenta analítica no controle de qualidade de rotina. Neste trabalho foram estudados três sistemas, onde, em dois deles, foram utilizadas medidas de absorbância e num deles medidas de reflectância difusa.

Os estudos com as amostras de caldo de cana de açúcar mostram que tanto o grau Brix como o grau Pol podem ser estimados através do uso da espectroscopia no infravermelho próximo. Os resultados obtidos mostram uma boa concordância entre o método de referência e o valor estimado. Portanto, a espectroscopia na região do infravermelho próximo permite que seja realizado o controle "in-line", "on-line" ou "at-line" deste produto, aumentando, consideravelmente a eficiência do controle de qualidade da indústria sucro-alcooleira.

O estudo de quantificação do mineral adicionado ao polímero mostrou resultados acima de quaisquer expectativas, mostrando a viabilidade do uso da técnica de espectroscopia no infravermelho próximo como uma nova ferramenta de análise para este sistema. Além disso, esta técnica permite que seja realizado um estudo em tempo real, bem como uma análise mais rigorosa da homogeneidade do material produzido. Por possuir características não lineares, os dados deste sistema foram úteis para destacar a importância dos modelos de inteligência artificial no tratamento de sistemas químicos específicos. Observou-se que, apesar dos modelos lineares possuírem um bom desempenho relativo, são inferiores aos de inteligência artificial. Os modelos de inteligência artificial, foi possível estimar a concentração do mineral a partir dos espectros de infravermelho próximo com a mesma precisão oferecida pela técnica de absorção atômica. Isso mostra que a técnica de espectroscopia no infravermelho próximo pode fornecer resultados com alta sensibilidade e precisão, quando associado ao pré-processamento dos dados e a um modelo de calibração multivariada adequado.

Os resultados alcançados na modelagem das amostras de origem biológica foram muito positivos. Este estudo indica claramente a possibilidade de determinar a concentração de glicose, triglicerídeos e colesterol em amostras de plasma humano, com a ordem de grandeza do erro obtido pelo método padrão de análise. Portanto, isso demonstra que é possível o uso da espectroscopia na região do infravermelho próximo como um método alternativo que permitirá a redução substancial do custo e tempo requerido para cada análise. Isto também pode permitir a análise simultânea de mais de um analito por espectro.

Esta Tese abre perspectivas para trabalhos futuros em amostras biológicas, com o objetivo principal de analisar outras espécies químicas importantes como : creatinina, ácido úrico, uréia, esteróides anabolizantes, proteína total, hormônios, etc.

Atualmente, vem se observando uma tendência na aplicação da técnica de reflectância difusa na região do infravermelho próximo, com o objetivo de analisar analitos de interesse biológico através de medidas diretas sobre o corpo humano. Um desenvolvimento neste segmento seria um avanço significativo para este tipo de análise.

Com relação às técnicas quimiométricas, existe a possibilidade de melhorar o desempenho do programa de rede neurogenética e desenvolver um programa híbrido Neuro-Fuzzy-Genético. Entretanto, acreditamos que, antes de desenvolver-lo é necessário provar que o programa do algoritmo genético realmente agregue vantagens reais ao processo de treinamento. Esse estudo indicou que os sistemas de treinamento com gradiente oferecem uma taxa de convergência mais rápida e resultados de alta qualidade. Em contrapartida, a rede treinada com o algoritmo genético mostrou desempenho inferior ou similar quando comparada com as técnicas derivativas de otimização.

Além disso, seria interessante realizar novos estudos comparativos entre as redes neural e Neuro-Fuzzy com o objetivo principal de avaliar o desempenho de cada uma das técnicas. Desta maneira, poderíamos ter mais subsídios para afirmar a superioridade da rede Neuro-Fuzzy na habilidade de modelar sistemas não lineares.

- [1] Sharaf, M. A., Illman, D. L., Kowalski, B. R. , *Chemometrics* , John Wiley & Sons , New York, (1986).
- [2] Massart, D. L., Vandeginste, B. G., Deming, S. N., Michotte, Y., Kaufman, L., *Chemometrics: a textbook*, Elsevier, New York, (1986).
- [3] Morgan, E. , *Chemometrics Experimental Design*, John Wiley & Sons, England, (1991).
- [4] Workman, J. J. JR. , Kowalski, B. R., Bro, R., “Review of chemometrics applied to spectroscopy; 1985 – 95, Part I“, *Appl. Spectrosc. Rev.*, **31**, 73, (1996).
- [5] Vandegiste, B. G. M., Massart, D. L., Buydens, L. M. C., Jong, S. de, Lewi, P. J., Smeyers-Verbeke, J., *Handbook of Chemometrics and Qualimetrics: Part B*, Elsevier, New York, (1998).
- [6] Otto, M., *Chemometrics: Statistics and Computer Application in Analytical Chemistry*, Wiley-VCH, New York, (1999).
- [7] Pal, N. R., “Soft computing for feature analysis”, *Fuzzy sets and Syst.*, **103**, 201, (1999).
- [8] Ni, Y., Liu, C., “Artificial neural networks and multivariate calibration for spectrophotometric differential kinetic determinations of food antioxidants”, *Anal. Chim. Acta*, **396**, 221, (1999).
- [9] Ding, Q., Small, G. W., Arnold, M. A., “ Evaluation of nonlinear model building strategies for the determination of glucose in biological matrices by near-infrared spectroscopy”, *Anal. Chim. Acta*, **384**, 333, (1999).
- [10] Dupuy, N., et. al., “Quantitative determination of polymer and mineral content in paper coatings by infrared spectroscopy. Improvements by non-linear treatments”, *Anal. Chim. Acta*, **335**, 79, (1996).

- [11] Huth-Fehre, Th., et. al., "NIR - Remote sensing and artificial neural networks for rapid identification of post consumer plastics", *J. of Mol. Struct.*, **348**, 143, (1995).
- [12] N. Dupuy, C. Ruckebush, L. Duponchel, P. Beurdeley-Saudou, B. Amram, J. P. Huvenne and P. Legrand, "Quantitative determination of polymer and mineral content in paper coatings by infrared spectroscopy. Improvements by non-linear treatments", *Anal. Chim. Acta*, **335**, 79, (1996).
- [13] Eustaquio, A., Graham, P., et. all. , "Quantification of paracetamol in intact tablets using near-infrared transmittance spectroscopy", *Analyst*, **123**, 2303, (1998).
- [14] Dubko, A., Dovletoglou, A., "Moisture assay of an antifungal by near-infrared diffuse reflectance spectroscopy". *J. Pharm. Biomed. Anal.*, **28**, 145, (2002).
- [15] Zhou, X., Hines, P., Borer, M. W., "Moisture determination in hygroscopy drug substance by near-infrared spectroscopy", *J. Pharm. Biomed. Anal.*, **17**, 219, (1998).
- [16] Candolfi, A., De Maesschalck, R., et. all., "The influence of data pre-processing in the pattern recognition of excipients near-infrared spectra", *J. Pharm. Biomed. Anal.*, **21**, 115, (1999).
- [17] Berntsson, O., Zackrisson G, Östling, G., "Determination of moisture in hard gelatin capsules using near-infrared spectroscopy: applications to at-line process control of pharmaceuticals" *J. Pharm. Biomed. Anal.*, **15**, 895, (1997).
- [18] Fischbacher, Ch., Jagemann, K. U., Danzer, K., "Enhancing calibration model for non-invasive near-infrared spectroscopic blood glucose determination", *Fresenius J. Anal. Chem.*, **359**, 78, (1997).
- [19] Shaw, R. A., Mantsch, H. H., *Encyclopedia of Analytical Chemistry*, John Wiley & Sons Ltd., Chichester, (2000).

- [20] Callis, J., Illman, D., Kowalski, B., "Process analytical chemistry ", *Anal. Chem.*, **59**, 624A, (1987).
- [21] Blaser, W., Bredeweg, R., Harner, R., Lapack, M., Leugers, A., Martin, D., Pell, R., Workman, J., Wright, L., "Process analytical chemistry", *Anal. Chem.*, **67**, 47A, (1995).
- [22] Draper, N. , Smith, H. , *Applied Regression Analysis* , John Wiley & Sons , New York, (1981).
- [23] Martens, H., Naes,T., *Multivariate Calibration*, John Wiley & Sons , New York, (1989).
- [24] De Maesschalck, R., Jouan-Rimbaud, D., Massart, D. L., "Tutorial the Mahalanobis distance", *Chemom. Intell. Lab. Syst.*, **50**, 1, (2000).
- [25] Mark, H. L, Tunnell, D., "Qualitative near-infrared reflectance analysis using Mahalanobis distances", *Anal. Chem.*, **57**, 1449, (1985).
- [26] Geladi, P., Kowalski, B. R., "Partial Least-Squares Regression: A Tutorial", *Anal. Chim. Acta*, **185**, 1, (1986).
- [27] Beebe, R.K, Kowalski, B.R, "An Introduction to Multivariate Calibration and Analysis", *Anal. Chem.*, **59**, 1007, (1987).
- [28] Wu, W., Guo, Q., Massart, D. L., "The robust normal variate transform for pattern recognition with near-infrared data", *Anal. Chim. Acta*, **382**, 87, (1999).
- [29] Bertran, E., Blanco, M. C., Maspoch, S., Ortiz, M. C., Sánchez, M. S., Sarabia,L. A. "Handling intrinsic non-linearity in near-infrared reflectance spectroscopy", *Chemom. Intell. Lab. Syst.*, **49**, 215, (1999).

- [30] Ruse, M., *Darwinism Defended. A Guide to the Evolution Controversies*, John Wiley & Sons , New York, (1982).
- [31] Susmita De, S. K. P., Ghosh, A. "Genotypic and phenotypic assortative mating in genetic algorithm", *J. Inf. Sciences*, **105**, 209, (1998).
- [32] Goldberg, D.E; *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley , Reading, (1989).
- [33] Lucasius, C. B., Kateman, G. , "Understanting and Using Genetic Algorithms Part I. Concepts, Properties and Context", *Chemom. Intell. Lab. Syst. , 19 , 1 , (1993).*
- [34] Lucasius, C. B. and Kateman, G., "Genetic algorithms for large scale optimization in chemometrics: an application", *Trends Anal. Chem.*, **10**, 254, (1991).
- [35] Holland, J. H., *Genetic Algorithms, Sci. Am. , 44, (1992).*
- [36] Li, T. H., Lucasius, C. B., Kateman, G., "Optimization of calibration data with the dynamic genetic algorithm", *Anal. Chim. Acta , 268, 123, (1992).*
- [37] Jang, J. S. R., Sun, C. T., Mizutani, E., *Neuro-Fuzzy and Soft Computing* , Prentice Hall, Upper Saddle River, (1997).
- [38] Zupan, J. , Gasteiger, J. , *Neural Networks for Chemists*, VCH, Weinheim, (1993).
- [39] Bos, M., Weber, H. T. , "Comparison of training of neural networks for quantitative X-ray fluorescence spectrometry by genetic algorithm and backward error propagation", *Anal. Chim. Acta, 247, 97, (1991).*
- [40] Hartnett, M. K., Bos, M., van-der-Linden, W. E., Diamond, D., "Determination of stability constants using genetic algorithms", *Anal. Chim. Acta, 316, 347, (1995).*

- [41] Jouan-Rimaud, D., Massart, D. L., Leardi, R., De-Noord, O. E., "Genetic algorithms as a tools for wavelength selection in multivariate calibration", *Anal. Chem.*, **67**, 4295, (1995).
- [42] Fontain, E., "The problem of atom-to-atom mapping. An application of Genetic Algorithms", *Anal. Chim. Acta*, **265**, 227, (1992).
- [43] Broudiscou, A., Leardi, R., Phan-Tan-Luu, R., "Genetic algorithm as a tool for selection of D-optimal design", *Chemom. Intell. Lab. Syst.*, **35**, 105, (1996).
- [44] Wienke, D., Lucasius, C., Ehrlich, M., Kateman, G., "Multi-criteria target vector optimization of analytical procedures using a genetic algorithm. II. of the photometric calibration graph of dry glucose sensors for quantitative clinical analysis", *Anal. Chim. Acta*, **271**, 253, (1993).
- [45] Leardi, R., "Application of a Genetic Algorithm to feature selections under full validation conditions and to outlier detection", *J. Chemom.*, **8**, 65, (1994).
- [46] Walczak, B., "Outlier detection in multivariate calibration", *Chemom. Intell. Lab. Syst.*, **28**, 259, (1995).
- [47] Lopes-Marques, R. M., Schoenmakers, P. J., Lucasius, C. B., Buyden, L., "Modelling chromatographic behaviour as a function of pH and solvent composition in reversed phase LC", *Chromatographia*, **36**, 83, (1993).
- [48] Hibbert, D. B., "Hybrid genetic algorithm for the estimation of kinetic parameters", *Chemom. Intell. Lab. Syst.*, **19**, 319, (1993).
- [49] Wienke, D., Lucasius, C., Kateman, G., "Multi-criteria target vector optimization of analytical procedures using a genetic algorithm. I. Theory, numerical simulations and application to atomic-emission spectroscopy", *Anal. Chim. Acta*, **265**, 211, (1992).

- [50] Blommers, M. J. J., Lucasius, C. B., Kateman, G., Kaptein, R., "Conformational analysis of a dinucleotide photodimer with the aid of the genetic algorithm", *Biopolymers*, **32**, 45, (1992).
- [51] Haykin, S., *Neural networks: a comprehensive foundation*, Prentice Hall, New York, (1999).
- [52] Vlasov, Y. G., Legin, A. V., Rudnistkaya, A. M., D'Amico, A., Di Natale, C., "Chemical analysis of multicomponent aqueous solutions using a system of nonselective sensors and artificial neural networks", *J. Anal. Chem.*, **52**, 1087, (1997).
- [53] Mello, C. A., "Redes Neurais e técnicas de pré-processamento de sinais em espectroscopia de reflectância no infravermelho próximo", Tese de Doutorado, IQ/ UNICAMP, Campinas, (1998).
- [54] Tian, Y., Zhang, J., Morris, J., "Optimal control of a batch emulsion copolymerisation reactor based on recurrent neural network models", *Chem. Eng. and Proc.*, **41**, 531, (2002).
- [55] Blanco, M., Coello, J., Iturriaga, H., MasPOCH, S., Porcel, M., "Simultaneous enzymatic spectrophotometric determination of ethanol and methanol by use of artificial neural networks for calibration", *Anal. Chim. Acta*, **398**, 83, (1999).
- [56] Goodacre, R., Neal, M. J., Kell, D. B., "Rapid and quantitative analysis of the pyrolysis mass spectra of complex binary and tertiary mixtures using multivariate calibration and artificial neural networks", *Anal. Chem.*, **66**, 1070, (1994).
- [57] Gemperline, P. J., JungHwan, C., Baker, B., Batchelor, B., Walker, D. S., "Determination of multicomponent dissolution profiles of pharmaceutical products by *in situ* fiber-optic UV measurements", *Anal. Chim. Acta*, **345**, 155, (1997).
- [58] Chan, H., Butler, A., Falck, D. M., Freund, M. S., "Artificial neural network processing of stripping analysis responses for identifying and quantifying heavy metals in the presence of intermetallic compound formation", *Anal. Chem.*, **69**, 2373, (1997).

- [59] Bos, A., Bos, M., van der Linden, W. E., "Artificial neural networks as a multivariate calibration tool: modeling the Fe/Cr/Ni system in X-ray fluorescence spectroscopy", *Anal. Chim. Acta*, **277**, 289, (1993).
- [60] van Leeuwen, J. A., Jonker, R. J. and Gill, R., "Octane number prediction based on gas chromatographic analysis with non-linear regression techniques", *Chemom. Intell. Lab. Syst.*, **25**, 325, (1994).
- [61] Bro, R., "Multiway Calibration. Multilinear PLS", *J. Chemom.*, **10**, 47, (1996).
- [62] Li, S., Hamilton, C., Gemperline, P. J., "Generalized rank annihilation method using similarity transformations", *Anal. Chem.*, **64**, 599, (1992).
- [63] Despagne, F., Massart, D. L., "Neural networks in multivariate calibration", *Analyst*, **123**, 157R, (1998).
- [64] Sexton, R. S., et. al., "Optimization of neural networks: A comparative analysis of the genetic algorithm and simulated annealing", *European J. of Operational Research*, **144**, 589, (1999).
- [65] Otto, M., "Fuzzy Sets: Applications to Analytical Chemistry", *Anal. Chem.*, **62**, 797A, (1990).
- [66] Otto, M., "Fuzzy theory: A promising tool for computerized chemistry", *Anal. Chim. Acta*, **235**, 169, (1990).
- [67] Shen, J., Shen, W., et. al., "Fuzzy neural nets with non-symmetric membership functions and applications in signal processing and image analysis functions", *Fuzzy sets and Syst.*, **80**, 965, (2000).
- [68] Otto, M., Bandemer, H., "Pattern recognition based on fuzzy observations for spectroscopic quality control and chromatographic fingerprinting", *Anal. Chim. Acta.*, **184**, 21, (1986).

- [69] Thuillard, M., "The development of algorithms for a smoke detector with neuro-fuzzy logic", *Fuzzy sets and Syst.*, **77**, 117, (1996).
- [70] Yea, B., et. al., "The concentration estimation of inflammable gases with a semiconductor gas sensor utilizing neural networks and fuzzy inference", *Sensors and Actuators.*, **B 41**, 121, (1997).
- [71] Steyer, J-P., et. al. , "Hybrid fuzzy neural network for diagnosis – application to the anaerobic treatment of wine distillery wasterwater in a fluized bed reactor", *Wat. Sci. Tech.*, **36**, 209, (1997).
- [72] Eerikäinen, T., Linko, P., Linko, S., Siimes, T. and Zhu, Y-H., "Fuzzy logic and neural network applications in food science and technology", *Trends in Food Science & Technology*, **4**, 237, (1993).
- [73] Otto, M., Bandemer, H., "Calibration with imprecise signals and concentrations based on fuzzy theory", *Chemom. Intell. Lab. Syst.*, **1**, 71, (1986).
- [74] Akay, M., Cohen, M., Hudson, D., "Fuzzy sets in life sciences", *Fuzzy sets and Syst.*, **90**, 219, (1997).
- [75] Walczak, B., Bauer-Wolf, W., Wegscheider, W., "A Neuro-fuzzy system for X-ray spectra interpretation", *Mikrochim. Acta*, **113**, 153, (1994).
- [76] Harris, C. J., et. al., "Neurofuzzy state estimators and their applications", *Annual Reviews in control*, **23**, 149, (1999).
- [77] Elling, J. W., Lahiri, S., Luck, J. P., et. all., " Hybrid artificial intelligence tools for assessing GC data", *Anal. Chem. New & Features*, **1**, 409A, (1997).
- [78] Yager, R. R., Filev, D. P., "Generation of Fuzzy rules by mountain clustering", *J. Intell. And Fuzzy Syst.*, **2**, 209, (1994).

- [79] Burns, D. A., Ciurczak, E. W., *Handbook of Near-Infrared Analysis*, Marcel Dekker, New York, (2001).
- [80] Workman JR., J. J., "Interpretative spectroscopy for near infrared", *Appl. Spectrosc. Rev.*, **31**, 251, (1996).
- [81] Willians, P., *Near-Infrared Technology in The Agricultural and Food Industries*, American Association of Cereal Chemists, St. Paul, Minnesota, USA, (1990).
- [82] Kauzmann, W., *Quantum Chemistry*, Academic Press, New York, (1957).
- [83] Barton, F. E., "Theory and principles of near infrared spectroscopy", *Spectrosc. Europe*, **3**,12, (2002).
- [84] Villarroya, I., Blanco, M., "NIR spectroscopy: a rapid-response analytical tool", *Trends in Anal. Chem.*, **21**, 240, (2002).
- [85] Zago, E. A., Silva, L. F., Bernardino, C. D., Amorim, H. V., *Métodos Analíticos para o Controle de Produção de Alcool e Açúcar*, Fermentec/ FEALQ/ ESALQ-USP, Piracicaba, (1996).
- [86] Rambla, F. J., Garrigues, S., de la Guardia, M., "PLS-NIR determination of total glucose, fructose, and sucrose in aqueous solution of fruit juices", *Anal. Chem. Acta*, **344**, 41, (1997).
- [87] Rodriguez-Saona, L. E., et al "Rapid analysis of sugars in fruit juices by FT-NIR spectroscopy", *Carbohydrate Research*, **336**, 63, (2001).
- [88] Rantanen, J., Lehtola, S., Rämetsä, P., Mannermaa, Jukka-Pekka, Yliruusi, J., "On-line monitoring of moisture content in an instrumented fluidized bed granulator with a multi-channel NIR moisture sensor", *Powder Technol.*, **99**, 163, (1998).

- [89] Soon, M. H., Pat, G. F., "Determination of SB 216469-S during tablet production using near-infrared reflectance spectroscopy" *J. Pharm. Biomed. Anal.*, **14**, 1681, (1996).
- [90] Jedvert, I., Josefson, M., Langkilde, F. "Quantification of active substance in a tablet by NIR and Raman spectroscopy" *J. Near Infrared Spectrosc.*, **6**, 279, (1998).
- [91] Sekulic, S. S., Walkeman, J., Doherty, P., Hailey, P. A., "Automated system for the on-line monitoring of powder blending process using near-infrared spectroscopy: Part II. Qualitative approaches to blend evaluation", *J. Pharm. Biomed. Anal.*, **17**, 1285, (1998).
- [92] Berntsson, O., Danielsson, L-G, Folestad, S., "Estimation of effective sample size when analysing powders with diffuse reflectance near-infrared spectrometry", *Anal. Chem. Acta*, **364**, 243, (1998).
- [93] Berntsson, O., Danielsson, L-G, Johansson, M. O., Folestad, S., "Quantitative determination of content in binary powder mixtures using diffuse reflectance near infrared spectrometry and multivariate analysis", *Anal. Chem. Acta*, **419**, 45, (2000).
- [94] Berntsson, O., Danielsson, L-G, Lagerholm, B., Folestad, S., "Quantitative in-line monitoring of powder blending by near infrared reflection spectroscopy", *Powder Technol.*, **123**, 185, (2002).
- [95] Woodward, A. M., et. al., "Rapid and non-invasive quantification of metabolic substrates in biological cell suspensions using non-linear dielectric spectroscopy with multivariate calibration and artificial neural networks. Principles and applications", *Bioelectrochemistry and Bionergetics*, **10**, 99, (1996).
- [96] Heise, H. M., Marbach, R., Bittner, A., "Clinical chemistry and near infrared spectroscopy: multicomponent assay for human plasma and its evaluation for the determination of blood substrates", *J. Near Infrared Spectrosc.*, **6**, 361, (1998).

- [97] Burtis, C. A., Ashwood, E. R., *Tietz Fundamentos de Química Clínica*, Guanabara Koogan S. A, Philadelphia, (1998).
- [98] Stryer, L., *Bioquímica*, 3^{ed.}, Guanabara Koogan, Rio de Janeiro, (1992).
- [99] Budínová, G., Salva, J., Volka, K., “ Application of molecular spectroscopy in the mid-infrared region to the determination of glucose and cholesterol in whole blood and in blood serum”, *Appl. Spectrosc.*, **5**, 631, (1997).
- [100] Marquardt, L. A., Arnold, M. A., Small, G. W., “Near-infrared spectroscopic measurement of glucose in a protein matrix”, *Anal. Chem.*, **65**, 3271, (1993).
- [101] Harzen, K. H., Arnold, M. A., Small, G. W., “Measurement of glucose and other analytes in undiluted human serum with near-infrared transmission spectroscopy”, *Anal. Chim. Acta*, **371**, 255, (1998).
- [102] Heise, H. M., Marbach, R., “Multivariate determination of glucose in whole blood by attenuated total infrared spectroscopy”, *Anal. Chem.*, **61**, 2009, (1989).