

FABRÍCIO JOSÉ PONTES

PROJETO OTIMIZADO DE REDES NEURAI ARTIFICIAIS PARA
PREDIÇÃO DA RUGOSIDADE EM PROCESSOS DE USINAGEM
COM A UTILIZAÇÃO DA METODOLOGIA DE PROJETO DE
EXPERIMENTOS

Tese apresentada à Faculdade de Engenharia do Campus de Guaratinguetá, Universidade Estadual Paulista, para a obtenção do título de Doutor em Engenharia Mecânica na área de Gestão e Otimização.

Orientador: Prof. Dr. Messias Borges Silva
Co-Orientador: Prof. Dr. Anderson Paulo de Paiva

Guaratinguetá
2011

P814p

Pontes, Fabrício José

Projeto otimizado de redes neurais artificiais para predição da rugosidade em processos de usinagem com a utilização da metodologia de projeto de experimentos / Fabrício José Pontes - Guaratinguetá : [s.n.], 2011.

259 f. : il.

Bibliografia: f. 236

Tese (doutorado) – Universidade Estadual Paulista, Faculdade de Engenharia de Guaratinguetá, 2011.

Orientador: Prof. Dr. Messias Borges Silva

Coorientador: Prof. Dr. Anderson Paulo de Paiva

1. Redes neurais (Computação) I. Título

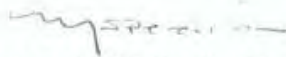
CDU 681.3

FABRÍCIO JOSÉ PONTES

ESTA TESE FOI JULGADA ADEQUADA PARA A OBTENÇÃO DO TÍTULO DE
“DOUTOR EM ENGENHARIA MECÂNICA”

PROGRAMA: ENGENHARIA MECÂNICA
ÁREA: GESTÃO E OTIMIZAÇÃO


APROVADA EM SUA FORMA FINAL PELO PROGRAMA DE PÓS-GRADUAÇÃO


Prof. Dr. Marcelo dos Santos Pereira
Coordenador

BANCA EXAMINADORA:


Prof. Dr. MESSIAS BORGES SILVA
Orientador / Unesp-Feg


Prof. Dr. MARCOS VALÉIRO RIBEIRO
Unesp - Feg


Prof. Dr. MARCELA A. G. MACHADO DE FREITAS
Unesp - Feg


Prof. Dr. DOMINGOS SÁVIO GIORDANI
EEL/USP


Prof. Dr. JOÃO ROBERTO FERREIRA
UNIFEI

Agosto de 2011

DADOS CURRICULARES

FABRÍCIO JOSÉ PONTES

NASCIMENTO	25.03.1972 – VARGINHA / MG
FILIAÇÃO	Jocias Pontes Maria Aparecida Pontes
1992/1996	Curso de Graduação em Engenharia Elétrica Escola de Engenharia da UFMG - Universidade Federal de Minas Gerais
2005/2006	Curso de Pós-graduação em Engenharia de Produção nível de Mestrado, no Instituto de Engenharia de Produção e Gestão da UNIFEI - Universidade Federal de Itajubá
2008/2011	Curso de Pós-graduação em Engenharia Mecânica, nível de Doutorado, na Faculdade de Engenharia do Campus de Guaratinguetá da Universidade Estadual Paulista.

DEDICATÓRIA

A Milena,
que em todos os momentos apoiou
e incentivou a realização deste trabalho.

AGRADECIMENTOS

Agradeço a Deus, Criador de Tudo, por me chamar à vida e me conceder os dons da Ciência e a Fortaleza para chegar até o fim deste projeto.

Agradeço a Maria Santíssima, por me haver protegido pelas estradas do caminho de Guratingetá nesse período de viagens frequentes.

A Milena, minha esposa, agradeço não somente pelo apoio e incentivo dados durante este trabalho, mas também pela compreensão pelo longo tempo que passei trabalhando. Espero poder retribuir ao menos um pouco de sua imensa ajuda.

A meus pais, Jocias e Maria Aparecida, agradeço por toda a educação, pelos valores, pela temperança e pela fé que me deram.

Agradeço ao Professor Messias Borges pelos muitos ensinamentos, pela confiança depositada, pelo profissionalismo e pela tranquilidade com que conduziu este projeto.

Ao meu amigo e segundo irmão Anderson, que forneceu uma co-orientação preciosa para a conclusão deste projeto, meu especial agradecimento.

Agradeço também à Universidade Federal de Itajubá, que me ofereceu a oportunidade de retomar meus estudos, e que, através do projeto PE024/2008 – Programa Pró-Engenharias, tornou possível a realização deste trabalho.

Agradeço a todos os professores do Departamento de Produção da FEG-UNESP. Os conhecimentos adquiridos durante as disciplinas e também durante os seminários da Pós-graduação certamente me dotaram de uma formação científica sólida e completa, que muito ajudará no desenvolvimento de minha carreira.

A meus sogros José Roberto e Sônia, agradeço pelo incentivo, por me cederem o espaço de trabalho e também pela compreensão durante o tempo em que estive na sua casa estudando, sem corresponder à hospitalidade recebida.

Agradeço a todos os meus amigos, de quem estive longe durante este longo período de Universidade. Estes anos foram de luta para mim como tem sido para eles. Espero que em breve possamos nos reencontrar com tranquilidade para celebrar tudo o que conquistamos.

Agradeço enfim à meus parentes e colegas, dos quais não foi possível estar próximo durante este período de estudos.

PONTES, F. J. **Projeto Otimizado de Redes Neurais Artificiais para predição da rugosidade em processos de usinagem com a utilização da Metodologia de Projeto de Experimentos**. 2011. 247 f. Tese (Doutorado em Engenharia Mecânica) – Faculdade de Engenharia do Campus de Guaratinguetá, UNESP – Univ. Estadual Paulista, Guaratinguetá, 2011.

RESUMO

O presente trabalho oferece contribuições à modelagem da rugosidade da peça em processos de usinagem por meio de Redes Neurais Artificiais. Propõe-se um método para o projeto de redes Perceptron Multi-Camada (Multi-Layer Perceptron, ou MLP) e Função de Base Radial (Radial Basis Function, ou RBF) otimizadas para a predição da rugosidade da peça (R_a). Desenvolve-se um algoritmo que utiliza de forma híbrida a metodologia do projeto de experimentos por meio das técnicas dos fatoriais completos e de Variações Evolucionárias em Operações (EVOP). A estratégia adotada é a de utilizar o projeto de experimentos na busca de configurações de rede que favoreçam estatisticamente o desempenho na tarefa de predição. Parâmetros de corte dos processos de usinagem são utilizados como entradas das redes. O erro médio absoluto em porcentagem (MAE%) do decil inferior das observações de predição para o conjunto de testes é utilizado como medida de desempenho dos modelos. Com o objetivo de validar o método proposto são empregados casos de treinamento gerados a partir de dados obtidos de trabalhos da literatura e de experimentos de torneamento do aço ABNT 12L14. O método proposto leva à redução significativa do erro de predição da rugosidade nas operações de usinagem estudadas, quando se compara seu desempenho ao apresentado por modelos de regressão, aos resultados relatados pela literatura e ao desempenho de modelos neurais propostos por um pacote computacional comercial para otimização de configurações de rede. As redes projetadas segundo o método proposto possuem dispersão dos erros de predição significativamente reduzidos na comparação com as redes propostas pelos estudos da literatura e pelo pacote computacional. Observa-se ainda que redes MLP atingem resultado estatisticamente superior aos obtidos pelas melhores redes RBF.

PALAVRAS-CHAVE: Rugosidade. Redes Neurais. Otimização de redes neurais. Projeto de Experimentos. EVOP.

PONTES, F. J. **Neural network optimization for prediction of surface roughness in machining processes through Design of Experiments**. 2011. 247 p. Doctoral Thesis – Faculdade de Engenharia do Campus de Guaratinguetá, UNESP - Univ. Estadual Paulista, Guaratinguetá, 2011.

ABSTRACT

The present work offers some contributions to the area of surface roughness modeling by Artificial Neural Networks in machining processes. It proposes a method for the project of networks of MLP (Multi-Layer Perceptron) and RBF (Radial Basis Function) architectures optimized for prediction of Average Surface Roughness (R_a). The method is expressed in the format of an algorithm employing two techniques from the DOE (Design of Experiments) methodology: Full factorials and Evolutionary Operations (EVOP). The strategy adopted consists in the systematic use of DOE in a search for network configurations that benefits performance in roughness prediction. Cutting parameters from machining operations are employed as network inputs. The mean absolute error in percentage (MAE%) of the lower decile of the predictions for the test set is used as a figure of merit for network performance. In order to validate the method, data sets retrieved from literature, as well as results of experiments with AISI/SAE free-machining steel, are employed to form training and test data sets for the networks. The proposed algorithm leads to significant reduction in prediction error for surface roughness when compared to the performance delivered by a regression model, by the networks proposed by the original studies data was borrowed from and when compared to neural models proposed by a software package intended to search automatically for optimal network configurations. In addition, networks designed according to the proposed algorithm displayed reduced dispersion of prediction errors in comparison to those suggested by literature and by the software package. It is also observed that networks of MLP architecture outperform networks of RBF architecture in the proposed task.

KEY-WORDS: Surface roughness. Neural Networks. Neural network optimization. DOE, EVOP.

LISTA DE FIGURAS

FIGURA 1 – Retirada de material em processo de usinagem.....	33
FIGURA 2 – Elementos em operação de torneamento.	34
FIGURA 3 – Sequência dos três instantes principais de um ciclo EDM.	36
FIGURA 4 – Desvios no formato de superfícies.	38
FIGURA 5 - Desvio médio aritmético R_a	39
FIGURA 6 – O neurônio de McCulloch & Pitts	44
FIGURA 7 – Conceito de sobre-ajuste de redes.	48
FIGURA 8 – Rede MLP	52
FIGURA 9 – Função de base radial.	56
FIGURA 10 – Rede neural de arquitetura RBF	57
FIGURA 11 – Ilustração do conceito de EVOP.....	78
FIGURA 12: Processo de construção e validação de modelos	80
FIGURA 13 - Distribuição anual das publicações revisadas	85
FIGURA 14 - Processos de usinagem estudados nas publicações revisadas	85
FIGURA 15 – Número de preditores utilizados para construção dos modelos nas publicações revisadas	94
FIGURA 16 – Natureza dos preditores utilizados para construção dos modelos de redes neurais	94
FIGURA 17 – Abordagem adotada para formação dos casos de treinamento.....	96
FIGURA 18 - Natureza das saídas de rede adotadas nas publicações revisadas	98
FIGURA 19 – Natureza das saídas de rede adotadas nas publicações revisadas.....	98
FIGURA 20 – Número de ocorrências de distintas arquiteturas de redes ao longo das publicações revisadas	100
FIGURA 21 – Abordagens adotadas para definição de topologias nas publicações revisadas	102
FIGURA 22 – Algoritmos adotados para o treinamento de redes MLP	104
FIGURA 23 – Critérios adotados para interrupção do treinamento de redes neurais de arquitetura MLP.....	105
FIGURA 24 – Grandezas adotadas para seleção de redes	107

FIGURA 25 – Grandezas adotadas para validação de redes	107
FIGURA 26 – Uso de conjuntos distintos de treinamento e teste.....	109
FIGURA 27 – Uso de terceiro conjunto de casos para validação	110
FIGURA 28 – Erro em predição de curvas sintéticas com método da Descendente mais inclinada.....	116
FIGURA 29 – Superfície em forma de sela obtida dos experimentos preliminares usando a técnica CCD	119
FIGURA 30 – Superfície com indicação de região de ótimo obtida dos experimentos ampliados usando a técnica CCD	120
FIGURA 31 – Curvas de nível com indicação de região de ótimo obtida dos experimentos ampliados usando a técnica CCD	120
FIGURA 32 – Superfície de erro obtida do experimento de saturação	121
FIGURA 33 – Efeitos principais –primeiro ciclo de otimização de redes MLP para dados de Sarkar et al. (2006)	154
FIGURA 34 – Efeitos de interação dupla – primeiro ciclo de otimização de redes MLP para dados de Sarkar et al. (2006).....	155
FIGURA 35 – Gráfico de Pareto – primeiro ciclo de otimização de redes MLP para dados de Sarkar et al. (2006) – Ciclo 1	155
FIGURA 36 – Distribuição dos pontos de medição nos corpos de prova.....	163
FIGURA 37 – Torno empregado para os experimentos de usinagem.	164
FIGURA 38 – Rugosímetro utilizado para medição da rugosidade.....	165
FIGURA 39 – Evolução de desempenho das melhores redes a cada ciclo – aplicação do método a curvas sintéticas – 30 casos de treinamento.	170
FIGURA 40 – Sumário gráfico da distribuição de erro da rede RBF otimizada – experimento com curvas sintéticas – 30 casos de treinamento	171
FIGURA 41 – Comparação entre rede otimizada RBF para curvas sintéticas e melhor rede obtida por Pontes (2006) – 30 casos de treinamento	172
FIGURA 42 – Gráfico ilustrativo da comparação entre a distribuição de predições entre rede RBF ótima e melhor rede obtida por Pontes (2006) – 30 casos de treinamento	173

FIGURA 43 – Comparação entre rede otimizada RBF para curvas sintéticas e melhor rede obtida por Pontes (2006) – 15 casos de treinamento	174
FIGURA 44 – Evolução de desempenho das melhores redes MLP a cada ciclo – aplicação do método ao conjunto de dados do Artigo 1.....	178
FIGURA 45 – Sumário gráfico da distribuição de erro da rede MLP ótima – experimento com dados do Artigo 1	179
FIGURA 46 – Resposta da rede MLP ótima para dados do Artigo 1 – AA versus Vc	180
FIGURA 47 – Resposta da rede MLP ótima para dados do Artigo 1 – AA versus f.....	180
FIGURA 48 – Resposta da rede MLP ótima para dados do Artigo 1 – AA versus ap	181
FIGURA 49 – Resposta da rede MLP ótima para dados do Artigo 1 – f versus ap.....	181
FIGURA 50 – Teste de hipótese para médias iguais de erro de predição entre a melhor rede MLP obtida pelo método e resultado do Artigo 1.....	182
FIGURA 51 – Diagrama de caixa do teste de hipótese entre erro de predição da melhor rede MLP obtida pelo método e resultado do Artigo 1.....	183
FIGURA 52 – Análise de regressão do conjunto de dados obtido do Artigo 1	184
FIGURA 53 – Teste de hipótese para médias iguais de erro de predição em MAE% entre a melhor rede MLP obtida pelo método e modelo de regressão - Artigo 1	185
FIGURA 54 – Diagrama de caixa do teste de hipótese para comparação entre a melhor rede MLP obtida pelo método e modelo de regressão para o Artigo 1.....	186
FIGURA 55 – Comparação entre erro de predição em MAE% da melhor rede MLP obtida pelo método e da melhor rede apontada pelo pacote computacional - Artigo 1	187
FIGURA 56 – Comparação entre a melhor rede MLP obtida pelo método e melhor rede apontada pelo pacote computacional para dados do Artigo 1	188
FIGURA 57 – Evolução de desempenho das melhores redes RBF a cada ciclo – aplicação do método ao conjunto de dados do Artigo 1.....	189
FIGURA 58 – Comparação entre redes otimizadas RBF e MLP - dados do Artigo 1	190
FIGURA 59 – Evolução de desempenho das melhores redes MLP a cada ciclo – aplicação do método ao conjunto de dados do Artigo 2.....	193
FIGURA 60 – Sumário gráfico da distribuição de erro da rede MLP ótima – experimento com dados do Artigo 2	194

FIGURA 61 – Teste de hipótese para médias iguais de erro de predição entre a melhor rede MLP obtida pelo método e resultado do Artigo 2.....	195
FIGURA 62 – Análise de regressão do conjunto de dados obtido do Artigo 2	196
FIGURA 63 – Teste de hipótese para médias iguais de erro de predição em MAE% entre a melhor rede MLP obtida pelo método e modelo de regressão - Artigo 2	197
FIGURA 64 – Diagrama de caixa do teste de hipótese para comparação entre a melhor rede MLP obtida pelo método e modelo de regressão para o Artigo 2.....	198
FIGURA 65 – Comparação entre erro de predição em MAE% da melhor rede MLP obtida pelo método e da melhor rede apontada pelo pacote computacional - Artigo 2....	199
FIGURA 66 – Comparação entre a melhor rede MLP obtida pelo método e melhor rede apontada pelo pacote computacional para dados do Artigo 2	200
FIGURA 67 – Evolução de desempenho das melhores redes RBF a cada ciclo – aplicação do método ao conjunto de dados do Artigo 2.....	201
FIGURA 68 – Comparação entre rede otimizada RBF e rede otimizada MLP para dados do Artigo 2	202
FIGURA 69 – Evolução de desempenho das melhores redes MLP a cada ciclo – aplicação do método ao conjunto de dados do Artigo 3.....	205
FIGURA 70 – Sumário gráfico da distribuição de erro da rede MLP ótima – experimento com dados do Artigo 3	206
FIGURA 71 – Teste de hipótese para médias iguais de erro de predição entre a melhor rede MLP obtida pelo método e resultado do Artigo 3.....	207
FIGURA 72 – Diagrama de caixa do teste de hipótese entre erro de predição da melhor rede MLP obtida pelo método e resultado do Artigo 3.....	208
FIGURA 73 – Análise de regressão do conjunto de dados obtido do Artigo 3	209
FIGURA 74 – Teste de hipótese para médias iguais de erro de predição em MAE% entre a melhor rede MLP obtida pelo método e modelo de regressão - Artigo 3	210
FIGURA 75 – Diagrama de caixa do teste de hipótese para comparação entre a melhor rede MLP obtida pelo método e modelo de regressão para o Artigo 3.....	211
FIGURA 76 – Comparação entre erro de predição em MAE% da melhor rede MLP obtida pelo método e da melhor rede apontada pelo pacote computacional - Artigo 3	212

FIGURA 77 – Comparação entre a melhor rede MLP obtida pelo método e melhor rede apontada pelo pacote computacional para dados do Artigo 3	213
FIGURA 78 – Evolução de desempenho das melhores redes RBF a cada ciclo – aplicação do método ao conjunto de dados do Artigo 3.....	214
FIGURA 79 – Evolução de desempenho das melhores redes MLP a cada ciclo – aplicação do método aos resultados dos experimentos de usinagem	218
FIGURA 80 – Sumário gráfico da distribuição de erro da rede MLP ótima – experimento com resultados dos experimentos de usinagem.....	219
FIGURA 81 – Análise de regressão do conjunto de dados resultante dos experimentos de usinagem	221
FIGURA 82 – Teste de hipótese para médias iguais de erro de predição em MAE% entre a melhor rede MLP obtida pelo método e modelo de regressão obtido dos resultados dos experimentos de usinagem.....	222
FIGURA 83 – Diagrama de caixa do teste de hipótese entre erro de predição da melhor rede MLP obtida pelo método e erro do modelo de regressão para resultados dos experimentos de usinagem.....	223
FIGURA 84 – Comparação entre melhor rede MLP obtida pelo método e melhor rede apontada pelo pacote computacional para resultados dos experimentos de usinagem.....	224
FIGURA 85 – Diagrama de caixa da comparação entre erro de predição da melhor rede MLP obtida pelo método e melhor rede apontada pelo pacote computacional para resultados dos experimentos de usinagem.....	224
FIGURA 86 – Evolução de desempenho das melhores redes RBF a cada ciclo – aplicação do método aos resultados dos experimentos de usinagem	225
FIGURA 87 – Comparação entre rede otimizada RBF e rede otimizada MLP para dados resultantes dos experimentos de usinagem	227

LISTA DE TABELAS

TABELA 1 – Níveis dos fatores durante a aplicação do método para otimização de redes para curvas sintéticas – 30 casos de treinamento – ciclos de 1 a 9	167
TABELA 2 – Níveis dos fatores durante a aplicação do método para otimização de redes para curvas sintéticas – 30 casos de treinamento – ciclos de 10 a 18	168
TABELA 3 – Pontos centrais e melhores configurações de rede obtidos ao longo da aplicação do método para curvas sintéticas – 30 casos de treinamento.	169
TABELA 4 – Configuração de rede neural RBF otimizada para predição da rugosidade para curvas sintéticas – 30 casos de treinamento.....	170
TABELA 5 – Desempenho de rede neural RBF otimizada para predição da rugosidade para curvas sintéticas – 30 casos de treinamento.....	171
TABELA 6 – Configuração de rede neural RBF otimizada para predição da rugosidade para curvas sintéticas – 15 casos de treinamento.....	174
TABELA 7 – Níveis dos fatores durante a aplicação do método para otimização de redes usando dados do Artigo 1	176
TABELA 8 – Pontos centrais e melhores configurações de rede obtidos ao longo da aplicação do método ao conjunto de dados do Artigo 1.....	177
TABELA 9 – Configuração de rede neural MLP otimizada para predição da rugosidade usando os dados do Artigo 1	178
TABELA 10 – Desempenho de rede neural MLP otimizada para predição da rugosidade usando os dados do Artigo 1	178
TABELA 11 – Configuração de rede neural RBF otimizada para predição da rugosidade com uso dos dados do Artigo 1.....	189
TABELA 12 – Desempenho de rede neural RBF otimizada para predição da rugosidade com uso dos dados do Artigo 1.....	189
TABELA 13 – Níveis dos fatores durante a aplicação do método para otimização de redes usando dados do Artigo 2.....	192
TABELA 14 – Pontos centrais e melhores configurações de rede obtidos ao longo da aplicação do método ao conjunto de dados do Artigo 2.....	192

TABELA 15 – Configuração de rede neural MLP otimizada para predição da rugosidade usando os dados do Artigo 2.....	193
TABELA 16 – Desempenho de rede neural MLP otimizada para predição da rugosidade com os dados do Artigo 2	194
TABELA 17 – Configuração de rede neural RBF otimizada para predição da rugosidade com uso dos dados do Artigo 2.....	201
TABELA 18 – Desempenho de rede neural RBF otimizada para predição da rugosidade usando os dados do Artigo 2.....	201
TABELA 19 – Níveis dos fatores durante a aplicação do método para otimização de redes usando dados do Artigo 3.....	204
TABELA 20 – Pontos centrais e melhores configurações de rede obtidos ao longo da aplicação do método ao conjunto de dados do Artigo 3.....	205
TABELA 21 – Configuração de rede neural MLP otimizada para predição da rugosidade usando os dados do Artigo 3.....	206
TABELA 22 – Desempenho de rede neural MLP otimizada para predição da rugosidade com os dados do Artigo 3	206
TABELA 23 – Configuração de rede neural RBF otimizada para predição da rugosidade com uso dos dados do Artigo 3.....	214
TABELA 24 – Desempenho de rede neural RBF otimizada para predição da rugosidade usando os dados do Artigo 3.....	215
TABELA 25 – Níveis dos fatores durante a aplicação do método para otimização de redes usando resultados dos experimentos de usinagem.....	217
TABELA 26 – Pontos centrais e melhores configurações de rede obtidos ao longo da aplicação do método aos resultados dos experimentos de usinagem.	218
TABELA 27 – Configuração de rede neural MLP otimizada para predição da rugosidade usando os resultados dos experimentos de usinagem.....	219
TABELA 28 – Desempenho de rede neural MLP otimizada para predição da rugosidade usando os resultados dos experimentos de usinagem.....	219
TABELA 29 – Configuração de rede neural RBF otimizada para predição da rugosidade com uso dos resultados dos experimentos de usinagem.....	226

TABELA 30 – Desempenho de rede neural RBF otimizada para predição da rugosidade com uso dos resultados dos experimentos de usinagem.....	226
TABELA 31 – Comparativo dos erros de predição entre os modelos de RNAs otimizados e erros obtidos por outros modelos	228

LISTA DE QUADROS

QUADRO 1 – Hipóteses presentes em uma ANOVA para dois fatores.	65
QUADRO 2 – Fórmulas para análise de variância para dois fatores em dois níveis.....	65
QUADRO 3 – Fórmulas para análise de variância em dois níveis.	67
QUADRO 4 - Sumário das decisões sobre redes tomadas nas publicações revisadas	88
QUADRO 5 - Sumário das decisões sobre redes tomadas nas publicações revisadas	89
QUADRO 6 – Fatores e níveis do arranjo interno - experimento exploratório MLP	124
QUADRO 7 – Fatores e níveis do arranjo externo - experimento exploratório MLP	124
QUADRO 8 – Fatores utilizados para a otimização de redes MLP	131
QUADRO 9 – Arranjo fatorial completo - a otimização de redes MLP	133
QUADRO 10 – Fatores utilizados para a otimização de redes RBF	134
QUADRO 11 – Arranjo fatorial completo utilizado durante a otimização de redes RBF.....	135
QUADRO 12 – Fatores de entrada da curva sintética utilizada para validação	158
QUADRO 13 – Fatores de entrada do conjunto de treinamento - Artigo 1.....	159
QUADRO 14 – Fatores de entrada do conjunto de treinamento - Artigo 2.....	160
QUADRO 15 – Fatores de entrada do conjunto de treinamento - Artigo 3.....	160
QUADRO 16 – Composição química do aço ABNT 12L14	161
QUADRO 17 – Fatores de entrada, ruídos e respectivos níveis utilizados nos experimentos de usinagem	162

LISTA DE ABREVIATURAS E SIGLAS

AISI	American Iron and Steel Institute
Al	Alumínio
Al ₂ O ₃	Trióxido de Alumínio
ANOVA	Analysis of Variance
Cr	Cromo
DOE	Design of Experiments
EDM	Electrical Discharge Machining
EVOP	Evolutionary Operations
GB	Giga Byte
HRC	Rockwell C Hardness
IJAMT	International Journal of Advanced Manufacturing Technology
JBSMSE	Journal of Brazilian Society of Mechanical Scientists and Engineers
MAE	Mean Absolute Error
MLP	Multi-Layer Perceptron
Nb	Nióbio
RAM	Random Access Memory (Memória de acesso aleatório)
RBF	Radial Basis Function
rpm	Rotações por minuto
RNA	Rede Neural Artificial
SiC	Carbeto de Silício
Ti	Titânio
TiAlN	Nitreto de Alumínio e Titânio
TRM	Taxa de Remoção de Material

LISTA DE SÍMBOLOS

a	Número de níveis do fator experimental A
A	Fator experimental genérico A
A	Ampere
ap	Profundidade de Usinagem
α	Probabilidade do Erro Tipo I
b	Número de níveis do fator experimental B
B	Fator experimental genérico B
β	Probabilidade do Erro Tipo II
β_i	Coefficiente do Modelo de Regressão
c0	Constante multiplicadora na formula 4 para Rugosidade
c1	Expoente da Velocidade de Corte na formula 4 para Rugosidade
c2	Expoente do Avanço de Corte na formula 4 para Rugosidade
c3	Expoente da Profundidade de Usinagem na formula 4 para Rugosidade
\vec{d}	Vetor resposta desejada para vetor de saída proveniente da camada oculta
e	Resíduo
exp	Função Exponencial
ϵ	Termo de Erro
Σ	Somatório
f	Avanço
f(.)	Função de Ativação
F	Estatística de teste utilizada pela Análise de Variância
$F_{\text{crítico}}$	Valor Crítico da Estatística de Teste F
ϕ_i	Função de Base Radial i
g	gramas
G	Matriz de Elemento (i,j) igual à entrada da i-ésima unidade de saída, quando o j-ésimo caso de treinamento é executado
G_+	Matriz Pseudo-Inversa
G^T	Transposta da Matriz G
h(x)	Função de saída da Rede Neural RBF

H_0	Hipótese nula
H_1	Hipótese alternativa
k	Número de fatores experimentais
K	Número de unidades radiais vizinhas
l_m	Comprimento de amostragem do perfil
L	Comprimento de amostragem
L	Valor da Função de Mínimos Quadrados
mm	Milímetro
mm/v	Milímetros por volta
m/min	Metros por minutos
m/s^2	Metros por Segundo ao quadrado
M	Linha Média
μm	Micrometro
μs	Microsegundo
$\bar{\mu}_j$	Hiper-Centro Vetorial da Função de Ativação j
n	Número de amostras ou de réplicas
n_d	Número de Parâmetros Livres da Rede / Dimensão da Função de Erro
N	Número de corridas experimentais
N	Número de vetores de entrada
r_e	Raio de Ponta da Ferramenta
R_a	Rugosidade Média
R_{max}	Rugosidade Máxima
R^2	Coeficiente de Determinação
S	Erro Residual
SS_t	Soma de Quadrados Total
SS_A	Soma de Quadrados Devida aos Tratamentos de linha (Fator A)
SS_B	Soma de Quadrados Devida aos Tratamentos de coluna (Fator B)
SS_{AB}	Soma de Quadrados Devida à Interação entre Fatores A e B
SS_E	Soma de Quadrados Devida ao Erro Experimental
σ	Parâmetro de Largura da Função Radial
\vec{i}	Espaço Vetorial Unidimensional

T_{on}	Período de descarga
T_{off}	Período de intervalo
v_c	Velocidade de corte
VB_{max}	Desgaste de flanco máximo
x_i	Entrada i da rede RBF
\bar{X}	Média amostral
\vec{x}	Vetor de entrada
\vec{x}^n	Espaço vetorial de entrada de n dimensões
χ^2	Distribuição qui-quadrado
y_i	Valor absoluto da ordenada do perfil medido em relação à linha média
y_i	Saída do modelo polynomial quadratico completo
\bar{y}_i	Média das n_i Observações no nível x_i
y_k	Saída da rede RBF
$y(\vec{x})$	Saída da rede RBF em função do vetor de entrada \vec{x}
$ y(x) $	Valores Absolutos dos Desvios Medidos em Relação à Linha Média
w_i	Valor do peso da sinapse i
w_0	Fator constante
%	Porcentagem
$1-\alpha$	Nível de confiança do teste
$1-\beta$	Poder do teste
°	Graus

SUMÁRIO

1	INTRODUÇÃO	26
1.1	Considerações Iniciais	26
1.2	Justificativa e Relevância	27
1.3	Objetivos	28
1.3.1	Objetivo Geral	28
1.3.2	Objetivos Específicos	28
1.4	Delimitações do Trabalho	29
1.5	Caráter Inédito	29
1.6	Metodologia	30
1.7	Organização do Texto	31
2	REVISÃO DA BIBLIOGRAFIA	33
2.1	Usinagem e Torneamento	33
2.2	Usinagem por Descargas Elétricas	35
2.3	Rugosidade da Peça	37
2.4	Redes Neurais Artificiais	42
2.4.1	Considerações Gerais	42
2.4.2	Histórico das Redes Neurais Artificiais	44
2.4.3	Aprendizado	46
2.4.4	Projeto de Redes Neurais	48
2.4.5	Redes de Arquitetura MLP (Perceptron Multi-Camada)	52
2.4.6	Treinamento de Redes MLP	53
2.4.7	Redes de Arquitetura RBF (Função de Base Radial)	55
2.4.8	Projeto e Treinamento de Redes RBF	58
2.5	Metodologia de Projeto e Análise de Experimentos	61
2.5.1	Considerações Gerais	61
2.5.2	Fatoriais Completos	62
2.5.3	Análise de Variância (ANOVA)	63
2.5.4	Testes para Significância	68
2.5.5	Modelos de Regressão Linear	68

2.5.6	Medidas de Adequação de Modelos de Regressão	74
2.5.7	Análise Residual	74
2.5.8	Testes para Falta de Ajustes Dos Modelos	75
2.5.9	Variações Evolucionárias em Operações (EVOP)	77
2.6	O Conceito de Modelagem	79
3	A MODELAGEM DA RUGOSIDADE DA PEÇA POR REDES NEURAIAS ARTIFICIAIS	81
3.1	Considerações Iniciais	81
3.2	Redes Neurais Aplicadas a Processos de Usinagem	81
3.3	Caracterização dos Trabalhos Analisados	84
3.4	Definição do Problema e Coleta de Dados	91
3.5	Seleção e Adequação de Modelos	99
3.6	Validação de Modelos	108
3.7	Conclusões	112
4	OTIMIZAÇÃO DE REDES NEURAIAS BASEADA EM PROJETO DE EXPERIMENTOS	114
4.1	Considerações Iniciais	114
4.2	Escopo da Proposta	114
4.3	Evolução do Trabalho	115
4.3.1	Concepção do Método Interativo	115
4.3.2	O Problema com Redes de Arquitetura RBF	123
4.3.3	A Definição da Variável de Saída	126
4.4	Apresentação e Detalhamento do Método	127
4.4.1	Conceito	127
4.4.2	Definições	128
4.4.3	Recomendações Bibliográficas observadas	129
4.4.4	Definição dos Fatores e Arranjos Experimentais	131
4.4.4.1	Fatores para experimentos com redes MLP	131
4.4.4.2	Arranjo experimental para redes MLP	132
4.4.4.3	Fatores para experimentos com redes RBF	134
4.4.4.4	Arranjo experimental para redes RBF	134

4.4.5	Algoritmo Geral	135
4.4.6	Detalhamento das Sub-rotinas	143
4.4.6.1	Estabelecer <i>Ponto Central do Arranjo</i>	143
4.4.6.2	Estabelecer <i>Intervalos Entre Níveis dos Fatores</i>	144
4.4.6.3	Estabelecer <i>Níveis dos Fatores do Arranjo</i>	146
4.4.6.4	Executar ciclo experimental de treinamento, seleção e validação das redes	146
4.4.6.5	Tabular resultados dos testes de validação	147
4.4.6.6	Analisar resultados dos testes de validação	147
4.4.6.7	Determinar <i>Melhor Configuração Ciclo Atual</i>	148
4.4.6.8	Estabelecer novo <i>Ponto central do Arranjo</i>	149
4.4.6.9	Reduzir <i>Intervalos Entre Níveis dos Fatores</i>	151
4.5	Validação do Método	151
5	APRESENTAÇÃO E ANÁLISE DOS RESULTADOS	153
5.1	Considerações Iniciais	153
5.2	Realização e Análise dos Experimentos	153
5.3	Caracterização dos Trabalhos Utilizados para Validação	157
5.3.1	Curvas Sintéticas	157
5.3.2	Artigos da Literatura	158
5.3.3	Experimentos de Usinagem	161
5.4	Aplicação do Método a Curvas Sintéticas	166
5.4.1	Valores Iniciais e Desenvolvimento	166
5.4.2	Resultados da Aplicação do Método	167
5.4.3	Comparação com Resultados do Trabalho Original	172
5.4.4	Repetição do Estudo Utilizando Conjunto de Dados Contendo 15 Casos de Treinamento	173
5.5	Aplicação do Método a Conjuntos de Dados Extraídos da Literatura	175
5.5.1	Aplicação do Método a Dados do Artigo N° 1	175
5.5.1.1	Valores Iniciais e Desenvolvimento	175
5.5.1.2	Resultados da Aplicação do Método	176
5.5.1.3	Comparação com Resultados do Trabalho Original	182

5.5.1.4	Comparação com Resultados de Modelo de Regressão Múltipla	183
5.5.1.5	Comparação com Resultados da Ferramenta Automática de Otimização de Redes	186
5.5.1.6	Reprodução do Estudo com Redes de Arquitetura RBF	188
5.5.2	Aplicação do Método a Dados do Artigo N° 2	191
5.5.2.1	Valores Iniciais e Desenvolvimento	191
5.5.2.2	Resultados da Aplicação do Método	191
5.5.2.3	Comparação com Resultados do Trabalho Original	194
5.5.2.4	Comparação com Resultados de Modelo de Regressão Múltipla	195
5.5.2.5	Comparação com Resultados da Ferramenta Automática de Otimização de Redes	198
5.5.2.6	Reprodução do Estudo com Redes de Arquitetura RBF	200
5.5.3	Aplicação do Método a Dados do Artigo N° 3	203
5.5.3.1	Valores Iniciais e Desenvolvimento	203
5.5.3.2	Resultados da Aplicação do Método	203
5.5.3.3	Comparação com Resultados do Trabalho Original	207
5.5.3.4	Comparação com Resultados de Modelo de Regressão Múltipla	208
5.5.3.5	Comparação com Resultados da Ferramenta Automática de Otimização de Redes	211
5.5.3.6	Reprodução do estudo com redes de arquitetura RBF	213
5.6	Aplicação do Método a Resultados Experimentais No Torneamento de Aço de Corte Fácil ABNT 12L14	216
5.6.1	Valores iniciais e desenvolvimento	216
5.6.2	Resultados da aplicação do método	216
5.6.3	Comparação com resultados de modelo de regressão múltipla	220
5.6.4	Comparação com resultados de ferramenta automática de otimização de redes	223
5.6.5	Reprodução do estudo com redes de arquitetura RBF	225
5.7	Considerações sobre os Resultados Obtidos	227

5.8	Comentários sobre o Custo Computacional da Aplicação do Método Proposto	229
6	CONCLUSÕES E SUGESTÕES PARA PESQUISAS FUTURAS	232
6.1	Conclusões	232
6.2	Sugestões para Pesquisas Futuras	233
	REFERÊNCIAS BIBLIOGRÁFICAS	236
APÊNDICE A	Conjuntos de dados formados a partir de curvas sintéticas para treinamento e teste de redes	250
APÊNDICE B	Conjuntos de dados extraídos do artigo 1 para treinamento e teste de redes	253
APÊNDICE C	Conjuntos de dados extraídos do artigo 2 para treinamento e teste de redes	255
APÊNDICE D	Conjuntos de dados extraídos do artigo 3 para treinamento e teste de redes	258
APÊNDICE E	Conjunto de testes resultante dos experimentos de usinagem com aço ABNT 12L14 e utilizado para teste de redes	259

1 INTRODUÇÃO

1.1 CONSIDERAÇÕES INICIAIS

A indústria de usinagem enfrenta atualmente um cenário caracterizado pela concorrência extrema, por margens operacionais estreitas e por um mercado consumidor cada vez mais exigente. Enfrenta também os desafios técnicos impostos por requisitos complexos de forma e por estreitos limites de especificação, pela frequente introdução de novos materiais de difícil usinagem, de novos tipos de ferramentas de corte e de novas variantes de operação (tais como o torneamento duro), que dão origem a processos governados por fenômenos físicos distintamente não-lineares e com objetivos conflitantes (KARPAT & ÖZEL, 2007; ARANTES, 2007; SINGH & RAO, 2007; CUS & ZUPERL, 2006).

Em função dos fatores mencionados, a obtenção de estimativas de parâmetros de usinagem tais como rugosidade e vida da ferramenta para novos processos ou materiais tem se tornado um importante diferencial competitivo. Diversos pesquisadores têm proposto a utilização de modelos de usinabilidade como solução para o problema (OKTEM, 2009; KARAYEL, 2009; CUS & ZUPERL, 2006).

Uma área para a qual modelos de usinabilidade têm sido intensivamente investigados é a de qualidade de superfície, caracterizada principalmente pela rugosidade de peça (ÖZEL & KARPAT, 2005). Trata-se de um requisito essencial ao consumidor em razão de seu impacto sobre o desempenho do produto (OKTEM, 2009; AMBROGIO *ET AL.*, 2008; TAMIZHARASAN *ET AL.*, 2006). O processo de formação da rugosidade é complexo, sendo governado por fatores físico-químicos, e é extremamente não linear (KARAYEL, 2009; SINGH & RAO, 2007; CUS & ZUPERL, 2006).

Uma abordagem freqüente na literatura para a modelagem da rugosidade tem sido o uso de redes neurais artificiais, ou RNAs, as quais correspondem a um dos paradigmas da inteligência artificial (BENARDOS & VOSNIAKOS, 2003). O uso de RNAs para o controle estatístico da rugosidade, no entanto, é caracterizado por grande

variação nas técnicas utilizadas e pela falta de consenso quanto à sua eficácia e eficiência (CORREA *ET AL.*, 2009; ÇAYDAS & HASÇALIK, 2007).

A proposta desta tese é desenvolver um método de projeto de Redes Neurais Artificiais de arquiteturas Perceptron Multi-Camada (*Multi-Layer Perceptron*, ou MLP) e Função de Base Radial (*Radial Basis Function*, ou RBF) otimizadas para a predição da rugosidade da peça (R_a) em processos de usinagem combinando distintos princípios da Metodologia de Projeto de Experimentos, tais como a técnica de Fatoriais Completos e a técnica de Variações Evolucionárias em Operações (*Evolutionary Operations*, ou EVOP).

1.2 JUSTIFICATIVA E RELEVÂNCIA

As justificativas para a realização do projeto provêm das necessidades enfrentadas pela indústria manufatureira e também dos problemas apontados por diversos autores com relação ao projeto e utilização de redes neurais para aplicação em processos de usinagem, e podem ser agrupadas nas seguintes categorias:

- ***Importância da modelagem nos processos de usinagem:*** em função da necessidade de eficiência econômica, diversos autores preconizam a utilização de modelos como funções objetivo para otimização e simplificação de processos (REDDY & RAO, 2005; PAIVA *ET AL.*, 2009; SINGH & RAO, 2007);
- ***Lacunas em métodos de projeto para redes neurais:*** muitos autores apontam para a inexistência de métodos sistemáticos de projeto como entrave ao uso de RNAs (ZHONG *ET AL.*, 2006; ZANCHETTIN *ET AL.*, 2005). Afirma-se também ser necessária grande quantidade de dados para a obtenção de modelos eficazes (DHOKIA *ET AL.*, 2008; CERVELLERA *ET AL.*, 2007; KARNIK *ET AL.*, 2008; AMBROGIO *ET AL.*, 2008; BAGCI & ISIK, 2006).
- ***Oportunidade de melhoria na qualidade dos estudos de modelagem da rugosidade por meio de RNAs:*** o estudo da literatura disponível sobre o tema revela extenso uso de tentativa e erro, desconsideração de princípios de neuro-

computação e estatística na modelagem, raro uso da metodologia de Projeto de Experimentos (DOE) e escasso emprego de arquiteturas de rede diferentes de MLPs para a modelagem (PONTES *ET AL.*, 2010; ERZURUMLU & OKTEM, 2007).

1.3 OBJETIVOS

1.3.1 OBJETIVO GERAL

O objetivo geral desta tese é desenvolver um método para o projeto de redes neurais com desempenho ótimo na predição da rugosidade da peça (R_a) em processos de usinagem. O método proposto envolverá a utilização da metodologia de Projeto de Experimentos (DOE) tais como Fatoriais Completos e Operação Evolucionária.

1.3.2 OBJETIVOS ESPECÍFICOS

Os objetivos específicos desta tese incluem:

- Obter modelos de predição de rugosidade da peça com desempenho superior ao apresentado por modelos de redes projetadas segundo a prática vigente;
- Obter modelos de predição de rugosidade da peça com desempenho superior ao apresentado por modelos de redes projetadas automaticamente por meio de ferramenta computacional;
- Obter modelos de predição de rugosidade da peça com desempenho superior ao apresentado por modelos obtidos isoladamente por modelos baseados nos métodos dos mínimos quadrados;
- Comparar o desempenho de predição de rugosidade de duas arquiteturas distintas de redes neurais, a saber: Perceptrons Multi-Camada e Funções de Base Radial.

1.4 DELIMITAÇÕES DO TRABALHO

Este trabalho estudará redes neurais de arquitetura Perceptron Multi-Camada e Função de Base Radial aplicadas à predição da rugosidade em processo de torneamento e usinagem por descarga elétrica.

Os resultados obtidos não deverão ser extrapolados para outras arquiteturas de redes neurais que não as incluídas nesta tese. Também não deverão ser extrapolados para outros processos de usinagem, materiais usinados ou tipos de ferramentas de corte.

1.5 CARÁTER INÉDITO

O presente projeto pretende inovar em relação ao estado atual da pesquisa sobre o tema nos seguintes aspectos:

- na definição da otimização de redes como sendo a busca por um domínio de parâmetros que resulte em topologias de rede com distribuição de erro favorável à ocorrência de erro mínimo de predição (em oposição à busca pela minimização de um estimador pontual de desempenho, tal como o erro médio);
- na utilização de uma estratégia híbrida de Projeto de Experimentos que inclui o uso de Fatoriais Completos e de uma adaptação da técnica de Operação Evolucionária ao projeto de redes neurais;
- na abordagem do problema do projeto de redes como um problema de elaboração de modelos, com solução baseada em fundamentos estatísticos e nas boas práticas recomendadas por autores do campo da neurocomputação;
- na proposta de um método sequencial e bem definido (sob a forma de algoritmo) para projeto de redes neurais para predição da rugosidade em processos de usinagem, que permita identificar configurações de rede com a máxima exatidão possível na predição para um dado conjunto de treinamento;
- na exploração da otimização de rede neurais da arquitetura Função de

Base Radial (RBF).

O caráter inédito desta tese é corroborado por artigos internacionais produzidos a partir dos resultados obtidos. Até o momento, dois artigos já foram publicados (PONTES *ET AL.*, 2010a; PONTES *ET AL.*, 2010b) nos seguintes periódicos: International Journal of Advanced Manufacturing Technology (IJAMT) and Journal of Brazilian Society of Mechanical Sciences and Engineering (JBSMSE). Alguns outros se encontram em processo de revisão.

1.6 METODOLOGIA

A análise da literatura revela que a maioria dos estudos sobre predição da rugosidade por redes neurais segue uma sequência comum de atividades que inclui a delimitação do problema, a definição de uma estratégia experimental, a coleta de dados, a escolha de uma arquitetura e topologia de rede, treinamento da rede, análise de dados, seleção do modelo de rede, e algum tipo de validação dos resultados. Esta sequência apresenta forte semelhança com o arcabouço conceitual de construção e validação de modelos proposto por Montgomery *et al.* (2008) e por Montgomery & Runger (2007) para a análise de regressão múltipla e previsão de séries temporais. O método experimental de pesquisa aparece como mais adequado, tendo em vista que é possível, pelo uso de ferramentas computacionais, manter todas as variáveis de projeto sob controle.

A metodologia do Projeto de Experimentos (DOE) será utilizada, mais especificamente as técnicas de Fatoriais Completos e a Operação Evolucionária. Os fatores do DOE (as variáveis independentes) serão os parâmetros de projeto das redes neurais. O planejamento experimental terá por meta identificar topologias e parâmetros de redes neurais que otimizem as habilidades de predição das redes; objetivará também avaliar a importância relativa de cada parâmetro de projeto no desempenho da rede e investigar as possíveis interações entre os níveis dos fatores de projeto. A variável dependente será a medida do desempenho das redes na predição da rugosidade. A grandeza utilizada para medição do desempenho será o erro médio

absoluto do decil inferior das observações de uma topologia de rede, expresso em porcentagem. Grandezas baseadas no erro médio absoluto são as mais utilizadas na literatura (PONTES *ET AL.*, 2010a). Os modelos de RNAs e os experimentos planejados para otimização de seus fatores serão realizados com a ajuda de ferramentas computacionais.

Inicialmente aplicar-se-á o método proposto a resultados provenientes de modelos simulados e, em seguida, a resultados experimentais obtidos da literatura disponível sobre predição de rugosidade por outros métodos. Para cada etapa, os resultados obtidos serão analisados e interpretados. Os resultados das predições serão comparados aos resultados obtidos da literatura e também aos resultados de predições de redes geradas por outros métodos. A validação do método proposto será realizada por meio de experimentos de usinagem, que darão origem a conjuntos de treinamento e testes. As redes ótimas obtidas pela aplicação do método serão aplicadas à predição da rugosidade da peça para o conjunto de testes e avaliadas por seu desempenho nesta tarefa.

Desta forma, pode-se classificar o trabalho a ser realizado, quanto à sua natureza, como pesquisa aplicada (APPOLINÁRIO, 2006). Em relação ao seu objetivo o trabalho se classifica como sendo Normativo (GIL, 2009; BERTRAND & FRANSOO, 2002), com abordagem quantitativa (CRESWELL & PLANO CLARK, 2007). O método utilizado será o experimental (BRYMAN, 1989) e também Modelagem e Simulação (BERTRAND & FRANSOO, 2002).

1.7 ORGANIZAÇÃO DO TEXTO

Esta tese está organizada em seis capítulos. Neste primeiro capítulo o problema da modelagem da rugosidade da peça por meio de redes neurais foi contextualizado. Foram apresentadas as justificativas para a escolha do tema e sua relevância, os objetivos, as delimitações do trabalho, seu caráter inédito e a metodologia de pesquisa a ser seguida, finalizando com este parágrafo sobre a organização do texto.

O Capítulo 2 traz uma revisão da bibliografia sobre processos de usinagem, Redes Neurais Artificiais e Projeto de Experimentos, incluindo aí uma breve discussão

sobre Variações Evolucionárias em Operações (EVOP).

O Capítulo 3 apresenta um levantamento do estado da arte da pesquisa na área da modelagem da rugosidade da peça por meio Redes Neurais Artificiais. Neste capítulo são mostradas as limitações atuais da pesquisa no campo e são apontadas as oportunidades de melhoria que orientaram o desenvolvimento do método proposto neste trabalho.

O Capítulo 4 descreve o desenvolvimento do projeto de pesquisa, detalha a proposta do método de projeto de Redes Neurais otimizadas para a predição da rugosidade da peça e define o planejamento experimental a ser seguido para avaliação e validação do método.

No Capítulo 5 se apresenta a aplicação inicial do método proposto à predição da rugosidade a modelos sintéticos obtidos a partir de resultados de operações de torneamento. Em seguida é apresentada a aplicação do método à predição da rugosidade a partir dos dados utilizados em estudos selecionados da literatura que tratam da predição da rugosidade por redes neurais projetadas com o uso de outras abordagens. Apresenta-se finalmente a validação do método proposto por meio de sua aplicação para a predição da rugosidade a partir de dados coletados experimentalmente em operações de torneamento. Para cada aplicação do método, são apresentados os resultados obtidos para redes neurais de arquitetura Perceptron Multi-Camada (MLP) e Função de Base Radial (RBF). Em cada caso serão mostradas também comparações com os resultados obtidos pelos trabalhos da literatura de onde foram extraídos os dados, comparações com os resultados obtidos por uma ferramenta automática de geração de modelos de redes e também com modelos baseados nos métodos de mínimos quadrados. No Capítulo 6 apresentam-se as conclusões obtidas e também as sugestões para futuras pesquisas na área.

2 REVISÃO DA BIBLIOGRAFIA

2.1 USINAGEM E TORNEAMENTO

Segundo Shaw (2004) usinagem é um processo de fabricação por geração de superfície através da retirada de material (cavaco) conferindo dimensão e forma à peça. Uma definição bastante ampla é a de Ferraresi (1977), o qual afirma que “operações de usinagem são as que, ao conferir à peça a forma, ou as dimensões ou o acabamento, ou qualquer combinação destes três itens, produzem cavaco”. Segundo Diniz *et al.* (2008), a retirada do cavaco envolve o cisalhamento concentrado ao longo de um plano chamado de plano de cisalhamento. Na Figura 1 é ilustrado o conceito de retirada de material.

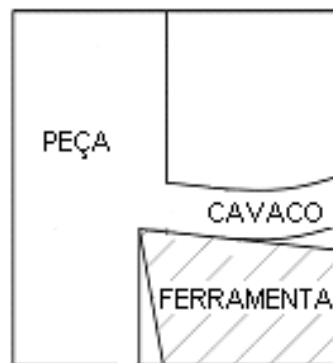


Figura 1 – Retirada de material em processo de torneamento. Adaptado de Trent (2000)

Segundo Trent (2000), o torneamento é a operação de usinagem mais comumente empregada em trabalhos de corte de metal. Ocorrem simultaneamente movimentos da peça (rotação) e da ferramenta (translação), como ilustrado na Figura 2. O torneamento é caracterizado pelos seus movimentos, que são: movimento de corte e movimento de avanço. É caracterizado também pelos seus parâmetros de corte: velocidade de corte (v_c), avanço (f) e profundidade de usinagem (ap).

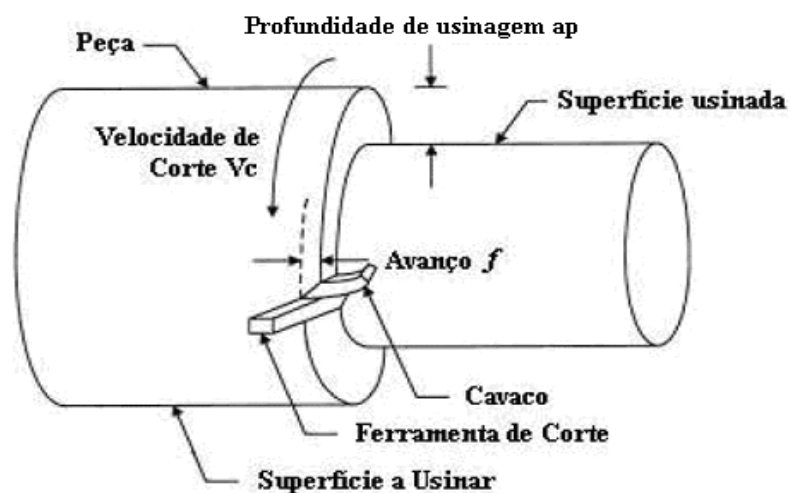


Figura 2 – Elementos em operação de torneamento. Adaptado de Amorim (2002)

O movimento de corte é o responsável pela remoção de material da peça. Ocorre entre a ferramenta e a peça. Sem a ocorrência simultânea do movimento de avanço, causa remoção de cavaco durante uma única rotação ou curso da ferramenta. O movimento de avanço é o que, juntamente com o movimento de corte, torna possível a remoção contínua ou repetida do cavaco, durante várias rotações ou cursos de ferramenta.

Segundo Diniz *et al.* (2008), a velocidade de corte é a velocidade tangencial instantânea resultante da rotação da peça em relação à ferramenta para a operação de torneamento, na qual os movimentos de avanço e corte ocorrem simultaneamente. Trent (2000) define a velocidade de corte como aquela com a qual a superfície a ser usinada passa pela aresta de corte da ferramenta. A velocidade de avanço é o produto da taxa de avanço pela rotação da ferramenta. A profundidade de usinagem ap é a profundidade ou largura de penetração da ferramenta em relação à peça, medida perpendicularmente ao plano de trabalho (DINIZ ET AL., 2008).

Cus & Zuperl (2006) afirmam que a operação de torneamento possui três objetivos conflitantes: taxa de produção, custo de operação e qualidade de usinagem. Segundo os autores, os três objetivos mencionados podem ser expressos como funções da velocidade de corte, do avanço e da profundidade de usinagem.

2.2 USINAGEM POR DESCARGAS ELÉTRICAS

A indústria de usinagem tem testemunhado a introdução de novos materiais com características térmicas, químicas e mecânicas superiores e dotados de maior resistência ao desgaste e à corrosão. Tais materiais trazem enorme benefício econômico em decorrência das melhorias de projeto e desempenho que propiciam aos produtos elaborados a partir deles. Por outro lado, processos tradicionais têm se mostrado incapazes de usinar tais materiais, ou, na melhor das hipóteses, incapazes de usiná-los de forma econômica (ARANTES, 2007). Segundo Haridy *et al.* (2011), processos de usinagem não-tradicionais, tais como a Usinagem por Descargas Elétricas, podem ser utilizados para manufaturar ligas de materiais de dureza extrema. O principal problema com estes processos, ainda segundo aqueles autores, é a determinação dos valores ótimos de usinagem.

Dentre os processos não convencionais de usinagem que têm sido aplicados para resolver este problema, o processo de Usinagem por Descargas Elétricas, ou EDM (do inglês *Electrical Discharge Machining*), ou ainda eletro-erosão, é um dos mais utilizados (MARKOPOULOS *ET AL.*, 2008; MANDAL *ET AL.*, 2007). Trata-se de um processo considerado especialmente útil para a usinagem de contornos complexos, peças diminutas e para a obtenção de elevada precisão para materiais que não são passíveis de usinagem por meio de processos tradicionais (HO & NEWMAN, 2003). De acordo com Mahapatra & Patnaik (2007), o processo de EDM é recomendado para a usinagem de materiais condutores de eletricidade quando uma alta precisão é requerida. EDM é largamente utilizada pelas indústrias aeroespacial, aeronáutica e automotiva, bem como para a manufatura de moldes e para a fabricação de componentes cirúrgicos (ASSARZADEH & GHOREISHI, 2008; MANDAL *ET AL.*, 2007).

O processo baseia-se na aplicação de uma diferença de potencial (em corrente contínua) entre duas placas condutoras de eletricidade, chamadas de eletrodo e peça, imersas em fluido dielétrico e separadas por pequena distância. Descargas elétricas ocorrem entre as placas abrindo um arco de plasma, o que faz com que o fluido passe ao estado gasoso e se torne eletrolítico. As descargas dão origem a um processo de

erosão na pequena região de descarga elétrica (ASSARZADEH & GHOREISHI, 2008).

As descargas duram alguns milionésimos de segundo, em um período chamado T_{on} . Segue-se um período no qual cessa a descarga e emergem os gases eletrolíticos, chamado de período T_{off} . O fluido dielétrico ocupa então o espaço do gás. Estando o fluido em temperatura inferior àquela do gás, o choque térmico causado origina uma micro-implosão que resulta na desagregação das partículas metálicas (KAMINSKI & CAPUANO, 1999). A sequência dos principais instantes de um ciclo EDM é mostrada na Figura 3.

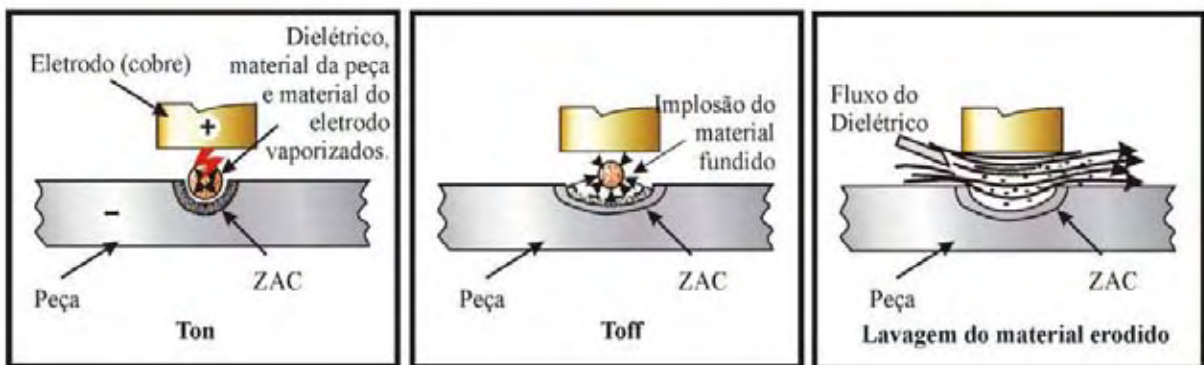


Figura 3 – Sequência dos três instantes principais de um ciclo EDM. Fonte: (ARANTES, 2007)

As principais grandezas envolvidas na operação de EDM são a tensão para abertura de arco; a tensão média de trabalho; a tensão de descarga; o período de descarga (T_{on}), o período de intervalo (T_{off}); o ciclo total de pulso; a frequência; o período de elevação da corrente; o período de queda da corrente; o período de retardo da ignição; a corrente de ignição e o pico de corrente. O rendimento da operação de EDM é caracterizado pela Taxa de Remoção de Material (TRM) da peça, é medido em volume de material removido da peça por unidade de tempo. Os fatores que determinam a TRM são os seguintes (ARANTES, 2007):

- Intensidade de corrente e tensão aplicadas (potência);
- Características físicas do material (ponto de fusão, condutibilidade

elétricas, microestrutura, etc.);

- Duração do pulso (T_{on}) e da pausa (T_{off}), ou seja, a frequência de aplicação da corrente;
- Geometria da ferramenta (eletrodo);
- Material da ferramenta;
- Tipo de lavagem;
- Tempo de pausa para limpeza.

2.3 RUGOSIDADE DA PEÇA

A qualidade da superfície da peça é um dos mais importantes requisitos do consumidor nos processos de usinagem. O principal indicador de tal requisito é a rugosidade (ÖZEL & KARPAT, 2005). Benardos & Vosniakos (2002) afirmam que a rugosidade da peça é uma medida da qualidade tecnológica de um produto e um fator que muito influencia em seu custo. O termo rugosidade da peça se refere a desvios da superfície nominal de terceira até sexta ordem. Segundo Benardos & Vosniakos (2003) desvios de terceira e quarta ordens são caracterizados por depressões, quebras e lapidações, que são influenciados pela forma e condições das arestas de corte, formação do cavaco e cinemática do processo. Desvios de quinta e sexta ordens, por sua vez, estão ligados à estrutura do material da peça usinada, a qual está ligada a mecanismos físico-químicos que agem sobre o grão e sobre o retículo cristalino (abrasão, difusão, oxidação, fadiga residual, entre outros). Desvios de diferentes ordens se sobrepõem para formar o perfil de rugosidade da peça (BENARDOS & VOSNIAKOS, 2002). O conceito é ilustrado na Figura 4.

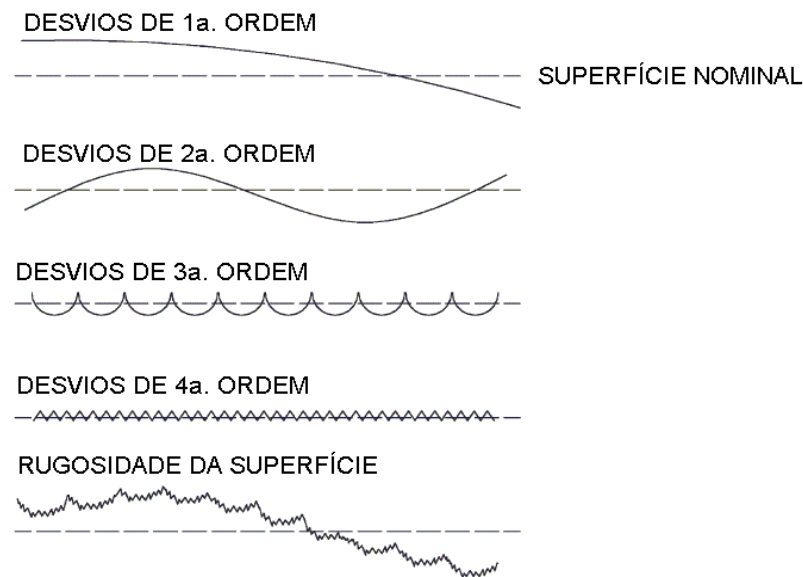


Figura 4 – Desvios no formato de superfícies. Adaptado de DIN 4670 (1982)

A rugosidade da peça é caracterizada pela forma como é feita sua medição. Os conceitos de rugosidade da peça são definidos pela norma ABNT NBR 6405/1985. No Brasil adota-se o sistema da “Linha Média” (M). Este é o sistema mais utilizado em todo o mundo. No sistema “M” todas as grandezas de medição são definidas a partir de uma linha paralela à direção geral do perfil, no comprimento de amostragem, seguindo uma linha que é a média entre os picos e reentrâncias de irregularidades.

O critério de medição da rugosidade adotado neste trabalho é o de profundidade de rugosidade. A grandeza medida é a rugosidade média de superfície (R_a). Esta é definida como o valor médio aritmético de todos os desvios do perfil em relação a uma linha média em um dado comprimento de amostragem. A rugosidade média pode ser expressa como na Equação 1 (ISO 4287/1, 2005).

$$R_a = \frac{1}{l_m} \int_0^{l_m} |y(x)| dx \quad (1)$$

em que R_a é a rugosidade média do perfil, l_m é o comprimento de amostragem do perfil empregado para medir-se a rugosidade e $|y(x)|$ corresponde aos valores absolutos dos desvios mensurados em relação à linha média. Para medições discretas, a rugosidade da peça é definida pela Equação 2 (ISO 4287/1, 2005).

$$R_a = \frac{1}{n} \sum_{i=1}^n |y_i| \quad (2)$$

na qual R_a é a rugosidade média de superfície, n é o número de amostras em um dado comprimento de amostragem L , e y_i são os valores absolutos das ordenadas do perfil efetivo (medido) em relação à linha média do comprimento de amostragem, conforme ilustrado na Figura 5.

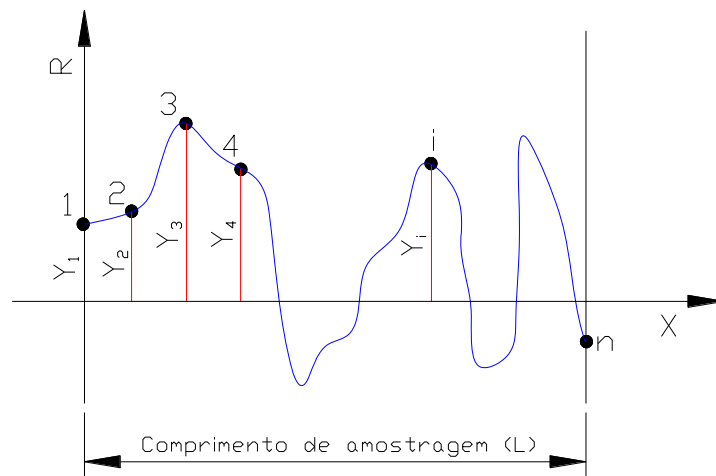


Figura 5 - Desvio médio aritmético R_a . Fonte: Agostinho *et al.* (1990)

Segundo o manual Sandvik (2005), o valor teórico da rugosidade máxima possível de ser obtido em um processo de torneamento é dado pela relação aproximada mostrada na Equação 3 (SANDVIK, 2005).

$$R_{\max} = \frac{f^2}{8r_e} \quad (3)$$

em que f denota o avanço e r_e o raio de ponta da ferramenta. Diniz *et al.* (2008) ressaltam, porém, que a rugosidade obtida na prática é normalmente superior a esse valor, devido a outros fatores tais como vibração, deformação e fluxo lateral do cavaco.

A rugosidade da peça define seu comportamento funcional (YANG *ET AL.*, 2010; CORREA *ET AL.*, 2009). Conforme apontado por Oktem *et al.* (2006) e por Chang & Lu (2006), ela desempenha um papel importante na determinação da qualidade de um produto usinado. A rugosidade é, portanto, um indicador, de acordo com Pal & Chakraborty (2005), do desempenho do processo de usinagem e deve ser

controlada dentro de limites de especificação adequados para operações de usinagem. Basheer *et al.* (2008) afirmam que as características das superfícies usinadas têm influência significativa sobre suas propriedades físicas. Novas aplicações em áreas da manufatura tão distintas quanto as indústrias automobilística, aeroespacial e biomédica, bem como a fabricação de matrizes e moldes, têm alimentado um rápido aumento na demanda por produtos com acabamento de alta qualidade (SHARMA *ET AL.*, 2008; ARANTES, 2007).

Os fenômenos que contribuem para a formação da rugosidade são complexos. Kilickap (2010) investigou a influência dos parâmetros de usinagem (velocidade de corte e avanço) e do ângulo de aproximação da ferramenta em processo de furação de materiais compósitos. O autor aplicou o método de Taguchi e utilizou a análise de variância (ANOVA) na busca por parâmetros ótimos de operação. Karayel (2009) declara que a rugosidade da peça depende de muitos fatores, incluindo parâmetros estruturais da máquina ferramenta, a geometria da ferramenta de corte, material das peças e da ferramenta de corte. Benardos & Vosniakos (2002) afirmam que os fatores que influenciam na qualidade da superfície são os seguintes: erros de montagem na torre e no insertos; variações periódicas na rigidez do conjunto peça-ferramenta de corte-máquina; desgaste da ferramenta de corte; formação da aresta postiça de corte e variações nas condições de corte. Ainda segundo Benardos & Vosniakos (2002), os principais fatores que influenciam a Rugosidade média da peça usinada são a profundidade de usinagem, o avanço, a velocidade de corte, o engajamento da ferramenta de corte, o desgaste desta, o uso de fluido de corte e os três componentes da força de corte. A importância dos parâmetros de corte e do desgaste da ferramenta na determinação da rugosidade pode ainda ser encontrada em diversos autores, como Shaw (2004), Özel & Karpuz (2005), Erzurumlu & Oktem (2007).

Segundo Diniz *et al.* (2008), operadores em geral não dispõem de meios de prever com exatidão a rugosidade de uma peça. Isso faz com que os critérios utilizados no chão de fábrica para troca de ferramentas sejam conservadores, levando a trocas de ferramentas bem antes do ponto em que as mesmas começam a gerar rugosidade acima da tolerância.

Benardos & Vosniakos (2003) listam quatro linhas de pesquisas dedicadas à

modelagem da rugosidade da peça: a modelagem segundo a teoria de usinagem; a investigação experimental em conjunto com técnicas de regressão; a abordagem do projeto de experimentos e técnicas que utilizam ferramentas oriundas do domínio da inteligência artificial.

Em relação aos modelos teóricos, Benardos & Vosniakos (2002) afirmam que muitos não são acurados o suficiente, e se aplicam a uma limitada gama de processos e condições de corte, o que se verifica em diversos trabalhos da literatura, como em Özel & Karpuz (2005), Oliveira (2003) e Amorim (2002). Um exemplo dos trabalhos envolvendo técnicas de projeto de experimento pode ser encontrado em Lela *et al.* (2009), o qual realizou experimentos com o objetivo de estabelecer um modelo entre a velocidade de corte, o avanço, a profundidade de usinagem e a Rugosidade Média da peça usinada em operações de fresamento. Dentre os exemplo do uso de técnicas de inteligência artificial encontra-se o trabalho de Kilickap *et al.* (2011), no qual se estuda a influência dos parâmetros de usinagem sobre a rugosidade da peça obtida em processo de furação de aço AISI 1045. Velocidade de corte, avanço e condições de lubrificação foram as variáveis de entrada. Um modelo matemático de predição da rugosidade foi desenvolvido com auxílio da metodologia de superfície de resposta. A equação derivada do método da superfície de resposta foi usada como função objetivo a ser minimizada por um algoritmo genético para a localização dos pontos de ocorrência de mínima rugosidade.

Ainda que não haja um modelo teórico único que torne possível a predição da rugosidade, a literatura aponta no sentido de que parâmetros de corte são determinantes na predição. Cus & Zuperl (2006) sugeriram modelos empíricos lineares e exponenciais para rugosidade da peça como funções da velocidade de corte (V_c), avanço (f) e profundidade de usinagem (ap). O modelo é mostrado na Equação 4.

$$R_a = c_0 (V_c^{c_1} f^{c_2} ap^{c_3}) \quad (4)$$

Na Equação 4, c_0 , c_1 , c_2 e c_3 são constantes específicas de uma dada combinação de material a ser usinado e ferramenta utilizada na usinagem.

2.4 REDES NEURAIAS ARTIFICIAIS

2.4.1 CONSIDERAÇÕES GERAIS

Uma rede neural “é um processador paralelo massivamente distribuído que possui uma propensão natural a armazenar conhecimento adquirido pela experiência e torná-lo disponível para uso”. O conhecimento da rede é adquirido por meio de um processo de aprendizado (HAYKIN, 2008). Para Bishop (2008), redes neurais artificiais são sistemas paralelos distribuídos compostos por unidades de processamento simples (chamadas de nós ou neurônios), que calculam determinadas funções matemáticas, em geral não-lineares, correspondendo a uma forma de computação não-algorítmica.

Um neurônio é uma unidade de processamento de informação composta de: um conjunto de sinapses, cada qual caracterizada por um valor de peso; um somador, responsável por somar os sinais de entrada devidamente multiplicados pelos pesos das sinapses; e uma função de ativação, que define e limita a saída do neurônio. Em uma rede neural, o conhecimento sobre um determinado problema é armazenado sob a forma dos pesos nas sinapses que interconectam os neurônios nas camadas de rede. (HAYKIN, 2007).

Redes neurais têm, em geral, três camadas distintas: a camada de entrada, uma ou mais camadas ocultas e uma camada de saída. Dados são apresentados aos neurônios da camada de entrada da rede, processados por uma ou mais camadas ocultas e os resultados são gerados na camada de saída da rede. Os neurônios da camada oculta e da camada de saída multiplicam cada um de seus sinais de entrada por pesos e utilizam uma função de ativação para determinar sua saída. Cada entrada de um neurônio tem um peso associado a si que determina sua intensidade.

A utilização de redes neurais na resolução de um dado problema envolve três etapas. Na fase de projeto, a arquitetura e os parâmetros da rede são estabelecidos. Na fase de aprendizado, exemplos são apresentados à rede para que esta, seguindo um algoritmo de treinamento, armazene conhecimento. Por último há uma fase de testes, em que o desempenho da rede em relação ao problema para o qual foi treinada é

avaliado de maneira independente.

A complexidade da curva modelada por uma rede neural é controlada de modo semi-paramétrico, através dos parâmetros com os quais a rede é configurada (MANUAL STATISTICA, 2005). Ao contrário de abordagens tradicionais de análise, que são baseadas em um número de idealizações e considerações teóricas, redes neurais artificiais possuem a capacidade de aprender por meio de exemplos e fazer interpolações (e em alguns casos extrapolações) do que aprenderam.

Redes neurais são aplicáveis a qualquer problema no qual uma relação entre variáveis de entrada (variáveis independentes) e saída (variáveis dependentes) exista e seja de natureza não linear. Exemplos disso são problemas de regressão múltipla, em que entradas e saída são variáveis numéricas contínuas, e também problemas de regressão logística, nos quais as entradas são variáveis numéricas contínuas e as saídas são variáveis discretas. Segundo Braga *et al.* (2000), redes neurais artificiais são capazes de atuar como mapeadores universais de funções multi-variáveis, com custo computacional que cresce apenas linearmente com o número de variáveis.

Conforme afirmam El-Mounayri *et al.* (2005) redes neurais possuem muitas propriedades que as tornam atraentes para a modelagem de problemas complexos de manufatura. Dentre tais propriedades, os autores destacam sua capacidade de agir como aproximadores universais de funções, sua resistência ao ruído ou à falta de dados, a capacidade de lidar simultaneamente com múltiplas variáveis não lineares e suas interações, e sua capacidade de generalização. Redes Neurais são especialmente úteis, segundo Ezugwu *et al.* (2005), para mapeamento de relações complexas cuja representação analítica seria difícil ou impraticável.

Redes neurais têm sido aplicadas em campos tão diversos quanto controle de processos (POTOCNIK & GRABEC, 2002), previsão de séries temporais (KHASHEI *ET AL.*, 2009; CHOUDURY *ET AL.*, 2008), reconhecimento de padrões, indústria química e biomédica, e em processos da manufatura como usinagem e soldagem, dentre outros. Quando aplicadas à previsão em processo de usinagem, as redes neurais também têm se mostrado úteis. Segundo Cus & Zuperl (2006), redes de arquitetura MLP provaram ser excelentes aproximadores universais para funções não lineares. É comum a utilização de redes implementadas por meio de software e inseridas como

sistema especialista dentro de sistemas industriais mais amplos. Diversos pacotes de software encontram-se disponíveis para projeto e geração de código para redes neurais.

2.4.2 HISTÓRICO DAS REDES NEURAIS ARTIFICIAIS

Redes neurais surgiram das pesquisas em inteligência artificial. Seu intento é o de imitar a tolerância às falhas e a capacidade de aprendizado de um sistema neural biológico, simulando a micro-estrutura e a dinâmica de aprendizado do cérebro humano. Segundo Hebb (1949), o aprendizado consiste principalmente em alterar a intensidade das sinapses, ou seja, das conexões existentes entre os neurônios que compõem o cérebro. Desta forma, partindo de unidades de processamento extremamente simples (os neurônios), o cérebro humano realiza tarefas muito complexas (KOVACS, 1996).

O primeiro modelo artificial de neurônio foi introduzido por McCulloch & Pitts (1943). Uma ilustração do conceito é mostrada na Figura 6.

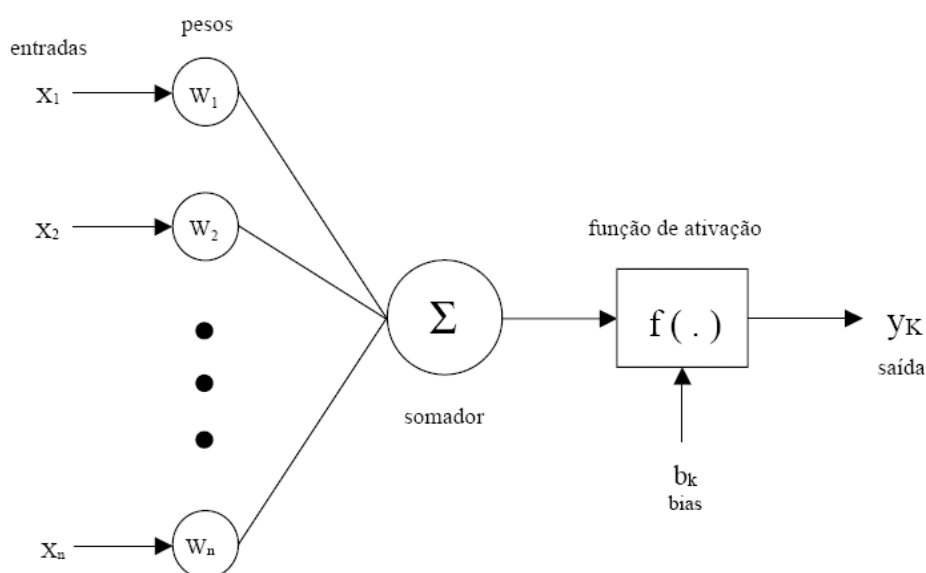


Figura 6 – O neurônio de McCulloch & Pitts (MCCULLOCH & PITTS, 1943).

O neurônio de McCulloch & Pitts é um modelo matemático com n entradas x_1, x_2, \dots, x_n e apenas um terminal de saída y_k . Cada terminal de entrada do neurônio traz um peso w_i associado a si. Os valores dos pesos podem ser positivos ou negativos. Eles

representam a intensidade de uma determinada entrada e estão relacionados à forma pela qual o neurônio deve considerar sinais que chegam por ela. A ativação do neurônio é obtida através da aplicação do resultado do produto escalar do vetor de entrada pelo vetor de pesos a um limiar de ativação b_k e a uma função de ativação $f(\cdot)$. O neurônio original de McCulloch & Pitts tem uma função linear binária, e dispara quando o resultado do produto escalar ultrapassa o limiar (BRAGA *ET AL.*, 2000).

Rosenblatt (1958), desenvolveu o Perceptron, uma rede formada por neurônios de McCulloch e Pitts com pesos de entrada ajustáveis. Tal rede podia ser treinada para classificar alguns tipos de padrões. O aprendizado do perceptron utilizava o conceito do aprendizado desenvolvido por Hebb. Alguns anos depois, porém, Minsky & Papert (1969) demonstraram matematicamente que o Perceptron era incapaz de solucionar problemas que não são linearmente separáveis. Seu trabalho levou a uma forte retração nas pesquisas em redes neurais durante os anos 70. Segundo Braga *et al.* (2000), a inexistência ou desconhecimento de algoritmos de treinamento para redes com uma ou mais camadas ocultas também contribuiu para a estagnação nas pesquisas com redes neurais durante o período.

O trabalho de Hopfield (1982), sobre memórias associativas, iniciou um novo período de investigação científica sobre o tema. Os trabalhos de Rumelhart *et al.* (1986) deram origem ao algoritmo de retro-propagação, removendo a limitação apontada anos antes por Minsky & Papert. Com efeito, enquanto uma rede sem uma camada oculta (como o Perceptron) pode resolver somente problemas linearmente separáveis, redes com camadas ocultas superam esta limitação.

Broomhead & Lowe (1988) apresentaram o procedimento para projeto de redes neurais acíclicas com a utilização de funções de base radial, que se tornaram conhecidas como redes RBF (do inglês *Radial Basis Function*). Tais redes provaram constituir-se em alternativas às redes Perceptron multicamadas (BRAGA *ET AL.*, 2000). Outras arquiteturas de redes neurais artificiais incluem redes neurais probabilísticas, redes GRNN (do inglês *Generalized Regression Neural Networks*), introduzidas por Specht (1991), redes SOFM (do inglês *Self-Organized Feature Map*) ou mapas de Kohonen (HAYKIN, 2008).

2.4.3 APRENDIZADO

A capacidade de aprender através de exemplos e de generalizar a informação aprendida é o atrativo principal da solução de problemas através de redes neurais artificiais, conforme afirma Braga *et al.* (2000). Por aprendizado entende-se o processo pelo qual os parâmetros de uma rede neural são ajustados através de uma forma continuada de estímulo pelo ambiente em que a rede está operando, sendo o tipo específico de aprendizado definido pela maneira particular como ocorrem os ajustes realizados nos parâmetros (HAYKIN, 2008). Nota-se dois paradigmas principais: aprendizado supervisionado e aprendizado não supervisionado.

No aprendizado supervisionado um conjunto de exemplos contendo valores de entradas e correspondentes saídas é reunido por um supervisor e apresentado à rede um determinado número de vezes. Os exemplos podem ter sua origem em dados históricos, ou modelos previamente existentes. Durante o treinamento, os parâmetros da rede são ajustados de acordo com algum algoritmo. O objetivo é o ajuste de pesos e limiares de forma a mapear-se o relacionamento entre as grandezas de entrada e saída (BRAGA *ET AL.*, 2000). Se o treinamento for bem realizado, a rede terá modelado a relação em suas sinapses, e poderá ser utilizada para realizar previsões para casos de entrada diferentes dos utilizados para treinamento, para os quais a saída não é conhecida de antemão pela rede. As já mencionadas redes MLP são exemplo do paradigma de aprendizado supervisionado.

No aprendizado não supervisionado, como o próprio nome sugere, não há um supervisor para acompanhar o processo de aprendizado. Apenas os padrões de entrada são apresentados à rede durante o treinamento. De acordo com Haykin (2008), na medida em que a rede identifica as regularidades estatísticas dos exemplos, ela forma representações internas que codificam características das entradas e permitem seu mapeamento nas saídas. Tal tipo de aprendizado pressupõe a existência de uma certa redundância dos dados de entrada, posto que, sem redundância, é impossível inferir padrões ou características dos mesmos (BRAGA *ET AL.*, 2000). Dentre os exemplos de aprendizado não supervisionado encontram-se as supra mencionadas redes SOFM e de Kohonen.

Dentro de cada paradigma, existem diferentes algoritmos de treinamento. Segundo Braga *et al.* (2000), algoritmo de treinamento é um conjunto de procedimentos bem definidos para adaptar os parâmetros de uma rede neural a fim de que ela possa aprender uma determinada função. O efeito dos algoritmos de treinamento é o ajuste dos valores dos pesos das sinapses e valores limiares de disparo, de modo a minimizar o erro de predição da rede. O erro de uma predição é definido como a diferença entre o valor real esperado para um determinado conjunto de entradas e o valor da predição da rede para o mesmo conjunto de entradas. Os erros de cada predição da rede para um dado conjunto podem ser reunidos e expressos por meio de alguma função para estimar-se o desempenho da rede neural. Dentre as funções que podem ser utilizadas para essa estimativa incluem-se o Erro Médio Absoluto, o Erro Médio Quadrático, dentre outros. O erro de predição da rede pode ser também representado como uma função de n_d dimensões, em que n_d é o número de parâmetros livres da rede. Examinado por esse ângulo, o objetivo de um algoritmo de treinamento é encontrar o ponto de mínimo absoluto desta função n -dimensional. Não há um método para projeto de redes neurais que garanta a obtenção de erro mínimo em tarefas de predição (HAYKIN, 2008).

O processo de treinamento pode ser visto como um problema de ajuste de curvas. Redes neurais podem ser compreendidas como uma técnica de interpolação não linear. A capacidade de generalização de uma rede neural é uma medida, portanto, da qualidade com que ela é capaz de interpolar entre pontos pertencentes ao conjunto utilizado para treinamento. Tal capacidade é afetada por três fatores: o tamanho e a qualidade do conjunto de treinamento, a arquitetura da rede e a complexidade da função que se deseja modelar (HAYKIN, 2008).

Um problema que pode ocorrer no emprego de redes é o chamado sobre-ajuste (ou *overfitting*, no original em inglês). Esse fenômeno se caracteriza por baixíssimo nível de erro na fase de treinamento e erros elevados na fase de testes, ou durante a operação. A rede memoriza as respostas para os casos de treinamento, mas é incapaz de generalizar quando apresentadas a novos casos. O sobre-ajuste pode ser causado pelo uso de uma arquitetura de rede mais complexa, ou seja, com mais neurônios e mais sinapses, do que o requerido pelo problema em questão. Deduz-se daí que a

determinação adequada do número de neurônios é essencial na busca por uma rede com capacidade de generalização adequada ao problema abordado. O conceito é exemplificado pela Figura 7, para uma rede hipotética cujo objetivo é modelar uma função de uma variável, em que os pontos correspondem a casos do conjunto de treinamento, a curva cheia representa a função modelada pela rede e a linha tracejada representa a função real que se desejava modelar.

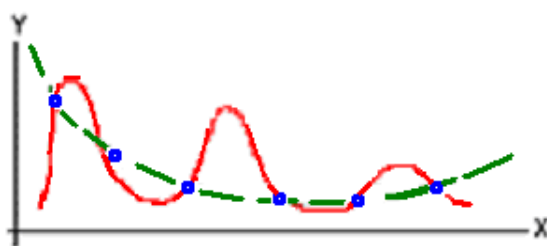


Figura 7 – Conceito de sobre-ajuste de redes. Adaptado do Manual Statistica (2005)

Outro problema que leva ao mesmo efeito é a ocorrência do sobre-aprendizado (ou *overlearning*, no original em inglês), que é característico das redes com aprendizado supervisionado. O sobre-aprendizado ocorre quando a rede passa por um número muito elevado de ciclos de treinamento, memoriza os exemplos apresentados e perde a capacidade de generalização. Existem algumas técnicas aplicáveis ao treinamento de redes MLP, dentre as quais a parada precoce (do inglês *early stopping*) cujo objetivo é interromper o treinamento em um ponto tal que evite a ocorrência do sobre-aprendizado (HAYKIN, 2008).

2.4.4 PROJETO DE REDES NEURAIAS

O desempenho de redes neurais é fortemente afetado pela configuração da rede. Ao contrário de técnicas de otimização linear, não existe um método sistemático para o projeto de uma rede neural ótima para um dado problema ou para um conjunto de dados, conforme reportado por diversos autores mesmo em publicações recentes (CORREA *ET AL.*, 2009; DHOKIA *ET AL.*, 2008; ZHONG *ET AL.*, 2006).

O projeto de uma rede neural envolve a escolha de quais variáveis serão

utilizadas como entradas, de quantas serão as variáveis de saída e quantos casos de teste serão utilizados para treinamento. Envolve também a escolha de uma determinada arquitetura de rede. Após a seleção da arquitetura, deve-se proceder ao projeto dos parâmetros da rede, que incluem a determinação do número de unidades na camada oculta, a escolha dos algoritmos de treinamento (que podem também ter seus próprios parâmetros de configuração), técnicas de poda e a estratégia de pré e pós-processamento dos dados. A etapa final do projeto envolve o treinamento da rede.

O conjunto de parâmetros de entrada na rede neural deve ser escolhido com base no conhecimento do projetista sobre a função que se deseja modelar. Em relação ao número de saídas, recomenda-se que seja unitário. Redes com múltiplas saídas podem sofrer o fenômeno do intercruzamento (ou *crosstalk*). Este é um fenômeno no qual a rede apresenta dificuldades no aprendizado, cuja origem é o fato dos neurônios tentarem modelar duas funções ao invés de uma (MANUAL STATISTICA, 2005).

Em relação ao número de casos de treinamento, não há um método seguro para a determinação da quantidade necessária para um treinamento bem sucedido de uma rede neural. A quantidade de casos depende da complexidade da função que se deseja modelar.

Existem diferentes arquiteturas de redes neurais, como as redes MLP (perceptron multi-camada), redes RBF (função de base radial), e outras. Redes neurais podem apresentar uma única camada oculta ou múltiplas camadas. As redes podem ser acíclicas (ou *feedforward*), nas quais saídas de um neurônio podem ser utilizadas como entrada somente pelas camadas seguintes; ou podem ser cíclicas (ou *feedback*), nas quais saídas de um neurônio podem ser usadas como entradas por neurônios da mesma camada ou até de camadas anteriores. Redes neurais podem ser conectadas parcial ou completamente. Neste último caso, que corresponde à grande maioria das aplicações, cada neurônio é conectado a todos os neurônios da camada seguinte da rede (BRAGA ET AL., 2000; HAYKIN, 2008).

A exatidão obtida pela rede depende do número de neurônios na camada oculta. Os neurônios localizados na camada oculta funcionam como detectores de características. O número de neurônios adequado para representar uma dada função depende de vários fatores como o número de exemplos utilizado, da complexidade da

função a ser mapeada e a distribuição estatística dos dados de treinamento. Um número excessivo de neurônios na camada oculta pode causar o já mencionado problema de sobre-ajuste. Um número insuficiente de neurônios também impossibilita que a rede capture a complexidade da função mapeada (BRAGA *ET AL.*, 2000).

Para uso em aplicações reais, redes neurais precisam ser dotadas de camadas para pré e pós-processamento de dados. Dados numéricos devem ser escalonados para processamento em uma rede neural. De acordo com Bishop (2007), isto se deve ao fato de que uma função de ativação tem uma faixa limitada de valores de saída. Em geral os valores de entrada são escalonados linearmente para utilização em aplicações reais. A rede neural é inserida logicamente entre duas camadas: uma camada de pré-processamento escalona os dados brutos antes que estes sejam enviados à camada de entrada da rede e uma camada de pós-processamento re-escalona os dados que emergem da camada de saída da rede para seu domínio original.

Para o pré-processamento de entradas numéricas, pode-se empregar um algoritmo como o Minimax (MANUAL STATISTICA, 2005). Este algoritmo apresenta dois parâmetros, um inferior e um superior. O algoritmo escalona os dados brutos de tal modo que o menor valor do conjunto de treinamento é convertido no valor do menor parâmetro, enquanto que o maior valor do conjunto de treinamento é convertido no valor do maior parâmetro. Os demais pontos do conjunto de treinamento são distribuídos linearmente entre aqueles dois extremos. Alternativamente pode-se empregar um algoritmo tal como o algoritmo Média/Desvio Padrão (MANUAL STATISTICA, 2005). Este algoritmo escalona dados linearmente, de tal modo que o valor médio do conjunto de treinamento é escalonado para um valor médio especificado, e o desvio padrão do conjunto de treinamento é escalonado para um desvio padrão especificado.

Para lidar com o problema do sobre-aprendizado, Haykin (2008) recomenda, no caso de redes MLP, a utilização de três conjuntos independentes de dados: um para treinamento, um para seleção e um para testes. O conjunto de seleção é mantido de fora do treinamento. Após cada época de treinamento, o conjunto de seleção é apresentado à rede. O treinamento é interrompido após a época em que a função erro da predição para o conjunto de seleção deixa de diminuir e começa a crescer, o que

indica o início do fenômeno do sobre-aprendizado. Os valores dos parâmetros livres recebem os valores da época anterior (correspondentes ao erro mínimo), e o treinamento se encerra. Os dados do conjunto de testes, por sua vez, não entram no treinamento nem na seleção da rede. Seu uso é recomendável como um teste final e independente para estimar a capacidade de generalização da rede.

No caso de redes RBF, em que a preocupação é o problema de sobre-ajuste, dois conjuntos independentes, um de treinamento e outro de testes, são suficientes. Os casos de teste não devem fazer parte do treinamento das redes. De acordo com Braga *et al.* (2000), outras técnicas para evitar a ocorrência do sobre-aprendizado incluem a ‘poda’ de redes (ou *pruning*), e os métodos de regularização.

É recomendável também que cada um dos conjuntos utilizados no treinamento seja, independentemente, representativo do modelo da função que se deseja modelar. Ainda que redes neurais sejam tolerantes ao ruído, recomenda-se ainda a retirada de *outliers* do conjunto de treinamento a ser utilizado (HAYKIN, 2008). Deve-se também considerar que dados de treinamento são relacionados a uma operação específica. Se o sistema ou a operação muda, os dados de treinamento não são mais válidos. Também é importante ter-se em conta que o conhecimento das redes neurais abrange somente o domínio dos casos de treinamento apresentados (MANUAL STATISTICA, 2005).

A ordem na qual os exemplos são apresentados à rede deve ser aleatória e variar a cada época de treinamento (HAYKIN, 1994). Recomenda-se também que, antes do início do treinamento, os pesos das conexões sinápticas e limiares de ativação obedeçam a uma distribuição uniforme dentro de uma faixa limitada de variação.

Sick (2002) recomenda conduzir mais de um experimento com cada configuração de rede. Com um único experimento, uma dada configuração pode apresentar um resultado correspondente a um mínimo local da função de erro, ou ter seu desempenho influenciado pela iniciação aleatória dos pesos de suas sinapses. Resultados assim obtidos podem superestimar ou subestimar o potencial da rede. Para uma avaliação estatística de medidas de desempenho de redes MLP, Haykin (2008) realizou 20 testes independentes com cada arquitetura de rede estudada.

2.4.5 REDES DE ARQUITETURA MLP (PERCEPTRON MULTI-CAMADA)

Redes com arquitetura perceptron multi-camada, ou MLP (do inglês Multi Layer Perceptron) constituem o modelo de redes neurais mais popular na literatura. Cada unidade desse tipo de rede efetua o produto escalar do seu vetor de entrada pelo vetor de pesos associados a cada entrada. O resultado é comparado a um valor de limiar. Em caso de disparo, o resultado do produto escalar é utilizado como variável independente de uma função de ativação, cuja variável dependente constituirá a saída do neurônio. A função sigmóide é a mais comumente utilizada como função de ativação, embora funções lineares, gaussianas e hiperbólicas também sejam utilizadas. Os neurônios são conectados em modo acíclico. Isso significa que neurônios de uma camada são ligados unicamente a neurônios da camada seguinte (HAYKIN, 2008). Na Figura 8 é ilustrado o conceito de rede MLP.

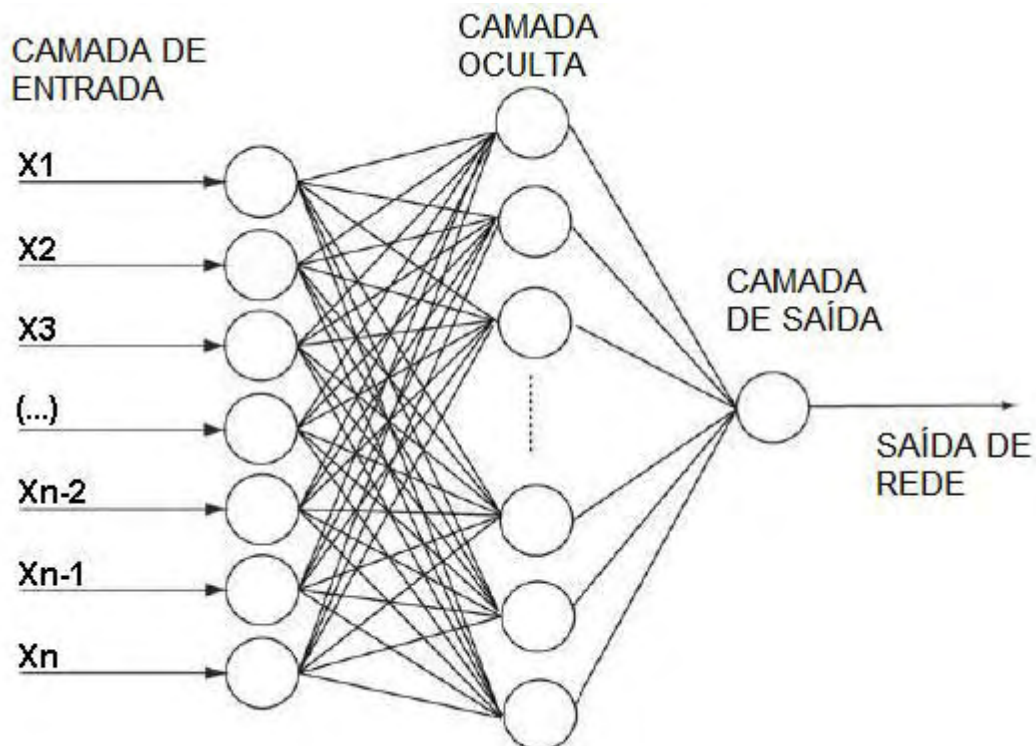


Figura 8 – Rede MLP. Adaptado de Özel & Karpaz (2005)

Segundo Bishop (2007), tal modelo de rede é capaz de modelar funções complexas. O número de camadas ocultas e o número de neurônios em cada uma

destas camadas determinam a complexidade da função que a rede é capaz de modelar. Os parâmetros livres do modelo da rede são os pesos das sinapses e o valor do limiar de disparo dos neurônios. Cybenko (1989), mostrou que uma rede MLP com uma camada oculta pode implementar qualquer função contínua. Duas camadas ocultas podem, segundo o mesmo autor, aproximar qualquer função.

2.4.6 TREINAMENTO DE REDES MLP

Dentre os algoritmos utilizados para treinamento de redes MLP o mais conhecido é o já mencionado algoritmo de retro-propagação (RUMELHART *ET AL.*, 1986), baseado na regra delta e no conceito de superfície de erro. O algoritmo procura minimizar o erro obtido ajustando pesos e limiares de ativação em busca dos pontos de mínimo da superfície de erro.

O treinamento através do algoritmo de retro-propagação repete-se por um certo número de vezes. As repetições são chamadas de épocas de treinamento. O algoritmo de retro-propagação pode ser executado baseado em exemplos ou em bateladas. No modo baseado em exemplos, o erro é retro-propagado, e os parâmetros livres são ajustados após a apresentação de cada exemplo. No modo baseado em bateladas, os parâmetros são ajustados somente após um ciclo completo de apresentação de exemplos, ciclo este que recebe o nome de época de treinamento (HAYKIN, 2008).

Os casos de treinamento são submetidos um de cada vez à rede. A saída da rede para cada caso é comparada ao valor de saída real do caso correspondente. Ao final de cada ciclo (após um exemplo ou após uma época, dependendo do modo de treinamento), os erros de predição são agrupados e uma função de erro é calculada. Este valor da função erro, juntamente com o valor calculado do gradiente da superfície de erro, é utilizado para ajustar os parâmetros livres da rede. Os valores dos pesos e limiares são ajustados no sentido do maior decremento do erro. Uma seqüência de ajustes nesta direção leva uma redução do erro para um mínimo (que pode ser um mínimo local ou absoluto). Ajustes mais amplos podem levar a uma convergência mais rápida, mas podem levar também à ultrapassagem do ponto de mínimo. Ajustes muito pequenos, por sua vez, requerem maior número de ciclos para a obtenção de um

mínimo. Na prática, os ajustes são proporcionais à derivada da função de erro, e também a um parâmetro denominado taxa de aprendizado. O algoritmo também prevê um termo denominado de momento, através do qual o ajuste em um sentido da função de erro é incrementado quando sucessivos ajustes ocorrem naquela direção (HAYKIN, 2008).

O treinamento ocorre, portanto, em duas etapas: a etapa direta (ou *forward*), na qual se define a saída da rede para um determinado padrão de entrada ou conjunto de padrões de entrada; e a etapa reversa (ou *backward*), que utiliza a diferença entre a saída de rede e a saída desejada, o erro da rede, para atualizar os parâmetros livres.

O número de vezes em que o processo é repetido pode ser definido de diferentes maneiras. O projetista da rede pode escolher por realizar um número arbitrário de épocas, pode realizar o treinamento até que o erro seja reduzido a um valor requerido ou pode realizar o treinamento até que o erro comece a crescer após atingir um mínimo.

Uma das desvantagens do algoritmo, segundo Braga *et al.* (2000), diz respeito à sua lentidão para a modelagem de funções complexas. Outra desvantagem, segundo Haykin (2008), é que o algoritmo pode estacionar em mínimos locais da superfície de erro. Diversas técnicas foram propostas para acelerar o algoritmo e reduzir a incidência de mínimos locais.

Outros algoritmos foram propostos para treinamento de redes MLP, como os algoritmos de gradiente conjugado, Levenberg-Marquardt, propagação rápida, quasi-Newton, dentre outros. Tais algoritmos procuram aumentar a eficiência na busca pelo mínimo global da função de erro, bem como acelerar o processo de treinamento das redes (MANUAL STATISTICA, 2005).

Uma abordagem comumente encontrada na literatura é o treinamento em duas fases, cada uma com um algoritmo distinto. O objetivo nestes casos é utilizar um algoritmo que convirja rapidamente à proximidade da região de mínimo e, a seguir, um segundo algoritmo com características de rapidez e robustez mais apropriadas à busca pela minimização do erro na região do mínimo.

2.4.7 REDES DE ARQUITETURA RBF (FUNÇÃO DE BASE RADIAL)

Funções de base radial representam uma classe de funções cujo valor diminui ou aumenta em função da distância a um ponto central. São utilizadas para tarefas de interpolação de conjuntos de pontos em espaços multi-dimensionais. Tal problema se caracteriza pelo mapeamento de um espaço vetorial \vec{x} de d dimensões em um espaço vetorial unidimensional t . O conjunto de dados consiste de N vetores de entrada \vec{x}^n , e seus correspondentes valores de \vec{t} . O objetivo é encontrar uma função $h(\vec{x})$, como na Equação 5 (BISHOP, 2007).

$$h(\vec{x}^n) = \vec{t}, \quad n = 1, \dots, N. \quad (5)$$

Segundo Bishop (2007), o uso de funções de base radial se mostrou apropriado na tarefa de interpolação, com o uso de conjuntos de N funções de base, uma para cada ponto, sendo as funções da forma da Equação 6 (BISHOP, 2007).

$$\phi(\|\vec{x} - \vec{x}^n\|) \quad (6)$$

em que ϕ é algum tipo de função não linear. O argumento da função ϕ é basicamente uma norma Euclidiana (uma distância) entre os vetores \vec{x} e \vec{x}^n . Uma das funções de base radial mais utilizada é a função gaussiana dada pela Equação 7 (BISHOP, 2007).

$$\phi_j(\mu) = -\exp\left(\frac{\|\vec{x} - \vec{\mu}_j\|^2}{2\sigma^2}\right) \quad (7)$$

em que \vec{x} corresponde a um dos vetores de entrada \vec{x}^n , com elementos x_i . $\vec{\mu}_j$ é um vetor que determina o hiper-centro para função ϕ_j , e possui elementos μ_{ij} , e σ representa um parâmetro que define a largura da mesma, isto é, a amplitude do decremento do valor de saída da função de ativação à medida que varia a distância ao

centro. O conceito de funções de base radial é ilustrado na Figura 9.

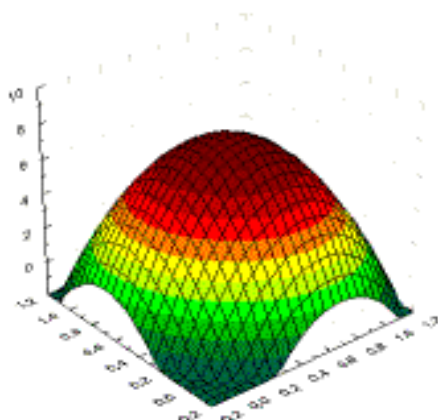


Figura 9 – Função de base radial. Fonte: Manual Statistica, 2005

Na Figura 9, a resposta de uma função de base radial é mostrada. Seu pico localiza-se no centro e decresce em função da distância a este. Redes RBF (do inglês *Radial Basis Function*) constituem uma arquitetura de redes neurais multi-camada na qual a função de ativação de cada unidade da camada oculta é uma função de base radial. O conceito de RBFs foi proposto pelos trabalhos de Broomhead & Lowe (1988). Segundo Braga *et al.* (2000), redes RBF utilizam como argumento da sua função de ativação a distância entre seus vetores de entrada e de pesos. Aos neurônios pertencentes à camada oculta de uma rede do tipo RBF dá-se o nome de unidades radiais.

Cada camada da rede desempenha uma função distinta. Uma rede RBF possui tipicamente uma única camada oculta formada por unidades radiais. Cada unidade radial modela uma função de ativação. A camada oculta agrupa os dados de entrada em clusters. Ela transforma um conjunto de padrões de entrada que não são linearmente separáveis em um conjunto de saídas linearmente separáveis. Uma camada oculta é suficiente para modelar qualquer função. As saídas da camada oculta são combinadas linearmente para formar a saída da rede (BISHOP, 2007).

A função radial em uso em redes RBF é, geralmente, uma função gaussiana, do tipo mostrado na Equação 7, na qual o vetor \vec{x} corresponde ao vetor de entrada da unidade radial e $\vec{\mu}_j$ representa o centro da função radial. A camada de saída das redes RBF contém geralmente neurônios que realizam o produto escalar das entradas e têm

função de ativação unitária. Em uma rede RBF com k unidades radiais na camada oculta e uma única saída, esta é dada pela Equação 8 (BISHOP, 1995).

$$y = \sum_{i=1}^k w_i \phi(\|\vec{x} - \vec{\mu}_i\|^2) + w_0 \quad (8)$$

em que \vec{x} e $\vec{\mu}_i$ são definidos como na Equação 7, k é o número de unidades radiais presentes na camada oculta, ϕ representa a função de ativação das unidades radiais, como, por exemplo, a função gaussiana representada pela Equação 7, w_i representa os pesos pelos quais a saída de uma unidade radial é multiplicada na camada de saída e w_0 um fator constante.

Na Figura 10 é ilustrado o conceito de rede RBF. Mostram-se as três camadas componentes: a camada de entrada, a camada oculta e a camada de saída. Na figura, $x_1, \dots, x_j, \dots, x_n$ correspondem aos elementos do vetor de entrada da rede. As unidades radiais são mostradas cada qual com sua função de ativação $\phi_1(\vec{x}), \dots, \phi_j(\vec{x}), \dots, \phi_m(\vec{x})$. Os valores $w_1, \dots, w_j, \dots, w_m$, representam os valores dos pesos das sinapses localizadas entre a camada oculta e a camada de saída. A saída da rede neural em função do vetor de entrada é representada por $y(\vec{x})$.

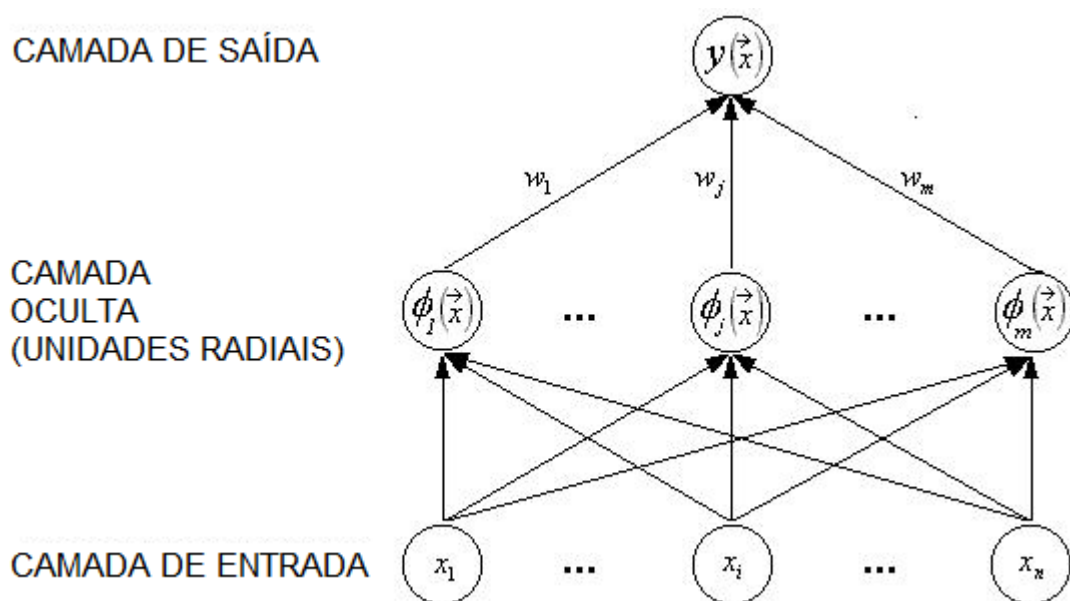


Figura 10 – Rede neural de arquitetura RBF. Adaptado de Bishop (2007)

As redes RBF têm algumas vantagens em relação a outras arquiteturas de rede, tais como sua estrutura mais simples, a rapidez do treinamento e a simplicidade da otimização da camada de saída. Às redes RBF se aplica também o teorema da aproximação universal de Cybenko (1989), que garante que uma única camada oculta é suficiente para aproximar uniformemente qualquer função contínua. Daí se infere que, para cada rede de arquitetura MLP aplicada a uma determinada tarefa, é possível encontrar-se uma rede RBF capaz de apresentar igual desempenho (HAYKIN, 2008).

2.4.8 PROJETO E TREINAMENTO DE REDES RBF

O projeto de redes RBF apresenta uma série de vantagens na comparação com o projeto de redes MLP. A primeira delas é que uma camada oculta é suficiente para modelar qualquer função. Isso facilita o projeto, pois elimina a etapa de determinação do número de camadas. Além disso, segundo Sick (2002), seu tempo de treinamento é muito curto se comparado ao das redes MLP aplicadas à mesma tarefa. Os trabalhos de Cus & Zuperl (2006) e de El-Mounayri *et al.* (2005) destacam a rapidez do treinamento de redes RBF. Afirmam os autores que a rapidez no treinamento de redes RBF constitui-se em importante diferencial quando da sua utilização em ambientes produtivos, cujos requisitos de tempo são severos, chegando mesmo a envolver o treinamento de redes em tempo de execução.

Outra vantagem é o fato de que a camada de saída é linear. Ela pode ser otimizada com a utilização de técnicas lineares, que não apresentam problemas de mínimos locais, como ocorre com redes MLP.

O projeto de uma rede RBF reduz-se, segundo Braga *et al.* (2000) e Bishop (2007), à decisão do número de unidades radiais e ao cálculo dos hiper-centros e dos valores dos parâmetros de largura das funções radiais em cada unidade.

Não se observa na literatura existente método para projeto de redes neurais RBF otimizadas. Segundo Braga *et al.* (2000), uma alternativa para o projeto de redes RBF é definir o número de neurônios como sendo igual ao número de casos de treinamento. Isso faz com que cada hiper-centro esteja situado sobre um vetor de entrada, e que a

rede RBF mapeie com exatidão vetores de entrada para as saídas desejadas. O autor ressalta, porém, que tal situação pode levar ao sobre-ajuste. Em relação aos parâmetros de largura das funções radiais, El-Mounayri *et al.* (2005) afirmam que inexistem uma regra geral para sua determinação.

O treinamento de redes RBF é considerado híbrido, pois é dividido em duas etapas. Na primeira etapa devem ser definidos os centros das funções radiais e seus parâmetros de largura. A definição dos parâmetros das funções radiais é feita por métodos não supervisionados de treinamento. A segunda etapa realiza o ajuste dos pesos das unidades da camada de saída. Como a saída da camada oculta é linearmente separável, os pesos podem ser otimizados por modelos lineares tais como a decomposição por valores singulares (BRAGA *ET AL.*, 2000).

Os dois métodos usualmente utilizados para determinação dos centros são a Sub-amostragem (ou *Sub-sampling*) e o algoritmo K-Médias (ou *K-Means*). No método de Sub-amostragem, pontos de treinamento escolhidos aleatoriamente são copiados como centros das unidades radiais (HAYKIN, 2008). Sendo aleatoriamente selecionados, eles representam estatisticamente a distribuição dos dados de treinamento.

Já no algoritmo K-Médias (BISHOP, 2007), há um esforço no sentido de definir um conjunto ótimo de pontos que sejam colocados como centróides dos conjuntos de dados de treinamento. Dadas K unidades radiais, o algoritmo ajusta as posições dos centros de modo que cada ponto pertença a um centro e esteja mais próximo dele do que de qualquer outro. O algoritmo busca também que cada um dos centros definidos realmente corresponda à posição central dos pontos a ele ligados. O algoritmo K-Médias divide o hiper-espaço dos padrões em K regiões. K vetores de entrada são aleatoriamente escolhidos para serem os centros dos grupos. Os demais vetores de entrada são atribuídos ao grupo de cujo centro estejam mais próximos. A seguir, os centros são recalculados para que correspondam à média dos vetores pertencentes ao grupo. Os vetores de entrada são rerepresentados à rede e associados ao centro mais próximo. Tal processo é repetido até que o algoritmo convirja, o que acontece quando os centros não têm suas posições alteradas em relação ao ciclo anterior (BRAGA *ET AL.*, 2000).

Após a determinação dos centros, deve-se determinar o parâmetro de largura da

função radial. Um valor muito pequeno fará com que a resposta da função de ativação decresça acentuadamente com a distância ao centro. A rede, neste caso, perde a capacidade de interpolar entre pontos conhecidos do conjunto de treinamento. Um valor alto, ao contrário, fará com que a resposta decresça suavemente, o que levaria à inexatidão na operação de interpolação. Isto pode ser visto como um tipo de manifestação do problema de sobre ou sub-ajuste (MANUAL STATISTICA, 2005).

Os algoritmos mais utilizados para a determinação do parâmetro de largura são o Isotrópico e o algoritmo dos K Mais Próximos (ou *K-Nearest*). No método Isotrópico, o valor do parâmetro de largura é selecionado heurísticamente a partir de um fator de escala para refletir o número de centros e o volume de hiper-espaço por eles ocupado (HAYKIN, 1994). Já no algoritmo dos K Mais Próximos o parâmetro de largura de cada unidade radial é calculado individualmente como a distância média às K unidades radiais vizinhas mais próximas (BISHOP, 1995). O valor de K deve ser definido pelo projetista da rede.

Ao final dessa primeira fase de treinamento, o passo seguinte corresponde à otimização da camada de saída, que pode ser feita empregando-se uma técnica padrão para otimização linear: a decomposição por valor singular. Este algoritmo usa a matriz pseudo-inversa para a determinação dos pesos da camada de saída. O algoritmo garante a determinação de um conjunto ótimo de pesos para a camada de saída que leve à minimização dos erros da rede para um dado conjunto de treinamento. Segundo o algoritmo, o vetor de pesos \vec{w} será dado pela Equação 9 (BISHOP, 2007).

$$\vec{w} = G_+ \vec{d} \quad (9)$$

na qual \vec{d} é o vetor que contém a resposta desejada para um dado vetor de saída proveniente da camada oculta, e na qual G_+ é a matriz pseudo-inversa definida pela Equação 10 (BISHOP, 2007).

$$G_+ = (G^T G)^{-1} G^T \quad (10)$$

em que G é uma matriz cujo elemento (i,j) corresponde à entrada da i -ésima

unidade de saída, quando o j -ésimo caso de treinamento é executado e G^T corresponde à transposta da matriz G .

2.5 METODOLOGIA DE PROJETO E ANÁLISE DE EXPERIMENTOS

2.5.1 CONSIDERAÇÕES GERAIS

A metodologia do Projeto de Experimentos (DOE) consiste em técnicas que objetivam planejar experimentos capazes de gerar dados apropriados para uma análise estatística que resulte em conclusões válidas e objetivas. Preconiza a variação simultânea dos fatores envolvidos em um experimento com o objetivo de construir modelos de previsão para as respostas de interesse. A cada diferente combinação de níveis dos fatores dá-se o nome de tratamento.

Montgomery & Runger (2007) citam a metodologia do DOE como contraponto para alguns tipos inadequados de experimentação tais como a estratégia de “Um Fator por Vez”, em que o nível de um único fator será trocado em cada ensaio. Os autores mostram que tal estratégia é deficiente e não garante que seja encontrada uma solução ótima para o modelo. É também ineficiente em determinar quais fatores afetam com maior intensidade a resposta de interesse e não permite a identificação e estimação de interações entre os níveis dos fatores testados.

Todo experimento, por simples que seja, envolve quatro atividades: conjectura, experimento, análise e conclusão. Uma maneira sistemática de avaliar a magnitude de várias fontes de variação que influenciam um processo deve iniciar com a identificação e seleção dos fatores que possam contribuir para a variação. Em seguida seleciona-se um modelo que inclua os fatores escolhidos e planejam-se experimentos eficientes para estimar seus efeitos. Devem-se utilizar resultados de experimentos exploratórios ou dados históricos para se avaliar o número de níveis adotados para cada fator. A ordem padrão utilizada pelos arranjos DOE gera condições experimentais balanceadas e ortogonais, de modo que os fatores investigados sejam experimentados uniformemente em cada um dos seus níveis (MONTGOMERY & RUNGER, 2007;

MONTGOMERY, 2001).

Dentre as características citadas por Montgomery & Runger (2007) como recomendáveis para um experimento planejado incluem-se as réplicas, a aleatorização e a blocagem. Réplicas são repetições de uma determinada corrida experimental. Elas permitem a obtenção de dados que possibilitam estimar a variabilidade da variável de resposta, a qual pode ser utilizada para avaliar a significância estatística do tratamento. A aleatorização é importante e desejável para garantir que a variação incontrolável seja diluída pelo arranjo proposto, reduzindo a chance de má interpretação dos resultados. A blocagem é uma técnica estatística que permite avaliar se a falta de homogeneidade interfere nos resultados. Deve ser empregada quando não for possível garantir a homogeneidade das condições ao longo da condução do experimento. Um bloco consiste de uma parte do experimento completo, que se presume ser mais uniforme do que o todo (PAIVA, 2006).

Os experimentos devem ser realizados de acordo com os arranjos experimentais do DOE, seguindo as premissas citadas anteriormente. Realizados os experimentos, estimam-se os efeitos dos fatores incluídos no modelo utilizando métodos estatísticos adequados. No final, os resultados devem ser interpretados e discutidos, e melhorias devem ser recomendadas.

2.5.2 FATORIAIS COMPLETOS

Um planejamento fatorial é completo quando todas as possíveis combinações entre os fatores envolvidos são experimentadas (BOX *ET AL.*, 1978). Em um experimento planejado segundo a técnica do fatorial completo cada fator é testado e re-testado igual número de vezes em cada um dos seus níveis, os quais são escolhidos de maneira apropriada dentro de sua amplitude normal de variação. Arranjos fatoriais completos cobrem todo o espaço experimental e correspondem ao arranjo experimental mais comum (PAIVA, 2006).

O número de experimentos é igual ao número de níveis experimentais elevado ao número de fatores. Para experimentos fatoriais em dois níveis, o número de corridas experimentais N para se avaliar os k fatores é dado por $N = 2^k$. Logo, enquanto o

número de fatores cresce linearmente, o número de corridas cresce exponencialmente.

Arranjos fatoriais completos podem ser gerados para qualquer quantidade de fatores e os níveis de cada fator se alternam nas colunas, segundo uma mesma ordem. Para a primeira coluna, os níveis se alteram a cada 2^0 experimento; para a segunda coluna, os níveis se alteram a cada 2^1 experimentos; para a terceira coluna, os níveis se alteram a cada 2^2 experimentos e assim por diante. O procedimento se repete para tantas colunas quantas forem os k fatores, até a k -ésima coluna.

2.5.3 ANÁLISE DE VARIÂNCIA (ANOVA)

Na metodologia de projeto de experimentos, o teste empregado para avaliar a significância da mudança de níveis de um fator ou de uma interação é um teste de hipótese para média. No caso dos fatoriais completos, este teste é a Análise de Variância, ou ANOVA (MONTGOMERY & RUNGER, 2007).

Quando dois níveis de um fator geram respostas médias estatisticamente iguais, assume-se que o fator não influencia a resposta de interesse. Quando, ao contrário, se detecta uma diferença significativa, o fator é importante. Para um fatorial completo com 2 fatores, A e B , em dois níveis (+1, -1), pode-se escrever o modelo mostrado na Equação 11.

$$Y_{ijk} = \mu + A_i + B_j + AB_{ij} + \varepsilon_{ijk} \quad (11)$$

em que i é o número de níveis de A , j é o número de níveis de B e k , o número de réplicas.

O termo Y_{ijk} é a (ijk) -ésima observação obtida no experimento, μ é a média geral, A_i é efeito do i -ésimo tratamento do Fator A , B_j o efeito do j -ésimo tratamento do Fator B , AB_{ij} , o efeito da ij -ésima interação AB entre os fatores e ε_{ijk} , um componente do erro aleatório.

A estatística de teste utilizada na ANOVA para verificar a igualdade entre as médias baseia-se na relação existente entre a variação “entre” tratamentos pela

variação “dentro” de um tratamento (MONTGOMERY & RUNGER, 2007). A variação “Dentro” provém dos repetidos ensaios executados com um dado tratamento. Calculando-se a média das réplicas, obtém-se a média dentro do tratamento. Quanto mais as réplicas diferirem desta média, maior será a variação dentro deste tratamento. Dividindo-se a variação “Entre” pela variação “Dentro” obtém-se a estatística de teste F , a qual, quando comparada com um valor crítico de F , definido segundo o nível de significância e o número de graus de liberdade da variação “Dentro”, possibilita a aceitação ou rejeição da hipótese nula de igualdade entre as médias dos níveis dos fatores.

Sempre que se faz uma inferência sobre uma população a partir de dados amostrais corre-se o risco de cometer algum tipo de erro. Hipótese é tão somente uma afirmação que se deseja testar, tendo como fonte de provas os dados provenientes da amostra. Existem dois tipos de erro que podem ser cometidos em um teste de hipóteses. O primeiro é o Erro Tipo I, que consiste em se rejeitar a hipótese nula quando ela é verdadeira. A probabilidade de se cometer este tipo de erro é α , denominado nível de significância do teste. O segundo é o Erro Tipo II, que consiste em se aceitar a hipótese nula quando ela for falsa. A probabilidade de se cometer este tipo de erro é β . Erros Tipo I e Erros Tipo II são mutuamente exclusivos.

A probabilidade de se aceitar a hipótese nula quando ela é verdadeira é $1-\alpha$, e é denominado nível de confiança do teste. A probabilidade de se rejeitar a hipótese nula quando ela for falsa é $1-\beta$, e é chamada de Poder do Teste. Com a amostra se calcula a estatística de teste, que quando comparada ao valor crítico determinado por α , conduz à aceitação ou rejeição da hipótese nula. Ao conjunto de todos os valores da estatística de teste que levam à rejeição da hipótese nula dá-se o nome de Região Crítica (MONTGOMERY, 2001).

Para se avaliar a significância do efeito dos níveis de um tratamento A , bem como de um tratamento B , faz-se necessário testar a hipótese da igualdade entre as respostas médias obtidas com os níveis dos dois fatores. Também é importante determinar se existe interação entre os dois tratamentos. As hipóteses necessárias neste caso estão representadas no Quadro 1.

Quadro 1 – Hipóteses presentes em uma ANOVA para dois fatores. Fonte: Adaptado de Montgomery & Runger (2007).

Hipóteses	Efeitos Principais	Interações
Hipótese Nula (H_0)	$H_0: A_1 = A_2$	$H_0: AB_{ij} = 0$
Hipótese Alternativa (H_1)	$H_1: A_1 \neq A_2$	$H_1: AB_{ij} \neq 0$

Para os dois tratamentos, a aceitação de H_0 significa que os efeitos na resposta obtidos com os dois níveis dos fatores A e B são iguais; a rejeição de H_0 , por sua vez, implica na aceitação da hipótese alternativa, de que os tratamentos são diferentes. Para interações, a aceitação de H_0 indica que a interação não é significativa.

Para se calcular os elementos de uma ANOVA para dois fatores, sendo o Fator A com a níveis, e o Fator B , com b níveis, para n réplicas, podem ser empregadas as fórmulas do Quadro 2.

Quadro 2 – Fórmulas para análise de variância para dois fatores em dois níveis. Fonte: Adaptado de Montgomery & Runger (2007)

Soma de cada linha (Y_{sl})	$\sum_{j=1}^b \sum_{k=1}^n y_{ijk}$	Média de cada linha (\bar{Y}_l)	$\frac{1}{bn} \sum_{j=1}^b \sum_{k=1}^n y_{ijk}$
Soma de cada coluna (Y_{sc})	$\sum_{i=1}^a \sum_{k=1}^n y_{ijk}$	Média de cada coluna (\bar{Y}_c)	$\frac{1}{an} \sum_{i=1}^a \sum_{k=1}^n y_{ijk}$
Soma dentro de cada célula (Y_{scel})	$\sum_{k=1}^n y_{ijk}$	Média dentro de cada célula (\bar{Y}_{cel})	$\frac{1}{n} \sum_{k=1}^n y_{ijk}$
Soma Geral (Y_{st})	$\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n y_{ijk}$	Média Geral (\bar{Y}_t)	$\frac{1}{abn} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n y_{ijk}$

Com estes somatórios, pode-se escrever a soma de quadrados total (SS_T) como mostrado na Equação 12.

$$\begin{aligned} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y}_t)^2 &= \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n [(\bar{y}_l - \bar{y}_t) + (\bar{y}_c - \bar{y}_t) + (\bar{y}_{cel} - \bar{y}_l - \bar{y}_c - \bar{y}_t) + (y_{ijk} - \bar{y}_t)]^2 \\ &= bn \sum_{i=1}^a (\bar{y}_l - \bar{y}_t)^2 + an \sum_{j=1}^b (\bar{y}_c - \bar{y}_t)^2 + n \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{cel} - \bar{y}_l - \bar{y}_c - \bar{y}_t) + \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y}_c)^2 \quad (12) \end{aligned}$$

Dessa maneira, a Soma de Quadrados Total do arranjo (SS_T) pode ser dividida em somas de quadrados devidos aos tratamentos de linha (Fator A) SS_A , aos tratamentos de coluna (Fator B) SS_B ; a soma de quadrados devida à interação entre A e B , SS_{AB} e a soma de quadrados devida ao erro experimental SS_E . Assim pode-se escrever a soma total dos quadrados como na Equação 13.

$$SS_T = SS_A + SS_B + SS_{AB} + SS_E \quad (13)$$

As equações das somas de quadrados podem ser ainda escritas como nas Equações 14, 15 e 16.

$$SS_T = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n y_{ijk}^2 - \frac{y_{st}^2}{abn} \quad (14)$$

$$SS_A = \frac{1}{bn} \sum_{i=1}^a y_{sl}^2 - \frac{y_{st}^2}{abn} \quad (15)$$

$$SS_B = \frac{1}{an} \sum_{j=1}^b y_{sc}^2 - \frac{y_{st}^2}{abn} \quad (16)$$

Para uma análise de variância com dois fatores, as somas dos quadrados de cada fator e da interação, junto dos respectivos números de graus de liberdade, médias quadráticas e estatísticas de teste são reunidos no Quadro 3.

Quadro 3 – Fórmulas para análise de variância em dois níveis. Fonte: Adaptado de Montgomery & Runger (2007)

Fonte de Variação	Soma de Quadrados	Graus de Liberdade	Média Quadrática	F _o
Fator A	SS_A	$a-1$	$MS_A = \frac{SS_A}{a-1}$	$F_0 = \frac{MS_A}{MS_E}$
Fator B	SS_B	$b-1$	$MS_B = \frac{SS_B}{b-1}$	$F_0 = \frac{MS_B}{MS_E}$
Interação AB	SS_{AB}	$(a-1)(b-1)$	$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)}$	$F_0 = \frac{MS_{AB}}{MS_E}$
Erro	SS_E	$ab(n-1)$	$MS_E = \frac{SS_E}{ab(n-1)}$	
Total	SS_T	$abn-1$		

Para se obter uma simplificação da soma de quadrados referentes à interação calcula-se primeiramente a soma de quadrados parciais SS_p , tal qual na Equação 17.

$$SS_p = \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^b y_{scel}^2 - \frac{y_{st}^2}{abn} \quad (17)$$

Subtrai-se SS_A e SS_B de SS_p , conforme mostrado pela Equação 18 e assim encontra-se a parcela de variação devido ao erro, conforme a Equação 19.

$$SS_{AB} = SS_p - SS_A + SS_B \quad (18)$$

$$SS_E = SS_T - SS_{AB} - SS_A - SS_B \quad (19)$$

O quadrado médio é o quociente entre a soma de quadrados e os graus de liberdade associados a cada fonte de variação. F_o representa a estatística de teste para a análise de dois grupos de dados.

2.5.4 TESTES PARA SIGNIFICÂNCIA

Este teste é realizado como um procedimento de ANOVA. Calculando-se a razão entre a média quadrática dos termos de regressão e a média quadrática do erro, encontra-se a estatística F . Comparando-se esta estatística com o valor crítico de F obtido a partir do nível de significância desejado, conclui-se sobre a significância do modelo. Se F for maior que $F_{crítico}$ (o que equivale a um valor P menor que o nível de significância), então o modelo é adequado (MONTGOMERY, 2001).

O teste individual de significância de cada coeficiente pode conduzir à otimização do modelo através da eliminação ou da adição de termos. Se o valor P do teste individual para os termos for inferior ao nível de significância, então, o termo é adequado ao modelo e deve, portanto, ser mantido. Ao contrário, o termo deve ser excluído se tal procedimento conduzir a um aumento do coeficiente de determinação R^2 conjuntamente com a diminuição do erro residual S e o valor de P referente à falta de ajuste do modelo for superior ao nível de significância. Além disso, a retirada de qualquer termo deve obedecer ao princípio da Hierarquia (MONTGOMERY, 2001). Este princípio postula que quando um termo de ordem alta é mantido no modelo, o de ordem baixa que o compõe também deve ser conservado. Por exemplo, se a interação AB é significativa, mas o efeito principal A não o é, então deve-se conservar ambos.

2.5.5 MODELOS DE REGRESSÃO LINEAR

O método dos mínimos quadrados ordinários (ou OLS, do inglês *Ordinary Least Squares*) é o algoritmo tipicamente usado para estimar os coeficientes de um modelo de regressão linear múltipla, fatoriais completos, fracionados ou superfícies de resposta.

Sejam consideradas as $n > k$ observações da variável de resposta disponíveis, assim como y_1, y_2, \dots, y_n . Assumindo-se que o valor esperado do erro seja zero, pode-se escrever como na Equação 20.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \varepsilon_i \quad (20)$$

O método dos mínimos quadrados baseia-se na escolha de valores para β na Equação 20, de modo que a soma dos quadrados dos erros seja minimizada. A função de mínimos quadrados pode ser escrita como na Equação 21.

$$L = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij} \right)^2 \quad (21)$$

A função L deve ser minimizada em função de $\beta_0, \beta_1, \dots, \beta_k$. Assim, igualando a zero as derivadas parciais da função L em relação a $\beta_0, \beta_1, \dots, \beta_k$ tem-se a Equação 22.

$$\frac{\partial L}{\partial \beta_0} = \frac{\partial L}{\partial \beta_j} = -2 \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij} \right) = 0 \quad (22)$$

Simplificando-se e generalizando-se a Equação 22, obtém-se o mostrado na Equação 23.

$$\hat{\beta}_0 \sum_{i=1}^n x_{ik} + \hat{\beta}_1 \sum_{i=1}^n x_{i1} x_{ik} + \hat{\beta}_2 \sum_{i=1}^n x_{i2} x_{ik} + \dots + \hat{\beta}_k \sum_{i=1}^n x_{ik}^2 = \sum_{i=1}^n x_{ik} y_i \quad (23)$$

Estas equações são chamadas de Equações normais de mínimos quadrados. Nota-se que existem $p = k+1$ equações, uma para cada coeficiente de regressão desconhecido. Em notação matricial, o sistema de Equações 23 pode ser escrito como mostrado nas Equações 24 e 25.

$$Y = \beta X + \varepsilon \quad (24)$$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \vdots \\ \varepsilon_k \end{bmatrix} \quad (25)$$

É possível notar que L pode ser expresso como na Equação 26. Como os estimadores de mínimos quadrados devem satisfazer à Equação 27, uma simplificação da mesma conduz à Equação 28.

$$L = y^T y - \beta^T X^T y - y^T X \beta + \beta^T X^T X \beta = y^T y - 2\beta^T X^T y + \beta^T X^T X \beta \quad (26)$$

$$\frac{\partial L}{\partial \beta} = -2X^T y + 2X^T X \hat{\beta} = 0 \quad (27)$$

$$X^T X \hat{\beta} = X^T y \quad (28)$$

Multiplicando-se ambos os lados da Equação 28 pela inversa da matriz $X^T X$, tem-se a Equação 29. Esta corresponde à expressão matricial das equações normais de mínimos quadrados, descritos anteriormente (MONTGOMERY, 2001).

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (29)$$

Modelos de regressão linear múltipla são freqüentemente utilizados como aproximações de funções. Isto é, como a verdadeira relação entre x e y não é conhecida, aproxima-se tal relacionamento por um modelo de regressão, que é relativamente adequado dentro de certa faixa de variação das variáveis independentes.

Em função da notação matricial, também se pode determinar uma fórmula para a estimação da variância, bem como estabelecer fórmulas para as somas dos quadrados.

Considere o modelo quadrático da Equação 30. Como se sabe, \hat{y}_i representa um valor ajustado, obtido a partir do modelo de regressão. A diferença entre o valor real

observado y_i e \hat{y}_i , denomina-se resíduo, e pode ser representado pela Equação 31.

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \beta_{ii} x_i^2 + \sum_{i < j} \beta_{ij} x_i x_j + \varepsilon \quad (30)$$

$$e = y - \hat{y} \quad (31)$$

Considerando-se a soma de quadrados para os resíduos, chega-se à Equação 32.

$$SS_E = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (e_i)^2 = e^T e \quad (32)$$

Substituindo-se $e = y - \hat{y} = y - X\hat{\beta}$, chega-se à Equação 33.

$$\begin{aligned} SS_E &= (y - X\hat{\beta})^T (y - X\hat{\beta}) = y^T y - \beta^T X^T y - y^T X\hat{\beta} + \hat{\beta}^T X^T X\hat{\beta} \\ SS_E &= y^T y - 2\beta^T X^T y + \hat{\beta}^T X^T X\hat{\beta} \end{aligned} \quad (33)$$

Como: $X^T X\hat{\beta} = X^T y$, a Equação 33 toma a forma da Equação 34.

$$SS_E = y^T y - \hat{\beta}^T X^T y \quad (34)$$

A Equação 34 é chamada soma de quadrados residual. Derivando-se uma forma matricial para a soma de quadrados total obtém-se o que é mostrado na Equação 35.

$$SS_T = \sum_{i=1}^n y_i^2 - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n} = y^T y - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n} \quad (35)$$

Para a soma de quadrados do modelo, tem-se o mostrado na Equação 36.

$$SS_m = SS_t - SS_e = y^T y - \frac{1}{n} \left(\sum_{i=1}^n y_i \right)^2 - \left(y^T y - \hat{\beta} X^T y \right) = \hat{\beta} X^T y - \frac{1}{n} \left(\sum_{i=1}^n y_i \right)^2 \quad (36)$$

O teste para a significância da regressão determina se existe uma relação linear entre a variável de resposta y e um subconjunto de regressores. As hipóteses apropriadas neste caso são as mostradas na Equação 37.

$$\begin{aligned} H_0 &= \beta_1 = \beta_2 = \beta_3 = \dots = \beta_k = 0 \\ H_1 &= \beta_j \neq 0 \end{aligned} \quad (37)$$

A rejeição da hipótese nula implica que, no mínimo, um dos regressores contribui significativamente para a explicação da variação de Y . Uma vez que raramente um modelo explica 100% da variação da resposta modelada, é provável que exista uma diferença entre o valor observado e o valor modelado. Se a hipótese nula, descrita na Equação 37 for verdadeira, então SQ_R/σ^2 será uma variável aleatória χ^2 , com k graus de liberdade. O número de graus de liberdade para esta variável é igual ao número de regressores presentes no modelo. Por sua vez, SQ_E/σ^2 segue o mesmo condicionamento, com $n-p$ graus de liberdade. As duas expressões anteriores são independentes. Logo, dividindo-se as duas variações encontra-se uma distribuição F . Portanto, a estatística de teste para a significância do modelo pode ser a descrita pela Equação 38.

$$F_0 = \frac{SQ_M/k}{SQ_E/(n-p)} = \frac{MQ_M}{MQ_E} \quad (38)$$

Deve-se rejeitar a hipótese nula se o valor calculado na Equação 38 for maior do que o valor de $F_{(\alpha, k, n-p)}$.

As hipóteses para testar a significância de qualquer coeficiente de regressão são as expressas pela Equação 39.

$$\begin{aligned} H_0 : \beta_j &= 0 \\ H_1 : \beta_j &\neq 0 \end{aligned} \quad (39)$$

Se a hipótese nula não for rejeitada, então o coeficiente poderá ser retirado do modelo. A estatística de teste utilizada para esta finalidade é dada pela Equação 40.

$$T_0 = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 \cdot C_{jj}}} \quad (40)$$

Na Equação 40 C_{jj} é o elemento da diagonal de $(X^T X)^{-1}$ correspondente a $\hat{\beta}_j$. Observa-se que o denominador da Equação 40 é o erro-padrão do coeficiente $\hat{\beta}_j$. Deve-se rejeitar a hipótese nula se $T_0 > t_{(\alpha/2, n-p)}$. Um intervalo de confiança para os coeficientes de regressão é dado pela Equação 41.

$$\hat{\beta}_j - t_{\alpha/2, n-p} \sqrt{\hat{\sigma}^2 C_{jj}} \leq \beta_j \leq \hat{\beta}_j + t_{\alpha/2, n-p} \sqrt{\hat{\sigma}^2 C_{jj}} \quad (41)$$

Do mesmo modo, pode-se também construir um intervalo de confiança para a resposta média de um determinado ponto como $x_0 = [1 \ x_{01} \ x_{02} \ \dots \ x_{0k}]^T$. A resposta média neste ponto é estimada por $\hat{\mu}_{Y|x_0} = x_0^T \hat{\beta}$, com variância $V(\hat{\mu}_{Y|x_0}) = \sigma^2 x_0^T (X^T X)^{-1} x_0$. Logo, um intervalo de confiança de $100(1-\alpha)\%$ é dado pela Equação 42.

$$\hat{\mu}_{Y|x_0} - t_{\alpha/2, n-p} \sqrt{\sigma^2 x_0^T (X^T X)^{-1} x_0} \leq \mu_{Y|x_0} \leq \hat{\mu}_{Y|x_0} + t_{\alpha/2, n-p} \sqrt{\sigma^2 x_0^T (X^T X)^{-1} x_0} \quad (42)$$

2.5.6 MEDIDAS DE ADEQUAÇÃO DOS MODELOS DE REGRESSÃO

A medida mais comum de adequação de um modelo é o coeficiente de determinação R^2 . Este termo representa o percentual de variação na resposta que é explicada pelo modelo construído. Associado a este coeficiente encontra-se o R^2 ajustado ($R^2_{adj.}$), que considera o fato de que R^2 tende a superestimar a quantidade atual de variação contabilizada para a população. Também é fato que a inclusão de muitos termos no modelo de regressão aumenta substancialmente o valor de R^2 . Se o modelo recebeu fatores adicionais desnecessários, haverá um incremento em R^2 , sem que haja, necessariamente, melhoria de informação na resposta. É por este motivo que o valor de R^2 ajustado é mais apropriado para se comparar modelos com diferentes quantidades de termos.

O valor de R^2 pode ser calculado aplicando-se a Equação 43.

$$R^2 = \frac{SQ_M}{SQ_T} = 1 - \frac{SQ_E}{SQ_T} \quad (43)$$

Já o R^2 ajustado é uma modificação que considera o número p de variáveis incluídas no modelo. Observando-se a Equação 44, nota-se que o valor ajustado decresce à medida que p aumenta.

$$R^2_{aj} = 1 - \left(\frac{n-1}{n-p} \right) (1 - R^2) \quad (44)$$

2.5.7 ANÁLISE RESIDUAL

Os resíduos de um modelo de regressão desenvolvem um importante papel no julgamento da sua adequação. Considerando-se alguns aspectos peculiares, o conjunto de resíduos deve ter distribuição normal com média zero e variância σ^2 , devem ser não correlacionados quando comparados a outros resíduos, à média dos tratamentos, às

variáveis preditoras e quando analisados ao longo do tempo. Devem ter padrão aleatório, não tendencioso e devem preferencialmente assumir a forma padronizada segundo a Equação 45.

$$d_i = \frac{e_i}{\sqrt{MQ_E}} = \frac{e_i}{\sqrt{\hat{\sigma}^2}} \quad (45)$$

A padronização dos resíduos cria um escalonamento para o desvio padrão, o que torna sua interpretação mais fácil. Outra forma de escalonamento é a de Student, dada pela Equação 46.

$$d_i = \frac{e_i}{\sqrt{\hat{\sigma}^2(1-h_{ii})}} \quad (46)$$

em que h_{ii} representa o i -ésimo elemento da diagonal de uma Matriz “Chapéu” H .

2.5.8 TESTES PARA FALTA DE AJUSTES DOS MODELOS

A presença de réplicas em um experimento permite que a soma dos quadrados do erro seja dividida em duas partes: o erro puro e a falta de ajuste. Quando o valor P do teste de falta de ajuste é superior ao nível de significância, então, não há evidência de que o modelo não explique adequadamente a variação da resposta (MONTGOMERY, 2001).

A adição de pontos centrais aos arranjos experimentais proporciona a obtenção de uma estimativa do erro experimental. De acordo com Montgomery (2001), este artifício permite que a soma de quadrados residual (SS_e) seja discriminada em dois componentes: (a) a soma de quadrados devida ao erro puro (SS_{pe}) e (b), a soma de quadrados devida à falta de ajuste do modelo escolhido (SS_{lof}) (*lof*, do inglês *Lack of Fit*). Assim, pode-se escrever a soma de quadrados residual como na Equação 47.

$$SS_e = SS_{pe} + SS_{lof} \quad (47)$$

Por suposição, admita-se que existam n_i observações de uma dada resposta de interesse no i -ésimo nível dos regressores x_i , $i = 1, 2, \dots, m$. Considere-se que y_{ij} denote a j -ésima observação de uma resposta no nível x_i , $i = 1, 2, \dots, m$ e $j = 1, 2, \dots, n_i$. Como existe um total de observações $n = \sum_{i=1}^m n_i$, então, o ij -ésimo resíduo será dado pela Equação 48.

$$y_{ij} - \hat{y}_i = (y_{ij} - \bar{y}_i) + (\bar{y}_i - \hat{y}_i) \quad (48)$$

em que \bar{y}_i é a média das n_i observações no nível x_i . Elevando-se ao quadrado ambos os lados da Equação 48 e somando-se cada i e j , obtém-se a Equação 49.

$$\underbrace{\sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \hat{y}_i)^2}_{SS_r} = \underbrace{\sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}_{SS_{ep}} + \underbrace{\sum_{i=1}^m n_i (\bar{y}_i - \hat{y}_i)^2}_{SS_{lof}} \quad (49)$$

O lado esquerdo da Equação 49 é a soma de quadrados residual SS_r . Os dois termos do lado direito são respectivamente, o erro puro SS_{ep} e a falta de ajuste do modelo SS_{lof} . Pode-se notar que SS_{ep} é obtido computando-se a correta soma de quadrados das observações repetidas em cada nível da variável independente (fator), em operações repetidas para cada um dos m níveis de x . Se a suposição de variância constante for satisfeita, SS_{ep} representa um modelo independente de medição do erro puro, porque apenas a variabilidade das respostas em nível dos fatores é utilizada para calculá-lo. Uma vez que existem $n_i - 1$ graus de liberdade associados ao erro puro em cada nível, o número total de graus de liberdade para o termo de erro puro é igual a $n - m$.

Como se nota, SS_{lof} é uma soma quadrática ponderada dos desvios encontrados entre a resposta média \bar{y}_i em um dado nível da variável independente e o correspondente valor ajustado \hat{y}_i . O valor ajustado \hat{y}_i é a estimativa fornecida pelo modelo de regressão escolhido para os valores das variáveis independentes em uma

dada observação. Se o valor ajustado \hat{y}_i possuir um valor numérico muito próximo ao valor das médias das respostas \bar{y}_i , então, há um forte indício de que o modelo de regressão escolhido é o mais correto. Caso contrário haverá a necessidade de se estudar outro modelo que apresente um melhor ajuste dos dados observacionais. Há $m-p$ graus de liberdade associados à SS_{lof} porque existem m níveis de x e p graus de liberdade perdidos em função dos p parâmetros que devem ser estimados pelo modelo.

A estatística de teste para a falta de ajuste pode, então, ser escrita na forma da Equação 50.

$$F_0 = \frac{SS_{lof} / (m - p)}{SS_{pe} / (n - m)} = \frac{MS_{lof}}{MS_{pe}} \quad (50)$$

Quando a estatística de F_0 assumir valores menores que o nível de significância escolhido, então, haverá evidência suficiente para se aceitar a hipótese nula de que a diferença entre o valor ajustado e a média observada é nula. Logo, não haverá falta de ajuste na estimativa proporcionada pelo modelo escolhido.

2.5.9 VARIAÇÕES EVOLUCIONÁRIAS EM OPERAÇÕES (EVOP)

Para além das motivações acadêmicas que deram origem ao DOE, a metodologia do Projeto de Experimentos foi desde cedo aplicada na indústria (a primeira aplicação prática do DOE ocorreu em 1930, na empresa British Textile). Após a Segunda Guerra mundial, a metodologia foi introduzida na indústria química e nos processos industriais de empresas nos Estados Unidos e Europa. Sua notória eficácia fez dela uma das principais ferramentas de melhoria de processos (PAIVA, 2006).

As ferramentas do projeto de experimentos deram origem a diversos métodos orientados à otimização de processos, dentre as quais os métodos de Taguchi (TAGUCHI *ET AL.*, 1987), a metodologia da Superfície de Resposta (MYERS & MONTGOMERY, 1995) e as Variações Evolucionárias em Operações, ou EVOP.

O método EVOP foi proposto por Box, em 1957 (BOX, 1957). Propõe o

monitoramento pelo uso do DOE e a otimização continuada de um dado processo, por meio do deslocamento do ponto de operação em direção ao ótimo (MYERS & MONTGOMERY, 1995). O método se apresenta como alternativa de otimização para solucionar as discrepâncias ocorridas entre resultados experimentos de pequena escala e os resultados obtidos em plantas fabris reais ao adotar-se os níveis recomendados pelos experimentos. Também se propõe a ser uma ferramenta de utilização periódica para reajuste dos pontos de operação em função de alterações sofridas pelos elementos envolvidos em um processo ao longo do tempo (BOX & DRAPER, 1998).

As variações evolucionárias consistem na introdução sistemática de pequenos desvios nos níveis correntes da entrada de um dado processo em unidades industriais reais. Um experimento planejado segundo a metodologia DOE é realizado. Trata-se, em geral, de um arranjo 2^k com um ponto central correspondente aos níveis correntes de operação. À execução de uma réplica do experimento dá-se o nome de ciclo EVOP.

As réplicas são repetidas e analisam-se os resultados dos experimentos. Busca-se uma diferença estatística significativa entre a saída do processo obtida aplicando-se os tratamentos e a saída originada pelo ponto corrente de operação (o ponto central do arranjo). Quando a análise indica que alguma variável ou interação têm efeito significativamente positivo sobre o processo (por análise de curvatura ou mudança na média), os níveis de operação devem ser deslocados para a combinação dos níveis que levou ao melhor resultado. O conceito é ilustrado na Figura 11.

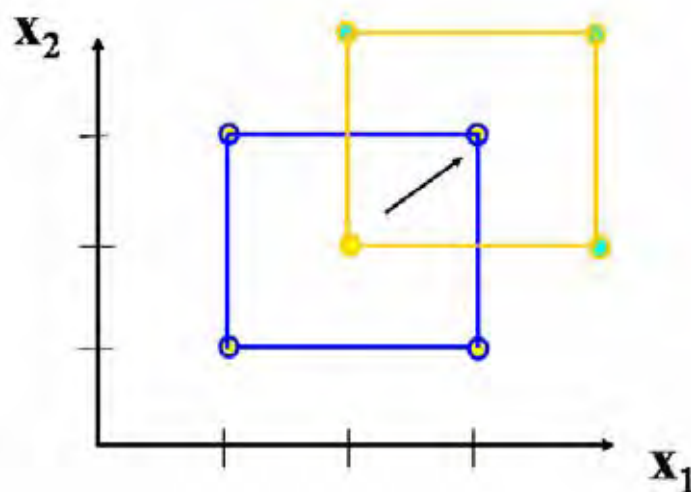


Figura 11 – Ilustração do conceito de EVOP

Após o estabelecimento do novo ponto de operação entende-se que uma fase do EVOP foi completada (MYERS & MONTGOMERY, 1995). Uma nova fase é então conduzida, centrada em torno do novo ponto de operação. As iterações continuam até que não se note melhoria significativa, indicando assim que o processo atingiu a região de ótimo. Os autores propõem a reaplicação do método após intervalos de tempo regulares, com o objetivo de verificar se o processo continua atuando na região do ótimo, ou se alguma condição (tal como desgastes em componentes de uma planta, ou variações na qualidade dos insumos) afetou o desempenho do mesmo (BOX & DRAPER, 1998).

Da forma como foi originalmente proposto, o EVOP se valia de planilhas e análises simplificadas dos efeitos dos fatores, que foram posteriormente substituídas pela realização da ANOVA com uso de pacotes computacionais. Sempre que a análise de variância indicar que certa combinação dos níveis testados conduz a uma melhoria do resultado do processo, segundo o nível de significância adotado, o ponto de operação deve ser deslocado para a combinação indicada. As fases EVOP são repetidas, caminhando-se em direção à região de ótimo, até o ponto em que a análise conclua que o ponto central é ponto de maior rendimento (MYERS & MONTGOMERY, 1995).

2.6 O CONCEITO DE MODELAGEM

Montgomery *et al.* (2008) propuseram um arcabouço conceitual para o desenvolvimento de modelos efetivos de predição para análises de regressão múltipla e para a predição de séries temporais. A construção e a validação de modelos, segundo os autores, deve seguir um conjunto bem definido de etapas. O primeiro passo, como pode ser visto na Figura 12, é a Definição do Problema.

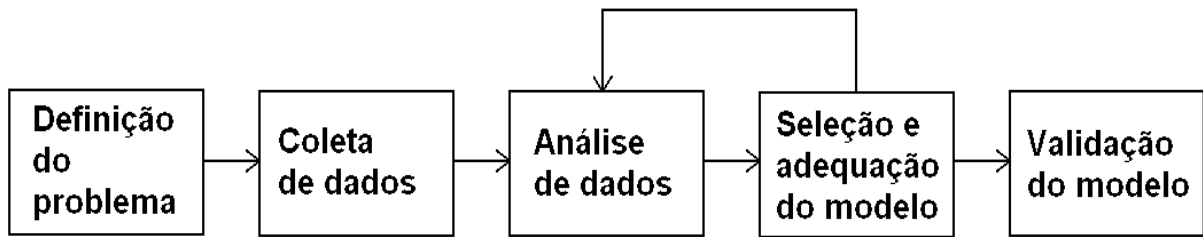


Figura 12: Processo de construção e validação de modelos – adaptado de Montgomery *et al.* (2008)

A etapa de Definição compreende a especificação do que será previsto pelo modelo, das variáveis independentes a serem utilizadas como preditores, e também a definição de como os usuários esperam utilizar o modelo resultante. A fase seguinte é a de Coleta de Dados que inclui a definição das técnicas utilizadas para coletar dados válidos e representativos do processo a ser modelado. A etapa de Análise dos dados se refere ao tratamento aplicado aos dados colhidos a fim de convertê-los em informação útil.

Sendo o comportamento de um modelo definido por sua arquitetura e por seus parâmetros livres, torna-se necessário realizar a Seleção e Adequação (ou Ajuste) do modelo. A Seleção do modelo está relacionada às regras empregadas para decidir-se em favor de uma determinada configuração para o modelo. A Adequação do modelo, por sua vez, corresponde aos ajustes realizados nos parâmetros livres do mesmo para garantir a máxima exatidão e precisão bem como as técnicas aplicadas na busca de valores ótimos para tais parâmetros.

A etapa final deve ser a Validação do modelo proposto. Nesta etapa, o modelo é aplicado para a predição da grandeza de interesse em casos não incluídos nas etapas anteriores de Seleção e Adaptação, com o objetivo de fornecer uma medida independente do desempenho do modelo. Em cada etapa citada, os autores citados recomendam fortemente o uso de técnicas estatísticas para garantir a desenvolvimento de modelos úteis e confiáveis.

3 A MODELAGEM DA RUGOSIDADE DA PEÇA POR REDES NEURAIIS ARTIFICIAIS

3.1 CONSIDERAÇÕES INICIAIS

Este capítulo sintetiza e analisa esforços de investigação que utilizam redes neurais em aplicações fora de linha para predição da rugosidade da peça. O objetivo é apresentar uma visão ampla das estratégias empregadas e problemas normalmente encontrados na literatura. Faz-se uma análise crítica do atual estágio da investigação científica sobre o tema. A análise feita se vale das recomendações de autores reconhecidos das áreas de neurocomputação e estatística para uma discussão sobre as boas práticas na modelagem da rugosidade da peça com o uso de RNAs.

A revisão inclui um conjunto de critérios de comparação, bem como avalia a forma como os resultados de pesquisa foram validados. As principais vertentes na literatura são identificadas e suas diferenças são comentadas.

Encontram-se incluídos os principais resultados publicados em Pontes *et al.* (2010a). O resultado da revisão mostra as principais lacunas existentes na literatura, estabelece um conjunto de boas práticas para a modelagem da rugosidade por meio de Redes Neurais e comprova o caráter inédito do método proposto nesta tese.

3.2 REDES NEURAIIS APLICADAS A PROCESSOS DE USINAGEM

O uso de RNAs em aplicações de usinagem, de acordo com Sick (2002), pode ser categorizado em aplicações em linha (ou *online*) ou fora de linha (ou *offline*). Em aplicações em linha, sensores e atuadores são instalados nas máquinas ferramentas e redes são treinadas em tempo real usando os parâmetros do processo medidos e inseridas em algum esquema de controle da operação de usinagem. Já nas aplicações fora de linha dados experimentais ou dados históricos são usados para treinar a rede com o objetivo de construir modelos para uso em planejamento ou otimização de processos. O uso técnicas de controle de qualidade fora de linha é um meio eficaz para

a melhoria da qualidade a custo relativamente baixo, segundo Antony (2005).

O potencial das RNAs para uso em processos de usinagem foi percebido na década de 1990. A fim de prever a vida da ferramenta no torneamento de ferro fundido com ferramentas de corte de material cerâmico, Ezugwu *et al.* (1995) utilizaram os resultados de experimentos envolvendo valores distintos de velocidade de corte e avanço para treinar uma MLP. Chao & Hwang (1997) realizaram um estudo similar. Os primeiros estudos sobre a modelagem da rugosidade da peça podem ser encontrados em Chien & Chou (2001), em que redes foram usadas para a predição da rugosidade da peça, força de corte e vida da ferramenta, em processo de torneamento de um aço inox 304. Tsai & Wang (2001) compararam arquiteturas distintas de RNAs para a predição da rugosidade da peça em processo de usinagem por descargas elétricas (EDM). Benardos & Vosniakos (2002) treinaram uma rede neural para prever a rugosidade da peça utilizando dados obtidos de experimentos conduzidos em uma fresadora com controle numérico (CNC) e planejados de acordo com técnicas DOE.

A aplicação de RNAs em processos de usinagem tem sido o tema de um número considerável de publicações. Em muitas delas pesquisadores têm chegado a conclusões favoráveis em relação a seu uso. Davim *et al.* (2008) sustentam que as redes neurais são capazes de capturar a não-linearidade característica do torneamento. Karpát & Özel (2007) falam sobre a dificuldade em gerar modelos analíticos explícitos das complexas relações existentes entre os parâmetros do processo de torneamento duro. As redes neurais, segundo os autores, constituem uma opção viável e prática para a modelagem. Os resultados obtidos por Özel *et al.* (2007) mostram que os modelos de redes neurais são capazes de prever, para uma série de condições de corte, o desgaste da ferramenta e padrões de rugosidade da peça; além disso, declaram os autores, podem também ser utilizadas para o planejamento inteligente do processo de torneamento duro. Oktem *et al.* (2005), utilizando dados obtidos experimentalmente, treinaram uma rede MLP utilizando o algoritmo BP para a predição da rugosidade. Os autores relataram excelentes resultados. Outros exemplos podem ser encontrados no trabalho de Assarzadeh & Ghoreishi (2008), cujo objetivo foi o de otimizar a rugosidade da peça em EDM usando redes neurais. Os autores declaram a eficácia do

uso do MLPs para a estimativa da taxa de remoção de material e para a predição de R_a . Em Hossain *et al.* (2008), um modelo de rede foi desenvolvido para a investigação e predição da relação entre os parâmetros de corte e rugosidade da peça durante o fresamento de acabamento em alta velocidade de uma liga de níquel Inconel 718. As redes apresentaram bom desempenho na predição. Outras abordagens incluem o trabalho de Panda & Mahapatra (2009) na qual componentes principais foram utilizados para a modelagem do desgaste de brocas em processo de furação. Os componentes principais dos parâmetros de perfuração foram calculados e as redes foram treinadas para predizê-los. Observou-se que as redes foram capazes de classificar as categorias baixo desgaste e alto desgaste com precisão de 90% e de prever o desgaste de flanco da broca com erro inferior a $\pm 6,5\%$.

Deve-se ressaltar, porém, que não há consenso sobre a experiência com redes neurais para a modelagem da rugosidade da peça. Em Çaydas & Ekici (2011) três diferentes tipos de Máquinas de Vetores de Suporte (ou SVM, do inglês Support Vector Machines) e redes neurais artificiais foram desenvolvidas e comparadas para estimar a rugosidade de superfície resultante do torneamento de aço austenítico AISI 304. Naquele trabalho, as SVM apresentaram melhor desempenho que as redes neurais. A falta de métodos sistemáticos de projeto para as redes neurais, o tempo gasto para identificação de topologias eficazes e a falta de repetibilidade dos resultados são apontados como um entrave à disseminação de seu uso (ZAIN *ET AL.*, 2010; DHOKIA *ET AL.*, 2008). Para Cervellera *et al.* (2007) e Karnik *et al.* (2008), a identificação de uma boa arquitetura de RNA requer várias tentativas de modelagem, consistindo em atividade longa e de resultado incerto. Ambrogio *et al.* (2008) citam a necessidade de grandes quantidades de dados para treinamento e validação como um limitador para a aplicação prática de redes neurais para predição em processos de usinagem. Já para Bagci & Işik (2006), o esforço computacional necessário para identificar e construir uma rede neural artificial com bom desempenho é superior ao requerido por outros métodos.

Dentre os esforços para a otimização de arquiteturas de redes neurais, encontra-se o trabalho de Zanchettin *et al.* (2005). Os autores, da área de neurocomputação, propuseram um método rudimentar com a utilização do DOE para a identificação do

fator mais influente sobre o desempenho de um sistema de inferência neuro-fuzzy. Exemplos de tentativas de otimização podem ser encontrados no trabalho de Mohana Rao *et al.* (2009), cujo objetivo é a modelagem da rugosidade da peça em processo de usinagem por descarga elétrica (*Electrical Discharge Machining*, ou EDM). Os autores utilizaram algoritmos genéticos para otimizar os pesos sinápticos da rede. Algoritmos genéticos foram usados naquele trabalho para a otimização dos pesos da rede. Ortiz-Rodrigues *et al.* (2006) propuseram a utilização dos métodos de Taguchi (uma técnica de DOE) para o projeto robusto de MLPs treinadas pelo algoritmo de retropropagação e desenvolveram uma estratégia experimental que enfatiza a otimização simultânea dos parâmetros da rede neural artificial sob diferentes condições de ruído. Em alguns estudos, abordagens heurísticas são utilizadas para definir os parâmetros (KOHLI & DIXIT, 2005). Em outros casos, uma estratégia de variar um parâmetro por vez é usada na busca de uma configuração adequada (FREDJ & AMAMOU, 2006).

Para a análise realizada a seguir, serão consideradas as etapas do arcabouço conceitual para elaboração de modelos para análise de regressão múltipla e previsão de séries temporais proposto por Montgomery & Runger (2007) e Montgomery *et al.* (2008). Os critérios propostos por aqueles autores foram adotados para categorização e análise crítica dos trabalhos da literatura.

3.3 CARACTERIZAÇÃO DOS TRABALHOS ANALISADOS

A busca de artigos foi realizada utilizando bases de pesquisa bibliográfica, tais como Elsevier, Springer, Taylor & Francis, Emerald, dentre outras. Publicações foram escolhidas para avaliação caso preenchessem os seguintes quesitos:

- foco na previsão de rugosidade da peça em processos de usinagem;
- utilização de redes neurais como técnica de modelagem;
- adoção de abordagem fora de linha (*off-line*) para o ajuste de modelos de rede.

Com base nos critérios listados, 45 publicações foram selecionadas e revisadas. Na Figura 13 é mostrada a distribuição cronológica dos trabalhos.

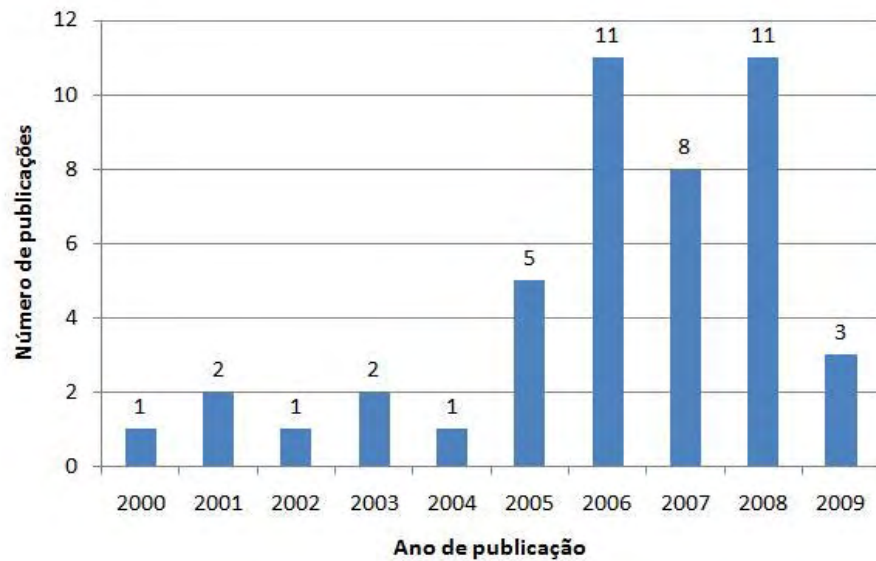


Figura 13 - Distribuição anual das publicações revisadas – adaptado de Pontes *et al.* (2010a)

Observa-se que a predição de rugosidade fora de linha na usinagem por meio de RNAs tem atraído um interesse sustentado por parte dos pesquisadores. Entre 2005 e 2008, houve um aumento no número de trabalhos publicados sobre o assunto (a pesquisa cobre o período que vai até o primeiro semestre de 2009). Na Figura 14 são indicados os processos específicos de usinagem investigados ao longo publicações selecionadas.

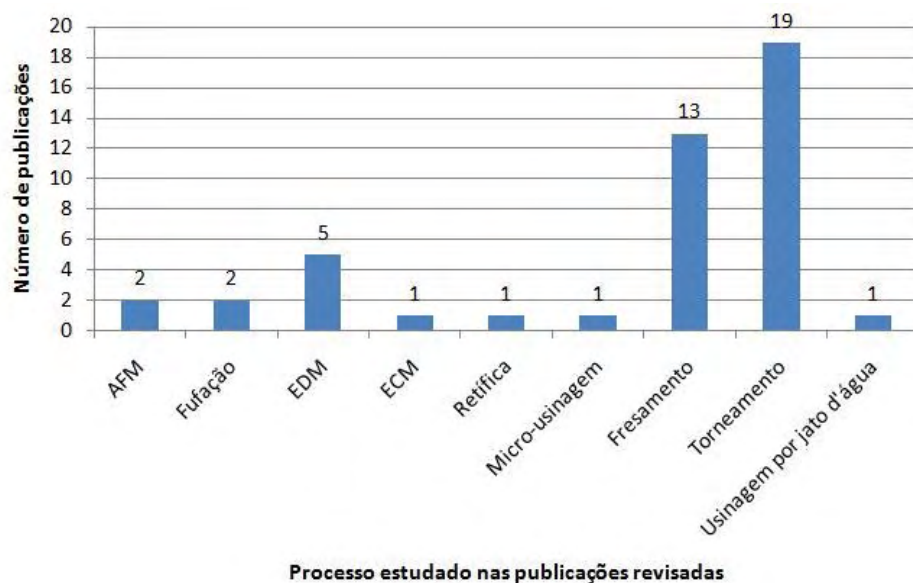


Figura 14 - Processos de usinagem estudados nas publicações revisadas – adaptado de Pontes *et al.* (2010a)

Trinta e seis dos quarenta e cinco trabalhos selecionados lidam com processos de usinagem convencional, dentre os quais torneamento (MALAKOOTI & RAMAN; 2000; CHIEN & CHOU; 2001; CHOUDURY & BARTARYA, 2003; LEE & CHEN, 2003; EZUGWU *ET AL.*, 2005; KOHLI & DIXIT, 2005; ÖZEL & KARPAT, 2005; PAL & CHAKRABORTY, 2005; BAGCI & IŞIK; 2006; CUS & ZUPERL, 2006; SONAR *ET AL.*; 2006; ZHONG *ET AL.*, 2006; AL-AHMARI, 2007; ÖZEL *ET AL.*, 2007; SARMA & DIXIT, 2007; BASHEER *ET AL.*, 2008; DAVIM *ET AL.*, 2008; SHARMA *ET AL.*, 2008; KARAYel, 2009), fresamento (BENARDOS & VOSNIAKOS, 2002; BALIC, 2004; WANG & HSU; 2005; CHANG & LU, 2006; ÖKTEM *ET AL.*, 2006; ERZURUMLU & ÖKTEM, 2007; JESUTHANAM *ET AL.*, 2007; BRUNI *ET AL.*, 2008; DHOKIA *ET AL.*, 2008; EL-SONBATY *ET AL.*, 2008; KUMANAN *ET AL.*, 2008; CORREA *ET AL.*, 2009; ÖKTEM; 2009), furação (SANJAY & JYOTHI, 2006; TSAO & HOCHENG, 2008), e retífica (FREDJ & AMAMOU, 2006).

Os nove trabalhos remanescentes tratam de processos de usinagem não-convencional, tais como usinagem por descargas elétricas, ou EDM (TSAI & WANG, 2001; SARKAR *ET AL.*, 2006; KUMAR & CHOUDURY, 2007; MANDAL *ET AL.*, 2007; MARKOPOULOS *ET AL.*, 2008), usinagem por jato abrasivo, ou AFM (ALI-TAVOLI *ET AL.*, 2006; MOLLAH & PRATIHAR, 2008), usinagem eletroquímica, ou ECM (ASOKAN *ET AL.*, 2008), micro-fresamento de acabamento (TANSEL *ET AL.*, 2005), e usinagem por jato de água (ÇAYDAS & HASÇALIK, 2008).

Os Quadros 4 e 5 sintetizam decisões tomadas pelos autores dos artigos. Indicam as arquiteturas adotadas em cada trabalho, bem como o número de entradas e saídas de rede utilizadas. O uso da velocidade de corte, avanço e profundidade de corte como entradas para a predição também é detalhado. Além disso, o Quadro 5 especifica a medida de rugosidade usada em cada trabalho. Nos quadros, a legenda N.E. indica que o item não foi especificado no trabalho. Os quadros trazem informação sobre o algoritmo de treinamento utilizado (para trabalhos com MLPs), sobre o uso de distintos grupos de treinamento e seleção, sobre a utilização de um terceiro conjunto de validação, e sobre a grandeza adotada para medir a exatidão de rede em cada

estudo. No Quadro 5, 'BP' indica treinamento pelo algoritmo de retropropagação e 'LM' e 'QN' significam uso dos algoritmos Levenberg-Marquadt e Quasi-Newton, respectivamente. O termo 'Novo' aponta para a proposição de um novo algoritmo de treinamento pelos autores do trabalho; o termo 'Par. Fractal' indica parâmetro de função fractal e o termo 'Class. Bin.' representa um índice de acerto em tarefa de classificação binária, e 'N. A.' indica que o item é não aplicável ao caso.

Quadro 4 - Sumário das decisões sobre redes tomadas nas publicações revisadas – adaptado de Pontes *et al.* (2010a)

Referência	Arquitetura RNA	Num. entradas	Vel. corte	Avanço	Profund. corte	Num. saídas
Cus & Zuperl, 2006	MLP/RBF	3	X	X	X	3
Basheer <i>et al.</i> , 2008	MLP	5		X	X	1
Sharma <i>et al.</i> , 2008	MLP	4	X	X	X	4
Özel & Karpat, 2005	MLP	7	X	X		1
Öktem, 2009	MLP	4	X	X	X	1
Karayel, 2009	MLP	3	X	X	X	1
Ezugwu <i>et al.</i> , 2005	MLP	4	X	X	X	1
Correa <i>et al.</i> , 2009	MLP	6		X	X	1
Çaydas & Hasçalik, 2008	MLP	5				1
Öktem <i>et al.</i> , 2006	MLP	5	X	X	X	1
Chang & Lu, 2006	Outro	6	X	X	X	1
Pal & Chakraborty, 2005	MLP	5	X	X	X	1
Erzurumlu & Öktem, 2007	MLP	5	X	X	X	1
Benardos & Vosniakos, 2002	MLP	8+	X	X	X	1
Davim <i>et al.</i> , 2008	MLP	3	X	X	X	2
Zhong <i>et al.</i> , 2006	MLP	7		X	X	2
Fredj & Amamou, 2006	MLP	8+	X		X	1
Al-Ahmari, 2007	MLP	4	X	X	X	3
Chien & Chou, 2001	MLP	3	X	X	X	N. E.
Tsai & Wang, 2001	MLP/RBF	N. E.				1
Özel <i>et al.</i> , 2007	MLP	6	X	X		1
Dhokia <i>et al.</i> , 2008	MLP/RBF	3	X	X	X	1
Bagci & Işik, 2006	MLP	3	X	X	X	1
Malakooti & Raman, 2000	MLP	3	X	X	X	4
Kohli & Dixit, 2005	MLP	4	X	X	X	1
Sonar <i>et al.</i> , 2006	RBF	4	X	X	X	1
Sarma & Dixit, 2007	MLP/RBF	3	X	X	X	N. E.
Choudury & Bartarya, 2003	MLP	3	X	X	X	1
Lee & Chen, 2003	MLP	3				1
Kumanan <i>et al.</i> , 2008	RBF	4	X	X	X	1
Balic, 2004	MLP	3				3
Bruni <i>et al.</i> , 2008	MLP	8+	X			1
El-Sonbaty <i>et al.</i> , 2008	MLP	5	X	X	X	1
Wang & Hsu, 2005	MLP	4	X	X	X	1
Jesuthanam <i>et al.</i> , 2007	MLP	4	X	X	X	1
Sanjay & Jyothi, 2006	MLP	3	X	X		1
Tsao & Hocheng, 2008	RBF	4	X	X	X	2
Kumar & Choudury, 2007	MLP	3				2
Markopoulos <i>et al.</i> , 2008	MLP	3				1
Sarkar <i>et al.</i> , 2006	MLP	6				3
Mandal <i>et al.</i> , 2007	MLP	3				3
Mollah & Pratihar, 2008	RBF	4				2
Asokan <i>et al.</i> , 2008	MLP	4				2
Tansel <i>et al.</i> , 2006	MLP	4	X	X	X	1
Ali-Tavoli <i>et al.</i> , 2006	Outra	2				2

Quadro 5 - Sumário das decisões sobre redes tomadas nas publicações revisadas
(continuação) – adaptado de Pontes *et al.* (2010a)

Referência	Medida rugosidade	Alg. Trein.	Conj. Distintos trein/teste	3º Conjunto validação	Figura de erro
Cus & Zuperl, 2006	R_a	BP	Sim	Não	Nenhuma
Basheer <i>et al.</i> , 2008	R_a	LM	Sim	Não	MAE(%)
Sharma <i>et al.</i> , 2008	R_a	BP	Sim	Sim	MSE
Özel & Karpat, 2005	R_a	LM	Sim	Não	RMSE
Öktem, 2009	R_a	BP	Sim	Não	RMSE
Karayel, 2009	R_a / R_{max}	BP	Sim	Sim	MAE(%)
Ezugwu <i>et al.</i> , 2005	R_a	LM	Sim	Sim	Correlação
Correa <i>et al.</i> , 2009	R_a	BP	Sim	Não	Proporção
Çaydas & Hasçalik, 2008	R_a	BP	Não	Não	MAE(%)
Öktem <i>et al.</i> , 2006	R_a	N. E.	Sim	Não	MAE(%)
Chang & Lu, 2006	R_a	N. A.	Sim	Não	MAE(%)
Pal & Chakraborty, 2005	R_a	BP	Sim	Não	MSE
Erzurumlu & Öktem, 2007	R_a	N. E.	Sim	Não	MAE(%)
Benardos & Vosniakos, 2002	R_a	LM	Sim	Sim	MSE
Davim <i>et al.</i> , 2008	R_a / R_t	BP	Sim	Não	MAE(%)
Zhong <i>et al.</i> , 2006	R_a / R_t	BP	Sim	Sim	MAE(%)
Fredj & Amamou, 2006	R_a	BP	Sim	Não	N. E.
Al-Ahmari, 2007	R_a	BP	Não	Não	MAE(%)
Chien & Chou, 2001	R_a	BP	Sim	Não	MAE(%)
Tsai & Wang, 2001	R_a	BP	N. E.	Não	MAE(%)
Özel <i>et al.</i> , 2007	R_a / R_t	LM	Sim	Não	Não
Dhokia <i>et al.</i> , 2008	R_a	LM	Sim	Sim	MAE(%)
Bagci & Işik, 2006	R_a	BP	Sim	Não	MAE(%)
Malakooti & Raman, 2000	R_a	BP	Sim	Não	Não
Kohli & Dixit, 2005	R_a	BP	Sim	Sim	MAE(%)
Sonar <i>et al.</i> , 2006	R_a	N. A.	Sim	N. E.	MAE(%)
Sarma & Dixit, 2007	R_a	N. E.	Sim	Não	Não
Choudury & Bartarya, 2003	R_a	Novo	Sim	Não	MAE(%)
Lee & Chen, 2003	R_a	BP	Sim	Não	MAE(%)
Kumanan <i>et al.</i> , 2008	N. E.	N. A.	Sim	Não	Não
Balic, 2004	N. E.	BP	Sim	Não	Não
Bruni <i>et al.</i> , 2008	R_a	BP	Não	Não	N. E.
El-Sonbaty <i>et al.</i> , 2008	Par. Fractal	LM	Sim	Sim	R2
Wang & Hsu, 2005	Class.Bin	QN	Não	Não	Não
Jesuthanam <i>et al.</i> , 2007	R_a	Novo	N. E.	Não	MAE(%)
Sanjay & Jyothi, 2006	R_a	BP	Não	Não	Não
Tsao & Hocheng, 2008	R_a	N. A.	Sim	Não	MAE(%)
Kumar & Choudury, 2007	R_a	LM	Sim	Não	MAE(%)
Markopoulos <i>et al.</i> , 2008	R_a	BP	Sim	Sim	Não
Sarkar <i>et al.</i> , 2006	R_a	BP	Sim	Não	MAE(%)
Mandal <i>et al.</i> , 2007	R_a	BP	Sim	Não	MAE(%)
Mollah & Pratihari, 2008	R_a	N. A.	Sim	Não	MAE(%)
Asokan <i>et al.</i> , 2008	R_a	N. E.	Sim	Não	MAE(%)
Tansel <i>et al.</i> , 2006	R_a	BP	Sim	Não	MAE(%)
Ali-Tavoli <i>et al.</i> , 2006	R_a	N. A.	Sim	Não	Não

O processo que apareceu por mais vezes entre as publicações revisadas foi o torneamento, responsável por aproximadamente 42% do total. O mesmo foi observado por Sick (2002) em uma revisão bibliográfica sobre o monitoramento de ferramentas de corte. Entre as publicações que lidam com o torneamento está o trabalho de Karayel (2009), onde um modelo de rede neural é proposto para a predição e controle da rugosidade da peça em um torno CNC. Davim *et al.* (2008) desenvolveram modelos de predição de rugosidade da peça utilizando redes de arquitetura MLP para investigar os efeitos das condições de corte, utilizando ferramentas de metal duro no torneamento de aço 9SMnPb28K (DIN). Em Basheer *et al.* (2008), um modelo de rede neural artificial é utilizado para prever a rugosidade da peça na usinagem de materiais compósitos. Encontram-se também aplicações que tratam de torneamento duro. Este processo tornou-se possível graças ao desenvolvimento de novos materiais para ferramentas de corte, tais como nitreto de boro cúbico e materiais cerâmicos (SINGH & RAO, 2007). O processo de torneamento duro se comporta de modo distinto do torneamento normal (KARPAT & ÖZEL, 2007). Quiza *et al.* (2008) sustentam que os métodos tradicionais de regressão apresentam graves limitações quando aplicados ao processo de torneamento duro devido à extrema não-linearidade que caracteriza o processo e à contínua introdução de novos materiais para ferramenta. Estes fatos parecem favorecer a utilização de RNA. Um exemplo pode ser encontrado em Özel *et al.* (2007) em que as redes neurais foram utilizadas para investigar a influência dos parâmetros de corte no desgaste de flanco da ferramenta e na qualidade da superfície no torneamento duro de aço AISI D2 (60 HRC) com pastilhas de cerâmica com geometria alisadora.

Em segundo lugar dentre os processos estudados nos artigos revistos aparece o fresamento, tema de 28% das publicações revistas. Oktem (2009) conduziu um estudo sobre a rugosidade da peça com objetivo de modelar e otimizar parâmetros de corte em processo de fresamento de acabamento de aço AISI 1040 usando ferramentas cerâmicas com revestimento de TiAlN com aplicação de fluido. O trabalho de Bruni *et al.* (2008) teve por objetivo a construção de modelos de predição da rugosidade da peça em fresamento de acabamento de face de aço inoxidável AISI 420 B. Outros processos convencionais de usinagem estão também representados, como no trabalho de Fredj & Amamou (2006) que propuseram o uso de RNAs para a predição da

rugosidade em processo de retificação.

São encontrados também exemplos de aplicações envolvendo a remoção de materiais não metálicos, como por exemplo, Tsao & Hocheng (2008), que fizeram uso de redes para prever forças de usinagem e rugosidade da peça no processo de furação de materiais compósitos. Dhokia *et al.* (2008), trabalharam com a usinagem de materiais macios, como o polipropileno. O seu trabalho foi direcionado para a criação de um modelo que ajudasse na identificação de condições de corte ótimas para a redução da rugosidade. Em Bagci & Isik (2006), uma RNA é empregada para estimar a rugosidade da peça no torneamento unidirecional de materiais compósitos reforçados com fibra de vidro.

RNAs têm sido utilizadas também para a predição da rugosidade em processos não-convencionais de usinagem. Sarkar *et al.* (2006), Mandal *et al.* (2007) e Markopoulos *et al.* (2008) propuseram modelos neurais para a predição de rugosidade da peça em operações de EDM. Sendo um processo já bem estabelecido na fabricação de ferramentas e moldes, EDM segue como um dos processos mais artesanais da indústria manufatureira (MARKOPOULOS *ET AL.*, 2008). Outro exemplo é o trabalho de Mollah & Pratihari (2008) no qual os autores empregaram redes RBF para modelagem das relações entrada-saída no processo de usinagem por jato abrasivo (AFM).

3.4 DEFINIÇÃO DO PROBLEMA E COLETA DE DADOS

As atividades iniciais para a elaboração de modelos, segundo Montgomery *et al.* (2008), correspondem à definição do problema e à coleta de dados. A definição do problema inclui a especificação do que será o objeto de modelagem, a estratégia a ser empregada, e o entendimento sobre como os usuários esperam utilizar o modelo a ser produzido. A coleta de dados inclui a seleção de variáveis preditoras e as técnicas utilizadas para obter dados úteis para a construção do modelo.

O problema em questão ao longo dos documentos analisados é o uso de RNAs como ferramenta de modelagem para a predição ou classificação da rugosidade da peça em processos de usinagem. Os parâmetros do processo são normalmente usados

como variáveis independentes. A expectativa geral, conforme relatado nos artigos, é a obtenção de modelos que apresentem erros de predição inferiores aos obtidos por outros métodos.

Aplicações de modelos compostos por RNAs em conjunto com outras ferramentas são observadas. A motivação para o uso de abordagens híbridas é a obtenção de resultados superiores aos que seriam obtidos separadamente por cada uma das técnicas envolvidas. Isto pode ser observado em Oktem (2009), onde um algoritmo genético foi empregado para explorar um modelo neural para rugosidade em busca de um conjunto de parâmetros de corte que minimizasse a rugosidade da peça. A mesma estratégia é observada em Tansel *et al.* (2006). Outros exemplos são os estudos de Oktem *et al.* (2006), Gao *et al.* (2008), e também Mollah & Pratihari (2008).

Dezoito trabalhos comparam modelos obtidos a partir de RNAs a modelos desenvolvidos a partir de outras técnicas, tais como análise de regressão múltipla ou metodologia de superfície resposta. Os modelos de rede são considerados como superiores em dez destes trabalhos. Exemplos podem ser encontrados em Asokan *et al.* (2008), Çaydas & Hasçalik (2008) e Tsao & Hocheng (2008). Al-Ahmari (2007) realizou testes estatísticos para determinar a diferença entre o desempenho de um modelo neural e um não-neural. O autor concluiu que a rede neural foi superior. Em seis estudos o desempenho dos modelos de RNA foi considerado como igual ao do modelo alternativo. Em dois artigos, o modelo concorrente superou o modelo neural. Esta foi a conclusão obtida por Correa *et al.* (2009), em que as redes neurais foram comparadas a redes Bayesianas (um método de classificação baseado em máquinas de aprendizagem) para a predição da rugosidade da peça em usinagem de alta velocidade.

Comparações entre diferentes tipos de redes neurais são realizadas em seis publicações (CHIEN & CHOU, 2001; TSAI & WANG, 2001; CUS & ZUPERL, 2006; SONAR *ET AL.*, 2006; ASOKAN *ET AL.*, 2008; DHOKIA *ET AL.*, 2008). Em Dhokia *et al.* (2008), os autores compararam o desempenho de duas arquiteturas de rede distintas, MLPs e RBF. Sua conclusão, baseada em testes de hipótese, é que as redes RBFs superaram as MLPs. Sonar *et al.* (2006) fizeram a mesma comparação. Naquele estudo, as MLPs superaram as RBFs. Tsai & Wang (2001) compararam seis configurações de rede, incluindo MLPs e RBFs treinadas por algoritmos distintos. O

coeficiente de Pearson (R^2) medido na etapa de validação foi utilizado para expressar os resultados. Embora uma grande quantidade de dados seja apresentada, o artigo não aponta claramente um vencedor. Deve-se salientar que tanto MLPs e RBF possuem propriedades de aproximadores universais de funções (HAYKIN, 1998). Em consequência, sempre haverá uma topologia de um modelo capaz de emular o desempenho de uma rede do outro tipo (PONTES *ET AL.*, 2010).

Abordagens menos usuais podem também ser encontradas. El-Sonbaty *et al.* (2008) desenvolveram modelos de RNAs para a análise e a predição da relação entre as condições de corte e parâmetros fractais em superfícies usinadas em operações de faceamento. Na obra de Balic (2004), a rugosidade é usada como entrada de uma rede MLP treinada para prever as melhores condições de corte, para utilização em planejamento de processo.

O passo seguinte, após a definição do problema e da estratégia de investigação, é a coleta de dados para o treinamento e teste das redes. Deve-se inicialmente definir o número e a natureza das variáveis independentes (ou preditores) a serem empregadas. Esta é uma questão importante na modelagem por RNAs. O conjunto de entradas deve, preferencialmente, incluir todas as variáveis significativas para a predição da rugosidade. Ao mesmo tempo, porém, autores de neurocomputação como Haykin (2008) recomendam o uso de um conjunto mínimo de entradas não correlacionadas para melhor modelagem. Na Figura 15 mostra-se o número de preditores adotados entre os 45 estudos avaliados.

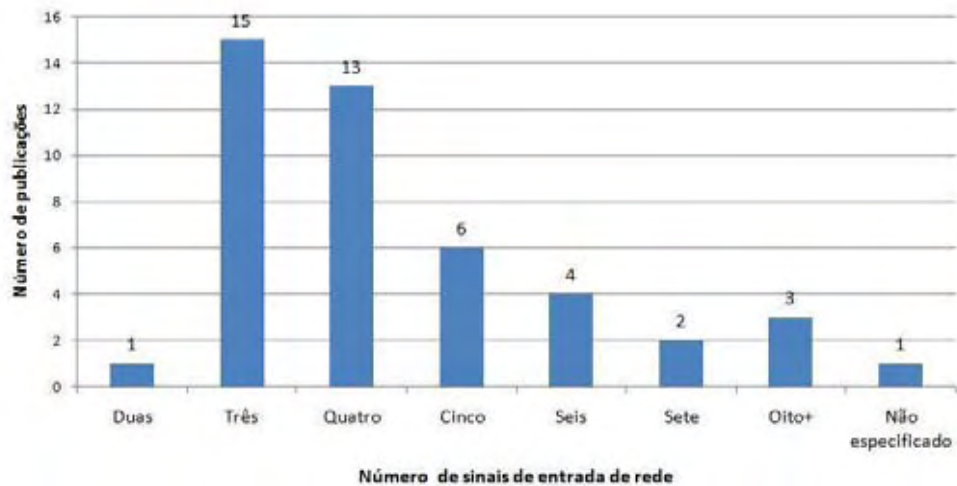


Figura 15 – Número de preditores utilizados para construção dos modelos nas publicações revisadas – adaptado de Pontes *et al.* (2010a)

O mínimo observado é de dois (ALI-TAVOLI *ET AL.*, 2006). Em três artigos, Benardos & Vosniakos (2002), Fredj & Amamou (2006) e Bruni *et al.* (2008), mais do que oito entradas são usadas. Praticamente todas as obras analisadas (43 de 45) definem claramente a natureza das variáveis independentes utilizadas. Nos dois artigos restantes, os textos sugerem que o trabalho foi baseado em parâmetros de corte. Na Figura 16 mostra-se o número de vezes que uma dada variável de entrada apareceu entre os 36 artigos que tratam processos de usinagem convencional.

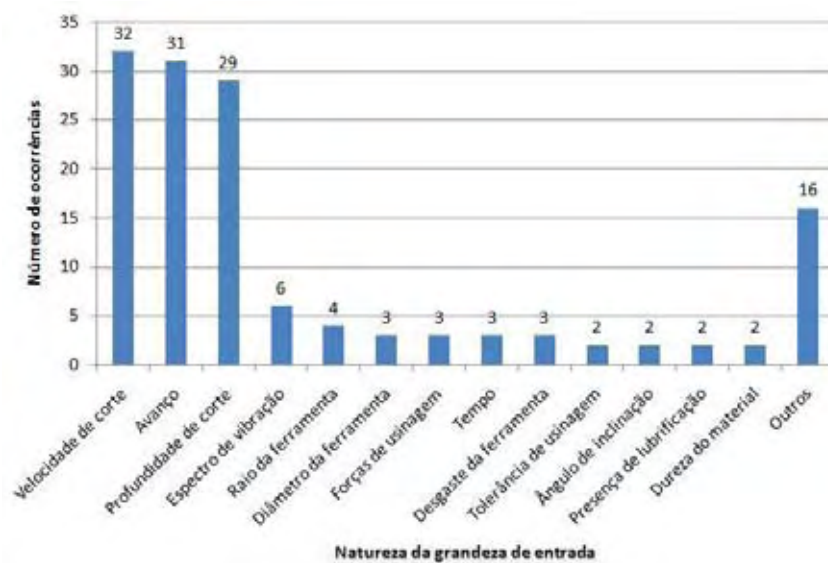


Figura 16 – Natureza dos preditores utilizados para construção dos modelos de redes neurais – adaptado de Pontes *et al.* (2010a)

Pode-se observar que as condições de corte (velocidade de corte, avanço e profundidade de corte) constituem a imensa maioria dos preditores utilizados, como exemplificado em Bagci & Isik (2006), Dhokia *et al.* (2008), Karayel (2009) e Oktem (2009). Apesar de menos freqüentes, outros parâmetros do processo são também empregados como entradas do modelo. O raio de ponta da ferramenta, por exemplo, é um dos fatores investigados por Al-Ahmari (2007). Sanjay & Jyothi (2006) utilizaram o diâmetro da broca, juntamente com as condições de corte em operação de furação, como preditores. O ângulo de inclinação ferramenta-peça é adotado como uma entrada por Zhong *et al.* (2006). Digno de nota é que em apenas três publicações a informação sobre o desgaste da ferramenta é utilizada como entrada, mesmo que muitos autores a considerem como um parâmetro influente na formação de rugosidade (ORABY *ET AL.*, 2004; CHANG & LU, 2006; KARPAT & ÖZEL, 2007). Entradas aparecendo apenas uma vez entre os artigos revisados foram agrupadas sob o rótulo "Outros". Essas entradas incluem o tamanho e volume de insertos em ferramentas, número de passes, tipos de pastilhas presentes na ferramenta, material da peça, a geometria de ponta de corte, a condição de lubrificação e comprimento de corte.

Trabalhos que tratam de processos de usinagem não convencionais fazem uso de informações específicas de processo para modelagem. Markopoulos *et al.* (2008) adotaram como entradas a corrente de pulso, duração de pulso e o tipo de material processado para a predição da rugosidade em EDM. Cinco tipos de aço foram testados e os outros fatores variaram sobre uma ampla faixa que cobriu desde o desbaste até operações de acabamento. O modelo para EDM proposto por Sarkar *et al.* (2006), adotou a duração do pulso, a duração da pausa, intensidade da corrente, tensão no eletrodo, volume do fluxo de fluido e distância entre eletrodo e peça como entradas para a predição da rugosidade em usinagem eletroquímica.

Vinte e sete obras justificam sua escolha de preditores. Em algumas as entradas são selecionadas com base em estudos anteriores, enquanto em outras, nenhuma justificativa é apresentada. Em Öktem *et al.* (2006), por exemplo, o intervalo de corte dos parâmetros é definido pelos limites fixados pelo fabricante da ferramenta utilizada. Em quatro estudos (MALAKOOTI & RAMAN, 2000; BENARDOS & VOSNIAKOS,

2002; ÖZEL & KARPAT, 2005; FREDJ & AMAMOU, 2006;), conjuntos distintos de preditores são comparados na busca da melhor combinação para a construção do modelo. Um exemplo dessa estratégia pode ser visto em Benardos & Vosniakos (2002) que utilizaram um arranjo de Taguchi (uma técnica DOE) para selecionar as entradas para predição de rugosidade em processo de fresamento.

Uma outra questão relativa à coleta de dados é a técnica adotada para a coleta. Cinco abordagens distintas podem ser encontradas entre as obras revistas, como ilustrado na Figura 17.

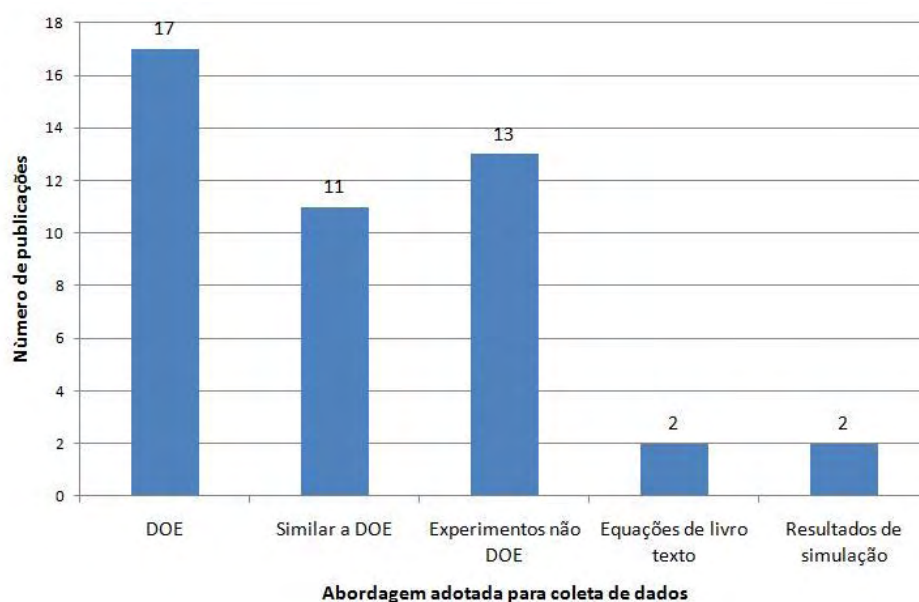


Figura 17 – Abordagem adotada para formação dos casos de treinamento – adaptado de Pontes *et al.* (2010a)

Em 17 publicações, o conjunto de treinamento é formado por resultados de experimentos planejados e executados de acordo com técnicas do DOE. Exemplos são encontrados em Oktem (2009), que utiliza os resultados de um arranjo fatorial completo. Çaydas & Hasçalik (2008) usaram como conjunto de treinamento os resultados de um arranjo de Taguchi. Cerca de um quarto dos trabalhos (onze publicações) utilizaram para formação dos conjuntos resultados de arranjos semelhantes aos de experimentos planejados. Nenhuma menção explícita, no entanto, é feita sobre o uso da metodologia, como visto, por exemplo, em Karayel (2009). Em alguns trabalhos, o arranjo real é apenas mencionado, mas não detalhado. Isto é o que

ocorre em Öktem *et al.* (2006) que utilizaram um notável conjunto de 243 exemplos coletados de operações de corte. Em quase um terço das publicações (treze estudos) o conjunto de treinamento é formado com dados obtidos de experimentos que não seguem a metodologia do Projeto de Experimentos. Em Sukhtomya & Tannock (2005) dados históricos colhidos de operações reais em linha de produção de lâminas para turbo-jatos foram retroativamente dispostos em um arranjo de Taguchi. Os dados do “arranjo” foram utilizados para o treinamento de uma rede neural que tinha por objetivo modelar o processo. Chien & Chou (2001) empregaram resultados experimentais obtidos a partir de 96 pontos de teste aleatoriamente selecionados entre os limites de condição de corte estabelecida. Em dois artigos, exemplos de treinamento foram gerados por simulações, e em dois outros, equações de livros didáticos foram empregadas para a geração de casos. Nota-se que DOE é prevalente na coleta de dados porque muitos papéis têm por objetivo comparar modelos de rugosidade obtidos por redes neurais a modelos obtidos por regressão múltipla obtidos usando resultados de experimentos planejados. O uso dos resultados de DOE para treinar as redes, nestes casos, tem por objetivo fornecer uma base comum para comparação entre os modelos concorrentes.

A conveniência do uso de técnicas DOE para formar conjuntos de treinamento é uma questão em aberto. Nenhum dos artigos revisados avaliou a influência da técnica usada para coleta de dados sobre o desempenho dos modelos. Outro ponto ainda insuficientemente explorado é o da eficiência relativa entre os dados necessários para a modelagem RNAs em comparação com outros métodos. Só Kohli & Dixit (2005) realizaram uma investigação sobre o número de casos necessários para produzir um modelo com uma exatidão específica para uma determinada combinação máquina, peça e ferramenta. Seu trabalho propõe um método para estabelecer o número mínimo de casos de treinamento necessário, na modelagem da rugosidade, para a obtenção de bom desempenho por parte da redes.

Outra classificação pode ser feita em relação à natureza e ao número de saídas dos modelos. Na Figura 18 é mostrado o número de saídas de rede empregado entre as publicações revistas.

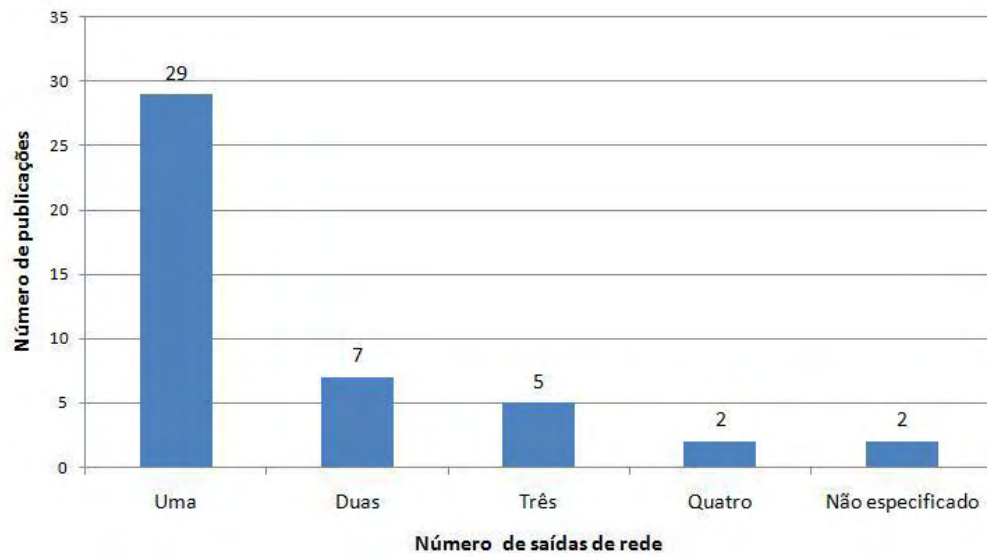


Figura 18 – Número de saídas de rede adotadas nas publicações revisadas – adaptado de Pontes *et al.* (2010a)

Na maior parte dos artigos, redes de uma única saída são empregadas. Na Figura 19 é mostrado um resumo da natureza das grandezas modeladas.

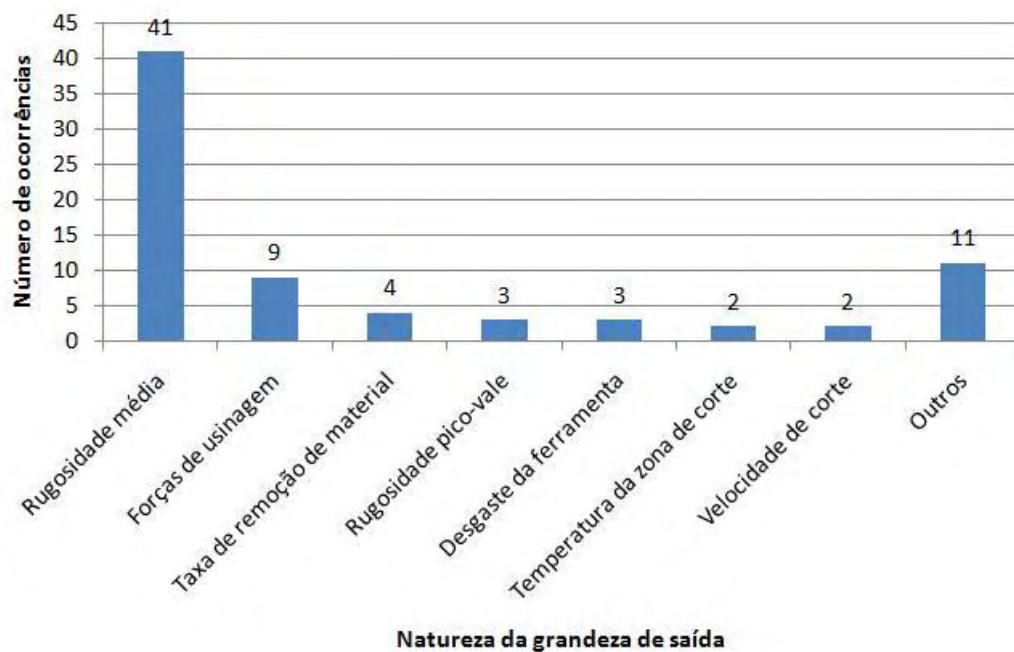


Figura 19 – Natureza das saídas de rede adotadas nas publicações revisadas – adaptado de Pontes *et al.* (2010a)

Observa-se que a rugosidade média é adotada para medição da rugosidade da peça em 41 obras, sendo responsável por cerca de 90% do total. Além de rugosidade média, redes neurais são comumente aplicadas para a predição de alguns outros parâmetros de interesse. Um exemplo pode ser encontrado em Davim *et al.* (2008), que desenvolveram modelos de RNAs para predizer rugosidade pico-vale (R_t). Mandal *et al.* (2007) adotaram três figuras de mérito (a taxa de remoção de material, o desgaste da ferramenta e rugosidade da peça) como medida do desempenho do processo. Sarkar *et al.* (2006) propuseram um modelo para o processo EDM que adotou como saídas do modelo a velocidade de corte, rugosidade da peça, e a distância entre peça e eletrodo.

Medidas de rugosidade que apareceram apenas uma vez foram agrupadas sob o rótulo "Outros". Essas saídas incluem rugosidade máxima, rugosidade média de dez pontos, parâmetros fractais, taxa de produção, custo de usinagem, consumo de energia, taxa de desgaste da ferramenta, e tempo de usinagem.

3.5 SELEÇÃO E ADEQUAÇÃO DE MODELOS

A etapa de Seleção e Adequação de Modelos consiste em escolher uma ou mais técnicas de predição e em estimar os parâmetros do modelo para ajustá-lo aos dados coletados (MONTGOMERY *ET AL.*, 2008). Após a coleta de dados, torna-se necessário selecionar e especificar a rede neural a ser utilizada. O projetista da rede escolhe o paradigma de aprendizagem a ser empregado, bem como a arquitetura de rede e a função de ativação dos neurônios. O conceito de Adequação, quando aplicado a RNAs, corresponde à definição de uma topologia de rede apropriada, ao treinamento à qual será submetida para aprender a relação entre variáveis preditoras e variáveis modeladas, bem como à análise dos dados realizada para comprovar a adequação do modelo.

O primeiro passo é a definição do esquema de aprendizagem e da arquitetura a ser utilizada. Na Figura 20 mostra-se o número de vezes que uma determinada arquitetura de RNA aparece entre as publicações revisadas.

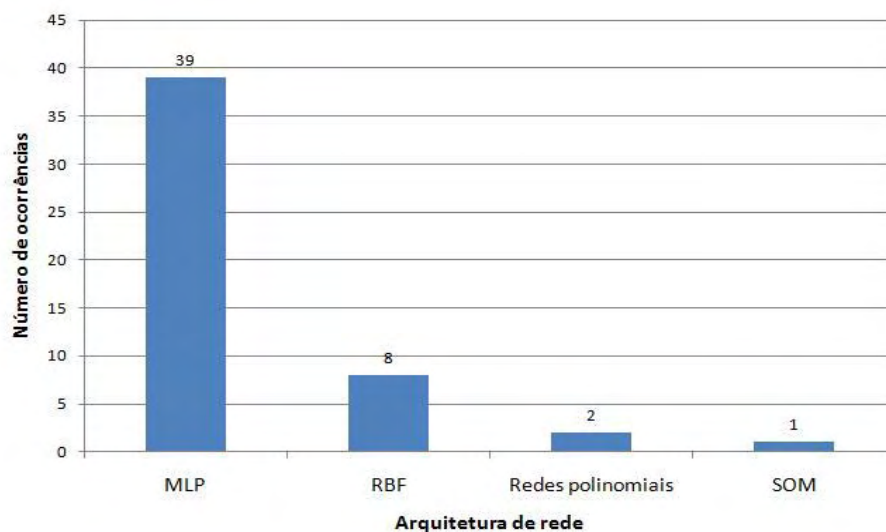


Figura 20 – Número de ocorrências de distintas arquiteturas de redes ao longo das publicações revisadas – adaptado de Pontes *et al.* (2010a)

Como alguns trabalhos empregam mais de um tipo de rede, com objetivo de comparar seus desempenhos, o número total de aparições é superior ao total de artigos revisados. Na Figura 20 observa-se que redes de arquitetura MLPs são usadas em 39 trabalhos (aproximadamente 87% do total). RBFs são usadas em oito publicações (cerca de 18%).

Nota-se que, entre os trabalhos revisados, redes de aprendizagem não supervisionada são raramente empregadas. Enquanto o uso de MLPs e RBF (exemplos dos paradigmas de aprendizagem supervisionada e híbrida, respectivamente) é generalizado, apenas três publicações empregam redes de aprendizagem não supervisionada. Um exemplo pode ser encontrado em Balic (2004), no qual mapas auto-organizáveis, ou SOM, do inglês *Self Organizing Maps* (KOHONEN, 1982; KOHONEN, 2001) são empregados como parte de um conjunto maior de redes para gerar programas de computador para fresamento e furação em centros de usinagem. Redes polinomiais (IVAKHNENKO, 1971), outro tipo de rede que usa aprendizagem não supervisionada, são propostos para a modelagem de rugosidade por Ali-Tavoli *et al.* (2006) e por Chang & Lu (2006). Para agrupamento de dados em situações onde as classes são desconhecidas *a priori*, a teoria de neurocomputação recomenda o uso de redes sem supervisão. Fica claro que o uso de redes de paradigma de aprendizagem não supervisionada para classificação da rugosidade é ramo pouco explorado na

literatura. Exemplos de tais arquiteturas incluem as já mencionadas redes SOM e também as redes ART (do inglês *Adaptive Resonance Theory Networks*, ou redes de ressonância adaptativa) (CARPENTER & GROSSBERG, 1987a; CARPENTER & GROSSBERG, 1987b).

Um exemplo de aplicação de redes recorrentes pode ser encontrado em Xiao & Zhu (2010), que investigaram a otimização da composição de materiais abrasivos não-amiantados utilizando técnicas de projeto de experimentos, metodologia de superfície de resposta e redes neurais artificiais. Foram estudados os efeitos de componentes e etapas do processo sobre três características de fricção do material abrasivo. Os autores treinaram redes de Elman utilizando dados medidos por um dinamômetro durante experimentos planejados segundo a metodologia DOE.

Quanto à função de ativação em uso com as redes da arquitetura MLP, 11 publicações fazem uso exclusivo da função tangente hiperbólica sigmóide. Sete trabalhos utilizaram a função logística sigmóide. Dois trabalhos comparam o efeito da função de ativação utilizada sobre a precisão da rede. Dezenove publicações não fazem qualquer menção à função de ativação em uso. Embora não haja evidência de diferença no desempenho relacionada à função de ativação empregada em MLPs, Haykin (2008) afirma que a função tangente hiperbólica, devido à sua forma simétrica, pode levar a uma convergência mais rápida durante o treinamento. Publicações que trabalham com redes de arquitetura RBF fazem uso explícito da função gaussiana, como visto em Tsai & Wang (2001), Sonar *et al.* (2006), Mollah & Pratihari (2008) e Tsao & Hocheng (2008), ou então não especificam a função de ativação utilizada.

Após a seleção da arquitetura de rede a ser utilizada, deve-se especificar os parâmetros livres da mesma. Existem muitas opções em termos de definição de topologia, algoritmos de treinamento, estratégias de treinamento e avaliação da adequação do modelo.

A definição de topologia é o inconveniente mais frequentemente relatado no uso de redes neurais ao longo dos artigos revistos. Autores envolvidos com a predição de rugosidade com RNAs citam repetidamente a falta de procedimentos sistemáticos para projeto como uma desvantagem na aplicação de redes neurais. Correa *et al.* (2009) declara: "Não existem métodos bem estabelecidos para a escolha dos parâmetros da

rede (número de camadas ocultas, número de neurônios na camada oculta, forma de funções de ativação).” Observações semelhantes são encontradas em Zhong *et al.* (2006) e Dhokia *et al.* (2008). A escolha de parâmetros de projeto adequados é essencial para o bom desempenho da rede. Por essa razão, é importante analisar como a questão é tratada na literatura. Os tratamentos dados pelos autores à definição da topologia de rede ao longo das publicações revisadas estão resumidos na Figura 21.

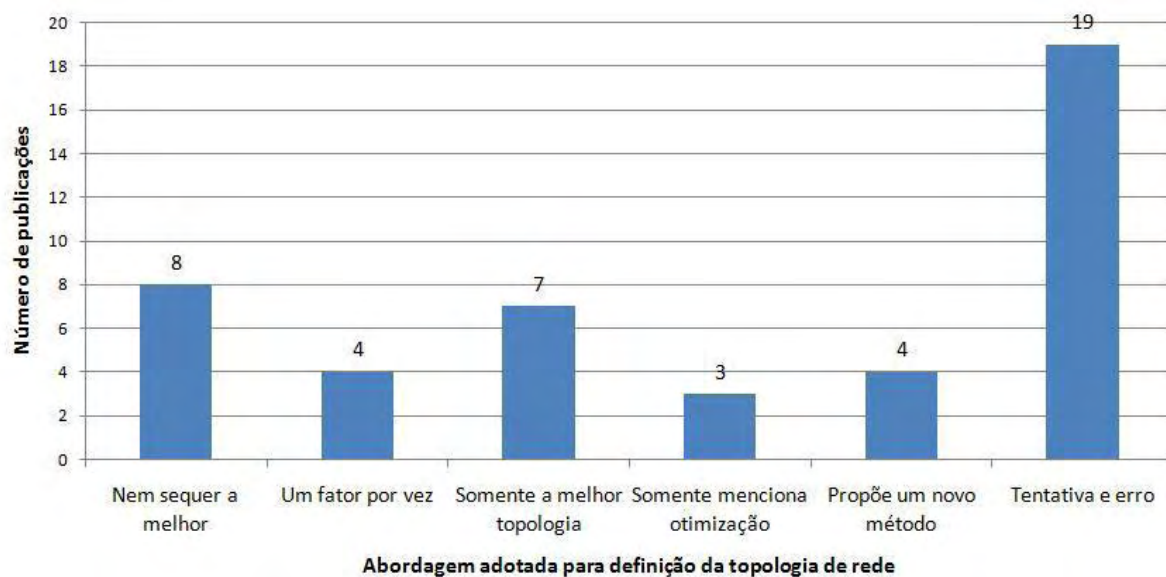


Figura 21 – Abordagens adotadas para definição de topologias nas publicações revisadas – adaptado de Pontes *et al.* (2010a)

Como visto na Figura 21, seis abordagens distintas podem ser encontradas. A mais comum é a definição por tentativa e erro, empregada em 19 trabalhos (aproximadamente 42% do total), como por exemplo em Zain *et al.* (2010) e Karayel (2009), para MLPs, e em Sonar *et al.* (2006) e Tsao & Hocheng (2008) para RBFs.

Autores propuseram métodos próprios para a definição da topologias em quatro publicações. Em Jesuthanam *et al.* (2007), um algoritmo genético foi utilizado para a escolha otimizada de pesos iniciais que são fornecidos como entrada para um algoritmo de otimização por enxame de partículas (*particle swarm optimization*), o qual por sua vez foi usado para a busca por pesos sinápticos ótimos. Os pesos resultantes foram aplicados à rede neural e o erro médio na predição foi comparado ao resultado esperado. O ciclo foi repetido até não se observar melhoria. Tansel *et al.* (2006) aplicaram algoritmos genéticos para encontrar topologias de rede otimizadas

para a predição de parâmetros de usinagem (dentre os quais a rugosidade) em operações de micro-usinagem. Sharma *et al.* (2008) apresentaram uma abordagem bastante peculiar em que o objeto da otimização é o número de épocas de treinamento em si. A estratégia resultou em redes com desempenho insatisfatório na predição da rugosidade.

Em poucas publicações, a topologia foi definida variando-se um parâmetro por vez como por exemplo, em Makeyev *et al.* (2008), em Mollah & Pratihari (2008), Kohli & Dixit (2005), Fredj & Amamou (2006) e Mandal *et al.* (2007).

Outras abordagens incluem o trabalho de Yang (2010), que utilizou métodos de programação linear para encontrar um modelo de rede neural com o melhor ajuste e com o mínimo número de neurônios na camada oculta. Neste estudo, o número de neurônios é tratado como o único fator relevante para a otimização do desempenho da rede, o que está em desacordo com a literatura.

Em três trabalhos a otimização é apenas mencionada. Sete trabalhos apresentam apenas a "melhor" topologia. Nestes, a configuração de parâmetros tais como o número de camadas ocultas, o número de neurônios nas camadas ocultas, ou o algoritmo de treinamento são fixados sem qualquer justificativa. Finalmente, em oito publicações, nem sequer a melhor topologia de rede é apresentada.

Algumas observações devem ser feitas sobre o assunto. Dado o impacto da topologia da rede sobre o desempenho do modelo neural, o fato de que as redes são projetadas por tentativa e erro em quase metade das publicações sugere que modelos de rugosidade obtidos nesses estudos podem estar longe de apresentar desempenho ótimo. O mesmo pode ser dito sobre os modelos apresentados sem evidência objetiva de otimização. Estratégias de variação de um parâmetro por vez também não são ideais para o projeto de rede, pois podem conduzir a soluções de sub-ótimas e são incapazes de detectar interações entre os fatores de projeto envolvidos (MONTGOMERY & RUNGER, 2007). A pior situação, no entanto, acontece nos casos em que nem mesmo a melhor topologia é apresentada.

Este problema é encontrado com frequência em publicações que pretendem comparar redes de arquiteturas distintas ou comparar redes com outros modelos. A ausência da otimização pode alterar drasticamente os resultados obtidos. Pode-se

aventar a hipótese de que o resultado de uma comparação seja apenas decorrência fortuita de uma má escolha de topologia de rede. Este é outro problema que pode levar a conclusões equivocadas em relação aos algoritmos, arquiteturas ou ao próprio uso de redes como uma opção para modelagem da rugosidade.

No intuito de aproveitar todo o potencial das RNAs para a modelagem, mais esforços devem ser investidos no projeto eficiente de redes. Um exemplo interessante sobre o tema, proveniente de outro campo de pesquisa, são as obras de Queiroz *et al.* (2007) e Balestrassi *et al.* (2009) em que a metodologia DOE é empregada com sucesso no projeto de redes MLP para previsão de séries temporais em aplicações relacionadas ao setor de energia elétrica. Em Guimarães & Silva (2007) e Guimarães *et al.* (2008), redes neurais foram utilizadas para modelar a descoloração de um tipo de corante em presença de H_2O_2 por processo de oxidação fotoquímica. Os autores propõem o uso do método Garson de partição (GARSON, 1991) para quantificar a influência das variáveis independentes associadas a cada neurônio de entrada no resultado do processo. O uso do DOE também é observado em Lasheras *et al.* (2010) e em Chen *et al.* (2010), para aplicações em outras áreas da manufatura.

Em seguida à definição da topologia, é necessário definir como o treinamento das redes será realizado. Em apenas duas publicações, Pal & Chadraborty (2005) e Mollah & Pratihari (2008), pode-se encontrar menção ao modo de treinamento adotado para MLPs. Em ambas, o modo de treinamento em batelada foi utilizado. Os algoritmos de treinamento em uso para redes MLP são apresentados na Figura 22.

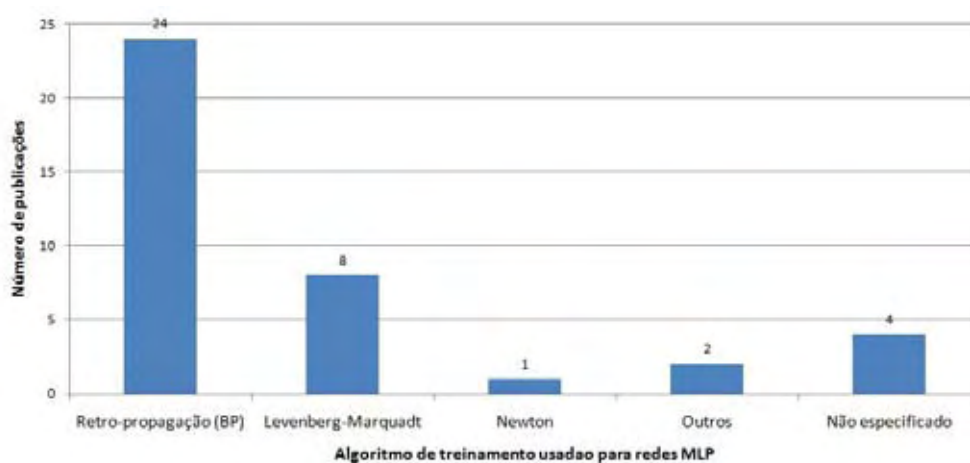


Figura 22 – Algoritmos adotados para o treinamento de redes MLP – adaptado de Pontes *et al.* (2010a)

Vinte e quatro trabalhos fazem uso do algoritmo de retropropagação padrão (RUMELHART *ET AL.*, 1986), indicado como BP na Figura 22. Um exemplo pode ser encontrado em Karayel (2009).

Variações do algoritmo de retropropagação e algoritmos de segunda ordem foram propostos para o treinamento de MLPs, com objetivo de aumentar a velocidade de convergência e exatidão. Citam-se, como exemplo, os algoritmos de Levenberg-Marquardt (HAGAN & MENHAJ, 1994) e Quasi-Newton (BATTITI, 1992). Entre os trabalhos analisados, o algoritmo Levenberg-Marquardt (citado como LM na Figura 22) é empregado em oito estudos, como em Ezugwu *et al.*(2005) e Basheer *et al.* (2008), sempre com bons resultados. Em um artigo (WANG & HSU, 2005), um algoritmo de Newton é empregado. Três trabalhos propõem novas abordagens para o treinamento (CHOUDURY & BARTARYA, 2003; JESUTHANAM *ET AL.*, 2007; KARAYEL, 2009), e em três outros, a escolha do algoritmo de treinamento não é clara.

O treinamento de uma MLP deve reduzir o erro de rede no conjunto de treinamento e ser interrompido quando o ponto de mínimo na superfície de erro é atingido. Os critérios adotados para a interrupção da fase de treinamento entre as publicações revisadas estão distribuídos conforme mostrado na Figura 23.

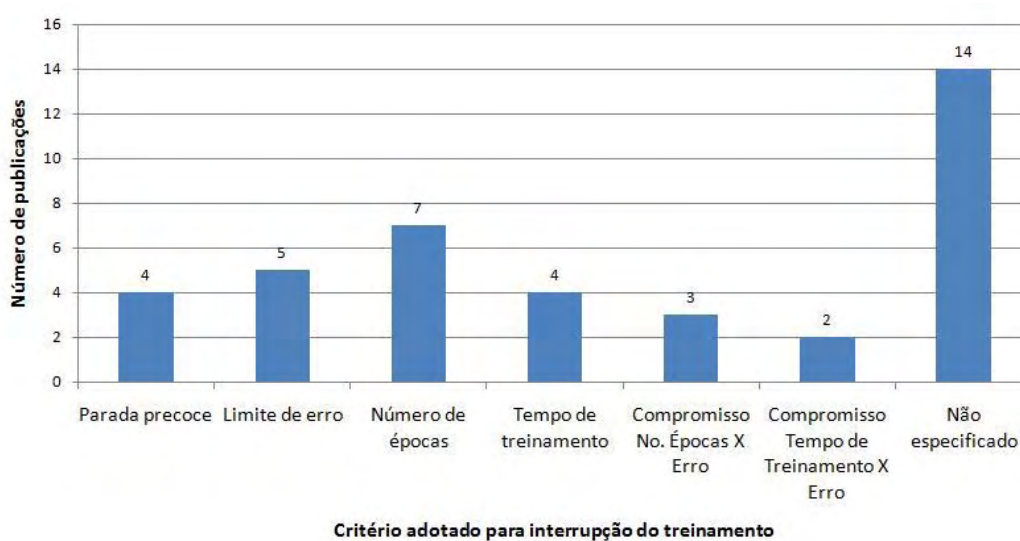


Figura 23 – Critérios adotados para interrupção do treinamento de redes neurais de arquitetura MLP – adaptado de Pontes *et al.* (2010a)

Quatro publicações fazem uso da interrupção precoce (BENARDOS & VOSNIAKOS, 2002; EZUGWU *ET AL.*, 2005; DHOKIA *ET AL.*, 2008; MARKOPOULOS *ET AL.*, 2008). Trata-se de uma técnica amplamente utilizada, segundo Haykin (2008), e consiste em verificar periodicamente o valor do erro da rede contra um conjunto de validação ao longo do treinamento. Este é interrompido quando o erro mínimo é alcançado e o valor do erro começa a crescer. Os critérios adotados para interrupção do treinamento dentre as publicações restantes não garantem a obtenção de erro mínimo. Em cinco trabalhos, o treinamento foi interrompido após o atingir-se um nível um erro de pré-especificado no conjunto de validação. Sete estudos interromperam o treinamento após um número pré-fixado de épocas de treinamento. Quatro outros estabeleceram um intervalo de tempo máximo para o treinamento. Em cinco publicações, um critério de compromisso foi adotado, e em quatorze obras não houve menção ao critério adotado para interrupção.

Além das recomendações anteriores, autores de neurocomputação sugerem o emprego de técnicas para otimizar a utilização de conjuntos de treinamento, tais como a validação cruzada, ou *K-fold cross-validation* (NGUYEN & WIDROW, 1990; HAYKIN, 2008). O uso de tal técnica pôde ser observado apenas em Correa *et al.* (2009).

Ao final das etapas previamente mencionadas, é necessário selecionar um modelo de rede. A seleção de uma rede baseia-se na análise das medidas de erro obtidas ao final da fase de treinamento. Redes que apresentam erro mínimo durante o treinamento são selecionadas e submetidas a testes de validação. Durante os testes de validação, a exatidão, precisão e capacidade de generalização do modelo devem ser estimadas usando dados não apresentados anteriormente à rede. Distintas figuras de mérito são utilizadas entre as publicações analisadas para expressar a exatidão dos modelos neurais nas fases de seleção e validação, como pode ser visto nas Figuras 24 e 25. Nestas figuras, MSE indica o uso do erro médio quadrático, RMSE significa raiz quadrada do erro médio quadrático, e SSE indica a soma dos erros médios quadráticos.

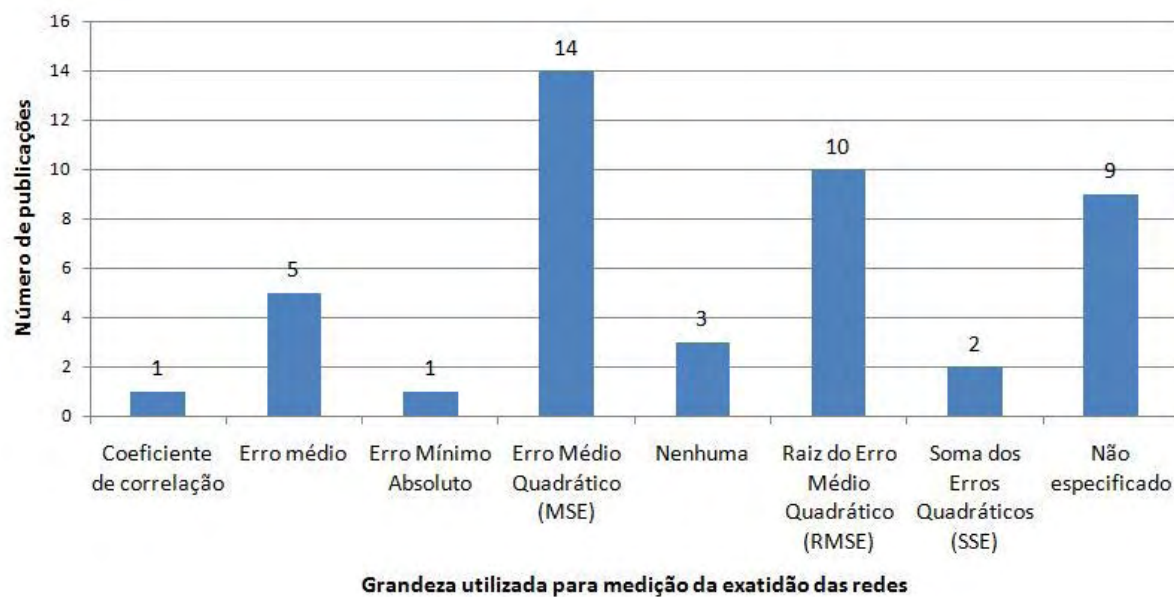


Figura 24 – Grandezas adotadas para seleção de redes – adaptado de Pontes *et al.* (2010a)

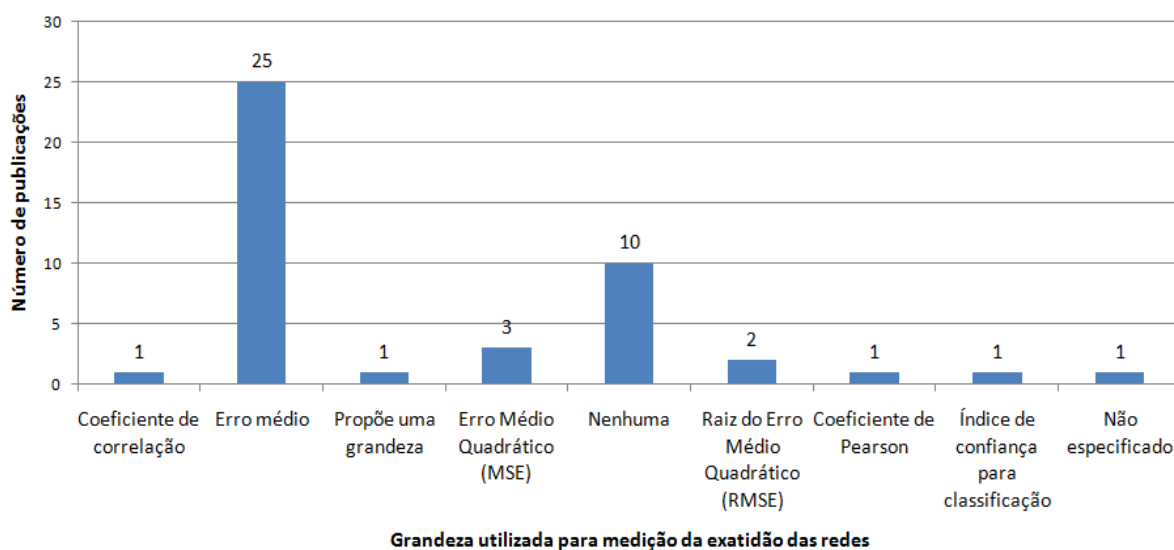


Figura 25 – Grandezas adotadas para validação de redes – adaptado de Pontes *et al.* (2010a)

Artigos que expressam a exatidão das redes por meio do erro médio absoluto ou erro médio em porcentagem foram agrupados sob o rótulo de "Erro Médio". Observa-se que figuras de mérito envolvendo erros quadráticos são prevalentes para a seleção de arquiteturas de rede. Por outro lado, figuras de mérito envolvendo erro médio são predominantes na fase de validação, seja para estimar a exatidão dos modelos ou para

comparar o desempenho de modelos neurais ao desempenho de modelos obtidos por outros métodos. O Erro médio é adotado como expressão final da exatidão das redes em 25 publicações (ou 56% do total).

3.6 VALIDAÇÃO DE MODELOS

O objetivo final de um modelo é fornecer informações exatas e predições confiáveis sobre as variáveis de interesse. A utilização de um modelo em situações práticas, portanto, requer uma avaliação da medida em que o modelo representa os fenômenos subjacentes à transformação de entradas em saídas. Isto é realizado por meio da validação do modelo. Conforme define Montgomery *et al.* (2008), validação consiste em aferir o desempenho do modelo na aplicação pretendida. Isto implica, segundo os autores, na obtenção de estimativas objetivas do erro esperado quando o modelo for aplicado à predição de dados distintos daqueles utilizados durante o treinamento. A aferição da confiabilidade associada a um modelo requer duas coisas: (1) testes independentes de validação após o treinamento, e (2) tratamento estatístico dos resultados dos testes.

Uma técnica fundamental à boa modelagem, de acordo com Montgomery *et al.* (2008), é o uso de distintos conjuntos de dados para ajuste e validação de modelos. Isto é necessário para avaliar-se o desempenho de um modelo quando exposto a novos dados e para comparar modelos distintos. O conceito, quando aplicado a RNAs, corresponde à divisão de dados disponíveis em conjuntos de treinamento e teste. Constitui uma das práticas elementares em neurocomputação. Na Figura 26 se ilustra como a questão é tratada entre as publicações revisadas. Em aproximadamente 84% dos trabalhos, a utilização de conjuntos distintos está claramente estabelecida. Em cinco artigos (ou 11% do total), o mesmo conjunto de dados é aplicado para treinar e testar as redes, e em dois estudos, a distinção entre conjuntos não é clara.

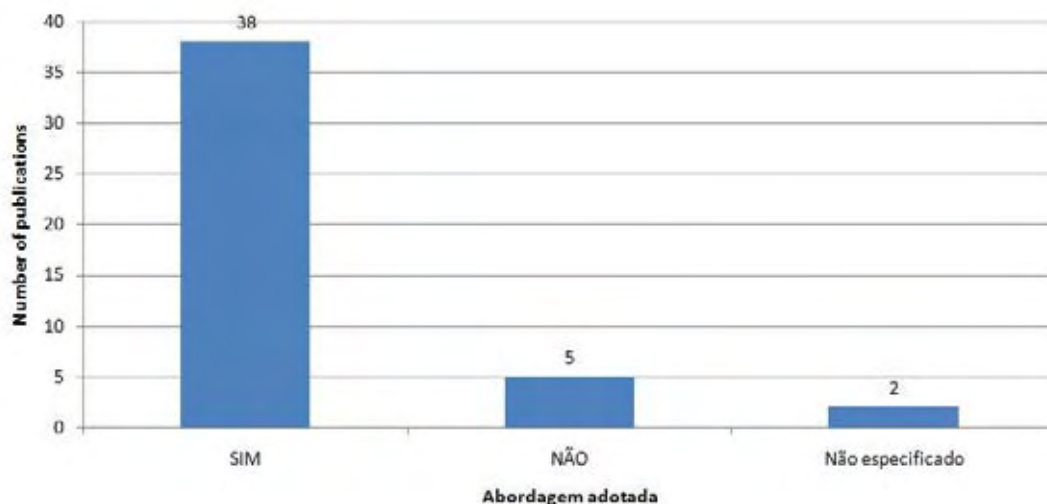


Figura 26 – Uso de conjuntos distintos de treinamento e teste – adaptado de Pontes *et al.* (2010a)

Um problema a evitar no uso de RNAs é a incidência do sobre-aprendizado (ou *overlearning*). Trata-se de um fenômeno que leva a uma redução na capacidade de generalização de uma rede. Ele pode surgir como resultado do excesso de treinamento. Algumas estratégias foram desenvolvidas para lidar com o problema. A mais comum, de acordo com Rumelhart *et al.* (1986) e com Haykin (2008), é o uso de uma forma básica de validação cruzada (NGUYEN & WIDROW, 1990), que consiste em dividir os dados coletados em três sub-grupos: estimação, validação e teste. O subconjunto de estimação é utilizado para treinamento da rede. O treinamento é interrompido periodicamente e a rede é testada contra o subconjunto de validação até que o erro de validação atinja um mínimo e comece a aumentar. A configuração de rede no ponto de erro mínimo é restaurada, a fase de treinamento é interrompida, e a rede é então aplicada à predição dos casos incluídos no subconjunto de teste. A técnica corresponde, na prática, a já mencionada interrupção precoce. A incidência do uso de tal esquema, entre os artigos analisados é apresentada na Figura 27.

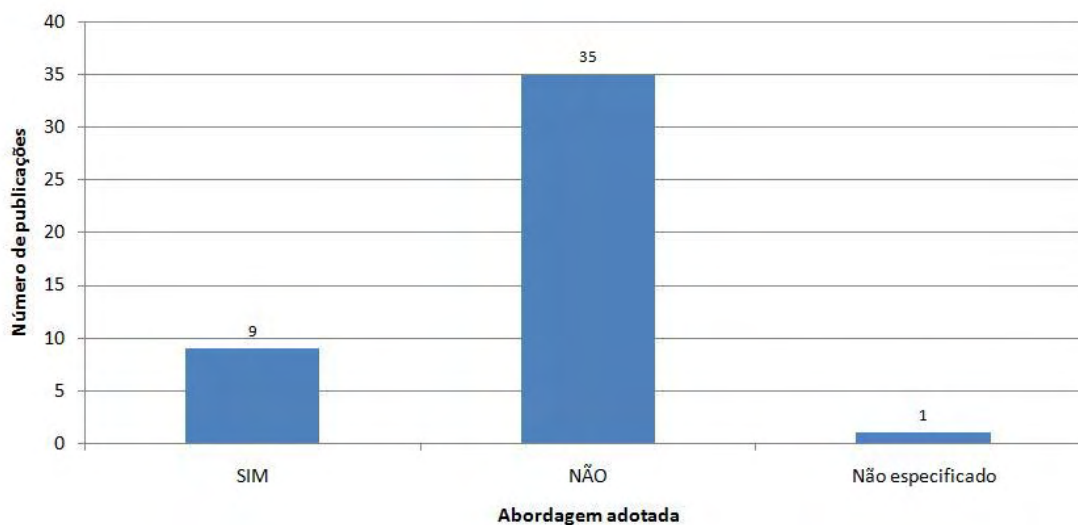


Figura 27 – Uso de terceiro conjunto de casos para validação – adaptado de Pontes *et al.* (2010a)

Na Figura 27 mostra-se que apenas 20% dos trabalhos fazem uso de um terceiro conjunto de dados. Um exemplo pode ser encontrado em Sharma *et al.* (2008). Nesse trabalho, 30 exemplos são usados para o treinamento, quatro para validação, e 17 são reservados para utilização como um conjunto de teste independente. Nota-se que, entre as publicações, outras técnicas para evitar o sobreajuste são raramente discutidas. Técnicas de regularização Bayesiana (FORESEE & HAGAN, 1997) foram utilizadas em apenas cinco casos (EZUGWU *ET AL.*, 2005; ÖZEL & KARPAT, 2005; FREDJ & AMAMOU, 2006; ÖZEL *ET AL.*, 2007; BASHEER *ET AL.*, 2008). O uso de técnicas de poda (REED, 1993) não foi observado.

Uma importante contribuição para a modelagem por RNAs em processos de usinagem foi dada por Sick (2002) em sua revisão sobre a utilização de redes neurais para a monitoração do desgaste de ferramenta em processo de torneamento. Além das duas questões discutidas anteriormente, Sick listou alguns requisitos práticos importantes para a validação do modelo neural. O autor recomenda a repetição de experimentos com uma dada topologia de rede a fim de reduzir influências estocásticas ligadas, por exemplo, à inicialização aleatória de pesos, à aplicação de sequências aleatórias de apresentação de exemplos de treinamento ou a partições aleatórias de dados. Tais fatores modificam a condição inicial e a trajetória descrita por uma rede sobre a hiper-superfície de erro durante o treinamento, como visto em Haykin (1998),

o que introduz um componente aleatório no processo de ajuste. É necessário, segundo o autor, adquirir amostras estatisticamente significativas do desempenho de uma rede a fim de estimar corretamente seu desempenho. A repetição de experimentos pode ser claramente detectada em apenas oito publicações, incluindo Kohli & Dixit (2005), que efetuaram dez repetições para cada topologia de rede sob teste.

Outro requisito para validação do modelo é a estimação da exatidão e da dispersão do modelo neural e sua expressão sob a forma de uma distribuição estatística. Autores recomendam que modelos sejam entregues com parâmetros tais como erro médio, variância, intervalos de confiança e intervalos de predição confiável (SICK, 2002; BISHOP, 2007). Tal prática quase nunca é seguida. Raros exemplos são encontrados em Kohli & Dixit (2005) e Sonar *et al.* (2006). Outro ponto importante é o fato de que não apenas erros de rede, mas também o erro incluído nos dados usados para treinamento influencia a precisão do modelo global. Estimativas de erro na coleta de dados devem ser apresentadas juntamente com as das redes. Entre as 45 publicações revisadas, não se observa a adoção de tal prática.

A aplicação de testes estatísticos para a comparação entre redes com estruturas, entradas, ou treinamento por algoritmos distintos é outra condição para sustentar as conclusões obtidas em predição por redes neurais. Como mencionado por Sick (2002), "uma decisão a favor ou contra um determinado número de neurônios na camada oculta deve ser baseada em um número adequado de repetições dos experimentos e na comparação dos resultados dessas repetições por meio de um teste estatístico". A prática é também recomendada pelo autor para comparações entre os modelos de RNAs e modelos obtidos por outros métodos. Em apenas dez publicações, ou aproximadamente 22% do total, a aplicação de testes estatísticos pôde ser observada. Uma boa abordagem é encontrada na obra de Al-Ahmari (2007). Para comparar os erros de predição da rugosidade de modelos obtidos por análise de regressão, metodologia de superfície de resposta e redes neurais no torneamento duro de aço austenítico AISI 302, o autor empregou testes *t* emparelhado e testes *F*. O mesmo autor comparou a dispersão de cada modelo por meio de testes de Levene.

Outras práticas básicas de neurocomputação básicas são negligenciadas entre os trabalhos revisados. A inicialização dos pesos é um ponto importante para a

convergência de redes, como observou Haykin (1998). O autor recomenda que pesos sejam uniformemente distribuídos dentro de um pequeno intervalo. Apenas cinco publicações mencionam a abordagem adotada para inicializar pesos. Em Bernardos & Vosniakos (2002), Mandal *et al.* (2007) e El-Sonbaty *et al.* (2008), a regra de Nguyen-Widrow (NGUYEN & WIDROW, 1990) é empregada, em acordo com as práticas recomendadas. Em Tsao & Hocheng (2008), os pesos são inicializados com o mesmo valor, e em Jesuthanam *et al.* (2007), um algoritmo genético é empregado para definir pesos iniciais. Além da inicialização, autores recomendam que a ordem de apresentação de exemplos de treinamento seja feita em ordem aleatória ao longo das épocas de treinamento. Não se encontra menção clara de tal prática nos trabalhos revisados.

3.7 CONCLUSÕES

A análise de cada categorização leva a algumas conclusões. A média de trabalhos revisados saltou de 1,4 no período entre 2000 e 2004 para 7,6 durante o período de 2005 a 2009. Há predominância de trabalhos que lidam com torneamento e fresamento. Os dois processos representam 71% de todos os trabalhos encontrados sobre o assunto. Redes usadas para predição da rugosidade possuem, em 75% dos casos, entre 3 e 5 entradas. Velocidade de corte, avanço e profundidade de corte são os indicadores mais utilizados para a modelagem da rugosidade. Juntos, eles respondem por 67% de todas as entradas utilizadas para a predição, enquanto os demais preditores estão espalhados em 26 categorias distintas. Os resultados de experimentos planejados segundo o DOE ou experimentos similares a DOE são empregados em 62% dos casos. Isto ocorre principalmente porque os pesquisadores em geral aproveitam resultados experimentais obtidos anteriormente para treinar e selecionar redes, e então comparar modelos obtidos por métodos dos quadrados mínimos e modelos neurais. Outra conclusão é de que redes de uma única saída são as mais utilizadas (64% do total). A rugosidade média (R_a) é a grandeza preferida para expressão da rugosidade, aparecendo em 54% das obras estudadas.

A predominância das redes MLP, que aparecem em 86% dos artigos analisados,

já foi mencionada. Outra constatação é a preferência dada ao algoritmo BP para o treinamento de MLPs. O algoritmo foi empregado em mais de metade (53%) dos estudos que fazem uso da arquitetura. Em 10% dos artigos que adotam a arquitetura MLP, o algoritmo utilizado para o treinamento não é especificado, o que impossibilita a reprodução do estudo. Um resultado surpreendente é o de que 42% das redes são projetadas por tentativa e erro e que esforços sistemáticos de otimização podem ser claramente observados em apenas 7% dos trabalhos.

Conclui-se também que as abordagens usadas para interromper o treinamento de MLPs variam bastante. Nenhuma abordagem individual atinge mais de 18% das obras que envolvem esse tipo de rede. Chama a atenção que a técnica utilizada não seja mencionada em 36% dos artigos que usam MLPs. Fica clara também a prevalência do uso de figuras de erro quadrático para seleção de redes. MSE, RMSE e SSE totalizam 57% do total. Grandezas baseadas em erro médio são as preferidas para expressar a exatidão final dos modelos de RNAs. Expressões do erro médio são adotadas em 55% das obras investigadas. Uma conclusão final é a de que a utilização de conjuntos distintos de treinamento e seleção é explicitamente adotada na grande maioria (84%) dos trabalhos revisados, em acordo com teoria da neurocomputação.

4 OTIMIZAÇÃO DE REDES NEURAIIS BASEADA EM PROJETO DE EXPERIMENTOS

4.1 CONSIDERAÇÕES INICIAIS

Este capítulo contém a apresentação do método para otimização de redes neurais para a predição da rugosidade com uso da metodologia de Projeto de Experimentos (DOE). O trabalho foi motivado pelos resultados obtidos por Pontes (2006), no qual a viabilidade da aplicação do DOE no projeto de redes RBF para predição da vida da ferramenta e da rugosidade foi investigada. Foi motivado também pelo ineditismo do tema e pela relevância da questão do projeto sistemático de redes, conforme exposto no capítulo contendo a revisão sobre o uso de redes para a predição em usinagem.

A questão de pesquisa a ser respondida é se um método de otimização baseado no Projeto de Experimentos pode identificar redes mais exatas na predição que as redes obtidas por métodos em uso na literatura, superiores a redes obtidas por meio de soluções especialistas utilizadas em pacotes computacionais de uso comercial e superiores a modelos de predição baseados em mínimos quadrados. O capítulo inicia com a delimitação do escopo do método, recapitula as etapas do trabalho de pesquisa, detalha o método para otimização e apresenta a proposta para sua validação, cujos resultados serão detalhados em capítulo posterior.

4.2 ESCOPO DA PROPOSTA

O método proposto se insere dentro do processo de elaboração de modelos proposto por Montgomery *et al.* (2008). Mais especificamente, o método é colocado como uma abordagem para implementação das fases de Seleção e Adequação do Modelo e de Validação do Modelo, quando o objetivo for a predição da rugosidade em processos de usinagem e a técnica de modelagem escolhida for uma rede neural de arquitetura MLP ou RBF.

Dadas as propriedades de armazenagem de conhecimento e de generalização das redes, e dada a não-linearidade característica da geração da rugosidade em processo de

usinagem, a aplicação do método não se restringe somente aos processos tratados ao longo deste trabalho, podendo ser aplicado também à predição da rugosidade em outros processos de usinagem, não exemplificados.

Além disso, o método não abrange a seleção dos preditores específicos de usinagem a serem utilizados para predição da rugosidade (a fase de Coleta de Dados, na concepção de Montgomery), deixando aberta ao praticante tal escolha. Mantém-se desta forma a generalidade de método, sem restringir sua aplicação aos processos de manufatura incluídos como exemplo. Deve-se ressaltar, porém, que o conhecimento do processo e a escolha adequada de preditores é fundamental para o sucesso de qualquer estratégia de modelagem da rugosidade por redes neurais. Autores do campo da neurocomputação, como Haykin (2008), afirmam que o conjunto de preditores deve ser formado por variáveis de processo independentes e não correlacionadas.

4.3 EVOLUÇÃO DO TRABALHO

4.3.1 CONCEPÇÃO DO MÉTODO ITERATIVO

A idéia inicial do projeto foi a de iniciar por um experimento exploratório com objetivo de identificar os fatores influentes no desempenho de redes RBF e a partir daí aplicar a metodologia de superfície de resposta (MSR) na busca por topologias ótimas de rede. Por meio da técnica da descendente mais inclinada (ou *steepest descent*, no original em inglês), seriam realizados experimentos sucessivos que se encerrariam ao chegar a um ponto no qual não se observasse melhoria no desempenho da rede. Ao final da descendente, um experimento planejado segundo a técnica do arranjo composto centralizado, ou CCD (do original em inglês *Central Composite Design*), seria realizado e por meio de uma ferramenta de otimização (como o Solver do Excel, por exemplo) a configuração de rede com erro mínimo seria encontrada.

A primeira tarefa foi a definição dos fatores e níveis a serem utilizados, e a definição do tipo de arranjo experimental a realizar durante os experimentos exploratórios. Arranjos de Taguchi, fatoriais mistos e fatoriais completos foram cogitados, bem como diversas combinações de variáveis. Depois de alguma reflexão

optou-se por partir diretamente para a descendente mais inclinada. O conjunto de treinamento utilizado foi o conjunto com 24 casos de treinamento e 6 casos de seleção correspondente a uma curva sintética utilizada por Pontes (2006). O conjunto de teste, composto por 120 casos, foi proveniente do mesmo trabalho. Diversas tentativas foram feitas, variando-se o número de fatores de projeto das redes entre dois e quatro (número de unidades radiais, a proporção entre o número de casos de treinamento e o número de casos de teste, valor do fator de abertura do algoritmo isotrópico para determinação do parâmetro de largura da função radial e o algoritmo de para localização dos hiper-centros das unidades radiais). O número de níveis de cada fator variou entre dois e três. Outro ciclo experimental completo seria conduzido para investigação do efeito de outro algoritmo para determinação do parâmetro de largura da função radial, o algoritmo dos K-mais próximos. Desta forma, rodou-se um arranjo experimental, calculou-se a direção da descendente mais inclinada e iniciou-se a busca pela rede com melhor desempenho. As primeiras tentativas apontaram para uma direção de melhoria, mas observou-se que o passo adotado para a descendente estava muito curto. Novas tentativas foram feitas com passo maior, com o objetivo de chegar mais rapidamente ao final da descendente. O resultado deste ciclo experimental é mostrado na Figura 28.

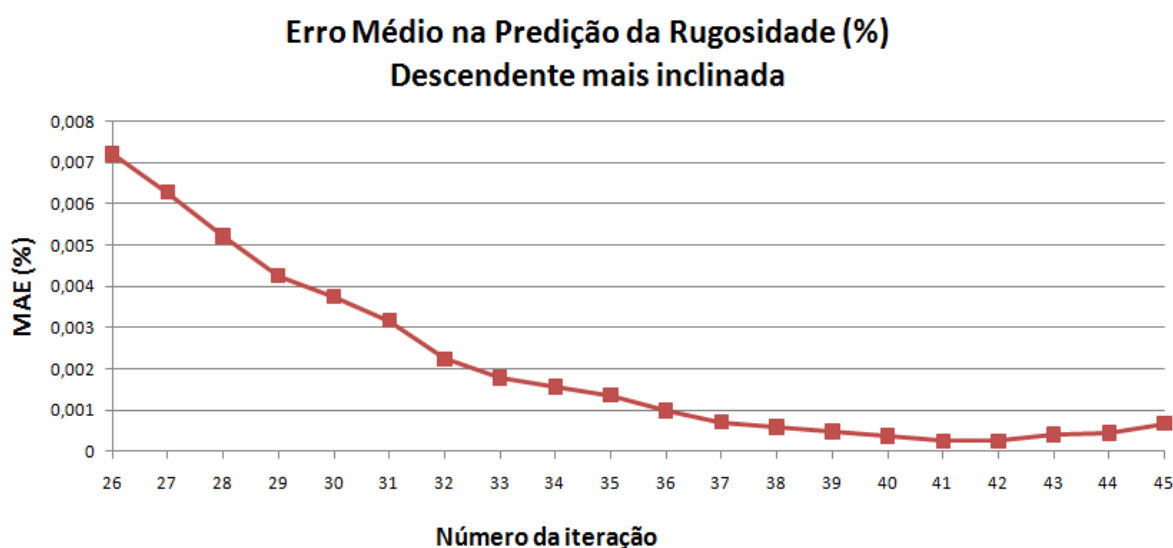


Figura 28 – Erro em predição de curvas sintéticas com método da Descendente mais inclinada

Os resultados obtidos foram promissores, mas a análise de variância deste experimento revelou ser a curvatura significativa em diversas execuções do arranjo, o que sugere a existência de múltiplos mínimos locais. Algumas outras tentativas foram feitas, aumentando-se a amplitude das unidades reais entre os níveis dos fatores, com objetivo de reduzir a sensibilidade do arranjo à curvatura. Isso eliminou a significância da curvatura, mas fez com que os resíduos se tornassem não-normais. Novas tentativas levaram à identificação de amplitude entre os níveis dos fatores que não levavam a mínimos locais e produziam resíduos normais. O cálculo da descendente mais inclinada, porém, apontava em direção contrária àquela que havia obtido bons resultados anteriormente, conduzindo a resultados ruins.

Os resultados insatisfatórios durante os experimentos levaram ao questionamento do uso da MSR para a otimização de redes neurais. Eles sugeriam que a superfície de erro da rede varia de forma irregular com a mudança dos parâmetros da rede, o que desqualifica o MSR como opção de otimização. Verificou-se que o mesmo problema havia sido detectado no trabalho de Siva *et al.* (2009), em um trabalho relacionado a processo de soldagem. Abandonou-se pois, esta abordagem.

A ideia que surgiu a seguir foi utilizar a técnica EVOP como ferramenta para localização da região de ótimo das redes. EVOP baseia-se na migração do centro dos arranjos na direção da melhoria, e prescinde do conceito de curvaturas para sua aplicação. A estratégia, neste caso, seria aplicar repetidamente arranjos fatoriais, deslocando o ponto central do arranjo na direção da melhoria, conforme indicado pela análise do experimento. Uma vez localizada a região do mínimo, aplicar-se-ia um arranjo CCD, com número reduzido de fatores e com centro localizado na melhor configuração indicada pelo EVOP. Se fosse comprovada a proximidade do ponto de mínimo (por meio da análise de significância da curvatura), utilizar-se-ia uma ferramenta de otimização (como o Solver, por exemplo) para localização do ponto de ótimo.

A primeira tentativa com a utilização do EVOP foi realizada utilizando-se níveis fixos de variação entre os fatores. Este esquema mostrou ser viável na busca pela configuração de erro mínimo, mas levou à execução de um grande número de iterações sem sinal de finalização.

Fez-se uma modificação no algoritmo com objetivo de realizar-se um EVOP em múltiplas fases. A primeira iniciava com diferenças amplas entre os níveis dos fatores. O objetivo era cobrir uma grande faixa de valores com número reduzido de iterações. Ao chegar-se ao ponto em que se deixava de observar melhoria, uma nova fase EVOP era iniciada, centrada na melhor configuração da fase anterior e reduzindo-se as diferenças entre os níveis dos fatores a uma fração da diferença anterior. A significância dos fatores sobre o desempenho da rede se reduzia à medida que diminuía a amplitude entre os níveis. As fases prosseguiam até um ponto em que nenhum fator era apontado pela análise de variância como estatisticamente significativo para o desempenho da rede. Neste ponto encerrava-se a otimização.

Como um exemplo, a aplicação deste método iniciou a primeira fase com o estabelecimento de um intervalo entre os níveis do fator parâmetro de largura do algoritmo isotrópico igual a 50. Na segunda fase, os intervalos foram reduzido a 10, e em seguida a 2 (os demais fatores também variaram proporcionalmente). Tal estratégia se mostrou bastante eficaz na busca pela região com erro mínimo. Após dois ou três ciclos o erro da rede nos casos de validação caía abruptamente. Nos ciclos posteriores, o desempenho das redes seguia melhorando incrementalmente.

Um novo problema ocorreu, todavia, ao realizar-se o experimento CCD para comprovar a localização do ponto de mínimo. Os autovalores apresentaram sinais opostos, indicando a existência de um ‘ponto de sela’, como mostrado na Figura 29.

Superfície de Erro - Experimento CCD preliminar

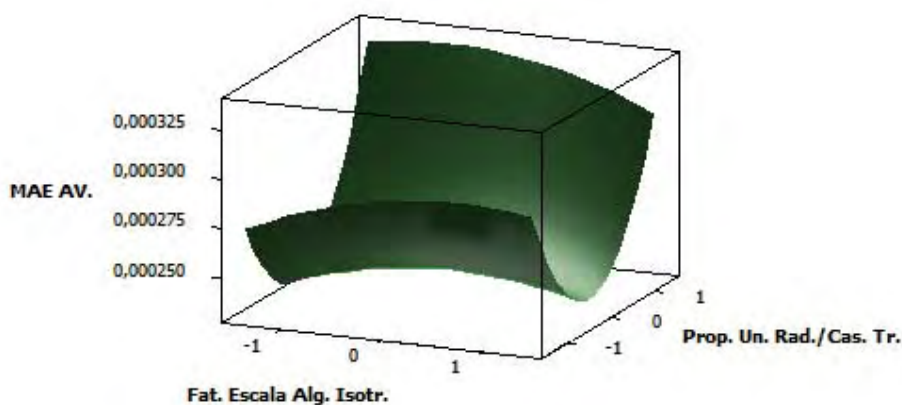


Figura 29 – Superfície em forma de sela obtida dos experimentos preliminares usando a técnica CCD

Aventou-se a possibilidade de que a ocorrência da ‘sela’ fosse devida ao fato da amplitude entre os níveis dos fatores ter sido muito reduzida na etapa final do EVOP e do arranjo CCD ter sido construído com os mesmos intervalos reduzidos. Novos arranjos CCD foram conduzidos, aumentando-se a amplitude dos intervalos entre os níveis dos fatores, centrados no ponto de mínimo encontrado após o final da última fase do EVOP. Observou-se que a ampliação dos intervalos entre os níveis dos fatores levou a dois autovalores positivos (indicando a existência de um ponto de mínimo). O resultado pode ser visualizado na Figura 30.

Superfície de Erro Experimentos CCD com região ampliada

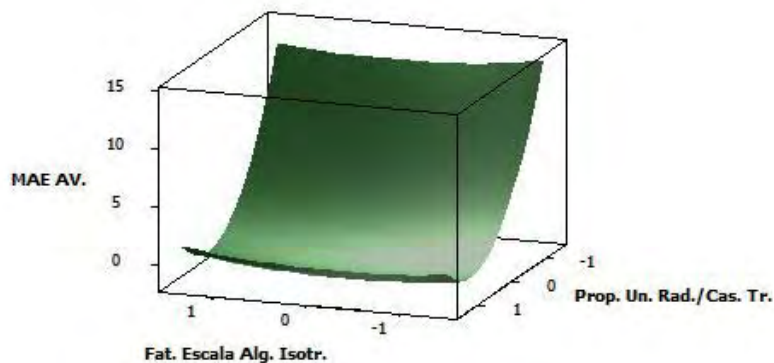


Figura 30 – Superfície com indicação de região de ótimo obtida dos experimentos ampliados usando a técnica CCD

Na Figuras 29 e 30, o eixo vertical indica o valor do erro médio absoluto em porcentagem. Na Figura 31 são mostradas as curvas de nível correspondentes à superfície da Figura 30. A região mais escura corresponde à região da ocorrência de erro mínimo.

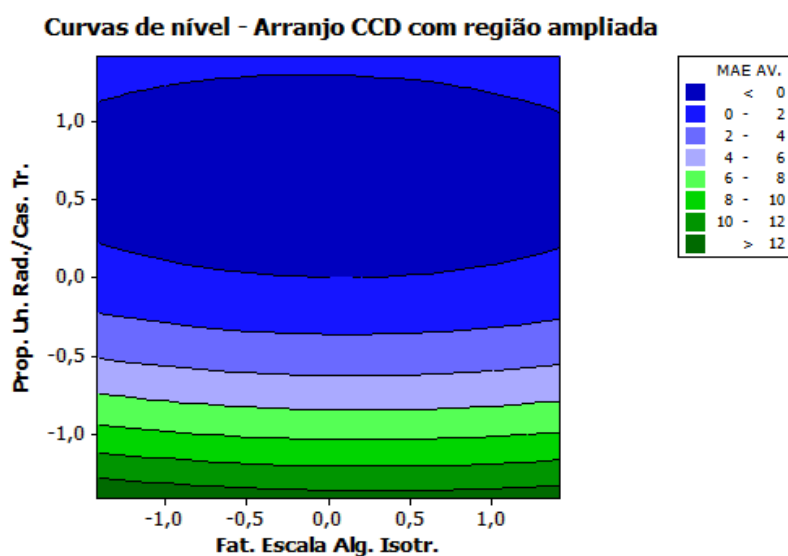


Figura 31 – Curvas de nível com indicação de região de ótimo obtida dos experimentos ampliados usando a técnica CCD

Infelizmente, porém, o indicador de falta de ajuste da regressão das curvas quadráticas obtidas com os arranjos CCD ampliados resultou muito elevado, revelando um ajuste pobre. Além disso, a comparação simples entre os pontos de mínimo obtidos pelo primeiro CCD (que levou ao ponto de sela) e o ponto revelado pelo Solver para o segundo CCD (que identificou um ponto de mínimo) revelou que as configurações apontadas eram diferentes e possuíam desempenho significativamente distinto. O desempenho da melhor configuração no primeiro CCD (em que ocorreu a ‘sela’) apresentava melhor desempenho do que o ponto sugerido pelo Solver no segundo CCD.

Ficou claro novamente que a irregularidade da superfície de erro da rede torna difícil a identificação de um único ponto de ótimo. Com objetivo de eliminar a dúvida, fez-se um experimento de saturação variando-se em intervalos de uma unidade todos os fatores em uma faixa ampla ao redor da região identificada como sendo a região do ótimo. A superfície de erro obtida no experimento de saturação é mostrada na Figura 32.

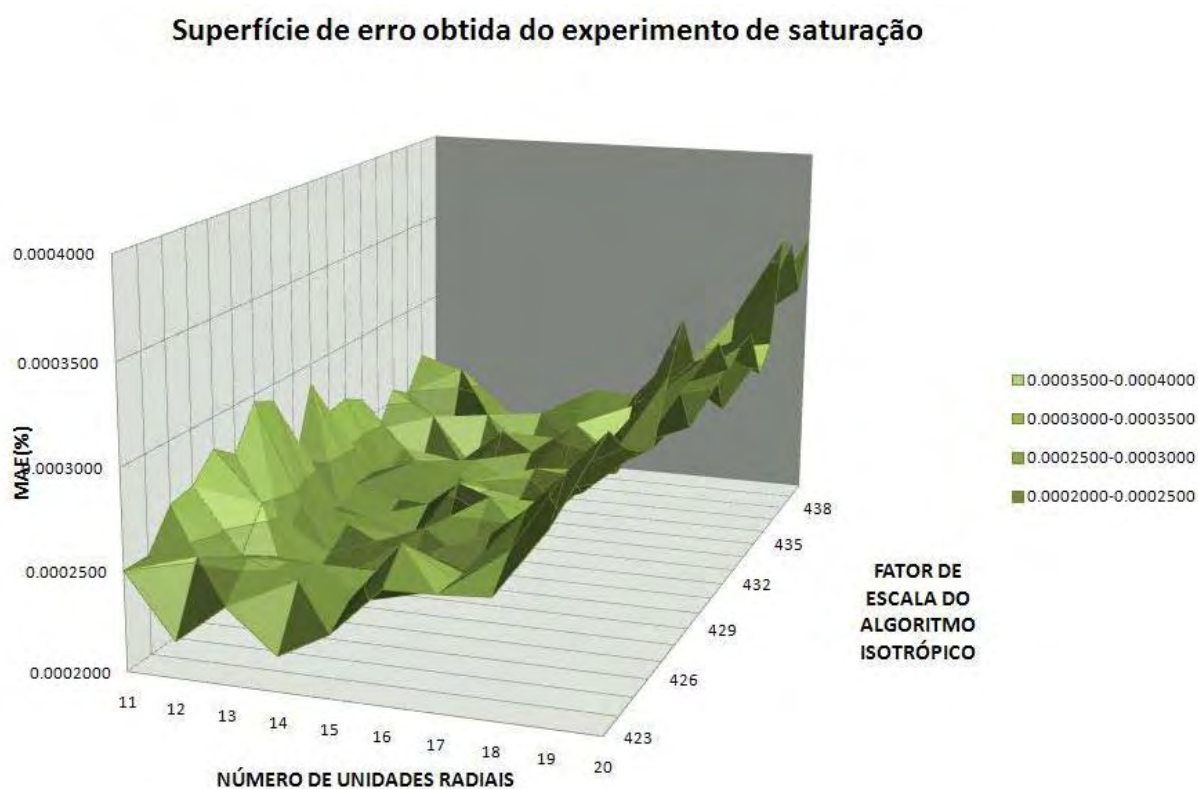


Figura 32 – Superfície de erro obtida do experimento de saturação

Este experimento consumiu bastante tempo. O resultado, porém, confirmou o que se intuía dos resultados dos experimentos anteriores:

- a superfície de erro do conjunto das arquiteturas de redes RBF é extremamente irregular, apresentando múltiplos mínimos locais;
- ao invés de um único ponto de mínimo, existe uma vasta de região de erros de predição reduzidos, dentro da qual variações das configurações individuais apresentam diferenças estatisticamente insignificantes de desempenho.

Tais resultados mostraram que, para a otimização de redes, a busca por uma região de ótimo é mais relevante que a busca por uma configuração individual com desempenho ótimo. Alterou-se desta forma o foco do trabalho, que passou a se concentrar na busca sistemática por um domínio de parâmetros de projeto que levassem à minimização do erro, abandonando-se a noção de identificar uma única configuração ótima.

Dentro dessa nova percepção, a estratégia mais eficaz para a busca da região do ótimo consistiu no uso do EVOP em múltiplas fases, com redução dos intervalos entre os níveis dos fatores ao se chegar a uma região na qual uma iteração indique não haver melhoria significativa no desempenho da rede. As fases EVOP se repetem utilizando este conceito até que nenhum fator apresente influência significativa sobre o desempenho da rede neural. A aplicação deste método demonstrou resultados excelentes.

O método foi então testado substituindo-se o algoritmo Isotrópico para determinação do parâmetro de largura pelo algoritmo dos K-mais próximos. O método convergiu e atingiu a região de erro mínimo como esperado, com bons resultados, embora inferiores aos demonstrados pelo algoritmo Isotrópico. Em mais uma rodada de testes, todo o processo foi repetido utilizando-se outro conjunto de treinamento, desta vez contendo 12 casos de treinamento e 3 de seleção (o conjunto de testes foi o mesmo, contendo 120 casos). Novamente o método convergiu. O algoritmo dos K-mais próximos, porém, apresentou resultado bastante inferior ao obtido pelo algoritmo Isotrópico, em decorrência da redução do número de casos de treinamento.

4.3.2 O PROBLEMA COM REDES DE ARQUITETURA RBF

O passo seguinte à definição do método e à aplicação do mesmo a dados simulados seria a verificação de sua aplicabilidade a dados obtidos da literatura. Esperava-se que o método conduzisse aos mesmos bons resultados obtidos para as curvas sintéticas. Não foi o que se observou, no entanto. As redes obtidas apresentavam valores elevados de erro, superiores aos obtidos pelos estudos que originaram os casos de treinamento e teste.

Este resultado pode ser atribuído à maior representatividade dos dados obtidos de operações reais na comparação com pontos gerados a partir de um modelo sintético, como foi o caso. As curvas sintéticas apresentam uma não-linearidade menor e seus dados não são afetados por fatores de ruído do processo, tais como potência e rigidez da máquina, temperatura de operação, dentre outras.

Sendo o desempenho das redes RBF insatisfatório, o trabalho foi reorientado no sentido de utilizar redes MLP para a predição da rugosidade. Como a arquitetura de redes MLP envolve um número bem maior de parâmetros que uma rede RBF, pensou-se em uma estratégia envolvendo a realização de um experimento exploratório que permitisse identificar os níveis de fatores qualitativos que beneficiassem a capacidade de predição da rede. Os níveis dos fatores qualitativos seriam então fixados nos valores recomendados pelo experimento exploratório. A otimização dos fatores quantitativos seria realizada utilizando-se a mesma abordagem baseada no EVOP.

Para tal experimento exploratório, decidiu-se pela aplicação de um arranjo misto L16 de Taguchi (TAGUCHI *ET AL.*, 1987), que permite explorar grande número de fatores com uma quantidade econômica de experimentos. O experimento foi planejado com uso um arranjo externo no qual foram acomodados dois outros fatores. Tal seleção baseou-se na estratégia utilizada por Balestrassi *et al.* (2009). Fatores e níveis utilizados nos arranjos interno e externo do experimento são mostrados nos Quadros 6 e 7. Nos quadros, BP indica algoritmo de retropropagação, LM indica o algoritmo Levenberg-Marquadt, CG indica o algoritmo do gradiente conjugado e QP indica o algoritmo Quick-Propagation. O experimento foi útil também para a estimativa de níveis razoáveis para a inicialização do EVOP para os fatores quantitativos.

Quadro 6 – Fatores e níveis do arranjo interno durante experimento exploratório com redes MLP

Código do fator	Descrição do fator	Nat. do fator	No. de níveis	Descrição dos níveis
P1	Alg. treinamento fase 1	Qual.	4	BP, LM, CG, QP
P2	Alg. treinamento fase 2	Qual.	4	Nenhum, LM, CG, QN
EP1	No. de épocas fase 1	Quant.	2	50, 100
H1	No. de neurônios na camada oculta 1	Quant.	2	0.2*No.casos treinamento, 0.5*No. casos treinamento
H2	No. de neurônios na camada oculta 2	Quant.	2	0, 0.5*H1
OF	Função de regressão de saída (função de ativação)	Qual.	2	Linear, logística sigmóide
LR	Taxa de aprendizado	Quant.	2	0.01, 0.1
IM	Método de inicialização	Qual.	2	Uniforme(0,1), Normal(0,1)
PU	Poda de neurônios	Qual.	2	Sem poda, poda de pesos inferiores a 0.05
PI	Poda de entradas	Qual.	2	Sem poda, poda de entradas com baixa influência após treinamento (taxa=1)

Quadro 7 – Fatores e níveis do arranjo externo durante experimento exploratório com redes MLP

Código do fator	Descrição do fator	Nat. do fator	No. de níveis	Código dos níveis
W1	Regularização de decaimento de peso fase 1	Qual.	2	Sem regularização, Reg. com fator decaimento=0.01
W2	Regularização de decaimento de peso fase 2	Qual.	2	Sem regularização, Reg. com fator decaimento=0.01

Alguns parâmetros qualitativos de projeto de redes de arquitetura MLP foram fixados nos melhores níveis apontados pelos resultados dos experimentos exploratórios. Tais tratamentos foram utilizados durante o restante da pesquisa. Os parâmetros fixados e respectivos tratamentos adotados foram os seguintes:

- Função de regressão de saída (função de ativação): logística sigmóide;
- Treinamento em duas fases; o algoritmo de treinamento da fase 1: retro-propagação (ou *Backpropagation*);
- Algoritmo de treinamento da fase 2: Gradiente Conjugado (ou *Conjugate Gradient*);
- Inicialização dos pesos: distribuição gaussiana com média igual a 0 e desvio-padrão igual a 1;
- Poda de entradas com pouca influência: sem poda;
- Poda de neurônios com pesos de saída: poda de pesos inferiores a 0.05;
- Regularização de decaimento de pesos nas fases 1 e 2 do treinamento: sem regularização.

O pré-processamento de entrada dos dados da rede foi realizado com o uso do algoritmo Minimax (MANUAL STATISTICA, 2005) com parâmetro inferior igual a 0 e superior igual a 1. Este algoritmo escalona os dados de tal modo que o menor valor do conjunto de treinamento é convertido para o valor do parâmetro inferior, o maior valor do conjunto de treinamento é convertido para o valor do parâmetro superior, e os demais pontos são distribuídos linearmente entre estes extremos. O pós-processamento dos dados foi realizado usando-se o algoritmo Média/Desvio Padrão (MANUAL STATISTICA, 2005) com parâmetros ‘Média’ igual a 0 e ‘Desvio Padrão’ igual a 1. Este algoritmo escalona linearmente valores, de modo que um valor de saída igual ao especificado pelo parâmetro ‘Média’ é convertido para o valor médio do conjunto de treinamento, enquanto que um desvio padrão para valores de saída é convertido para um desvio padrão do conjunto de treinamento.

Uma vez fixados os fatores qualitativos das redes MLP e definidos quais fatores quantitativos seriam sujeitos à otimização, retomou-se a aplicação do método aos dados obtidos de trabalhos da literatura. Observou-se desempenho sensivelmente superior ao obtido pelas redes RBF.

4.3.3 A DEFINIÇÃO DA VARIÁVEL DE SAÍDA

Enfrentou-se ainda outra questão para o aperfeiçoamento do método. A variável de saída escolhida inicialmente (e utilizada para análise dos experimentos de otimização de redes) foi a média do erro médio absoluto das redes expresso em porcentagem, chamada de MAE%.

A análise criteriosa dos resultados de predição obtidos pelas redes, porém, revelou outro fato para qual não se havia atentado anteriormente. Topologias individuais de rede têm seus pesos inicializados de maneira aleatória segundo uma determinada distribuição estatística (no caso deste estudo, uma distribuição Gaussiana, ou normal). São também submetidas à apresentação de exemplos em ordem aleatória. Estas variáveis introduzem um efeito estocástico na hiper-trajetória seguida pela rede sobre a hiper-superfície de erro durante o treinamento. Isto fez com que o desempenho final de uma mesma topologia de rede apresente uma distribuição de erros de predição em torno de um valor médio, ao invés de um único valor médio do erro.

Foi percebido que havia uma distinção entre a minimização da média do erro das redes, que se perseguia até então, e a proposta do projeto de conceber um método para identificar um domínio de parâmetros de rede que minimizem o erro. Em vários casos, redes com média de erro maior, porém com dispersão também maior, levaram à identificação de redes com menor erro na predição da rugosidade. Em outras palavras, minimizar o erro médio de predição não é a estratégia mais adequada para identificação de redes ótimas. Tal observação sugere haver interação entre a média e a dispersão na otimização das redes.

Esta constatação alterou a percepção que se tinha do problema. O objetivo foi então alterado da busca pela minimização de médias de erro na predição da rugosidade para uma busca topologias de rede apresentando uma distribuição de valores extremos mínimos que fosse favorável à minimização dos erros na predição da rugosidade. Com o objetivo de adequar o trabalho a esta nova percepção, redefiniu-se a variável de saída a ser minimizada como sendo a média dos erros médio absolutos do decil inferior do conjunto dos resultados das sequências inicialização-treinamento-seleção-validação conduzidas com uma determinada topologia de rede, expressa em porcentagem, e

referenciada por MAE-DI%. As réplicas dos arranjos experimentais foram igualmente redefinidas. Uma réplica foi definida como a repetição independente da sequência inicialização-treinamento-seleção-validação de uma topologia de rede por 100 vezes. A variável de saída foi coletada para análise da seguinte forma:

- colocou-se em ordem crescente os erros médios absolutos em porcentagem (MAE%) de predição dos casos de validação obtidos pela topologia de rede durante as 100 repetições que compõem uma réplica;
- tomou-se da lista ordenada os 10 menores valores de MAE% (o decil inferior);
- calculou-se o erro médio absoluto do decil inferior (MAE-DI%);
- transportou-se o valor da média do decil inferior para a ferramenta de análise de experimentos, como saída do experimento da linha correspondente à réplica.

Uma vez computados e transcritos os resultados de uma réplica, realizava-se a análise dos experimentos, com o objetivo de detectar os valores dos fatores que minimizassem a variável de saída.

A solução foi testada em experimentos preliminares e garantiu a convergência do método e a obtenção de erros de predição muito reduzidos, sendo então adotada como variável de saída pelo restante da pesquisa. Esta nova abordagem motivou inclusive a retomada da aplicação do método para a otimização de redes de arquitetura RBF, em adição à sua aplicação a redes MLP, utilizando a nova variável. Completou-se desta forma o aperfeiçoamento do método.

4.4 APRESENTAÇÃO E DETALHAMENTO DO MÉTODO

4.4.1 CONCEITO

No trabalho realizado por Pontes (2006) a aplicabilidade da utilização do DOE para o projeto de redes RBF foi estudada. Os resultados indicaram um bom potencial

para uso do Projeto de Experimentos e indicaram a possibilidade da otimização de redes por meio da metodologia.

O método proposto consiste em um algoritmo iterativo. Partindo de um conjunto de configurações iniciais de rede, a técnica EVOP (Operações Evolucionárias) é utilizada na busca por uma configuração de rede que apresente uma distribuição de erro na predição da rugosidade que favoreça a ocorrência de erro mínimo de predição da rugosidade, para um dado conjunto de treinamento. O método é proposto para redes de arquitetura MLP e RBF. Os parâmetros de projeto das redes foram utilizados como os fatores experimentais do EVOP.

A otimização ocorre em fases caracterizadas pela amplitude dos intervalos entre os níveis dos fatores. O objetivo de uma fase é identificar a rede mais apropriada usando-se a amplitude escolhida. Ao final de cada fase, a amplitude entre os níveis é reduzida. O processo se repete até que a análise de variância indique não haver fator estatisticamente significativo para a otimização da rede, ou antes, caso o usuário considere satisfatório o nível de otimização alcançado após uma fase e interrompa a otimização.

A definição dos níveis iniciais de cada fator do experimento foi elaborada com base em experimentos exploratórios, de modo a garantir a convergência do método em um número razoável de iterações. Os fatores que não foram sujeitos à otimização foram também fixados com base no resultado de experimentos exploratórios.

4.4.2 DEFINIÇÕES

Tratamento - Dentro do conceito proposto para otimização de redes, cada tratamento corresponde um determinado conjunto de parâmetros de projeto de rede.

Arranjo experimental - conjunto de experimentos da metodologia DOE utilizado para identificar a melhor configuração de rede para distintos tratamentos. Para o propósito deste trabalho, foram utilizados arranjos fatoriais completos com duas réplicas.

Ciclo - Um ciclo experimental corresponde à sequência de atividades composta por definição de níveis dos fatores, montagem, execução e análise do arranjo

experimental.

Fase - Uma fase corresponde a um conjunto de ciclos ao longo dos quais os intervalos entre os níveis dos fatores permanecem constantes. Ao final de uma fase, refinam-se os intervalos entre os níveis dos fatores ou encerra-se a execução do algoritmo, caso sejam satisfeitas as condições para tal.

4.4.3 RECOMENDAÇÕES BIBLIOGRÁFICAS OBSERVADAS

A revisão da bibliografia sobre o tema revelou diversas categorias de problemas relativos aos métodos estatísticos e aos conceitos básicos de neurocomputação ao longo das várias publicações revisadas. Com o objetivo de evitar tais problemas neste trabalho, uma série de cuidados foi tomada no planejamento e execução dos experimentos com redes neurais.

Em relação aos princípios da neurocomputação, foram levadas em conta recomendações de autores da área, como Haykin (2008) e Bishop (2007), dentre outros. Diversos cuidados experimentais devem ser seguidos para que a execução do método respeite os princípios da neurocomputação e produza resultados consistentes. Tais cuidados foram adotados neste trabalho e são os seguintes:

- A inicialização dos pesos da rede foi feita com valores reduzidos e com pequena dispersão;
- O conjunto de dados de treinamento foi dividido em três sub-conjuntos (treinamento, seleção e teste), sendo que o conjunto de seleção foi utilizado juntamente com a técnica da interrupção precoce para detecção de sobre-aprendizado;
- Utilizou-se validação cruzada (troca aleatória de casos entre os sub-conjuntos de teste e seleção durante o treinamento) ao longo das épocas de treinamento;
- Dentro de cada ciclo experimental, a ordem de aplicação de um determinado tratamento foi aleatória;

- Em cada ciclo experimental, cada tratamento foi aplicado de maneira independente por 200 vezes; isso significa que cada configuração de rede teve seus pesos inicializados, foi treinada, selecionada e submetida a um experimento de confirmação de maneira independente por aquele número de vezes.
- Cada conjunto de 100 repetições foi utilizado para computação da variável de saída de uma réplica do arranjo.
- Em cada repetição, a ordem de apresentação dos exemplos de treinamento foi aleatória;
- Os experimentos de confirmação corresponderam à etapa de validação do modelo e utilizaram os casos do conjunto de testes, que não foram utilizados durante o treinamento e seleção das redes;
- Utilização de métodos de regularização de pesos e poda.

A mencionada repetição da aplicação de um tratamento cumpre o objetivo de obter uma amostra estatisticamente significativa do comportamento de uma dada configuração, dado que a mesma rede exibe uma distribuição dos erros de predição em decorrência dos fatores estocásticos associados à inicialização e ao treinamento. Este elevado número de repetições foi possível graças à utilização de recursos computacionais que permitiram a configuração rápida dos experimentos e sua repetição automática. Um ciclo experimental típico consome aproximadamente 15 minutos.

Em relação aos requisitos para validação estatística de trabalhos que aplicam redes neurais para a predição de rugosidade em processos de usinagem, foram levadas em contas as recomendações feitas por Sick (2002).

- Análises de resultados e comparações entre resultados de distintas configurações de rede foram feitas utilizando-se testes estatísticos;
- Comparações entre resultados de predição obtidos por redes e resultados obtidos por outros métodos foram realizadas utilizando-se testes estatísticos;
- O modelo resultante foi expresso em termos de seus parâmetros

estatísticos;

- Procurou-se (sempre que possível) incluir entre os resultados o erro associado aos conjuntos de treinamento à expressão final do erro do modelo.

Além dos cuidados mencionados, cumpre reafirmar que o método foi baseado no uso da metodologia do Projeto de Experimentos (DOE) que prevê a otimização simultânea dos fatores. Incorporou-se ao trabalho, desta maneira, o estudo dos efeitos das interações entre os níveis dos fatores utilizados no experimento, e evitou-se a armadilha da otimização de um fator por vez.

4.4.4 DEFINIÇÃO DOS FATORES E ARRANJOS EXPERIMENTAIS

4.4.4.1 Fatores para experimentos com redes MLP

Os fatores utilizados para a otimização de redes de arquitetura MLP são os indicados no Quadro 8.

Quadro 8 – Fatores utilizados para a otimização de redes MLP

Código do fator	Descrição do fator	Unidade
H1	Número de neurônios na camada oculta 1	Não aplicável
H2	Número de neurônios na camada oculta 2	Não aplicável
E1	Número de épocas de treinamento na fase 1	Não aplicável
LR	Taxa de aprendizado na fase 1	Não aplicável
E2	Número de épocas de treinamento na fase 2	Não aplicável

4.4.4.2 Arranjo experimental para redes MLP

O arranjo experimental utilizado para a otimização de redes MLP é mostrado no Quadro 9.

Trata-se de um arranjo fatorial completo com cinco fatores e dois níveis, em duas réplicas, com dois pontos centrais. Conforme já mencionado, cada réplica consistiu na repetição independente do treinamento e seleção de um tratamento (uma configuração de rede) por 100 vezes. O resultado de cada réplica foi computado a partir das observações obtidas das repetições. O Quadro 9 mostra o arranjo experimental utilizado durante o processo de otimização. Os níveis dos fatores estão codificados em -1, 0 e +1. Os valores dos níveis de cada fator em unidades não-codificadas são calculados a cada ciclo do método iterativo.

Quadro 9 – Arranjo fatorial completo utilizado durante a otimização de redes MLP

	Código do Fator				
	H1	H2	E1	LR	E2
1	-1	-1	-1	-1	-1
2	+1	-1	-1	-1	-1
3	-1	+1	-1	-1	-1
4	+1	+1	-1	-1	-1
5	-1	-1	+1	-1	-1
6	+1	-1	+1	-1	-1
7	-1	+1	+1	-1	-1
8	+1	+1	+1	-1	-1
9	-1	-1	-1	+1	-1
10	+1	-1	-1	+1	-1
11	-1	+1	-1	+1	-1
12	+1	+1	-1	+1	-1
13	-1	-1	+1	+1	-1
14	+1	-1	+1	+1	-1
15	-1	+1	+1	+1	-1
16	+1	+1	+1	+1	-1
17	-1	-1	-1	-1	+1
18	+1	-1	-1	-1	+1
19	-1	+1	-1	-1	+1
20	+1	+1	-1	-1	+1
21	-1	-1	+1	-1	+1
22	+1	-1	+1	-1	+1
23	-1	+1	+1	-1	+1
24	+1	+1	+1	-1	+1
25	-1	-1	-1	+1	+1
26	+1	-1	-1	+1	+1
27	-1	+1	-1	+1	+1
28	+1	+1	-1	+1	+1
29	-1	-1	+1	+1	+1
30	+1	-1	+1	+1	+1
31	-1	+1	+1	+1	+1
32	+1	+1	+1	+1	+1
33	0	0	0	0	0

4.4.4.3 Fatores para experimentos com redes RBF

Os fatores utilizados para a otimização de redes de arquitetura RBF são os indicados no Quadro 10:

Quadro 10 – Fatores utilizados para a otimização de redes RBF

Código do fator	Descrição do fator	Unidade
ISO	Fator de escala do algoritmo Isotrópico para determinação do parâmetro de largura da função radial	Não aplicável
UR	Número de unidades radiais	Não aplicável
CEN	Algoritmo para determinação do hiper-centro das unidades radiais - fator qualitativo, com dois níveis: sub-amostragem (SS) e K-Médias (KM)	Não aplicável

4.4.4.4 Arranjo experimental para redes RBF

Observando-se os fatores envolvidos na otimização de redes RBF, nota-se que um dos fatores (CEN) é um fator qualitativo, o que contraria a premissa de evolução do ponto central de operação que caracteriza o EVOP. Foi necessário encontrar uma solução para estimar a influência do algoritmo de determinação dos hiper-centros em conjunto com os fatores quantitativos.

A solução encontrada foi realizar dois arranjos para cada ciclo experimental envolvendo redes RBF. Um arranjo fatorial completo misto de dois fatores com três níveis, contendo os fatores quantitativos, foi utilizado. Este arranjo foi executado duas vezes, cada uma utilizando um nível do tratamento qualitativo (SS ou KM). A cada ciclo foram analisados ambos os arranjos em busca da melhor configuração de rede.

Os experimentos foram realizados em duas réplicas. Assim como para redes MLP, cada réplica consistiu na repetição independente do treinamento e seleção de um tratamento (uma configuração de rede) por 100 vezes e o resultado de cada réplica foi computado a partir das observações obtidas das repetições. O Quadro 11 mostra o arranjo experimental utilizado durante o processo de otimização. Os níveis dos fatores

estão codificados em 1, 2 e 3. Os valores reais são calculados a cada ciclo do método iterativo.

Quadro 11 – Arranjo fatorial completo - otimização de redes RBF

	Código do Fator	
	ISO	UR
1	1	1
2	1	2
3	1	3
4	2	1
5	2	2
6	2	3
7	3	1
8	3	2
9	3	3

4.4.5 ALGORITMO GERAL

O algoritmo geral para otimização das redes neurais para a predição da rugosidade em processos da usinagem é mostrado a seguir. Escreveu-se o algoritmo em linguagem estruturada (FARRER *ET AL.*, 1999). Nesta notação, variáveis de trabalho são registradas em negrito. Instruções matemáticas são representadas por símbolos e diretivas são notadas por palavras escritas em letras maiúsculas. A instrução representada por $A \leftarrow B$ significa “variável **A** recebe o valor de **B**”, onde **B** pode representar o valor de uma variável ou um valor explícito. A instrução representada por $A \leftarrow A + I$ significa “**A** é incrementada em uma unidade”. Decisões a serem tomadas ao longo do algoritmo são representadas pelos blocos de palavras reservadas *SE-SENÃO-FIM SE*. As instruções incluídas entre a diretiva *SE* e a diretiva *SENÃO* são executadas caso a condição mostrada na linha *SE* seja satisfeita; caso contrário, são executadas as instruções incluídas entre a diretiva *SENÃO* e a diretiva *FIM SE*.

INÍCIO ALGORITMO**DECLARAR:*****Ponto Central do Arranjo,******Melhor Configuração Ciclo Atual,******Melhor Configuração da Fase,******Melhor Configuração Geral, registros compostos por 5 campos:****Número de neurônios na camada oculta 1, número inteiro não negativo;**Número de neurônios na camada oculta 2, número inteiro não negativo;**Número de épocas de treinamento na fase 1, número inteiro não negativo;**Taxa de aprendizado na fase 1, número real positivo ≥ 0.001 ;**Número de épocas de treinamento na fase 2, número inteiro não negativo;****Conjunto de Pontos Centrais de Arranjo, vetor de registros compostos por 5 campos:****Número de neurônios na camada oculta 1, número inteiro não negativo;**Número de neurônios na camada oculta 2, número inteiro não negativo;**Número de épocas de treinamento na fase 1, número inteiro não negativo;**Taxa de aprendizado na fase 1, número real positivo;**Número de épocas de treinamento na fase 2, número inteiro não negativo;****Intervalos Entre Níveis dos Fatores, registro composto por 5 campos:****Intervalo Entre Número de neurônios na camada oculta 1, número inteiro positivo;**Intervalo Entre Número de neurônios na camada oculta 2, número inteiro positivo;**Intervalo Entre Número de épocas de treinamento na fase 1, número inteiro positivo;**Intervalo Entre Taxa de aprendizado na fase 1, número real positivo;**Intervalo Entre Número de épocas de treinamento na fase 2, número inteiro positivo;*

CONTINUAÇÃO ALGORITMO

Níveis dos Fatores do Arranjo, registro composto por 10 campos:

Nível inferior Número de neurônios na camada oculta 1, número inteiro não negativo;

Nível inferior Número de neurônios na camada oculta 2, número inteiro não negativo;

Nível inferior Número de épocas de treinamento na fase 1, número inteiro não negativo;

Nível inferior Taxa de aprendizado na fase 1, número real positivo;

Nível inferior Número de épocas de treinamento na fase 2, número inteiro não negativo;

Nível superior Número de neurônios na camada oculta 1, número inteiro não negativo;

Nível superior Número de neurônios na camada oculta 2, número inteiro não negativo;

Nível superior Número de épocas de treinamento na fase 1 número inteiro não negativo;

Nível superior Taxa de aprendizado na fase 1, número real positivo;

Nível superior Número de épocas de treinamento na fase 2, número inteiro negativo;

Número Máximo de Fases,

Índice de Fase,

Índice de Ciclo, números inteiros positivos;

Final de Fase, binário (FALSO ou VERDADEIRO);

Melhor Resultado Ciclo Atual,

Melhor Resultado Ciclo Anterior,

Melhor Resultado da Fase,

Melhor Resultado Geral, números reais não negativos;

CONTINUAÇÃO ALGORITMO**Estabelecer Ponto Central do Arranjo;****Melhor Configuração Ciclo Atual \leftarrow Ponto Central do Arranjo;****Melhor Configuração da Fase \leftarrow Ponto Central do Arranjo;****Melhor Resultado Geral $\leftarrow \infty$;****Escolher Número Máximo de Fases;****PARA Índice de Fase $\leftarrow 1$ até Número Máximo de Fases REPITA:****Final de Fase \leftarrow FALSO;****Conjunto de Pontos Centrais de Arranjo \leftarrow nulo;****Melhor Resultado Ciclo Anterior $\leftarrow \infty$;****Melhor Resultado Ciclo Atual $\leftarrow \infty$;****Melhor Resultado da Fase $\leftarrow \infty$;****Estabelecer Intervalos Entre Níveis dos Fatores;****Índice de Ciclo $\leftarrow 1$;****ENQUANTO Final de Fase = FALSO REPITA:****Armazenar Ponto Central do Arranjo no Conjunto de Pontos Centrais de Arranjo;****Estabelecer Níveis dos Fatores do Arranjo;****Executar ciclo experimental de treinamento, seleção e validação das redes;****Tabular resultados dos testes de validação;****Analisar resultados dos testes de validação;****SE pelo menos um fator significativo para a redução do erro FAÇA:****Determinar Melhor Configuração Ciclo Atual;****Melhor Resultado Ciclo Atual \leftarrow resultado da Melhor Configuração Ciclo Atual;****SE Melhor Resultado Ciclo Atual < Melhor Resultado Ciclo Anterior FAÇA:****Melhor Configuração da Fase \leftarrow Melhor Configuração Ciclo Atual;****Melhor Resultado da Fase \leftarrow Melhor Resultado Ciclo Atual**

O algoritmo inicia pela declaração das variáveis de trabalho e por sua inicialização. Neste ponto declaram-se as estruturas de dados que serão manipuladas para a obtenção do resultado.

Observa-se que o algoritmo inclui dois laços de repetição (dois *loops*). O laço externo, limitado pelas palavras reservadas *PARA* e *FIM PARA*, corresponde às fases do algoritmo. Ao longo de cada fase o intervalo entre os níveis dos fatores é constante. O laço interno, indicado pelas palavras reservadas *ENQUANTO* e *FIM ENQUANTO*, corresponde aos ciclos que integram cada fase, ao longo dos quais o centro do arranjo vai sendo deslocado, segundo a técnica do EVOP, em direção a uma configuração de rede com erro mínimo.

Após a declaração das variáveis, o algoritmo inicia a primeira fase, entrando no laço externo. As condições para o final de fase e os valores de referência para o erro da rede são inicializados e inicia-se a execução do primeiro ciclo, no laço interno.

Em cada ciclo um arranjo experimental é montado, executado e analisado. Se algum fator for considerado como estatisticamente significativo pela análise, a melhor configuração e o valor de seu erro de predição são armazenados. O valor do erro de predição é comparado ao menor erro de predição já encontrado. Se houver melhoria do resultado (ou seja, se houver redução no valor do erro em relação ao menor erro encontrado anteriormente) a melhor configuração do ciclo é armazenada como a melhor configuração de rede para predição. O ponto central do arranjo é deslocado para uma posição calculada com base na melhor configuração de rede encontrada e outro ciclo se inicia, com o arranjo experimental construído ao redor do novo ponto central.

Os ciclos se repetem enquanto houver melhoria do resultado ou até que o ponto central do arranjo calculado para o próximo ciclo seja um ponto já utilizado como ponto central em algum ciclo anterior da mesma fase. Quando qualquer destas condições ocorrer, encerra-se a fase. Os intervalos entre os níveis dos fatores são reduzidos, o histórico de pontos centrais de arranjos da fase é apagado e outra fase se inicia, ao longo da qual os valores reduzidos dos intervalos entre os níveis dos fatores serão utilizados.

As fases se sucedem (repetições do laço externo) até que o número máximo de

fases seja alcançado, quando o algoritmo é finalizado. O processo também pode ser encerrado antes de atingir-se o número máximo de fases, caso a análise de variância realizada em um determinado ciclo indique que nenhum fator é estatisticamente significativo para a redução do erro da rede. Isto pode ocorrer à medida que os intervalos entre os níveis do fatores são reduzidos.

Ao final da execução do algoritmo, a configuração de rede armazenada como a melhor rede para a predição é apontada como o melhor modelo neural para predição da rugosidade. Seus parâmetros estatísticos armazenados são utilizados para caracterização do modelo neural para a predição.

O algoritmo listado anteriormente é aplicável à otimização de redes neurais de arquitetura MLP. Para a otimização de redes RBF, é necessário substituir a declaração de algumas variáveis, como se mostra a seguir. Além disso, em cada ciclo, é necessário executar duas vezes o arranjo experimental (uma com cada nível do fator CEN) e escolher a melhor configuração com base no resultado da análise de ambos os arranjos.

DECLARAR:

Ponto Central do Arranjo, registro composto por 2 campos:

Fator de Escala Algoritmo Isotrópico, número inteiro positivo;

Número de unidades radiais, número inteiro positivo;

Melhor Configuração Ciclo Atual,

Melhor Configuração da Fase,

Melhor Configuração Geral, registros compostos por 3 campos:

Fator de Escala Algoritmo Isotrópico, número inteiro positivo;

Número de unidades radiais, número inteiro positivo;

Algoritmo Determinação Hiper-Centros, binário (SS ou KM);

Conjunto de Pontos Centrais de Arranjo, vetor de registros compostos por 2 campos:

Fator de Escala Algoritmo Isotrópico, número inteiro positivo;

Número de unidades radiais, número inteiro positivo;

DECLARAR:

Intervalos Entre Níveis dos Fatores, registro composto por 2 campos:

Intervalo Entre Fator de Escala Algoritmo Isotrópico, número inteiro positivo;

Intervalo Entre Número de unidades radiais, número inteiro positivo;

Níveis dos Fatores do Arranjo, registro composto por 6 campos:

Nível inferior Fator de Escala Algoritmo Isotrópico, número inteiro positivo;

Nível inferior Número de unidades radiais, número inteiro positivo;

Nível médio Fator de Escala Algoritmo Isotrópico, número inteiro positivo;

Nível médio Número de unidades radiais, número inteiro positivo;

Nível superior Fator de Escala Algoritmo Isotrópico, número inteiro positivo;

Nível superior Número de unidades radiais, número inteiro positivo;

Diversas operações listadas ao longo algoritmo estão expressas em alto nível e representam na verdade uma série de instruções. Tais operações correspondem a sub-rotinas do algoritmo, e são detalhadas nas seções seguintes.

4.4.6 DETALHAMENTO DAS SUB-ROTINAS

4.4.6.1 Estabelecer *Ponto Central do Arranjo*

Para redes neurais de arquitetura MLP, *Ponto Central do Arranjo* deve ser inicializado como mostrado a seguir. Os valores foram escolhidos com base nos resultados de experimentos exploratórios, em conjunto com os intervalos iniciais dos níveis dos fatores, por propiciarem o crescimento das redes em direção à configuração ótima em um número razoável de ciclos.

Ponto Central do Arranjo:

Número de neurônios na camada oculta 1 ← 8;

Número de neurônios na camada oculta 2 ← 8;

Número de épocas de treinamento na fase 1 ← 32;

Taxa de aprendizado na fase 1 ← 0.4;

Número de épocas de treinamento na fase 2 ← 64;

Para redes neurais de arquitetura RBF, *Ponto Central do Arranjo* deve ser inicializado como mostrado a seguir:

Ponto Central do Arranjo:

Fator de Escala Algoritmo Isotrópico ← 64

Número de unidades radiais ← $0.5 * \text{número de casos de treinamento}$

Onde *número de casos de treinamento* é número de exemplos incluído no subconjunto de treinamento da rede neural. Caso o $0.5 * \text{número de casos de treinamento}$ não resulte em um número inteiro positivo, *Número de unidades radiais* deverá ser arredondado para o inteiro positivo mais próximo.

4.4.6.2 Estabelecer *Intervalos Entre Níveis dos Fatores*

Para redes neurais de arquitetura MLP, *Intervalos Entre Níveis dos Fatores* devem ser inicializados como mostrado a seguir. Os valores foram escolhidos com base nos resultados de experimentos exploratórios, por propiciarem a cobertura de uma ampla faixa de valores em um número reduzido de ciclos.

SE Índice de Fase = 1 FAÇA:

Intervalos Entre Níveis dos Fatores:

Intervalo Entre Número de neurônios na camada oculta 1 ← 8

Intervalo Entre Número de neurônios na camada oculta 2 ← 8

Intervalo Entre Número de épocas de treinamento na fase 1 ← 32

Intervalo Entre Taxa de aprendizado na fase 1 ← 0.4

Intervalo Entre Número de épocas de treinamento na fase 2 ← 64

SENÃO FAÇA:

Intervalos Entre Níveis dos Fatores:

Intervalo Entre Número de neurônios na camada oculta 1 ←

*Intervalo Entre Número de neurônios na camada oculta*0.5;*

Intervalo Entre Número de neurônios na camada oculta 2 ←

*Intervalo Entre Número de neurônios na camada oculta 2*0.5;*

Intervalo Entre Número de épocas de treinamento na fase 1 ←

*Intervalo Entre Número de épocas de treinamento na fase 1*0.5;*

Intervalo Entre Taxa de aprendizado na fase 1 ←

*Intervalo Entre Taxa de aprendizado na fase 1*0.5;*

Intervalo Entre Número de épocas de treinamento na fase 2 ←

*Intervalo Entre Número de épocas de treinamento na fase 2*0.5;*

FIM “SE Índice de Fase (...)”;

Para redes neurais de arquitetura RBF, *Ponto Intervalos Entre Níveis dos Fatores* devem ser inicializados como mostrado a seguir:

SE Índice de Fase = 1 FAÇA:

Intervalos Entre Níveis dos Fatores:

Intervalo Entre Fator de Escala Algoritmo Isotrópico ← 32

Intervalo Entre Número de unidades radiais ← 0.1 * número de casos de treinamento

SENÃO FAÇA:

Intervalos Entre Níveis dos Fatores:

Intervalo Entre Fator de Escala Algoritmo Isotrópico ← 32

*Intervalo Entre Fator de Escala Algoritmo Isotrópico**0.5;

Intervalo Entre Número de unidades radiais ← 0.1 * número de casos

*Intervalo Entre Número de unidades radiais**0.5;

FIM “SE Índice de Fase (...)”;

Onde *número de casos de treinamento* é número de exemplos incluído no subconjunto de treinamento da rede neural. Caso o intervalo calculado não resulte em um número inteiro, *Intervalo Entre Número de unidades radiais* deverá ser arredondado para o inteiro positivo mais próximo.

Observe-se também que *Intervalo Entre Número de unidades radiais* é definido como inteiro positivo, e logo, o valor mínimo para o fator é igual a 1.

4.4.6.3 Estabelecer *Níveis dos Fatores do Arranjo*

No início de cada ciclo experimental, os *Níveis dos Fatores do Arranjo* devem ser estabelecidos como mostrado a seguir.

PARA cada fator REPITA:

Nível inferior do fator ←

(Ponto central do fator – Intervalo Entre Nível do Fator)

Nível superior do fator ←

(Ponto central do fator + Intervalo Entre Nível do Fator)

FIM “PARA cada fator (...);

4.4.6.4 Executar ciclo experimental de treinamento, seleção e validação das redes

Os ciclos experimentais devem ser executado segundo a seqüência mostrada a seguir:

PARA cada combinação possível entre Níveis dos Fatores do Arranjo REPITA:

Construir uma rede neural utilizando a combinação de fatores experimentais e os tratamentos definidos para os parâmetros fixos;

FIM “PARA cada combinação (...);

Em ordem aleatória, PARA cada rede neural construída REPITA:

PARA Índice ← 1 até 200 REPITA:

Inicializar os pesos da rede neural;

Treinar a rede neural usando a técnica da interrupção precoce, apresentando os exemplos do sub-conjunto de treinamento em ordem aleatória;

Aplicar a rede neural treinar para predizer a rugosidade para os casos incluídos no sub-conjunto de validação;

Calcular e armazenar os erros médios absolutos (MAE%) para o conjunto de validação;

FIM “PARA Índice (...);

FIM “Em ordem aleatória, PARA cada rede (...);

4.4.6.5 Tabular resultados dos testes de validação

Os resultados obtidos dos testes de validação das redes segundo a seqüência mostrada a seguir:

PARA cada rede neural construída REPITA:

Armazenar resultados das predições de 1 a 100 em Réplica 1;

Armazenar resultados das predições de 101 a 200 em Réplica 2;

Colocar em ordem crescente os erros de predição da Réplica 1;

Colocar em ordem crescente os erros de predição da Réplica 2;

Média Réplica 1 ← média dos 10 menores erros MAE% da Réplica 1;

Média Réplica 2 ← média dos 10 menores erros MAE% da Réplica 2;

Na primeira réplica do arranjo, preencher a célula corresponde ao resultado do tratamento com Média Réplica 1;

Na segunda réplica do arranjo, preencher a célula corresponde ao resultado do tratamento com Média Réplica 2;

FIM “PARA cada rede (...)”;

4.4.6.6 Analisar resultados dos testes de validação

A análise dos arranjos fatoriais deve ser feita com utilização da técnica da análise de variância, ou ANOVA (MONTGOMERY & RUNGER, 2007). Os valores-p (*p-values*) resultantes da análise devem ser utilizados para a determinação dos fatores e interações significativos para a redução do erro da rede. Além dos valores-p, o princípio da Hierarquia (MONTGOMERY, 2001) deve ser levado em conta na determinação da significância de um fator. Os valores dos efeitos dos fatores e interações significativos devem ser armazenados para determinar qual a melhor configuração de rede.

4.4.6.7 Determinar *Melhor Configuração Ciclo Atual*

A determinação da *Melhor Configuração Ciclo Atual* deve seguir o procedimento padrão para análise de experimentos, conforme a seqüência a seguir:

Para cada fator ou interação significativos, recuperar o valor absoluto do efeito do fator ou interação;

***Melhor Configuração Ciclo Atual** ← Calcular combinação de níveis dos fatores que minimize a soma algébrica dos efeitos de fatores e interações significativos.*

Sendo a variável de saída uma representação do erro das redes na predição da rugosidade, deve calcular a combinação de fatores que minimize algebricamente a soma dos efeitos de fatores e interações significativos. O cálculo é feito utilizando uma técnica padrão de otimização de funções lineares, o método Simplex (BAZARAA *ET AL.*, 2004). A combinação de níveis dos fatores resultante da minimização dos efeitos pelo Simplex corresponde à melhor configuração de rede do ciclo experimental.

4.4.6.8 Estabelecer novo *Ponto central do Arranjo*

Uma vez finalizado um ciclo experimental, o *Ponto central do Arranjo* para o ciclo seguinte deve ser estabelecido como mostrado a seguir.

PARA cada fator FAÇA:

SE Análise de variância indica que fator deve ser decrementado FAÇA:

SE Índice de Ciclo = 1 FAÇA:

Ponto central do fator ←

(Ponto central do fator – Intervalo Entre Nível do Fator)

SENÃO FAÇA:

SE Ponto central do fator – Intervalo Entre Nível do Fator não transgredir o limite do Nível inferior do fator FAÇA:

Ponto central do fator ←

(Ponto central do fator – Intervalo Entre Nível do Fator)

FIM “SE Ponto central (...);

FIM “SE Índice de Ciclo (...)”

SENÃO FAÇA:

SE Análise indica que fator deve ser incrementado FAÇA:

Ponto central do fator ←

(Ponto central do fator + Intervalo Entre Nível do Fator)

FIM “SE Análise indica (...) incrementado”;

FIM “SE Análise indica (...) decrementado”

FIM “PARA cada fator (...);

No algoritmo apresentado, se o resultado da análise de variância recomendar uma redução no nível do fator, deve-se proceder como segue:

- Se o ciclo experimental for o primeiro ciclo de uma fase, o ponto central do fator para o próximo ciclo recebe o resultado do valor atual menos o intervalo entre níveis definido para o fator em questão;

- Se o ciclo experimental não for o primeiro de uma fase, deve-se verificar se a subtração do ponto central atual do fator pelo intervalo entre níveis do fator não transgredir o limite inferior definido para o fator. Se não houver transgressão, o ponto central é decrementado; caso contrário, o ponto central para o fator em questão não é alterado;

Se, por outro lado, o resultado da análise de variância recomendar um incremento no nível do fator, o ponto central do fator para o próximo ciclo recebe o resultado do valor atual mais o intervalo entre níveis definido para o fator em questão.

No caso de redes RBF, para os quais um arranjo de três níveis é utilizado para os fatores quantitativos, pode ocorrer também a situação da análise não recomendar qualquer mudança de nível de um fator para o próximo ciclo. Neste caso, nenhuma das condições lógicas será satisfeita, e o ponto central do arranjo para o fator permanecerá inalterado. Desta forma o teste '*SE Ponto Central do Arranjo \notin Conjunto de Pontos Centrais de Arranjo*' do Algoritmo Geral resultará falso (pois sendo o novo ponto central igual ao anterior, já estará contido no conjunto de pontos centrais), o que ocasionará o fim da fase. A próxima fase, neste caso, iniciará a partir do mesmo ponto central.

4.4.6.9 Reduzir *Intervalos Entre Níveis dos Fatores*

Intervalos Entre Níveis dos Fatores devem ser reduzidos ao final de cada fase como mostrado a seguir.

PARA cada fator definido como inteiro não negativo OU inteiro positivo FAÇA:

*SE $(0.5 * \text{Intervalo Entre Níveis do Fator}) > 1$ FAÇA:*

Intervalo Entre Níveis do Fator \leftarrow

*$0.5 * \text{Intervalo Entre Níveis do Fator}$*

SENÃO FAÇA:

Intervalo Entre Níveis do Fator $\leftarrow 1$;

*FIM “SE $0.5 * \text{Intervalo} (...)$ ”*

Intervalo Entre Níveis do Fator;

FIM “PARA cada fator definido como inteiro(...);

PARA cada fator definido como número real positivo FAÇA:

Intervalo Entre Níveis do Fator \leftarrow

*$0.5 * \text{Intervalo Entre Níveis do Fator}$*

FIM “PARA cada fator definido como número real (...);

4.5 VALIDAÇÃO DO MÉTODO

O método proposto foi validado por meio de três etapas de pesquisa experimental. A validação seguiu a estratégia de utilizar dados inicialmente disponíveis para a concepção e refinamento do método, evoluindo em direção a dados reais coletados de experimentos de usinagem. Para os experimentos de otimização de redes neurais, utilizou-se o pacote computacional Statistica®.

A primeira etapa consistiu na aplicação do método a curvas sintéticas, já utilizadas anteriormente durante o trabalho de mestrado. A razão para a utilização de curvas sintéticas neste primeiro momento foi sua pronta disponibilidade. Por meio de seu uso, foi possível trabalhar no desenvolvimento do método ao mesmo tempo em que se fazia a revisão da literatura em busca de trabalhos contendo dados apropriados para a etapa seguinte de

validação. Para as curvas sintéticas, fez-se a modelagem somente com redes RBF. Foi feita comparação entre os resultados obtidos pela melhor topologia de rede e os resultados do estudo original.

A segunda etapa da validação foi composta pela aplicação do método aos resultados de artigos selecionados da literatura. Por conterem dados utilizados em aplicações reais, o uso destes artigos representou uma oportunidade valiosa de incrementar a maturidade e efetividade do método antes de investir na realização de um experimento real de usinagem. A modelagem foi feita por meio de redes MLP e RBF. Foram feitas comparações entre os resultados obtidos pelas duas arquiteturas, entre os resultados da melhor topologia de rede e os resultados do estudo original, entre os resultados da melhor topologia e os resultados da melhor rede obtida por uma solução especialista de um pacote computacional, e entre os resultados da melhor arquitetura de rede e os resultados de um modelo de mínimos quadrados, obtido por regressão múltipla. As comparações realizadas lançaram mão da utilização de testes estatísticos para fornecer suporte às conclusões obtidas.

A terceira e última etapa da validação do método consistiu em sua aplicação a resultados de experimentos reais de usinagem, conduzidos no laboratório da UNIFEI – Universidade Federal de Itajubá. Também neste caso modelagem também foi feita por meio de redes MLP e RBF. Foram feitas comparações entre os resultados obtidos pelas duas arquiteturas, entre os resultados da melhor topologia e os resultados da melhor rede obtida por uma solução especialista de um pacote computacional, e entre os resultados da melhor arquitetura de rede e os resultados de um modelo de mínimos quadrados, obtido por regressão. Assim como na segunda etapa de validação, lançou-se mão de testes estatísticos para o embasamento das conclusões obtidas.

5 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

5.1 CONSIDERAÇÕES INICIAIS

Este capítulo apresenta os resultados obtidos por meio da aplicação do método proposto. Inicialmente são caracterizados os trabalhos e experimentos de usinagem utilizados para a validação. Em seguida é mostrada a evolução das redes ao longo do processo de otimização, indicando-se os níveis dos fatores utilizados em cada ciclo e a trajetória descrita pela variável de saída. O melhor modelo resultante é especificado em termos de seus parâmetros e de seu desempenho. São conduzidas comparações com os resultados dos trabalhos originais, no caso das curvas sintéticas e dos trabalhos da literatura. Para os experimentos realizados com base no dados coletados da literatura e dos experimentos de usinagem, os resultados de predição das redes obtidas pela aplicação do método são comparados também aos resultados de predição obtidos por meio de regressão múltipla e por modelos de redes neurais obtidos por meio do IPS, uma ferramenta automática para otimização de redes neurais contida no pacote computacional Statistica®. São realizadas também comparações entre os desempenhos obtidos pelas melhores redes MLP e RBF. Em todos os casos, apenas os resultados de predição das redes para os casos de teste (ou seja, casos não utilizados para treinamento) são utilizados como material para a análise.

5.2 REALIZAÇÃO E ANÁLISE DOS EXPERIMENTOS

Um cuidado tomado durante a análise, segundo recomendações de vários autores (MONTGOMERY, 2001; BOX *ET AL.*, 1978; NOORDIN *ET AL.*, 2004), foi a análise dos resíduos das respostas das réplicas. Segundo Montgomery (2001), os resíduos devem ser normais, independentes (não-correlacionados) e equitativamente distribuídos. Após a análise feita para cada experimento foram analisados os resíduos da predição da rugosidade, por meio do teste de Anderson-Darling, ao nível de significância de 0,05. Para os casos em que o teste indicou que os resíduos de um dado experimento não seguiam a distribuição normal, fez-se a remoção dos *outliers*, repetiu-

se a análise dos efeitos e resíduos. A operação foi repetida até que os resíduos apresentassem distribuição normal. Ao final desta operação realizou-se a análise de variância. Para efeitos da análise aqui apresentada, foram considerados significativos fatores e interações cujo valor p obteve valor inferior ao nível de significância adotado, que foi igual a 0,05.

A ferramenta utilizada para a realização das análises estatísticas foi o pacote computacional Minitab®. A cada ciclo, transportou-se para a ferramenta os resultados dos experimentos com redes neurais e efetuou-se a análise do arranjo correspondente. Os resultados da análise determinaram, pela soma dos efeitos significativos (principais e de interação), a configuração de rede com menor erro de predição, que passava então a ser utilizada como base para o cálculo do centro do arranjo seguinte, conforme os algoritmos descritos no capítulo 4 . Nas Figuras 33, 34 e 35 mostra-se, a título de exemplo, um gráfico de efeitos principais, um gráfico dos efeitos de interação dupla e um gráfico de Pareto dos efeitos. As Figuras são resultado do primeiro ciclo do algoritmo de otimização de redes MLP utilizando os dados extraídos de Sarkar *et al.* (2006) para treinamento das redes.

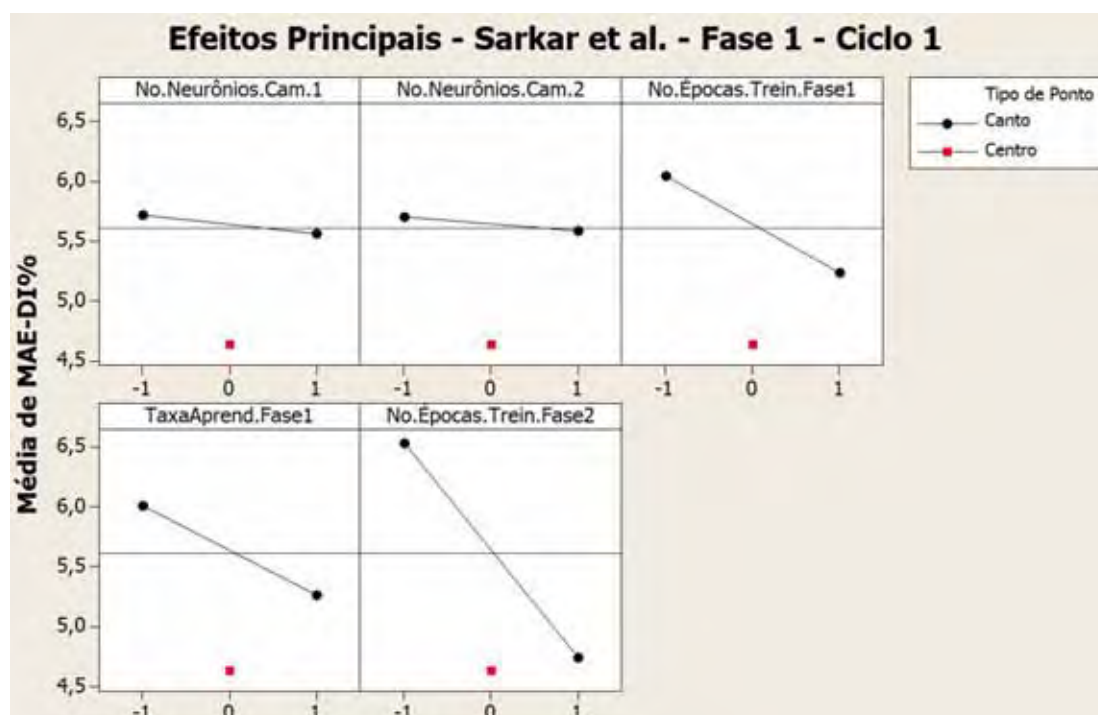


Figura 33 – Efeitos principais –primeiro ciclo de otimização de redes MLP para dados de Sarkar *et al.* (2006)

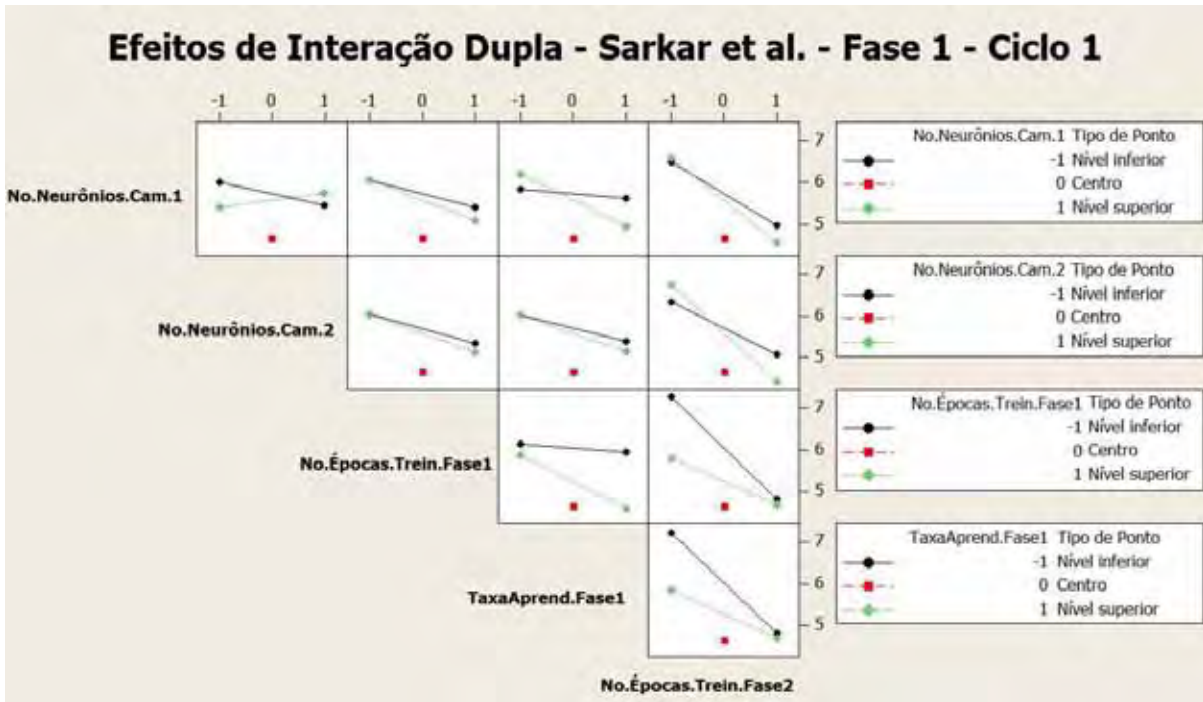


Figura 34 – Efeitos de interação dupla – primeiro ciclo de otimização de redes MLP para dados de Sarkar *et al.* (2006)

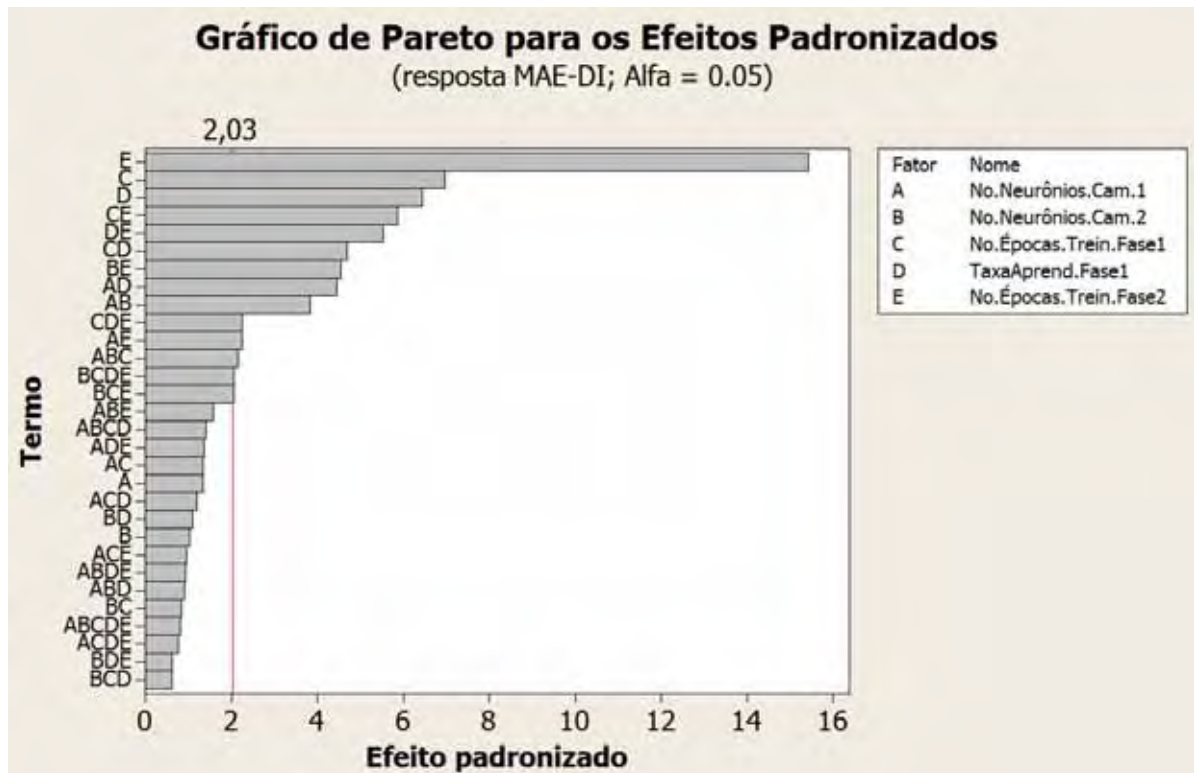


Figura 35 – Gráfico de Pareto – primeiro ciclo de otimização de redes MLP para dados de Sarkar *et al.* (2006) – Ciclo 1

Na Figura 33 são ilustrados os efeitos causados pela variação dos níveis dos fatores de entrada sobre os valores da variável de saída. Para cada fator experimental são mostrados, no eixo das abscissas, os valores codificados assumidos pelo fator durante o arranjo experimental (-1, 0 e 1). No eixo das ordenadas são mostrados os valores médios da variável de saída obtidos nos experimentos do arranjo realizados com o fator de entrada naquele nível. A significância do efeito de um fator é medida pelo teste de Análise de Variância.

Na Figura 34 são detalhados os fatores de interação dupla. As interações podem ser traduzidas como a amplificação, ou a inibição, do efeito da variação de um determinado fator sobre a variável de saída em decorrência do nível assumido por um outro fator. A capacidade de detectar interações entre fatores é um dos pontos fortes da metodologia de Projeto de Experimentos. Pode-se observar na figura os efeitos de interação dupla. No eixo horizontal são mostrados os valores codificados (-1, 0 e 1) assumidos durante o arranjo experimental pelo fator de entrada cujo nome está abaixo da célula do gráfico. Cada linha no interior de uma célula corresponde a um valor codificado (-1, 1) assumido pelo fator de entrada cujo nome está ao lado da célula do gráfico. Os pontos representam os valores médios da variável de saída quando os dois fatores (ao lado e abaixo da célula) estão em um determinado nível. A declividade das linhas indica em quanto o valor médio da variável de saída é afetado quando um fator (representado por uma linha) assume um nível e o outro fator (representado pelos pontos -1, 0 e 1) varia. Assim como para os efeitos principais, a significância do efeito de interação é medida pela Análise de Variância. Pode-se visualizar graficamente a importância de um efeito de interação pela comparação entre declividade das retas no gráfico de interações. Quanto mais significativo o efeito de interação sobre a saída, mais próximas da perpendicular estão as retas.

Para experimentos projetados com mais de dois fatores, existe a possibilidade da ocorrência de efeitos de ordem superior a dois. Na Figura 35 mostra-se o conjunto dos efeitos significativos para o exemplo em questão. O gráfico indica os efeitos principais e os efeitos de interação que são considerados como estatisticamente significantes pela Análise de Variância. São considerados como significantes para a redução do erro de predição, ao nível de significância de 0.05, aqueles efeitos que ultrapassam o valor

indicado pela linha vermelha. Pode-se notar como, além de efeitos principais e efeitos de interação dupla, alguns efeitos de interação tripla até mesmo de interação quádrupla têm efeito significativo sobre o erro de predição da rede. Além disso, segundo o princípio da hierarquia (MONTGOMERY, 2001), um fator de entrada deve ser considerado como significativo se qualquer efeito de interação no qual estiver envolvido for significativo, independentemente da significância do efeito principal do fator. O estabelecimento de todos os fatores de entrada significativos nos níveis que maximizam a soma de efeitos na direção da minimização do erro de predição da rede determina o hiper-ponto central para o ciclo experimental seguinte, conforme detalhado no capítulo 4.

5.3 CARACTERIZAÇÃO DOS TRABALHOS UTILIZADOS PARA VALIDAÇÃO

5.3.1 CURVAS SINTÉTICAS

As curvas sintéticas utilizadas correspondem a conjuntos de treinamento utilizados por Pontes (2006) em estudo no qual se avaliou a viabilidade da utilização do DOE para projetos de rede para predição da vida da ferramenta e da rugosidade. Os modelos em questão foram gerados a partir de um modelo polinomial quadrático completo obtido por Paiva *et al.* (2005) em experimento de torneamento do aço SAE/ABNT 52100 – 55 HRC. Foram utilizadas ferramentas com insertos de cerâmica mista ($Al_2O_3 + TiC$), classe Sandvik GC 6050, recoberta com TiN, com geometria ISO CNGA 120408 S01525. O experimento foi planejado e executado segundo a Metodologia da Superfície de Resposta combinada com a técnica do fatorial completo. Os valores de rugosidade contidos nas curvas sintéticas estão compreendidos entre $0,219 \mu m$ e $0,879 \mu m$. O modelo em questão apresentou elevado ajuste aos dados experimentais, o que motivou sua utilização. Os fatores de entrada utilizados para predição da rugosidade naquele trabalho são listados no Quadro 12. A variável de saída do modelo é a rugosidade média (R_a), em μm .

Quadro 12 – Fatores de entrada da curva sintética utilizada para validação do método

Código do fator	Descrição do fator	Unidade
V_c	Velocidade de corte	<i>m/min</i>
f	Avanço	<i>mm/v</i>
a_p	Profundidade de usinagem	<i>mm</i>

Foram realizados experimentos com dois conjuntos de dados para o treinamento. O primeiro contém 40 casos, dos quais 30 são usados como casos de treinamento e 10 como casos de seleção em cada época de treinamento. O segundo conjunto contém 20 casos, sendo 15 casos de treinamento e 5 casos de seleção em cada época de treinamento. O conjunto de teste é o mesmo para ambos os conjuntos de treinamento, e contém 90 casos. Os dados são mostrados no Apêndice A.

5.3.2 ARTIGOS DA LITERATURA

A seleção dos artigos da literatura usados na validação do método procurou abranger distintos processos de usinagem, diferentes tamanhos de conjuntos de treinamento e distintas abordagens para formação de tais conjuntos. O primeiro artigo selecionado, aqui referenciado como Artigo 1, foi o trabalho de Sharma et al. (2008). O objetivo do trabalho foi o de prever parâmetros tais como as forças de corte e rugosidade da superfície em processo de torneamento duro. O material utilizado foram corpos de prova de Adamita, com dureza igual a 467 HV. As ferramentas utilizadas continham insertos de carbetto (CCMT090304) e deposição por vapores químicos de TiCN (4 μm) e Al_2O_3 (8 μm). Os dados foram coletados em experimentos cujos arranjos são semelhantes aos usado em DOE. Os valores de rugosidade medidos durante os experimentos conduzidos pelos autores estão compreendidos entre 0,84 μm e 7,49 μm . Os resultados foram usados para a construção de um modelo de rede neural para predição da rugosidade. Os autores propuseram um método próprio para otimização da topologia de rede.

Os fatores de entrada utilizados para predição da rugosidade naquele trabalho são

mostrados no Quadro 13. A variável de saída do modelo, utilizada para fins de comparação, é a rugosidade média (R_a), em μm .

Quadro 13 – Fatores de entrada do conjunto de treinamento extraído do Artigo 1.

Fonte: Sharma *et al.* (2008)

Código do fator	Descrição do fator	Unidade
V_c	Velocidade de corte	m/min
f	Avanço	mm/v
a_p	Profundidade de usinagem	mm
AA	Ângulo de aproximação	°

O conjunto de treinamento contém 34 casos (sendo 30 de treinamento e 4 de seleção, a cada época) e o conjunto de teste, 17 casos. O Apêndice B detalha o conjunto de dados extraído do trabalho e utilizado para treinamento e testes das redes neurais.

O segundo artigo selecionado, designado como Artigo 2, foi o trabalho de Kohli & Dixit (2005). Um dos experimentos conduzidos pelos autores foi de torneamento a seco de aço carbono (0.35%) de dureza 130 HV, com 0.35% de carbono, por meio de ferramentas com insertos de carbetto. A máquina utilizada foi um torno HMT NH-26, com motor de 11kW, que possibilita a adoção de 23 velocidades entre 40 e 2040 *rpm*. O objetivo foi encontrar uma topologia de rede MLP para predição da rugosidade da peça. Os dados foram coletados em experimentos cujos arranjos semelhantes aos usado em DOE. Os valores de rugosidade medidos durante os experimentos conduzidos pelos autores estão compreendidos entre 1,79 μm e 6,74 μm . Os dados de entrada e os resultados daquele artigo foram utilizados para treinamento e teste das redes neurais projetadas segundo o método proposto.

Os fatores de entrada utilizados para predição da rugosidade no Artigo 2 são discriminados no Quadro 14. A variável de saída do modelo, que constitui o objetivo da predição, é a rugosidade média (R_a), em μm . O conjunto de treinamento utilizado contém 30 casos (sendo 21 casos de treinamento e 9 de seleção, a cada época), enquanto o conjunto de teste contém 80 casos. O Apêndice C lista o conjunto de dados

extraído do trabalho.

Quadro 14 – Fatores de entrada do conjunto de treinamento extraído do Artigo 2.

Fonte: Kohli & Dixit (2005)

Código do fator	Descrição do fator	Unidade
V_c	Velocidade de corte	m/min
f	Avanço	mm/v
a_p	Profundidade de usinagem	mm
A	Vibração radial	m/s^2

O terceiro estudo da literatura, doravante referenciado como Artigo 3, foi realizado por Sarkar *et al.* (2006), que trabalharam com o processo de usinagem por descarga elétrica (EDM) de uma liga de alumínio e titânio (Ti = 44.5%, Al = 2%, Cr = 2%, Nb = 0.3%). A dureza dos corpos de prova utilizados foi de 148 HV. A máquina usada foi um modelo Electra Supercut 734 Series 2000 com controle numérico. Os autores propuseram uma estratégia para a otimização do processo. Uma rede neural treinada pelo algoritmo de retro-propagação foi desenvolvida para modelar o processo de usinagem e então empregada para a geração de modelos multidimensionais para a identificação de parâmetros ótimos de processo. Os fatores de entrada utilizados para predição da rugosidade no trabalho são listados no Quadro 15.

Quadro 15 – Fatores de entrada do conjunto de treinamento extraído do Artigo 3.

Fonte: Sarkar *et al.* (2006)

Código do fator	Descrição do fator	Unidade
T_{on}	Duração do pulso	μs
T_{off}	Duração da pausa	μs
I_p	Corrente de pico	A
WT	Tensão do fio	g
SV	Tensão elétrica de referência no servo motor	V
FR	Taxa de fluxo do dielétrico	Kg/cm^2
MS	Velocidade de usinagem	mm/min

A variável de saída do modelo é a rugosidade média (R_a), em μm . Os valores de rugosidade medidos durante os experimentos conduzidos pelos autores estão compreendidos entre 2,31 μm e 3,16 μm . O conjunto de treinamento contém 18 casos (sendo 15 casos de treinamento e 3 de seleção, a cada época) e o conjunto de teste contém 6 casos. Ambos são listados no Apêndice D.

5.3.3 EXPERIMENTOS DE USINAGEM

Os experimentos de usinagem realizados para validação do método corresponderam à execução de um ciclo de experimentos de torneamento de aço de corte fácil ABNT 12L14. A motivação para a escolha do processo e do material usinado foi sua grande aplicação no setor industrial. Aços de corte fácil são amplamente aplicados na produção de peças em tornos automáticos multifuso. Entre os tipos de aços de corte fácil, o aço ABNT/SAE 12L14 destaca-se por possuir uma maior demanda de produção e elevada usinabilidade (SALGADO JUNIOR, 2010). O material dos corpos de provas utilizados nos ensaios foi o aço ABNT 12L14, com dureza igual a 142 HV, fornecido pela Empresa Aços Villares S.A., cuja composição química é mostrada no Quadro 16.

Quadro 16 – Composição química do aço ABNT 12L14

%C	%Si	%Mn	%P	%S	%Cr	%Ni	%Cu	%Al	%Mo	%Pb
0,090	0,030	1,240	0,046	0,273	0,150	0,080	0,260	0,001	0,020	0,280

Os experimentos foram planejados por meio de uma adaptação da estratégia sugerida por Taguchi para projeto de parâmetro robusto. Utilizou-se um arranjo cruzado composto por um arranjo interno do tipo CCD (ao invés do arranjo ortogonal de Taguchi), formado pelas variáveis de controle, e outro arranjo externo, para a investigação da influência de variáveis de ruído. Apesar de serem variáveis não controláveis no processo de usinagem real, os ruídos adotados são controláveis no ambiente experimental, o que permitiu a investigação de sua influência sobre o processo.

O arranjo CCD empregado contém 8 pontos fatoriais, 6 pontos axiais e 3 pontos centrais e utiliza um valor de ρ igual a 1,682. Foram calculados os valores de média e variância para a variável de resposta. O arranjo externo é um fatorial completo com níveis mistos $2^2 \times 3$. Os fatores de entrada e ruído considerados, bem como os níveis experimentais adotados são detalhados no Quadro 17. A variável de saída do modelo é a rugosidade média (R_a), em μm .

No Quadro 17, a esbeltez da peça corresponde à divisão do comprimento do corpo de prova pelo seu diâmetro. Sendo o comprimento constante para todos os corpos de prova utilizados, definiu-se com esbeltas ou finas as peças que possuem diâmetro inicial igual a 30 mm e como não esbeltas (ou grossas) as que possuem diâmetro inicial igual a 50 mm. A razão para investigação desta variável de ruído se baseia na hipótese de que corpos de prova mais esbeltos sofram maior influência da vibração durante o processo de usinagem e, conseqüentemente, apresentem valores maiores de rugosidade da peça.

Quadro 17 – Fatores de entrada, ruídos e respectivos níveis utilizados nos experimentos de usinagem

Código do fator	Classificação	Descrição	Unidade	Níveis		
				-1	0	+1
V_c	Fator	Velocidade de corte	<i>m/min</i>	220	280	340
f	Fator	Avanço	<i>mm/v</i>	0,080	0,100	0,120
a_p	Fator	Profundidade de usinagem	<i>mm</i>	0,700	0,950	1,200
E	Ruído	Esbeltez (comprimento/diâmetro)	-	Grossa ($\phi 50mm$)	-	Fina ($\phi 30mm$)
VB	Ruído	Desgaste da ferramenta	<i>mm</i>	Nova (0 mm)	-	Usada (0,30 mm)
P	Ruído	Posição de medição	-	Contra-ponta	Centro	Castanha

A inclusão do fator denominado desgaste de ferramenta teve por objetivo investigar a influência do uso de ferramentas novas e usadas. Foram consideradas como novas ferramentas com arestas de corte sem qualquer desgaste. As ferramentas usadas empregadas possuíam desgaste medido na aresta de corte de aproximadamente $VB_{Max} = 0,30$ mm.

O fator chamado de posição de medição foi incluído no experimento como uma maneira indireta de avaliar o efeito da vibração sobre a rugosidade da peça da peça durante o processo de torneamento. A hipótese foi a de que, quanto maior o afastamento do ponto de fixação do corpo de prova na máquina (o engastamento), maior seria o efeito da vibração e maior a rugosidade da peça. Três regiões foram adotadas para medição da rugosidade: (i) região próxima ao contra-ponta (CP), onde há apenas apoio simples; (ii) região central (CE) e (iii), região próxima às castanhas (CA), onde ocorre o engastamento. As posições são ilustradas na Figura 36.

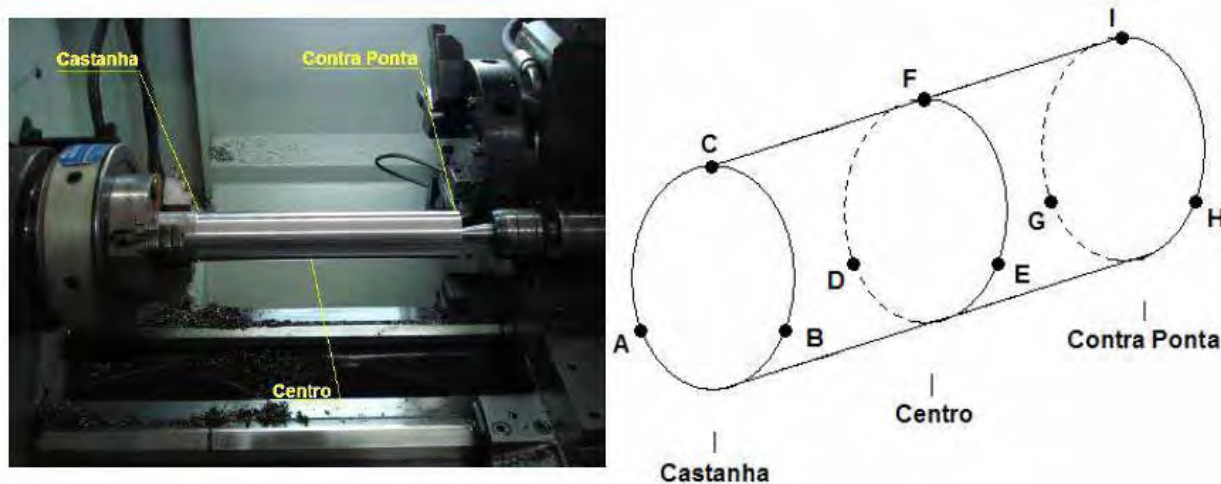


Figura 36 – Distribuição dos pontos de medição nos corpos de prova. Adaptado de Salgado Junior, 2010

Os experimentos foram conduzidos nos laboratórios de usinagem da Universidade Federal de Itajubá, em um torno CNC NARDINI, modelo LOGIC 175, com comando numérico CNC MCS Série 500, potência de 7,5 CV, rotação máxima do eixo árvore de 4000 rpm, torre porta-ferramenta com 8 posições, diâmetro máximo 100 mm e comprimento máximo torneável de 300 mm, entre pontas. A máquina

empregada é mostrada na Figura 37.



Figura 37 – Torno empregado para os experimentos de usinagem. Fonte: Salgado Junior, 2010

As ferramentas de corte utilizadas neste trabalho eram constituídas por metal duro classe ISO P35 revestido com três coberturas (Ti(C,N), Al₂O₃, TiN), classe GC 4035 Sandvik, na geometria ISO SNMG 09 03 04 – PM. O suporte utilizado é do tipo ISO DSBNL 1616H 09. Os corpos de prova possuíam as dimensões Φ 50 x 295 mm (peças grossas), Φ 30 x 295 mm (peças finas), rigorosamente identificados de acordo com sua origem para que não ocorressem trocas durante a execução dos ensaios.

Foram usinados ao todo 68 corpos de prova, sendo que cada um possui 4 pontos de medição distantes 90° entre si em relação a seção transversal em cada uma das 3 regiões de medição (contra-ponta, centro e castanha), como indicado na Figura 36. Em cada um dos 12 pontos de medição de cada corpo de prova foram tomadas 3 medidas. Desta forma, cada combinação de ruído (esbeltez, desgaste e posição) fornece um valor médio calculado conforme exemplificado pela Equação 51.

$$\bar{y}_{ijk} = \frac{y_{a1} + y_{a2} + y_{a3} + y_{b1} + y_{b2} + y_{b3} + y_{c1} + y_{c2} + y_{c3} + y_{d1} + y_{d2} + y_{d3}}{12} \quad (51)$$

A partir do valor médio calculado na equação anterior, calculou-se a média dos

valores da rugosidade como mostrado na Equação 52.

$$\mu_{ij} = \frac{\sum_{k=1}^{12} y_{ijk}}{12} \quad (52)$$

Foram realizadas 144 medições para se obter os valores de μ_{ij} em cada condição de usinagem. Assim sendo, foram executadas 2448 medições para a variável de resposta R_a .

Para a medição dos valores de rugosidade em cada corpo de prova, após o torneamento e resfriamento dos mesmos, foi utilizado um rugosímetro MITUTOYO SurfTest SJ-201P. Os valores medidos para R_a foram obtidos simultaneamente, durante o percurso da unidade de avanço e posteriormente registrados em planilha eletrônica. O instrumento utilizado é mostrado na Figura 38. Os valores de rugosidade medidos durante os experimentos conduzidos pelos autores estão compreendidos entre $0,59 \mu m$ e $10,48 \mu m$.



Figura 38 – Rugosímetro utilizado para medição da rugosidade

Os resultados obtidos foram pré-processados para remoção dos *outliers* extremos (observações que ultrapassassem 3 intervalos interquartis de distância em relação às extremidades do primeiro e terceiro quartis de um diagrama de caixa). A coleção final de medições formou os conjuntos de treinamento e seleção para validação do método.

Do total de exemplos (2448), dois terços foram utilizados como conjunto de treinamento e um terço como conjunto de seleção.

Um experimento adicional de confirmação foi realizado. Este experimento consistiu de 24 operações de corte em condições aleatórias entre os níveis adotados para os fatores e ruídos. Os resultados deste experimento de confirmação não foram utilizados durante o treinamento das redes, e constituíram o conjunto de testes. Desta forma, o conjunto de treinamento gerado contém 2448 casos (sendo 1632 de treinamento e 816 de seleção, a cada época) enquanto o conjunto de teste contém 24 casos de teste. O Apêndice E lista os casos de teste.

5.4 APLICAÇÃO DO MÉTODO A CURVAS SINTÉTICAS

5.4.1 VALORES INICIAIS E DESENVOLVIMENTO

Para a aplicação do método à otimização de redes neurais utilizando o conjunto de formado a partir de curvas sintéticas, contendo 30 casos de treinamento, foram adotados os seguintes valores iniciais para o algoritmo:

Número Máximo de Fases = 3

Ponto Central do Arranjo:

Fator de Escala Algoritmo Isotrópico ← 50

Número de unidades radiais ← 15

Intervalos Entre Níveis dos Fatores:

Intervalo Entre Fator de Escala Algoritmo Isotrópico ← 50

Intervalo Entre Número de unidades radiais ← 3

Os valores de inicialização do Fator de Escala do Algoritmo Isotrópico e do intervalo entre níveis não seguiram o especificado nos algoritmos ‘*Estabelecer Ponto Central do Arranjo*’ e ‘*Estabelecer Intervalos Entre Níveis dos Fatores*’, descritos no capítulo 4. Isso se deve ao fato destes experimentos terem sido conduzidos antes da finalização da concepção do método.

A otimização levou à realização de 18 ciclos experimentais. A primeira fase

Tabela 2 – Níveis dos fatores durante a aplicação do método para otimização de redes para curvas sintéticas – 30 casos de treinamento – ciclos de 10 a 18

Fator	Nível	Fase 1		Fase 2			Fase 3			
		Número do ciclo								
		10	11	12	13	14	15	16	17	18
ISO	1	450	400	440	430	420	410	420	428	430
	0	500	450	450	440	430	420	430	430	432
	-1	550	500	460	450	440	430	440	432	434
UR	1	15	7	12	9	12	9	12	14	13
	0	23	15	15	12	15	12	15	15	14
	-1	30	23	18	15	18	15	18	16	15
CEN	1	KM	KM	KM	KM	KM	KM	KM	KM	KM
	-1	SS	SS	SS	SS	SS	SS	SS	SS	SS

Nas Tabelas 1 e 2 são detalhados os níveis experimentais de cada fator empregados ao longo dos ciclos. Os ciclos que correspondem ao início e ao fim de uma fase são indicados em cores. As colunas em verde correspondem ao primeiro ciclo de uma fase. As colunas em azul correspondem ao fim de uma fase por repetição de um ponto central e as colunas em vermelho correspondem ao último ciclo de uma fase encerrada por falta de evidência estatística de redução do erro. Na Tabela 3, mostrada a seguir, são indicados os níveis dos fatores ISO e UR que compõem o centro dos dois arranjos fatoriais executados a cada ciclo para otimização de redes RBF (cada arranjo fatorial utiliza um nível do fator CEN). A tabela contém a indicação da melhor configuração obtida ao final de cada ciclo, nas colunas agrupadas sob o título ‘*Melhor Configuração Obtida no ciclo*’. Os níveis dos fatores quantitativos apontados como os melhores servem como base para o estabelecimento do ponto central dos arranjos para o próximo ciclo, como especificado pelo algoritmo *Estabelecer novo Ponto Central do Arranjo*, descrito no capítulo 4. Sob a coluna ‘*Variável de Saída (MAE-DI%)*’ são mostrados os resultados de predição obtidos pela melhor configuração de rede de cada ciclo. O mesmo código de cores utilizado nas tabelas anteriores é utilizado na Tabela 3 para indicar os ciclos que correspondem ao início e ao fim de cada fase.

Tabela 3 – Pontos centrais e melhores configurações de rede obtidos ao longo da aplicação do método para curvas sintéticas – 30 casos de treinamento.

Fase	Ciclo	Centro do arranjo (unidades naturais)		Melhor configuração obtida no ciclo (unidades codificadas)			Variável de saída (MAE-DI%)
		ISO	UR	ISO	UR	CEN	
1	1	50	15	1	1	1	0,002659
	2	100	23	1	-1	-1	0,001174
	3	150	15	1	1	1	0,001303
	4	200	23	1	-1	1	0,000425
	5	250	15	1	1	1	0,000696
	6	300	23	1	-1	-1	0,000245
	7	350	15	1	1	1	0,000469
	8	400	23	1	-1	1	0,000241
	9	450	15	1	1	1	0,000406
	10	500	23	-1	-1	1	0,000226
		<i>11</i>	<i>450</i>	<i>15</i>	<i>Repetição do arranjo ciclo 9. Fim da fase</i>		
2	12	450	15	-1	-1	1	0,000229
	13	440	12	-1	1	1	0,000228
	14	430	15	-1	-1	1	0,000224
	15	420	12	-1	1	1	0,000221
		<i>16</i>	<i>430</i>	<i>15</i>	<i>Repetição do arranjo ciclo 14. Fim da fase</i>		
3	17	430	15	1	-1	1	0,000226
	<i>18</i>	<i>432</i>	<i>14</i>	<i>Sem fator significativo. Fim do algoritmo</i>			<i>0,000215</i>

Na Figura 39 é ilustrada a evolução de desempenho obtida ao longo do processo de otimização.



Figura 39 – Evolução de desempenho das melhores redes a cada ciclo – aplicação do método a curvas sintéticas – 30 casos de treinamento.

Para cada ciclo (no eixo das abscissas) é indicado o desempenho da melhor configuração de rede obtida (no eixo das ordenadas), expresso em MAE-DI%. Nota-se a expressiva redução no erro de predição ocorrido nos quatro primeiros ciclos. Ao final da aplicação do método a melhor configuração de rede obtida foi a especificada pela Tabela 4. O desempenho do melhor modelo de rede RBF é sumarizado na Tabela 5.

Tabela 4 – Configuração de rede neural RBF otimizada para predição da rugosidade para curvas sintéticas – 30 casos de treinamento

Fator de projeto	Nível ótimo do fator
Fator de escala do algoritmo Isotrópico para determinação do parâmetro de largura da função radial	432
Número de unidades radiais	14
Algoritmo para determinação do hiper-centro das unidades radiais	K-Médias

Tabela 5 – Desempenho de rede neural RBF otimizada para predição da rugosidade para curvas sintéticas – 30 casos de treinamento

Medida de Desempenho	Erro de predição (MAE-DI%)
Valor médio	0,000215
Desvio-padrão	0,000062
Valor mínimo	0,000130

A repetição independente da sequência inicialização-treinamento- teste do modelo por 200 vezes (sendo 100 em cada réplica experimental) tornou possível a obtenção de dados relativos à distribuição estatística do erro de predição da rugosidade (em MAE%) associado a uma dada configuração de rede. O sumário gráfico da distribuição resultante da aplicação do modelo ótimo de rede à predição do conjunto dos casos de teste é ilustrado na Figura 40.

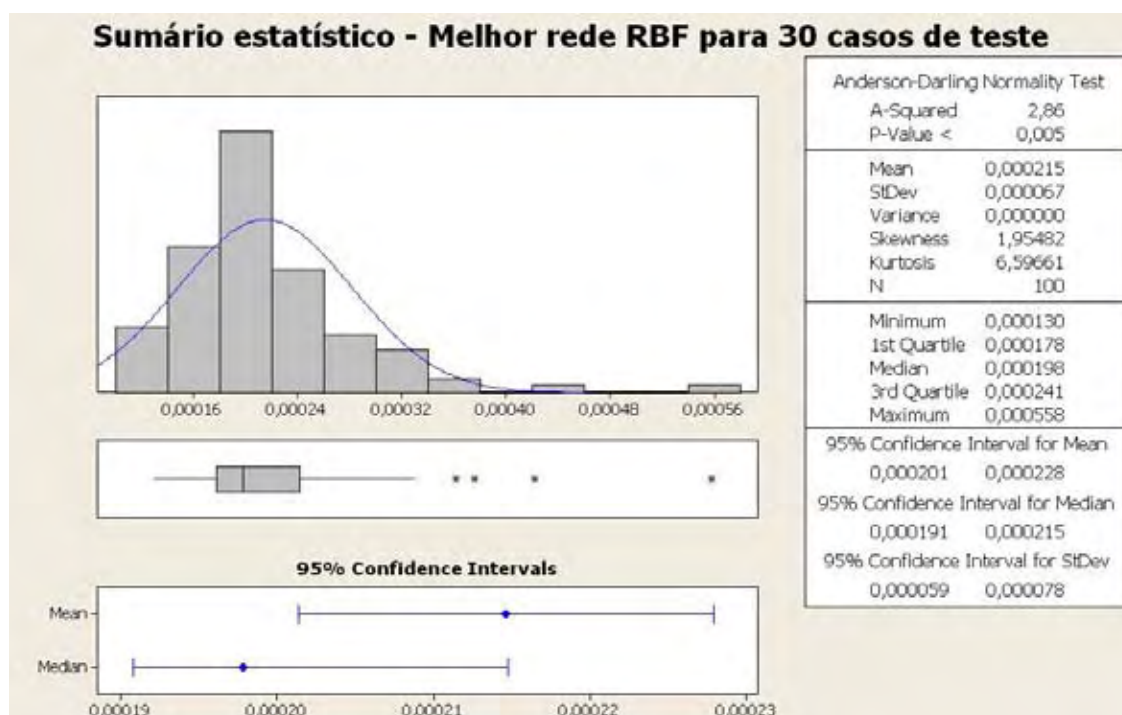


Figura 40 – Sumário gráfico da distribuição de erro da rede RBF otimizada – experimento com curvas sintéticas – 30 casos de treinamento

5.4.3 COMPARAÇÃO COM RESULTADOS DO TRABALHO ORIGINAL

Os resultados da melhor configuração de rede obtida por meio da aplicação do método proposto foram comparados aos resultados obtidos por Pontes (2006) com redes RBF utilizando os mesmos conjuntos de treinamento, seleção e teste. O valor médio de MAE-DI% obtido pela melhor rede daquele trabalho foi igual a 0,241512%, com desvio-padrão igual a 0,040828%. A comparação estatística entre as médias foi feita utilizando-se análise de variância. Comparou-se o desempenho das duas redes na predição da rugosidade dos casos de teste. O resultado é mostrado na Figura 41, na qual se ilustra a tabela da ANOVA extraída do Minitab®.

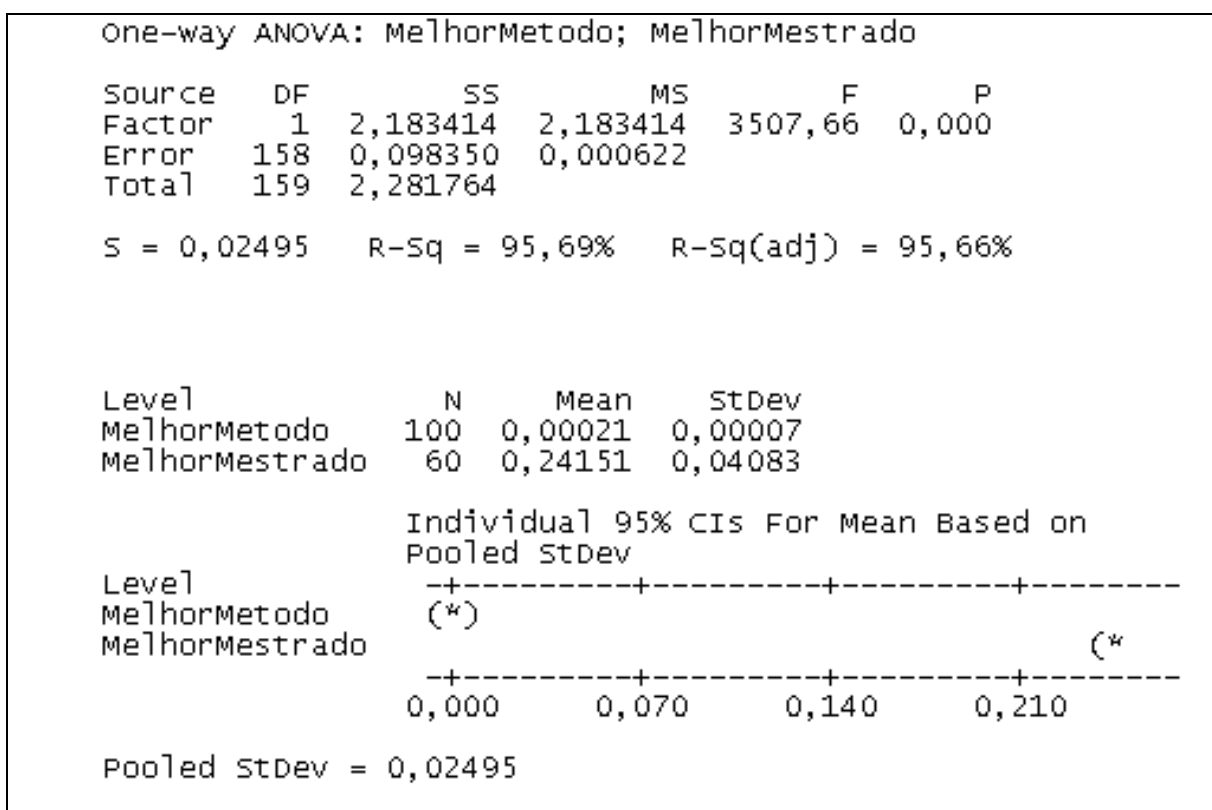


Figura 41 – Comparação entre rede otimizada RBF para curvas sintéticas e melhor rede obtida por Pontes (2006) – 30 casos de treinamento

Conforme se observa, o valor p resultante da aplicação da ANOVA apresenta valor igual 0,000, Isto indica que, ao nível de significância adotado (igual a 0,05) existe forte evidência estatística de diferença entre o desempenho das duas redes.

Conclui-se que o método proposto foi capaz de identificar uma configuração de rede RBF com desempenho superior ao obtido no estudo original. O resultado da comparação é adicionalmente ilustrado na Figura 42, que contém os diagramas de Caixa das redes sob comparação.

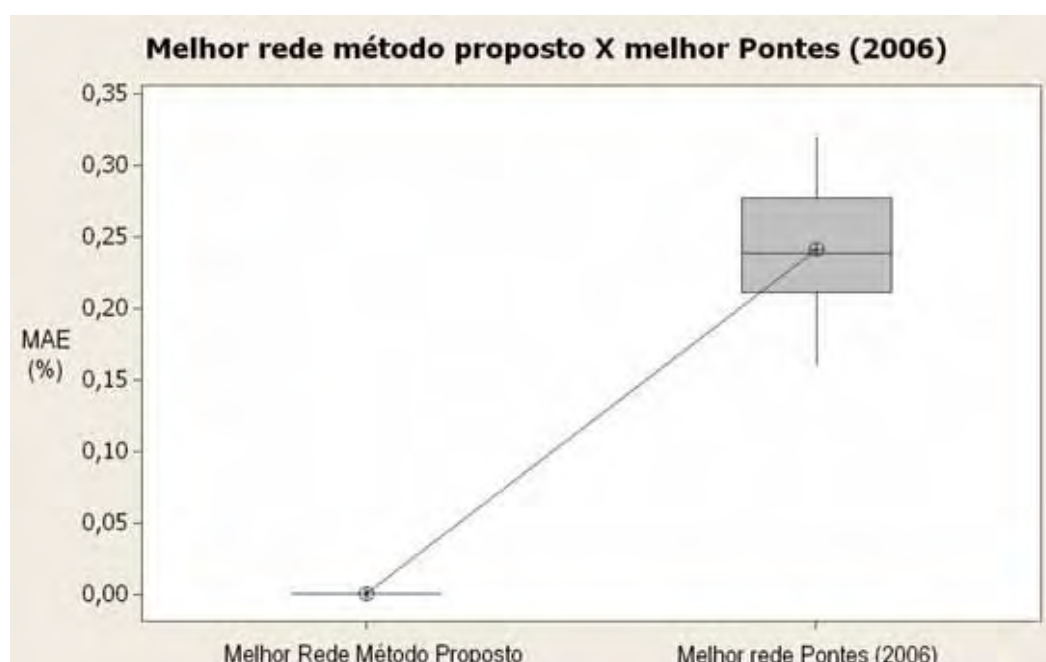


Figura 42 – Gráfico ilustrativo da comparação entre a distribuição de predições entre rede RBF ótima e melhor rede obtida por Pontes (2006) – 30 casos de treinamento

Nesta figura, o eixo das ordenadas indica o erro médio das redes em MAE-DI%. Além de apresentar erro de predição muito inferior, nota-se que a rede obtida pela aplicação do método proposto apresenta dispersão muito reduzida na comparação com aquela obtida por Pontes (2006).

5.4.4 REPETIÇÃO DO ESTUDO UTILIZANDO CONJUNTO DE DADOS CONTENDO 15 CASOS DE TREINAMENTO

O estudo foi repetido para o conjunto de dados contendo apenas 15 casos de treinamento. Os valores de inicialização e o processo de análise foram os mesmos utilizados anteriormente. Ao todo, foram conduzidos 17 ciclos experimentais, sendo dez na primeira fase, quatro na segunda e três na terceira fase, que se encerrou por falta de efeitos significativos. A melhor configuração de rede obtida é mostrada na

de treinamento.

A exatidão obtida na predição para curvas sintéticas é bastante elevada. Deve-se lembrar porém, que tais curvas são resultantes de um modelo quadrático, cuja não linearidade é bastante reduzida na comparação com resultados de experimentos reais. Desta forma, pode-se esperar erros maiores quando da aplicação do método a casos reais. A utilidade do uso das curvas sintéticas na primeira fase do trabalho foi possibilitar a concepção do método enquanto se buscavam conjuntos de dados na literatura e se preparam os experimentos de torneamento com aço 12L14.

5.5 APLICAÇÃO DO MÉTODO A CONJUNTOS DE DADOS EXTRAÍDOS DA LITERATURA

5.5.1 APLICAÇÃO DO MÉTODO AOS DADOS DO ARTIGO Nº 1

5.5.1.1 Valores iniciais e desenvolvimento

Para a aplicação do método utilizando o conjunto de dados do Artigo 1 aplicou-se o algoritmo de otimização a redes de arquitetura MLP. Utilizou um *Número Máximo de Fases* igual a 4. *Ponto Central do Arranjo* e *Intervalos Entre Níveis dos Fatores* foram inicializados conforme valores especificados nos algoritmos do capítulo 4.

A otimização levou à realização de 9 ciclos experimentais. A primeira fase encerrou-se após o ciclo 2, quando a análise apontou como ponto central do arranjo para o ciclo seguinte um conjunto de níveis dos fatores já utilizado como ponto central em ciclo anterior da mesma fase. A segunda fase encerrou-se após o ciclo 5, pelo mesmo motivo. A terceira fase encerrou-se ao final do ciclo 7, quando a análise dos resultados deixou de apontar redução do erro. A quarta fase se encerrou ao final do ciclo 9, pelo mesmo motivo.

De maneira diversa do parágrafo anterior, não foram considerados os ciclos cujos arranjos seriam iguais aos executados em ciclo anterior. Ao invés disso, a listagem de ciclos passa diretamente ao arranjo do primeiro ciclo da fase seguinte.

5.5.1.2 Resultados da aplicação do método

Os resultados da aplicação do método à otimização de redes neurais MLP para o conjunto de dados do Artigo 1 são mostrados na Tabela 7.

Tabela 7 – Níveis dos fatores durante a aplicação do método para otimização de redes usando dados do Artigo 1

		Fase 1		Fase 2			Fase 3		Fase 4	
		Número do ciclo								
Fator	Nível	1	2	3	4	5	6	7	8	9
H1	1	0	8	4	8	12	10	12	13	12
	0	8	16	8	12	16	12	14	14	13
	-1	16	24	12	16	20	14	16	15	14
H2	1	0	0	4	0	0	2	0	1	0
	0	8	8	8	4	4	4	2	2	1
	-1	16	16	12	8	8	6	4	3	2
E1	1	0	0	16	0	0	8	0	4	8
	0	32	32	32	16	16	16	8	8	12
	-1	64	64	48	32	32	24	16	12	16
LR	1	0,001	0,400	0,200	0,001	0,001	0,100	0,001	0,050	0,100
	0	0,400	0,800	0,400	0,200	0,200	0,200	0,100	0,100	0,150
	-1	0,800	1,200	0,600	0,400	0,400	0,300	0,200	0,150	0,200
E2	1	0	0	32	0	0	16	0	8	0
	0	64	64	64	32	32	32	16	16	8
	-1	128	128	96	64	64	48	32	24	16

Nesta tabela são mostrados os níveis experimentais de cada fator empregados ao longo dos ciclos. Os ciclos que correspondem ao início e ao fim de uma fase são indicados em cores. As colunas em verde correspondem ao primeiro ciclo de uma fase. As colunas em azul correspondem ao fim de uma fase por repetição de um ponto central e as colunas em vermelho correspondem ao último ciclo de uma fase encerrada por falta de evidência estatística de redução do erro. Na Tabela 8, mostrada a seguir, são indicados os níveis dos fatores que compõem o centro do arranjo fatorial executado a cada ciclo. A tabela contém a indicação da melhor configuração obtida ao

final do ciclo, nas colunas agrupadas sob o título ‘*Melhor Configuração Obtida no ciclo*’. Os níveis dos fatores quantitativos apontados como os melhores servem como base para o estabelecimento do ponto central dos arranjos para o próximo ciclo, conforme especificado pelo algoritmo *Estabelecer novo Ponto Central do Arranjo*, no capítulo 4. Sob a coluna ‘*Variável de Saída (MAE-DI%)*’ são mostrados os resultados de predição obtidos pela melhor configuração de rede obtida no ciclo. O mesmo código de cores utilizado na Tabela 7 é utilizado na Tabela 8 para indicar os ciclos que correspondem ao início e ao fim de cada fase.

Tabela 8 – Pontos centrais e melhores configurações de rede obtidos ao longo da aplicação do método ao conjunto de dados do Artigo 1.

Fase	Ciclo	Centro do arranjo (unidades naturais)					Melhor configuração obtida no ciclo (unidades codificadas)					Variável de saída (MAE- DI%)
		H1	H2	E1	LR	E2	H1	H2	E1	LR	E2	
1	1	8	8	32	0,400	64	1	-1	-1	1	-1	14,104
	2	16	8	32	0,800	64	-1	-1	-1	-1	-1	13,718
2	3	8	8	32	0,400	64	1	-1	-1	-1	-1	14,805
	4	12	4	16	0,200	32	1	-1	-1	-1	-1	14,198
	5	16	4	16	0,200	32	-1	-1	-1	-1	-1	14,127
3	6	12	4	16	0,200	32	1	-1	-1	-1	-1	13,635
	7	14	2	8	0,100	16	-1	-1	1	-1	-1	13,956
4	8	14	2	8	0,100	16	-1	-1	1	1	-1	12,797
	9	13	1	12	0,150	8	1	-1	-1	-1	-1	13,513

Na Figura 44 é ilustrada a evolução de desempenho obtida ao longo do processo de otimização. Para cada ciclo (no eixo das abscissas) mostra-se o desempenho da melhor configuração de rede obtida (no eixo das ordenadas), expresso em MAE-DI%. Ao final da otimização a melhor configuração de rede obtida foi a especificada pela Tabela 9, sendo seu desempenho sumarizado na Tabela 10.

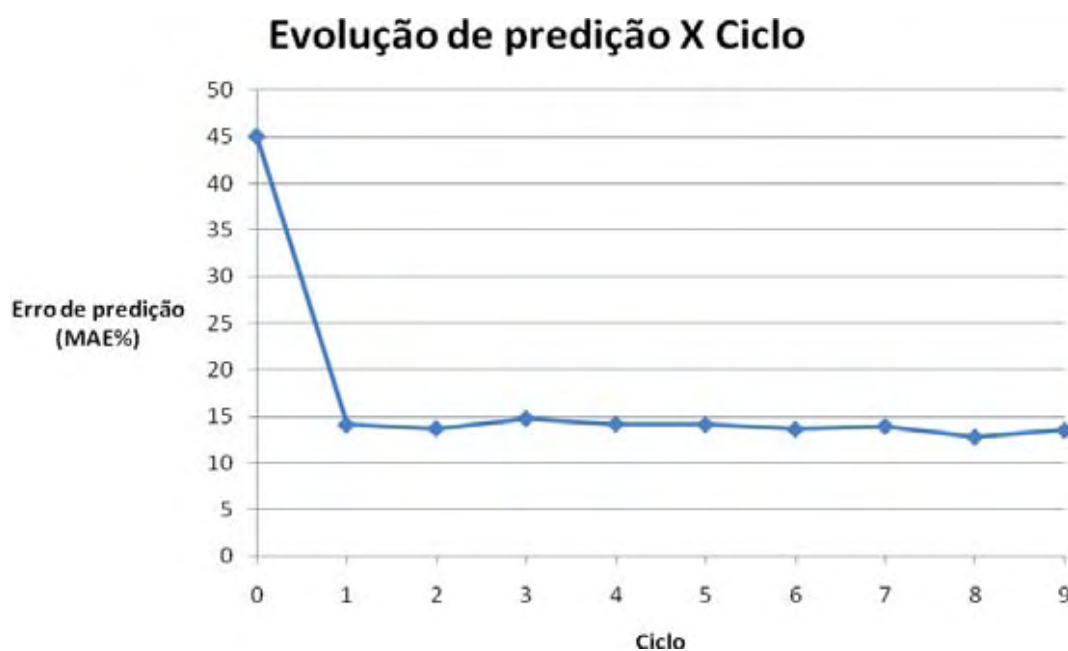


Figura 44 – Evolução de desempenho das melhores redes MLP a cada ciclo – aplicação do método ao conjunto de dados do Artigo 1

Tabela 9 – Configuração de rede neural MLP otimizada para previsão da rugosidade usando os dados do Artigo 1

Fator de projeto	Nível ótimo do fator
Número de neurônios na camada oculta 1	13
Número de neurônios na camada oculta 2	1
Número de épocas de treinamento na fase 1	12
Taxa de aprendizado na fase 1	0,150
Número de épocas de treinamento na fase 2	8

Tabela 10 – Desempenho de rede neural MLP otimizada para previsão da rugosidade usando os dados do Artigo 1

Medida de Desempenho	Erro de previsão (MAE-DI%)
Valor médio	12,797
Desvio-padrão	15,230
Valor mínimo	10,470

A repetição, da sequência inicialização-treinamento-teste do modelo por 200 vezes tornou possível a obtenção de dados relativos à distribuição estatística do erro de predição da rugosidade (em MAE%) associado a uma dada configuração de rede. O sumário gráfico da distribuição resultante da aplicação do modelo ótimo de rede MLP à predição do conjunto dos casos de teste é ilustrado na Figura 45.

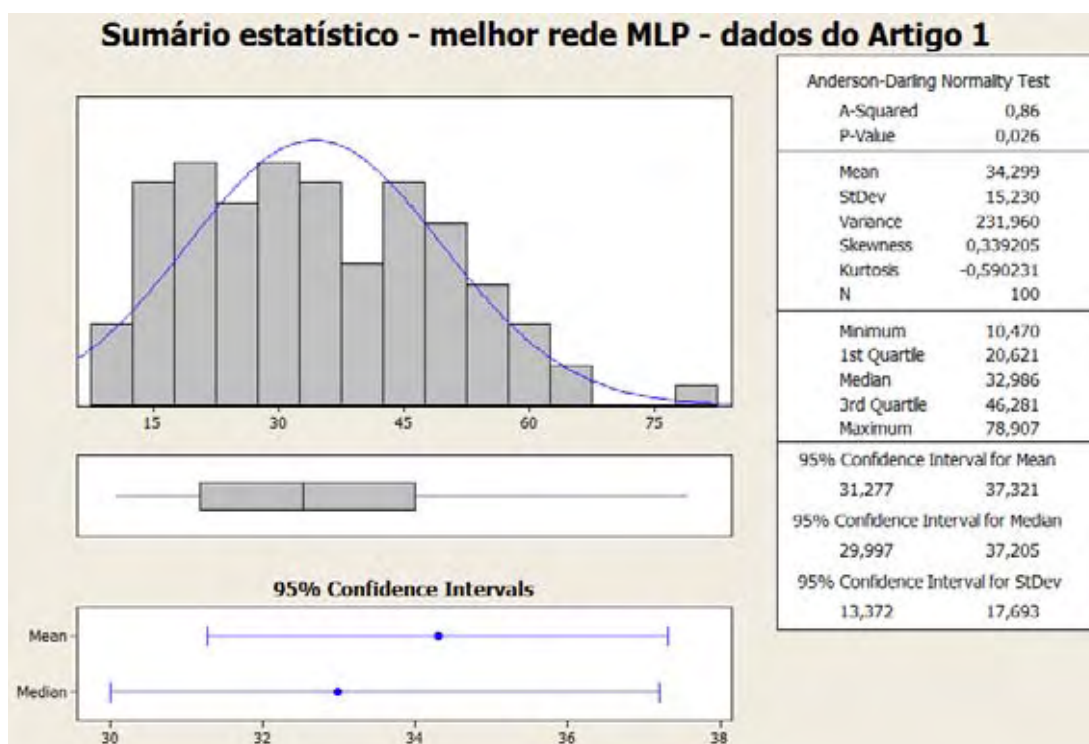


Figura 45 – Sumário gráfico da distribuição de erro da rede MLP ótima – experimento com dados do Artigo 1

O modelo de rede ótima é capaz de gerar superfícies de resposta para o processo sob estudo. Nas Figuras 46, 47, 48 e 49 são mostrados exemplos de superfícies de resposta geradas pela configuração ótima de rede MLP. Percebe-se a capacidade da rede de mapear a não-linearidade característica do processo de formação da rugosidade.

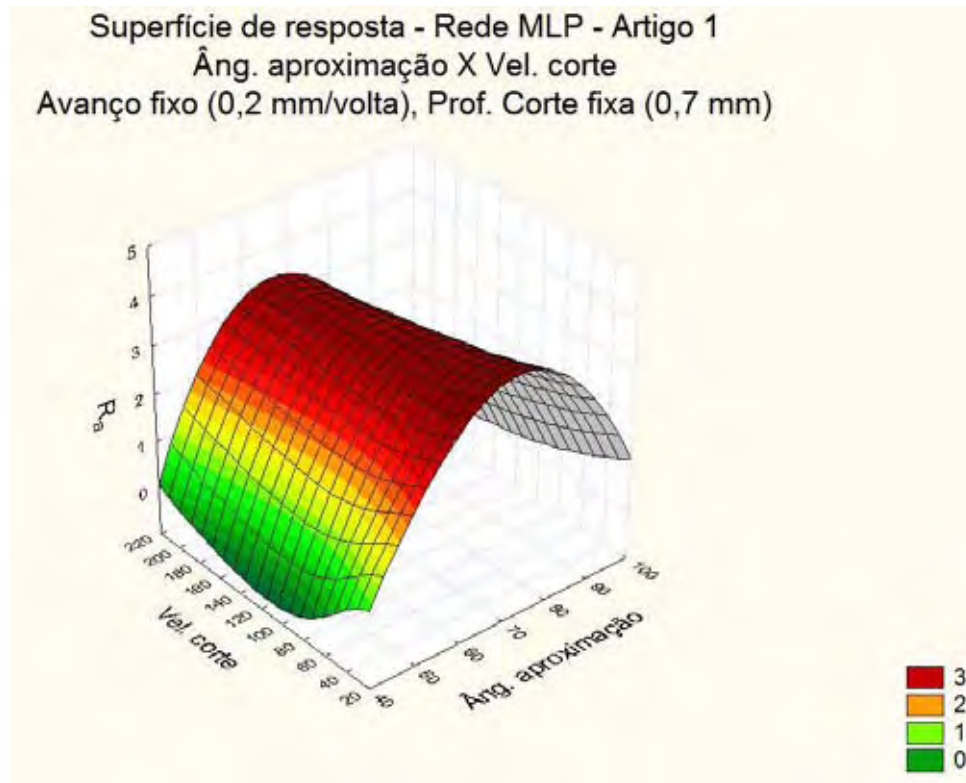


Figura 46 – Resposta da rede MLP ótima para dados do Artigo 1 – AA versus V_c

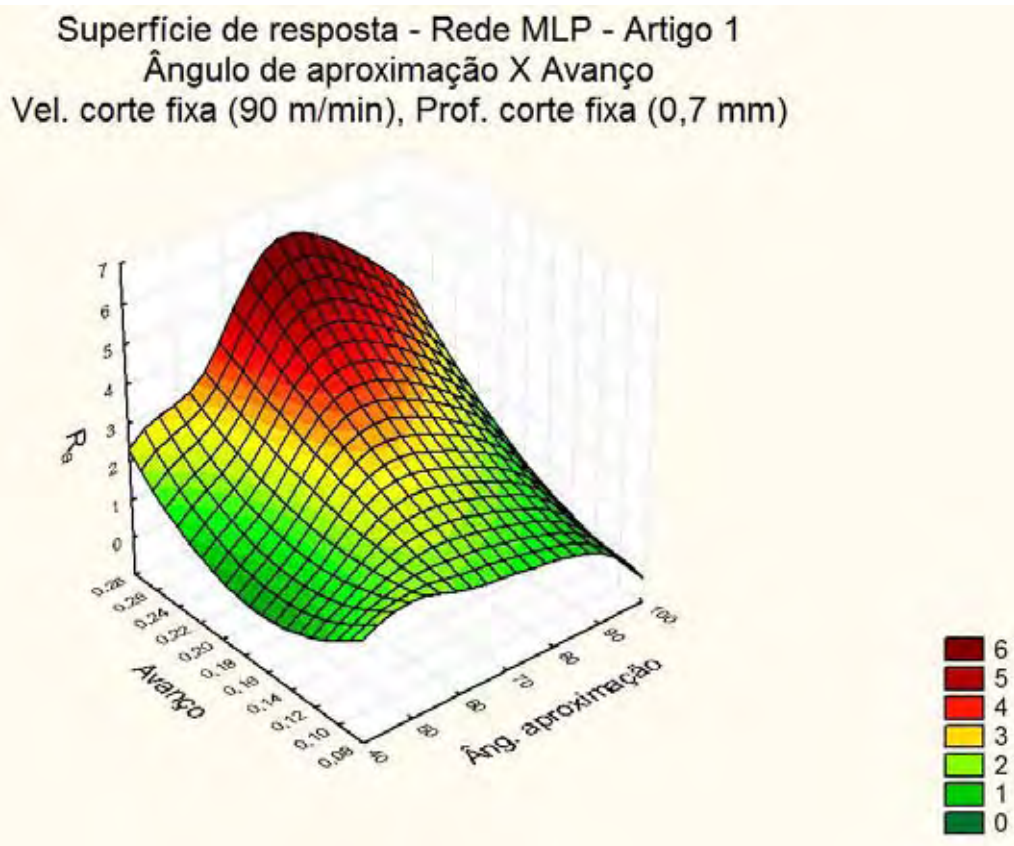


Figura 47 – Resposta da rede MLP ótima para dados do Artigo 1 – AA versus f

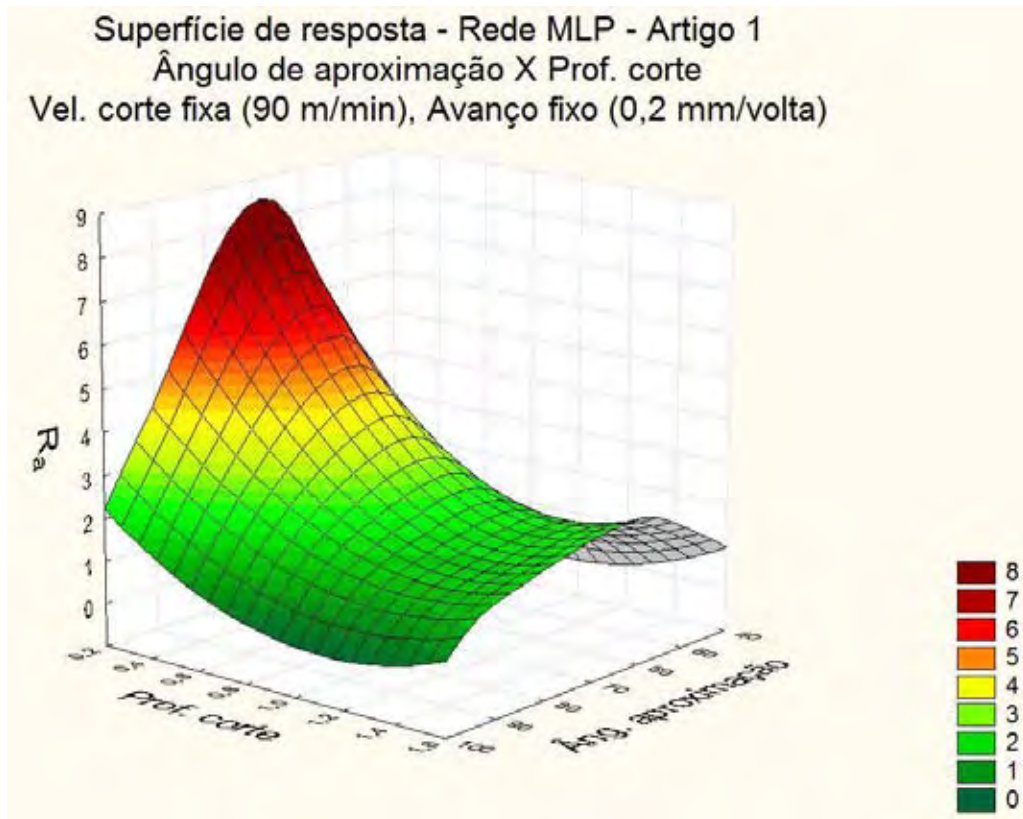


Figura 48 – Resposta da rede MLP ótima para dados do Artigo 1 – AA versus a_p

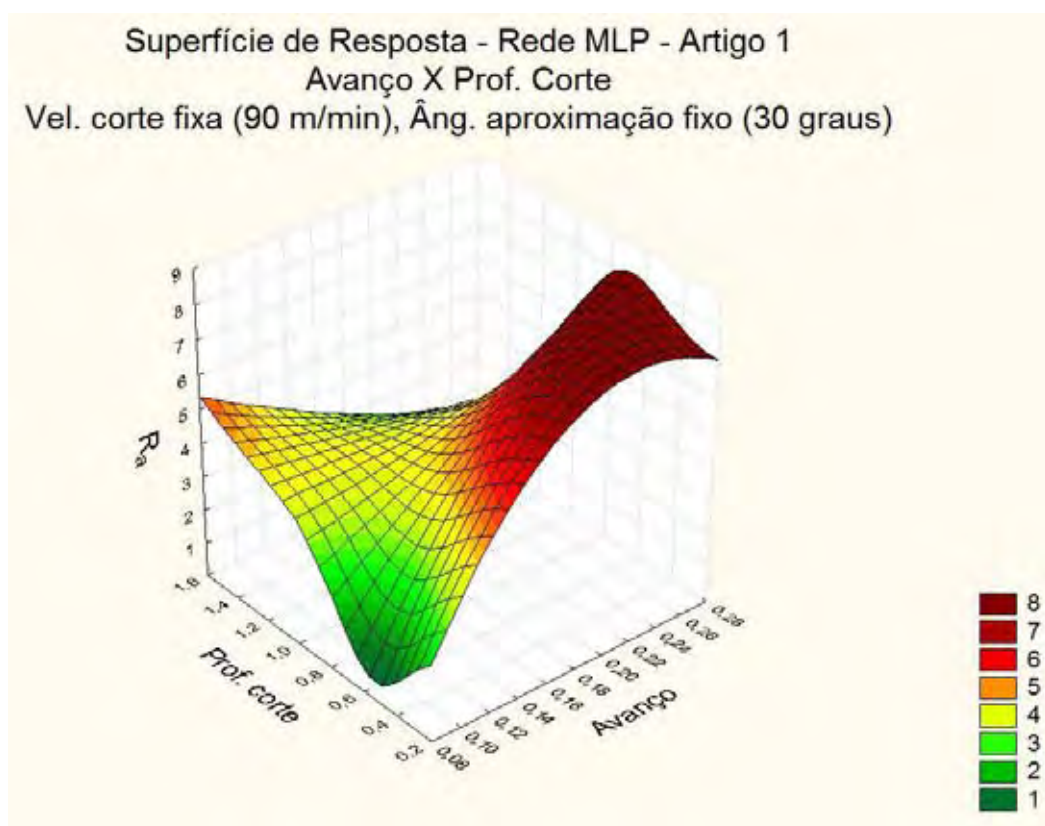


Figura 49 – Resposta da rede MLP ótima para dados do Artigo 1 – f versus a_p

5.5.1.3 Comparação com resultados do trabalho original

Comparou-se o desempenho da melhor configuração de rede obtida por meio da aplicação do método proposto ao resultado obtido pelo estudo que originou os dados utilizados. O valor do erro de predição atingido pela melhor rede na predição do conjunto de casos de teste foi comparado ao valor correspondente calculado com base nos resultados do artigo da literatura, por meio de um teste de hipóteses. Utilizou-se o teste-t bilateral para hipótese nula de médias de erro iguais entre as redes ao nível de significância de 0,05. O valor da variável de saída da melhor rede obtida é igual a 12,797%. O valor correspondente, calculado para o Artigo 1, (utilizado como valor de comparação para o teste de hipótese) é igual a 72,108%. O resultado do teste, extraído do Minitab®, é mostrado na Figura 50.

```

One-sample T: Metodo proposto
Test of mu = 72,1085 vs not = 72,1085

Variable      N      Mean    StDev   SE Mean      95% CI      T
Metodo proposto  10  12,5398  1,0047   0,3177  (11,8210; 13,2585)  -187,49

Variable      P
Metodo proposto  0,000

```

Figura 50 – Teste de hipótese para médias iguais de erro de predição entre a melhor rede MLP obtida pelo método e resultado do Artigo 1

Como se observa na figura, o valor p resultante do teste é igual a 0,000, o que leva à rejeição da hipótese nula de médias iguais. Há forte evidência estatística de que o desempenho obtido pela configuração de rede resultante do método de otimização proposto é superior ao desempenho da rede proposta no artigo original. O resultado do teste de hipótese também é ilustrado na Figura 51, na qual se mostra o diagrama de caixa para o teste de hipótese. Na figura, \bar{X} representa a média da variável de saída (medida em MAE-DI%) para a rede otimizada e H_0 representa o resultado declarado no artigo da literatura.

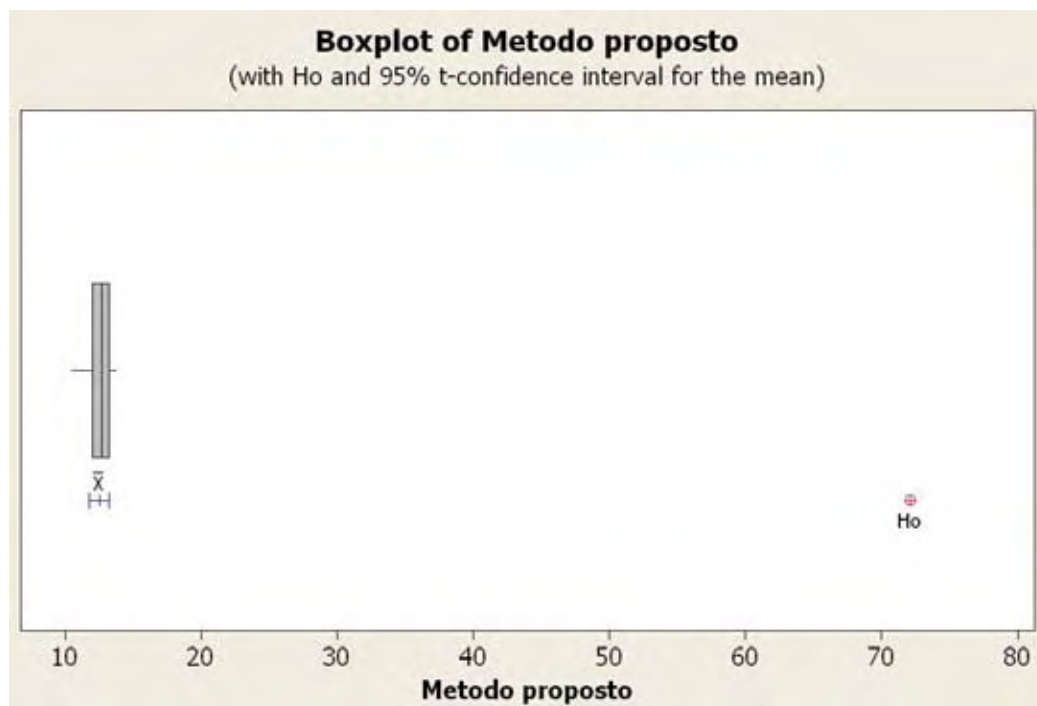


Figura 51 – Diagrama de caixa do teste de hipótese entre erro de predição da melhor rede MLP obtida pelo método e resultado do Artigo 1

Além do resultado estatístico superior obtido pela melhor configuração de rede na comparação com o resultado do estudo original, observou-se que a mesma configuração apresentou resultado ainda melhor em algumas instâncias do teste. Em algumas instâncias, observou-se que a configuração otimizada chegou a apresentar erro mínimo igual a 10,470%.

5.5.1.4 Comparação com resultados de modelo de regressão múltipla

O desempenho da melhor configuração de rede obtida foi comparado ao desempenho de um modelo de predição elaborado por meio de regressão múltipla. A elaboração do modelo de regressão seguiu as recomendações de Montgomery & Runger (2007): iniciando-se do modelo mais simples, abandonava-se um modelo em favor de outro quando a adição de um regressor ao modelo original reduzia a soma quadrática do erro do novo modelo por um valor no mínimo igual à média quadrática do erro do modelo original. Repetiu-se tal operação para todas as combinações possíveis de regressores, e desta forma, chegou-se ao modelo mostrado na Equação 53,

onde os regressores AA , V_c , f e a_p são definidos como no Quadro 13, e estão expressos em unidades naturais.

$$R_a = 233 - 0,0236*AA - 3,31*V_c - 1088*f - 185*a_p - 165*f*a_p + 55,1*V_c *f*a_p \quad (53)$$

Na Figura 52 mostra-se os resultados da análise de regressão efetuada sobre o modelo escolhido, com uso do pacote computacional Minitab®.

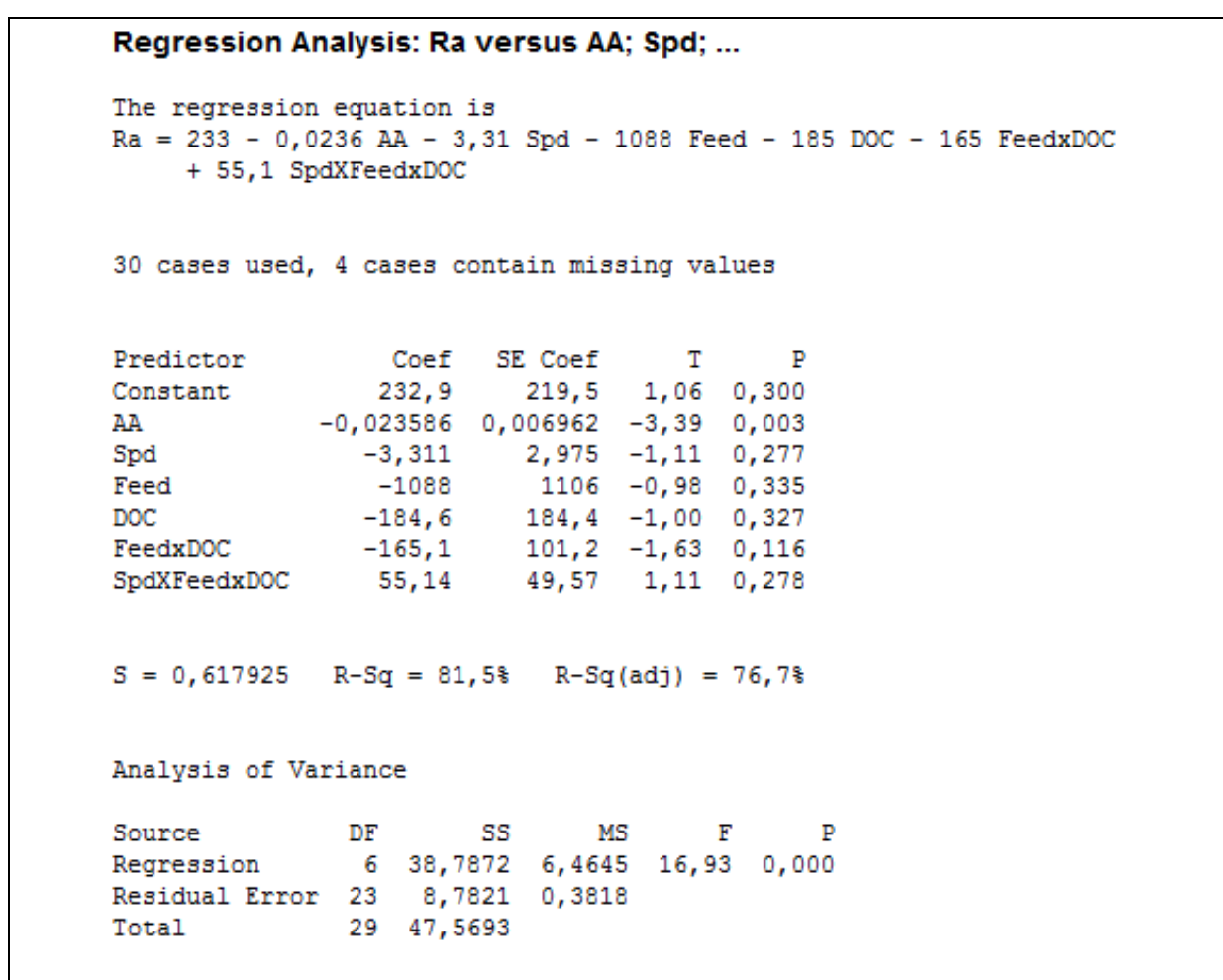


Figura 52 – Análise de regressão do conjunto de dados obtido do Artigo 1

Observa-se o bom ajuste da curva, indicado pelo valor de $R-Sq(adj)$ igual a 76,7% e pelo valor p da análise de variância para significância da regressão, igual a 0,000. O teste de normalidade de Anderson-Darling foi aplicado sobre os resíduos e não indicou qualquer evidência de que os resíduos não sigam a distribuição normal.

A comparação entre a rede e o modelo de regressão foi feita por meio de um teste de hipótese entre os valores do erro de predição da melhor configuração de rede para o conjunto de teste (em MAE-DI%) e o valor do erro de predição obtido pelo modelo de regressão. Utilizou-se o teste-t bilateral para hipótese nula de médias de erro iguais ao nível de significância de 0,05. O valor do erro de predição calculado para o modelo de regressão (utilizado como valor de comparação para o teste de hipótese) é igual a 24,873%. O resultado do teste, extraído do Minitab®, é mostrado na Figura 53.

One-sample T: Metodo proposto						
Test of mu = 24,8726 vs not = 24,8726						
variable	N	Mean	StDev	SE Mean	95% CI	T
Metodo proposto	10	12,5398	1,0047	0,3177	(11,8210; 13,2585)	-38,82
variable	P					
Metodo proposto	0,000					

Figura 53 – Teste de hipótese para médias iguais de erro de predição em MAE-DI% da melhor rede MLP obtida pelo método e modelo de regressão para dados do Artigo 1

Como se observa na Figura 53, o valor p resultante do teste é igual a 0,000, o que leva à rejeição da hipótese nula de médias iguais. Este resultado evidencia, ao nível de significância adotado, superioridade do desempenho da rede obtida por meio do método proposto. O resultado também é ilustrado na Figura 54, na qual se mostra o diagrama de caixa para o teste de hipótese. Na figura, \bar{X} representa a média da variável de saída (medida em MAE-DI%) para a rede otimizada e H_0 representa o resultado obtido pelo modelo de regressão.

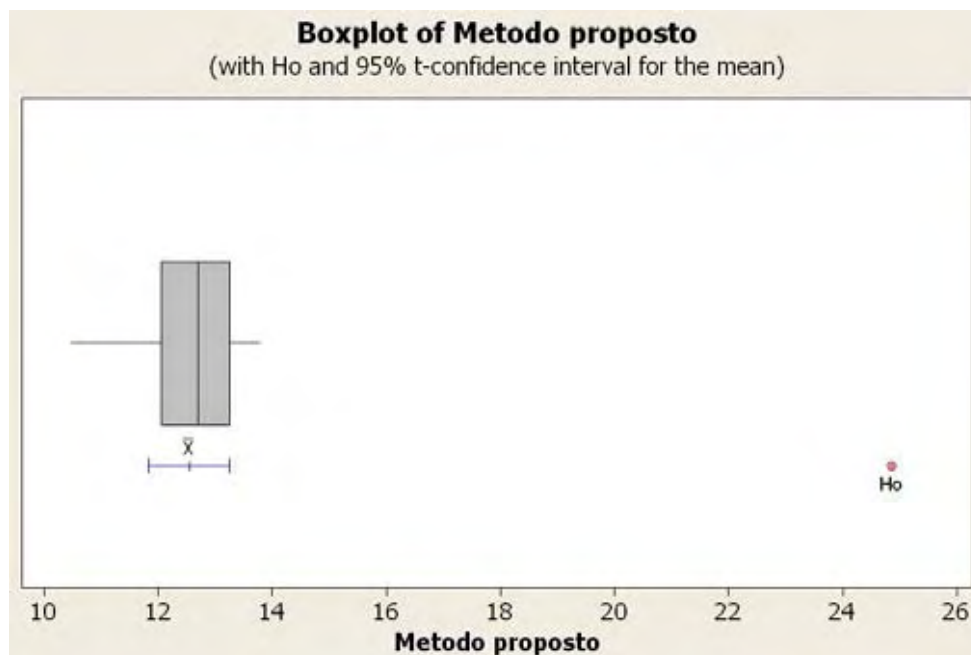


Figura 54 – Diagrama de caixa do teste de hipótese para comparação entre a melhor rede MLP obtida pelo método e modelo de regressão para o Artigo 1

5.5.1.5 Comparação com resultados da ferramenta automática de otimização de redes

Uma outra comparação foi realizada entre a melhor configuração de rede MLP obtida por meio da aplicação do método e uma arquitetura de rede otimizada pelo pacote computacional. Em todos os casos relatado neste trabalho, o IPS foi utilizado com limites superiores de procura extremamente altos (por exemplo: 100.000 arquiteturas de redes, ou 1000 neurônios na camada oculta) com objetivo de maximizar a probabilidade da ferramenta encontrar uma arquitetura ótima.

O valor da variável de saída (em MAE-DI%) medido para a rede proposta pela ferramenta, foi igual a 14,170%. Utilizou-se o teste de análise de variância entre para comparar as previsões de erro das duas configurações de rede. O resultado é ilustrado na Figura 55, na qual se mostra a tabela da ANOVA.

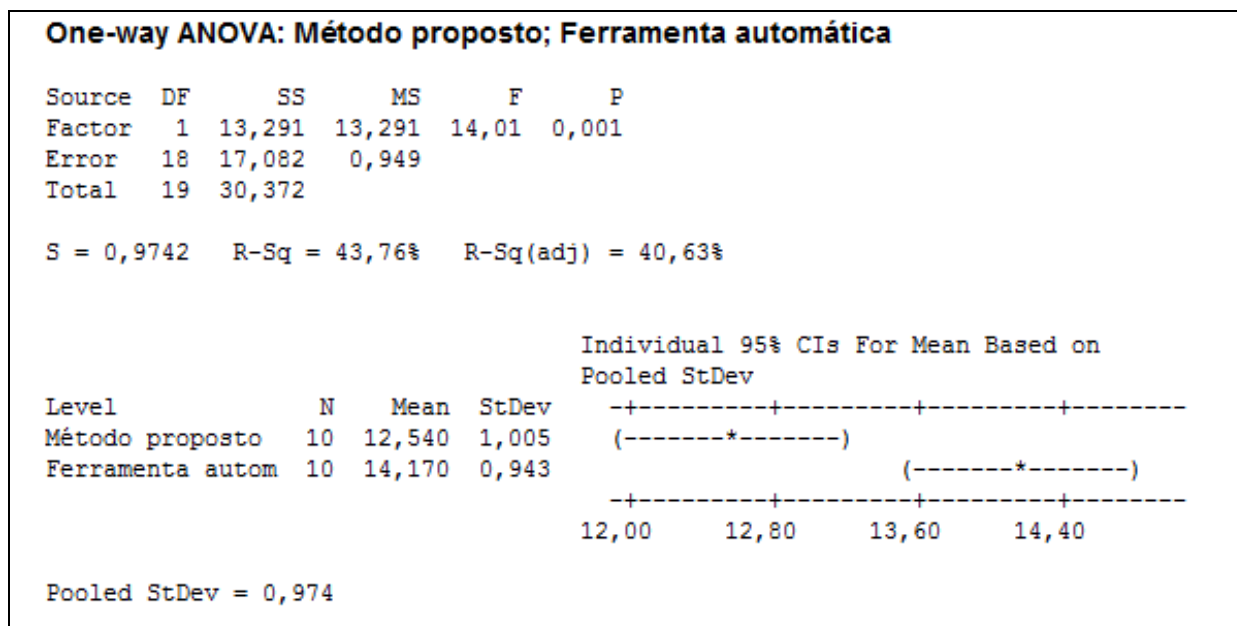


Figura 55 – Comparação entre erro de predição em MAE-DI% da melhor rede MLP obtida pelo método e da melhor rede apontada pelo pacote computacional para dados do Artigo 1

O resultado da ANOVA indica que, ao nível de significância adotado (igual a 0,05), existe evidência de diferença estatística entre os erros de predição das configurações de redes sob comparação. O erro de predição da rede obtida por meio do método é inferior ao obtido por meio do pacote computacional. Observa-se, portanto, que o método proposto foi capaz de identificar uma configuração de rede com desempenho superior ao da ferramenta comercial utilizada. A comparação é adicionalmente ilustrada na Figura 56, na qual se mostra o diagrama de caixa das observações do erro de predição para as duas configurações de rede, para o conjunto dos casos de teste. Observa-se a clara diferença entre o desempenho das redes sob comparação.

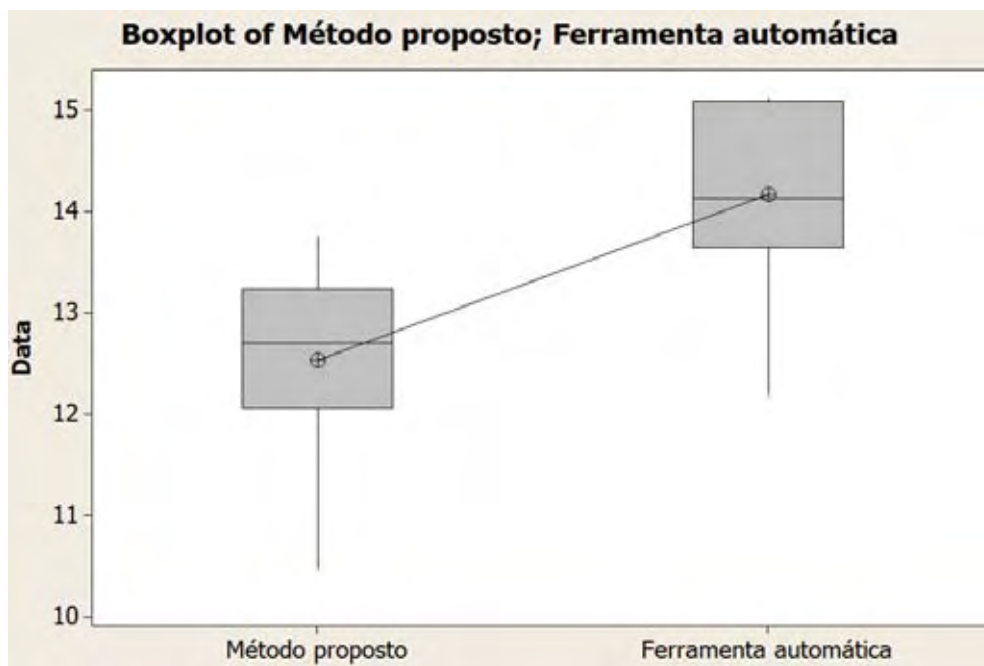


Figura 56 – Comparação entre a melhor rede MLP obtida pelo método e melhor rede apontada pelo pacote computacional para dados do Artigo 1

5.5.1.6 Reprodução do estudo com redes de arquitetura RBF

A aplicação do método à otimização de redes RBF foi realizada de modo similar ao empregado para a otimização de redes MLP. O *Número Máximo de Fases* utilizado foi igual a 4. *Ponto Central do Arranjo* e *Intervalos Entre Níveis dos Fatores* foram inicializados conforme especificado nos algoritmos do capítulo 4. A otimização foi realizada em 11 ciclos experimentais. A primeira fase encerrou-se após o ciclo 4, quando a análise apontou como ponto central do arranjo para o ciclo seguinte o mesmo conjunto de níveis dos fatores utilizado como ponto central naquele ciclo. A segunda fase encerrou-se após o ciclo 7, quando a análise dos resultados deixou de apontar redução do erro. A terceira e quarta fases encerraram-se ao final dos ciclos 9 e 11, respectivamente, pela mesma razão que determinou o fim da primeira fase.

Na Figura 57 é ilustrada a evolução de desempenho obtida ao longo do processo de otimização. Para cada ciclo (no eixo das abscissas) é indicado o desempenho da melhor configuração de rede obtida (no eixo das ordenadas), expresso em MAE-DI%. Ao final da aplicação do método a melhor configuração de rede obtida foi a

especificada pela Tabela 11. O desempenho do melhor modelo de rede RBF é sumarizado na Tabela 12.



Figura 57 – Evolução de desempenho das melhores redes RBF a cada ciclo – aplicação do método ao conjunto de dados do Artigo 1

Tabela 11 – Configuração de rede neural RBF otimizada para previsão da rugosidade com uso dos dados do Artigo 1

Fator de projeto	Nível ótimo do fator
Fator de escala do algoritmo Isotrópico para determinação do parâmetro de largura da função radial	1
Número de unidades radiais	21
Algoritmo para determinação do hiper-centro das unidades radiais	Sub-amostragem

Tabela 12 – Desempenho de rede neural RBF otimizada para previsão da rugosidade com uso dos dados do Artigo 1

Medida de Desempenho	Erro de previsão (MAE-DI%)
Valor médio	17,367
Desvio-padrão	16,916
Valor mínimo	15,325

Quando se compara os resultados da melhor rede RBF àqueles da melhor rede MLP por meio de análise de variância, ao nível de significância de 0,05, nota-se que o modelo RBF possui erro de predição significativamente superior. Isto pode ser observado pelo valor p resultante da ANOVA, igual a 0,000, mostrado na Figura 58. Seu desempenho é, portanto, inferior ao da MLP, quando aplicada ao mesmo conjunto de testes.

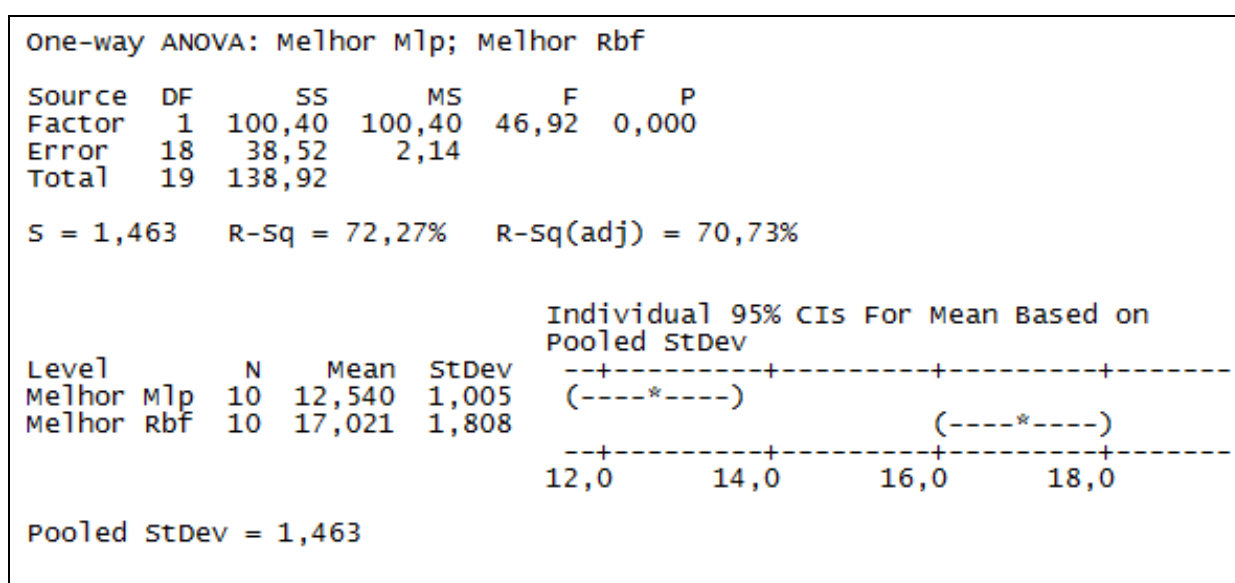


Figura 58 – Comparação entre redes otimizadas RBF e MLP - dados do Artigo 1

A comparação entre o desempenho da melhor rede RBF e o resultado do estudo original foi realizada por meio do teste de hipótese já descrito para a comparação feita com as redes MLP. O valor p resultante foi igual a 0,000, o que indica forte evidência de desempenho superior da configuração obtida por meio do método de otimização.

O mesmo teste anteriormente empregado para comparar o desempenho da melhor rede MLP ao desempenho do modelo de regressão foi repetido para a rede RBF. A equação de regressão utilizada foi a mesma. O teste resultou em um valor p igual a 0,000, o que evidencia a superioridade do modelo RBF obtido por meio do método proposto em relação ao modelo de regressão.

Uma outra comparação foi feita entre o desempenho da rede RBF obtida pelo método e a rede obtida por meio da aplicação do pacote computacional para otimização de redes, por meio da análise de variância, ao nível de significância de

0,05. A rede proposta pela ferramenta apresentou um valor da variável de saída, medido em MAE-DI%, igual a 38,527%. O valor p resultante da ANOVA foi igual a 0,000, o que evidencia uma diferença estatística significativa em favor da configuração de rede obtida pelo método proposto.

5.5.2 APLICAÇÃO DO MÉTODO AOS DADOS DO ARTIGO Nº 2

5.5.2.1 Valores iniciais e desenvolvimento

Para a aplicação do método utilizando o conjunto de dados do Artigo 2, utilizou o algoritmo de otimização com um *Número Máximo de Fases* igual a 4. Os valores de *Ponto Central do Arranjo e Intervalos Entre Níveis dos Fatores* foram inicializados conforme especificado nos algoritmos do capítulo 4.

A otimização foi obtida com a realização de 9 ciclos experimentais. A primeira fase encerrou-se após o ciclo 2, quando a análise dos resultados deixou de apontar redução do erro. A segunda, terceira e quarta fases encerraram-se após os ciclos 4, 6 e 9, respectivamente, pela mesma razão.

5.5.2.2 Resultados da aplicação do método

Os resultados da aplicação do método à otimização de redes neurais MLP para o conjunto de dados do Artigo 2 são mostrados a seguir. Na Tabela 13 são mostrados os níveis experimentais de cada fator empregados ao longo dos ciclos. Da mesma forma que nas seções anteriores, as cores verde e vermelha das colunas indicam os ciclos que correspondem ao início e ao fim de uma fase por falta de evidência estatística de redução do erro. Na Tabela 14 são mostrados os níveis dos fatores que compõem o centro do arranjo fatorial executado a cada ciclo para otimização das redes. O significado das colunas é o mesmo das tabelas similares mostradas anteriormente. Sob a coluna '*Variável de Saída (MAE-DI%)*' são mostrados os resultados de predição obtidos pela melhor configuração de rede de cada ciclo. O início e fim de cada ciclo

são indicados pelas cores verde e vermelha de cada linha, respectivamente.

Tabela 13 – Níveis dos fatores durante a aplicação do método para otimização de redes usando dados do Artigo 2

Fator	Nível	Fase 1		Fase 2		Fase 3		Fase 4		
		Número do ciclo								
		1	2	3	4	5	6	7	8	9
H1	1	0	8	12	8	10	12	13	12	11
	0	8	16	16	12	12	14	14	13	12
	-1	16	24	20	16	14	16	15	14	13
H2	1	0	0	4	0	2	0	1	0	0
	0	8	8	8	4	4	2	2	1	1
	-1	16	16	12	8	6	4	3	2	2
E1	1	0	32	48	32	40	48	52	56	52
	0	32	64	64	48	48	56	56	60	56
	-1	64	96	80	64	56	64	60	64	60
LR	1	0,001	0,001	0,200	0,001	0,100	0,200	0,250	0,200	0,250
	0	0,400	0,400	0,400	0,200	0,200	0,300	0,300	0,250	0,300
	-1	0,800	0,800	0,600	0,400	0,300	0,400	0,350	0,300	0,350
E2	1	0	64	96	64	80	96	104	112	104
	0	64	128	128	96	96	112	112	120	112
	-1	128	192	160	128	112	128	120	128	120

Tabela 14 – Pontos centrais e melhores configurações de rede obtidos ao longo da aplicação do método ao conjunto de dados do Artigo 2.

Fase	Ciclo	Centro do arranjo (unidades naturais)					Melhor configuração obtida no ciclo (unidades codificadas)					Variável de saída (MAE-DI%)
		H1	H2	E1	LR	E2	H1	H2	E1	LR	E2	
1	1	8	8	32	0,400	64	1	-1	1	-1	1	18,525
	2	16	8	64	0,400	128	-1	-1	-1	-1	1	18,986
2	3	16	8	64	0,400	128	-1	-1	-1	-1	-1	17,438
	4	12	4	48	0,200	96	-1	-1	1	1	-1	17,973
3	5	12	4	48	0,200	96	1	-1	1	1	1	16,908
	6	14	2	56	0,300	112	-1	-1	1	-1	-1	17,109
4	7	14	2	56	0,300	112	-1	-1	1	-1	1	17,829
	8	13	1	60	0,250	120	-1	-1	-1	1	-1	17,674
	9	12	1	56	0,300	112	1	-1	-1	-1	-1	17,801

Na Figura 59 é ilustrada a evolução de desempenho obtida ao longo do processo de otimização. Para cada ciclo (no eixo das abscissas) é indicado o desempenho da melhor configuração de rede obtida (no eixo das ordenadas), expresso em MAE-DI%. Ao final da aplicação do método a melhor configuração de rede obtida foi a especificada pela Tabela 15. O desempenho do melhor modelo de rede MLP é resumido na Tabela 16.



Figura 59 – Evolução de desempenho das melhores redes MLP a cada ciclo – aplicação do método ao conjunto de dados do Artigo 2

Tabela 15 – Configuração de rede neural MLP otimizada para predição da rugosidade usando os dados do Artigo 2

Fator de projeto	Nível ótimo do fator
Número de neurônios na camada oculta 1	14
Número de neurônios na camada oculta 2	2
Número de épocas de treinamento na fase 1	56
Taxa de aprendizado na fase 1	0,300
Número de épocas de treinamento na fase 2	112

Tabela 16 – Desempenho de rede neural MLP otimizada para predição da rugosidade com os dados do Artigo 2

Medida de Desempenho	Erro de predição (MAE-DI%)
Valor médio	16,908
Desvio-padrão	7,227
Valor mínimo	14,418

O sumário gráfico da distribuição do erro resultante da aplicação das 200 instâncias do modelo ótimo de rede à predição do conjunto dos casos de teste é ilustrado na Figura 60.

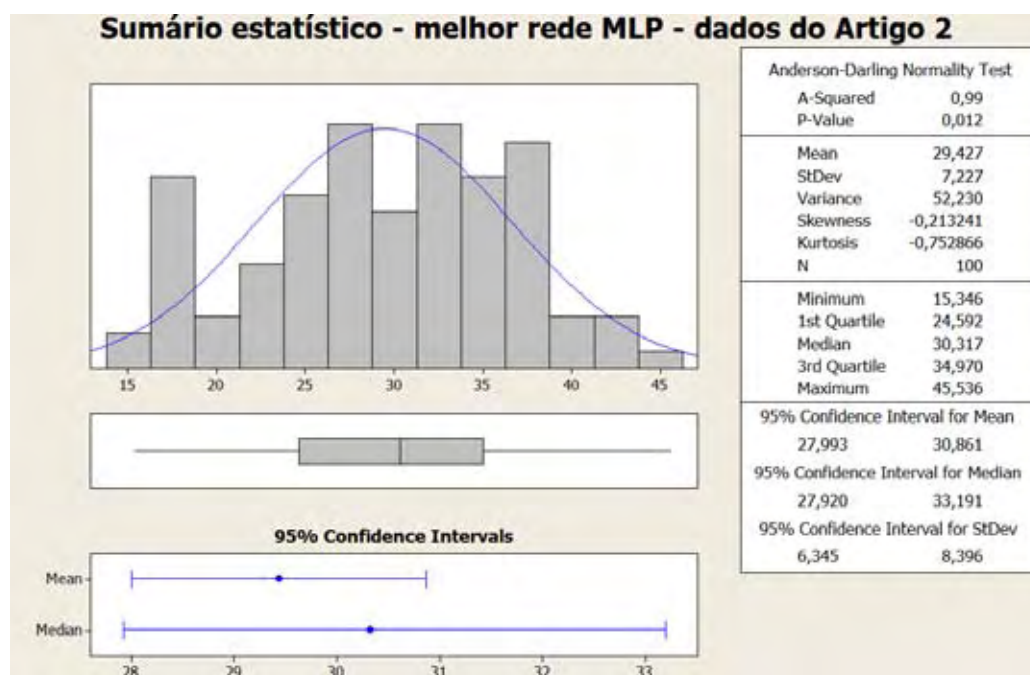


Figura 60 – Sumário gráfico da distribuição de erro da rede MLP ótima – experimento com dados do Artigo 2

5.5.2.3 Comparação com resultados do trabalho original

Comparou-se o desempenho da melhor configuração de rede obtida por meio da aplicação do método proposto ao resultado obtido pelo estudo que originou os dados utilizados. O valor do erro de predição atingido pela melhor rede na predição do

conjunto de casos de teste foi comparado ao valor correspondente calculado com base nos resultados do artigo da literatura, por meio de um teste de hipóteses. Utilizou-se o teste-t bilateral para hipótese nula de médias de erro iguais entre as redes ao nível de significância de 0,05. O valor da variável de saída da melhor rede obtida é igual a 16,908%. O valor correspondente, calculado para o Artigo 2, (utilizado como valor de comparação para o teste de hipótese) é igual a 15,028%. O resultado do teste, extraído do Minitab®, é mostrado na Figura 61.

One-Sample T: Metodo proposto							
Test of mu = 15,0282 vs not = 15,0282							
Variable	N	Mean	StDev	SE Mean	95% CI	T	P
Metodo proposto	10	16,5170	1,0966	0,3468	(15,7326; 17,3014)	4,29	0,002

Figura 61 – Teste de hipótese para médias iguais de erro de predição entre a melhor rede MLP obtida pelo método e resultado do Artigo 2

Como se observa na figura, o valor p resultante do teste é igual a 0,002, o que leva à rejeição da hipótese nula de médias iguais. Há, neste caso, evidência estatística de que média do erro de predição da configuração otimizada é superior à da rede proposta no artigo original. Observou-se porém que, apesar do resultado estatístico desfavorável obtido pela para a variável de saída, a melhor configuração de rede apresentou resultado melhor do que o declarado pelo artigo em algumas instâncias do teste, obtendo erro mínimo igual a 14,418%.

5.5.2.4 Comparação com resultados de modelo de regressão múltipla

O desempenho da melhor configuração de rede obtida foi comparado ao desempenho de um modelo de predição elaborado por meio de regressão múltipla, elaborado segundo os mesmos princípios utilizados para obtenção do modelo de regressão a partir dos dados do Artigo 1. Obteve-se desta forma o modelo mostrado na Equação 54, onde os regressores A , V_c , f e a_p são definidos como no Quadro 14, e estão

expressos em unidades naturais.

$$R_a = 1,01 + 0,00766*V_c + 0,396*a_p + 8,88*f + 0,219*A - 0,0325*V_c*f - 0,00145*V_c*A - 5,30*a_p*f + 0,515*f*A \quad (54)$$

Na Figura 62 mostra-se os resultados da análise de regressão efetuada sobre o modelo escolhido, com uso do pacote computacional Minitab®.

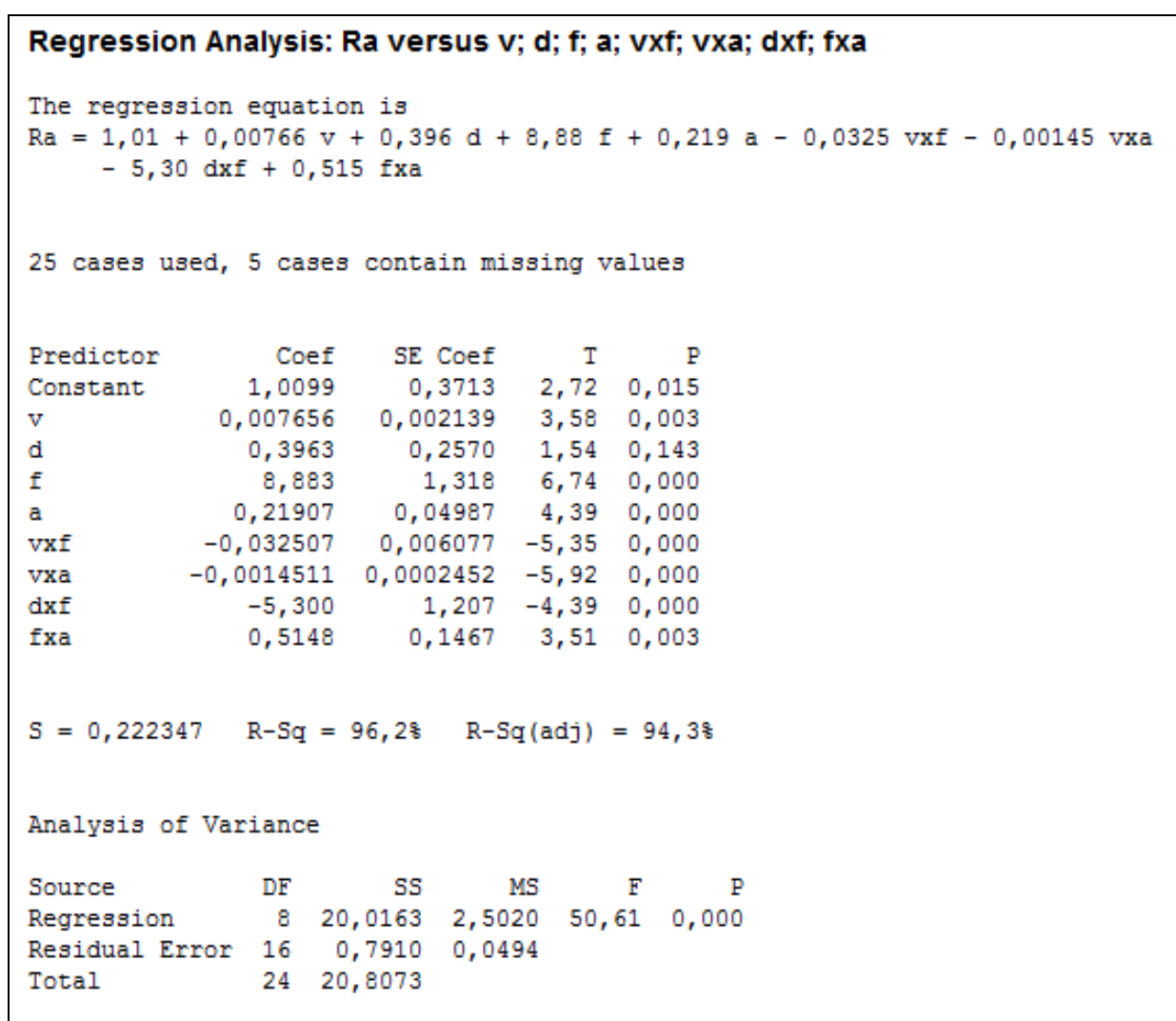


Figura 62 – Análise de regressão do conjunto de dados obtido do Artigo 2

Observa-se na Figura 62 o excelente ajuste da curva, indicado pelo valor de $R\text{-}Sq(adj)$ igual a 94,3% e pelo valor p da análise de variância para significância da regressão, igual a 0,000. O teste de normalidade de Anderson-Darling foi aplicado

sobre os resíduos e não indicou qualquer evidência de que os resíduos não sigam a distribuição normal.

A comparação entre a rede e o modelo de regressão foi feita por meio de um teste de hipótese entre os valores do erro de predição da melhor configuração de rede para o conjunto de teste (em MAE-DI%) e o valor do erro de predição obtido pelo modelo de regressão. Utilizou-se o teste-t bilateral para hipótese nula de médias de erro iguais ao nível de significância de 0.05. O valor do erro de predição calculado para o modelo de regressão (utilizado como valor de comparação para o teste de hipótese) é igual a 42,167%. O resultado do teste, extraído do Minitab®, é mostrado na Figura 63.

```

One-sample T: Método proposto
Test of mu = 42,1671 vs not = 42,1671

Variable      N    Mean   StDev  SE Mean      95% CI          T
Método proposto  10  16,9085  0,9534   0,3015  (16,2265; 17,5905)  -83,78

Variable      P
Método proposto  0,000

```

Figura 63 – Teste de hipótese para médias iguais de erro de predição em MAE-DI% da melhor rede MLP obtida pelo método e modelo de regressão para dados do Artigo 2

Como se observa na figura, o valor p resultante do teste é igual a 0,000, o que leva à rejeição da hipótese nula de médias iguais. Este resultado evidencia, ao nível de significância adotado, superioridade do desempenho da rede obtida por meio do método proposto. O resultado também é ilustrado na Figura 64, na qual se mostra o diagrama de caixa para o teste de hipótese. Na figura, \bar{X} representa a média da variável de saída (medida em MAE-DI%) para a rede otimizada e H_0 representa o resultado obtido pelo modelo de regressão.

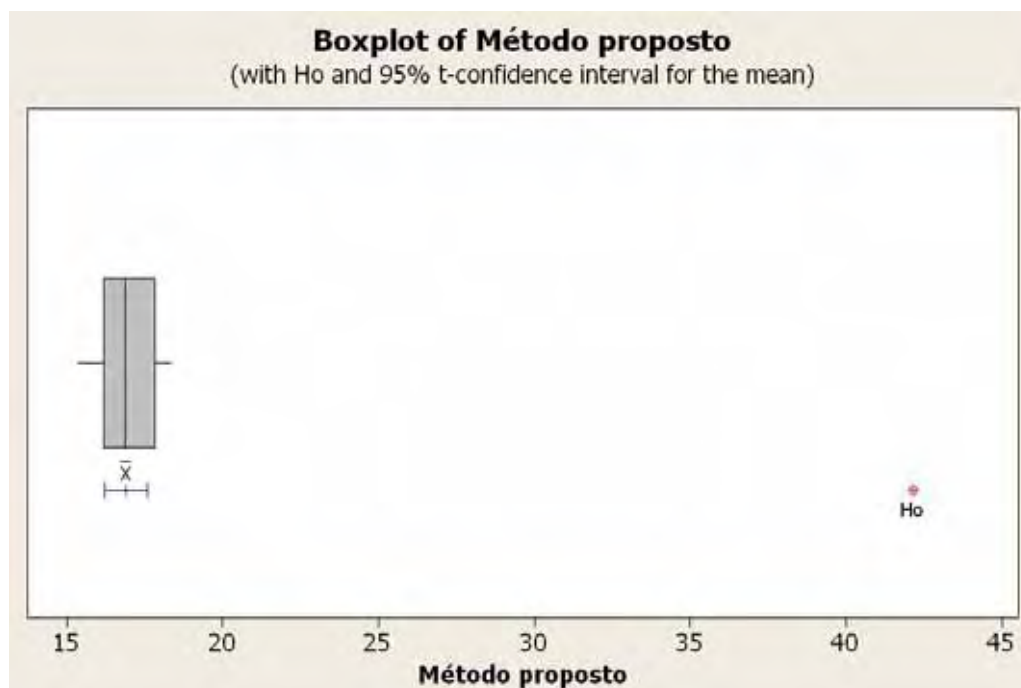


Figura 64 – Diagrama de caixa do teste de hipótese para comparação entre a melhor rede MLP obtida pelo método e modelo de regressão para o Artigo 2

5.5.2.5 Comparação com resultados da ferramenta automática de otimização de redes

Uma outra comparação foi realizada entre a melhor configuração de rede MLP obtida por meio da aplicação do método e uma arquitetura de rede otimizada pelo pacote computacional. O valor da variável de saída (em MAE-DI%) medido para a rede proposta pela ferramenta, foi igual a 25,288%. Utilizou-se o teste de análise de variância entre para comparar as previsões de erro das duas configurações de rede. O resultado é mostrado na Figura 65, na qual se mostra a tabela da ANOVA.

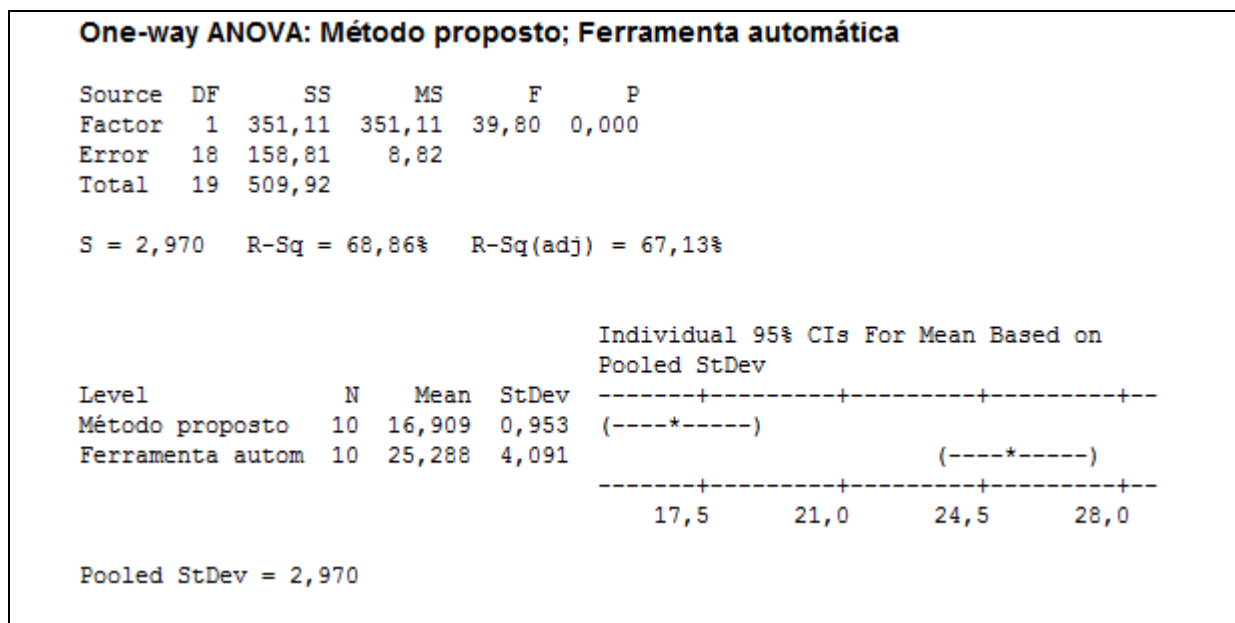


Figura 65 – Comparação entre erro de predição em MAE-DI% da melhor rede MLP obtida pelo método e da melhor rede apontada pelo pacote computacional para dados do Artigo 2

O resultado da ANOVA indica que, ao nível de significância adotado (igual a 0,05), existe evidência de diferença estatística entre os erros de predição das configurações de redes sob comparação. O erro de predição da rede obtida por meio do método é inferior ao obtido por meio do pacote computacional. Conclui-se, portanto, que o método proposto identificou uma configuração de rede com desempenho superior ao da ferramenta comercial utilizada. O resultado da comparação é adicionalmente ilustrado na Figura 66, na qual se mostra o diagrama de caixa das observações do erro de predição para as duas configurações de rede, para o conjunto dos casos de teste. Pode-se notar a diferença entre as médias e a distribuição dos erros de predição das redes sob comparação.

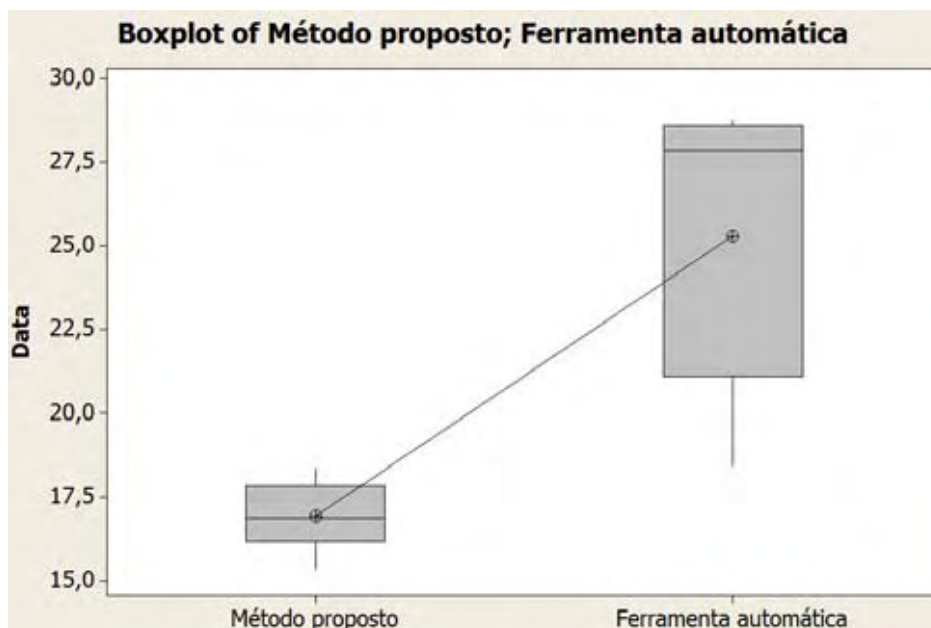


Figura 66 – Comparação entre a melhor rede MLP obtida pelo método e melhor rede apontada pelo pacote computacional para dados do Artigo 2

5.5.2.6 Reprodução do estudo com redes de arquitetura RBF

A aplicação do método à otimização de redes RBF foi realizada de modo similar ao empregado para a otimização de redes MLP. O *Número Máximo de Fases* utilizado foi igual a 4. Os valores de *Ponto Central do Arranjo* e *Intervalos Entre Níveis dos Fatores* foram inicializados conforme especificado nos algoritmos do capítulo 4. A otimização foi finalizada em 10 ciclos experimentais. A primeira fase encerrou-se após o ciclo 2, quando a análise apontou como ponto central do arranjo para o ciclo seguinte o mesmo conjunto de níveis dos fatores utilizado como ponto central naquele ciclo. A segunda fase encerrou-se após o ciclo 3, pelo mesmo motivo. As fases três e quatro foram encerradas após os ciclos 8 e 10, respectivamente, quando a análise dos resultados deixou de apontar redução do erro.

Na Figura 67 é mostrada a evolução de desempenho obtida ao longo do processo de otimização. Para cada ciclo (no eixo das abscissas) é indicado o desempenho da melhor configuração de rede obtida (no eixo das ordenadas), expresso em MAE-DI%.

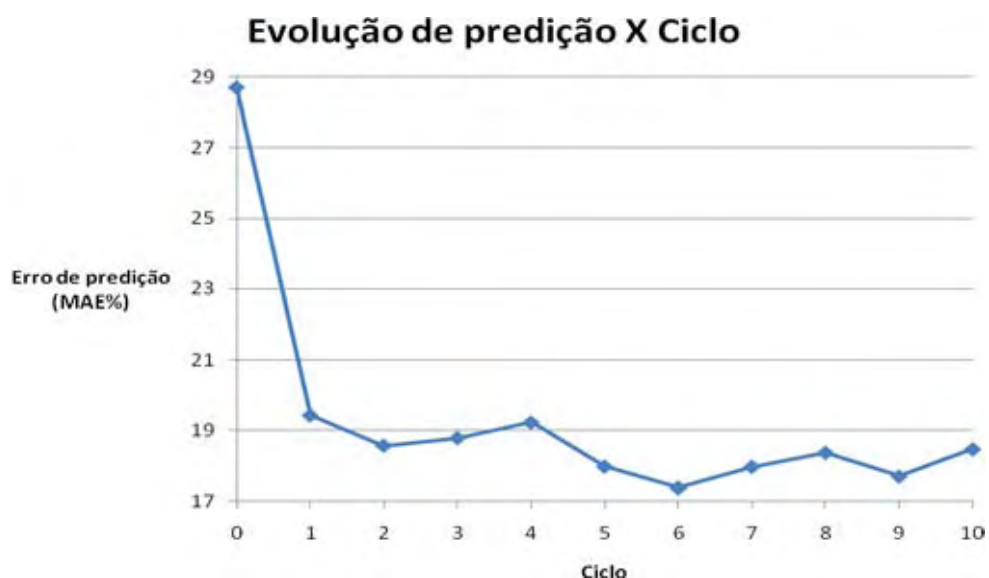


Figura 67 – Evolução de desempenho das melhores redes RBF a cada ciclo – aplicação do método ao conjunto de dados do Artigo 2

Ao final da aplicação do método a melhor configuração de rede obtida foi a especificada pela Tabela 17. O desempenho do melhor modelo de rede RBF é resumido na Tabela 18.

Tabela 17 – Configuração de rede neural RBF otimizada para previsão da rugosidade com uso dos dados do Artigo 2

Fator de projeto	Nível ótimo do fator
Fator de escala do algoritmo Isotrópico para determinação do parâmetro de largura da função radial	88
Número de unidades radiais	9
Algoritmo para determinação do hiper-centro das unidades radiais	K-Médias

Tabela 18 – Desempenho de rede neural RBF otimizada para previsão da rugosidade usando os dados do Artigo 2

Medida de Desempenho	Erro de previsão (MAE-DI%)
Valor médio	17,380
Desvio-padrão	21,817
Valor mínimo	16,432

Ao comparar-se resultados da melhor rede RBF àqueles da melhor MLP por análise de variância, ao nível de significância de 0,05, nota-se que não há evidência estatística de diferença entre o desempenho dos modelos, quando aplicados ao mesmo conjunto de testes. Isto se observa pelo valor p resultante da ANOVA, igual a 0,386, mostrado na Figura 68.

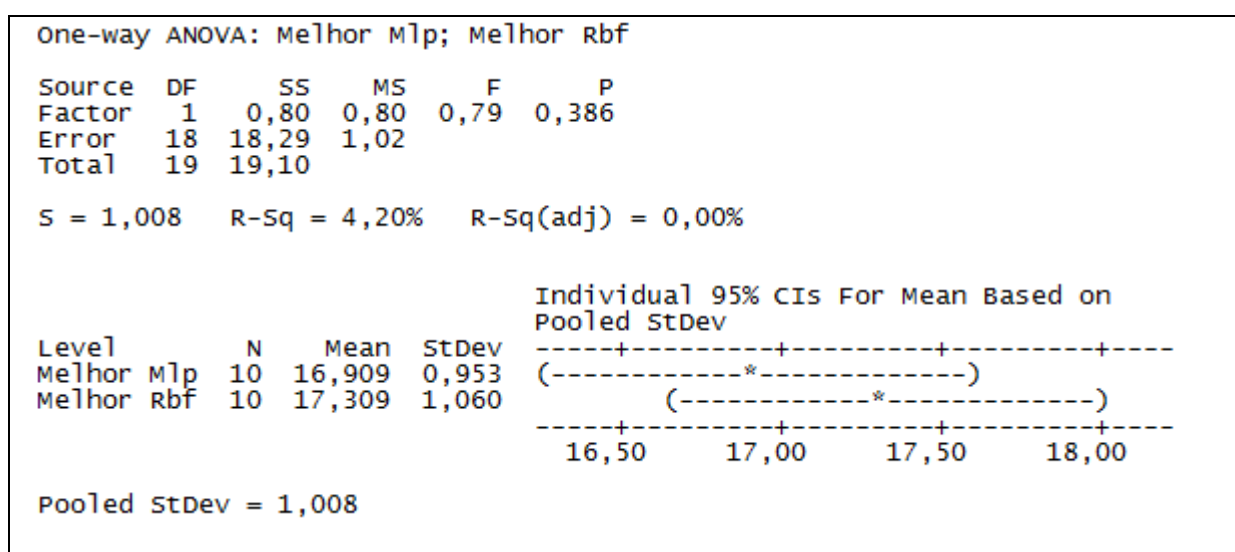


Figura 68 – Comparação entre rede otimizada RBF e rede otimizada MLP para dados do Artigo 2

A comparação entre o desempenho da melhor rede RBF e o resultado do estudo original foi realizada por meio do teste de hipótese já descrito para a comparação feita com as redes MLP. O valor p resultante foi igual a 0,000. Neste caso, tem-se forte evidência de diferença entre as redes, favorável ao resultado do estudo original.

O mesmo teste anteriormente empregado para comparar o desempenho da melhor rede MLP ao desempenho do modelo de regressão foi repetido para a rede RBF. A equação de regressão utilizada foi a mesma. O teste resultou em um valor p igual a 0,000, o que evidencia a superioridade do modelo RBF obtido por meio do método proposto em relação ao modelo de regressão.

Uma outra comparação foi feita entre o desempenho da rede RBF obtida pelo método e a rede obtida por meio da aplicação do pacote computacional para otimização de redes, por meio da análise de variância, ao nível de significância de

0,05. A rede proposta pela ferramenta apresentou um valor da variável de saída, medido em MAE-DI%, igual a 25,090%. O valor p resultante da ANOVA foi igual a 0,000, o que evidencia superioridade da configuração de rede obtida pelo método proposto.

5.5.3 APLICAÇÃO DO MÉTODO AOS DADOS DO ARTIGO Nº 3

5.5.3.1 Valores iniciais e desenvolvimento

A aplicação do método utilizando o conjunto de dados do Artigo 3 adotou um *Número Máximo de Fases* igual a 4. Os valores de *Ponto Central do Arranjo* e *Intervalos Entre Níveis dos Fatores* foram definidos conforme especificado nos algoritmos do capítulo 4.

A otimização levou à realização de 11 ciclos experimentais. As fases se encerraram após os ciclos 2, 5, 8 e 11, quando a análise dos resultados deixou de apontar redução do erro.

5.5.3.2 Resultados da aplicação do método

Os resultados da aplicação do método à otimização de redes neurais MLP para o conjunto de dados do Artigo 3 são mostrados na Tabela 19.

Tabela 19 – Níveis dos fatores durante a aplicação do método para otimização de redes usando dados do Artigo 3

Fator	Nível	Fase 1		Fase 2		Fase 3		Fase 4				
		Número do ciclo										
		1	2	3	4	5	6	7	8	9	10	11
H1	1	0	8	12	16	20	22	24	22	23	22	21
	0	8	16	16	20	24	24	26	24	24	23	22
	-1	16	24	20	24	28	26	28	26	25	24	23
H2	1	0	0	4	0	0	2	0	0	1	0	0
	0	8	8	8	4	4	4	2	2	2	1	1
	-1	16	16	12	8	8	6	4	4	3	2	2
E1	1	0	32	48	32	48	56	64	56	60	64	60
	0	32	64	64	48	64	64	72	64	64	68	64
	-1	64	96	80	64	80	72	80	72	68	72	68
LR	1	0,001	0,400	0,600	0,400	0,600	0,700	0,600	0,500	0,550	0,600	0,650
	0	0,400	0,800	0,800	0,600	0,800	0,800	0,700	0,600	0,600	0,650	0,700
	-1	0,800	1,200	1,000	0,800	1,000	0,900	0,800	0,700	0,650	0,700	0,750
E2	1	0	64	96	64	96	112	96	80	88	96	104
	0	64	128	128	96	128	128	112	96	96	104	112
	-1	128	192	160	128	160	144	128	112	104	112	120

Na Tabela 19 são detalhados os níveis experimentais de cada fator empregados ao longo dos ciclos. Assim como nas seções anteriores, as cores verde e vermelha das colunas indicam os ciclos que correspondem, respectivamente, ao início e ao fim de uma fase. Na Tabela 20 são mostrados os níveis dos fatores que compõem o centro do arranjo fatorial executado a cada ciclo para otimização das redes. O significado das colunas é o mesmo das tabelas similares mostradas anteriormente. Sob a coluna 'Variável de Saída (MAE-DI%)' são mostrados os resultados de predição obtidos pela melhor configuração de rede de cada ciclo. As cores verde e vermelha das linhas indicam os ciclos que correspondem, respectivamente, ao início e ao fim de uma fase.

Tabela 20 – Pontos centrais e melhores configurações de rede obtidos ao longo da aplicação do método ao conjunto de dados do Artigo 3.

Fase	Ciclo	Centro do arranjo (unidades naturais)					Melhor configuração obtida no ciclo (unidades codificadas)					Variável de saída (MAE- DI%)
		H1	H2	E1	LR	E2	H1	H2	E1	LR	E2	
1	1	8	8	32	0,400	64	1	-1	1	1	1	3,570
	2	16	8	64	0,800	128	-1	1	1	-1	-1	4,037
2	3	16	8	64	0,800	128	1	-1	-1	-1	-1	3,807
	4	20	4	48	0,600	96	1	-1	1	1	1	3,707
	5	24	4	64	0,800	128	1	-1	-1	-1	1	3,724
3	6	24	4	64	0,800	128	1	-1	1	-1	-1	3,877
	7	26	2	72	0,700	112	-1	-1	-1	-1	-1	3,321
	8	24	2	64	0,600	96	-1	-1	-1	-1	-1	3,697
4	9	24	2	64	0,600	96	-1	-1	1	1	1	3,867
	10	23	1	68	0,650	104	-1	-1	-1	1	1	3,796
	11	22	1	64	0,700	112	-1	-1	1	1	-1	3,832

Na Figura 69 pode ser observada a evolução de desempenho obtida ao longo do processo de otimização. Para cada ciclo (no eixo das abscissas) é indicado o desempenho da melhor configuração de rede obtida (no eixo das ordenadas), expresso em MAE-DI%. Ao final da aplicação do método a melhor configuração de rede obtida foi a especificada pela Tabela 21. O desempenho do melhor modelo de rede MLP é sumarizado na Tabela 22.



Figura 69 – Evolução de desempenho das melhores redes MLP a cada ciclo – aplicação do método ao conjunto de dados do Artigo 3

Tabela 21 – Configuração de rede neural MLP otimizada para predição da rugosidade usando os dados do Artigo 3

Fator de projeto	Nível ótimo do fator
Número de neurônios na camada oculta 1	24
Número de neurônios na camada oculta 2	0
Número de épocas de treinamento na fase 1	64
Taxa de aprendizado na fase 1	0,600
Número de épocas de treinamento na fase 2	96

Tabela 22 – Desempenho de rede neural MLP otimizada para predição da rugosidade com os dados do Artigo 3

Medida de Desempenho	Erro de predição (MAE-DI%)
Valor médio	3,321
Desvio-padrão	2,095
Valor mínimo	1,399

O sumário gráfico da distribuição do erro resultante da aplicação das 200 instâncias do modelo ótimo de rede à predição do conjunto dos casos de teste é ilustrado na Figura 70.

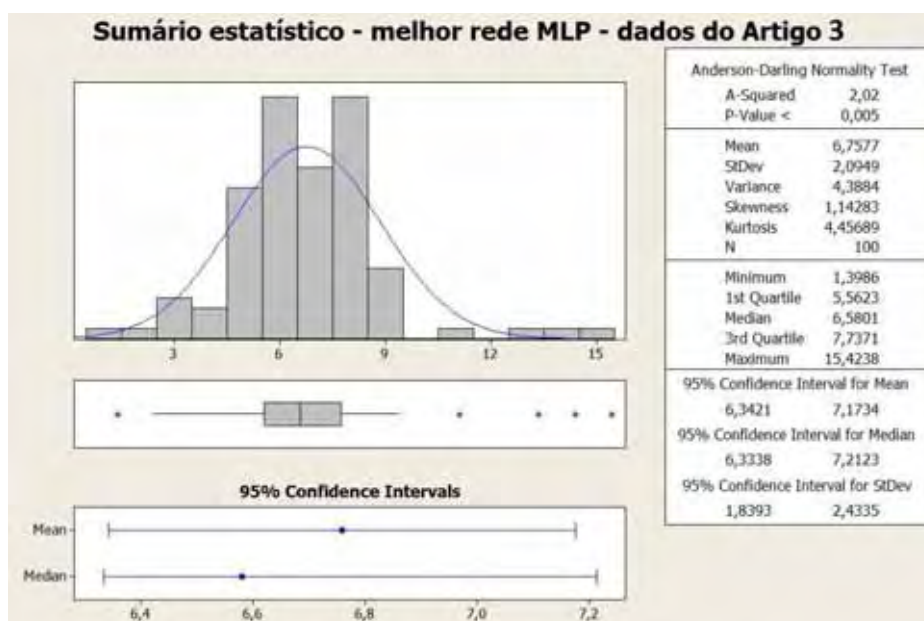


Figura 70 – Sumário gráfico da distribuição de erro da rede MLP ótima – experimento com dados do Artigo 3

5.5.3.3 Comparação com resultados do trabalho original

Comparou-se o desempenho da melhor configuração de rede obtida por meio da aplicação do método proposto ao resultado obtido pelo estudo que originou os dados utilizados. O valor do erro de predição atingido pela melhor rede na predição do conjunto de casos de teste foi comparado ao valor correspondente calculado com base nos resultados do artigo da literatura, por meio de um teste de hipóteses. Utilizou-se o teste-t bilateral para hipótese nula de médias de erro iguais entre as redes ao nível de significância de 0,05. O valor da variável de saída da melhor rede obtida é igual a 3,321%. O valor correspondente, calculado para o Artigo 3, (utilizado como valor de comparação para o teste de hipótese) é igual a 4,899%. O resultado do teste, extraído do Minitab®, é mostrado na Figura 71.

One-sample T: Método proposto						
Test of mu = 4,89936 vs not = 4,89936						
Variable	N	Mean	StDev	SE Mean	95% CI	T
Método proposto	10	3,32069	0,87063	0,27532	(2,69787; 3,94350)	-5,73
Variable	P					
Método proposto	0,000					

Figura 71 – Teste de hipótese para médias iguais de erro de predição entre a melhor rede MLP obtida pelo método e resultado do Artigo 3

Como se observa na figura, o valor p resultante do teste é igual a 0,000, o que leva à rejeição da hipótese nula de médias iguais. Há forte evidência estatística de que o desempenho obtido pela configuração de rede resultante do método de otimização proposto é superior ao desempenho da rede proposta no artigo original. O resultado do teste de hipótese também é ilustrado na Figura 72, na qual se mostra o diagrama de caixa para o teste de hipótese. Na figura, \bar{X} representa a média da variável de saída (medida em MAE-DI%) para a rede otimizada e H_0 representa o resultado declarado no artigo da literatura.

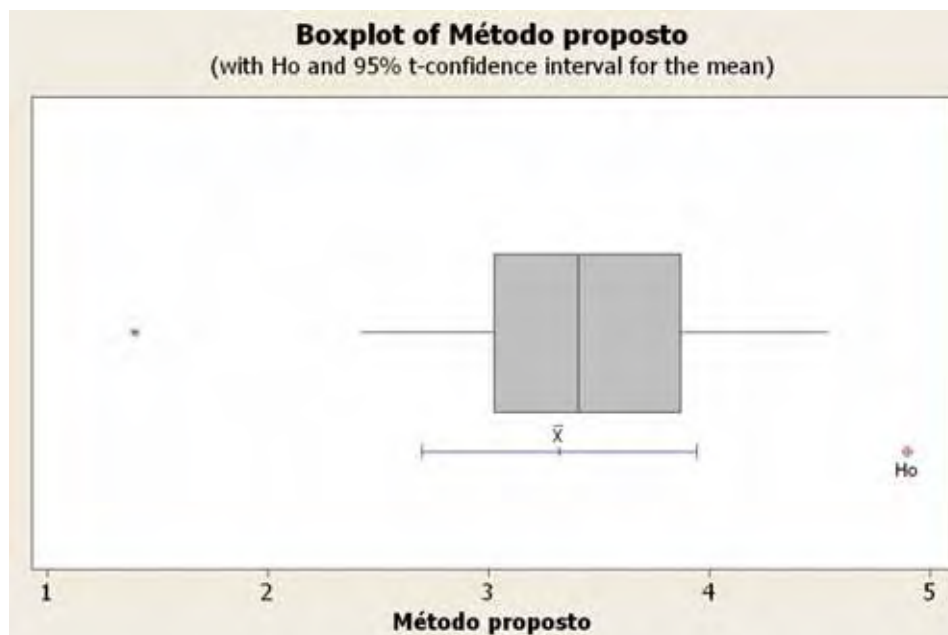


Figura 72 – Diagrama de caixa do teste de hipótese entre erro de predição da melhor rede MLP obtida pelo método e resultado do Artigo 3

Observou-se incidentalmente que, além do resultado estatístico obtido pela melhor configuração de rede para a variável de saída ser superior ao do estudo original, a mesma apresentou resultado ainda melhor em algumas instâncias do teste, chegando a apresentar erro mínimo igual a 1,399%.

5.5.3.4 Comparação com resultados de modelo de regressão múltipla

O desempenho da melhor configuração de rede obtida foi comparado ao desempenho de um modelo de predição elaborado por meio de regressão múltipla, elaborado segundo os mesmos princípios utilizados para obtenção do modelo de regressão a partir dos dados do Artigo 1. Obteve-se desta forma o modelo mostrado na Equação 55, onde os regressores T_{on} , T_{off} , I_p , WT , SV , FR e MS são definidos como no Quadro 15, e estão expressos em unidades naturais.

$$\begin{aligned}
R_a = & 4,30 - 0,134*T_{ON} + 0,0358*T_{OFF} - 0,00182*I_P - 0,00536*WT \\
& + 0,0469*SV - 0,0671*FR + 0,103*MS + 0,158*T_{ON}*MS \\
& - 0,000948*T_{OFF}*SV + 0,000013*I_P*WT - 0,00231 I_P*FR \\
& + 0,00372*I_P*MS + 0,000414*WT*FR - 0,000203*WT*MS \quad (55)
\end{aligned}$$

Na Figura 73 mostra-se os resultados da análise de regressão efetuada sobre o modelo escolhido, com uso do pacote computacional Minitab®. Observa-se o excelente ajuste da curva, indicado pelo valor de $R\text{-}Sq(adj)$ igual a 99,9% e pelo valor p da análise de variância para significância da regressão, igual a 0,000. O teste de normalidade de Anderson-Darling foi aplicado sobre os resíduos e não indicou qualquer evidência de que os resíduos não sigam a distribuição normal.

Regression Analysis: Ra versus TON; TOFF; ...					
The regression equation is					
Ra = 4,30 - 0,134 TON + 0,0358 TOFF - 0,00182 IP - 0,00536 WT + 0,0469 SV					
- 0,0671 FR + 0,103 MS + 0,158 TONxMS - 0,000948 TOFFxSV + 0,000013 IPxWT					
- 0,00231 IPxFR + 0,00372 IPxMS + 0,000414 WTxFR - 0,000203 WTxMS					
Predictor	Coef	SE Coef	T	P	
Constant	4,3002	0,1686	25,50	0,000	
TON	-0,13351	0,07655	-1,74	0,179	
TOFF	0,035836	0,001485	24,13	0,000	
IP	-0,0018185	0,0006530	-2,78	0,069	
WT	-0,0053619	0,0001698	-31,57	0,000	
SV	0,046862	0,004262	11,00	0,002	
FR	-0,06710	0,01587	-4,23	0,024	
MS	0,10273	0,02749	3,74	0,033	
TONxMS	0,15750	0,03564	4,42	0,022	
TOFFxSV	-0,0009484	0,0001662	-5,71	0,011	
IPxWT	0,00001343	0,00000051	26,37	0,000	
IPxFR	-0,00230735	0,00008631	-26,73	0,000	
IPxMS	0,0037153	0,0001427	26,03	0,000	
WTxFR	0,00041402	0,00001695	24,42	0,000	
WTxMS	-0,00020298	0,00002215	-9,17	0,003	
S = 0,00556537 R-Sq = 100,0% R-Sq(adj) = 99,9%					
Analysis of Variance					
Source	DF	SS	MS	F	P
Regression	14	1,008307	0,072022	2325,29	0,000
Residual Error	3	0,000093	0,000031		
Total	17	1,008400			

Figura 73 – Análise de regressão do conjunto de dados obtido do Artigo 3

A comparação entre a rede e o modelo de regressão foi feita por meio de um teste de hipótese entre os valores do erro de predição da melhor configuração de rede para o conjunto de teste (MAE-DI%) e o valor do erro de predição obtido pelo modelo de regressão. Utilizou-se o teste-t bilateral para hipótese nula de médias de erro iguais ao nível de significância de 0.05. O valor do erro de predição calculado para o modelo de regressão (utilizado como valor de comparação para o teste de hipótese) é igual a 11,173%. O resultado do teste, extraído do Minitab®, é mostrado na Figura 74.

One-Sample T: Método proposto						
Test of mu = 11,1734 vs not = 11,1734						
Variable	N	Mean	StDev	SE Mean	95% CI	T
Método proposto	10	3,32069	0,87063	0,27532	(2,69787; 3,94350)	-28,52
Variable	P					
Método proposto	0,000					

Figura 74 – Teste de hipótese para médias iguais de erro de predição em MAE-DI% da melhor rede MLP obtida pelo método e modelo de regressão para dados do Artigo 3

Conforme se observa na figura, o valor p resultante do teste é igual a 0,000, o que leva à rejeição da hipótese nula de médias iguais. Este resultado evidencia, ao nível de significância adotado, superioridade do desempenho da rede obtida por meio do método proposto. O resultado é adicionalmente ilustrado na Figura 75, na qual se mostra o diagrama de caixa para o teste de hipótese. Na figura, \bar{X} representa a média da variável de saída (medida em MAE-DI%) para a rede otimizada e H_0 representa o resultado obtido pelo modelo de regressão.

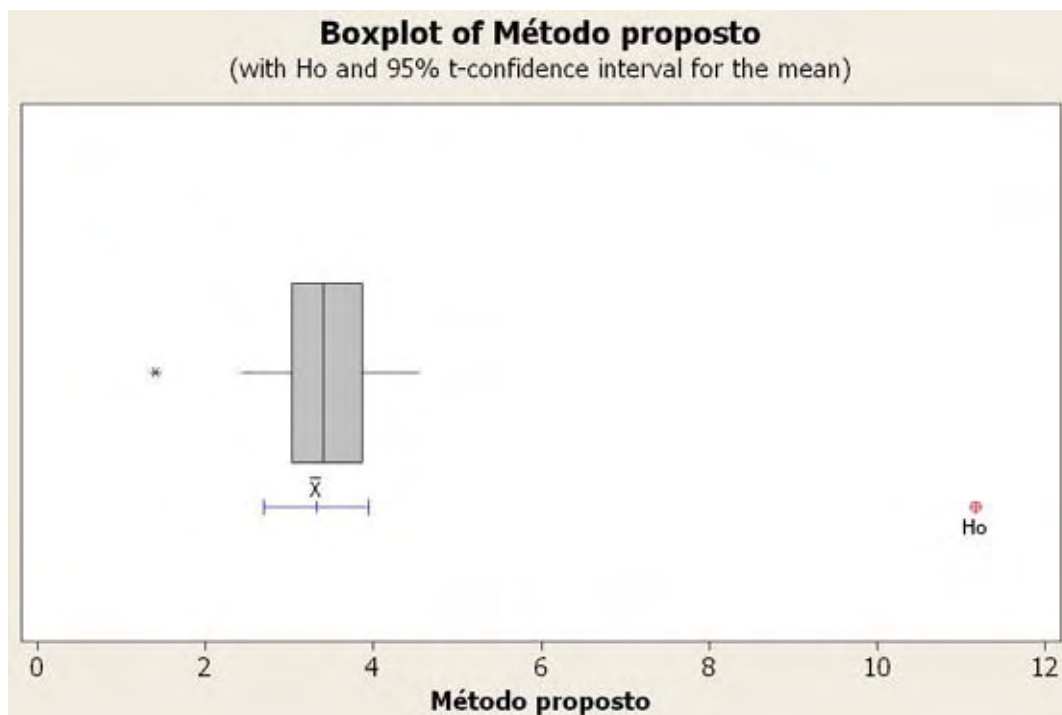


Figura 75 – Diagrama de caixa do teste de hipótese para comparação entre a melhor rede MLP obtida pelo método e modelo de regressão para o Artigo 3

5.5.3.5 Comparação com resultados da ferramenta automática de otimização de redes

Uma outra comparação foi realizada entre a melhor configuração de rede MLP obtida por meio da aplicação do método e uma arquitetura de rede otimizada pelo pacote computacional. O valor da variável de saída (em MAE-DI%) medido para a rede proposta pela ferramenta, foi igual a 6,189%. Utilizou-se o teste de análise de variância entre para comparar as predições de erro das duas configurações de rede. O resultado é mostrado na Figura 76, onde se pode observar a tabela da ANOVA.

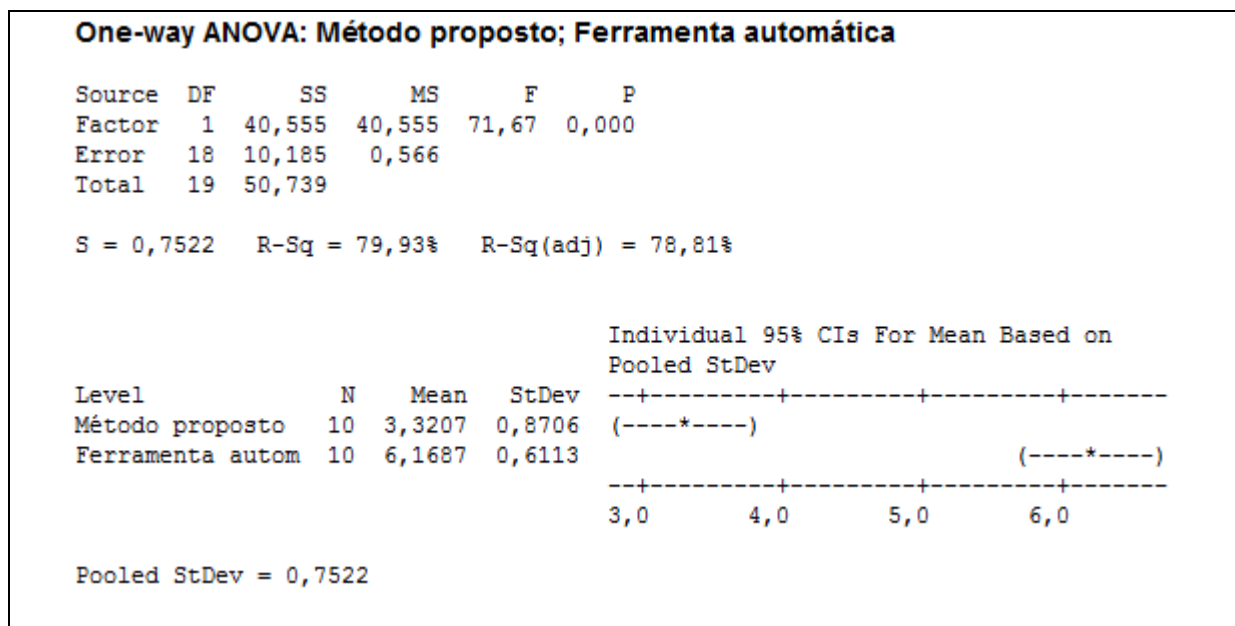


Figura 76 – Comparação entre erro de predição em MAE-DI% da melhor rede MLP obtida pelo método e da melhor rede apontada pelo pacote computacional para dados do Artigo 3

O resultado da ANOVA indica que, ao nível de significância adotado (igual a 0,05), há evidência de diferença estatística entre os erros de predição das configurações de redes sob comparação. O erro de predição da rede obtida por meio do método é inferior ao obtido por meio do pacote computacional. Conclui-se que o método proposto foi capaz de identificar uma configuração de rede com desempenho superior ao da ferramenta comercial utilizada. A comparação é adicionalmente ilustrada na Figura 77, na qual se mostra o diagrama de caixa das observações do erro de predição para as duas configurações de rede, para o conjunto dos casos de teste.

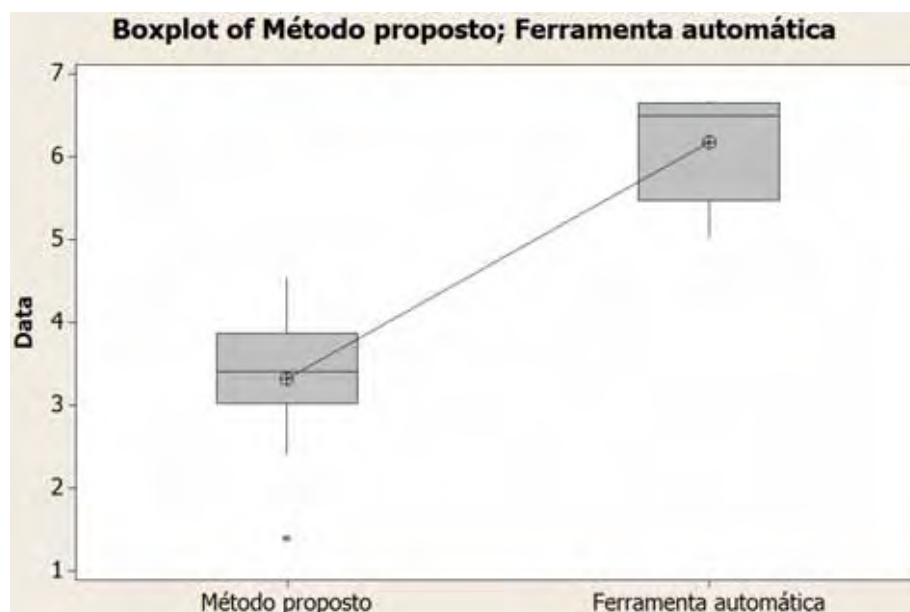


Figura 77 – Comparação entre a melhor rede MLP obtida pelo método e melhor rede apontada pelo pacote computacional para dados do Artigo 3

5.5.3.6 Reprodução do estudo com redes de arquitetura RBF

A aplicação do método à otimização de redes RBF foi realizada de modo similar ao empregado para a otimização de redes MLP. O valor de *Número Máximo de Fases* utilizado foi igual a 4. Os valores de *Ponto Central do Arranjo e Intervalos Entre Níveis dos Fatores* foram inicializados conforme especificado nos algoritmos do capítulo 4. O processo foi realizado em realização de 8 ciclos experimentais. A primeira fase encerrou-se após o ciclo 2, quando a análise dos resultados deixou de apontar redução do erro. A segunda fase encerrou-se após o ciclo 3, quando a análise apontou como ponto central do arranjo para o ciclo seguinte o mesmo conjunto de níveis dos fatores utilizado como ponto central naquele ciclo. A terceira fase encerrou-se após o ciclo 6, quando a análise dos resultados deixou de apontar redução do erro. A quarta fase encerrou-se ao final do ciclo 8, quando a análise apontou como ponto central do arranjo para o ciclo seguinte o mesmo conjunto de níveis dos fatores utilizado como ponto central naquele ciclo.

Na Figura 78 é ilustrada a evolução de desempenho obtida ao longo do processo de otimização. Para cada ciclo (no eixo das abscissas) é indicado o desempenho da

melhor configuração de rede obtida (no eixo das ordenadas), expresso em MAE-DI%. Ao final da aplicação do método a melhor configuração de rede obtida foi a especificada pela Tabela 23. O desempenho do melhor modelo de rede RBF é sumarizado na Tabela 24.



Figura 78 – Evolução de desempenho das melhores redes RBF a cada ciclo – aplicação do método ao conjunto de dados do Artigo 3

Tabela 23 – Configuração de rede neural RBF otimizada para predição da rugosidade com uso dos dados do Artigo 3

Fator de projeto	Nível ótimo do fator
Fator de escala do algoritmo Isotrópico para determinação do parâmetro de largura da função radial	64
Número de unidades radiais	6
Algoritmo para determinação do hiper-centro das unidades radiais	K-Médias

Tabela 24 – Desempenho de rede neural RBF otimizada para predição da rugosidade usando os dados do Artigo 3

Medida de Desempenho	Erro de predição (MAE-DI%)
Valor médio	3,855
Desvio-padrão	1,215
Valor mínimo	3,307

Quando se compara os resultados da melhor rede RBF àqueles da melhor rede MLP por meio de análise de variância, ao nível de significância de 0,05, nota-se que não há evidência estatística de diferença entre o desempenho dos modelos quando aplicados ao mesmo conjunto de testes, como indicado pelo valor p resultante da ANOVA, igual a 0,090.

A comparação entre o desempenho da melhor rede RBF e o resultado do estudo original foi realizada por meio do teste de hipótese já descrito para a comparação feita com as redes MLP. O valor p resultante foi igual a 0,000, o que indica forte evidência de desempenho superior da configuração obtida por meio do método proposto.

O mesmo teste anteriormente empregado para comparar o desempenho da melhor rede MLP ao desempenho do modelo de regressão foi repetido para a rede RBF. A equação de regressão utilizada foi a mesma. O teste resultou em um valor p igual a 0,000, o que evidencia a superioridade do modelo RBF obtido por meio do método proposto em relação ao modelo de regressão.

Uma outra comparação foi feita entre o desempenho da rede RBF obtida pelo método e a rede obtida por meio da aplicação do pacote computacional para otimização de redes, por meio da análise de variância, ao nível de significância de 0,05. A rede proposta pela ferramenta apresentou um valor da variável de saída, medido em MAE-DI%, igual a 5,920%. O valor p resultante da ANOVA foi igual a 0,000, o que evidencia superioridade da configuração de rede RBF obtida pelo método proposto.

5.6 APLICAÇÃO DO MÉTODO A RESULTADOS EXPERIMENTAIS NO TORNEAMENTO DE AÇO DE CORTE FÁCIL ABNT 12L14

5.6.1 VALORES INICIAIS E DESENVOLVIMENTO

Para a aplicação do método utilizando o conjunto de dados formado a partir dos resultados do experimento de usinagem aplicou-se o algoritmo de otimização a redes de arquitetura MLP. Utilizou um *Número Máximo de Fases* igual a 4. Os valores de *Ponto Central do Arranjo* e *Intervalos Entre Níveis dos Fatores* foram inicializados conforme especificado nos algoritmos do capítulo 4.

A otimização foi obtida com a realização de 13 ciclos experimentais. A primeira fase encerrou-se após o ciclo 3, quando a análise dos resultados deixou de apontar redução do erro. A segunda fase encerrou-se após o ciclo 6, quando o resultado da análise determinou a volta ao ponto central do ciclo anterior. A terceira e quarta fases encerraram-se após os ciclos 9 e 13, respectivamente, quando a análise dos resultados deixou de apontar redução do erro.

5.6.2 RESULTADOS DA APLICAÇÃO DO MÉTODO

Os resultados da aplicação do método à otimização de redes neurais para o conjunto de dados formado a partir dos resultados do experimento de usinagem são mostrados a seguir. A Tabela 25 contém a informação sobre os níveis experimentais de cada fator empregados ao longo dos ciclos. As cores verde e vermelha das colunas da tabela representam os ciclos que correspondem, respectivamente, ao início e ao fim de uma fase.

Tabela 25 – Níveis dos fatores durante a aplicação do método para otimização de redes usando resultados dos experimentos de usinagem

Fator	Nível	Fase 1		Fase 2			Fase 3			Fase 4				
		Número de ciclo												
		1	2	3	4	5	6	7	8	9	10	11	12	13
H1	1	1	16	1	8	1	8	4	8	4	6	4	6	4
	0	16	32	16	16	8	16	8	12	8	8	6	8	6
	-1	32	48	32	24	16	24	12	16	12	10	8	10	8
H2	1	0	16	32	40	48	40	52	48	52	54	56	54	56
	0	16	32	48	48	56	48	56	52	56	56	58	56	58
	-1	32	48	64	56	64	56	60	56	60	58	60	58	60
E1	1	0	0	32	48	32	16	40	48	56	60	64	60	56
	0	32	32	64	64	48	32	48	56	64	64	68	64	60
	-1	64	64	96	80	64	48	56	64	72	68	72	68	64
LR	1	0,001	0,001	0,400	0,600	0,400	0,200	0,500	0,400	0,300	0,350	0,300	0,350	0,400
	0	0,400	0,400	0,800	0,800	0,600	0,400	0,600	0,500	0,400	0,400	0,350	0,400	0,450
	-1	0,800	0,800	1,200	1,000	0,800	0,600	0,700	0,600	0,500	0,450	0,400	0,450	0,500
E2	1	0	0	0	32	0	0	16	32	16	24	32	40	32
	0	64	64	64	64	32	32	32	48	32	32	40	48	40
	-1	128	128	128	96	64	64	48	64	48	40	48	56	48

Na Tabela 26 são mostrados os níveis dos fatores que compõem o centro do arranjo fatorial executado a cada ciclo para otimização das redes. O significado das colunas é o mesmo das tabelas similares mostradas anteriormente. Sob a coluna 'Variável de Saída (MAE-DI%)' são mostrados os resultados de predição obtidos pela melhor configuração de rede de cada ciclo. As cores verde e vermelha das linhas da tabela representam os ciclos que correspondem, respectivamente, ao início e ao fim de uma fase. Na Figura 79 é mostrada a evolução de desempenho obtida ao longo do processo de otimização. Para cada ciclo (no eixo das abscissas) é indicado o desempenho da melhor configuração de rede obtida (no eixo das ordenadas), expresso em MAE-DI%.

Tabela 26 – Pontos centrais e melhores configurações de rede obtidos ao longo da aplicação do método aos resultados dos experimentos de usinagem.

Fase	Ciclo	Centro do arranjo (unidades naturais)					Melhor configuração obtida no ciclo (unidades codificadas)					Variável de saída (MAE- DI%)
		H1	H2	E1	LR	E2	H1	H2	E1	LR	E2	
1	1	16	16	32	0,400	64	1	1	-1	-1	-1	27,844
	2	32	32	32	0,400	64	-1	1	1	1	-1	26,948
	3	16	48	64	0,800	64	1	-1	1	1	-1	28,164
2	4	16	48	64	0,800	32	-1	1	-1	-1	-1	27,976
	5	8	56	48	0,600	32	1	-1	-1	-1	-1	27,231
	6	16	48	32	0,400	32	-1	1	1	1	-1	26,934
3	7	8	56	48	0,600	32	1	-1	1	-1	1	26,341
	8	12	52	56	0,500	48	-1	1	1	-1	-1	26,229
	9	8	56	64	0,400	32	-1	1	-1	1	1	27,060
4	10	8	56	64	0,400	32	-1	1	1	-1	1	26,995
	11	6	58	68	0,350	40	1	-1	-1	1	1	26,279
	12	8	56	64	0,400	48	-1	1	-1	1	-1	25,760
	13	6	58	60	0,450	40	1	1	1	1	1	26,541



Figura 79 – Evolução de desempenho das melhores redes MLP a cada ciclo – aplicação do método aos resultados dos experimentos de usinagem

Ao final da aplicação do método a melhor configuração de rede obtida foi a especificada pela Tabela 27. O desempenho do melhor modelo de rede MLP é

sumarizado na Tabela 28. O sumário gráfico da distribuição do erro resultante da aplicação das 200 instâncias do modelo ótimo de rede à predição do conjunto dos casos de teste é ilustrado na Figura 80.

Tabela 27 – Configuração de rede neural MLP otimizada para predição da rugosidade usando os resultados dos experimentos de usinagem

Fator de projeto	Nível ótimo do fator
Número de neurônios na camada oculta 1	6
Número de neurônios na camada oculta 2	58
Número de épocas de treinamento na fase 1	60
Taxa de aprendizado na fase 1	0,450
Número de épocas de treinamento na fase 2	40

Tabela 28 – Desempenho de rede neural MLP otimizada para predição da rugosidade usando os resultados dos experimentos de usinagem

Medida de Desempenho	Erro de predição (MAE-DI%)
Valor médio	25,760
Desvio-padrão	12,555
Valor mínimo	20,536

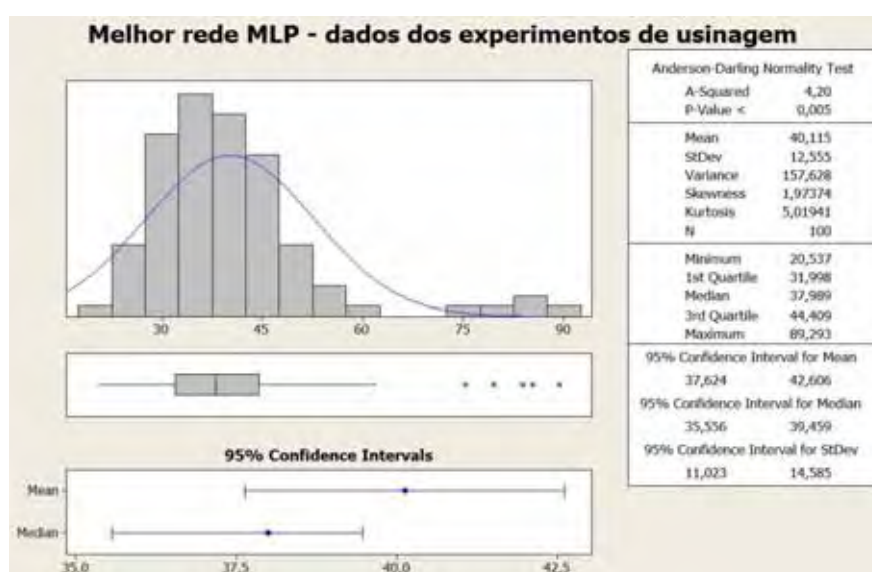


Figura 80 – Sumário gráfico da distribuição de erro da rede MLP ótima – experimento com resultados dos experimentos de usinagem

5.6.3 COMPARAÇÃO COM RESULTADOS DE MODELO DE REGRESSÃO MÚLTIPLA

O desempenho da melhor configuração de rede obtida foi comparado ao desempenho de um modelo de predição elaborado por meio de regressão múltipla, elaborado segundo os mesmos princípios utilizados para obtenção do modelo de regressão a partir dos dados do Artigo 1. Obteve-se desta forma o modelo mostrado na Equação 56, onde os regressores V_c, f, a_p, E, VB e P são definidos como no Quadro 17, e estão expressos em unidades naturais.

$$\begin{aligned}
 R_a = & 3,19 - 0,0111 * V_c - 4,87 * f + 5,87 * a_p - 0,541 * E - 4,53 * VB - 2,05 * P \\
 & + 0,00474 * V_c * E + 0,00839 * V_c * VB + 0,00522 * V_c * P - 19,5 * f * AP \\
 & + 33,9 * f * VB + 10,4 * f * P - 5,08 * a_p * E - 7,57 * a_p * VB - 1,97 * a_p * P \\
 & - 0,112 * E * P + 1,69 * VB * P \\
 & - 0,00396 * V_c * a_p * E + 0,0237 * V_c * a_p * VB \\
 & + 0,00301 * V_c * E * VB - 0,00393 * V_c * VB * P + 4,16 * f * a_p * E \\
 & + 4,01 * f * a_p * P - 1,42 * f * E * VB - 0,24 * f * E * P - 8,52 * f * VB * P \\
 & + 8,94 * a_p * E * VB + 2,68 * a_p * E * P + 3,76 * a_p * VB * P \\
 & + 0,0957 * V_c * f * a_p * E - 0,0291 * V_c * f * a_p * VB \\
 & - 0,0214 * V_c * a_p * E * VB - 0,0116 * V_c * a_p * VB * P \\
 & - 2,60 * a_p * E * VB * P \\
 & - 0,0871 * V_c * F * AP * E * D - 0,0503 * V_c * F * AP * E * P \\
 & + 0,0062 * V_c * F * AP * D * P + 0,00705 * V_c * AP * E * D * P \\
 & - 16,9 * F * AP * E * D * P + 0,0709 * V_c * f * AP * E * VB * P
 \end{aligned} \tag{56}$$

Na Figura 81 mostra-se os resultados da análise de regressão efetuada sobre o modelo escolhido, com uso do pacote computacional Minitab®. Observa-se o elevado nível de ajuste da curva, indicado pelo valor de $R-Sq(adj)$ igual a 82,3% e pelo valor p

da análise de variância para significância da regressão, igual a 0,000. O teste de normalidade de Anderson-Darling foi aplicado sobre os resíduos e não indicou qualquer evidência de que os resíduos não sigam a distribuição normal.

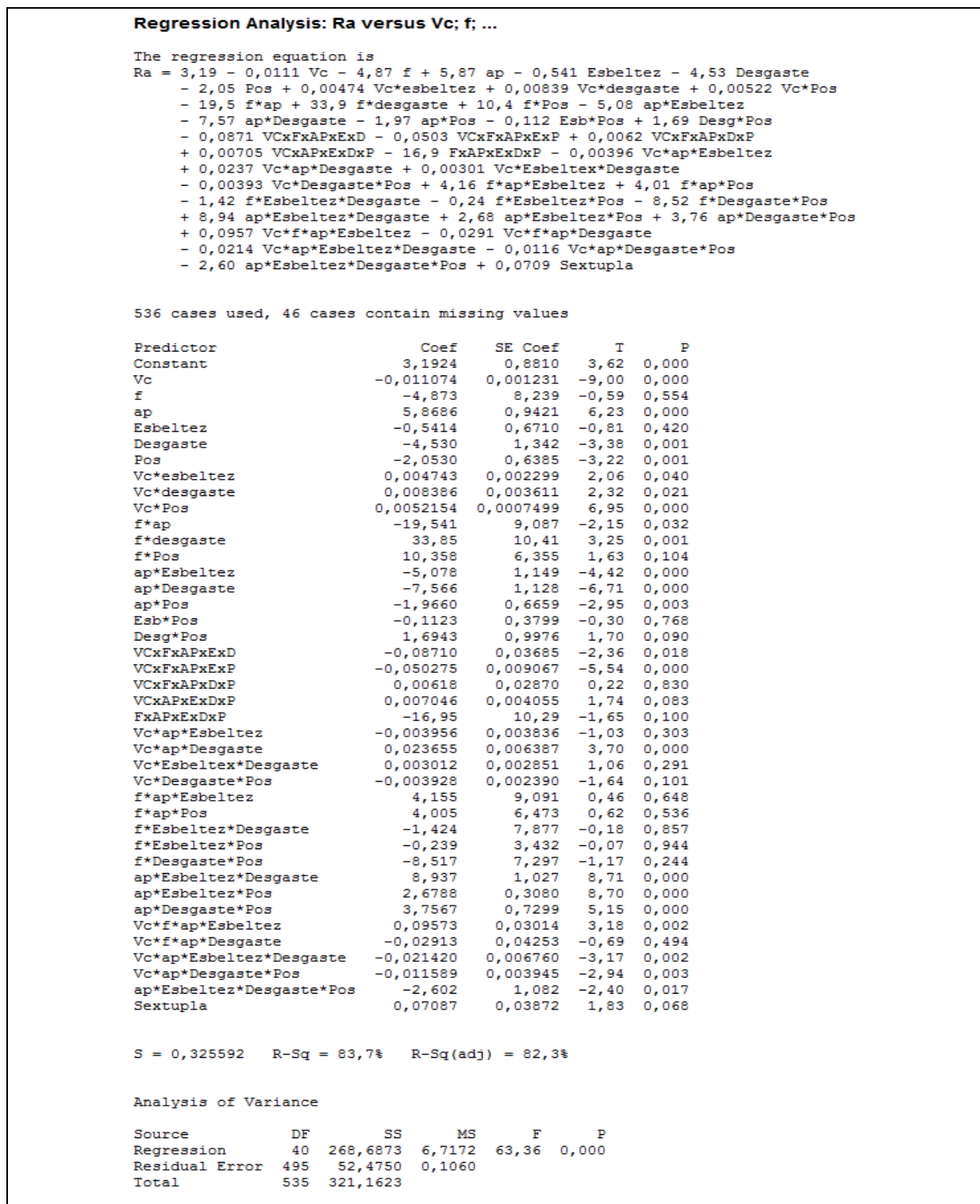


Figura 81 – Análise de regressão do conjunto de dados resultante dos experimentos de usinagem

A comparação entre a rede e o modelo de regressão foi feita por meio de um teste de hipótese entre os valores do erro de predição da melhor configuração de rede para o conjunto de teste (em MAE-DI%) e o valor do erro de predição obtido pelo modelo de regressão. Utilizou-se o teste-t bilateral para hipótese nula de médias de erro iguais ao nível de significância de 0,05. O valor do erro de predição calculado para o modelo de regressão (utilizado como valor de comparação para o teste de hipótese) é igual a 33,721%. O resultado do teste, extraído do Minitab®, é mostrado na Figura 82.

One-Sample T: Método proposto						
Test of mu = 33,7215 vs not = 33,7215						
Variable	N	Mean	StDev	SE Mean	95% CI	T
Método proposto	10	25,7325	2,1205	0,6706	(24,2156; 27,2495)	-11,91
Variable	P					
Método proposto	0,000					

Figura 82 – Teste de hipótese para médias iguais de erro de predição em MAE-DI% da melhor rede MLP obtida pelo método e modelo de regressão obtido dos resultados dos experimentos de usinagem

Como se observa na figura, o valor p resultante do teste é igual a 0,000, o que leva à rejeição da hipótese nula de médias iguais. Este resultado evidencia, ao nível de significância adotado, superioridade do desempenho da rede obtida por meio do método proposto. O resultado também é ilustrado na Figura 83, na qual se mostra o diagrama de caixa para o teste de hipótese. Na figura, \bar{X} representa a média da variável de saída (medida em MAE-DI%) para a rede otimizada e H_0 representa o resultado obtido pelo modelo de regressão.

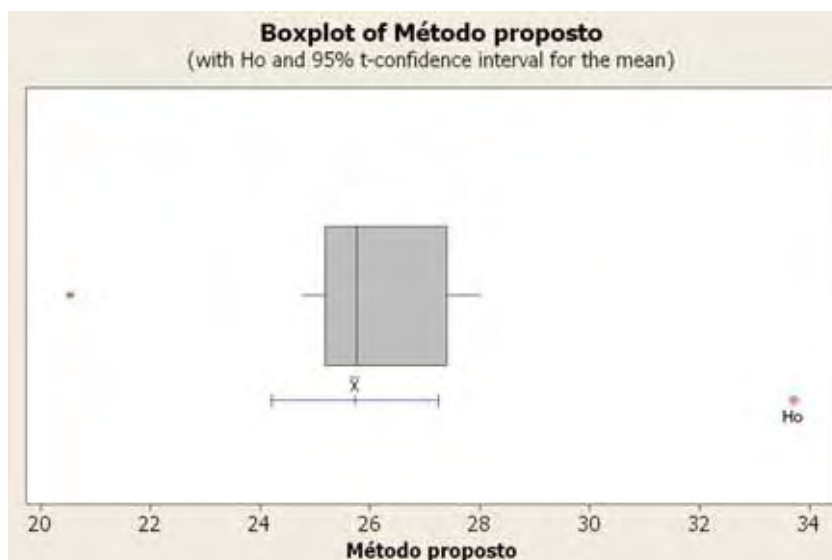


Figura 83 – Diagrama de caixa do teste de hipótese entre erro de predição da melhor rede MLP obtida pelo método e erro do modelo de regressão para resultados dos experimentos de usinagem

5.6.4 COMPARAÇÃO COM RESULTADOS DA FERRAMENTA AUTOMÁTICA DE OTIMIZAÇÃO DE REDES

Uma outra comparação foi realizada entre a melhor configuração de rede MLP obtida por meio da aplicação do método e uma arquitetura de rede otimizada pelo pacote computacional. O valor da variável de saída (em MAE-DI%) medido para a rede proposta pela ferramenta, foi igual a 30,994%. Utilizou-se o teste de análise de variância entre para comparar as predições de erro das duas configurações de rede. O resultado é mostrado na Figura 84, na qual se mostra a tabela da ANOVA.

O resultado da ANOVA indica que, ao nível de significância adotado (igual a 0,05), existe forte evidência de diferença entre os erros de predição das configurações de redes sob comparação. O erro de predição da rede obtida por meio do método é inferior ao obtido por meio do pacote computacional. O método proposto foi capaz de identificar uma configuração de rede com desempenho superior ao da ferramenta comercial utilizada.

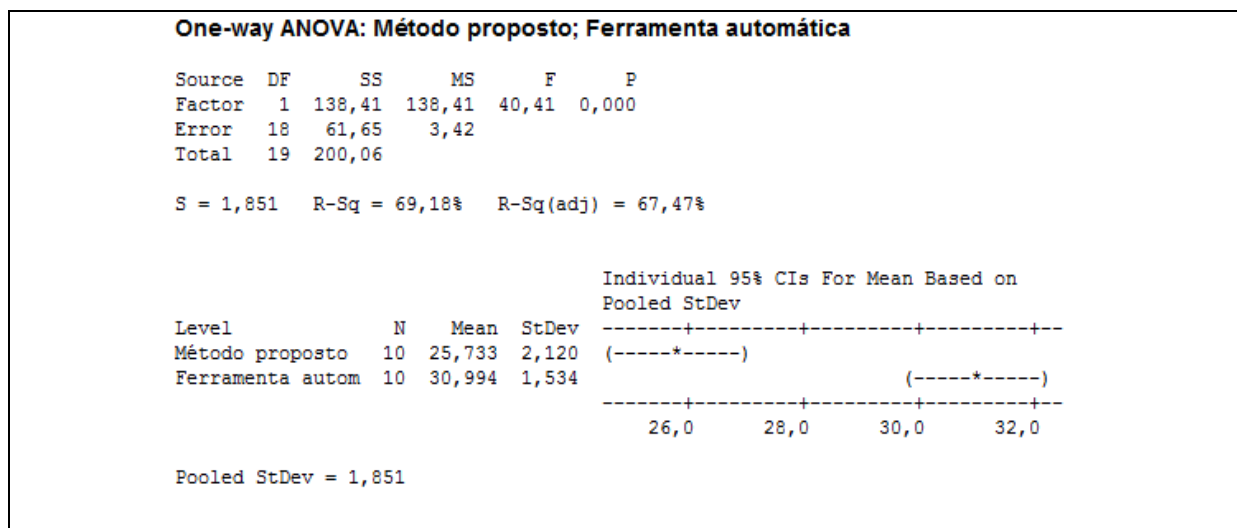


Figura 84 – Comparação entre melhor rede MLP obtida pelo método e melhor rede apontada pelo pacote computacional para resultados dos experimentos de usinagem

A comparação é adicionalmente ilustrada na Figura 85, na qual se mostra o diagrama de caixa das observações do erro de predição para as duas configurações de rede, para o conjunto dos casos de teste.

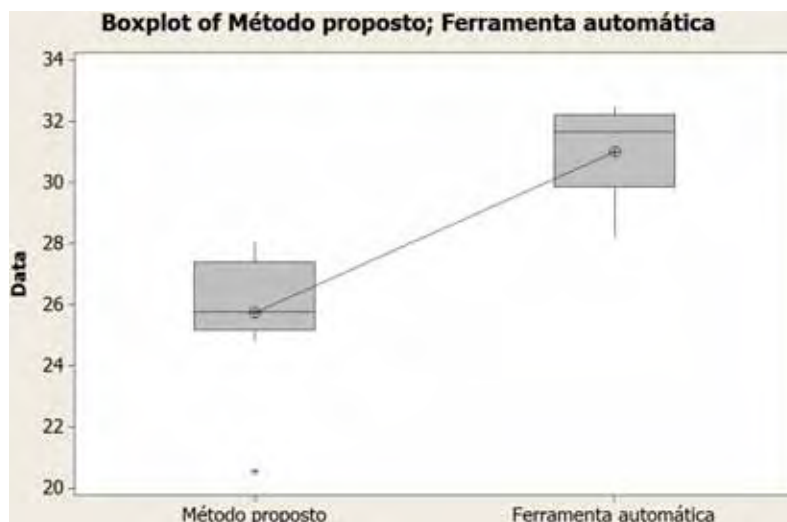


Figura 85 – Diagrama de caixa da comparação entre erro de predição da melhor rede MLP obtida pelo método e melhor rede apontada pelo pacote computacional para resultados dos experimentos de usinagem

5.6.5 REPRODUÇÃO DO ESTUDO COM REDES DE ARQUITETURA RBF

A aplicação do método à otimização de redes RBF foi realizada de modo similar ao empregado para a otimização de redes MLP. O *Número Máximo de Fases* utilizado foi igual a 5. Os valores de *Ponto Central do Arranjo* e *Intervalos Entre Níveis dos Fatores* foram inicializados conforme especificado nos algoritmos do capítulo 4. A otimização foi realizada em 14 ciclos experimentais. A primeira fase encerrou-se após o ciclo 6, quando a análise apontou como ponto central do arranjo para o ciclo seguinte o mesmo conjunto de níveis dos fatores utilizado como ponto central naquele ciclo. A segunda e a terceira fases se encerraram após os ciclos 8 e 10, respectivamente, pela mesma razão. A quarta e quinta fases foram encerradas após os 12 e 14, respectivamente, quando a análise dos resultados deixou de apontar redução do erro.

Na Figura 86 é mostrada a evolução de desempenho obtida ao longo do processo de otimização. Para cada ciclo (no eixo das abscissas) é indicado o desempenho da melhor configuração de rede obtida (no eixo das ordenadas), expresso em MAE-DI%.

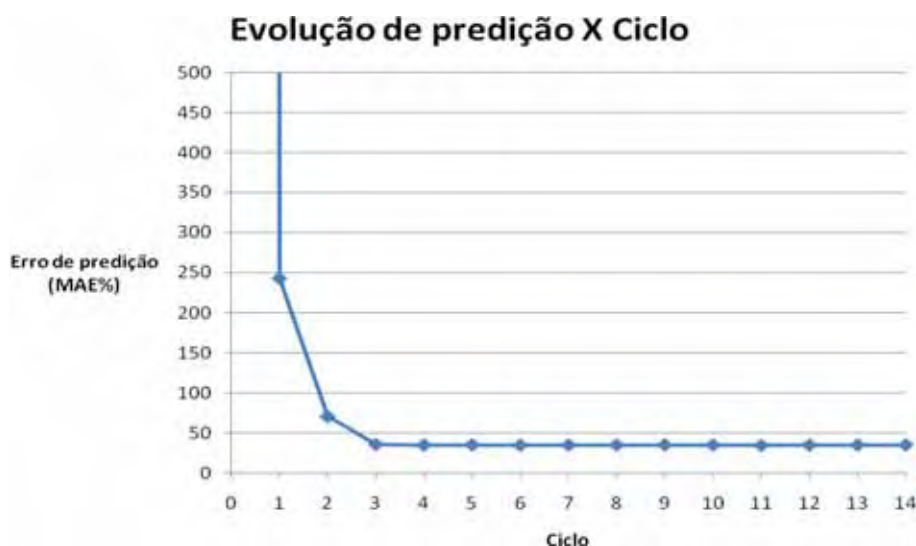


Figura 86 – Evolução de desempenho das melhores redes RBF a cada ciclo – aplicação do método aos resultados dos experimentos de usinagem

Ao final da aplicação do método a melhor configuração de rede obtida foi a especificada pela Tabela 29. O desempenho do melhor modelo de rede RBF é sumarizado na Tabela 30.

Tabela 29 – Configuração de rede neural RBF otimizada para predição da rugosidade com uso dos resultados dos experimentos de usinagem

Fator de projeto	Nível ótimo do fator
Fator de escala do algoritmo Isotrópico para determinação do parâmetro de largura da função radial	132
Número de unidades radiais	78
Algoritmo para determinação do hiper-centro das unidades radiais	KM

Tabela 30 – Desempenho de rede neural RBF otimizada para predição da rugosidade com uso dos resultados dos experimentos de usinagem

Medida de Desempenho	Erro de predição (MAE-DI%)
Valor médio	34,257
Desvio-padrão	203591,644
Valor mínimo	33,148

Quando se compara os resultados da melhor rede RBF àqueles da melhor rede MLP por meio de análise de variância, ao nível de significância de 0,05, nota-se que o modelo RBF possui erro de predição superior. Isto pode ser observado pelo valor p resultante da ANOVA, igual a 0,000, mostrado na Figura 87. Seu desempenho é, portanto, inferior ao da MLP, quando aplicada ao mesmo conjunto de testes.

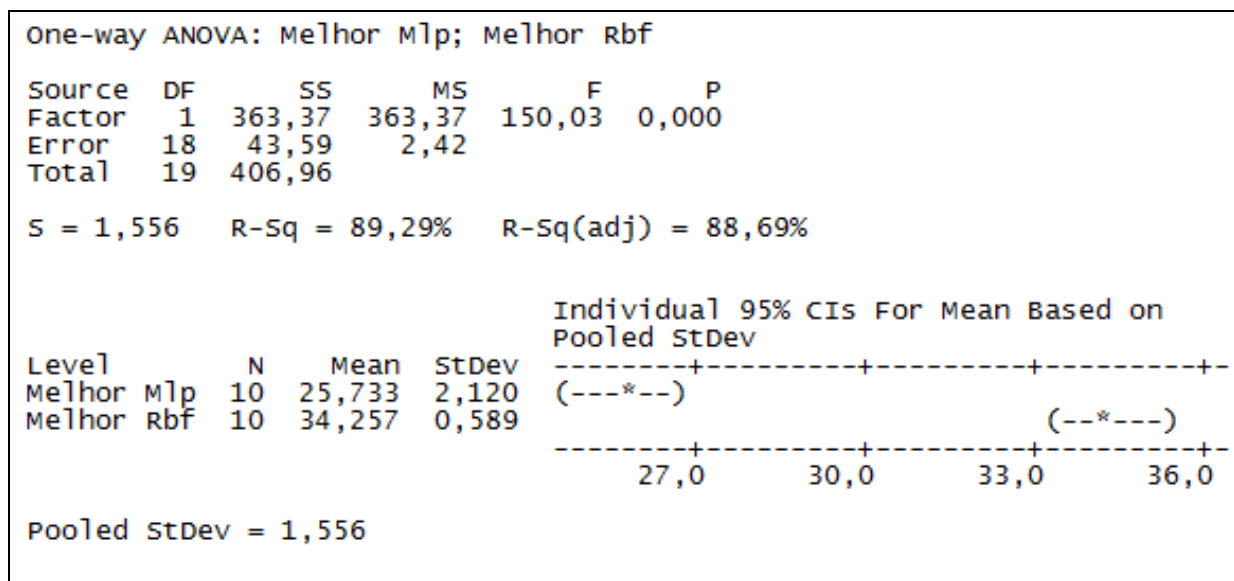


Figura 87 – Comparação entre rede otimizada RBF e rede otimizada MLP para dados resultantes dos experimentos de usinagem

O mesmo teste empregado para comparar o desempenho da melhor rede MLP ao desempenho do modelo de regressão foi repetido para a rede RBF. A equação de regressão utilizada foi a mesma. Neste caso, o erro de predição do modelo de regressão é inferior ao valor correspondente da rede RBF. O teste resultou em um valor p igual a 0,018, evidenciando, neste caso, a superioridade do modelo de regressão.

Uma outra comparação foi feita entre o desempenho da rede RBF obtida pelo método e a rede obtida por meio da aplicação do pacote computacional para otimização de redes, por meio da análise de variância ao nível de significância de 0,05. A rede proposta pela ferramenta apresentou um valor da variável de saída, medido em MAE-DI%, igual a 42,924%. O valor p resultante da ANOVA foi igual a 0,000, o que fornece evidência de diferença estatística significativa em favor da configuração de rede obtida pelo método proposto.

5.7 CONSIDERAÇÕES SOBRE OS RESULTADOS OBTIDOS

Na Tabela 31 é mostrado um sumário dos resultados de predição dos modelos de RNA obtidos a partir da aplicação do método proposto. São indicados os valores obtidos por modelos de regressão, os valores obtidos da literatura (quando aplicável), e os valores da estatística de erros da arquitetura ótima de rede apontada pelo método de

otimização. Os valores mostrados correspondem à medida final do erro de predição, medidos em MAE-DI%. Na coluna relativa aos resultados obtidos pelas redes MLP são mostrados ainda, entre parênteses, o melhor desempenho obtido pela arquitetura otimizada.

Tabela 31 – Comparativo dos erros de predição entre os modelos de RNAs otimizados e erros obtidos por outros modelos

	Literatura	Regressão	Ferramenta computacional		Método proposto	
			RBF	MLP	RBF	MLP
Curva sintética 30 casos	0,241512	--	--	--	0,000215	--
Curva sintética 15 casos	0,980715	--	--	--	0,000508	--
Artigo 1	72,108	24,873	14,170	38,527	17,367	12,797 (10,470)
Artigo 2	15,028	42,167	25,286	25,090	17,380	16,908 (14,418)
Artigo 3	4,899	11,173	3,189	5,920	3,855	3,321 (1,399)
Experimentos de torneamento	--	33,721	30,994	42,924	34,257	25,760 (20,536)

Os resultados obtidos mostram que as redes projetadas segundo o método proposto apresentam desempenho bastante superior ao apresentado por modelos de regressão. Permite concluir também que as configurações de rede MLP obtidas por meio da aplicação do método obtiveram níveis de erro de predição estatisticamente inferiores aos relatados pela literatura em duas das três comparações efetuadas. Mesmo no caso no qual a estatística do erro de predição da melhor rede foi superior ao

resultado relatado pelo trabalho da literatura (na aplicação do método aos resultados do Artigo 2), a rede projetada segundo o método apresentou nível de erro de predição inferior ao relatado na literatura em diversos testes. Como a ferramenta utilizada para a criação de redes permite armazenar e recuperar instâncias específicas de configurações sob teste, pode-se concluir que foi possível identificar uma configuração de rede com resultado superior ao relatado pela literatura em todas as comparações efetuadas. Os resultados mostram também que as redes geradas através da aplicação do método obtiveram estatísticas de erro inferiores e, portanto, desempenho superior ao das redes propostas por uma ferramenta computacional comercial para otimização de configurações de rede (Statistica IPS®).

Uma observação adicional se faz necessária sobre a diferença na ordem de grandeza dos erros de predição observados. Em relação aos experimentos de torneamento do aço ABNT 12L14, por exemplo, mesmo tendo obtido resultados estatisticamente superiores aos dos métodos concorrentes, o modelo de RNA otimizado pelo método proposto apresentou medida de erro da ordem de 25%. Tal dimensão pode ser atribuída ao material utilizado e à severidade das condições experimentais utilizadas durante as operações de usinagem. O aço ABNT 12L14, devido à sua usinabilidade, caracteriza-se pela extrema não-linearidade nos processos que conduzem à formação da rugosidade. Adicionalmente, a dificuldade para a confecção de um modelo ao mesmo tempo genérico e exato é amplificada pela grande variedade de condições operacionais e de ruído incluídas nos experimentos. Desta forma, tendo o método de otimização proposto apresentado resultados superiores (ou no mínimo iguais) aos métodos concorrentes, os resultados sugerem que as diferenças de ordem de grandeza observadas devem residir em características específicas das operações de corte que deram origem aos conjuntos de treinamento utilizados.

5.8 COMENTÁRIOS SOBRE O CUSTO COMPUTACIONAL DA APLICAÇÃO DO MÉTODO PROPOSTO

Para a elaboração do método foi criado um sistema de *scripts* e arquivos de lote (*'batch files'*, em inglês) do Statistica®, que permitiram a repetição independente dos

ensaios com cada configuração de rede de maneira automatizada. Os scripts foram usados também para a execução de arranjos experimentais completos, aumentando em muito a produtividade do trabalho e permitindo dessa forma a cobertura em tempo hábil do vasto domínio das variáveis independentes investigadas durante o projeto.

A programação de um experimento era feita com a montagem dos arquivos contendo casos de treinamento, seleção e testes próprios do artigo ou experimento de usinagem para o qual se deseja encontrar arquiteturas de rede ótimas. Os passos seguintes correspondiam à montagem dos *scripts* para o tipo de arranjo a ser executado e a configuração dos *batch files* que reduziam a configuração das macros a cada ciclo à simples digitação dos valores não codificados dos parâmetros experimentais. Uma vez programado o ambiente, o *script* do ciclo era executado. Ao final da execução, os resultados eram transferidos para um conjunto de planilhas elaborado de modo a fornecer os dados já expressos nas unidades corretas para uso do Minitab®, que fazia a análise do arranjo experimental.

A tarefa de programação toma de 4 a 6 horas de trabalho, em média, e deve ser feito uma vez, no início de cada processo de otimização. O trabalho foi conduzido em uma máquina com processador Intel® Core 2 Quad, com 1GB de RAM. A execução do *script* de cada arranjo experimental leva aproximadamente 10 minutos, e o tempo requerido para execução aumenta com o aumento do número de casos de treinamento. A tarefa de processamento dos resultados de um arranjo, realizada após a execução do *script* do Statistica®, requer aproximadamente meia hora de trabalho. Desta forma, o custo computacional por ciclo é de aproximadamente uma hora. Nos estudos de caso realizados neste trabalho, portanto, o custo computacional para otimização ficou entre 9 e 13 horas.

O custo computacional para o desenvolvimento do método foi proporcional ao número de ciclos experimentais efetuado durante o projeto. O trabalho começou pelo desenvolvimento do método para RBF usando curvas sintéticas. A atividade foi reorientada quando as RBFs se revelaram inapropriadas para a predição de resultados experimentais colhidos na literatura. Foi realizada então uma série de ciclos experimentais exploratórios com uso de um arranjo ortogonal de Taguchi, com objetivo de detectar os fatores relevantes para otimização de redes MLP. Na sequência

foram conduzidas séries de experimentos exploratórios com objetivo de selecionar o melhor arranjo experimental e a variável de saída para otimização de redes MLP. Após essas definições, foram realizados os ciclos experimentais efetivos para a geração dos resultados. Nesta etapa, o método, já maduro, foi aplicado para otimização de MLPs e RBFs para as curvas sintéticas, resultados da literatura e resultados dos experimentos de torneamento com aço ABNT 12L14, conforme já descrito. Ao todo, 356 arranjos experimentais foram executados até a conclusão do trabalho. Considerando-se o tempo requerido para geração e configuração de scripts, pode-se estimar que o custo computacional requerido para a finalização do trabalho ficou em torno de 400 horas.

6 CONCLUSÕES E SUGESTÕES PARA PESQUISAS FUTURAS

6.1 CONCLUSÕES

A aplicação do método proposto para o projeto de Redes Neurais artificiais para a predição da rugosidade em processos de usinagem levou à identificação de topologias de rede que apresentam os seguintes resultados:

- reduções do erro de predição de 48,5%, 59,9%, 70,3% e 39,1%, na comparação com os modelos de regressão para os Artigos 1, 2 e 3 e para os experimentos de torneamento com aço ABNT 12L14;
- reduções do erro de predição de 82,3%, 4,1% e 71,5%, na comparação com os resultados relatados nos artigos 1, 2 e 3;
- reduções de estatística de erro de predição iguais a 9,7%, 33,1%, 46,3% e 33,7% para os conjuntos de testes formados pelos dados Artigos 1, 2 e 3 e para os dados dos experimentos de torneamento com aço ABNT 12L14, na comparação com as redes projetadas pelo IPS, uma ferramenta computacional de uso comercial desenvolvida especificamente para a otimização de topologias de rede;
- dispersão dos erros de predição significativamente reduzidas na comparação com as redes propostas pelos estudos da literatura e pelo pacote computacional, o que pode ser corroborado por teste para comparação entre a variância das observações;

Os resultados obtidos pelo método foram consistentemente superiores aos obtidos pelo método dos mínimos quadrados, superior aos resultados obtidos da literatura e aos obtidos pelo IPS. Foram atingidos, desta forma, os objetivos específicos estabelecidos, no que tange ao desempenho do método proposto.

Outro objetivo específico do trabalho corresponde à comparação de desempenho entre arquiteturas de rede. O confronto entre o desempenho de predição das melhores configurações de rede MLP e RBF para os dados obtidos dos estudos da literatura e

para os resultados dos experimentos de usinagem indicam que, em dois casos (dados do Artigo 1 e dados dos experimentos com aço ABNT 12L14) as redes MLP atingiram resultado estatisticamente superior aos das melhores redes RBF. Nos dois casos restantes (dados dos Artigos 2 e 3) a análise não evidenciou diferença significativa de erro de predição entre as duas arquiteturas, ainda que o erro de predição da MLP tenha sido menor. Deve-se mencionar porém que, em função do pouco tempo exigido para o seu treinamento, a aplicação do método a redes RBF pode ser realizada em um intervalo de tempo bastante inferior ao requerido para redes MLP.

Ressalta-se que o trabalho realizado teve natureza investigativa. A suposição inicial era de que redes projetadas segundo o método proposto poderiam apresentar desempenho superior na predição da rugosidade. De posse dos resultados obtidos nos experimentos realizados, pode-se afirmar que a suposição inicial mostrou-se verdadeira.

Deve-se salientar que as conclusões obtidas referentes à exatidão do método proposto não devem ser extrapoladas para arquiteturas de redes neurais não abordadas neste trabalho. Por outro lado, os resultados alcançados e a generalidade do método justificam sua aplicação a outras arquiteturas de rede e também a dados obtidos de outros processos de usinagem, com distintos materiais e ferramentas.

Conclui-se que o método proposto para a otimização de arquitetura de redes neurais leva à redução significativa do erro de predição da rugosidade em operações de usinagem quando comparado às técnicas atualmente utilizadas. Constitui também opção eficaz para o projeto sistemático modelos baseados em redes neurais para predição da rugosidade, suprimindo desta forma a lacuna apontada na literatura sobre o tema.

6.2 SUGESTÕES PARA PESQUISAS FUTURAS

Os questionamentos e alternativas surgidos durante o projeto oferecem diversas possibilidades para a continuação deste trabalho:

- Investigação da hipótese de que uma versão aumentada do algoritmo evolutivo ora proposto, abrangendo um número maior de fatores, possa

levar a reduções ainda maiores no erro de predição da rugosidade; a lista de fatores adicionais para otimização pode incluir a natureza dos algoritmos de treinamento das fases 1 e 2, a poda de entradas e de neurônios, a regularização de decaimento de pesos nas fases 1 e 2 de treinamento, o uso de outras funções de transferência e pré-processamento de dados, dentre outros;

- O aumento do número de fatores leva a um aumento exponencial do número de corridas experimentais por arranjo, e conseqüentemente do tempo necessário para execução do processo de otimização. Uma sugestão para resolver este problema seria o desenvolvimento de um pacote computacional a partir dos algoritmos propostos, que permitisse reduzir a carga de trabalho do operador à inicialização dos valores do algoritmo, e que efetue de maneira automatizada todo o processo de treinamento, teste e análise dos resultados;
- Investigação de estratégias experimentais utilizadas para formação de conjuntos de dados de treinamento. A grande maioria dos trabalhos analisados forma os conjuntos segundo arranjos baseados em projeto de experimentos. Outras estratégias, tais como distribuição uniforme de casos entre os limites das variáveis de entrada poderiam ser comparadas aos arranjos tipicamente utilizados, e os resultados de predição após a aplicação do método de otimização aqui proposto poderiam ser comparados em busca da melhor estratégia para coleta de dados;
- Adaptação do método de otimização aqui proposto para outras arquiteturas de redes, tais como mapas auto-organizáveis de memória, por exemplo. O objetivo, neste caso, seria comparar o desempenho de outros tipos de rede ao obtido pelas redes obtidas ao longo deste trabalho.
- Aplicação do método a outros processos de usinagem, como o fresamento e a furação. Em função de sua generalidade, a aplicabilidade do método a outros processos de manufatura que requerem a modelagem de relações não-lineares entre variáveis independentes e dependentes, tais como a soldagem, também pode ser explorada. Para tanto, basta alterar o conjunto

de dados utilizado para treinamento e testes das redes;

- Exploração do método Garson de partição para quantificar a influência das variáveis independentes associadas a cada neurônio de entrada no resultado do processo de usinagem. Nenhuma aplicação do dito método foi encontrada para processos de manufatura;
- Investigação mais aprofundada da especificação de modelos de redes neurais por meio de distribuições estatísticas de valores extremos. A expressão das propriedades de topologias de redes em termos de distribuições estatísticas é uma área ainda inexplorada da literatura.

REFERÊNCIAS BIBLIOGRÁFICAS

AGOSTINHO, O. L.; RODRIGUES, A. C.; LIRANI, J. **Tolerâncias, Ajustes, Desvios e Análise de Dimensões**. Editora Edgar Blucher, São Paulo, 1990.

AL-AHMARI, A. M. A. Predictive machinability models for a selected hard material in turning operations, **Journal of Materials Processing Technology**, v.190, p.305–311, 2007.

ALI-TAVOLI, M.; NARIMAN-ZADEH, N.; KHAKHALI, A.; MEHRAN, M. Multi-Objective Optimization Of Abrasive Flow Machining Processes Using Polynomial Neural Networks And Genetic Algorithms, **Machining Science and Technology**, v.10, p.491-510, 2006.

AMORIM, H. J. **Estudo da relação entre velocidade de corte, desgaste de ferramenta, rugosidade e forças de usinagem em torneamento com ferramenta de metal duro**. 2002. 114f. Dissertação de mestrado – UFRGS – Universidade Federal do Rio Grande do Sul, 2002.

AMBROGIO, G.; FILICE, L.; SHIVPURI, R.; UMBRELLO, D. Application of NN technique for predicting the in-depth residual stresses during hard machining of AISI 52100 steel. **International Journal of Material Forming**, v.1, p.39–45, 2008.

ANTONY, J. Multiple response optimization using Taguchi methodology and neuro-fuzzy based model, **Journal of Manufacturing Technology Management**, v.17, n.7, p.908-925, 2005.

APPOLINÁRIO, F. **Metodologia da ciência – filosofia e prática da pesquisa**. Editora Pioneira Thomson Learning, São Paulo, 2006.

ARANTES, L. J. **Desenvolvimento e Avaliação do Processo Híbrido de Usinagem por Descargas Elétricas e Jato de Água**. 2007. 125f. Tese de doutorado – Universidade Federal de Uberlândia, Programa de Pós-Graduação em Engenharia Mecânica, 2007.

ASOKAN, P.; RAVI KUMAR R.; JEYAPPAUL, R.; SANTHI, M. Development of multi-objective optimization models for electrochemical machining process, **International Journal of Advanced Manufacturing Technology**, v.39, p.55–63, 2008.

ASSARZADEH, S.; GHOREISHI, M. Neural-network-based modeling and optimization of the electro-discharge machining process. **International Journal of Advanced Manufacturing Technology**, v.39, p.488–500, 2008.

BAGCI, E.; IŞIK, B. Investigation of surface roughness in turning unidirectional GFRP composites by using RS methodology and ANN. **International Journal of Advanced Manufacturing Technology**, v.31, p.10–17, 2006.

BALESTRASSI, P. P.; POPOVA, E.; PAIVA, A. P.; MARANGON LIMA, J. W. Design of experiments on neural network's training for nonlinear time series forecasting. **Neurocomputing**, v.72, n.4–6, p.1160–1178, 2009.

BALIC, J. Neural-Network-Based Numerical Control for Milling Machine, **Journal of Intelligent and Robotic Systems**, v.40, p.343–358, 2004.

BASHEER, A. C.; DABADE, U.A.; SUHAS, S.J.; BHANUPRASAD, V. V. Modeling of surface roughness in precision machining of metal matrix composites using ANN. **Journal of Materials Processing Technology**, v.197, p.439–444, 2008.

BATTITI, R. First and Second-Order Methods for Learning: Between Steepest Descent and Newton's Method, **Neural Computation**, v.4, n.2, p.141-166, 1992.

BAZARAA, M. S.; JARVIS, J. J.; SHERALI, H.D. **Linear Programming and Network Flows**, Wiley, Hoboken, 2004.

BENARDOS, P. G.; VOSNIAKOS, G. C. Predicting surface roughness in machining: a review. **International Journal of Machine Tools and Manufacture**, v.43, p.833–844, 2003.

BENARDOS, P. G.; VOSNIAKOS, G. C. Prediction of surface roughness in CNC face milling using neural networks and Taguchi's design of experiments. **Robotics and Computer Integrated Manufacturing**, v.18, p.343-354, 2002.

BERTRAND, J. W. M.; FRANSOO, J. C. Modelling and simulation: operations management research methodologies using quantitative modeling. **International Journal of Operations & Production Management**, v. 22, n.2, p. 241-264, 2002.

BISHOP, C. M. (2007) **Pattern recognition and machine learning**, 1st edn. Springer Business-Media, New York, 2007.

BONIFÁCIO, M. E. R.; DINIZ, A. E. Monitoring the tool life in finish turning using vibration signals. **Revista Brasileira de Ciências Mecânicas**, v.16, n.1, p.58-71, 1994.

BOX, G. E. P. Evolutionary Operation: a method for increasing industrial productivity. **Applied Statistics**, v.6, p.81–101, 1957.

BOX, G. E. P.; DRAPER, N. R. **Evolutionary Operation – A Statistical Method for Process Improvement**, Wiley Interscience, New York, 1998.

BOX, G. E. P.; HUNTER, W. G.; HUNTER, J. S. **Statistics for Experimenters**. John Wiley & Sons, New York, 1978.

BRAGA, A. P.; CARVALHO, A. P. L. F.; LUDERMIR, T. B. **Redes Neurais Artificiais: Teoria e Aplicações**. LTC Livros Técnicos Editora S. A., Rio de Janeiro, 2000.

BROOMHEAD, D. S.; LOWE, D. Multivariable functional interpolation and adaptive networks. **Complex Systems**, v.2, p.321-355, 1988.

BRUNI, C.; D'APOLITO, L.; FORCELESE, A.; GABRIELLI, F.; SIMONCINI, M. Surface roughness modelling in finish face milling under MQL and dry cutting conditions, **International Journal of Material Forming**, v.1, n.1, p.503-506, 2008.

BRYMAN, A. **Research methods and organization studies (contemporary social research)**, 1^a ed. Routledge, London, 1989.

CARPENTER, G. A.; GROSSBERG, S. A massively parallel architecture for a self-organizing neural pattern recognition machine, **Computer Vision, Graphics and Image Processing**, v.37, p.54-115, 1987.

CARPENTER, G. A.; GROSSBERG, S. ART2: Self-organization of stable category recognition codes for analog input patterns, **Applied Optics**, v.26, n.23, p.4919-4930, 1987.

CERVELLERA, C.; WEN, A.; CHEN, V. C. P. Neural network and regression spline value function approximations for stochastic dynamic programming. **Computers & Operations Research**, v.34, p.70-90, 2007.

CHANG, C. K.; LU, H. S. Study on the prediction model of surface roughness for side milling operations. **International Journal of Advanced Manufacturing Technology**, v.29, p. 867-878, 2006.

CHAO, P. Y.; HWANG, Y. D. An improved neural network model for the prediction of cutting tool life, **Journal of Intelligent Manufacturing**, v.8, p.107-117, 1997.

CHEN, W.C.; HSU, Y.Y.; HSIEH, L.F.; TAI, P.H. A systematic optimization approach for assembly sequence planning using Taguchi method, DOE, and BPNN, **Expert Systems with Applications**, v.37, p.716-726, 2010.

CHIEN, W. T.; CHOU, C. Y. The predictive model for machinability of 304 stainless steel, **Journal of Materials Processing Technology**, v.118, p.442-447, 2001.

CHOUDHURY, S. K.; BARTARYA, G. Role of temperature and surface finish in predicting tool wear using neural network and design of experiments, **International Journal of Machine Tools & Manufacture**, v.43, p.747-753, 2003.

CHOUDHURY, J.P.; SARKAR, B.; MUKHERJEE, S.K. Forecasting of engineering manpower through fuzzy associative memory neural network with ARIMA: a comparative study, **Neurocomputing**, v.47, p.241-257, 2002.

CORREA, M.; BIELZA, C.; PAMIES-TEIXEIRA, J. Comparison of Bayesian networks and artificial neural networks for quality detection in a machining process. **Expert Systems Applications**, v.36, n.3, p.7270–7279, 2009.

CRESWELL, J. W.; PLANO CLARK, V. L. **Designing and conducting mixed methods research**. Sage Publications, California 2007.

CUS, F.; ZUPERL, U. Approach to optimization of cutting conditions by using artificial neural networks. **Journal of Materials Processing Technology**, v.173, p. 281–290, 2006.

ÇAYDAS, U.; EKICI, S. Support vector machines models for surface roughness prediction in CNC turning of AISI 304 austenitic stainless steel, **Journal of Intelligent Manufacturing**, DOI 10.1007/s10845-010-0415-2.

ÇAYDAS, U.; HASÇALIK, A. A study on surface roughness in abrasive waterjet machining process using artificial neural networks and regression analysis method. **Journal of Materials Processing Technology**, v.202, n.1–3, p.574–582, 2008.

CYBENKO, G. Approximation by superpositions of a sigmoidal function. **Mathematics of Control, Signals and Systems**, v.2, p.304-314, 1989.

DAVIM, J. P.; GAITONDE, V. N.; KARNIK, S. R. Investigations into the effect of cutting conditions on surface roughness in turning of free machining steel by ANN models, **Journal of Materials Processing Technology**, v.205, p.16–23, 2008.

DHOKIA, V. G.; KUMAR, S.; VICHARE, P.; NEWMAN, S. T., ALLEN, R. D. Surface roughness prediction model for CNC machining of polypropylene. **Proceedings of the Institution of Mechanical Engineers Part B: J. Engineering Manufacture**. v.222, p.137–153, 2008.

DAVIM, J. P. A note on the determination of optimal cutting conditions for surface finish obtained in turning using design of experiments. **Journal of Materials Processing Technology**, v.116, p.305-308, 2001.

DIN 4760: **Form deviations, concepts; classification systems**, Deutsches Institut Für Normung, e. V., 1982.

DINIZ, A. E.; MARCONDES, F. C.; COPPINI, N. L. **Tecnologia da Usinagem dos Materiais**. Artliber, São Paulo, 2008.

EL-MOUNAYRI, H.; KISHAWY, H.; BRICENO, J. Optimization of CNC ball end milling: a neural network-based model. **Journal of Materials Processing Technology** 166:50–62, 2005.

EL-SONBATY, I. A.; KHASHABA, U. A.; SELMY, A. I.; ALI, A. I. Prediction of surface roughness profiles for milled surfaces using an artificial neural network and fractal geometry approach, **Journal Of Materials Processing Technology**, v.200, p.271–278, 2008.

ERZURUMLU, T.; OKTEM, H. Comparison of response surface model with neural network in determining the surface quality of moulded parts. **Materials and Design**, v.28, p.459–465, 2007.

EZUGWU, E.O.; FADARE, D. A.; BONNEY, J.; DA SILVA, R. B.; SALES, W. F. Modelling the correlation between cutting and process parameters in high-speed machining of Inconel 718 alloy using an artificial neural network. **International Journal of Machine Tools and Manufacture**, v.45, p.1375–1385, 2005.

EZUGWU, E. O.; ARTHUR, S. J.; HINES, E. L. Tool-wear prediction using artificial neural networks. **Journal of Materials Processing Technology**, v.49, p.255–264, 1995.

FARRER, H.; BECKER, C. G.; FARIA, E. C.; DE MATOS, H. F.; DOS SANTOS, M. A., MAIA, M. L. **Algoritmos Estruturados, 3^a ed.** LTC Editora, Rio de Janeiro, 1999.

FERRARESI, D. **Fundamentos da usinagem dos metais.** Ed. Edgard Blucher, São Paulo, 1977.

FORESEE, F. D.; HAGAN, M. T. Gauss–Newton approximation to Bayesian learning. **Proceedings of the International Joint Conference on Neural Networks**, v.2, p.1930–1935, 1997.

FREDJ, N. B.; AMAMOU, R. Ground surface roughness prediction based upon experimental design and neural network models. **International Journal of Advanced Manufacturing Technology**, v.31, p.24–36, 2006.

GAO, Q.; ZHANG, Q.; SU, S.; ZHANG, J. Parameter optimization model in electrical discharge machining process, **Journal of Zhejiang University SCIENCE A**, v.9, n.1, p.104–108, 2008.

GARSON, G. D. Interpreting neural network connection weights, **Artificial Intelligence Expert**, v.6, n.7, p.47–51, 1991.

GIL, A. C. **Como elaborar projetos de pesquisa, 4. ed.** Atlas, São Paulo, 2009.

GUIMARÃES, O. L. C.; CHAGAS, M. H. R.; FILHO, D. N. V.; SIQUEIRA, A. F.; FILHO, H. J. I.; AQUINO, H. O. Q.; SILVA, M. B. Discoloration process modeling by neural network, **Chemical Engineering Journal**, v.140, p.71–76, 2008.

GUIMARÃES, O. L. C.; SILVA, M. B. Hybrid neural model for decoloration by UV/H₂O₂ involving process variables and structural parameters characteristics to azo dyes, **Chemical Engineering and Processing**, v.46, p.45–51, 2007.

HAGAN, M.; MENHAJ, M. Training feed forward networks with the Marquadt algorithm, **IEEE Transactions on Neural Networks**, v.5, n.6, p.989-993, 1994.

HARIDY, S.; GOUDA, S.A.; WU, Z. An integrated framework of statistical process control and design of experiments for optimizing wire electrochemical turning process, **International Journal of Advanced Manufacturing Technology**, v.53, n.1-4, p.191-207, 2011.

HAYKIN, S. **Neural networks and learning machines**, 3rd edn. Pearson Prentice Hall, New Jersey, 2008.

HAYKIN, S. **Neural Networks - A comprehensive foundation**, 2nd ed., Prentice Hall, New York, 1998.

HEBB, D. O. **The Organization of Behavior**. Wiley, New York, 1949.

HO, K. H.; NEWMAN, S. T. State of the art electrical discharge machining (EDM). **International Journal of Machine Tools & Manufacture**, v.43, p.1287–1300, 2003.

HOPFIELD, J. J. Neural Networks and physical systems with emergent collective properties. **Proceedings of the National Academy of Sciences**, v.79, p.2554-2558, 1982.

HOSSAIN, M. I.; AMIN, A. K. M. N.; PATWARI, A. U. Development of an Artificial Neural Network Algorithm for Predicting the Surface Roughness in End Milling of Inconel 718 Alloy, In: IEEE ICCCE 2008. May 13-15, 2008, Kuala Lumpur, Malaysia. **Proceedings of the International Conference on Computer and Communication Engineering**, Kuala Lumpur: IEEE, 2008, p.1321-1324.

ISO 4287:1997/Cor 2:2005. **Geometrical Product Specifications (GPS) - Surface texture: Profile method - Terms, definitions and surface texture parameters**, International Organization for Standardization, 2005.

IVAKHNENKO, A. G. Polynomial Theory of Complex Systems, **IEEE Transactions on Systems, Man and Cybernetics**, v.1, n.4, p.364–378, 1971.

JESUTHANAM, C. P.; KUMANAN, S.; ASOKAN, P. Surface Roughness Prediction Using Hybrid Neural Networks, **Machining Science and Technology**, v.11, p.271-286, 2007.

KAMINSKI, P. C.; CAPUANO, M. N. **Usinagem de Micro-Furos pelo Processo de Eletroerosão por Penetração**. Revista OESP Metal-Mecânica, Ano 4, n.25, OESP Mídia, São Paulo, p.42-47, 1999.

KARAYEL, D. Prediction and control of surface roughness in CNC lathe using artificial neural network. **Journal of Materials Processing Technology**, v.209, p.3125–3137, 2009.

KARNIK, S. R.; GAITONDE, V. N.; DAVIM, J. P. A comparative study of the ANN and RSM modeling approaches for predicting burr size in drilling. **International Journal of Advanced Manufacturing Technology**, v.38, n.9–10, p.868–883, 2008.

KARPAT, Y.; ÖZEL, T. Multi-objective optimization for turning processes using neural network modeling and dynamic neighborhood particle swarm optimization. **International Journal of Advanced Manufacturing Technology**, v.35, p.234–247, 2007.

KILICKAP, E.; HUSEYINOGLU, M.; YARDIMEDEN, A. Optimization of drilling parameters on surface roughness in drilling of AISI 1045 using response surface methodology and genetic algorithm, **International Journal of Advanced Manufacturing Technology**, v.52, n.1-4, p.79-88, 2011.

KILICKAP, E., Optimization of cutting parameters on delamination based on Taguchi method during drilling of GFRP composite, **Expert Systems with Applications**, v.37, p.6116-6122, 2010.

KHASHEI, M.; BIJARI, M.; ARDALI, G.A.R. Improvement of Auto-Regressive Integrated Moving Average models using Fuzzy logic and Artificial Neural Networks (ANNs), **Neurocomputing**, v.72, p.956-967, 2009.

KOHLI, A.; DIXIT, U. S. A neural-network-based methodology for the prediction of surface roughness in a turning process. **International Journal of Advanced Manufacturing Technology**, v.25, p.118–129, 2005.

KOHONEN, T. **Self Organizing Maps**, 3rd ed., Springer, Berlin, Heidelberg, New York, 2001.

KOHONEN, T. Self-organized formation of topologically correct feature maps, **Biological Cybernetics**, v.43, p.49-59, 1982.

KOVÁCS, Z. L. **Redes Neurais Artificiais - Fundamentos e Aplicações: um texto básico**. Colegium Cognitio/Editora Edição Acadêmica, São Paulo, 1996.

KUMANAN, S.; JESUTHANAM, C. P.; ASHOK KUMAR, R. Application of multiple regression and adaptive neuro fuzzy inference system for the prediction of surface roughness, **International Journal of Advanced Manufacturing Technology**, v.35, p.778–788, 2008.

KUMAR, S.; CHOUDURY, S. K. Prediction of wear and surface roughness in electro-discharge diamond grinding, **Journal of Materials Processing Technology**, v.191, p.206–209, 2007.

LASHERAS, F.S.; VILÁN, J.A.V.; NIETO, P.J.G.; DEL COZ DÍAZ, J.J. The use of design of experiments to improve a neural network model in order to predict the thickness of the chromium layer in a hard chromium plating process, **Mathematical and Computer Modeling**, v.52, p.1169-1176, 2010.

LEE, S. S.; CHEN, J. C. On-line surface roughness recognition system using artificial neural networks system in turning operations, **International Journal of Advanced Manufacturing Technology**, v.22, p.498–509, 2003.

LELA, B.; BAJIC, D.; JOZIC, S. Regression analysis, support vector machines, and Bayesian neural network approaches to modeling surface roughness in face milling, **International Journal of Advanced Manufacturing Technology**, v.42, n.11-12 , p.1082-1088, 2009.

MAHAPATRA, S. S.; PATNAIK, A. Optimization of wire electrical discharge machining (WEDM) process parameters using Taguchi method. **International Journal of Advanced Manufacturing Technology**, v.34, p.911–925, 2007.

MAKEYEV, O.; SAZONOV, E.; BAIDYK, T.; MARTÍN, A. Limited receptive area neural classifier for texture recognition of mechanically treated metal surfaces, **Neurocomputing**, v.71, p. 1418-1421, 2008.

MALAKOOTI, B.; RAMAN, V. An interactive multi-objective artificial neural network approach for machine setup optimization, **Journal of Intelligent Manufacturing**, v.11, p.41-50, 2000.

MANDAL, D.; PAL, S. K.; SAHA, P. Back propagation neural network based modeling of multi-responses of an electrical discharge machining process. **International Journal of Knowledge-based Intelligent Engineering Systems**, v.11, p.381–390, 2007.

MANUAL STATISTICA (Statistical Software) – Versão 7.1 – Statsoft Inc., 2005.

MARKOPOULOS, A. P.; MANOLAKOS, D. E.; VAXEVANIDIS, N. M. Artificial neural network models for the prediction of surface roughness in electrical discharge machining. **Journal of Intelligent Manufacturing**, v.19, p.283–292, 2008

McCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **Bulletin of Mathematical Biophysics**, v.5, p.115-133, 1943.

MOHANA RAO, G. K.; RANGAJANARDHAA G.; HANUMANTHA RAO, D.; SREENIVASA RAO, M. Development of hybrid model and optimization of surface roughness in electric discharge machining using artificial neural networks and genetic algorithm. **Journal of Materials Processing Technology**, v.209, p.1512–1520, 2009.

MOLLAH, A. A.; PRATIHAR, D. K. Modeling of TIG welding and abrasive flow machining processes using radial basis function networks, **International Journal of Advanced Manufacturing Technology**, v.37, p.937–952, 2008.

MONTGOMERY, D. C.; JENNINGS, C. L.; KULAHCI, M. **Introduction to time series analysis and forecasting**. Wiley, Hoboken, 2008.

MONTGOMERY, D. C.; RUNGER, G. C. **Applied statistics and probability for engineers**, 4^a ed. John Wiley & Sons, Hoboken, 2007.

MONTGOMERY, D. C. **Design and Analysis of Experiments**, 5ed., John Wiley & Sons, Hoboken, 2001.

MINSKY, M. L.; PAPERT, S. A. **Perceptrons: an introduction to computational geometry**. MIT Press, Cambridge, MA, 1969.

MYERS, R. H.; MONTGOMERY, D.C. **Response Surface Methodology: process and Product optimization Using Designed Experiments**, Wiley, New York, 1995.

NGUYEN, D.; WIDROW, B. Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights, **Proceedings of the International Joint Conference on Neural Networks**, v.3, p.21–26, 1990.

ÖKTEM, H. An integrated study of surface roughness for modelling and optimization of cutting parameters during end milling operation. **International Journal of Advanced Manufacturing Technology**, p.43, n.9–10, p.852–861, 2009.

OKTEM, H.; ERZURUMLU, T.; ERZINCANLI, F. Prediction of minimum surface roughness in end milling mold parts using neural network and genetic algorithm, **Materials and Design**, v.27, p.735–744, 2006.

OLIVEIRA, A. J. **Otimização das condições de usinagem visando o torneamento a seco do aço ABNT 1045 em operação de desbaste**. 2003. 115f. Dissertação de Mestrado – Unicamp – Universidade Estadual de Campinas, 2003.

ORABY, S. E.; AL-ASKARI, A.; AL-MESHAIEI, E. A. Quantitative and qualitative evaluation of surface roughness-tool wear correlation in turning operations, **Kuwait Journal of Science and Engineering**, v.31, n.1, p.219-244, 2004.

ORTIZ-RODRÍGUEZ, J. M.; MARTÍNEZ-BLANCO, M. R.; VEGA-CARRILLO H. R. Robust design of artificial neural networks applying the Taguchi methodology and DOE. In: IEEE CERMA 2006. September 26–29, 2006, Cuernavaca, Morelos, Mexico. **Proceedings of the Electronics, Robotics and Automotive Mechanics Conference**, Morelos: IEEE, 2006, v.2, p.131–136.

ÖZEL, T.; KARPAT, Y. Predictive modeling of surface roughness and tool wear in hard turning using regression and neural networks. **International Journal of Machine Tools and Manufacture**, v.45, p.467–479, 2005.

ÖZEL, T.; KARPAT, Y.; FIGUEIRA, L.; DAVIM, J. P. Modelling of surface finish and tool flank wear in turning of AISI D2 steel with ceramic wiper inserts, **Journal of Materials Processing Technology**, v.189, p.192–198, 2007.

PAIVA, A. P.; PAIVA, E. J.; FERREIRA, J. R.; BALESTRASSI, P. P.; COSTA, S. C. A multivariate mean square error optimization of AISI 52100 hardened steel turning. **International Journal of Advanced Manufacturing Technology**, v.43, n.7–8, p.631–643, 2009.

PAIVA, A. P. **Metodologia de Superfície de Resposta Multivariada – Uma proposta de otimização para processos de manufatura com múltiplas respostas correlacionadas**. 2006. 279f. Tese de Doutorado – UNIFEI – Universidade Federal de Itajubá, Programa de Pós-Graduação em Engenharia Mecânica, 2006.

PAIVA, A. P.; PEREIRA, J. C. C.; BALESTRASSI, P. P.; FERREIRA, J. R. . Application of design of experiment methodology on turning parameters of the SAE 52100 hardened steel. In: 18th International Congress of Mechanical Engineering. November 6-11, 2005, Ouro Preto. **Proceedings of the 18 th International Congress of Mechanical Engineering**, Ouro Preto, 2005. v.1. p.1-8.

PAL, S. K.; CHAKRABORTY, D. Surface roughness prediction in turning using artificial neural network, **Neural Computing & Applications**, v.14, p.319–324, 2005.

PANDA, S. S.; MAHAPATRA, S. S. PCA fused NN approach for drill wear prediction in drilling mild steel specimen. In: IEEE ICCSIT 2009. August 8-11, 2009, Beijing, China. **Proceedings of the 2nd IEEE International Conference on Computer Science and Information Technology**, Beijing: IEEE, 2009, p.85-89.

PONTES, F. J. **Predição da vida de ferramentas e da rugosidade da peça por Redes Neurais RBF projetadas com uso da Metodologia de Projeto de Experimentos**. 2006. 166f. Dissertação de mestrado – UNIFEI – Universidade Federal de Itajubá, 2006.

PONTES, F. J.; FERREIRA, J. R.; SILVA, M. B.; PAIVA, A. P.; BALESTRASSI, P. P. Artificial neural networks for machining process surface roughness modeling. **International Journal of Advanced Manufacturing Technology**, v.49, n.9-12, p.879-902, 2010a.

PONTES, F. J.; SILVA, M. B.; FERREIRA, J. R.; PAIVA, A. P.; BALESTRASSI, P. P.; SCHÖNHORST, G. B. A DOE Based Approach for the Design of RBF Artificial Neural Networks Applied to Prediction of Surface Roughness in AISI 52100 Hardened Steel Turning. **Journal of Brazilian Society of Mechanical Sciences and Engineering (Impresso)** v. 32, n.5, p. 503-510, 2010.

POTOCNIK, P.; GRABEC, I. Nonlinear model predictive control of a cutting process, **Neurocomputing**, v.43, p.107-126, 2002.

QUEIROZ, A. F.; OLIVEIRA, F. A.; MARANGON LIMA, J. W.; BALESTRASSI, P. P. Simulating Electricity Spot Prices in Brazil Using Neural Network and Design of Experiments. In: IEEE Power Tech 2007, July 1-5, 2007, Lausanne, Switzerland. **Proceedings of the International Conference on Power System Technology**, Lausanne: IEEE, 2007, p.2029-2034.

QUIZA, R.; FIGUEIRA, L.; DAVIM, J. P. Comparing statistical models and artificial neural networks on predicting the tool wear in hard machining D2 AISI steel. **International Journal of Advanced Manufacturing Technology**, v.37, p.641-648, 2008.

REDDY, N. S. K.; RAO, P. V. Selection of optimum tool geometry and cutting conditions using a surface roughness prediction model for end milling. **International Journal of Advanced Manufacturing Technology**, v.26, p.1202-1210, 2005.

REED, R. Pruning algorithms—a survey. **IEEE Transactions on Neural Networks**, v.4, n.5, p.740-746, 1993.

ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. **Psychology Review**, v.65, p.386-408, 1958.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **Nature**, v.323, p.533-536, 1986.

SALGADO JUNIOR, A. R. **Otimização de múltiplos duais correlacionados no processo de torneamento de aço de corte fácil ABNT 12L14**. 2010. 166f. Dissertação de mestrado – UNIFEI – Universidade Federal de Itajubá, 2010.

SANDVIK-Coromant, **Manual Técnico de Usinagem**, Sandviken, Suécia, 2005.

SANJAY, C.; JYOTHI, C. A study of surface roughness in drilling using mathematical analysis and neural networks, **International Journal of Advanced Manufacturing Technology**, v.29, p.846–852, 2006.

SARKAR, S.; MITRA, S.; BHATTACHARYYA, B. Parametric optimisation of wire electrical discharge machining of γ titanium aluminide alloy through an artificial neural network model, **International Journal of Advanced Manufacturing Technology**, v.27, p.501–508, 2006.

SARMA, D. K.; DIXIT, U. S. A comparison of dry and air-cooled turning of grey cast iron with mixed oxide ceramic tool, **Journal of Materials Processing Technology**, v.190, p.160–172, 2007.

SHARMA, V. S.; DHIMAN, S.; SEHGAL, R.; SHARMA, S. K. Estimation of cutting forces and surface roughness for hard turning using neural networks. **Journal of Intelligent Manufacturing**, v.19, p.473–483, 2008.

SHAW, M. **Metal Cutting Principles**. Oxford University Press, New York, 2004.

SICK, B. On-line and indirect tool wear monitoring in turning with artificial neural networks: a review of more than a decade of research. **Mechanical Systems and Signal Processing**, v.16, n.4, p.487–546, 2002.

SINGH, D.; RAO, P. V., A surface roughness prediction model for hard turning process. **International Journal of Advanced Manufacturing Technology**, v.32, p.1115–1124, 2007.

SIVA, K.; MURUGAN, N.; LOGESH, R. Optimization of weld bead geometry in plasma transferred arc hardfaced austenitic stainless steel plates using genetic algorithm, **International Journal of Advanced Manufacturing Technology**, v.41, p.24-30, 2009.

SONAR, D. K.; DIXIT, U. S.; OJHA, D. K. The application of a radial basis function neural network for predicting the surface roughness in a turning process, **International Journal of Advanced Manufacturing Technology**, v.27, p.661–666, 2006.

SPECHT, D. F. A general regression neural network. **IEEE Transactions on Neural Networks**, v.2, n.6, p.568-576, 1991.

SUKTHOMYA, W.; TANNOCK, J.D.T. Taguchi experimental design for manufacturing process optimization using historical data and a neural network process model, **International Journal of Quality & Reliability Management**, v.22, n.5, p.485-502, 2005.

TAGUCHI, G.; ELSAYED, E. A.; HSIANG, T. **Engenharia da Qualidade em Sistemas de Produção**. McGraw-Hill, São Paulo, 1987.

TAMIZHARASAN, T.; SEVARAJ, T.; HAQ, A. N. Analysis of tool wear and surface finish in hard turning. **International Journal of Advanced Manufacturing Technology**, v.28, p.671–679, 2006.

TANSEL, I. N.; OZCELIK, B.; BAOA, W. Y.; CHEN, P.; RINCON, D.; YANG, S. Y.; YENILMEZ, A. Selection of optimal cutting conditions by using GONNS, **International Journal of Machine Tools & Manufacture**, v.46, p.26–35, 2006.

TRENT, E. M., WRIGHT, P. K. **Metal Cutting**. Butterworth Heinemann, Woburn, MA: 2000.

TSAI, K. M.; WANG, P. J. Predictions on surface finish in electrical discharge machining based upon neural network models, **International Journal of Machine Tools & Manufacture**, v.41, p.1385–1403, 2001.

TSAO, C. C.; HOCHENG, H. Evaluation of thrust force and surface roughness in drilling composite material using Taguchi analysis and neural network, **Journal of Materials Processing Technology**, v.203, p.342–348, 2008.

WANG, S. G.; HSU, Y. L. One-pass milling machining parameter optimization to achieve mirror surface roughness, **Proceedings of the Institution of Mechanical Engineers Part B: J Engineering Manufacture**, v.219, p.177-181, 2005.

XIAO, G.; ZHU, Z. Friction materials development by using DOE/RSM and artificial neural network, **Tribology International**, v.43, p.218-227, 2010.

YANG, F. Neural network metamodeling for cycle time-throughput profiles in manufacturing, **European Journal of Operational Research**, v.205, p.172-185, 2010.

YANG, S.H.; NATARAJAN, U.; SEKAR, M.; PALANI, S. Prediction of surface roughness in turning operations by computer vision using neural network trained by differential evolution algorithm, **International Journal of Advanced Manufacturing Technology**, v.51, p.965-971, 2010.

ZAIN, A.M.; HARON, H.; SHARIF, S. Prediction of surface roughness in the end milling machining using Artificial Neural Network, **Expert Systems with Applications**, v.37, p.1755-1768, 2010.

ZANCHETTIN, C.; MINKU, F. L.; LUDERMIR, T. B. Design of experiments in neuro-fuzzy systems. In: IEEE CONFERENCE ON HYBRID INTELLIGENT SYSTEMS 2005. December 6–9, 2005, Rio de Janeiro. **Proceedings of the 5th International Conference on Hybrid Intelligent Systems**, Rio de Janeiro: IEEE, 2005, p.218–226.

ZHONG, Z. W.; KHOO, L. P.; HAN, S. T. Prediction of surface roughness of turned surfaces using neural networks. **International Journal of Advanced Manufacturing Technology**, v.28, p.688–693, 2006.

APÊNDICE A: CONJUNTOS DE DADOS FORMADOS A PARTIR DE CURVAS SINTÉTICAS PARA TREINAMENTO E TESTE DE REDES

Quadro A1 – Conjunto com 30 casos de treinamento e 10 casos de seleção

V_c	f	a_p	R_a
-0,266	0,026	-1,263	0,387
0,056	-0,928	-1,097	0,281
0,211	0,903	-0,260	0,535
-0,996	-0,631	0,669	0,256
-0,422	-0,876	0,890	0,266
-1,619	-0,759	-1,464	0,404
1,250	-0,678	0,466	0,402
-0,692	-0,096	-1,061	0,387
-1,304	-1,602	-1,377	0,337
-1,443	-1,323	1,191	0,199
0,810	1,489	-0,856	0,693
0,828	-0,800	-0,572	0,310
-1,514	-0,745	-1,088	0,367
0,456	-1,373	-0,992	0,289
0,768	-0,932	1,002	0,380
0,478	0,834	1,118	0,520
0,414	-1,271	-0,263	0,308
-1,336	1,144	0,482	0,607
0,837	0,322	0,984	0,457
-1,293	0,692	-0,072	0,525
0,236	1,326	0,592	0,637
1,546	-0,343	-0,036	0,405
1,028	0,571	-1,526	0,429
0,238	0,659	0,149	0,474
-1,060	-0,076	-0,551	0,385
1,130	0,364	1,561	0,517
-1,149	1,149	1,212	0,540
-1,084	-1,401	1,003	0,229
-1,186	-0,525	1,241	0,228
0,602	-0,265	-0,882	0,325
1,204	-1,028	-0,467	0,339
-1,580	1,567	-0,431	0,870
0,786	0,166	0,525	0,418
1,205	1,356	0,038	0,668
-1,290	0,517	-0,517	0,518
-0,819	1,099	1,174	0,530
-0,289	0,965	1,428	0,499
0,638	-0,937	-0,493	0,304
0,260	-1,312	-0,401	0,298
-0,464	-1,611	0,417	0,289

Quadro A2 – Conjunto com 15 casos de treinamento e 5 casos de seleção

V_c	f	a_p	R_a
-0,266	0,026	-1,263	0,387
0,056	-0,928	-1,097	0,281
0,211	0,903	-0,260	0,535
-0,996	-0,631	0,669	0,256
-0,422	-0,876	0,890	0,266
-1,619	-0,759	-1,464	0,404
1,250	-0,678	0,466	0,402
-0,692	-0,096	-1,061	0,387
-1,304	-1,602	-1,377	0,337
-1,443	-1,323	1,191	0,199
0,810	1,489	-0,856	0,693
0,828	-0,800	-0,572	0,310
-1,514	-0,745	-1,088	0,367
0,456	-1,373	-0,992	0,289
0,768	-0,932	1,002	0,380
1,204	-1,028	-0,467	0,339
-1,580	1,567	-0,431	0,870
0,786	0,166	0,525	0,418
1,205	1,356	0,038	0,668
-1,290	0,517	-0,517	0,518

Quadro A3 – Conjunto com 90 casos de teste

V_c	f	a_p	R_a
0,221	1,047	0,569	0,560
1,044	0,843	1,513	0,581
0,849	-1,561	-1,270	0,296
0,600	1,308	0,119	0,642
-1,047	1,545	-0,758	0,836
0,834	-0,219	-0,707	0,338
0,977	-1,497	1,002	0,445
-1,607	-0,140	-1,225	0,469
-1,501	1,092	0,462	0,601
-1,197	1,207	-0,019	0,665
-1,147	0,561	-0,484	0,516
0,745	1,604	0,127	0,735
0,182	1,432	0,655	0,666
-0,979	0,199	0,871	0,344
-0,718	0,018	0,674	0,330
-1,441	-1,455	0,373	0,247
-0,052	1,193	-0,520	0,628
-0,279	0,050	-1,513	0,398
-0,413	0,508	0,650	0,421
0,459	0,642	1,205	0,481
0,336	0,284	-1,524	0,402
-0,154	-0,888	0,386	0,284
-1,352	-1,118	0,222	0,255
-0,224	0,342	1,230	0,381
-0,708	1,277	0,127	0,650
-0,505	0,541	0,480	0,433
-1,425	0,243	1,368	0,310
-1,199	0,494	0,247	0,447
1,284	0,005	0,785	0,457
1,547	-0,453	-0,612	0,356
-1,267	-1,055	-0,753	0,309
1,552	-0,973	-0,668	0,347

Quadro A3 – Conjunto com 90 casos de teste – continuação

<i>Vc</i>	<i>f</i>	<i>ap</i>	<i>Ra</i>
-1,607	-0,918	-0,056	0,284
-0,277	-0,425	1,104	0,291
-0,848	1,006	-0,749	0,633
-0,898	1,095	0,190	0,598
-0,886	1,303	0,080	0,670
1,051	-1,585	1,133	0,476
0,935	-1,017	-0,351	0,327
0,302	0,765	0,660	0,496
0,930	1,373	-0,462	0,658
0,968	-1,387	-0,670	0,326
-0,137	-1,374	-0,833	0,284
-1,374	-0,721	1,059	0,220
0,608	-1,372	-0,188	0,328
0,996	0,696	-0,882	0,475
0,938	0,853	-1,115	0,506
0,905	-1,412	0,061	0,368
1,468	-1,013	-0,865	0,326
-0,818	-0,362	0,962	0,270
0,484	-0,025	-1,436	0,345
-1,434	1,430	0,275	0,729
0,453	-1,014	1,488	0,362
0,123	0,958	0,209	0,541
0,102	0,985	-0,788	0,568
-0,592	0,704	0,412	0,472
-1,068	-0,049	1,053	0,294
-1,596	-0,618	-0,998	0,380
0,244	-1,254	0,269	0,313
1,168	0,698	1,331	0,562
-0,485	1,283	-0,332	0,669
1,167	0,711	0,031	0,513
1,131	-1,553	-0,175	0,386
-1,225	-0,220	-0,428	0,364
-1,395	0,842	0,702	0,502
-0,270	0,640	0,327	0,461
0,099	-1,256	-0,009	0,298
1,414	-1,511	-0,816	0,354
-0,287	-0,889	-0,216	0,282
-1,225	1,563	-0,929	0,879
0,809	1,478	0,352	0,697
-0,174	0,117	0,291	0,367
-1,443	0,634	-0,751	0,582
0,618	1,581	0,968	0,723
0,780	0,291	1,511	0,463
-1,343	1,157	1,323	0,533
1,568	0,483	1,251	0,573
0,520	-0,110	1,210	0,384
1,581	0,355	1,609	0,583
0,333	-0,712	1,023	0,331
-0,959	0,829	-1,034	0,612
-0,749	-1,527	-1,558	0,309
0,404	1,075	-0,098	0,577
-0,859	-0,202	-0,190	0,339
-0,529	1,299	-0,370	0,679
-0,030	1,523	-0,346	0,727
-1,612	0,091	-0,731	0,469
-1,504	-1,362	-0,560	0,301
1,319	-1,385	1,401	0,519
1,333	-0,475	-0,084	0,378

APÊNDICE B: CONJUNTOS DE DADOS EXTRAÍDOS DO ARTIGO 1 PARA TREINAMENTO E TESTE DE REDES

Quadro B1 – Conjunto com 30 casos de treinamento e 4 casos de seleção Fonte:

Sharma *et al.* (2008)

AA	Vc	f	ap	Ra
45,00	36,6	0,1	0,3	2,8
45,00	36,6	0,1	0,9	3,88
45,00	36,6	0,1	1,5	3,92
45,00	81,7	0,1	0,6	1,34
45,00	196,0	0,1	0,6	1,25
45,00	36,6	0,17	0,6	3,33
45,00	36,6	0,27	0,6	6,98
60,00	36,6	0,1	0,3	1,47
60,00	36,6	0,1	0,9	1,44
60,00	36,6	0,1	1,5	1,59
60,00	51,5	0,1	0,6	1,54
60,00	196,0	0,1	0,6	1,61
60,00	36,6	0,17	0,6	3,25
60,00	36,6	0,27	0,3	1,43
60,00	36,6	0,27	0,9	1,48
60,00	36,6	0,27	1,5	1,53
75,00	36,6	0,1	0,6	1,67
75,00	51,5	0,1	0,6	1,48
75,00	81,7	0,1	0,6	1,61
75,00	196,0	0,1	0,6	1,28
75,00	36,6	0,17	0,6	3,52
75,00	36,6	0,27	0,6	8,04
90,00	36,6	0,1	0,3	1,25
90,00	36,6	0,1	0,6	1,16
90,00	36,6	0,1	1,2	1,1
90,00	51,5	0,1	0,6	1,24
90,00	81,7	0,1	0,6	1,02
90,00	126,6	0,1	0,6	0,98
90,00	36,6	0,13	0,6	1,51
90,00	36,6	0,21	0,6	2,82
45,00	51,5	0,1	0,6	1,4
45,00	36,6	0,13	0,6	1,93
60,00	36,6	0,1	1,2	1,49
60,00	36,6	0,13	0,6	1,93

Quadro B2 – Conjunto com 17 casos de teste

Fonte: Sharma *et al.* (2008)

<i>AA</i>	<i>Vc</i>	<i>f</i>	<i>ap</i>	<i>Ra</i>
45,00	36,6	0,1	0,6	3,82
45,00	36,6	0,1	1,2	3,9
45,00	126,6	0,1	0,6	0,84
45,00	36,6	0,21	0,6	4,57
60,00	36,6	0,1	0,6	1,53
60,00	81,7	0,1	0,6	1,56
60,00	126,6	0,1	0,6	1,6
60,00	36,6	0,21	0,6	4,73
60,00	36,6	0,27	0,6	7,49
60,00	36,6	0,27	1,2	1,28
75,00	126,6	0,1	0,6	1,5
75,00	36,6	0,13	0,6	2,07
75,00	36,6	0,21	0,6	5,04
90,00	36,6	0,1	0,9	1,07
90,00	36,6	0,1	1,5	1,16
90,00	196,0	0,1	0,6	0,94
90,00	36,6	0,17	0,6	2,12

APÊNDICE C: CONJUNTOS DE DADOS EXTRAÍDOS DO ARTIGO 2 PARA TREINAMENTO E TESTE DE REDES

Quadro C1 – Conjunto com 21 casos de treinamento e 9 casos de seleção

Fonte: Kohli & Dixit (2005)

V_c	a_p	f	a	R_a
102,87	0,1	0,32	5,75	4,89
102,87	0,1	0,32	6,42	4,72
235,84	0,1	0,04	5,91	3,43
235,84	0,1	0,04	4,41	2,5
226,87	0,1	0,32	4,98	3,53
226,87	0,1	0,32	5,24	3,28
236,48	1,0	0,04	6,2	2,43
236,48	1,0	0,04	7,03	2,48
106,26	0,1	0,04	2,45	2,32
106,26	0,1	0,04	2,86	2,53
96,12	1,0	0,04	12,84	3,38
96,12	1,0	0,04	6,83	2,72
97,04	1,0	0,32	5,16	3,65
97,04	1,0	0,32	5,11	3,46
238,41	1,0	0,32	11,47	2,43
238,41	1,0	0,32	13,5	2,24
191,49	0,6	0,16	19,4	3,2
130,89	0,1	0,2	8,58	6,16
197,92	0,3	0,12	8,32	2,22
138,19	0,6	0,24	11,13	3,99
232,0	0,1	0,16	7,51	2,4
130,31	0,1	0,24	5,87	5,62
197,42	0,6	0,12	14,12	2,87
184,06	0,1	0,3	6,81	4,7
87,15	0,6	0,32	5,94	6,26
221,67	0,6	0,16	15,87	2,33
89,9	0,6	0,28	12,06	5,72
233,55	0,6	0,08	5,59	2,41
212,27	1,0	0,12	8,8	2,35
228,15	0,1	0,28	5,55	3,52

Quadro C2 – Conjunto com 80 casos de teste

Fonte: Kohli & Dixit (2005)

<i>Vc</i>	<i>ap</i>	<i>f</i>	<i>a</i>	<i>Ra</i>
184,06	0,1	0,3	6,81	4,7
160,42	0,1	0,16	4,89	2,22
197,42	0,6	0,12	14,12	2,87
228,15	0,1	0,28	5,55	3,62
203,86	0,6	0,28	8,69	2,79
103,32	0,1	0,28	6,29	4,18
223,67	1,0	0,08	8,66	2,65
120,37	1,0	0,16	7,53	2,38
135,7	0,3	0,32	4,51	3,58
132,06	0,1	0,12	6,42	4,08
201,88	0,1	0,08	6,62	2,59
159,65	0,1	0,12	6,04	3,74
126,97	0,6	0,2	6,7	3,49
224,64	0,3	0,2	5,16	2,31
129,14	0,1	0,32	5,14	3,56
196,9	0,1	0,2	6,81	2,8
105,81	0,1	0,08	2,61	2,36
131,47	0,1	0,16	9,5	4,94
202,86	0,1	0,04	4,62	2,21
128,48	1,0	0,04	4,93	2,54
192,97	0,3	0,12	8,32	2,22
131,53	0,6	0,16	6,72	2,92
191,4	0,6	0,16	19,4	3,2
157,37	0,1	0,28	7,28	4,42
218,7	0,3	0,28	5,71	3,18
215,72	0,6	0,24	6,74	1,85
197,92	0,6	0,32	7,69	2,62
103,33	0,6	0,08	8,53	3,94
222,16	1,0	0,08	7,84	2,46
114,5	1,0	0,24	8,61	2,8
233,28	0,1	0,12	4,54	2,31
194,95	0,3	0,16	9,92	2,06
203,86	0,6	0,28	8,69	2,79
104,9	0,1	0,12	5,71	4,39
104,53	0,1	0,16	4,93	3,29
104,23	0,1	0,2	6,38	5,96
102,87	0,1	0,32	6,17	6,74
132,64	0,1	0,08	4,86	2,97
129,72	0,1	0,28	5,06	4,71
166,5	0,1	0,08	8,35	4,2

Quadro C2 – Conjunto com 80 casos de teste

Fonte: Kohli & Dixit (2005) – continuação

<i>Vc</i>	<i>ap</i>	<i>f</i>	<i>a</i>	<i>Ra</i>
158,89	0,1	0,2	6,41	5,22
163,45	0,1	0,24	4,38	3,68
199,89	0,1	0,12	5,52	2,31
197,92	0,1	0,16	4,84	2,15
196,9	0,1	0,2	6,81	2,8
188,52	0,1	0,28	6,79	3,81
184,06	0,1	0,32	6,48	4,91
235,84	0,1	0,04	5,92	3,43
234,56	0,1	0,08	5,83	2,29
229,44	0,1	0,2	11,16	3,26
226,87	0,1	0,32	4,98	3,53
131,47	0,3	0,04	11,5	4,43
129,72	0,3	0,08	6,92	2,61
126,22	0,3	0,16	8,45	3,62
150,33	0,3	0,08	5,96	2,27
147,11	0,3	0,12	6,26	3,26
144,83	0,3	0,16	4,72	2,52
142,55	0,3	0,2	4,6	2,66
140,27	0,3	0,24	4,44	2,84
137,98	0,3	0,28	4,62	3,5
135,7	0,3	0,32	4,51	3,57
236,51	0,3	0,04	3,02	1,9
233,55	0,3	0,08	3,3	1,97
221,67	0,3	0,24	5,17	2,5
230,58	0,3	0,12	3,89	2,18
227,61	0,3	0,16	4,45	2,58
212,76	0,3	0,32	5,79	3,26
142,93	0,6	0,04	7,83	3,04
134,11	0,6	0,28	8,87	3,82
127,1	0,32	0,6	9,63	4,69
203,36	0,08	0,6	11,15	2,23
179,61	0,24	0,6	7,19	2,02
173,68	0,28	0,6	7,72	3,47
167,73	0,32	0,6	6,81	3,53
227,61	0,12	0,6	8,92	2,64
183,57	0,08	1,0	10,1	2,55
166,49	0,16	1,0	9,46	1,79
158,89	0,2	1,0	10,51	2,37
151,36	0,24	1,0	9,98	2,5
136,08	0,32	1,0	10,75	3,51

APÊNDICE D: CONJUNTOS DE DADOS EXTRAÍDOS DO ARTIGO 3 PARA TREINAMENTO E TESTE DE REDES

Quadro D1 – Conjunto com 15 casos de treinamento e 3 casos de seleção

Fonte: Sarkar *et al.* (2006)

T_{ON}	T_{OFF}	I_P	WT	SV	FR	MS	Ra
1,1	14,0	120,0	900,0	2,0	7,0	2,67	2,77
1,1	20,0	170,0	1140,0	6,0	8,5	2,15	2,78
1,1	30,0	220,0	1380,0	10,0	10,0	1,38	2,78
1,6	14,0	120,0	1140,0	6,0	10,0	2,24	2,85
1,6	20,0	170,0	1380,0	10,0	7,0	2,12	2,81
1,6	30,0	220,0	900,0	2,0	8,5	2,08	2,97
0,8	14,0	170,0	900,0	10,0	10,0	1,84	2,33
0,8	20,0	220,0	1140,0	2,0	7,0	1,69	2,48
0,8	30,0	120,0	1380,0	6,0	8,5	0,98	2,3
1,1	14,0	220,0	1380,0	6,0	7,0	2,1	2,82
1,1	20,0	120,0	900,0	10,0	8,5	1,56	2,64
1,1	30,0	170,0	1140,0	2,0	10,0	1,78	2,78
1,6	14,0	170,0	1380,0	2,0	8,5	2,58	2,87
1,6	20,0	220,0	900,0	6,0	10,0	2,56	2,92
1,6	30,0	120,0	1140,0	10,0	7,0	2,25	3,03
0,8	14,0	220,0	1140,0	10,0	8,5	1,69	2,39
0,8	20,0	120,0	1380,0	2,0	10,0	1,14	2,3
0,8	30,0	170,0	900,0	6,0	7,0	1,07	2,42

Quadro D2 – Conjunto com 6 casos de teste

Fonte: Sarkar *et al.* (2006)

T_{ON}	T_{OFF}	I_P	WT	SV	FR	MS	Ra
1,6	30,0	170,0	900,0	2,0	10,0	2,24	3,16
0,8	20,0	170,0	1380,0	6,0	8,5	1,14	2,55
0,8	14,0	220,0	1140,0	2,0	8,5	2,1	2,31
1,6	14,0	170,0	900,0	2,0	7,0	2,9	2,84
1,6	14,0	220,0	900,0	6,0	8,5	2,87	2,72
1,1	30,0	220,0	900,0	6,0	8,5	1,65	2,62

APÊNDICE E: CONJUNTO DE TESTES RESULTANTE DOS EXPERIMENTOS DE USINAGEM COM AÇO ABNT 12L14 UTILIZADO PARA TESTE DE REDES

Quadro E1 – Conjunto com 24 casos de teste para redes, resultante do experimento de
usinagem de confirmação com aço ABNT 12L14

<i>Vc</i>	<i>f</i>	<i>ap</i>	<i>Esbeltez</i>	<i>Desgaste</i>	<i>Posição</i>	<i>Ra</i>
251	0,086	0,6	Grossa	Nova	CP	0,905
251	0,086	0,6	Grossa	Nova	CE	1,053
251	0,086	0,6	Grossa	Nova	CA	1,114
251	0,086	0,6	Grossa	Usada	CP	2,103
251	0,086	0,6	Grossa	Usada	CE	2,673
251	0,086	0,6	Grossa	Usada	CA	2,771
251	0,086	0,6	Fina	Nova	CP	0,786
251	0,086	0,6	Fina	Nova	CE	0,759
251	0,086	0,6	Fina	Nova	CA	0,782
251	0,086	0,6	Fina	Usada	CP	1,609
251	0,086	0,6	Fina	Usada	CE	1,454
251	0,086	0,6	Fina	Usada	CA	1,69
345	0,086	0,68	Grossa	Nova	CP	1,428
345	0,086	0,68	Grossa	Nova	CE	1,358
345	0,086	0,68	Grossa	Nova	CA	1,351
345	0,086	0,68	Grossa	Usada	CP	1,844
345	0,086	0,68	Grossa	Usada	CE	2,061
345	0,086	0,68	Grossa	Usada	CA	2,167
345	0,086	0,68	Fina	Nova	CP	1,843
345	0,086	0,68	Fina	Nova	CE	0,796
345	0,086	0,68	Fina	Nova	CA	0,826
345	0,086	0,68	Fina	Usada	CP	1,774
345	0,086	0,68	Fina	Usada	CE	1,728
345	0,086	0,68	Fina	Usada	CA	1,918