

UNIVERSIDADE FEDERAL DE ITAJUBÁ
Pró-Reitoria de Pesquisa e Pós-Graduação
Programa de Pós-Graduação em Engenharia de Produção

**PREDIÇÃO DA VIDA DE FERRAMENTAS E DA
RUGOSIDADE DA PEÇA POR REDES NEURAIS
RBF PROJETADAS COM USO DA
METODOLOGIA DE PROJETO DE
EXPERIMENTOS**

Fabício José Pontes

Dissertação submetida ao Programa de Pós-Graduação em Engenharia de Produção como requisito parcial à obtenção do título de *Mestre em Engenharia de Produção*.

Orientador: Prof. João Roberto Ferreira, Dr.

Co-orientador: Prof. Anderson Paulo de Paiva, Dr.

Itajubá

2006

Ficha catalográfica elaborada pela Biblioteca Mauá –
Bibliotecária Margareth Ribeiro- CRB_6/1700

P814p

Pontes, Fabrício José

Predição da vida de ferramentas e da rugosidade da peça por
Redes Neurais RBF projetadas com uso da Metodologia de Pro_
jeto de Experimentos / Fabrício José Pontes. -- Itajubá (MG):
[s.n.], 2006.

166p. : il.

Orientador: Prof. Dr. João Roberto Ferreira.

Co-orientador: Prof. Dr. Anderson Paulo de Paiva.

Dissertação (Mestrado) - Universidade Federal de Itajubá.

1. Redes neurais. 2. Projeto de Experimentos (DOE). 3. Vida
da ferramenta. 4. Rugosidade da peça. I. Ferreira, João Roberto,
orient. II. Paiva, Anderson Paulo de, co-orient. III. Universidade
Federal de Itajubá. IV. Título.

CDU 004.032.26(043)

UNIVERSIDADE FEDERAL DE ITAJUBÁ
Pró-Reitoria de Pesquisa e Pós-Graduação
Programa de Pós-Graduação em Engenharia de Produção

Fabício José Pontes

**PREDIÇÃO DA VIDA DE FERRAMENTAS E DA
RUGOSIDADE DA PEÇA POR REDES NEURAIS
RBF PROJETADAS COM USO DA
METODOLOGIA DE PROJETO DE
EXPERIMENTOS**

Dissertação aprovada por banca examinadora em 16 de novembro de 2006, conferindo
ao autor o título de *Mestre em Engenharia de Produção*

Banca Examinadora:

Prof. Dr. João Roberto Ferreira (Orientador)

Prof. Dr. Anderson Paulo de Paiva (Co-orient.)

Prof. Dr. José Arnaldo Barra Montevechi

Prof. Dr. Reginaldo Teixeira Coelho

Itajubá

2006

*A Milena,
que me apoia e incentiva desde o início,
dedico o resultado do meu trabalho.*

AGRADECIMENTOS

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

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

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

A Milena, muitíssimo obrigado por me fazer acreditar mais em minha capacidade, por me incentivar a voltar aos estudos, pela compreensão por todos os dias e noites que passei trabalhando, e pelos conselhos tão úteis durante este trabalho.

Ao meu amigo e segundo irmão Anderson, que acreditou no meu potencial, confiou em mim e foi o grande responsável por minha volta aos estudos, meu especial agradecimento.

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

Agradeço a todos os professores do Departamento de Engenharia de Produção da UNIFEI. Os conhecimentos adquiridos nestes dois anos de relacionamento acadêmico certamente me tornaram um profissional mais completo e também uma pessoa com visão mais ampla da realidade.

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 enfim pela compreensão a todos os meus parentes e amigos, dos quais não foi possível estar próximo durante este período de estudos.

SUMÁRIO

1. Introdução	1
1.1. Considerações Iniciais	1
1.2. Objetivos	3
1.3. Justificativas	3
1.4. Limitações	5
1.5. Problemas de Pesquisa	6
1.6. Proposições	6
1.7. Metodologia de Pesquisa	7
1.8. Variáveis de Estudo	8
1.8.1. Dependentes (Y)	8
1.8.2. Independentes (X)	8
1.9. Estrutura do Trabalho	9
2. Vida da Ferramenta e Rugosidade	11
2.1. Usinagem e Torneamento	11
2.2. Desgaste e Vida da Ferramenta	12
2.3. Rugosidade da Peça	16
2.4. Considerações Finais	20
3. Redes Neurais Artificiais	21
3.1. Considerações Iniciais	21
3.2. Histórico das Redes Neurais Artificiais	23
3.3. Aprendizado	25
3.4. Projeto de Redes Neurais	27
3.5. Redes de Arquitetura MLP (Perceptron Multi-Camada)	31
3.6. Treinamento de Redes MLP	32
3.7. Redes de Arquitetura RBF (Função de Base Radial)	33
3.8. Projeto e Treinamento de Redes RBF	36
3.9. Redes Neurais Aplicadas à Tarefa de Predição	39
3.9.1. Considerações Gerais	39
3.9.2. Estudos com Redes Neurais Aplicadas à Predição	41
3.10. Considerações Finais	48
4. Metodologia de Projeto e Análise de Experimentos	49
4.1. Considerações Iniciais	49
4.2. Metodologia de Taguchi e Arranjos Ortogonais	50
4.2.1. Projeto Robusto	50
4.2.2. Planejamento de Experimentos no Método de Taguchi	52
4.2.3. Análise de Experimentos no Método de Taguchi	53
4.3. Fatoriais Completos	54
4.4. Análise de Variância (ANOVA)	55
4.5. Teste para Significância do Modelo	59
4.6. Teste para Significância Individual dos Coeficientes do Modelo	59
4.7. Modelos de Regressão Linear	59
4.8. Medidas de Adequação de Modelos de Regressão	64
4.9. Análise Residual	65
4.10. Testes para a Falta de Ajuste dos Modelos	66
4.11. Considerações Finais	68

5. Procedimento Experimental	69
5.1. Considerações Iniciais	69
5.2. Delimitação do Escopo e Evolução do Trabalho	69
5.3. Definição dos Fatores Experimentais e de Seus Níveis	71
5.4. Organização dos Experimentos	72
5.5. Geração dos Dados Experimentais	73
5.5.1. Considerações Gerais	73
5.5.2. Geração dos Dados para Vida da Ferramenta	73
5.5.3. Geração dos Dados para Rugosidade Média (R_a)	75
5.6. Planejamento Experimental	76
5.6.1. Experimentos com Arranjos Ortogonais de Taguchi	76
5.6.2. Experimentos com Fatoriais Completos Mistos	77
5.6.3. Experimentos Definitivos com Fatoriais Completos	79
5.7. Execução dos Experimentos	85
6. Análise dos Resultados para Vida da Ferramenta	87
6.1. Considerações Iniciais	87
6.2. Resultados dos Experimentos com Arranjos Ortogonais de Taguchi	87
6.3. Resultados dos Experimentos com Fatoriais Completos Mistos	96
6.4. Resultados dos Experimentos Definitivos com Fatoriais Completos	102
6.5. Análise dos Resultados dos Experimentos Definitivos com Fatoriais Completos	108
7. Análise dos Resultados para Rugosidade Média (R_a)	118
7.1. Considerações Iniciais	118
7.2. Resultados dos Experimentos com Arranjos Ortogonais de Taguchi	118
7.3. Resultados dos Experimentos com Fatoriais Completos Mistos	127
7.4. Resultados dos Experimentos Definitivos com Fatoriais Completos	133
7.5. Análise dos Resultados dos Experimentos Definitivos com Fatoriais Completos	139
8. Conclusão	149
8.1. Considerações Iniciais	149
8.2. Conclusões sobre Vida da Ferramenta e Rugosidade da Peça	149
8.3. Recomendações para Trabalhos Futuros	153
Apêndice A – Tópicos em Estatística e Simulação	154
A.1 Inferência Sobre a Média de Uma População com Variância Conhecida	154
A.2 Teste de Levene	154
A.3 Diagrama de Pareto	155
A.4 Transformações de Box-Cox	156
A.5 Meta-Modelos e Simulação	157
Referências Bibliográficas	161

LISTA DE FIGURAS

Figura 2.1 – Retirada de material em processo de usinagem	11
Figura 2.2 – Elementos em operação de torneamento	12
Figura 2.3 – Desgastes em ferramenta após atingir critério de fim de vida	14
Figura 2.4 – Quebra de ferramenta	14
Figura 2.5 – Desvios no formato de superfícies	16
Figura 2.6 - Desvio médio aritmético R_a	18
Figura 3.1 – O neurônio de McCulloch e Pitts	24
Figura 3.2 – Conceito de sobre-ajuste de redes	27
Figura 3.3 – Rede MLP	31
Figura 3.4 – Função de base radial	34
Figura 3.5 – Rede neural de arquitetura RBF	36
Figura 6.1 – Efeitos principais sobre a média da resposta Razão de Desvios Padrão para conjunto de treinamento de 300 casos	89
Figura 6.2 – Efeitos principais sobre a função sinal-ruído para conjunto de treinamento de 300 casos	89
Figura 6.3 – Efeitos principais sobre o desvio padrão da resposta Razão de Desvios Padrão para conjunto de treinamento de 300 casos	90
Figura 6.4 – Box-plot da resposta Razão de Desvios Padrão para as melhores configurações obtidas em cada experimento envolvendo arranjos ortogonais de Taguchi para vida da ferramenta	95
Figura 6.5 – Efeitos principais para a resposta Razão de Desvios Padrão, para experimento com arranjos fatoriais mistos envolvendo 400 casos de treinamento para vida da ferramenta	97
Figura 6.6 – Efeitos de interação para a resposta Razão de Desvios Padrão, para experimento com arranjos fatoriais mistos envolvendo 400 casos de treinamento para vida da ferramenta	97
Figura 6.7 - Transformação Box-Cox para resposta Razão de Desvios Padrão, experimento para vida da ferramenta com arranjos fatoriais completos utilizando 48 casos de treinamento	104
Figura 6.8 - Efeitos principais para resposta Razão de Desvios Padrão, experimento para vida da ferramenta com arranjos fatoriais completos utilizando 48 casos de treinamento	105

Figura 6.9 - Efeitos de interação para resposta Razão de Desvios Padrão, experimento para vida da ferramenta com arranjos fatoriais completos utilizando 48 casos de treinamento	105
Figura 6.10 - Análise da normalidade dos resíduos, experimento para vida da ferramenta com arranjos fatoriais completos utilizando 48 casos de treinamento	106
Figura 6.11 - Diagrama de Pareto ($\alpha=0,05$), experimento para vida da ferramenta com arranjos fatoriais completos utilizando 48 casos de treinamento	108
Figura 6.12 – Diagrama de rede RBF obtida durante o experimento para vida da ferramenta com a utilização de 60 casos de treinamento	110
Figura 6.13 – Diagrama de superfície de erro da melhor rede obtida para vida da ferramenta com a utilização de 60 casos de treinamento	111
Figura 6.14 – Resultado do teste de comparação entre as variâncias das melhores configurações de rede obtidas para vida da ferramenta durante experimentos utilizando 48 e 60 casos de treinamento	115
Figura 6.15 – Box-plot das melhores configurações de rede obtidas para vida da ferramenta durante os experimentos definitivos com arranjos fatoriais completos	116
Figura 6.16 – Box-plot das melhores configurações de rede de 24 a 500 casos obtidas para vida da ferramenta durante os experimentos definitivos com arranjos fatoriais completos	117
Figura 6.17 – Comparação de desempenho da predição da vida da ferramenta feita entre as melhores configurações de rede e método da matriz Pseudo-Inversa para otimização linear	117
Figura 7.1 – Efeitos principais sobre a média da resposta Razão de Desvios Padrão para conjunto de treinamento de 300 casos	120
Figura 7.2 – Efeitos principais sobre a função sinal-ruído para conjunto de treinamento de 300 casos	120
Figura 7.3 – Efeitos principais sobre o desvio padrão da resposta Razão de Desvios Padrão para conjunto de treinamento de 300 casos	121
Figura 7.4 – Box-plot da resposta Razão de Desvios Padrão para as melhores configurações obtidas em cada experimento envolvendo arranjos ortogonais de Taguchi para Rugosidade Média (R_a)	126
Figura 7.5 – Efeitos principais para a resposta Razão de Desvios Padrão, para experimento com arranjos fatoriais mistos envolvendo 500 casos de treinamento para Rugosidade Média (R_a)	128

Figura 7.6 – Efeitos de interação para a resposta Razão de Desvios Padrão, para experimento com arranjos fatoriais mistos envolvendo 500 casos de treinamento para Rugosidade Média (R_a)	128
Figura 7.7 - Transformação Box-Cox para resposta Razão de Desvios Padrão, experimento para Rugosidade Média com arranjos fatoriais completos utilizando 60 casos de treinamento	135
Figura 7.8 - Efeitos principais para resposta Razão de Desvios Padrão, experimento para Rugosidade Média com arranjos fatoriais completos utilizando 60 casos de treinamento	136
Figura 7.9 - Efeitos de interação para resposta Razão de Desvios Padrão, experimento para Rugosidade Média com arranjos fatoriais completos utilizando 60 casos de treinamento	136
Figura 7.10 - Análise da normalidade dos resíduos, experimento para Rugosidade Média com arranjos fatoriais completos utilizando 60 casos de treinamento	137
Figura 7.11 - Diagrama de Pareto ($\alpha=0,05$), experimento para Rugosidade Média com arranjos fatoriais completos utilizando 60 casos de treinamento	139
Figura 7.12 – Diagrama de rede RBF obtida para o experimento para Rugosidade Média com a utilização de 120 casos de treinamento	141
Figura 7.13 – Diagrama de superfície de erro da melhor rede obtida para Rugosidade Média com a utilização de 120 casos de treinamento	142
Figura 7.14 – Resultado do teste de comparação entre as variâncias das melhores configurações de rede obtidas para Rugosidade Média durante experimentos utilizando 60 e 120 casos de treinamento	146
Figura 7.15 – Box-plot das melhores configurações de rede obtidas para Rugosidade Média durante os experimentos definitivos com arranjos fatoriais completos	147
Figura 7.16 – Box-plot das melhores configurações de rede de 24 a 500 casos obtidas para Rugosidade Média durante os experimentos definitivos com arranjos fatoriais completos	148
Figura 7.17 – Comparação de desempenho da predição da Rugosidade Média feita entre as melhores configurações de rede e método da matriz Pseudo-Inversa para otimização linear	148
Figura A-1 – Diagrama de Pareto	156
Figura A-2 – Ilustração do conceito da transformação de Box-Cox	157
Figura A.3 – Ilustração do conceito de meta-modelos	159

LISTA DE TABELAS

Tabela 6.1 – Valores Médios de Razão de Desvios Padrão obtidos durante os experimentos envolvendo Arranjos Ortogonais de Taguchi para Vida da Ferramenta	88
Tabela 6.2 – Valores do desvio padrão da resposta Razão de Desvios Padrão obtidos durante os experimentos envolvendo Arranjos Ortogonais de Taguchi para Vida da Ferramenta	88
Tabela 6.3 – Configurações Apontadas como as melhores para Predição da Vida da Ferramenta, para cada experimento envolvendo Arranjos Ortogonais de Taguchi	90
Tabela 6.4 – Configurações de Níveis com Melhor Desempenho para os experimentos envolvendo Arranjos Ortogonais de Taguchi e Respectivos Valores de Razão de Desvios Padrão para Predição da Vida da Ferramenta	92
Tabela 6.5 – Valores P Resultantes do Teste para Inferência Sobre a Média Entre Pares de Amostras para as melhores configurações obtidas em cada experimento envolvendo Arranjos Ortogonais de Taguchi para Vida da Ferramenta	92
Tabela 6.6 – Valores P Resultantes do Teste de Levene Entre Pares de Amostras para as melhores configurações obtidas em cada experimento envolvendo Arranjos Ortogonais de Taguchi para Vida da Ferramenta	94
Tabela 6.7 – Análise de Variância para experimento com Arranjos Fatoriais Mistos envolvendo 400 casos de treinamento para Vida da Ferramenta	98
Tabela 6.8 – Valores Médios dos Mínimos Quadrados dos Efeitos Principais, para o experimento envolvendo 12 casos de treinamento, com uso de Fatoriais Mistos, para Vida da Ferramenta	99
Tabela 6.9 – Valores Médios dos Mínimos Quadrados dos Efeitos Principais, para o experimento envolvendo 15 casos de treinamento, com uso de Fatoriais Mistos, para Vida da Ferramenta	99
Tabela 6.10 – Valores Médios dos Mínimos Quadrados dos Efeitos Principais, para o experimento envolvendo 24 casos de treinamento, com uso de Fatoriais Mistos, para Vida da Ferramenta	100
Tabela 6.11 – Valores Médios dos Mínimos Quadrados dos Efeitos Principais, para o experimento envolvendo 30 casos de treinamento, com uso de Fatoriais Mistos, para Vida da Ferramenta	100

Tabela 6.12 – Valores Médios dos Mínimos Quadrados dos Efeitos Principais, para o experimento envolvendo 48 casos de treinamento, com uso de Fatoriais Mistos, para Vida da Ferramenta	100
Tabela 6.13 – Valores Médios dos Mínimos Quadrados dos Efeitos Principais, para o experimento envolvendo 60 casos de treinamento, com uso de Fatoriais Mistos, para Vida da Ferramenta	100
Tabela 6.14 – Valores Médios dos Mínimos Quadrados dos Efeitos Principais, para o experimento envolvendo 120 casos de treinamento, com uso de Fatoriais Mistos, para Vida da Ferramenta	101
Tabela 6.15 – Valores Médios dos Mínimos Quadrados dos Efeitos Principais, para o experimento envolvendo 150 casos de treinamento, com uso de Fatoriais Mistos, para Vida da Ferramenta	101
Tabela 6.16 – Valores Médios dos Mínimos Quadrados dos Efeitos Principais, para o experimento envolvendo 240 casos de treinamento, com uso de Fatoriais Mistos, para Vida da Ferramenta	101
Tabela 6.17 – Valores Médios dos Mínimos Quadrados dos Efeitos Principais, para o experimento envolvendo 300 casos de treinamento, com uso de Fatoriais Mistos, para Vida da Ferramenta	101
Tabela 6.18 – Valores Médios dos Mínimos Quadrados dos Efeitos Principais, para o experimento envolvendo 400 casos de treinamento, com uso de Fatoriais Mistos, para Vida da Ferramenta	102
Tabela 6.19 – Valores Médios dos Mínimos Quadrados dos Efeitos Principais, para o experimento envolvendo 500 casos de treinamento, com uso de Fatoriais Mistos, para Vida da Ferramenta	102
Tabela 6.20 – Médias e desvios padrão dos Experimentos para Vida da Ferramenta com Fatorial Completo utilizando 12, 15 e 24 casos de treinamento. Fonte: Minitab® 13.0	102
Tabela 6.21 – Médias e desvios padrão dos Experimentos para Vida da Ferramenta com Fatorial Completo utilizando 30, 48 e 60 casos de treinamento. Fonte: Minitab® 13.0	103
Tabela 6.22 – Médias e desvios padrão dos Experimentos para Vida da Ferramenta com Fatorial Completo utilizando 120, 150 e 240 casos de treinamento. Fonte: Minitab®13.0	103
Tabela 6.23 – Médias e desvios padrão dos Experimentos para Vida da Ferramenta com Fatorial Completo utilizando 300, 400 e 500 casos de treinamento. Fonte: Minitab®13.0	103
Tabela 6.24 – Efeitos Estimados dos Coeficientes, experimento para Vida Ferramenta com Arranjo Fatorial Completo utilizando 48 casos de treinamento	107

Tabela 6.25 – Análise de Variância, experimento para Vida da Ferramenta com Arranjo Fatorial Completo utilizando 48 casos de treinamento	107
Tabela 6.26 – Níveis dos Fatores Apontados como os melhores para Predição da Vida da Ferramenta, para os experimentos definitivos envolvendo Arranjos Fatoriais Completos	109
Tabela 6.27 – Configurações Apontadas como as melhores para Predição da Vida da Ferramenta, para cada experimento envolvendo Arranjos Fatoriais Completos	109
Tabela 6.28 – Sumário dos Efeitos Significativos para Vida da Ferramenta, obtidos dos experimentos definitivos com Fatoriais Completos	112
Tabela 6.29 – Valores Médios da resposta Razão de Desvios Padrão, desvios padrão e Erros-Padrão da Média para as melhores Redes obtidas para Predição da Vida da Ferramenta durante o ciclo experimental definitivo com uso de Fatoriais Completos	113
Tabela 6.30 – Valores P obtidos dos Testes de ANOVA para Comparação das Médias das Respostas das melhores configurações de Rede para Vida da Ferramenta	114
Tabela 6.31 – Valores P Resultantes do Teste de Levene para Comparação das Variâncias das melhores configurações de Rede para Vida da Ferramenta	115
Tabela 7.1 – Valores Médios de Razão de Desvios Padrão obtidos durante os experimentos envolvendo Arranjos Ortogonais de Taguchi para Rugosidade Média (R_a)	119
Tabela 7.2 – Valores do Desvio padrão da resposta Razão de Desvios Padrão obtidos durante os experimentos envolvendo Arranjos Ortogonais de Taguchi para Rugosidade Média (R_a)	119
Tabela 7.3 – Configurações Apontadas como as melhores para Predição da Rugosidade Média (R_a), para cada experimento envolvendo Arranjos Ortogonais de Taguchi	121
Tabela 7.4 – Configurações de Níveis com Melhor Desempenho para os experimentos envolvendo Arranjos Ortogonais de Taguchi e Respectivos Valores de Razão de Desvios Padrão para Predição da Rugosidade Média (R_a)	123
Tabela 7.5 – Valores P Resultantes do Teste para Inferência Sobre a Média Entre Pares de Amostras para as melhores configurações obtidas em cada experimento envolvendo Arranjos Ortogonais de Taguchi para Rugosidade Média (R_a)	123
Tabela 7.6 – Valores P Resultantes do Teste de Levene Entre Pares de Amostras para as melhores configurações obtidas em cada experimento envolvendo Arranjos Ortogonais de Taguchi para Rugosidade Média (R_a)	125
Tabela 7.7 – Análise de Variância para experimento com Arranjos Fatoriais Mistos envolvendo 500 casos de treinamento para Rugosidade Média (R_a)	129

Tabela 7.8 – Valores Médios dos Mínimos Quadrados dos Efeitos Principais, para o experimento envolvendo 12 casos de treinamento, com uso de Fatoriais Mistos, para Rugosidade Média (R_a)	130
Tabela 7.9 – Valores Médios dos Mínimos Quadrados dos Efeitos Principais, para o experimento envolvendo 15 casos de treinamento, com uso de Fatoriais Mistos, para Rugosidade Média (R_a)	130
Tabela 7.10 – Valores Médios dos Mínimos Quadrados dos Efeitos Principais, para o experimento envolvendo 24 casos de treinamento, com uso de Fatoriais Mistos, para Rugosidade Média (R_a)	131
Tabela 7.11 – Valores Médios dos Mínimos Quadrados dos Efeitos Principais, para o experimento envolvendo 30 casos de treinamento, com uso de Fatoriais Mistos, para Rugosidade Média (R_a)	131
Tabela 7.12 – Valores Médios dos Mínimos Quadrados dos Efeitos Principais, para o experimento envolvendo 48 casos de treinamento, com uso de Fatoriais Mistos, para Rugosidade Média (R_a)	131
Tabela 7.13 – Valores Médios dos Mínimos Quadrados dos Efeitos Principais, para o experimento envolvendo 60 casos de treinamento, com uso de Fatoriais Mistos, para Rugosidade Média (R_a)	131
Tabela 7.14 – Valores Médios dos Mínimos Quadrados dos Efeitos Principais, para o experimento envolvendo 120 casos de treinamento, com uso de Fatoriais Mistos, para Rugosidade Média (R_a)	132
Tabela 7.15 – Valores Médios dos Mínimos Quadrados dos Efeitos Principais, para o experimento envolvendo 150 casos de treinamento, com uso de Fatoriais Mistos, para Rugosidade Média (R_a)	132
Tabela 7.16 – Valores Médios dos Mínimos Quadrados dos Efeitos Principais, para o experimento envolvendo 240 casos de treinamento, com uso de Fatoriais Mistos, para Rugosidade Média (R_a)	132
Tabela 7.17 – Valores Médios dos Mínimos Quadrados dos Efeitos Principais, para o experimento envolvendo 300 casos de treinamento, com uso de Fatoriais Mistos, para Rugosidade Média (R_a)	132
Tabela 7.18 – Valores Médios dos Mínimos Quadrados dos Efeitos Principais, para o experimento envolvendo 400 casos de treinamento, com uso de Fatoriais Mistos, para Rugosidade Média (R_a)	133

Tabela 7.19 – Valores Médios dos Mínimos Quadrados dos Efeitos Principais, para o experimento envolvendo 500 casos de treinamento, com uso de Fatoriais Mistos, para Rugosidade Média (R_a)	133
Tabela 7.20 – Médias e desvios padrão dos Experimentos para Rugosidade Média com Fatorial Completo utilizando 12, 15 e 24 casos de treinamento. Fonte: Minitab® 13.0	133
Tabela 7.21 – Médias e desvios padrão dos Experimentos para Rugosidade Média com Fatorial Completo utilizando 30, 48 e 60 casos de treinamento. Fonte: Minitab® 13.0	134
Tabela 7.22 – Médias e desvios padrão dos Experimentos para Rugosidade Média com Fatorial Completo utilizando 120, 150 e 240 casos de treinamento. Fonte: Minitab®13.0	134
Tabela 7.23 – Médias e desvios padrão dos Experimentos para Rugosidade Média com Fatorial Completo utilizando 300, 400 e 500 casos de treinamento. Fonte: Minitab® 13.0	134
Tabela 7.24 – Efeitos Estimados dos Coeficientes, experimento para Rugosidade Média (R_a) com Arranjo Fatorial Completo utilizando 48 casos de treinamento	138
Tabela 7.25 – Análise de Variância, experimento para Rugosidade Média (R_a) com Arranjo Fatorial Completo utilizando 48 casos de treinamento	138
Tabela 7.26 – Níveis dos Fatores Apontados como os melhores para Predição da Rugosidade Média (R_a), para os experimentos definitivos envolvendo Arranjos Fatoriais Completos	140
Tabela 7.27 – Configurações Apontadas como as melhores para Predição da Rugosidade Média (R_a), para cada experimento envolvendo Arranjos Fatoriais Completos	140
Tabela 7.28 – Sumário dos Efeitos Significativos para Rugosidade Média (R_a), obtidos dos experimentos definitivos com Fatoriais Completos	143
Tabela 7.29 – Valores Médios da resposta Razão de Desvios Padrão, desvios padrão e Erros-Padrão da Média para as melhores Redes obtidas para Predição da Rugosidade durante o ciclo experimental definitivo com uso de Fatoriais Completos	144
Tabela 7.30 – Valores P obtidos dos Testes de ANOVA para Comparação das Médias das Respostas das melhores configurações de Rede para Rugosidade Média	145
Tabela 7.31 – Valores P resultantes do Teste de Levene entre pares de amostras para as melhores configurações obtidas em cada experimento do ciclo experimental definitivo para Rugosidade Média (R_a) com o uso de Fatoriais Completos	147

LISTA DE QUADROS

Quadro 4.1 – Arranjo L8 para experimento com um fator com quatro níveis e dois fatores com dois níveis, sem investigação da interação	53
Quadro 4.2 – Hipóteses Presentes em uma ANOVA para dois Fatores	56
Quadro 4.3 – Fórmulas para Análise de Variância para dois Fatores em dois Níveis	57
Quadro 4.4 – Fórmulas para Análise de Variância em dois Níveis	58
Quadro 5.1 – Arranjo L8 para o ciclo experimental envolvendo Arranjos Ortogonais de Taguchi	76
Quadro 5.2 – Fatores e Níveis Associados aos Fatores dos experimentos envolvendo Arranjos Ortogonais de Taguchi	77
Quadro 5.3 –Número de Unidades Radiais em cada experimento envolvendo Arranjos Ortogonais de Taguchi	77
Quadro 5.4 – Arranjo utilizado durante o ciclo experimental envolvendo Fatoriais Completos Mistos	78
Quadro 5.5 – Fatores e Níveis Associados aos Fatores dos Experimentos com a utilização de Arranjos Fatoriais Completos Mistos	79
Quadro 5.6 –Número de Unidades Radiais em cada experimento com a utilização de Fatoriais Completos Mistos	79
Quadro 5.7 – Arranjo utilizado durante o ciclo experimental definitivo envolvendo Fatoriais Completos	80
Quadro 5.8 – Níveis dos Fatores utilizados nos experimentos definitivos com Fatorial Completo para Vida da Ferramenta, experimentos com 12 e 15 casos de treinamento	81
Quadro 5.9 – Níveis dos Fatores utilizados nos experimentos definitivos com Fatorial Completo para Vida da Ferramenta, experimentos com 24 e 30 casos de treinamento	81
Quadro 5.10 – Níveis dos Fatores utilizados nos experimentos definitivos com Fatorial Completo para Vida da Ferramenta, experimentos com 48 e 60 casos de treinamento	81
Quadro 5.11 – Níveis dos Fatores utilizados nos experimentos definitivos com Fatorial Completo para Vida da Ferramenta, experimentos com 120 e 150 casos de treinamento	82
Quadro 5.12 – Níveis dos Fatores utilizados nos experimentos definitivos com Fatorial Completo para Vida da Ferramenta, experimentos com 240 e 300 casos de treinamento	82
Quadro 5.13 – Níveis dos Fatores utilizados nos experimentos definitivos com Fatorial Completo para Vida da Ferramenta, experimentos com 400 e 500 casos de treinamento	82

Quadro 5.14 – Níveis dos Fatores utilizados nos experimentos definitivos com Fatorial Completo para Rugosidade Média, experimentos com 12 e 15 casos de treinamento	83
Quadro 5.15 – Níveis dos Fatores utilizados nos experimentos definitivos com Fatorial Completo para Rugosidade Média, experimentos com 24 e 30 casos de treinamento	83
Quadro 5.16 – Níveis dos Fatores utilizados nos experimentos definitivos com Fatorial Completo para Rugosidade Média, experimentos com 48 e 60 casos de treinamento	83
Quadro 5.17 – Níveis dos Fatores utilizados nos experimentos definitivos com Fatorial Completo para Rugosidade Média, experimentos com 120 e 150 casos de treinamento	84
Quadro 5.18 – Níveis dos Fatores utilizados nos experimentos definitivos com Fatorial Completo para Rugosidade Média, experimentos com 240 e 300 casos de treinamento	84
Quadro 5.19 – Níveis dos Fatores utilizados nos experimentos definitivos com Fatorial Completo para Rugosidade Média, experimentos com 400 e 500 casos de treinamento	84

LISTA DE ABREVIATURAS E SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
ANOVA	<i>Analysis of Variance</i>
ART2	<i>Adaptive Resonance Theory 2</i>
CNC	<i>Computer Numerical Control</i>
DIN	<i>Deutsches Institut für Normung</i>
DOE	<i>Design of Experiments</i>
F	Distribuição de Fisher e Snedecor
GRNN	<i>General Regression Neural Network</i>
HRC	<i>Rockwell C Hardness</i>
ISO	<i>International Organization for Standardization</i>
MLP	Multi Layer Perceptron
MQ Ajustada	Média Quadrática Ajustada
OLS	<i>Ordinary Least Squares</i>
P	Valor P
RBF	<i>Radial Basis Function</i>
Relação S/R	Relação Sinal Ruído
RSM	<i>Response Surface Methodology</i>
R ² adj	Coefficiente de Determinação Ajustado
SAE	<i>Society of Automotive Engineers</i>
S. D. Ratio	<i>Standard Deviation Ratio</i>
SOFM	<i>Self Organized Feature Map</i>
SQ Ajustada	Soma Quadrática Ajustada
SQ Seq	Soma Quadrática Seqüencial
StDev	<i>Standard Deviation</i>
t	Distribuição T de Student
T	Distribuição T de Student

LISTA DE SÍMBOLOS

Al_2O_3	Trióxido de Alumínio
a_p	Profundidade de Usinagem
α	Probabilidade do Erro Tipo I
β	Probabilidade do Erro Tipo II
β_i	Coefficiente do Modelo de Regressão
C	Constante Multiplicadora da Fórmula de Taylor
C1	Expoente da Velocidade de Corte na fórmula (5) para Rugosidade
C2	Expoente do Avanço na fórmula (5) para Rugosidade
C3	Expoente da Profundidade de Usinagem na fórmula (5) 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
$F_{critico}$	Valor Crítico da Estatística de Teste F
f_M	Função Meta-Modelar
f(.)	Função de Ativação
ϕ_i	Função de Base Radial i
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
GHz	GigaHertz
G^T	Transposta da Matriz G
H_0	Hipótese Nula
H_1	Hipótese Alternativa
$h(x)$	Função de saída da Rede Neural RBF
K	Número de Unidades Radiais Vizinhas
KT	Profundidade Máxima de Cratera
l	Expoente da Velocidade de Corte na fórmula de Taylor
L	Valor da Função de Mínimos Quadrados
\log_{10}	Função Logaritmo Decimal

l_m	Comprimento de Amostragem do Perfil
L8	Arranjo ortogonal de Taguchi com oito corridas experimentais
λ	Parâmetro da Transformação Box-Cox
m	Expoente do Avanço na Fórmula de Taylor
M	Linha Média
MB	Megabytes
mm	Milímetro
mm/v	Milímetros por Volta
MS _i	Média Quadrática do Fator i
m/min	Metros por Minuto
μ	Média Populacional
μm	Micrômetro
$\bar{\mu}_j$	Hiper-Centro Vetorial da Função de Ativação j
n	Expoente da Profundidade de Usinagem na Fórmula de Taylor
N	Número de Vetores de Entrada
n _a	Número de amostras
n _d	Número de Parâmetros Livres da Rede / Dimensão da Função de Erro
η	Função sinal-ruído
R _a	Rugosidade Média
RAM	Random Access Memory
r _e	Raio de Ponta da Ferramenta
R _{max}	Rugosidade Máxima
R ²	Coefficiente 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	Desvio padrão de uma População
σ	Parâmetro de Largura da Função Radial
\vec{t}	Espaço Vetorial Unidimensional
T	Vida da Ferramenta
TiC	Carboneto de Titânio

TiN	Nitreto de Titânio
Ti6Al4V	Liga de Titânio-Alumínio
VBB	Desgaste de flanco médio
VBMax	Desgaste de flanco máximo
VN	Desgaste de Entalhe
V_c	Velocidade de corte
X	Entrada de função, variável independente
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	Saída de função, variável dependente
y_i	Valor absoluto da ordenada do perfil medido em relação à linha média
y_i	Desvio em Relação à Característica da Qualidade
y_i	Saída do modelo polinomial quadrático 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	Distribuição de Levene
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
$\forall i$	Para todo i

RESUMO

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. Itajubá 166p. Dissertação de Mestrado – Universidade Federal de Itajubá (2006).*

A predição da vida da ferramenta é importante para garantir conformidade e evitar danos à peça e à máquina. A qualidade de superfície é requisito essencial para produtos usinados. Seu principal indicador é a rugosidade da peça. Predizer de modo exato a vida da ferramenta pode representar redução de custos e aumento da produtividade. A predição da rugosidade, por outro lado, pode contribuir para melhoria da qualidade do produto e para minimização de tempos e custos. Realizar tais predições é difícil devido à não linearidade associada aos processos de desgaste e de formação da rugosidade. Redes neurais mostram-se eficazes para predição em processos que envolvem não linearidade, como é o caso dos processos mencionados. Este trabalho estuda o desempenho e a variabilidade de redes neurais RBF (função de base radial), projetadas com auxílio da metodologia de projeto de experimentos (DOE), na predição da vida da ferramenta e da rugosidade no torneamento do aço SAE/ABNT 52100 (55 HRC) com ferramenta de cerâmica mista. Os fatores utilizados no planejamento experimental são parâmetros de projeto das redes neurais (número de unidades radiais, algoritmo para determinação de centros da função radial e algoritmo para determinação do parâmetro de largura da função radial). Os parâmetros de corte são utilizados como entradas das redes. A grandeza de saída utilizada para medida do desempenho é a Razão de Desvios Padrão da fase de testes das redes. São investigados os efeitos dos fatores de projeto e do tamanho do conjunto de treinamento sobre o desempenho das redes na predição. Para tanto são executados experimentos com diferentes quantidades de casos de treinamento. Possíveis efeitos de interação entre fatores de projeto das redes RBF também são investigados. Os resultados do estudo são expressos na forma de parâmetros de projeto de rede para cada conjunto de treinamento utilizado. As melhores redes obtidas pelo uso do método proposto apresentam exatidão e precisão crescentes com o aumento do número de exemplos. Outro resultado é a estimação da importância relativa dos fatores de projeto no desempenho das redes. O trabalho chega à conclusão de que os efeitos de interação entre níveis dos fatores envolvidos são significativos para o desempenho das redes RBF nas tarefas propostas. Uma comparação de desempenho entre as redes neurais e um método de ajuste de curvas e otimização linear indica superioridade das redes na modelagem da vida da ferramenta e da rugosidade. As conclusões sugerem que a metodologia de projeto de experimentos pode constituir uma abordagem sistemática para projeto de redes neurais superior à procura de configurações por tentativa e erro, ou à estratégia de variar-se um parâmetro por vez.

Palavras-chave: *Redes Neurais, Projeto de Experimentos (DOE), Vida da ferramenta, Rugosidade da peça, Aço 52100 (55 HRC).*

ABSTRACT

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. Itajubá 166p. Dissertação de Mestrado – Universidade Federal de Itajubá (2006).*

Prediction of tool life is important to ensure conformity of the product and to avoid damage to the product itself and to the machine. Surface roughness is an essential requirement for machined products. The main figure for that requirement is surface roughness. Accurate forecasts of tool life may lead to cost reduction and increase in productivity. Prediction of roughness may contribute to improve product quality and to reduce production times and costs. To perform such predictions is a difficult task due to intrinsic non-linearity that characterizes the processes involved. Neural networks are proven tools for prediction in processes involving non-linearity, as it is the case for prediction of tool life and roughness. This work is a study on the performance and variability of RBF (Radial Basis Function) neural networks, projected with support of the methodology of Design of Experiments (DOE) applied to prediction of tool life and surface roughness in the turning of a SAE 52100 hardened steel (55 HRC) with mixed ceramic tool. The factors employed in experimental planning are parameters of project of the networks (number of radial units, algorithm for determination of radial centers and algorithm for determination of the spread factor). Cutting process parameters are employed as input for the networks. The output variable chosen to measure network performance is the S. D. Ratio obtained during the testing phase of the networks. Effects of project parameters and of size of training set on network performance are investigated. In order to do that, experiments with training sets of different sizes are conducted. Possible effects of interaction among factors are also investigated. Results are expressed as network project parameters, for each training set employed. The best networks obtained by means of the proposed method show accuracy and precision increasingly better as the number of available examples for training grows up. Other results are the ranking of relative importance of project factors in network performance. The work concludes that interaction effects among levels of factors involved are statistically significant to the performance of RBF networks in the proposed tasks. A comparison between RBF networks and a linear optimization method points to superior performance of the networks in modeling tool life and roughness. Conclusions suggest that the approach of using the methodology of Design of Experiments (DOE) as a tool for project of networks for prediction of tool life and surface roughness may constitute a better option than the trial and error approach or than the strategy of varying one factor at a time in the search for high performance network configurations.

Keywords: *Neural networks, Design of Experiments (DOE), Tool Life, Surface Roughness, SAE 52100 hardened steel (55 HRC).*

1. INTRODUÇÃO

1.1 Considerações Iniciais

Materiais e ligas dos mais diversos tipos passam pelo processo de usinagem. A maioria das formas de peças produzidas no mundo da engenharia são usinadas.

Segundo Trent (2000), a competição econômica leva à evolução tecnológica e à pesquisa de novos materiais para ferramental bem como de métodos que otimizem o processo de usinagem. Como afirmam Benardos e Vosniakos (2003), engenheiros envolvidos em processos de manufatura enfrentam dois problemas práticos principais: a determinação dos valores de parâmetros do processo que leva à obtenção da qualidade que se almeja para o produto e a maximização do desempenho do sistema de manufatura utilizando os recursos disponíveis. Dentro desse contexto, a capacidade de prever a duração da vida de ferramentas e a rugosidade da peça é de grande utilidade para a indústria metal-mecânica, e constitui-se em um problema que tem atraído a atenção de muitos pesquisadores.

Exemplos do impacto da evolução tecnológica são a recente tendência à redução ou eliminação do uso de fluidos de corte e a introdução de novos materiais de ferramentas, que têm motivado trabalhos de pesquisa em que se objetiva modelar os efeitos sobre a vida da ferramenta e sobre a rugosidade causados por essas novas condições de operação. Isto pode ser observado no trabalho de Arenas Miranda (2003), no qual se propõe um método para obtenção de velocidades ótimas para furação bem como se estudam os efeitos da furação sem uso de fluidos de corte sobre o acabamento da peça de aço-liga ABNT 4340. O surgimento de novos materiais, para os quais não existem modelos, também motiva o estudo do desgaste e da vida de ferramentas. Ferreira (1999) realizou trabalho de investigação das propriedades de usinagem de materiais compósitos reforçados com fibras de carbono. Paiva et al. (2005) realizaram um estudo detalhado sobre a aplicação da metodologia de projeto de experimentos (DOE) na modelagem da vida da ferramenta no processo de usinagem de um aço SAE/ABNT 52100 – 55 HRC com ferramenta de cerâmica mista. Já Ezugwu et al. (2005) afirmam que as desejáveis características de extrema resistência presentes em superligas são também obstáculos para a usinagem destes materiais. Segundo os autores, considerável esforço tem sido feito no sentido de obter-se mais conhecimento sobre os processos de usinagem dos mesmos, com o objetivo de tornar tais operações mais eficazes e econômicas.

A predição da vida da ferramenta é valiosa para evitar a não conformidade da peça, danos a ela e à máquina, que podem levar à perda de produtividade e aumento dos custos de produção, segundo Özel e Nadgir (2002). Lee et al. (1996) afirmam que tal tarefa, no entanto, é extremamente difícil graças à não linearidade associada aos processos de desgaste.

Paiva et al. (2005) afirmam que o estabelecimento de um relacionamento funcional entre vida da ferramenta e os parâmetros de corte em vários níveis de velocidade de corte, avanço e profundidade de usinagem implicam em grande número de experimentos, que torna os custos com experimentação proibitivos. A troca antecipada de ferramentas de corte devida à incerteza na predição da duração da vida da mesma é comum nos processos de usinagem, relatam Diniz et al. (2005). Segundo Sick (2002), a supervisão do desgaste e a predição da vida da ferramenta são tarefas difíceis no contexto do monitoramento em processos de usinagem. Diversos estudos que objetivam a predição da vida da ferramenta utilizam-se com sucesso de redes neurais (Alajmi, Oraby e Esat, 2005).

A qualidade de superfície é um dos mais importantes requisitos do cliente quanto aos produtos usinados. A rugosidade da peça, segundo Özel e Karpat (2005), é o principal indicador de tal requisito. Ela é uma medida da qualidade tecnológica de um produto e fator que muito influencia no custo da manufatura. Benardos e Vosniakos (2003) afirmam que a rugosidade da peça é largamente empregada como índice de qualidade de produto e, em muitos casos, requisito técnico para produtos da indústria mecânica. Os autores sustentam que a qualidade de superfície é de grande importância no comportamento funcional de uma peça.

Vários modelos teóricos têm sido propostos para a predição da rugosidade. Tais modelos não são exatos e se aplicam, em sua maioria, a faixa limitada de condições de corte, ou devem ser usados em conjunto com diagramas e tabelas. Há necessidade, por parte da indústria, de ferramentas que permitam a predição da rugosidade da peça antes da usinagem da mesma, e que possam, ao mesmo tempo, ser facilmente empregadas no chão de fábrica, contribuindo na minimização de tempos e custos (BENARDOS e VOSNIAKOS, 2002). Um estudo de Benardos e Vosniakos (2003) descreve as principais estratégias utilizadas na busca pela melhoria na predição da rugosidade. Observa-se a obtenção de bons resultados com o emprego de redes neurais artificiais para tal tarefa.

A proposta do presente trabalho consiste em realizar estudo da predição da vida de uma ferramenta específica assim como da rugosidade superficial com a utilização de redes neurais artificiais projetadas segundo a Metodologia de Projeto de Experimentos (DOE). Far-se-á uso dos arranjos ortogonais de Taguchi bem como da técnica de Fatoriais Completos

como ferramenta para o projeto dos parâmetros das redes que as leva a bom desempenho na predição.

1.2 Objetivos

O objetivo deste trabalho é estudar o desempenho e a variabilidade de redes neurais artificiais de arquitetura RBF (função de base radial), projetadas com auxílio da metodologia de projeto de experimentos (DOE), na predição da vida da ferramenta e da rugosidade da peça no torneamento do aço SAE/ABNT 52100 com ferramenta de cerâmica mista. Os fatores a serem utilizados para o planejamento experimental são os parâmetros de projeto das redes neurais.

Serão investigados os efeitos dos fatores no desempenho das redes na tarefa de predição. A influência do tamanho do conjunto de treinamento sobre o desempenho das redes também será analisada, para o que serão executados experimentos com conjuntos de treinamento de diferentes tamanhos. Possíveis efeitos de interação entre os fatores de projeto das redes RBF também serão investigados.

1.3 Justificativas

Segundo Sanjay et al. (2005), o monitoramento de processos de usinagem é problema clássico e ainda não solucionado. De acordo com Choudury e Bartarya (2003), o desgaste de ferramentas corresponde a grande limitação na automação em operações de usinagem. Sustentam os autores que qualidade do produto, estratégias de trocas de ferramentas, custos com ferramentas e produtividade são todas influenciadas pelo desgaste das mesmas. Turk et al (2004) afirmam que, em processos de prensagem, uma predição acurada da vida de uma ferramenta pode representar uma redução de até 30% nos custos de produção. Para Özel e Karpaz (2005), em operações envolvendo torneamento pesado, a indústria utiliza ferramentas de corte até a totalidade de sua vida em apenas um terço das operações. Melhorias no planejamento do processo para torneamento pesado, segundo os autores, são necessárias para redução de custos e aumento da produtividade. Tansel et al. (2006) destacam a dificuldade de se encontrar expressões analíticas ou empíricas para calcular condições ótimas de operação que evitem a quebra prematura de ferramentas e reduzam a rugosidade da peça.

De acordo com Sick (2002), os custos com ferramentas e com *setup* podem ser sensivelmente reduzidos com o uso das mesmas até o limite de sua vida. O autor afirma que uma previsão exata do desgaste da ferramenta permite o ajuste da sua posição com o objetivo

de atender às especificações geométricas, bem como permite controlar a taxa de desgaste para garantir certo nível de qualidade da superfície da peça usinada (medida por sua rugosidade).

Segundo Cus e Zuperl (2006), os fatores que influenciam o planejamento da tecnologia de usinagem são o tipo de operação (torneamento, fresamento, entre outros), os parâmetros da máquina utilizada (rigidez e potência, entre outros), parâmetros de corte, características do material da peça e da ferramenta de corte utilizada. Dentre estes, segundo os autores, os parâmetros de corte são os mais influentes. Com efeito, observa-se na literatura uma tendência à utilização de tais parâmetros para predição da vida e da rugosidade.

A aplicação de modelos analíticos para predição da vida da ferramenta pode ser recomendada se os mesmos existirem e quando se trata de modelos testados e validados. Sick (2002) reporta, porém, que a experiência com modelos analíticos é muito pobre. Modelos analíticos apresentam, segundo o autor, duas desvantagens fundamentais:

- Constantes dos modelos precisam ser determinadas experimentalmente para cada combinação diferente de condições de corte, tipo de ferramenta e material da peça;
- Alguns modelos usam como entradas sinais que são de medição impossível ou podem ser medidos somente com grande investimento.

Em relação à rugosidade, Benardos e Vosniakos (2003) afirmam que os fenômenos que levam à sua formação são extremamente complexos e repletos de interações entre os fatores, o que torna sua previsão difícil. Afirmam ainda que a estratégia mais comum adotada pela indústria consiste na seleção de parâmetros de processo conservadores, os quais nem garantem a obtenção do nível de acabamento desejado nem proporcionam altas taxas de remoção de metal.

Existem diversos modelos teóricos e experimentais para predição da rugosidade superficial. De acordo com Benardos e Vosniakos (2003) tais modelos são baseados em convenções e idealizações, que conduzem a erros e limitações na sua aplicabilidade. Os mesmos autores defendem que modelos gerados a partir de técnicas de inteligência artificial (tais como as redes neurais) tendem a ser mais realistas e exatos.

Técnicas típicas da área de inteligência artificial, como sistemas especialistas, para otimização em processos de usinagem têm sido observados na literatura (Batista, 2000). Sick (2002) relata a efetividade das redes neurais na predição da vida de ferramentas em processos de torneamento. Benardos e Vosniakos (2003) realizaram uma extensa revisão bibliográfica sobre a predição da rugosidade em processo de usinagem. Seu trabalho também aponta para a adequação das técnicas baseadas no uso de redes neurais para a tarefa de predição. Benardos e Vosniakos (2003) citam avanços mais recentes na área como o desenvolvimento de novos

algoritmos de treinamento e aumento do poder computacional como algumas das razões que têm incrementado o interesse pela mesma.

A revisão da literatura indica que redes neurais podem apresentar resultados superiores aos obtidos por métodos de regressão linear múltipla em tarefas de predição em usinagem. Choudhury e Bartarya (2003), compararam a técnica de projeto de experimentos com técnicas envolvendo redes neurais na predição do desgaste de ferramentas. Seu trabalho apontou que redes neurais apresentam melhor desempenho na tarefa proposta que a técnica de projeto de experimentos. El-Mounayri et al (2005) afirmam que, comparadas aos métodos de computação tradicionais, redes neurais são robustas e versáteis.

A utilização do processo de torneamento para investigação da vida de ferramenta e da rugosidade é justificada, segundo Sick (2002), porque o processo de torneamento é menos complexo que outros processos de usinagem tais como furação e fresamento, devido ao fato de que somente uma aresta de corte está presente.

A escolha de redes de arquitetura RBF (função de base radial) como objeto de pesquisa justifica-se pelo fato de que, segundo Haykin (1994), tais redes são de projeto e treinamento mais simples do que os correspondentes para redes MLP (perceptron multi-camada) e são capazes de emular o desempenho daquele tipo de rede. Justifica-se também pelo fato de que poucos estudos utilizam tal tipo de rede para tarefas de predição em operações de usinagem. Sick (2002), realizou extensa revisão bibliográfica sobre o tema da predição da vida de ferramentas em processos de torneamento envolvendo redes neurais. Sua pesquisa analisou 138 artigos sobre o tema. Em somente seis dos trabalhos as redes RBF foram utilizadas. Outra motivação é o fato de que não foi possível encontrar na literatura trabalho que envolvesse uma busca sistemática por uma arquitetura otimizada para redes RBF aplicadas à predição da vida da ferramenta e acabamento superficial.

1.4 Limitações

Este trabalho tem como foco a predição da vida da ferramenta para processo de torneamento do aço SAE/ABNT 52100 – 55 HRC com ferramenta de cerâmica mista. Serão estudadas somente redes neurais de arquitetura RBF (Função de Base Radial). As conclusões obtidas por este trabalho são aplicáveis apenas aos materiais e arquitetura de rede utilizados.

Os dados utilizados para treinamento e teste das redes neurais serão obtidos através de um modelo polinomial quadrático completo, obtido e validado em experimento planejado e executado segundo a Metodologia de Superfície de Resposta. Isto se deve ao fato de que não

foi possível encontrar na literatura um banco de dados de vida da ferramenta e da rugosidade com dimensões e características adequadas para basear-se o estudo pretendido.

1.5 Problemas de Pesquisa

Ainda que diversos autores tenham abordado o tema da predição da vida de ferramentas e do acabamento superficial, em poucos se nota a utilização de redes de arquitetura RBF para tal predição. Também não se encontra na literatura um planejamento experimental com o objetivo de buscar uma arquitetura de rede RBF que apresente alto desempenho na tarefa de predição.

Dessa forma, este trabalho se propõe a responder às seguintes questões:

- a) Redes neurais de arquitetura RBF são adequadas para a tarefa de predição da vida de ferramentas e da rugosidade da peça?
- b) Quais seriam as configurações de redes RBF com alto desempenho para tais tarefas?
- c) Como o tamanho do conjunto de treinamento das redes influencia seu desempenho?
- d) Qual seria o número mínimo de casos de treinamento necessário para se obter boas predições de uma rede?
- e) Caso as redes RBF mostrem-se adequadas para a tarefa proposta, existe alguma interação entre os fatores que compõem o projeto da rede neural RBF?

1.6 Proposições

A aplicabilidade das redes de arquitetura RBF será investigada por meio de experimentos planejados segundo a metodologia de Projeto de Experimentos (DOE). Experimentos com conjuntos de treinamento de diferentes tamanhos serão conduzidos para atingir o objetivo de investigar-se o efeito do tamanho dos ditos conjuntos sobre o desempenho das redes.

O projeto de cada rede neural será desenvolvido e detalhado. As redes projetadas serão simuladas com a utilização de ferramenta computacional (Statistica 7.0®) e seu desempenho avaliado através do emprego de ferramentas estatísticas aplicadas aos resultados obtidos. Para atestar a qualidade da predição, os resultados serão comparados com os resultados obtidos por um modelo de otimização linear aplicado à mesma tarefa.

É esperado obter-se os parâmetros de redes neurais capazes de prever com exatidão a vida da ferramenta e a rugosidade para o processo em estudo. Espera-se também que as Redes Neurais apresentem previsão melhor do que aquela obtida por métodos de otimização linear, uma vez que os modelos de vida da ferramenta e da rugosidade são tipicamente não lineares.

1.7 Metodologia de Pesquisa

A metodologia de pesquisa utilizada será a pesquisa experimental. As técnicas utilizadas serão a simulação computacional e o projeto de experimentos.

A pesquisa experimental, segundo Bryman (1989) permite o estabelecimento de relações de causalidade, possibilitam ao pesquisador manter o controle sobre as variáveis independentes e é dotada de validade interna. Tais características vêm de encontro aos objetivos desta dissertação.

O trabalho a ser realizado constitui-se em um tipo de meta-modelagem, na qual a rede neural será utilizada para modelar a função da vida da ferramenta e da rugosidade, segundo dados obtidos de um modelo já existente. Segundo Bertrand e Fransoo (2002), modelos quantitativos são baseados em um conjunto de variáveis que podem assumir valores dentro de uma faixa especificada, e para as quais relações quantitativas e causais entre as mesmas são definidas.

Segundo Bertrand e Fransoo (2002), um passo muito importante em pesquisa com o uso de simulação é a elaboração do planejamento experimental. Todos os fatores dentro do modelo científico que possam ter impacto na solução devem ser identificados e simulados dentro de uma faixa de valores suficientemente larga. Além disso, segundo os mesmos autores, o número de fatores considerados deve ser mantido tão baixo quanto possível, a fim de propiciar a realização de uma simulação eficiente e análise efetiva dos dados obtidos.

O projeto de experimentos (DOE) consiste em técnicas que objetivam planejar experimentos capazes de gerar dados apropriados para análises estatísticas que resultem 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 (MONTGOMERY, 2001).

1.8 Variáveis de Estudo

1.8.1 Dependentes (Y)

A grandeza de saída utilizada como medida do desempenho e para a comparação da influência dos diferentes fatores no desempenho das redes foi a Razão de Desvios Padrão da fase de testes das redes neurais. Em um problema de regressão, a Razão de Desvios Padrão (do inglês *Standard Deviation Ratio* ou *S. D. Ratio*) é definida como sendo a relação entre o desvio padrão dos resíduos da predição pelo desvio padrão dos dados experimentalmente obtidos. Assim, quanto mais próximo de zero o valor da Razão de Desvios Padrão, melhor a capacidade de predição da rede neural. A Razão de Desvios Padrão corresponde a 1 menos a variância explicada pelo modelo.

1.8.2 Independentes (X)

As variáveis independentes a serem empregadas como fatores no planejamento experimental correspondem aos parâmetros de modelagem da rede neural de arquitetura RBF e são as seguintes:

- a) Número de unidades radiais;
- b) Algoritmo para determinação de centros da função radial;
- c) Algoritmo para determinação do parâmetro de largura da função radial;

Todas elas correspondem a fatores discretos, cujos níveis serão variados seguindo o planejamento para a realização dos experimentos. O número de unidades radiais corresponderá a uma fração do número de casos no conjunto de treinamento.

Serão testados dois algoritmos para determinação de centros da função radial: o algoritmo de sub-amostragem e o algoritmo das K-Médias (Bishop, 1995).

Em relação ao algoritmo para determinação do parâmetro de largura da função radial, serão testados dois algoritmos: o algoritmo Isotrópico (Haykin, 1994) e o algoritmo dos K vizinhos mais próximos (Bishop, 1995). O algoritmo Isotrópico é caracterizado por um parâmetro chamado fator de escala do desvio e o algoritmo dos K vizinhos mais próximos é caracterizado pelo parâmetro K. Os dois poderão assumir valores discretos.

1.9 Estrutura do Trabalho

Este trabalho apresenta-se dividido em oito capítulos. Este primeiro trata das considerações iniciais sobre o tema, objetivos, justificativas, delimitação do tema, problemas de pesquisa, metodologia e definição das variáveis de trabalho.

O capítulo 2 traz uma revisão bibliográfica sobre usinagem e torneamento. São definidos os termos Vida da Ferramenta e Rugosidade da Peça, cujas predições são o tema deste trabalho. São discutidos brevemente os mecanismos formadores do desgaste e da rugosidade superficial em operações de torneamento.

No capítulo 3 encontra-se uma revisão bibliográfica sobre redes neurais artificiais. Nele, discutem-se as principais características que tornam redes neurais interessantes para o problema da predição. Elabora-se um breve panorama da evolução das redes e são discutidas as características e propriedades das redes de arquitetura MLP (perceptron multi-camada). Tal discussão foi incluída por tratar-se do modelo mais didático e mais amplamente utilizado em aplicações de predição ligadas à usinagem. Em seguida são apresentadas funções de base radial e redes neurais de arquitetura RBF. O capítulo encerra-se com uma ampla revisão sobre as principais estratégias e trabalhos relacionados à utilização de redes neurais aplicadas à tarefa de predição da vida da ferramenta e do acabamento superficial.

O capítulo 4 apresenta uma revisão sobre a metodologia de projeto de experimentos. São apresentadas as bases históricas da metodologia, os principais conceitos e é apresentada a fundamentação teórica do projeto de experimentos (DOE). Os arranjos ortogonais de Taguchi são trabalhados, bem como as técnicas de fatorial completo. As técnicas para análise dos resultados obtidos por meio de experimentos planejados são apresentadas.

O capítulo 5 discute o planejamento experimental. Os cuidados experimentais para obtenção de resultados estatisticamente válidos são destacados. Descrevem-se os fatores e níveis empregados em cada experimento. A geração dos dados utilizados para treinamento e testes das redes neurais é discutida. Aborda-se a seqüência na qual os experimentos são realizados, as decisões tomadas durante o processo e suas justificativas. São descritos os arranjos ortogonais e os arranjos fatoriais empregados nos experimentos. O capítulo encerra-se com uma descrição do ambiente utilizado para os experimentos e uma descrição de sua execução e da tabulação dos dados obtidos.

O capítulo 6 traz os resultados obtidos para vida da ferramenta. O capítulo 7 traz os resultados obtidos para rugosidade da peça. Nos dois capítulos são apresentados os resultados

dos experimentos com arranjos ortogonais de Taguchi, os resultados dos experimentos exploratórios e definitivos com fatoriais completos. A análise detalhada e a avaliação dos resultados também são apresentadas.

O capítulo 8 contém as conclusões do trabalho. À luz da literatura pesquisada e com base nos resultados da análise realizada sobre os dados obtidos dos experimentos planejados são tecidas conclusões sobre as questões propostas. Aponta-se também para direções que podem ser foco de trabalhos que dêem continuidade a este estudo.

O apêndice A contém a complementação da argumentação principal. Ele traz a definição e discussão de elementos de estatística e modelagem utilizados no corpo do trabalho.

2 VIDA DA FERRAMENTA E RUGOSIDADE

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. (2005), a retirada do cavaco envolve o cisalhamento concentrado ao longo de um plano chamado de plano de cisalhamento. A figura 2.1 ilustra o conceito de retirada de material.

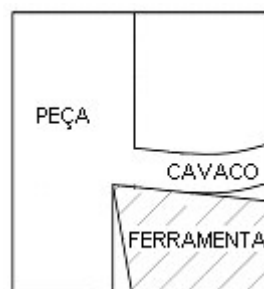


Figura 2.1 – Retirada de material em processo de usinagem. Adaptado de Trent (2000)

Segundo Trent (2000), o torneamento é a operação de usinagem mais comumente empregada em trabalhos de corte de metal. Amorim (2002), afirma que o corte em operações de torneamento é normalmente contínuo e pode ser descrito em coordenadas cilíndricas. Ocorrem simultaneamente movimentos da peça (rotação) e da ferramenta (translação), como ilustrado na figura 2.2. 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, avanço e profundidade de usinagem.

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. Pode ser definido também como o percurso realizado pelo movimento de avanço em cada volta ou curso da ferramenta (DINIZ et al., 2005).

Segundo Diniz et al. (2005), a velocidade de corte é a velocidade tangencial instantânea resultante da rotação da ferramenta em torno da peça 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 a_p é a profundidade ou largura de penetração da ferramenta em relação à peça, medida perpendicularmente ao plano de trabalho (DINIZ et al., 2005).

Em relação à sua otimização, Cus e Zuperl (2006) afirmam que a operação de torneamento é problema que possui três objetivos conflitantes: taxa de produção, custo de operação e qualidade de usinagem. Ainda segundo os autores, os três objetivos mencionados podem ser expressos como funções da velocidade de corte, do avanço e da profundidade de usinagem.

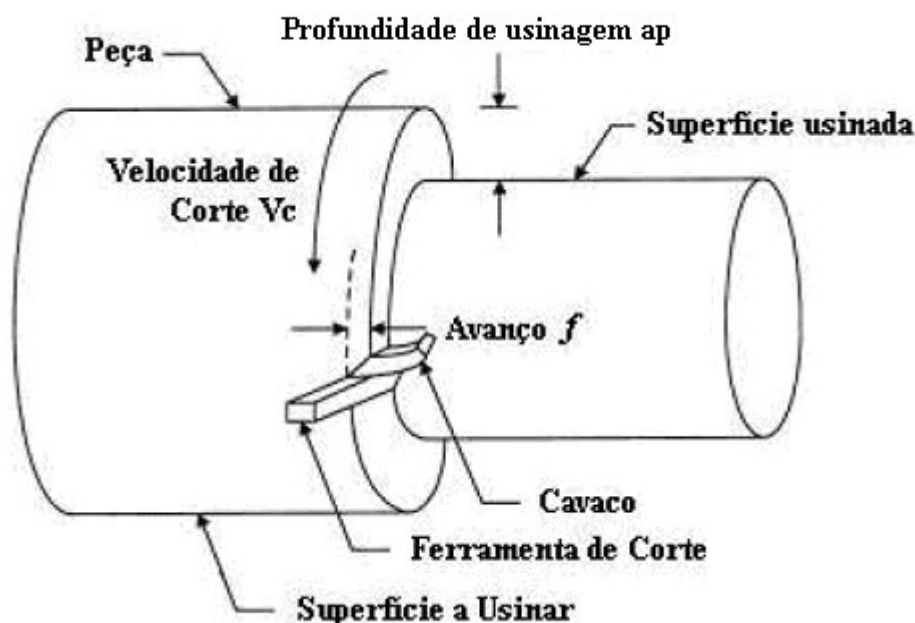


Figura 2.2 – Elementos em operação de torneamento. Adaptado de Amorim (2002)

2.2 Desgaste e Vida da Ferramenta

O fim de vida da ferramenta é geralmente causado pelo aumento dos desgastes nela ocorridos. Segundo Amorim (2002), ferramentas podem ser substituídas por duas causas principais: avarias ou desgaste excessivo.

Os principais desgastes e avarias que limitam a vida da ferramenta são listados a seguir (DINIZ et al., 2005; TRENT, 2000; SHAW, 2004):

- a) **Desgaste de flanco:** é o tipo de desgaste mais comum. Ocorre na superfície de folga da ferramenta e é causado pelo atrito ou abrasão entre a ferramenta e a peça. O limite máximo permitido desse desgaste está ligado ao acabamento superficial desejado para a peça, pois quanto maior o seu valor médio maior será a deterioração da aresta de corte. Com isso piores serão o acabamento superficial e a tolerância dimensional da peça;
- b) **Desgaste de cratera:** acontece na superfície de saída da ferramenta. Pode ser causado por difusão, oxidação e adesão. Seu aumento pode provocar a quebra da ferramenta;
- c) **Desgaste de entalhe:** segundo Dos Santos (2001), pode ocorrer na aresta principal de corte, na extremidade do contato cavaco-ferramenta e na superfície de saída da ferramenta. Ocorre juntamente com o desgaste de cratera e flanco e é caracterizado por sulcos profundos que são provocados principalmente pelos mecanismos de desgaste por oxidação e abrasão;
- d) **Deformação plástica da aresta de corte:** A pressão aplicada à ponta da ferramenta, somada à alta temperatura, gera a deformação plástica da aresta de corte. Provoca deficiência do controle de cavacos e deterioração do acabamento superficial da peça. Dependendo da sua intensidade, pode levar à quebra da aresta de corte.

Outros tipos de avarias tais como trincas, lascamentos e quebras, podem surgir durante o torneamento de aços endurecidos. Decorrem das condições da operação, que incluem rigidez da máquina, resistência da ferramenta, interrupção abrupta do corte, irregularidades na peça, dentre outras causas.

Entre os principais mecanismos causadores do desgaste incluem-se a aresta postiça de corte, a aderência, a abrasão mecânica, a difusão e a oxidação. A figura 2.3 ilustra uma situação de desgaste em ferramenta.

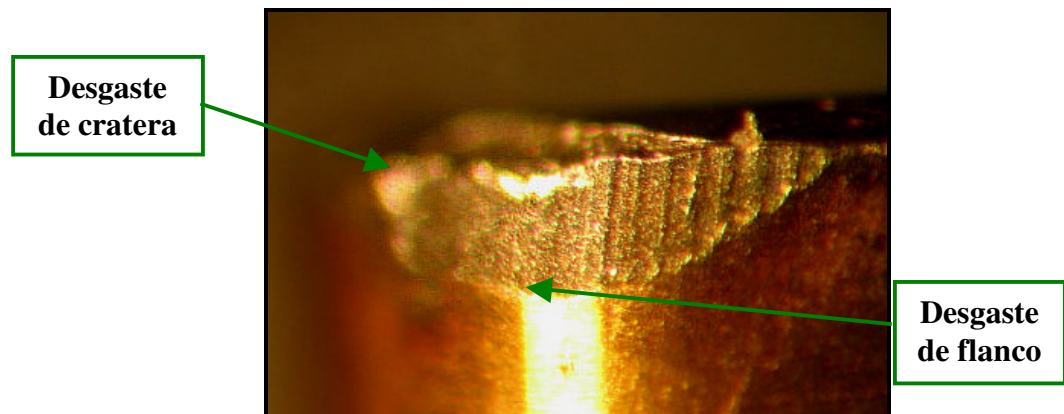


Figura 2.3 – Desgastes em ferramenta após atingir critério de fim de vida. Adaptado de Pereira (2006)

Além dos desgastes, ferramentas também podem sofrer avarias. Avarias são potenciais causadoras de prejuízos, pois podem levar à quebra súbita da aresta de corte da ferramenta, também conhecida como falha catastrófica. A figura 2.4 ilustra tal situação. Tais avarias podem ser provocadas pelos próprios desgastes sofridos pela ferramenta ou por razões tais como abrasão, aderência, difusão, oxidação, choques térmicos ou mecânicos, defeitos microestruturais das ferramentas ou deformação plástica da aresta de corte (DINIZ ET AL., 2005).

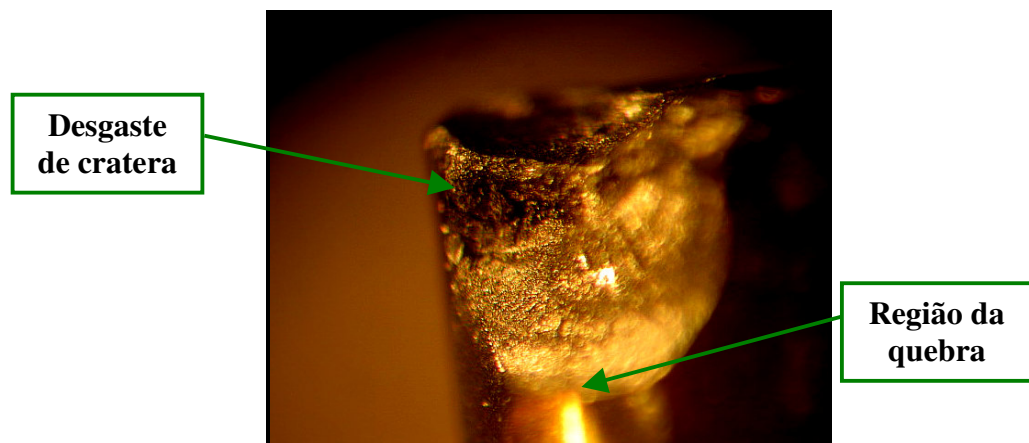


Figura 2.4 – Quebra de ferramenta. Adaptado de Pereira (2006)

Segundo Alajmi et al. (2005), os desgastes de ferramentas e sua relação com os parâmetros de corte são de grande importância no estudo dos processos de usinagem. A detecção e predição dos desgastes antes que estes provoquem danos à superfície usinada é de valia expressiva para evitar-se a perda da peça em trabalho e danos ao suporte da ferramenta e à torre de fixação da máquina, com conseqüente perda de produtividade. Os parâmetros de corte são de grande importância na determinação dos desgastes sofridos pela ferramenta. Diniz et al. (2005), afirmam que a velocidade de corte é o parâmetro mais influente no

desgaste, seguido pelo avanço e pela profundidade de usinagem, nessa ordem. Apesar disso, a predição do desgaste de ferramentas com o uso de modelos teóricos é extremamente difícil, segundo Lee et al (1996), em razão das características de não-linearidade dos mecanismos de desgaste.

A vida da ferramenta é definida, segundo Diniz et al. (2005), como o tempo em que a mesma trabalha efetivamente até perder sua capacidade de corte, dentro de algum critério estabelecido. Ferraresi (1977) a define como o tempo em que a ferramenta de corte trabalha, efetivamente, sem perder o corte ou até que atinja um critério de fim de vida previamente estabelecido. Os critérios enumerados pela norma ISO 3685 (1993) para o fim de vida de ferramentas de metal duro, aço rápido e cerâmicas são:

- Desgaste de flanco médio, $VBB = 0,3\text{mm}$;
- Desgaste de flanco máximo, $VBMax = 0,6\text{mm}$;
- Profundidade máxima de cratera, $KT = 0,06 + 0,3f$, onde f é o avanço em mm/v;
- Desgaste de entalhe, $VN = 1,0\text{mm}$;
- Falha catastrófica.

O desgaste de flanco é o que apresenta maior risco de danos à peça e que exige mais potência de corte, razões pelas quais é o mais utilizado como critério para determinação do fim de vida da ferramenta (AMORIM, 2002).

Também influenciam na determinação da vida da ferramenta, conforme Alajmi et al (2005), o tipo e dureza dos materiais que constituem a ferramenta e a superfície usinada. A relação entre vida da ferramenta e as variáveis independentes do processo de usinagem pode ser expressa pela fórmula expandida da equação de Taylor, mostrada na equação (1) (Choudury e El-Baradie, 1998; Shaw, 2004).

$$T = C \left(Vc^l f^m ap^n \right) \quad (1)$$

Na equação (1), T é a vida da ferramenta em minutos, Vc é a velocidade de corte em m/min, f é o avanço, dado em mm/v, e ap é a profundidade de usinagem, medida em mm. C , l , m e n são constantes características do trio ferramenta, peça e máquina.

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 E KARPAT, 2005). Benardos e 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 superficial se refere a desvios da superfície nominal de terceira até sexta ordem. Segundo Benardos e Vosniakos (2003) desvios de terceira e quarta ordens são caracterizados por depressões, quebras e dilapidações, 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 superfície, de acordo com a figura 2.5 (BENARDOS E VOSNIAKOS, 2002).

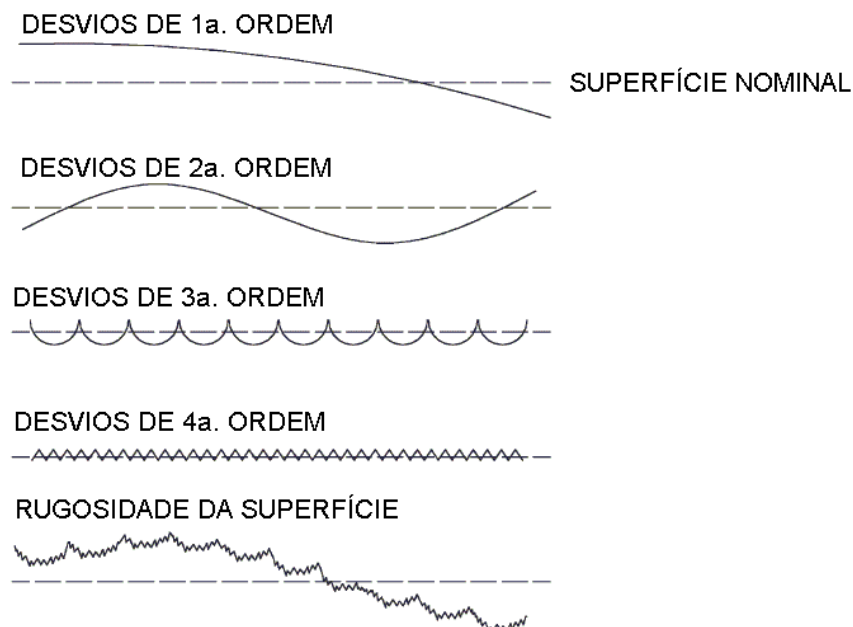


Figura 2.5 – Desvios no formato de superfícies. Adaptado de DIN 4670 (1982)

A rugosidade superficial é caracterizada pela forma como é feita sua medição. 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 (2) (ISO 4287/1, 1997).

$$R_a = \frac{1}{l_m} \int_0^{l_m} |y(x)| dx \quad (2)$$

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 superficial é definida pela equação (3) (ISO 4287/1, 1997).

$$R_a = \frac{1}{n} \sum_{i=1}^n |y_i| \quad (3)$$

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. O conceito é ilustrado na figura 2.6.

ferramenta na determinação da rugosidade pode ainda ser encontrada em diversos autores, como Shaw (2004), Özel e Karpaz (2005), Erzurumlu e Oktem (2006), Davim (2001).

Segundo Diniz et al. (2005), operadores em geral não dispõem de meios de aferir 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 e Vosniakos (2002) afirmam que há necessidade, por parte da indústria, de ferramentas que permitam estimar a rugosidade superficial antes da usinagem da peça, e que sejam, ao mesmo tempo, facilmente utilizáveis em ambientes de chão de fábrica.

Benardos e Vosniakos (2003) listam quatro linhas de pesquisas dedicadas à modelagem da rugosidade superficial: 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 e Vosniakos (2002) afirmam que muitos não são acurados o suficiente, e se aplicam a uma limitada gama de processos e condições de corte. Amorim (2002) afirma que o modelo teórico da rugosidade baseado na geometria da peça apresenta boa aproximação para a rugosidade apenas no início da vida da ferramenta. Isso se deve, segundo o autor, ao fato do modelo não levar em conta o desgaste da aresta de corte ao longo do tempo. Oliveira (2003), em um trabalho experimental relacionado à influência dos fluidos de corte nas condições de usinagem, observou que os valores experimentais observados foram superiores aos valores teóricos desde o início da operação.

Um modelo teórico para a rugosidade da peça foi estudado por Özel e Karpaz (2005). Ele não leva em conta imperfeições no processo, tais como vibração da ferramenta e aderência dos cavacos. Os autores concluem que a rugosidade superficial em experimentos onde se utiliza baixa taxa de avanço não corresponde à prevista pelo modelo teórico (ÖZEL e KARPAT, 2005).

Em relação à investigação experimental, observa-se o trabalho de Bonifácio e Diniz (1994), os quais realizaram experimentos nos quais se estuda a influência da velocidade de corte, da profundidade de usinagem e do avanço nos parâmetros de usinagem R_a e R_{max} para torneamento de um aço 4340. O estudo concluiu que os fatores citados têm, em diferentes níveis, influência na determinação da rugosidade da peça.

Um exemplo dos trabalhos envolvendo técnicas de projeto de experimento pode ser encontrado em Davim (2001), o qual realizou experimentos com a utilização de técnicas de

Taguchi, 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 torneamento. Seus resultados indicam uma maior influência da velocidade de corte, seguida pelo avanço.

Ainda que não haja um modelo teórico único que torne possível a predição da rugosidade, a literatura aponta no sentido de que os parâmetros de corte são determinantes na predição. Cus e Zuperl (2006) e Fang e Safi-Jahanshahi (1997), sugeriram modelos empíricos lineares e exponenciais para a rugosidade de superfície como funções da velocidade de corte (V_c), avanço (f) e profundidade de usinagem (ap). Tal modelo é mostrado na equação (5).

$$R_a = c_o (V_c^{c_1} f^{c_2} ap^{c_3}) \quad (5)$$

Na equação (5), 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 Considerações Finais

Diante do exposto algumas observações podem ser feitas. A primeira é de que a predição da vida da ferramenta e da rugosidade da peça é tema de pesquisa relevante, que tem atraído a atenção de muitos pesquisadores. A segunda é que a predição de tais grandezas é problema complexo. Os esforços na área estão distribuídos em diversas linhas de pesquisa. O próximo capítulo mostra que técnicas de inteligência artificial tais como as redes neurais, em conjunto com técnicas de planejamento experimental, apresentam-se como áreas de pesquisa promissoras para a tarefa de predição.

3 REDES NEURAIS ARTIFICIAIS

3.1 Considerações Iniciais

Segundo Haykin (1994), 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.

Na definição de Braga et al. (2000), 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. Procuram reproduzir as funções das redes neurais biológicas, através da imitação de seu comportamento básico e de sua dinâmica. Ambos os sistemas são formados por unidades de computação paralela e distribuída que trocam informação por meio de conexões sinápticas.

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. As funções de ativação são geralmente diferenciáveis, fato este que é crítico para o sucesso dos algoritmos de treinamento. Diversos tipos de funções de ativação são utilizados (HAYKIN, 1994).

Redes neurais têm, em geral, três camadas distintas: a camada de entrada, uma ou mais camadas escondidas e uma camada de saída. Dados são apresentados aos neurônios da camada de entrada da rede, processados por uma ou mais camadas escondidas e os resultados são gerados na camada de saída da rede. A camada de entrada de uma rede neural não é realmente composta de neurônios artificiais. Tais unidades, em geral, servem somente como portas de entrada para dados oriundos do meio externo. 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.

Redes neurais constituem-se em uma poderosa técnica de modelagem. Segundo Ezugwu et al. (2005), elas permitem a modelagem de relações cuja descrição com a utilização de modelos físicos seria difícil. 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.

As redes 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. O teorema de Kolmogorov-Nielsen garante que, para uma função contínua qualquer, existe uma implementação exata com uma rede neural de três camadas. Segundo Kovacs (1996), no entanto, este teorema apenas prova a existência de tal rede, não havendo método formal que permita chegar à configuração paramétrica da mesma.

Em anos recentes, tem havido grande interesse pela utilização de redes neurais. Elas têm sido aplicadas em campos tão diversos quanto mercado de ações, medicina, engenharia, dentre outros. Quando aplicadas à predição em processo de usinagem, as redes neurais também têm se mostrado úteis. Segundo Cus e Zuperl (2006), redes de arquitetura MLP provaram ser excelentes aproximadores universais para funções não lineares. Alajmi et al (2005) afirmam que elas têm sido empregadas com sucesso em processos de usinagem para a predição de desgaste de ferramentas e na simulação da natureza complexa dos processos de corte.

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.

3.2 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. Elas simulam a micro-estrutura do cérebro humano. O cérebro é composto por bilhões de neurônios interconectados. Cada neurônio é formado por uma estrutura de entrada (os dendritos), um núcleo celular e uma estrutura de saída (os axônios). Axônios de uma célula conectam-se aos dendritos de outra célula por meio de sinapses. A tarefa dos neurônios é a propagação de sinais eletroquímicos. Um neurônio é ativado quando os sinais eletroquímicos em seus dendritos ultrapassam um certo limiar de disparo. Quando ativado, um neurônio dispara um sinal eletroquímico por meio das suas sinapses para outros neurônios, os quais podem, por sua vez, ser ativados (KOVACS, 1996).

A intensidade do sinal recebido por um neurônio depende da eficácia das suas sinapses de entrada. Cada sinapse contém um espaçamento entre seus terminais, no qual neurotransmissores são posicionados para transmitir um sinal de um neurônio a outro. Segundo Hebb (1949), o aprendizado consiste principalmente em alterar a intensidade das sinapses. 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 e Pitts, em 1943. Uma ilustração do conceito de neurônios McCulloch e Pitts é mostrada na figura 3.1.

O neurônio de McCulloch e 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. Pesos 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 e Pitts tem uma função linear binária, e dispara quando o resultado do produto escalar ultrapassa o limiar (BRAGA et al., 2000).

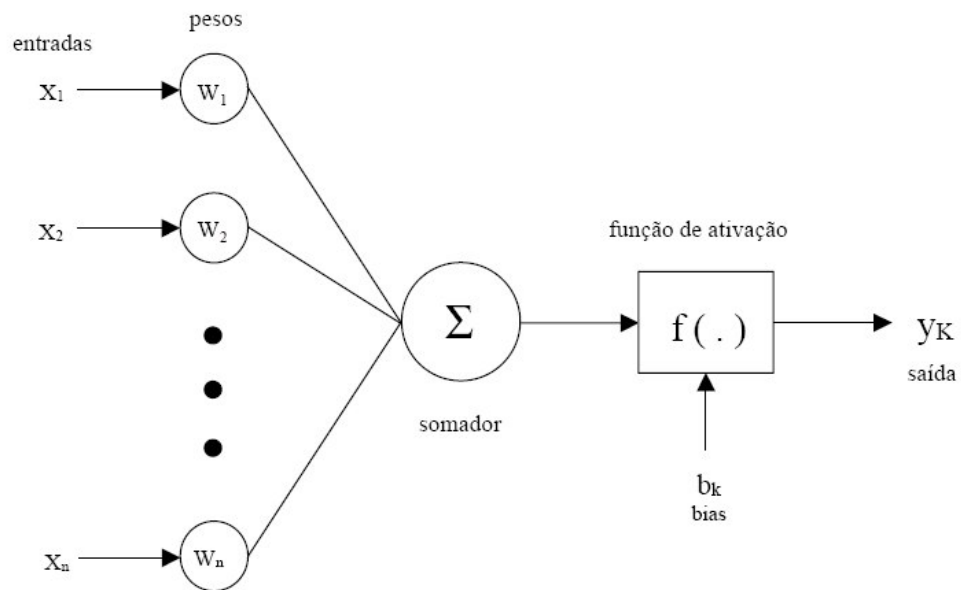


Figura 3.1 – O neurônio de McCulloch e Pitts. Adaptado de Dos Santos (2001)

Ainda que um único neurônio seja de limitada utilidade, uma rede formada por tais unidades pode ser treinada para realizar tarefas mais complexas. Rosenblatt, em 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. Em 1969, porém, Minsky e Papert 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 escondidas também contribuiu para a estagnação nas pesquisas com redes neurais durante aquele período.

O trabalho de Hopfield, em 1982, sobre memórias associativas, iniciou um novo período de investigação científica sobre o tema. Em 1986, os trabalhos de Rumelhart, Hinton e Williams deram origem ao algoritmo de retro-propagação, removendo a limitação apontada anos antes por Minsky e Papert. Com efeito, enquanto uma rede sem uma camada oculta (como o perceptron) pode resolver somente problemas linearmente separáveis, redes com camadas escondidas superam esta limitação.

Broomhead and 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 (*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, 1994). Tais arquiteturas não serão discutidas neste trabalho, cujo foco será a utilização de redes RBF. Apenas breve discussão sobre redes MLP será desenvolvida posto que correspondem a uma arquitetura tradicional e largamente empregada no campo de redes neurais. A compreensão do funcionamento de tal arquitetura é útil para a compreensão do funcionamento de redes RBF, donde se justifica a revisão sobre o assunto.

3.3 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). É tarefa principal de uma rede neural aprender um modelo do ambiente no qual se encontra inserida e manter esse modelo suficientemente consistente com o mundo real tal que possa cumprir com exatidão e precisão a tarefa para a qual foi projetada (HAYKIN, 1994).

Por fase de 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, 1994). 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 certo 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.

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 (1994), 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 é possível somente quando existe uma certa redundância dos dados de entrada, posto que, sem redundância, é impossível inferir padrões ou características dos mesmos (Braga et al., 2000).

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 a mesma possa aprender uma determinada função. Conforme mencionado, o efeito dos algoritmos de treinamento é o ajuste dos valores dos pesos das sinapses e valores limiares de disparo. O ajuste é feito de tal 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 valor RMS dos erros de predição, ou o valor da Razão de Desvios Padrão. 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, 1994).

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 a mesma 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, 1994).

Um problema que pode ocorrer no emprego de redes é o chamado sobre-ajuste (*overfitting*). Esse fenômeno se caracteriza por alto desempenho na fase de treinamento e desempenho pobre na fase de testes, ou durante a operação. É como se a rede houvesse

memorizado as respostas para os casos de treinamento, mas fosse 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, do que o requerido pelo problema em questão. O conceito é exemplificado pela figura 3.2, 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.

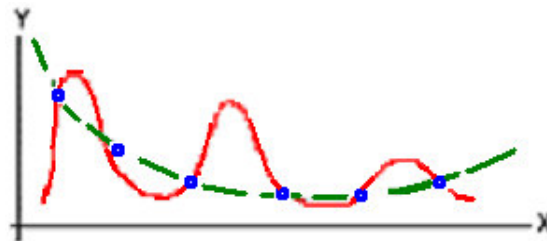


Figura 3.2 – Conceito de sobre-ajuste de redes. Adaptado do Manual Statistica (2005)

Um número elevado de neurônios implica em um número elevado de pesos e limiares, que são parâmetros livres da rede. Uma tal rede pode modelar uma função mais complexa do que a função que corresponde realmente ao fenômeno que se quer modelar, levando ao sobre-ajuste. Uma rede sobre-ajustada pode apresentar baixíssimo erro durante o treinamento e erro elevado em operação real. Deduz-se então que a determinação adequada do número de neurônios é essencial na busca por uma rede com capacidade de generalização adequada ao problema abordado.

Um outro problema que leva ao mesmo efeito é a ocorrência do sobre-aprendizado (*overlearning*), 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, cujo objetivo é interromper o treinamento em um ponto tal que evite a ocorrência do sobre-aprendizado (HAYKIN, 1994).

3.4 Projeto de Redes Neurais

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 (KHAW et al., 1995).

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), 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 (*crosstalk*). Este é um fenômeno no qual os neurônios apresentam dificuldades no aprendizado, cuja origem é o fato dos neurônios tentarem modelar duas funções ao invés de uma (MANUAL STATISTICA, 2005). Em grande parte dos casos observados na literatura, as redes neurais apresentam uma única saída.

Em relação ao número de casos de teste, 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 (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, 1994).

A exatidão obtida pela rede depende do número de neurônios na camada oculta. Os neurônios localizados na camada oculta funcionam como detectores de características. A maioria dos trabalhos encontrados na literatura determina este número empiricamente. 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. Não se deve usar um número excessivo de neurônios na camada oculta, pois isso pode causar o já mencionado problema de sobre-ajuste. O oposto também é válido: não se deve utilizar um número muito baixo de neurônios na camada oculta (BRAGA et al., 2000).

Segundo Kim e Yum (2003), o processo convencional para o projeto de redes neurais é um processo de tentativa e erro. Em geral, projetistas de redes neurais variam a arquitetura de rede, o algoritmo de treinamento utilizado ou os parâmetros da rede (número de neurônios, parâmetros do algoritmo de treinamento) um número de vezes em busca de redes que apresentem bom desempenho na tarefa proposta. Os autores afirmam que tal prática é claramente ineficiente, consome muito tempo e pode falhar na identificação de uma arquitetura aceitável em tempo razoável. Por essa razão, alguns autores têm buscado métodos que racionalizem o projeto de redes neurais.

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 (1995), 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 Estatística, 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 como o Média/Desvio Padrão (Manual Estatística, 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 (1994) 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. O conjunto de testes, por sua vez, não entra no treinamento da rede. Seu uso é recomendável como um teste final e independente para estimar a capacidade de generalização da rede. Os dados pertencentes a este conjunto não devem entrar na fase de treinamento nem na fase de seleção.

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. É recomendado ainda que a ordem na qual os exemplos são apresentados à rede seja aleatória (HAYKIN, 1994). De acordo com Braga et al. (2000), outras técnicas para evitar a ocorrência do sobre-aprendizado incluem a ‘poda’ de redes (*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, 1994).

É útil levar-se em conta que dados de treinamento são relacionados a uma dada operação. 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 (1994) realizou 20 testes independentes com cada arquitetura de rede estudada.

3.5 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, ou *feed-forward*. Isso significa que neurônios de uma camada são ligados unicamente a neurônios da camada seguinte (HAYKIN, 1994). A figura 3.3 ilustra o conceito de redes MLP.

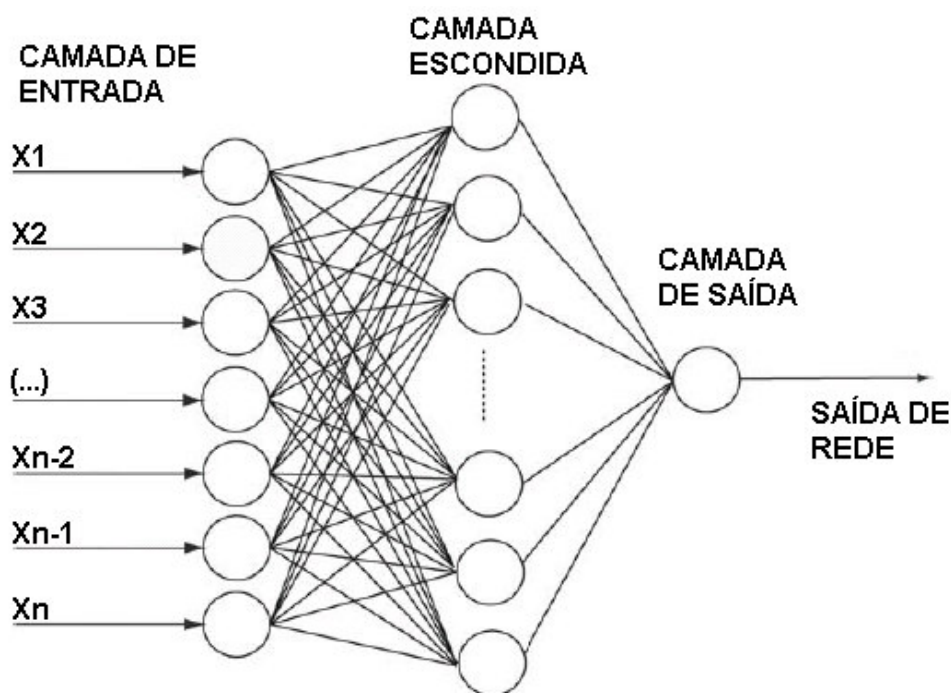


Figura 3.3 – Rede MLP. Adaptado de Özel e Karpaz (2005)

Segundo Bishop (1995), tal modelo de rede é capaz de modelar funções complexas. O número de camadas escondidas 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 escondidas podem, segundo o mesmo autor, aproximar qualquer função.

3.6 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, Hinton e Williams, 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, 1994).

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, 1994).

O treinamento ocorre, portanto, em duas fases: a fase direta (*forward*), na qual se define a saída da rede para um determinado padrão de entrada ou conjunto de padrões de

entrada; e a fase reversa (*backward*), que utiliza a diferença entre a saída de rede e a 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 (1994), 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).

3.7 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 (6) (BISHOP, 1995).

$$h(\vec{x}^n) = \vec{t}, \quad n = 1, \dots, N. \quad (6)$$

Segundo Bishop (1995), 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 (7) (BISHOP, 1995).

$$\phi(\|\vec{x} - \vec{x}^n\|) \quad (7)$$

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 \bar{x} e \bar{x}^n . Uma das funções de base radial mais utilizadas é a função gaussiana dada pela equação (8) (BISHOP, 1995).

$$\phi_j(\mu) = -\exp\left(\frac{\|\bar{x} - \bar{\mu}_j\|^2}{2\sigma^2}\right) \quad (8)$$

em que \bar{x} corresponde a um dos vetores de entrada \bar{x}^n , com elementos x_i . $\bar{\mu}_j$ 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 3.4. Na figura, a resposta de uma função de base radial é mostrada. Seu pico localiza-se no centro e decresce em função da distância a este.

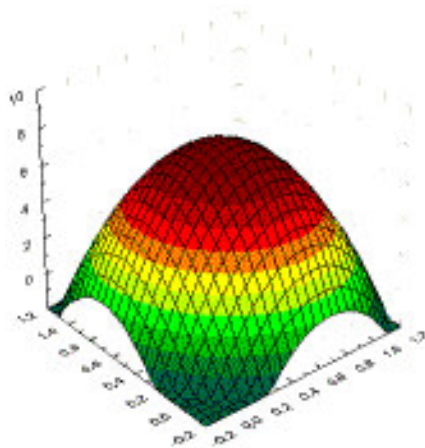


Figura 3.4 – Função de base radial. Fonte: Manual Statistica, 2005

Redes RBF (*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 and 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. Uma restrição importante é a de que redes RBF são incapazes de extrapolação (BISHOP, 1995).

A camada de saída de redes RBF geralmente é linear. Isso significa que a função de ativação dos neurônios daquela camada é unitária. A saída de um neurônio da camada de saída nestas condições é simplesmente o produto escalar do vetor de saída da camada oculta pelo vetor de pesos da unidade (BRAGA et al., 2000).

A função radial em uso em redes RBF é, geralmente, uma função gaussiana, do tipo mostrado na equação (8), na qual o vetor \bar{x} corresponde ao vetor de entrada da unidade radial e $\bar{\mu}_j$ 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 (9) (BISHOP, 1995).

$$y = \sum_{i=1}^k w_i \phi(\|\bar{x} - \bar{\mu}_i\|^2) + w_0 \quad (9)$$

em que \bar{x} e $\bar{\mu}_i$ são definidos como na equação (8), 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 (8), 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.

A figura 3.5 ilustra o conceito de uma 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 $\phi_1(\bar{x}), \dots, \phi_j(\bar{x}), \dots, \phi_m(\bar{x})$. 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 $y(\bar{x})$.

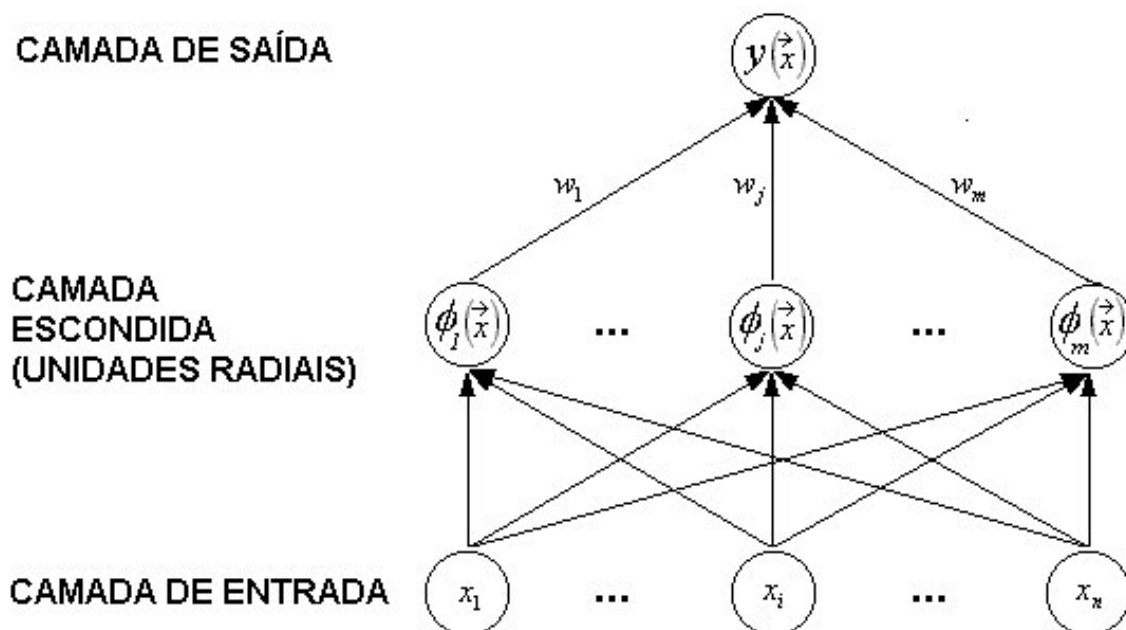


Figura 3.5 – Rede neural de arquitetura RBF. Adaptado de BaseGroup Lab (2005)

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. Tal modelo de rede é apropriado para desempenhar diversas tarefas, dentre as quais se inclui a aproximação de funções. 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, 1994).

3.8 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.

Uma segunda vantagem é o fato de que, segundo Sick (2002), seu tempo de treinamento é muito curto se comparado ao das redes MLP aplicadas à mesma tarefa. Este é um grande atrativo para o emprego de redes RBF, segundo o autor. Os trabalhos de Cus e Zuperl (2006) e de El-Mounayri et al. (2005) destacam a rapidez do treinamento de redes RBF como vantagem decisiva sobre as redes MLP. 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 (1995), 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, ou *overfitting*. Em relação aos parâmetros de largura das funções radiais, El-Mounayri et al (2005) afirmam que inexistente 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 os parâmetros de largura das mesmas. 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 (*Sub-sampling*) e o algoritmo K-Médias (*K-Means*). No método de Sub-amostragem, pontos de treinamento escolhidos aleatoriamente são copiados como centros das unidades radiais (Haykin, 1994). Sendo aleatoriamente selecionados, eles representam estatisticamente a distribuição dos dados de treinamento.

Já no algoritmo K-Médias (Bishop, 1995), 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 da mesma. 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 (*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 (10) (BISHOP, 1995).

$$\vec{w} = G_+ \vec{d} \quad (10)$$

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 (11) (Bishop, 1995).

$$G_+ = (G^T G)^{-1} G^T \quad (11)$$

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 .

3.9 Redes Neurais aplicadas à tarefa de predição

3.9.1 Considerações Gerais

Segundo Sick (2002), redes neurais artificiais são usadas com frequência na modelagem de dependências não-lineares envolvendo desgaste das ferramentas e condições de corte. Existem duas abordagens para a modelagem de vida da ferramenta com o auxílio de redes neurais: uma com a utilização de sensores e outra sem a utilização de sensores. Os métodos que não utilizam sensores foram os mais estudados até o momento. Redes do tipo MLP são as mais amplamente utilizadas na predição da vida da ferramenta, correspondendo a 70% do total de trabalhos (SICK, 2002). Segundo Benardos e Vosniakos (2003) redes neurais de arquitetura MLP (perceptron multi-camada) treinadas segundo o algoritmo de retro-propagação também são as mais comumente utilizadas em tarefas de predição da rugosidade da peça.

O potencial e os percalços para predizer a vida de ferramentas com o uso de redes neurais artificiais são temas já discutidos na literatura por mais de uma década. Ezugwu et al (1995) já apontavam para a dificuldade na determinação de parâmetros de projeto para a obtenção de bons resultados com redes MLP, e citava-se a falta de métodos apropriados que guiassem tal escolha. Ressaltava-se também a dificuldade proporcionada pela falta de casos de treinamento em quantidade suficiente e também para a escolha de um algoritmo adequado para o pré e pós-processamento dos dados. Em Peterson et al. (1995), já se observa um esforço no sentido de mensurar o desempenho de redes neurais MLP na aproximação de funções. Nota-se naquele trabalho a preocupação em reduzir os erros de predição e também com a seleção de uma arquitetura de rede que apresentasse respostas com variabilidade reduzida.

A literatura indica oportunidade de melhoria nos modelos de predição da vida da ferramenta e rugosidade em processos de torneamento implementados com a utilização de redes neurais artificiais. Em ampla revisão bibliográfica realizada por Sick (2002) envolvendo mais de cem estudos sobre o tema, o autor afirma que na maioria dos trabalhos os parâmetros de rede e métodos de treinamento são fixados arbitrariamente, sem investigação de alternativas que pudessem otimizar seu desempenho. Além disso, muitos trabalhos somente fazem breve menção a investigações realizadas em busca da melhor arquitetura, sem descrição de quaisquer detalhes do método usado. É comum apresentar-se somente a melhor rede encontrada. Em muitos estudos o número de entradas da rede é desnecessariamente elevado (SICK, 2002).

Sick (2002) lista práticas que são desejáveis para a boa qualidade de trabalhos científicos nesta área, tais como: o uso de conjuntos de treinamento e de teste distintos para as redes durante o experimento; a realização de repetições das corridas experimentais, para possibilitar a computação de médias e variâncias para estimação da exatidão e da variabilidade da rede e também para evitar a influência de efeitos ligados à iniciação aleatória dos pesos sobre os resultados; e, por fim, a aplicação de testes estatísticos para casos em que diferentes configurações são comparadas. Tais requisitos são importantes, segundo o mesmo autor, para prover embasamento estatístico que sustente as conclusões obtidas. Muito embora seu trabalho fosse orientado à predição da vida da ferramenta, as práticas listadas podem ser igualmente aplicadas aos estudos orientados à predição da rugosidade superficial.

Observa-se que muitos trabalhos que envolvem tarefas de predição em processos de usinagem com a utilização de redes neurais têm o duplo objetivo de estudar a vida da ferramenta e também a rugosidade superficial. Em muitos estudos, o foco é a predição do desgaste da ferramenta, ou a medição das forças de corte, ou algum outro parâmetro do processo que constitua uma indicação do fim de vida da ferramenta. Alguns estudos utilizam técnicas da metodologia de projeto de experimentos (DOE) como ferramenta para otimizar algum aspecto das redes neurais. Nas descrições a seguir, procurou-se agrupar os trabalhos pelo seu foco (vida, desgaste ou rugosidade). Também foram agrupados estudos envolvendo redes RBF, bem como aqueles que utilizam alguma técnica de DOE como ferramenta para otimização de redes.

3.9.2 Estudos com Redes Neurais aplicadas à predição

O acompanhamento das condições da ferramenta com utilização de redes neurais pode ser realizado através de monitoração (baseada em sensores) ou de predição (sem a utilização de sensores). Na monitoração, um sistema de aquisição de sinais é combinado a uma rede neural, que é treinada em tempo de execução para a tarefa. Na predição, a rede é treinada com dados históricos ou provenientes de um modelo. A rede é usada para prever vida ou rugosidade antes do início da operação, no momento em que sejam determinados os parâmetros do processo. A monitoração não é facilmente implementável, dada a complexidade da medida dos sinais durante o processo. Os custos envolvidos na montagem de um sistema de monitoração são bem mais elevados do que aqueles que não utilizam sensores (SICK, 2002).

A monitoração pode ser direta ou indireta (SICK, 2002). A monitoração direta baseia-se na utilização de sensores óticos, radioativos, de proximidade ou piezo-elétricos para medição do desgaste durante o processo. A monitoração indireta, segundo Özel e Karpat (2005), procura obter medidas de outros sinais que fornecem indicações de desgastes na ferramenta tais como forças de corte, vibração, condições de corte, dentre outros.

Axinte (2006), realizou um trabalho utilizando redes neurais probabilísticas (PNN) para classificação automática de avarias em ferramentas utilizadas em processos de brochamento. Seu trabalho monitorou a força de corte e utilizou a mesma como sinal para predição da condição da ferramenta.

Outro exemplo da utilização de redes em conjunto com sensores para a detecção do desgaste de ferramentas é o trabalho de Abu-Mahfouz (2003), que utilizou dados de vibração coletados de acelerômetros como entradas de redes para monitoração do desgaste em processo de furação.

A monitoração do desgaste de ferramentas pode se valer da correlação entre sinais provenientes de sensores e o desgaste da ferramenta. Liu e Altintas (1999) empregaram redes neurais de arquitetura MLP para monitoração baseada em sensores do desgaste em operações de torneamento. Os autores enumeram diversos sinais que podem ser utilizados com tal fim, dentre os quais: as forças de corte, as emissões acústicas, a corrente do motor e a vibração. Os autores ressaltam, porém, que há relação de dependência entre os sinais listados e os parâmetros de corte utilizados, tais como velocidade de corte, avanço e profundidade de usinagem. A mesma abordagem pode ser encontrada em Das et al. (1997).

Um exemplo do paradigma da utilização de monitoração para detecção do desgaste de ferramentas pode ser encontrado em Silva et al (1998). O objetivo do trabalho foi o de desenvolver um sistema de monitoração de desgaste no qual técnicas de processamento de sinais, redes neurais e sistemas especialistas fossem usados. As redes foram utilizadas para classificar padrões estatísticos e sinais do domínio da frequência oriundos da amostragem realizada por sensores acoplados à máquina. O estudo utilizou redes de arquitetura SOFM (mapas auto-organizáveis), e também redes do tipo ART2, dois tipos de rede associados aos métodos de aprendizado não supervisionado. Os autores concluíram que ambas as arquiteturas foram capazes de classificar com desempenho aceitável.

Um exemplo de monitoração indireta pode ser encontrado em Ghasempoor-Nobandgany (1998), no qual redes neurais estáticas de arquitetura MLP, treinadas segundo o algoritmo do gradiente conjugado, são utilizadas juntamente com um sistema de aquisição de dados para predição das forças de corte. As forças de corte são, naquele trabalho, utilizadas como indicador indireto de medição do desgaste da ferramenta. O autor reporta bons resultados na tarefa proposta.

A utilização de redes para a predição em processos de usinagem é abundante na literatura. Tansel et al. (2006) realizaram dois experimentos nos quais utilizaram redes de arquitetura MLP otimizadas através de algoritmo genético para a predição de condições ótimas de operação. O objetivo no primeiro experimento, envolvendo micro-usinagem, foi o de que a rede predissesse condições de corte que levassem à manutenção de forças de corte dentro de uma determinada faixa de valores. As entradas de rede utilizadas neste experimento foram a profundidade de usinagem, o avanço e o raio da ferramenta. No segundo experimento, o objetivo foi o de que as redes predissessem condições de corte que levassem ao melhor compromisso entre rugosidade e vida da ferramenta. Neste caso, as entradas utilizadas foram o avanço, a rotação, profundidade radial de usinagem e a tolerância. Os autores utilizaram equações provenientes de simulação para geração de casos de treinamento e testes das redes neurais utilizadas em seu trabalho.

Durmus et al. (2006) utilizaram redes neurais para a predição de perda de material por desgaste na ferramenta e para a predição da rugosidade na usinagem de ligas de alumínio. Seus resultados indicam bom desempenho das redes na tarefa proposta. Panda et al (2006) aplicaram redes MLP à tarefa de predição do desgaste de flanco em operações de furação, utilizando parâmetros do processo tais como rotação e diâmetro da broca, avanço, força de avanço, torque e espessura do cavaco, obtendo bons resultados.

Erzurumlu e Oktem (2006) realizaram um trabalho de comparação entre o desempenho de um modelo obtido através do método da superfície de resposta (RSM) com a predição obtida de redes neurais artificiais para rugosidade em partes moldadas. As redes utilizadas foram de arquitetura MLP, treinadas segundo o algoritmo de retro-propagação. Seu trabalho indicou que redes neurais tiveram resultado ligeiramente superior ao modelo polinomial utilizado.

De acordo com Cus e Zuperl (2006), o treinamento de redes neurais pode ser parcialmente baseado sobre resultados extraídos de modelos pré-existentes, analíticos ou empíricos. Os autores compararam redes de arquitetura MLP e RBF na predição da rugosidade em operações de torneamento. Seu trabalho indica que redes RBF são confiáveis e extremamente rápidas.

Em Sonar et al (2006) encontra-se um estudo que aplica redes RBF à predição da rugosidade da peça baseada nos parâmetros de corte. Seu trabalho procurou comparar o desempenho de redes RBF com o de redes MLP aplicadas à mesma tarefa. Seu trabalho apontou para um desempenho ligeiramente inferior das redes RBF na comparação com a outra arquitetura. Não se encontra em tal trabalho, porém, indícios de uma busca sistemática por uma configuração otimizada de rede.

Alajmi, Oraby e Esat (2005) compararam algumas arquiteturas de redes neurais, RBF entre elas, na predição da vida da ferramenta. Seus resultados indicaram um melhor desempenho para redes MLP na comparação com redes RBF. Os dados utilizados como conjunto de treinamento em seu trabalho foram obtidos de experimentos de corte realizados segundo a metodologia de projeto de experimentos. Não se nota, porém, um método sistemático para o projeto de redes, o que pode explicar o fraco desempenho obtido pelas redes RBF.

El-Mounayri et al. (2005) aplicaram redes RBF à predição de forças de corte em operações de fresamento de topo com ponta esférica. As variáveis de entrada utilizadas foram velocidade, o avanço e a profundidade radial e axial de usinagem. Os resultados indicam que as redes atingiram alto grau de exatidão na tarefa proposta. Além disso, os autores destacam a facilidade na definição e a rápida convergência daquele tipo de rede.

Ezugwu et al. (2005) utilizaram redes neurais de arquitetura MLP, treinadas segundo o algoritmo de Levenberg-Marquardt e utilizando regularização Bayesiana para predição das relações entre parâmetros de corte e de processo no torneamento de ligas de níquel. As entradas utilizadas foram a velocidade de corte, o avanço, a profundidade de usinagem, o tempo de corte, a pressão do fluido de corte. As saídas utilizadas foram parâmetros tais como

as componentes das forças de corte, consumo de potência do motor, rugosidade superficial, desgaste médio e máximo de flanco. Redes diferentes foram treinadas para cada saída requerida. O desempenho das redes neurais na predição foi excelente, segundo os autores.

Özel e Karpat (2005), realizaram um trabalho envolvendo a predição de vida de ferramentas e de rugosidade com a utilização de redes neurais de arquitetura MLP, com a utilização do algoritmo de treinamento de Levenberg-Marquardt para a predição da Rugosidade Média da peça usinada e compararam os resultados àqueles obtidos por regressão múltipla. Procurou-se também comparar o desempenho de redes com saída única ao desempenho de redes com duas saídas, uma para vida e outra para rugosidade. Como entradas, foram usadas, além dos parâmetros de corte, as forças de usinagem, a dureza do material da peça, a geometria da ferramenta e o comprimento de corte. Sua conclusão foi de que redes com saída única apresentam melhor desempenho do que as redes com duas saídas.

Sanjay et al. (2005) empregaram redes de arquitetura MLP para detecção de desgaste em brocas. As entradas utilizadas foram o diâmetro da broca, o avanço, a velocidade de corte, o torque, o tempo de usinagem e a força de avanço. Durmus et al. (2006) empregaram redes neurais à tarefa de predição de desgaste da ferramenta e Rugosidade Média de peças de ligas de alumínio AA6351.

Choudury e Bartarya (2003) utilizaram dados obtidos em experimentos de corte realizados segundo a metodologia do DOE como entradas para o treinamento de redes neurais para predição. Seu trabalho utilizou-se de redes neurais de arquitetura MLP treinadas segundo o algoritmo de retro-propagação. Os parâmetros de corte (velocidade de corte, avanço e profundidade de usinagem) foram utilizados como entradas, enquanto que as saídas de rede foram definidas como sendo o desgaste de flanco, o acabamento da superfície e a temperatura da zona de corte. O trabalho realiza uma comparação entre o modelo polinomial obtido e o resultado de predição das redes neurais. Os autores são incisivos ao concluir que a rede neural apresenta melhor desempenho que o modelo polinomial.

Benardos e Vosniakos (2002) realizaram um estudo para a predição da Rugosidade Média da peça em fresamento de faceamento utilizando redes neurais artificiais que foram treinadas através de dados obtidos de experimentos planejados segundo técnicas do método Taguchi. Os autores utilizaram redes MLP treinadas segundo o algoritmo Levenberg-Marquardt.

Tosun e Özler (2002), realizaram trabalho envolvendo redes neurais MLP treinadas segundo o algoritmo de retro-propagação para predição da vida de ferramentas de corte utilizadas na usinagem de aço-manganês em altas temperaturas. O trabalho envolveu a coleta

de dados para a vida da ferramenta em diferentes condições de avanço, profundidade de usinagem, velocidades de corte e temperaturas de superfície. Os resultados coletados foram utilizados para a elaboração de um modelo de regressão múltipla e também como entradas para o treinamento da rede neural. O trabalho conclui pela superioridade das redes neurais na comparação com os modelos de regressão múltipla.

Özel e Nadgir (2002) realizaram um trabalho sobre a predição do desgaste de flanco na usinagem de um aço H13 (55 HRC) utilizando redes neurais MLP treinadas pelo algoritmo de retro-propagação. As entradas utilizadas foram a velocidade e a força de corte, o avanço, o tempo e a força de avanço. A saída de rede no trabalho foi o desgaste de flanco, medido em mm. Segundo os autores, as redes foram capazes de prever com precisão o desgaste de ferramentas utilizando-se as forças de usinagem como entradas.

Chien e Chou (2001), utilizaram redes neurais de arquitetura MLP treinadas pelo algoritmo de retro-propagação para predição da rugosidade, força de corte e vida da ferramenta. Seu trabalho apontou para bons resultados de predição mesmo com a utilização de um conjunto de treinamento de tamanho reduzido.

Em Dos Santos (2001) encontra-se uma aplicação das redes neurais à predição do desgaste de ferramentas e da rugosidade em processo de fresamento do aço ABNT 1045. Os parâmetros utilizados como entradas são: a velocidade de corte, o avanço por dente, a profundidade de usinagem, o comprimento em balanço da fresa, o posicionamento da mesa em relação à mesma, a potência efetiva de corte e a vibração. Os resultados do trabalho indicam que as redes proporcionaram estimativas de elevada confiabilidade e baixo erro a partir de um número relativamente pequeno de ensaios.

Tsai e Wang (2001) compararam seis tipos de redes neurais e um sistema híbrido contendo uma rede neural e lógica nebulosa na predição de rugosidade. Suas conclusões apontaram para um melhor desempenho de redes neurais MLP com funções de ativação do tipo tangente hiperbólica.

Tansel et al. (2006) propuseram um sistema baseado em métodos de codificação e redes neurais empregadas na tarefa de classificação, para monitoração contínua do desgaste da ferramenta durante a usinagem de micro partes não-metálicas. Seu trabalho utilizou redes neurais probabilísticas (PNN), bem como redes MLP. O baixo custo, a confiabilidade e simplicidade de implementação foram apontados pelos autores como as principais vantagens do sistema.

Chao e Hwang (1997) realizaram trabalho envolvendo redes neurais MLP treinadas segundo o algoritmo de retro-propagação para predição da vida de ferramentas de corte. Nota-

se naquele estudo a preocupação com o tratamento estatístico adequado dos resultados obtidos.

A utilização da metodologia de projeto de experimentos em conjunto com redes neurais pode ser encontrada em Kim e Yum (2003). Os autores realizaram um trabalho com uso da metodologia de Taguchi com o objetivo de otimizar o projeto de redes neurais de arquitetura MLP treinadas pelo algoritmo de retro-propagação. O trabalho define parâmetros de projeto das redes neurais como fatores e também algumas variáveis de ruído para o planejamento experimental com os arranjos ortogonais. Sua conclusão é de que o projeto sistematizado de redes neurais pode oferecer um melhor desempenho do que a abordagem tradicional, de tentativa e erro.

Benardos e Vosniakos (2002), utilizaram o método de Taguchi para selecionar um conjunto de entradas de redes neurais que conduzisse à melhor predição da rugosidade superficial em processo de fresamento de faceamento de ligas de alumínio. Cada combinação de entradas foi apresentada por três vezes a três arquiteturas de rede diferentes. As redes utilizadas foram do tipo MLP, treinadas com o uso do algoritmo de Levenberg-Marquardt. Seu trabalho obteve sucesso em descobrir uma arquitetura de rede que apresentasse bons resultados.

A utilização da metodologia de projeto de experimentos para a determinação de parâmetros de redes neurais pode ser encontrada em Khaw et al. (1995). Os autores utilizaram a metodologia de Taguchi para o projeto de redes MLP utilizadas na definição de políticas operacionais para sistemas de manufatura. A pesquisa procurou obter redes neurais com maior precisão e menor tempo de treinamento. Levou-se em conta também a existência de ruídos, presentes sob a forma de pesos aleatórios nas conexões entre os neurônios da rede no início do treinamento. A abordagem do trabalho considera a rede como uma “caixa-preta”. Fatores tais como o número de neurônios na camada oculta, tamanho do conjunto de treinamento e valores dos coeficientes de aprendizado são variados, como o objetivo de otimizar-se o desempenho da rede. Em Peterson et al (1995) encontra-se uma aplicação dos arranjos ortogonais de Taguchi com o objetivo de investigar-se o efeito de diversos fatores sobre a performance de redes neurais MLP treinadas segundo o algoritmo de retro-propagação, aplicadas à tarefa de aproximação de funções. O trabalho procurou investigar os efeitos do ruído presente nos conjuntos de treinamento e teste, da densidade dos casos de treinamento e de diferentes números de neurônios na camada oculta. Os autores concluíram que menores níveis de ruído são melhores para o desempenho da rede assim como um conjunto de treinamento mais denso. Em relação à arquitetura, o estudo conclui que a melhor opção

corresponde à rede com duas camadas escondidas. É interessante notar, em tal trabalho, o uso de conceitos de projeto de experimentos (DOE) aplicados à determinação de parâmetros para redes neurais.

Outras estratégias de uso de redes neurais artificiais podem ser observadas na literatura. Tansel et al. (2006) utilizaram redes neurais MLP co-localizadas em conjuntos para predição das condições ótimas de corte em operações de micro-usinagem e fabricação de moldes. Zawada-Tomkiewicz (2001) realizou trabalho voltado à monitoração contínua de desgaste de ferramentas utilizando aquisição digital de imagens. Seu estudo fez uso do resultado do processamento digital de tais imagens como entrada para redes neurais, com o objetivo identificar padrões de desgaste de ferramentas.

Em outros processos industriais, a presença de redes neurais aplicadas à tarefa de predição também é notável. O uso de redes na modelagem de funções, aplicado à previsão de parâmetros de interesse na soldagem de ligas de titânio, é encontrado em Casalino et al. (2005). Dados oriundos de experimentos realizados segundo a metodologia de Taguchi, contendo três fatores com dois níveis cada, foram utilizados para treinamento de redes de arquitetura MLP. O objetivo foi fazer com que, partindo dos dados reais, a rede pudesse interpolar dados experimentais incompletos, para permitir a análise dos mesmos segundo a técnica de análise de variância (ANOVA). Turk et al. (2004) aplicaram redes à tarefa de predição do desgaste de matrizes e do número crítico de ciclos de operação acima do qual a ferramenta não consegue mais manter a peça dentro das especificações em processo de moldagem.

Um trabalho de interesse envolvendo redes neurais, ainda que não aplicadas a processos de usinagem é o estudo de Piñeros Garcet et al. (2006). No estudo os autores utilizaram redes de arquitetura RBF como meta-modelos para aplicação em substituição a um modelo numérico de drenagem de nitrato em solos. O trabalho foi proposto por seus autores como validação da metodologia de desenvolvimento e validação de meta-modelos por eles proposta. Seus resultados indicam que as redes RBF constituem-se em alternativa válida para utilização como meta-modelos.

3.10 Considerações Finais

As revisões histórica e bibliográfica desenvolvidas neste capítulo contextualizam as redes neurais e sua aplicação em tarefas de monitoração e predição da vida da ferramenta, desgaste e rugosidade da peça. A farta literatura sobre o assunto indica a relevância do tema. Observa-se também o limitado uso feito até o momento de redes de arquitetura RBF para a predição em processos de usinagem. Tais fatos motivam o uso de tal arquitetura de redes na execução deste trabalho.

4 METODOLOGIA DE PROJETO E ANÁLISE DE EXPERIMENTOS

4.1 Considerações Iniciais

A metodologia do projeto de experimentos foi desenvolvida entre 1920 e 1930 por Fisher, sendo posteriormente aperfeiçoada por outros pesquisadores como Box, Hunter e Taguchi. Ronald A. Fisher foi o inovador no uso dos métodos estatísticos e da análise de dados. Em 1933, levou seu trabalho para a Universidade de Londres, onde o DOE (do inglês *Design of Experiments*) foi aplicado com sucesso em experimentos agrícolas e em ciências biológicas. A primeira aplicação prática do DOE data de 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).

Segundo Özel e Karpaz (2005), toda a literatura relevante envolvendo a quantificação de impactos de parâmetros de usinagem na qualidade final da peça abarcam algum tipo de projeto de experimentos que leve a um modo sistemático de quantificação dos efeitos de um número finito de parâmetros.

O 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, 2001).

Montgomery e Runger (2003) 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.

Montgomery e Runger (2003) afirmam que 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, segundo Montgomery (2001), 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.

Dentre as características citadas por Montgomery e Runger (2003) 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.

4.2 Metodologia de Taguchi e arranjos ortogonais

4.2.1 Projeto Robusto

Dentre as várias abordagens de engenharia de qualidade para produtos e processos, encontra-se o método de Taguchi et al.(1987). Segundo Ross (1991), o método corresponde a uma abordagem ‘fora da linha’ da engenharia da qualidade, por ter como objetivo assegurar, ainda durante a fase de projeto e concepção, a obtenção de boa qualidade de produtos e bom desempenho de processos. Sua principal premissa é de que a variância de um processo resulta

em perda de qualidade. Os métodos de Taguchi distinguem-se de outras abordagens de engenharia de qualidade por alguns conceitos que lhe são específicos:

- a) Função perda da qualidade;
- b) Funções sinal-ruído;
- c) Arranjos ortogonais.

Ross (1991) afirma que a qualidade de um produto ou serviço pode ser objetivamente medida pelo valor do desvio de alguma característica funcional do mesmo em relação a um valor esperado. A este desvio dá-se o nome de função perda da qualidade. Tais desvios são originados por fatores de variação do processo. Aos fatores que são controláveis, dá-se o nome de parâmetros de projeto. Por oposição, àqueles que não o são dá-se o nome de ruídos.

A variância, segundo Taguchi et al.(1987), provém de fontes de variação que podem ser controláveis ou não. Seu método propõe produzir um processo que seja robusto em relação aos fatores perturbadores, ao invés de buscar a eliminação dos mesmos. Robustez, segundo Taguchi et al. (1987), é entendida aqui como a redução da sensibilidade do processo (ou do produto) à presença de fatores de perturbação. A robustez é medida, dentro da metodologia, por meio de uma relação sinal-ruído.

A proposta do método inclui a determinação e minimização da função perda da qualidade bem como a maximização da relação sinal-ruído através do emprego de técnicas de projeto de experimentos envolvendo arranjos ortogonais.

Segundo Benardos e Vosniakos (2002), a metodologia de Taguchi abrange uma série de passos:

- Formulação do problema;
- Formulação dos objetivos do experimento;
- Seleção das características da qualidade e sistemas para sua medição;
- Seleção dos fatores que podem influenciar as características da qualidade;
- Seleção dos níveis dos fatores;
- Seleção do arranjo ortogonal apropriado;
- Seleção das interações que podem afetar as características da qualidade;
- Distribuição dos fatores e interações pelo arranjo ortogonal escolhido;
- Condução dos experimentos, segundo os arranjos ortogonais escolhidos;
- Análise dos resultados experimentais;
- Condução de um experimento de confirmação.

4.2.2 Planejamento de Experimentos no método de Taguchi

Segundo Liao (2004), o método de Taguchi é uma abordagem tradicional para o projeto de experimentos robustos, que procura obter a melhor combinação de fatores e níveis com o menor custo para atender aos requisitos dos consumidores. Segundo o autor, no método de Taguchi os parâmetros de projeto (aqueles que podem ser controlados pelo projetista) e os parâmetros de ruído (fatores que não estão sob o controle do projetista, tais como fatores ambientais) são considerados como influentes na qualidade do produto. O autor ressalta, porém, que o julgamento de engenharia é uma ferramenta primária para aplicação bem sucedida do método.

A estratégia de projeto de experimentos utilizada no método de Taguchi é baseada em arranjos ortogonais. Eles correspondem a um tipo de experimento fatorial fracionário, nos quais nem todas as combinações possíveis entre fatores e níveis são testadas. Tais arranjos fatoriais são úteis para a estimação dos efeitos principais dos fatores e de efeitos de interação selecionados sobre o processo. O objetivo primeiro deste tipo de estratégia é obter o máximo de informação sobre o efeito dos fatores com um mínimo de corridas experimentais (ROSS, 1991). Segundo Khaw et al. (1995), os arranjos ortogonais permitem a análise das relações entre um conjunto numeroso de parâmetros de projeto com o menor número possível de experimentos.

Em adição ao fato de requerer um número menor de experimentos, os arranjos ortogonais ainda permitem testar fatores com número misto de níveis. Assim, é possível fazer-se experimentos contendo alguns fatores com dois níveis, e outros com quatro níveis, por exemplo. Para um experimento planejado contendo três fatores, sendo um deles em quatro níveis e os demais em dois níveis cada, sem investigar-se a interação entre os fatores, o arranjo ortogonal L8 de Taguchi é como mostrado no quadro 4.1. Os numerais na coluna ‘Número do experimento’ indicam o número da corrida experimental e o numerais nas colunas ‘Fator A’, ‘Fator B’ e ‘Fator C’ indicam os níveis do respectivo fator.

Número do Experimento	Fator A	Fator B	Fator C
1	1	1	1
2	1	2	2
3	2	1	1
4	2	2	2
5	3	1	2
6	3	2	1
7	4	1	2
8	4	2	1

Quadro 4.1 – Arranjo L8 para experimento com um fator com quatro níveis e dois fatores com dois níveis, sem investigação da interação. Fonte: Minitab® Statistical Software Release 13.0

4.2.3 Análise de experimentos no método de Taguchi

A função perda da qualidade, dentro do método de Taguchi, varia conforme o tipo do problema estudado. Problemas podem ser do tipo: “quanto menor melhor”, “quanto maior melhor”, “nominal é melhor”. Neste trabalho, o objetivo foi a minimização da resposta analisada, caracterizando um problema do tipo “quanto menor, melhor”. Para tal tipo de problema, a relação sinal-ruído a ser maximizada é expressa pela equação (12) (ROSS, 1991; TAGUCHI ET AL., 1987).

$$\eta = -10 \log_{10} \left[\frac{1}{n} \sum_{i=1}^n y_i^2 \right] \quad (12)$$

em que η é o valor da relação sinal-ruído, y_i é valor do desvio em relação à característica da qualidade cuja tolerância é do tipo “quanto menor, melhor” e n corresponde ao número de experimentos realizados.

Os dados obtidos experimentalmente são analisados segundo suas médias, valor da relação sinal-ruído e desvios padrão das corridas. O objetivo da análise é obter-se os níveis dos fatores que levam à minimização da função-perda da qualidade e para a maximização da relação sinal-ruído (ROSS, 1991).

As técnicas de projeto de experimentos no método de Taguchi, segundo Benardos e Vosniakos (2003), têm sua maior utilidade como procedimento exploratório para determinar a significância de cada fator e para identificar os níveis dos fatores que levam ao desempenho esperado, sem a formulação de nenhum tipo de modelo.

Interações selecionadas podem ser estudadas, se colunas dos arranjos ortogonais forem dedicadas às mesmas. Isso implicará, no entanto, em aumento do número de corridas

experimentais necessárias, reduzindo assim a principal vantagem dos arranjos que é o número reduzido de experimentos. Por outro lado, segundo Ross (1991), a utilização de fatores em todas as colunas faz com que os efeitos de possíveis interações existentes entre eles sejam mascarados, isto é, agrupados aos efeitos principais. O autor alerta para o fato de que a redução do número de ensaios implica na perda de informação. Observa-se deste modo que o planejamento de experimentos com o uso dos arranjos ortogonais implica em um compromisso entre a resolução e o número de experimentos.

Em alguns casos, encontram-se aplicações nas quais não se utiliza toda a metodologia, senão as ferramentas que são de interesse para a solução de um determinado problema. Um exemplo é o trabalho de Casalino et al. (2005). Naquele trabalho, os arranjos ortogonais de Taguchi são utilizados como um meio de obter-se informação para modelar o processo de soldagem de ligas de titânio Ti6Al4V com um mínimo de corridas experimentais. A análise de Taguchi é realizada sobre os valores das médias e razão sinal-ruído. Sendo aquele um experimento exploratório, no entanto, outros passos da metodologia que não eram do interesse do trabalho não foram realizados, como o experimento de confirmação. Benardos e Vosniakos (2003) afirmam que, dada sua generalidade e fundamentação estatística, certas tarefas da metodologia Taguchi podem ser utilizadas isoladamente em problemas de engenharia nos quais seja necessário promover uma redução inicial do espaço de busca.

4.3 Fatoriais Completos

Um planejamento fatorial é completo quando todas as possíveis combinações entre os fatores envolvidos são experimentadas (BOX, HUNTER E HUNTER, 1978). Em um experimento planejado segundo a técnica do fatorial completo cada fator é testado e re-testado um 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.

4.4 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 (ANOVA) (Montgomery e Runger, 2003).

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 (13).

$$Y_{ijk} = \mu + A_i + B_j + AB_{ij} + \varepsilon_{ijk} \quad (13)$$

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 (Between) pela variação dentro de um tratamento (Within) (Montgomery e Runger, 2003). 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. De acordo com Montgomery e Runger (2003) 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 (Montgomery, 2001).

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 4.2.

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

Quadro 4.2 – Hipóteses presentes em uma ANOVA para dois fatores. Fonte: Adaptado de Montgomery e Runger (2003).

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 4.3 (Montgomery e Runger, 2003).

Soma de cada linha (Y_{st})	$\sum_{j=1}^b \sum_{k=1}^n y_{ijk}$	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)	$\frac{1}{abn} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n y_{ijk}$

Quadro 4.3 – Fórmulas para análise de variância para dois fatores em dois níveis. Fonte: Adaptado de Montgomery e Runger (2003)

Com estes somatórios, pode-se escrever a soma de quadrados total (SS_t) como mostrado na equação (14).

$$\begin{aligned} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y}_t)^2 &= \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n [(\bar{y}_l - \bar{y}_t) + (\bar{y}_c - \bar{y}_t) + (\bar{y}_{cel} - \bar{y}_l - \bar{y}_c - \bar{y}_t) + (y_{ijk} - \bar{y}_t)]^2 \\ &= bn \sum_{i=1}^a (\bar{y}_l - \bar{y}_t)^2 + an \sum_{j=1}^b (\bar{y}_c - \bar{y}_t)^2 + n \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{cel} - \bar{y}_l - \bar{y}_c - \bar{y}_t) + \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y}_c)^2 \end{aligned} \quad (14)$$

Dessa maneira, a Soma de Quadrados Total do arranjo 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 (15).

$$SS_T = SS_A + SS_B + SS_{AB} + SS_E \quad (15)$$

As equações das somas de quadrados podem ser ainda escritas como nas equações (16), (17) e (18).

$$SS_T = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n y_{ijk}^2 - \frac{y_{st}^2}{abn} \quad (16)$$

$$SS_A = \frac{1}{bn} \sum_{i=1}^a y_{st}^2 - \frac{y_{st}^2}{abn} \tag{17}$$

$$SS_B = \frac{1}{an} \sum_{j=1}^b y_{sc}^2 - \frac{y_{st}^2}{abn} \tag{18}$$

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 4.4.

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 _o = $\frac{MS_A}{MS_E}$
Fator B	SS _B	b-1	MS _B = $\frac{SS_B}{b-1}$	F _o = $\frac{MS_B}{MS_E}$
Interação AB	SS _{AB}	(a-1)(b-1)	MS _{AB} = $\frac{SS_{AB}}{(a-1)(b-1)}$	F _o = $\frac{MS_{AB}}{MS_E}$
Erro	SS _E	ab(n-1)	MS _E = $\frac{SS_E}{ab(n-1)}$	
Total	SS _T	abn-1		

Quadro 4.4 – Fórmulas para análise de variância em dois níveis. Fonte: Adaptado de Montgomery e Runger (2003)

Para se obter uma simplificação da soma de quadrados referentes à interação calcula-se primeiramente a soma de quadrados parciais, tal qual na equação (19).

$$SS_p = \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^b y_{scel}^2 - \frac{y_{st}^2}{abn} \tag{19}$$

Subtrai-se SS_A e SS_B de SS_p, conforme mostrado pela equação (20) e assim encontra-se a parcela de variação devido ao erro, conforme a equação (21).

$$SS_{AB} = SS_p - SS_A + SS_B \tag{20}$$

$$SS_E = SS_T - SS_{AB} - SS_A - SS_B \tag{21}$$

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.

4.5 Teste para significância do modelo

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).

4.6 Teste para significância individual dos coeficientes do modelo

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 o modelo deve conservar ambos.

4.7 Modelos de regressão linear

O método dos mínimos quadrados ordinários (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 (22).

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \varepsilon_i \quad (22)$$

O método dos mínimos quadrados baseia-se na escolha de valores para β na equação (22), de modo que a soma dos quadrados dos erros seja minimizada. A função de mínimos quadrados pode ser escrita como na equação (23).

$$L = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij} \right)^2 \quad (23)$$

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 (24).

$$\frac{\partial L}{\partial \beta_0} = \frac{\partial L}{\partial \beta_j} = -2 \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij} \right) = 0 \quad (24)$$

Simplificando-se e generalizando-se a equação (24), obtém-se o mostrado na equação (25).

$$\hat{\beta}_0 \sum_{i=1}^n x_{ik} + \hat{\beta}_1 \sum_{i=1}^n x_{i1} x_{ik} + \hat{\beta}_2 \sum_{i=1}^n x_{ik} x_{i2} + \dots + \hat{\beta}_k \sum_{i=1}^n x_{ik}^2 = \sum_{i=1}^n x_{ik} y_i \quad (25)$$

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 (25) pode ser escrito como mostrado nas equações (26) e (27).

$$Y = \beta X + \varepsilon \quad (26)$$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \vdots \\ \varepsilon_k \end{bmatrix} \quad (27)$$

É possível notar que L pode ser expresso como na equação (28). Como os estimadores de mínimos quadrados devem satisfazer à equação (29), uma simplificação da mesma conduz à equação (30).

$$L = y^T y - \beta^T X^T y - y^T X \beta + \beta^T X^T X \beta = y^T y - 2\beta^T X^T y + \beta^T X^T X \beta \quad (28)$$

$$\frac{\partial L}{\partial \beta} = -2X^T y + 2X^T X \hat{\beta} = 0 \quad (29)$$

$$X^T X \hat{\beta} = X^T y \quad (30)$$

Multiplicando-se ambos os lados da equação (30) pela inversa da matriz $X^T X$, tem-se a equação (31). Esta corresponde à expressão matricial das equações normais de mínimos quadrados, descritos anteriormente (Montgomery, 2001).

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (31)$$

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 (32). 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 (33).

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \beta_{ii} x_i^2 + \sum_{i < j} \beta_{ij} x_i x_j + \epsilon \quad (32)$$

$$e = y - \hat{y} \quad (33)$$

Considerando-se a soma de quadrados para os resíduos, chega-se à equação (34).

$$SS_E = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (e_i)^2 = e^T e \quad (34)$$

Substituindo-se $e = y - \hat{y} = y - X\hat{\beta}$, chega-se à equação (35).

$$\begin{aligned} SS_E &= (y - X\hat{\beta})^T (y - X\hat{\beta}) = y^T y - \beta^T X^T y - y^T X\hat{\beta} + \hat{\beta}^T X^T X\hat{\beta} \\ SS_E &= y^T y - 2\beta^T X^T y + \hat{\beta}^T X^T X\hat{\beta} \end{aligned} \quad (35)$$

Como: $X^T X\hat{\beta} = X^T y$, a equação (35) toma a forma da equação (36).

$$SS_E = y^T y - \hat{\beta}^T X^T y \quad (36)$$

A equação (36) 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 (37).

$$SS_T = \sum_{i=1}^n y_i^2 - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n} = y^T y - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n} \quad (37)$$

Para a soma de quadrados do modelo, tem-se o mostrado na equação (38).

$$SS_m = SS_T - SS_e = y^T y - \frac{1}{n} \left(\sum_{i=1}^n y_i\right)^2 - (y^T y - \hat{\beta}^T X^T y) = \hat{\beta}^T X^T y - \frac{1}{n} \left(\sum_{i=1}^n y_i\right)^2 \quad (38)$$

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 (39).

$$\begin{aligned} H_0 &= \beta_1 = \beta_2 = \beta_3 = \dots = \beta_k = 0 \\ H_1 &= \beta_j \neq 0 \end{aligned} \quad (39)$$

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 (39) 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 (40).

$$F_0 = \frac{SQ_M/k}{SQ_E/(n-p)} = \frac{MQ_M}{MQ_E} \quad (40)$$

Deve-se rejeitar a hipótese nula se o valor calculado na equação (40) 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 (41).

$$\begin{aligned} H_0 &: \beta_j = 0 \\ H_1 &: \beta_j \neq 0 \end{aligned} \quad (41)$$

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 (42).

$$T_0 = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 \cdot C_{jj}}} \quad (42)$$

Na equação (42) C_{jj} é o elemento da diagonal de $(X^T X)^{-1}$ correspondente a $\hat{\beta}_j$. Observa-se que o denominador da equação (42) 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 (43).

$$\hat{\beta}_j - t_{\alpha/2, n-p} \sqrt{\hat{\sigma}^2 C_{jj}} \leq \beta_j \leq \hat{\beta}_j + t_{\alpha/2, n-p} \sqrt{\hat{\sigma}^2 C_{jj}} \quad (43)$$

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- α)% é dado pela equação (44).

$$\hat{\mu}_{Y|x_0} - t_{\alpha/2, n-p} \sqrt{\sigma^2 x_0^T (X^T X)^{-1} x_0} \leq \mu_{Y|x_0} \leq \hat{\mu}_{Y|x_0} + t_{\alpha/2, n-p} \sqrt{\sigma^2 x_0^T (X^T X)^{-1} x_0} \quad (44)$$

4.8 Medidas de adequação de 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 haver, 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 (45).

$$R^2 = \frac{SQ_M}{SQ_T} = 1 - \frac{SQ_E}{SQ_T} \quad (45)$$

Já o R^2 ajustado é uma modificação que considera o número p de variáveis incluídas no modelo. Observando-se a equação (46), nota-se que o valor ajustado decresce à medida que p aumenta.

$$R_{aj}^2 = 1 - \left(\frac{n-1}{n-p} \right) (1 - R^2) \quad (46)$$

4.9 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 (47).

$$d_i = \frac{e_i}{\sqrt{MQ_E}} = \frac{e_i}{\sqrt{\hat{\sigma}^2}} \quad (47)$$

A padronização dos resíduos cria um escalonamento para o desvio padrão, o que torna sua interpretação mais fácil. Uma outra forma de escalonamento é a de Student, dada pela equação (48).

$$d_i = \frac{e_i}{\sqrt{\hat{\sigma}^2(1-h_{ii})}} \quad (48)$$

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

4.10 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 (49).

$$SS_e = SS_{pe} + SS_{lof} \quad (49)$$

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 (50).

$$y_{ij} - \hat{y}_i = (y_{ij} - \bar{y}_i) + (\bar{y}_i - \hat{y}_i) \quad (50)$$

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 (49) e somando-se cada i e j , obtém-se a equação (51).

$$\underbrace{\sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \hat{y}_i)^2}_{SS_r} = \underbrace{\sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}_{SS_{ep}} + \underbrace{\sum_{i=1}^m n_i (\bar{y}_i - \hat{y}_i)^2}_{SS_{lof}} \quad (51)$$

O lado esquerdo da equação (51) 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 um 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 (52).

$$F_0 = \frac{SS_{lof} / (m - p)}{SS_{pe} / (n - m)} = \frac{MS_{lof}}{MS_{pe}} \quad (52)$$

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.

4.11 Considerações Finais

Neste capítulo fez-se breve introdução à metodologia de projeto de experimentos (DOE). Foram discutidos o planejamento e a análise de experimentos com uso da metodologia de Taguchi e também com o uso das técnicas de Fatoriais Completos. Os testes estatísticos utilizados na análise experimental foram definidos e seus princípios apresentados. Tais testes serão utilizados para a análise dos resultados obtidos nos experimentos realizados, de cujo planejamento trata o próximo capítulo.

5 PROCEDIMENTO EXPERIMENTAL

5.1 Considerações Iniciais

Este capítulo descreve o procedimento experimental seguido durante este trabalho. Inicialmente são feitas considerações a respeito da evolução e delimitação do escopo. A seguir são definidos os fatores e respectivos níveis de variação envolvidos em cada ciclo experimental. O capítulo prossegue com a discussão dos cuidados experimentais para prover o estudo do necessário embasamento estatístico que permita sustentar as conclusões obtidas.

A geração dos dados experimentais é discutida. A seguir, são detalhados os arranjos experimentais utilizados em cada ciclo experimental. Finalmente, apresenta-se o modo como foi executada cada fase do trabalho.

5.2 Delimitação do escopo e evolução do trabalho

A proposta inicial deste trabalho foi a de comparar o desempenho de diferentes arquiteturas de redes neurais na predição da vida de ferramentas e da rugosidade da peça no processo de torneamento. Pensou-se em comparar redes RBF (função de base radial), redes GRNN (redes neurais para regressão generalizada) e redes MLP (perceptron multi-camada). Para estas últimas, pensou-se em avaliar diferentes algoritmos de treinamento. Como a quantidade de experimentos a ser realizada seria muito elevada, optou-se por realizar um planejamento experimental com a utilização dos arranjos ortogonais de Taguchi. Tais arranjos permitiriam a obtenção de informações com um número reduzido de experimentos.

Os resultados dos experimentos conduzidos segundo os arranjos ortogonais, no entanto, levaram a uma mudança de objetivos. Em primeiro lugar, constatou-se que a realização de um planejamento experimental apropriado, que cobrisse todas as arquiteturas descritas anteriormente e seus principais algoritmos de treinamento, geraria uma quantidade enorme de experimentos e dados, mesmo com a utilização dos fatoriais fracionários de Taguchi. Isso tornaria inviável a execução do trabalho e a análise de resultados em tempo hábil. Em segundo lugar, os experimentos envolvendo redes RBF e arranjos ortogonais mostraram que o desempenho de tais redes na tarefa de predição era bastante promissor. Este fato encorajou a mudança de foco do trabalho para este tipo de rede. Além disso, os resultados dos experimentos envolvendo redes RBF e arranjos ortogonais de Taguchi revelaram alguns

indícios da existência de interações entre os fatores, os quais não puderam ser devidamente estimados naquele ciclo experimental devido ao arranjo ortogonal adotado. Pensou-se também na possibilidade de aprofundar o estudo por meio da investigação de outros níveis de variação do número de unidades radiais presentes nas redes.

Aventou-se a hipótese da realização de novo ciclo experimental com a utilização de fatoriais completos mistos. Tais arranjos permitiriam investigar mais níveis de variação do número de unidades radiais bem como a existência de interação entre os fatores. Pelas razões descritas, optou-se pela alteração do escopo da pesquisa, que foi restrito ao estudo das redes RBF e passou a utilizar arranjos fatoriais completos para o planejamento experimental.

Foi realizado um segundo ciclo experimental envolvendo fatoriais completos com mais de dois níveis, nos quais apenas redes de arquitetura RBF foram testadas. A faixa de variação do número de unidades radiais das redes foi aumentada. Após a execução do ciclo, porém, detectou-se que a diferença nos efeitos dos tratamentos era muito elevada. Este fato tornava significativos todos os efeitos principais e todas as interações. O objetivo de identificar fatores ou interações influentes com diferentes números de casos de treinamento não pôde ser alcançado com os experimentos realizados durante este ciclo.

Por outro lado, pôde-se notar que algumas configurações que apresentavam bons resultados em exatidão, estabilidade e variabilidade eram coincidentes entre os vários experimentos realizados, e que tais configurações apresentavam resultados mais próximos entre si. A diferença de desempenho entre os resultados obtidos por tais configurações não podia, no entanto, ser devidamente analisada com o arranjo fatorial misto até então utilizado. Por essa razão optou-se pela execução de um terceiro ciclo experimental, com a utilização de arranjos fatoriais completos mais simples.

Para o terceiro ciclo experimental, os níveis do segundo ciclo que levavam a um mau desempenho foram descartados, e mantidos apenas os dois níveis que levaram aos melhores desempenhos. O critério formalmente adotado para seleção foi o de manter os dois níveis do fator, cujos efeitos principais levassem aos menores valores mínimos quadrados, durante a análise dos experimentos do segundo ciclo. O arranjo experimental utilizado foi um fatorial completo de três fatores e apenas dois níveis. Os resultados deste terceiro ciclo experimental foram os utilizados para a análise estatística dos resultados que levaram às conclusões obtidas neste trabalho.

5.3 Definição dos fatores experimentais e de seus níveis

Os fatores incluídos nos experimentos foram o número de unidades radiais na camada oculta da RBF, o algoritmo para determinação dos centros da função radial e o algoritmo para determinação do parâmetro de largura da função radial. Os níveis dos fatores ‘algoritmo para determinação do parâmetro de largura da função radial’ e ‘algoritmo para determinação dos centros da função radial’ foram definidos como sendo os diferentes algoritmos empregados para as operações. Nota-se, portanto, que dois dos fatores envolvidos nos experimentos são de natureza qualitativa.

Em relação ao fator ‘número de unidades radiais’, Haykin (1994) afirma que ele não pode ser maior do que o número de casos de treinamento, nem muito pequeno quando comparado ao número de casos. Como o autor sugere que a definição deva basear-se em uma proporção, os níveis dos fatores foram definidos como proporções entre o número de unidades radiais e o número de casos de treinamento disponíveis.

O fator ‘algoritmo para determinação dos centros da função radial’ foi definido como um fator de dois níveis. Foram testados dois algoritmos distintos: o algoritmo de Sub-amostragem e o algoritmo das K-Médias. A escolha de tais algoritmos é justificada por Bishop (1995), que cita ambos como os mais comumente utilizados na determinação dos centros radiais. Cada algoritmo foi adotado como um nível do fator.

Para o fator ‘Algoritmo para determinação do parâmetro de largura da função radial’ foram testados dois algoritmos distintos: o algoritmo Isotrópico e o algoritmo dos K-Mais Próximos. Tais algoritmos são os mais adotados para a definição de tal parâmetro, sendo citados por Bishop (1995). Um valor muito pequeno do parâmetro de largura fará com que a resposta da função de ativação decresça acentuadamente com a distância ao seu 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 inexactidão na operação de interpolação (MANUAL STATISTICA, 2005).

Para o algoritmo Isotrópico decidiu-se investigar a influência do seu fator de escala. Desta forma, dois níveis foram selecionados para teste com base em experimentos preliminares. Para o algoritmo dos K Mais Próximos, por sua vez, desejava-se investigar a influência do valor de K. Novamente dois níveis do fator foram selecionados para teste com base em experimentos preliminares. Decidiu-se então agrupar algoritmos e níveis dos respectivos parâmetros como um único fator de quatro níveis.

5.4 Organização dos experimentos

Um dos objetivos do trabalho foi o de investigar a influência da variação do tamanho dos conjuntos de treinamento no desempenho das redes. Isto foi feito através da realização de experimentos com conjuntos de treinamento de tamanhos diferentes dentro de cada ciclo experimental. Foram planejados e executados três ciclos experimentais: um ciclo de oito experimentos com arranjos ortogonais de Taguchi, um ciclo de doze experimentos exploratórios com arranjos fatoriais completos mistos e um ciclo de doze experimentos com arranjos fatoriais completos em dois níveis. Define-se como um experimento, a execução de um arranjo experimental completo para um dado conjunto de treinamento.

As entradas utilizadas para as redes neurais foram os parâmetros de corte (velocidade de corte, avanço e profundidade de usinagem). De acordo com Sick (2002), estas são as grandezas mais utilizadas como entradas de redes neurais em aplicações de predição em processos de torneamento.

Segundo Sick (2002), o uso de redes neurais com uma única saída é a prática mais comum em aplicações de predição da vida da ferramenta com uso de redes neurais. Segundo o autor, o uso de redes com mais de uma saída se justifica somente quando há evidência de correlação entre os parâmetros de saída da rede. Optou-se neste trabalho por implementar redes com uma única saída, também como meio de evitar-se a possibilidade da ocorrência efeitos de inter cruzamento (MANUAL STATISTICA, 2005).

As recomendações dadas por Sick (2002), para a realização de um trabalho científico de qualidade envolvendo predição da vida da ferramenta com uso de redes neurais em processos de torneamento foram levadas em conta no desenvolvimento do trabalho. Tais recomendações são as seguintes:

- a) O uso de diferentes conjuntos de treinamento e teste para experimentos de simulação em redes neurais;
- b) A repetição dos experimentos para evitar que efeitos tais como iniciação aleatória de pesos e seqüências aleatórias de padrões de treinamento influam no resultado do experimento;
- c) Computação de médias e variâncias das saídas para medidas de desempenho;
- d) A aplicação de testes estatísticos para casos em que redes provenientes de diferentes experimentos são comparadas.

5.5 Geração dos dados experimentais

5.5.1 Considerações gerais

Para a obtenção de bom desempenho na tarefa de predição é necessário que haja uma quantidade de exemplos suficiente para que a rede possa modelar a função representativa da vida da ferramenta ou Rugosidade Média (R_a) da peça usinada. O tamanho do conjunto de treinamento necessário, porém, depende da complexidade da função mapeada, não existindo um método para sua definição. Além disso, tal conjunto deve conter uma distribuição equitativa de casos que seja representativa da função a ser modelada (HAYKIN, 1994).

Em razão da falta de recursos para executar um experimento para geração de dados que reunisse as condições enunciadas no parágrafo anterior, tentou-se encontrar na literatura um banco de dados que fosse amplo e houvesse sido gerado em condições experimentais adequadas à análise estatística, o que resultou infrutífero. Por essas razões resolveu-se utilizar casos de treinamento simulados para treinamento e testes das redes. Os casos de treinamento utilizados podem ser considerados como meta-modelos, segundo a definição de Piñeros Garcet et al. (2006).

5.5.2 Geração dos dados para vida da ferramenta

Para geração dos casos utilizados nos experimentos de predição da vida da ferramenta fez-se uso de um modelo polinomial quadrático completo obtido por Paiva et al. (2005) em um experimento de torneamento do aço SAE/ABNT 52100 – 55 HRC com ferramentas com insertos de cerâmica mista ($Al_2O_3 + TiC$), classe Sandvik GC 6050, recoberta com TiN, com geometria ISO CNGA 120408 S01525. As operações de torneamento foram realizadas em um Torno CNC Nardini Logic 175, com potência máxima de eixo de 7,5 CV; rotação máxima de 4000 rpm; torre com oito posições e torque máximo de 200 Kgf.m. Como suporte foi adotado o modelo ISO DCLNL 1616H12. Os corpos de prova utilizados nos ensaios têm dimensões de 49 mm de diâmetro e 50 mm de comprimento e utilizam Aço SAE/ABNT 52100 (Villares), com dureza de 55 HRC. Utilizou-se como critério do fim de vida o valor de $V_{BMax} = 0,3mm$. Tal experimento foi planejado e executado segundo a Metodologia da Superfície de Resposta combinada com a técnica do fatorial completo. Nesses ensaios foram variados os parâmetros de usinagem, adotando-se a velocidade de corte (V_c) entre os níveis

200 m/min e 240 m/min, avanço de usinagem (f) entre os níveis 0,05 mm/v e 0,10 mm/v e profundidade de usinagem (a_p) entre os níveis 0,15 mm e 0,30 mm.

O modelo resultante foi validado estatisticamente por meio de testes de análise de variância (ANOVA) e apresentou um alto nível de ajuste, com R^2 adj = 85,0%. A equação do modelo para a vida da ferramenta é apresentada, em unidades codificadas, na equação (53) (PAIVA ET AL. 2005).

$$y = 7,968 - 1,251V_c - 2,341f - 1,639a_p + 0,234V_c^2 + 1,547f^2 + 0,422a_p^2 + 0,750V_c f + 0,075V_c.a_p + 0,675f.a_p \quad (53)$$

Para o presente trabalho foram gerados, com o uso do software estatístico Minitab®, um arquivo com 620 conjuntos de dados (V_c, f, a_p, y) a partir da equação (53), sendo y o valor correspondente à vida da ferramenta, dado em minutos, prevista segundo o modelo utilizado. Os dados foram gerados seguindo uma distribuição uniforme entre os níveis do experimento original de Paiva et al. (2005). Para cada parâmetro de corte, foram aplicados testes de Levene a todos os pares distintos de conjuntos de treinamento, e verificou-se que, ao nível de significância de 0,05 não havia evidência de diferença de variância entre os conjuntos.

A partir do arquivo original contendo os 620 conjuntos de dados, foram gerados doze conjuntos de treinamento de diferentes tamanhos. Os conjuntos formados contêm 12, 15, 24, 30, 48, 60, 120, 150, 240, 300, 400 e 500 casos de treinamento cada um. Cada conjunto de treinamento é um sub-conjunto do arquivo original. Assim, o conjunto de treinamento contendo 12 casos é formado pelos 12 primeiros casos do arquivo original; o conjunto de treinamento contendo 15 casos é formado pelos 15 primeiros casos do arquivo original, e assim por diante. A todos os pares possíveis de conjuntos de treinamento foram aplicados testes de Levene e verificou-se que, ao nível de significância de 0,05 não havia evidência de diferença de variância entre os mesmos.

Dos 620 casos gerados a partir da equação (53), 120 foram reservados e utilizados unicamente como conjunto de testes. Estes casos não participaram do treinamento das redes e foram usados para testar de maneira independente o desempenho das redes na predição da vida da ferramenta após a fase de treinamento.

5.5.3 Geração dos dados para Rugosidade Média (R_a)

Para geração dos casos utilizados nos experimentos para predição da rugosidade da peça fez-se uso de um modelo polinomial quadrático completo obtido por Pereira (2006) no experimento descrito no item 5.5.2. Para as medições necessárias, utilizou-se um Rugosímetro Taylor Hobson, modelo Surtronic 3⁺ e um Micrômetro Mitutoyo. Para a Rugosidade Média (R_a) da peça usinada, adotou-se como parâmetro a medição da rugosidade após o décimo passe de usinagem.

O modelo resultante foi validado estatisticamente por meio de testes de análise de variância (ANOVA) e apresentou um alto nível de ajuste, com R^2 adj = 89,1%. A equação do modelo para a Rugosidade Média (R_a) da peça é apresentada, em unidades codificadas, na equação (54) (PEREIRA, 2006).

$$R_a = 0,357 + 0,016Vc + 0,134f - 0,008ap + 0,021Vc^2 + 0,066f^2 + 0,001ap^2 - 0,026Vc.f + 0,050Vc.ap - 0,017f.ap \quad (54)$$

Para a efetivação deste trabalho foram gerados, com o uso do software estatístico Minitab®, um arquivo com 620 conjuntos de dados (Vc , f , ap , R_a) a partir da equação (54), sendo R_a a Rugosidade Média da superfície usinada, em μm , prevista segundo o modelo utilizado. Os dados foram gerados seguindo uma distribuição uniforme entre os níveis do experimento original de Pereira (2006). Para cada parâmetro de corte, foram aplicados testes de Levene a todos os pares possíveis de conjuntos de treinamento, e verificou-se que, ao nível de significância de 0,05 não havia evidência de diferença de variância entre os conjuntos.

A partir do arquivo original contendo os 620 conjuntos de dados, foram gerados doze conjuntos de treinamento de diferentes tamanhos. Os conjuntos formados contêm 12, 15, 24, 30, 48, 60, 120, 150, 240, 300, 400 e 500 casos de treinamento cada um. Cada conjunto de treinamento é um sub-conjunto do arquivo original. Assim, o conjunto de treinamento contendo 12 casos é formado pelos 12 primeiros casos do arquivo original; o conjunto de treinamento contendo 15 casos é formado pelos 15 primeiros casos do arquivo original, e assim por diante. A todos os pares possíveis de conjuntos de treinamento foram aplicados testes de Levene e verificou-se que, ao nível de significância de 0,05 não havia evidência de diferença de variância entre os mesmos.

Dos 620 casos gerados a partir da equação (54), 120 foram reservados e utilizados unicamente como conjunto de testes. Estes casos não participaram do treinamento das redes e foram usados para testar de maneira independente o desempenho das redes na predição da Rugosidade Média (R_a) após a fase de treinamento.

5.6 Planejamento Experimental

5.6.1 Experimentos com Arranjos Ortogonais de Taguchi

Para o ciclo experimental inicial planejado segundo os arranjos ortogonais de Taguchi, foram realizados oito experimentos para vida da ferramenta e oito experimentos para rugosidade superficial. Cada experimento utilizou um conjunto de treinamento de tamanho diferente. Os conjuntos de treinamento utilizados contêm 24, 30, 48, 60, 240, 300, 400 e 500 casos de treinamento cada um.

As proporções utilizadas como níveis do fator ‘número de unidades radiais’ foram $\frac{1}{2}$ e um inteiro do número de casos de treinamento disponível em cada experimento. O número de níveis dos demais fatores foi definido como descrito no item 5.3.

O arranjo ortogonal utilizado para os experimentos foi um arranjo L8 de Taguchi com três fatores, sendo um fator com quatro níveis e dois fatores com dois níveis. O arranjo é ilustrado no quadro 5.1. Os fatores e seus respectivos níveis são sumarizados no quadro 5.2.

Número da corrida	Algoritmo para determinação do parâmetro de largura da função radial	Algoritmo para determinação de centros da função radial	Número de unidades radiais
1	1	1	1
2	1	2	2
3	2	1	1
4	2	2	2
5	3	1	2
6	3	2	1
7	4	1	2
8	4	2	1

Quadro 5.1 – Arranjo L8 para o ciclo experimental envolvendo arranjos ortogonais de Taguchi. Fonte: Minitab® 13.0

Fator	Número de níveis	Nível 1	Nível 2	Nível 3	Nível 4
Algoritmo para determinação do parâmetro de largura da função radial	4	Alg. Isotrópico - Fator de escala = 1	Alg. Isotrópico - Fator de escala = 10	K-Mais Próximos K= 5	K-Mais Próximos K=10
Algoritmo para determinação de centros da função radial	2	Sub-amostragem	K-Médias	-	-
Número de unidades radiais	2	Igual a ½ do número de casos de treinamento	Igual ao número de casos de treinamento	-	-

Quadro 5.2 – Fatores e níveis associados aos fatores dos experimentos envolvendo arranjos ortogonais de Taguchi

O quadro 5.3 explicita os valores absolutos do número de unidades radiais correspondentes aos níveis 1 e 2 para aquele fator, em cada experimento.

Número de casos no conjunto de treinamento	Nível 1	Nível 2
24	12	24
30	15	30
48	24	48
60	30	60
240	120	240
300	150	300
400	200	400
500	250	500

Quadro 5.3 – Número de unidades radiais em cada experimento envolvendo arranjos ortogonais de Taguchi

5.6.2 Experimentos com Fatoriais Completos Mistos

Para o ciclo experimental planejado com a utilização de fatoriais completos mistos, foram realizados doze experimentos para vida da ferramenta e doze experimentos para rugosidade superficial. Cada experimento utilizou um conjunto de treinamento de tamanho diferente. Os conjuntos de treinamento utilizados contêm 12, 15, 24, 30, 48, 60, 120, 150, 240, 300, 400 e 500 casos de treinamento cada um.

As proporções utilizadas como níveis do fator ‘número de unidades radiais’ foram $\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$ e um inteiro do número de casos de treinamento disponível em cada experimento. O número de níveis dos demais fatores foi definido como descrito no item 5.3.

O arranjo fatorial utilizado para cada experimento foi um arranjo fatorial completo com três fatores, dois deles distribuídos em quatro níveis e um fator distribuído em dois níveis, totalizando 32 corridas experimentais em cada experimento. O arranjo é ilustrado no quadro 5.4. Os fatores e seus respectivos níveis são sumarizados no quadro 5.5.

Número da corrida	Algoritmo para determinação do parâmetro de largura da função radial	Número de unidades radiais	Algoritmo para determinação de centros da função radial
1	1	1	1
2	1	1	2
3	1	2	1
4	1	2	2
5	1	3	1
6	1	3	2
7	1	4	1
8	1	4	2
9	2	1	1
10	2	1	2
11	2	2	1
12	2	2	2
13	2	3	1
14	2	3	2
15	2	4	1
16	2	4	2
17	3	1	1
18	3	1	2
19	3	2	1
20	3	2	2
21	3	3	1
22	3	3	2
23	3	4	1
24	3	4	2
25	4	1	1
26	4	1	2
27	4	2	1
28	4	2	2
29	4	3	1
30	4	3	2
31	4	4	1
32	4	4	2

Quadro 5.4 – Arranjo utilizado durante o ciclo experimental envolvendo fatoriais completos mistos.

Fonte: Minitab® 13.0

Fator	Número de níveis	Nível 1	Nível 2	Nível 3	Nível 4
Algoritmo para determinação do parâmetro de largura da função radial	4	Alg. Isotrópico - Fator de escala = 1	Alg. Isotrópico - Fator de escala = 10	K-Mais Próximos K= 5	K-Mais Próximos K=10
Número de unidades radiais	4	Igual a $\frac{1}{4}$ do número de casos de treinamento	Igual a $\frac{1}{2}$ do número de casos de treinamento	Igual a $\frac{3}{4}$ do número de casos de treinamento	Igual ao número de casos de treinamento
Algoritmo para determinação de centros da função radial	2	Sub-amostragem	K-Médias	-	-

Quadro 5.5 – Fatores e níveis associados aos fatores dos experimentos com a utilização de arranjos fatoriais completos mistos

O quadro 5.6 explicita os valores absolutos do número de unidades radiais correspondentes aos níveis 1, 2, 3 e 4, para aquele fator, em cada experimento.

Número de casos no conjunto de treinamento	Nível 1	Nível 2	Nível 3	Nível 4
12	3	6	9	12
15	4	8	12	15
24	6	12	18	24
30	8	15	23	30
48	12	24	36	48
60	15	30	45	60
120	30	60	90	120
150	38	75	113	150
240	60	120	180	240
300	75	150	225	300
400	100	200	300	400
500	125	250	375	500

Quadro 5.6 – Número de unidades radiais em cada experimento com a utilização de fatoriais completos mistos

5.6.3 Experimentos definitivos com Fatoriais Completos

Para o ciclo experimental definitivo planejado com a utilização de fatoriais completos, foram realizados doze experimentos para vida da ferramenta e doze experimentos para rugosidade superficial. Cada experimento utilizou um conjunto de treinamento de tamanho diferente. Os conjuntos de treinamento utilizados contêm 12, 15, 24, 30, 48, 60, 120, 150, 240, 300, 400 e 500 casos de treinamento cada um.

O arranjo fatorial utilizado para cada experimento foi um arranjo fatorial com três fatores distribuídos em dois níveis, totalizando oito corridas em cada experimento. Tal arranjo é ilustrado no quadro 5.7.

Número da corrida	Algoritmo para determinação do parâmetro de largura da função radial	Número de unidades radiais	Algoritmo para determinação de centros da função radial
1	-1	-1	-1
2	+1	-1	-1
3	-1	+1	-1
4	+1	+1	-1
5	-1	-1	+1
6	+1	-1	+1
7	-1	+1	+1
8	+1	+1	+1

Quadro 5.7 – Arranjo utilizado durante o ciclo experimental definitivo envolvendo fatoriais completos.
Fonte: Minitab® 13.0

Os níveis do fator ‘algoritmo para determinação de centros da função radial’ foram os mesmos utilizados durante o ciclo experimental anterior. Os níveis do fator ‘número de unidades radiais’ foram reduzidos a dois. As proporções utilizadas como níveis do fator em cada experimento correspondem aos dois níveis que apresentaram menor valor de mínimos quadrados no experimento correspondente do ciclo experimental realizado com o uso de fatoriais mistos. Os níveis do fator ‘algoritmo para determinação do parâmetro de largura da função radial’ foram reduzidos a dois. Da mesma forma os dois níveis remanescentes deste fator correspondem aos dois níveis que apresentaram menor valor de mínimos quadrados no experimento correspondente do ciclo experimental realizado com o uso de fatoriais mistos.

Assim sendo, os níveis dos fatores para os experimentos deste ciclo não foram os mesmos em todos os experimentos. Os níveis dos fatores para os experimentos relativos a vida da ferramenta com fatoriais completos são mostrados nos quadros 5.8 a 5.13. Os níveis dos fatores para os experimentos relativos à Rugosidade Média (R_a) com fatoriais completos são mostrados nos quadros 5.14 a 5.19.

	Tamanho do conjunto de treinamento			
	12		15	
	Níveis do fator:		Níveis do fator:	
	-1	+1	-1	+1
Algoritmo para determinação do parâmetro de largura da função radial	Alg. Isotrópico - Fator de escala = 10	K-Mais Próximos K=10	Alg. Isotrópico - Fator de escala = 10	K-Mais Próximos K=10
Número de unidades radiais	6	9	8	12
Algoritmo para determinação de centros da função radial	Sub-amostragem	K-Médias	Sub-amostragem	K-Médias

Quadro 5.8 – Níveis dos fatores utilizados nos experimentos definitivos com fatorial completo para vida da ferramenta, experimentos com 12 e 15 casos de treinamento

	Tamanho do conjunto de treinamento			
	24		30	
	Níveis do fator:		Níveis do fator:	
	-1	+1	-1	+1
Algoritmo para determinação do parâmetro de largura da função radial	Alg. Isotrópico - Fator de escala = 10	K-Mais Próximos K=10	Alg. Isotrópico - Fator de escala = 10	K-Mais Próximos K=10
Número de unidades radiais	12	18	15	23
Algoritmo para determinação de centros da função radial	Sub-amostragem	K-Médias	Sub-amostragem	K-Médias

Quadro 5.9 – Níveis dos fatores utilizados nos experimentos definitivos com fatorial completo para vida da ferramenta, experimentos com 24 e 30 casos de treinamento

	Tamanho do conjunto de treinamento			
	48		60	
	Níveis do fator:		Níveis do fator:	
	-1	+1	-1	+1
Algoritmo para determinação do parâmetro de largura da função radial	Alg. Isotrópico - Fator de escala = 10	K-Mais Próximos K=10	Alg. Isotrópico - Fator de escala = 10	K-Mais Próximos K=10
Número de unidades radiais	24	36	15	30
Algoritmo para determinação de centros da função radial	Sub-amostragem	K-Médias	Sub-amostragem	K-Médias

Quadro 5.10 – Níveis dos fatores utilizados nos experimentos definitivos com fatorial completo para vida da ferramenta, experimentos com 48 e 60 casos de treinamento

	Tamanho do conjunto de treinamento			
	120		150	
	Níveis do fator:		Níveis do fator:	
	-1	+1	-1	+1
Algoritmo para determinação do parâmetro de largura da função radial	Alg. Isotrópico - Fator de escala = 10	K-Mais Próximos K=10	Alg. Isotrópico - Fator de escala = 10	K-Mais Próximos K=10
Número de unidades radiais	30	60	38	75
Algoritmo para determinação de centros da função radial	Sub-amostragem	K-Médias	Sub-amostragem	K-Médias

Quadro 5.11 – Níveis dos fatores utilizados nos experimentos definitivos com fatorial completo para vida da ferramenta, experimentos com 120 e 150 casos de treinamento

	Tamanho do conjunto de treinamento			
	240		300	
	Níveis do fator:		Níveis do fator:	
	-1	+1	-1	+1
Algoritmo para determinação do parâmetro de largura da função radial	Alg. Isotrópico - Fator de escala = 1	Alg. Isotrópico - Fator de escala = 10	Alg. Isotrópico - Fator de escala = 10	K-Mais Próximos K=5
Número de unidades radiais	60	120	75	150
Algoritmo para determinação de centros da função radial	Sub-amostragem	K-Médias	Sub-amostragem	K-Médias

Quadro 5.12 – Níveis dos fatores utilizados nos experimentos definitivos com fatorial completo para vida da ferramenta, experimentos com 240 e 300 casos de treinamento

	Tamanho do conjunto de treinamento			
	400		500	
	Níveis do fator:		Níveis do fator:	
	-1	+1	-1	+1
Algoritmo para determinação do parâmetro de largura da função radial	Alg. Isotrópico - Fator de escala = 10	K-Mais Próximos K=10	Alg. Isotrópico - Fator de escala = 1	Alg. Isotrópico - Fator de escala = 10
Número de unidades radiais	100	200	125	250
Algoritmo para determinação de centros da função radial	Sub-amostragem	K-Médias	Sub-amostragem	K-Médias

Quadro 5.13 – Níveis dos fatores utilizados nos experimentos definitivos com fatorial completo para vida da ferramenta, experimentos com 400 e 500 casos de treinamento

	Tamanho do conjunto de treinamento			
	12		15	
	Níveis do fator:		Níveis do fator:	
	-1	+1	-1	+1
Algoritmo para determinação do parâmetro de largura da função radial	Alg. Isotrópico - Fator de escala = 10	K-Mais Próximos K=10	Alg. Isotrópico - Fator de escala = 10	K-Mais Próximos K=10
Número de unidades radiais	6	9	8	12
Algoritmo para determinação de centros da função radial	Sub-amostragem	K-Médias	Sub-amostragem	K-Médias

Quadro 5.14 – Níveis dos fatores utilizados nos experimentos definitivos com fatorial completo para Rugosidade Média, experimentos com 12 e 15 casos de treinamento

	Tamanho do conjunto de treinamento			
	24		30	
	Níveis do fator:		Níveis do fator:	
	-1	+1	-1	+1
Algoritmo para determinação do parâmetro de largura da função radial	Alg. Isotrópico - Fator de escala = 10	K-Mais Próximos K=10	Alg. Isotrópico - Fator de escala = 10	K-Mais Próximos K=10
Número de unidades radiais	12	18	15	23
Algoritmo para determinação de centros da função radial	Sub-amostragem	K-Médias	Sub-amostragem	K-Médias

Quadro 5.15 – Níveis dos fatores utilizados nos experimentos definitivos com fatorial completo para Rugosidade Média, experimentos com 24 e 30 casos de treinamento

	Tamanho do conjunto de treinamento			
	48		60	
	Níveis do fator:		Níveis do fator:	
	-1	+1	-1	+1
Algoritmo para determinação do parâmetro de largura da função radial	Alg. Isotrópico - Fator de escala = 10	K-Mais Próximos K=10	Alg. Isotrópico - Fator de escala = 10	K-Mais Próximos K=5
Número de unidades radiais	24	36	30	45
Algoritmo para determinação de centros da função radial	Sub-amostragem	K-Médias	Sub-amostragem	K-Médias

Quadro 5.16 – Níveis dos fatores utilizados nos experimentos definitivos com fatorial completo para Rugosidade Média, experimentos com 48 e 60 casos de treinamento

	Tamanho do conjunto de treinamento			
	120		150	
	Níveis do fator:		Níveis do fator:	
	-1	+1	-1	+1
Algoritmo para determinação do parâmetro de largura da função radial	Alg. Isotrópico - Fator de escala = 1	Alg. Isotrópico - Fator de escala = 10	Alg. Isotrópico - Fator de escala = 10	K-Mais Próximos K=10
Número de unidades radiais	30	60	38	75
Algoritmo para determinação de centros da função radial	Sub-amostragem	K-Médias	Sub-amostragem	K-Médias

Quadro 5.17 – Níveis dos fatores utilizados nos experimentos definitivos com fatorial completo para Rugosidade Média, experimentos com 120 e 150 casos de treinamento

	Tamanho do conjunto de treinamento			
	240		300	
	Níveis do fator:		Níveis do fator:	
	-1	+1	-1	+1
Algoritmo para determinação do parâmetro de largura da função radial	Alg. Isotrópico - Fator de escala = 10	K-Mais Próximos K=5	Alg. Isotrópico - Fator de escala = 10	K-Mais Próximos K=10
Número de unidades radiais	60	120	75	150
Algoritmo para determinação de centros da função radial	Sub-amostragem	K-Médias	Sub-amostragem	K-Médias

Quadro 5.18 – Níveis dos fatores utilizados nos experimentos definitivos com fatorial completo para Rugosidade Média, experimentos com 240 e 300 casos de treinamento

	Tamanho do conjunto de treinamento			
	400		500	
	Níveis do fator:		Níveis do fator:	
	-1	+1	-1	+1
Algoritmo para determinação do parâmetro de largura da função radial	Alg. Isotrópico - Fator de escala = 1	Alg. Isotrópico - Fator de escala = 10	Alg. Isotrópico - Fator de escala = 10	K-Mais Próximos K=10
Número de unidades radiais	100	200	125	250
Algoritmo para determinação de centros da função radial	Sub-amostragem	K-Médias	Sub-amostragem	K-Médias

Quadro 5.19 – Níveis dos fatores utilizados nos experimentos definitivos com fatorial completo para Rugosidade Média, experimentos com 400 e 500 casos de treinamento

5.7 Execução dos experimentos

Todos os experimentos foram conduzidos com a utilização do software Statística®, versão 7.1. A máquina utilizada foi um micro-computador dotado de processador Intel Pentium 4® de 1.7 GHz e 256 MB de memória RAM, configurado com o sistema operacional Microsoft Windows 2000®. Procedimentos automatizados (chamados de macros) do Statística® foram desenvolvidos para a execução de cada experimento. Cada procedimento contém: a configuração da rede sob teste, os comandos para a iniciação aleatória da mesma, o endereço do arquivo contendo os casos de treinamento a serem utilizados para a condução daquele experimento específico, os comandos para a execução da simulação e o endereço para escrita do arquivo contendo os resultados do experimento. Uma vez terminada sua confecção, tais procedimentos podem ser armazenados e executados pelo número de vezes que seja necessário. O uso dos procedimentos automatizados em muito incrementou a produtividade do trabalho.

Segundo Sick (2002), a falta de repetições é um dos fatores que impede uma eficaz análise estatística que sustente as conclusões dos estudos realizados, envolvendo redes neurais na tarefa de predição em processos de torneamento. Para evitar-se tal problema, cada configuração de rede foi iniciada, treinada e testada de maneira independente por um certo número de vezes em cada ciclo experimental, utilizando-se os procedimentos automatizados. Nos experimentos envolvendo arranjos ortogonais de Taguchi o número de repetições adotado foi de cinquenta. Nos demais experimentos, planejados com a utilização de fatoriais completos, esse número foi de sessenta repetições.

Em relação às etapas de pré e pós-processamento, os dados foram linearmente escalonados. Para os dados de entrada foi utilizado o algoritmo *Minimax* (Manual Statística, 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. Para o pós-processamento dos dados, utilizou-se o algoritmo Média/Desvio Padrão (Manual Statística, 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. O efeito prático foi o de converter os dados de entrada para o domínio das redes e os dados de saída de volta para o domínio do problema sob estudo.

Uma vez terminada a configuração de todos os procedimentos automatizados para um dado ciclo experimental e verificado seu correto funcionamento e sua conformidade com o planejamento experimental adotado, procedia-se à execução dos experimentos. Em virtude do automatismo do processo, cada um dos três ciclos experimentais foi realizado em bloco único. Cada ciclo foi executado do início ao fim, sem interrupções, em uma única máquina. Por esse motivo a técnica de blocagem não foi adotada na análise dos resultados experimentais.

A saída dos experimentos se deu sob a forma de arquivos de dados produzidos pelo software Statistica®. Os arquivos contêm os resultados da predição das redes para os casos de teste utilizados, expressos em Razão de Desvios Padrão, que foi a grandeza utilizada para medir-se a exatidão, estabilidade e variabilidade das predições. Os dados contidos nos arquivos gerados foram tabulados e inseridos no software estatístico Minitab 13.0®, software no qual foram realizadas as análises estatísticas pertinentes a cada experimento conduzido.

6 ANÁLISE DOS RESULTADOS PARA VIDA DA FERRAMENTA

6.1 Considerações Iniciais

Neste capítulo são apresentados os resultados obtidos para vida da ferramenta. Inicialmente são apresentados os resultados obtidos do ciclo experimental envolvendo arranjos ortogonais de Taguchi e discutidos seus resultados. A seguir são apresentados os resultados obtidos do ciclo experimental planejado com arranjos fatoriais mistos. Finalmente são apresentados os resultados obtidos durante o ciclo experimental definitivo realizado com a utilização de arranjos fatoriais completos. Os resultados deste último ciclo experimental são discutidos em detalhes. A análise dos resultados foi realizada com a utilização do software estatístico Minitab®, versão 13.0.

6.2 Resultados dos experimentos com arranjos ortogonais de Taguchi

A tabela 6.1 traz as médias da resposta Razão de Desvios Padrão para todas as corridas dos oito experimentos realizados. A tabela 6.2 traz os desvios padrão associados às respostas das corridas.

Fez-se a análise prevista pelo método Taguchi para as médias da resposta, para valor da relação sinal-ruído e para o desvio padrão. A análise considera a dimensão da diferença entre o maior e o menor efeito calculado para o nível de um fator.

Para os experimentos realizados, a análise fornece os valores dos efeitos principais de cada fator sobre cada uma das grandezas sob análise (média da resposta, relação sinal-ruído e variância), bem como o *ranking* dos impactos dos fatores sobre aquelas grandezas. Estas informações estão sumarizadas nas tabelas contidas nos arquivos anexados a este trabalho.

Nota-se que, em todos os casos e para as três análises, o algoritmo para determinação do parâmetro de largura da função radial foi indicado como sendo o fator mais influente. O número de unidades radiais na camada oculta é indicado como o segundo fator mais importante para redução da variância em todos os experimentos. Em relação aos valores da média das respostas e do valor da relação sinal-ruído, nota-se se uma alternância entre os fatores na segunda e terceira colocações entre os experimentos.

Tabela 6.1 – Valores médios da Razão de Desvios Padrão obtidos durante os experimentos envolvendo arranjos ortogonais de Taguchi para vida da ferramenta

Número da corrida	Número de casos no conjunto de treinamento							
	24	30	48	60	240	300	400	500
1	0,506837	0,471353	0,476249	0,447445	0,465651	0,496682	0,567953	0,590641
2	6,522662	34,945944	34,990795	30,443682	12,141536	121,962782	202,662244	6,133433
3	0,018459	0,013033	0,007509	0,005075	0,000136	0,000052	0,000029	0,000026
4	0,037715	0,014187	0,000882	0,048351	0,000045	0,000070	0,000057	0,000030
5	0,216641	2,239549	0,105665	6,083101	2,866522	22,883455	133,477726	0,268465
6	0,183767	0,136221	0,123134	0,118419	0,057907	0,051184	0,054017	0,058393
7	0,095379	0,122809	0,048547	0,051541	0,178457	130,204854	0,016078	0,303463
8	0,122167	0,086081	0,069006	0,058230	0,025310	0,021892	0,021095	0,019557

Tabela 6.2 – Valores do desvio padrão da resposta Razão de Desvios Padrão obtidos durante os experimentos envolvendo arranjos ortogonais de Taguchi para vida da ferramenta

Número da corrida	Número de casos no conjunto de treinamento							
	24	30	48	60	240	300	400	500
1	0,121496	0,126519	0,070439	0,071011	0,045228	0,038607	0,027564	0,025632
2	38,259227	115,005910	132,711707	152,159041	44,944093	350,666782	785,570957	36,013976
3	0,007346	0,005115	0,001654	0,001828	4,840E-05	2,456E-05	1,411E-05	1,409E-05
4	0,030172	2,661E-10	0,000108	0,137427	2,438E-05	0,000128	3,773E-05	1,821E-05
5	0	0	4,992E-09	0	0	7,707E-07	3,083E-06	0
6	0,054226	0,032440	0,038706	0,025966	0,010930	0,010314	0,011293	0,009779
7	2,607E-09	0	0	1,683E-09	5,214E-09	0	5,645E-10	7,374E-09
8	0,060683	0,023477	0,022489	0,013327	0,006848	0,005653	0,005642	0,004112

As figuras 6.1 a 6.3 mostram, a título de exemplo, os gráficos dos efeitos principais para o experimento conduzido com o conjunto de treinamento contendo 300 casos. Cada figura é dividida em três partes. Cada parte está relacionada a um fator, indicado na parte superior da figura. A parte inferior mostra os níveis para cada fator. Por eles pode-se ter uma idéia da importância relativa de cada um dos efeitos. Quanto maior a diferença entre os valores dos efeitos principais em dado nível de um fator sobre a grandeza analisada (valores médios de Razão de Desvios Padrão, relação sinal-ruído e desvio padrão da resposta), maior a influência do mesmo. Através das figuras pode-se ver também os níveis que a análise indica serem aqueles que levarão a rede ao melhor desempenho. O conjunto dos gráficos relativos ao ciclo experimental envolvendo arranjos ortogonais de Taguchi pode ser visto nos arquivos anexados a este trabalho.

Em concordância com o que é mostrado nas figuras 6.1 a 6.3 e analisando as tabelas resultantes da análise de Taguchi, os níveis dos fatores apontados como sendo os que mais contribuem para a minimização da resposta e do desvio padrão e para a maximização da robustez do sistema são os sumarizados na tabela 6.3. Para os valores médios da Razão de

Desvios Padrão encontrou-se que, exceto para um tamanho do conjunto de treinamento, as combinações apontadas são as mesmas. No caso da relação sinal-ruído, em seis dos casos a configuração apontada como sendo a que leva à maior robustez é a mesma. Em se tratando da variância, a configuração apontada como sendo a que mais contribui para sua redução foi a mesma para todos os experimentos.

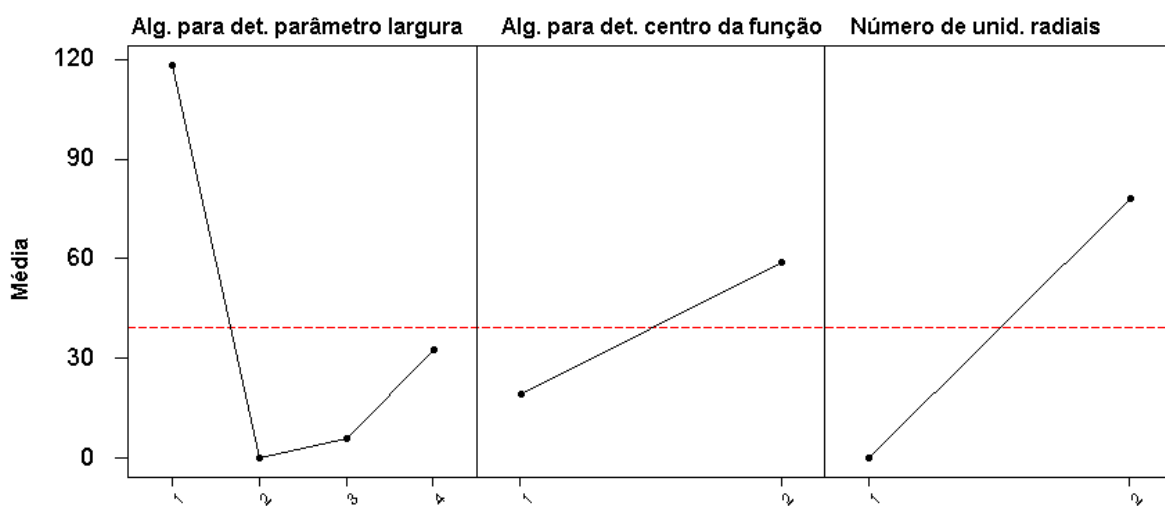


Figura 6.1 – Efeitos principais sobre a média da resposta Razão de Desvios Padrão para conjunto de treinamento de 300 casos. Fonte: Minitab 13.0®

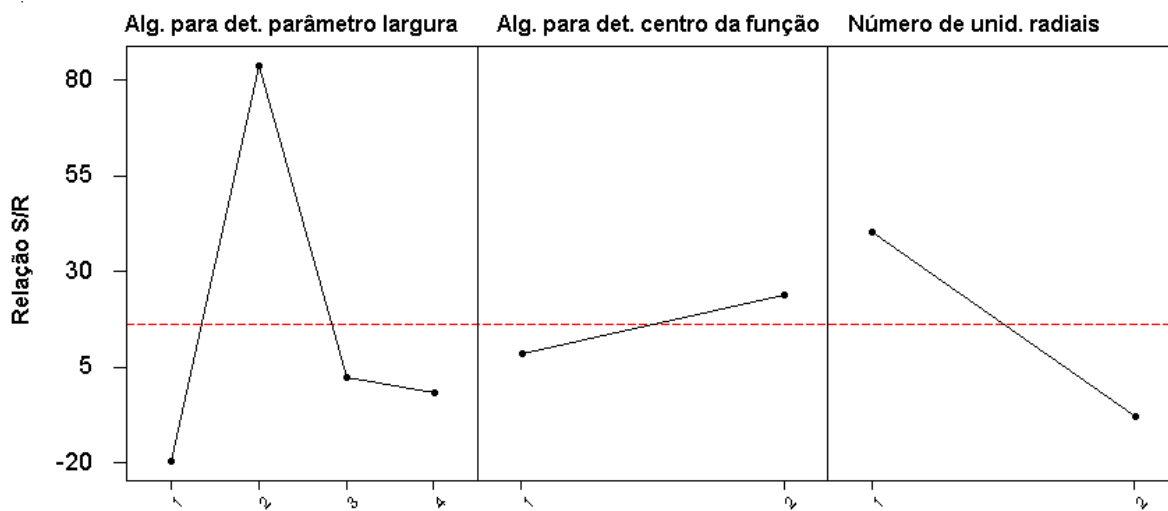


Figura 6.2 – Efeitos principais sobre a função sinal-ruído para conjunto de treinamento de 300 casos. Fonte: Minitab 13.0®

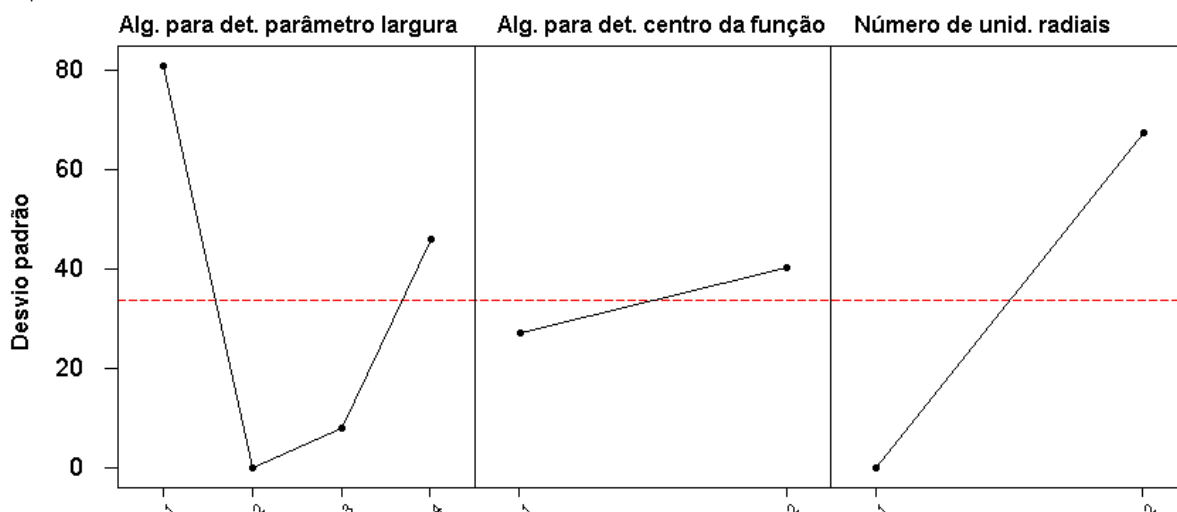


Figura 6.3 – Efeitos principais sobre o desvio padrão da resposta Razão de Desvios Padrão para conjunto de treinamento de 300 casos. Fonte: Minitab 13.0®

Tabela 6.3 – Configurações apontadas como as melhores para predição da vida da ferramenta, para cada experimento envolvendo arranjos ortogonais de Taguchi

Número de casos no conjunto de treinamento	Combinação para valores médios da Razão de Desvios Padrão	Combinação para valores da relação sinal-ruído	Combinação para valores do desvio padrão
24	2:1:1	2:1:1	2:1:1
30	2:1:1	2:1:1	2:1:1
48	2:1:1	2:1:1	2:1:1
60	4:1:1	2:1:1	2:1:1
240	2:1:1	2:2:1	2:1:1
300	2:1:1	2:2:1	2:1:1
400	2:1:1	2:1:1	2:1:1
500	2:1:1	2:1:1	2:1:1

Para cada experimento foram aplicados testes de hipótese da estatística Z (Montgomery e Runger, 2003) para inferência sobre a média de uma população. Testes da estatística t de Student e análise de variância não foram utilizados porque testes estatísticos preliminares não apresentaram evidência estatística de igualdade de variância entre as amostras, nem de as mesmas possuírem uma distribuição normal, o que seria condição para utilização daquelas técnicas. Dentro de um dado experimento, cada amostra foi comparada por meio do teste Z ao valor da média de cada uma das demais amostras ao nível de significância de 0,05. Testou-se, ao nível de significância de 0,05 a hipótese nula de que a média da amostra testada fosse igual ao valor da outra média. O objetivo destes testes foi o de estabelecer estatisticamente qual configuração apresentou a menor média da resposta Razão de Desvios Padrão dentre as configurações testadas.

Fazendo-se uma comparação entre os resultados obtidos através da análise de Taguchi e aqueles obtidos nos testes Z foram encontradas algumas discrepâncias. Nos experimentos com 48 e 240 casos de treinamento, a análise apontou para uma configuração que não era a que possuía a menor média da resposta entre as corridas daqueles experimentos. Percebeu-se também que, no experimento com 60 casos de treinamento, a análise indicou uma configuração que não foi testada durante o procedimento experimental.

Para estes casos discrepantes, fez-se nova comparação estatística com auxílio do teste Z. Foram comparadas as amostras das configurações indicadas pela análise de Taguchi com as médias das corridas apontadas pelo primeiro teste Z como as mais baixas do experimento em questão. Para o caso do experimento com 60 casos de treinamento, em que a análise de Taguchi apontou para uma configuração que não fazia parte do arranjo ortogonal, montou-se a configuração indicada pela análise e realizou-se uma corrida experimental extra, com o conjunto de treinamento apropriado e com o mesmo número de repetições das demais corridas. Após isso, fez-se o segundo teste, cujos resultados são mostrados em tabela contida nos arquivos anexados a este trabalho. Testou-se, ao nível de significância de 0,05 a hipótese nula de a média da amostra indicada pela análise de Taguchi ser igual ao valor da média da configuração com melhor desempenho durante as corridas experimentais. Do resultado, nota-se que as configurações indicadas pela análise de Taguchi possuem médias de Razão de Desvios Padrão maiores (e portanto, desempenho pior) que as médias das melhores configurações experimentais, como se verifica pelo valor P igual a zero. Tal fato pode indicar a existência de interações entre os fatores do experimento, o que não pôde ser confirmado devido ao arranjo ortogonal escolhido para os experimentos.

Assim sendo, as melhores configurações obtidas para a predição da vida da ferramenta durante os experimentos com arranjos ortogonais de Taguchi foram as mostradas na tabela 6.4. Mostram-se também os respectivos valores médios da resposta Razão de Desvios Padrão bem como os valores da relação sinal-ruído e desvios padrão da resposta para tais configurações.

Em relação à tabela 6.4, a simples observação da mesma não permite afirmar que uma resposta difere de outra. Com o propósito de comparar estatisticamente as melhores configurações obtidas, foram aplicados testes para comparação de médias e variâncias para todos os pares possíveis de configurações.

Tabela 6.4 – Configurações de níveis com melhor desempenho para os experimentos envolvendo arranjos ortogonais de Taguchi e respectivos valores de Razão de Desvios Padrão para predição da vida da ferramenta

Número de casos no conjunto de treinamento	Algoritmo para determinação do parâmetro de largura da função radial	Algoritmo para determinação dos centros da função radial	Número de unidades radiais	Valor médio da resposta Razão de Desvios Padrão	Desvio Padrão
24	Alg. isotrópico Fator de escala = 10	Sub-amostragem	12	0,018459	0,007346
30	Alg. isotrópico Fator de escala = 10	Sub-amostragem	15	0,013033	0,005115
48	Alg. isotrópico Fator de escala = 10	K-Médias	48	0,000882	0,000108
60	Alg. isotrópico Fator de escala = 10	Sub-amostragem	30	0,005075	0,001828
240	Alg. isotrópico Fator de escala = 10	K-Médias	240	0,000045	2,438E-05
300	Alg. isotrópico Fator de escala = 10	K-Médias	150	0,000052	2,456E-05
400	Alg. isotrópico Fator de escala = 10	Sub-amostragem	200	0,000029	1,411E-05
500	Alg. isotrópico Fator de escala = 10	Sub-amostragem	250	0,000026	1,409E-05

Para comparação entre as médias dos valores da resposta Razão de Desvios Padrão obtidas pelas melhores configurações recorreu-se outra vez a um teste de inferência sobre a média de uma população (teste Z). Foram comparadas, ao nível de significância de 0,05, as médias de Razão de Desvios Padrão da melhor configuração obtida em um experimento com o valor médio obtido no experimento com o conjunto de treinamento de tamanho imediatamente inferior. Os resultados destes testes podem ser observados na tabela 6.5.

Tabela 6.5 – Valores P resultantes do teste para inferência sobre a média entre pares de amostras para as melhores configurações obtidas em cada experimento envolvendo arranjos ortogonais de Taguchi para vida da ferramenta

Número de casos no conjunto de treinamento maior	Valores médios da resposta para o conjunto de treinamento maior	Número de casos no conjunto de treinamento menor	Valores médios da resposta para o conjunto de treinamento menor	Valor P
30	0,013033	24	0,018459	0,000
48	0,000882	30	0,013033	0,000
60	0,005075	48	0,000882	0,000
240	0,000045	60	0,005075	0,000
300	0,000052	240	0,000045	0,060
400	0,000029	300	0,000052	0,000
500	0,000026	400	0,000029	0,192

Da análise da tabela 6.5, nota-se que existe evidência, ao nível de significância adotado, de que a média da resposta Razão de Desvios Padrão para a melhor configuração obtida para 30 casos é inferior à média da melhor configuração obtida para o conjunto contendo 24 casos de treinamento. Da mesma maneira existe evidência de que a média da melhor configuração para o conjunto de treinamento contendo 48 casos é inferior àquela da melhor configuração obtida para conjunto de 30 casos de treinamento.

Nota-se, porém, que a melhor configuração obtida com o conjunto de treinamento contendo 60 casos apresenta forte evidência de que sua média é superior àquela obtida da melhor configuração para o experimento envolvendo 48 casos. A partir desse ponto, nota-se que os valores médios de Razão de Desvios Padrão obtidos pela melhor configuração para um determinado número de casos no conjunto de treinamento apresentam evidência de que sua média é inferior à obtida por aquela do conjunto de treinamento de tamanho imediatamente inferior. Isso é verdade até a comparação entre as médias das respostas das melhores configurações obtidas para os conjuntos de treinamento contendo 400 e 500 casos, que não apresentam evidência de diferença significativa entre si.

Como a melhor configuração de rede obtida para o conjunto contendo 60 casos de teste apresentou média superior à obtida pela melhor configuração obtida para 48 casos, repetiu-se o teste comparando a média obtida entre a melhor configuração de rede para 60 casos e aquela obtida pela melhor configuração obtida para 30 casos, ao nível de significância de 0,05. O resultado é mostrado em tabela contida nos arquivos anexados a este trabalho. Desta última comparação observa-se que há forte evidência, ao nível de significância adotado, de que a rede treinada com 60 casos apresenta média de Razão de Desvios Padrão inferior à rede treinada com 30 casos.

Para testar as variâncias entre as configurações, foram aplicados testes de Levene para hipótese nula de variâncias iguais entre todos os pares de respostas possíveis da tabela 6.4, ao nível de significância de 0,05. Optou-se por esse teste para comparação entre variâncias pelo fato de testes preliminares indicarem não haver evidência de que as amostras da tabela 6.4 sigam a distribuição normal. Os resultados dos testes de Levene podem ser observados na tabela 6.6, na qual um valor P superior ao nível de significância adotado indica que não há evidência estatística de diferença entre as variâncias e um valor P inferior a 0,05 indica forte evidência de diferença entre as variâncias.

A análise da tabela 6.6 indica que não há evidência de diferença na variância entre as melhores configurações de rede obtidas para os conjuntos de treinamento com 24 e 30 casos.

Os dois conjuntos, no entanto, possuem uma variância maior que a de todos os demais conjuntos.

O resultado dos testes evidencia que não há, ao nível de significância adotado, diferença entre a variância da melhor configuração obtida para o conjunto de treinamento com 48 casos e a variância das melhores configurações de rede obtidas para os conjuntos contendo 240, 300, 400 e 500 casos. Há também forte evidência de que a melhor configuração obtida para 48 casos possui variância menor que a variância do conjunto contendo 60 casos. Nota-se que a variância obtida para este conjunto é bastante reduzida. A melhor configuração obtida para 60 casos, por seu turno, apresenta forte evidência de variância superior aos conjuntos com maior número de casos.

Tabela 6.6 – Valores P resultantes do teste de Levene entre pares de amostras para as melhores configurações obtidas em cada experimento envolvendo arranjos ortogonais de Taguchi para vida da ferramenta

	Número de casos no conjunto de treinamento	24	30	48	60	240	300	400
Número de casos no conjunto de treinamento	Variâncias observadas	5,4E-05	2,6E-05	1,1E-08	3,3E-06	5,9E-10	6,0E-10	1,99E-10
		Valores P obtidos do teste de Levene						
30	2,6E-05	0,094	-	-	-	-	-	-
48	1,1E-08	0,000	0,000	-	-	-	-	-
60	3,3E-06	0,000	0,000	0,000	-	-	-	-
240	5,9E-10	0,000	0,000	0,966	0,000	-	-	-
300	6,0E-10	0,000	0,000	0,930	0,000	0,587	-	-
400	1,99E-10	0,000	0,000	0,619	0,000	0,030	0,002	-
500	1,99E-10	0,000	0,000	0,666	0,000	0,056	0,004	0,591

No que diz respeito à melhor configuração de rede obtida para o conjunto contendo 240 casos de treinamento, o teste fornece evidência estatística de que não há diferença entre sua variância e aquela dos conjuntos contendo 300 e 500 casos de treinamento. O teste sugere que a variância deste conjunto é superior à variância observada no conjunto contendo 400 casos, embora o valor P observado, igual a 0,03, não esteja distante do nível de significância adotado.

O resultado dos testes de Levene para a melhor configuração obtida para o conjunto contendo 300 casos de treinamento fornece evidência, ao nível de significância de 0,05 de que a sua variância é superior à observada nos conjuntos contendo 400 e 500 casos. Já em relação ao conjunto contendo 400 casos de treinamento, o teste não evidenciou diferença significativa

entre a variância da melhor configuração obtida para o conjunto e aquela do conjunto contendo 500 casos.

Com exceção do experimento envolvendo 60 casos de treinamento, para as melhores configurações obtidas, observou-se uma tendência à redução da média de Razão de Desvios Padrão e da variância à medida que aumentou o número de casos de treinamento disponível para a rede. Isso sugere um melhor desempenho das redes na predição da vida da ferramenta à medida que aumenta o número de casos de treinamento disponíveis. O gráfico de *box-plot* para as melhores configurações obtidas é mostrado na figura 6.4. Os pontos indicam a média da resposta da melhor configuração de rede obtida para um conjunto de treinamento. Os asteriscos indicam *outliers*. Os limites de cada caixa indicam, a partir de baixo, o 1º, 2º e 3º quartis das observações de predição de cada rede. Na figura pode-se visualizar que a média e a dispersão são maiores nos experimentos envolvendo 24 e 30 casos de teste, e que a média cai com o aumento do número de casos disponíveis para treinamento. Nota-se também o incremento no valor da média e na dispersão observados na melhor configuração obtida para o experimento envolvendo 60 casos de treinamento.

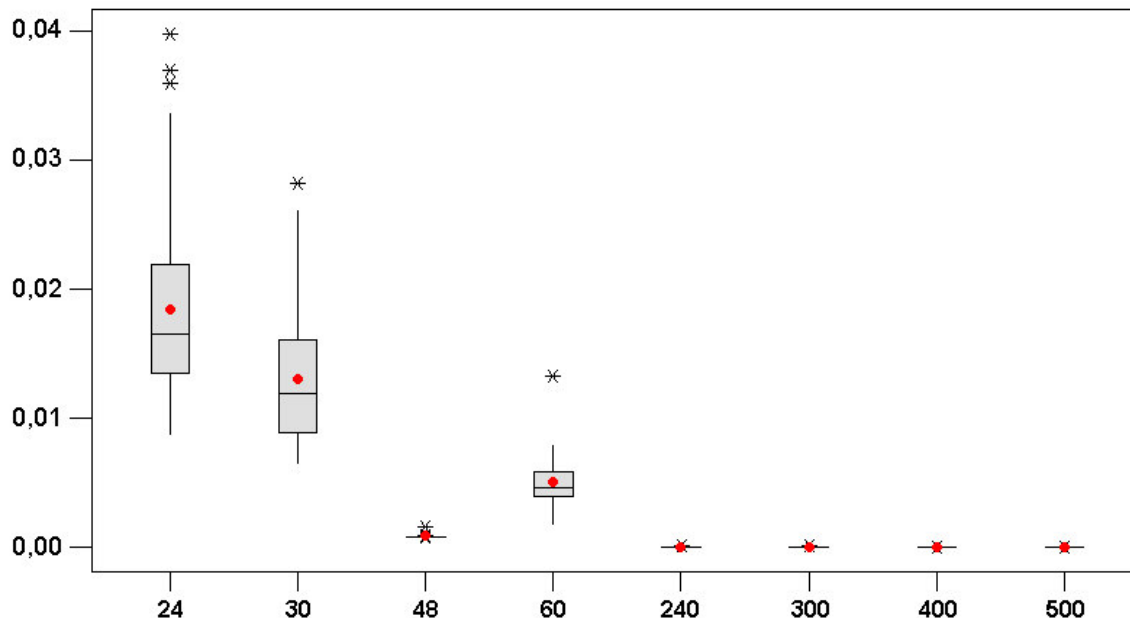


Figura 6.4 – *Box-plot* da resposta Razão de Desvios Padrão para as melhores configurações obtidas em cada experimento envolvendo arranjos ortogonais de Taguchi para vida da ferramenta. Fonte: Minitab 13.0®

Os resultados deste ciclo experimental indicam que mesmo para situações envolvendo um número pequeno de casos de treinamento (como o que envolve apenas 24 casos), as redes

apresentam um bom desempenho na tarefa de predição para os casos de teste apresentados, como evidenciam os valores mostrados pela tabela 6.4. Isso sugere que redes RBF podem constituir-se numa alternativa válida e economicamente viável para a tarefa de predição da vida da ferramenta. Além disso, as discrepâncias entre os resultados experimentais e os resultados da análise indicam a possibilidade da existência de interações entre os níveis dos fatores, o que não pôde ser confirmado devido ao arranjo ortogonal utilizado. Os resultados também levantaram dúvidas sobre qual seria o desempenho das redes com conjuntos contendo menos que 24 casos de treinamento e também contendo números de casos entre 60 e 240 casos. Questionou-se também se outras proporções entre número de unidades radiais e o número de casos de treinamento poderiam levar as redes a apresentar melhores resultados.

Tais resultados e as questões levantadas durante o experimento levaram ao planejamento e execução do segundo ciclo experimental com a utilização de fatoriais completos mistos, com um maior número de conjuntos de treinamento e mais níveis para o fator número de unidades radiais.

6.3 Resultados dos experimentos com Fatoriais Completos mistos

O segundo ciclo experimental teve por objetivos estimar o valor dos efeitos dos fatores sobre o desempenho das redes para predição da vida da ferramenta, estimar o efeito de possíveis interações entre os níveis dos fatores sob teste e avaliar o desempenho das redes com a utilização de mais níveis do fator número de unidades radiais.

Com efeito, os experimentos deste ciclo permitiram estimar os níveis dos fatores que mais influem na resposta Razão de Desvios Padrão, bem como estimar a existência de interações entre os fatores. As figuras 6.5 e 6.6 mostram os efeitos principais e as interações obtidas dos experimentos envolvendo um conjunto de treinamento composto por 400 casos.

Os gráficos dos efeitos principais e de interação para os demais experimentos deste ciclo com a utilização de arranjos fatoriais mistos podem ser encontrados nos arquivos anexados a este trabalho.

A tabela 6.7 contém o resultado da análise de variância realizada para o experimento envolvendo 400 casos de treinamento. As tabelas contendo o resultado da análise de variância para os demais experimentos realizados durante este ciclo podem ser encontradas nos arquivos anexados a este trabalho.

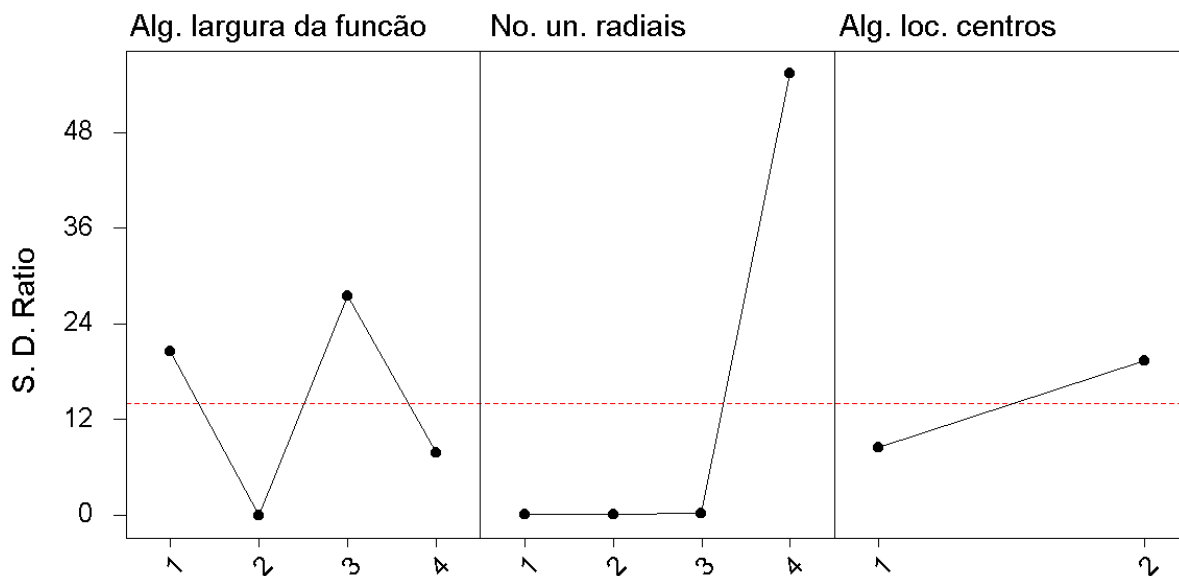


Figura 6.5 – Efeitos principais para a resposta Razão de Desvios Padrão, para experimento com arranjos fatoriais mistos envolvendo 400 casos de treinamento para vida da ferramenta. Fonte: Minitab 13.0®

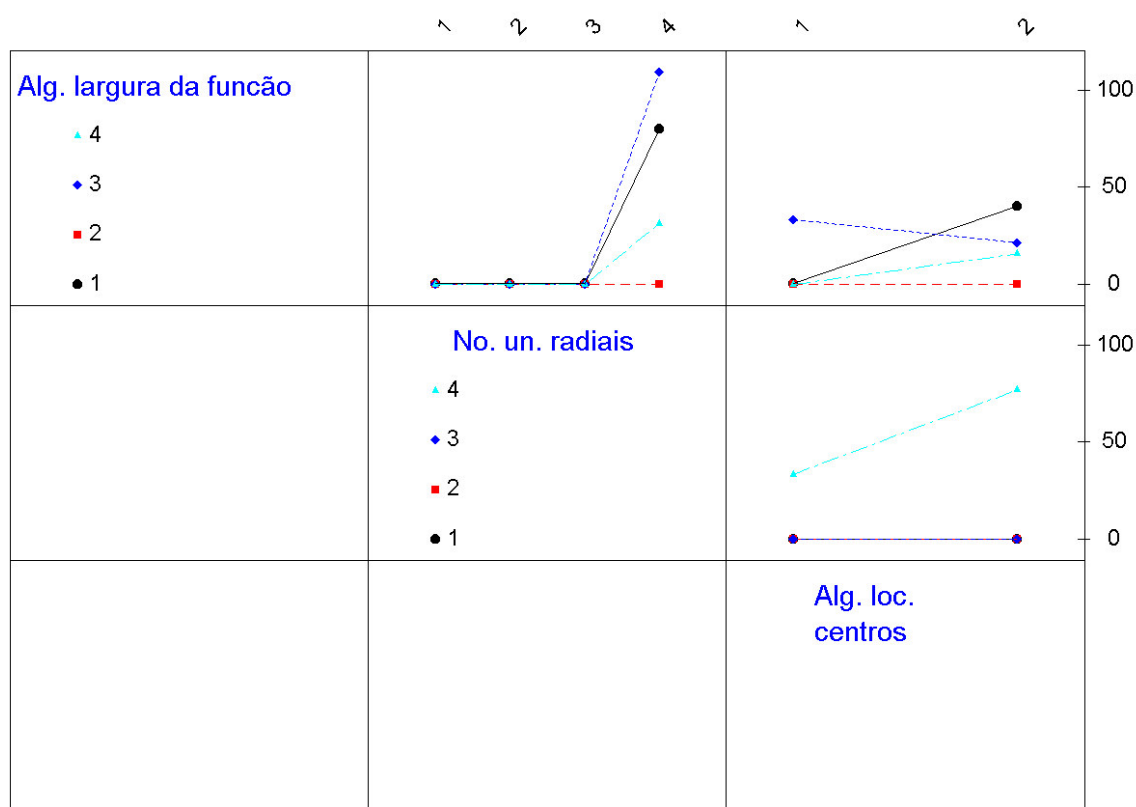


Figura 6.6 – Efeitos de interação para a resposta Razão de Desvios Padrão, para experimento com arranjos fatoriais mistos envolvendo 400 casos de treinamento para vida da ferramenta. Fonte: Minitab 13.0®

A análise da tabela 6.7 permite concluir, ao nível de significância adotado de 0,05, que todos os três fatores são significativos para a resposta Razão de Desvios Padrão, o que se

infeire pelo valor P observado nas linhas dedicadas aos efeitos principais. Do mesmo modo observa-se que todas as interações são significativas, o que se infere pelo valor P observado nas linhas dedicadas aos efeitos de interação, que é inferior ao nível de significância adotado.

Tabela 6.7 – Análise de variância para experimento com arranjos fatoriais mistos envolvendo 400 casos de treinamento para vida da ferramenta. Fonte: Minitab 13.0®

Fonte	Graus de liberdade	SQ Seq.	SQ Ajustada	MQ Ajustada	F	Valor P
Algoritmo para determinação do parâmetro de largura da função	3	36597	36597	12199	5,11	0,002
Número de unidades radiais	3	183300	183300	61100	25,59	0,000
Algoritmo para localização dos centros	1	9554	9554	9554	4,00	0,046
Algoritmo para determinação do parâmetro de largura da função*Número de unidades radiais	9	107828	107828	11981	5,02	0,000
Algoritmo para determinação do parâmetro de largura da função*Algoritmo para localização dos centros	3	29987	29987	9996	4,19	0,006
Número de unidades radiais*Algoritmo para localização dos centros	3	28778	28778	9593	4,02	0,008
Algoritmo para determinação do parâmetro de largura da função*Número de unidades radiais*Algoritmo para localização dos centros	9	90161	90161	10018	4,20	0,000
Erro	288	687676	687676	2388		
Total	319	1173881				

A análise dos resultados deste ciclo experimental permitiu concluir que as interações entre os níveis dos fatores são significativas. Mostrou também que existe grande diferença no desempenho das redes em função dos níveis adotados. Algumas configurações mostraram ótimos resultados, enquanto que outras apresentaram péssimos resultados. Quase todos os efeitos e interações foram significativos.

Os resultados descritos levantaram questionamentos sobre a eficácia do ciclo experimental. A grande diferença de desempenho existente entre as configurações fez com que os efeitos dos tratamentos fossem elevados. Isso levou a análise a apontar todos, ou quase todos os efeitos como significativos para o desempenho da rede. Questionou-se então se as significâncias dos efeitos poderiam estar artificialmente elevadas pelas grandes diferenças no desempenho. Concluiu-se, neste ponto, que a presença de configurações com desempenho ruim no planejamento experimental não agregou valor à análise.

Os fatos descritos motivaram a realização de um terceiro ciclo experimental com fatoriais completos com três fatores e dois níveis. Neste terceiro ciclo, para aqueles fatores testados anteriormente em quatro níveis, foram utilizados apenas os dois níveis dos fatores que apresentaram melhores resultados durante o ciclo com fatoriais mistos.

Desta forma aqueles fatores anteriormente distribuídos em quatro níveis tiveram esse número reduzido a dois. Foram mantidos os níveis dos efeitos que apresentaram os menores valores de mínimos quadrados durante este ciclo. As tabelas de 6.8 a 6.19 trazem os valores dos mínimos quadrados dos efeitos principais de cada fator, para os doze experimentos realizados durante este ciclo. Em cada experimento destacam-se os valores das médias dos mínimos quadrados dos níveis dos fatores selecionados para o terceiro ciclo experimental.

Em suma, dois foram os principais resultados obtidos deste ciclo experimental. O primeiro foi a evidência de que interações entre os níveis dos fatores podem ser significativas para o desempenho da rede. O segundo foi a identificação dos níveis dos fatores que levam as redes a melhores desempenhos, provendo a base para planejamento e execução do ciclo experimental definitivo.

Tabela 6.8 – Valores médios dos mínimos quadrados dos efeitos principais, para o experimento envolvendo 12 casos de treinamento, com uso de fatoriais mistos, para vida da ferramenta. Fonte: Minitab 13.0®

Médias dos mínimos quadrados para os efeitos principais			
Nível	Algoritmo para determinação dos parâmetros de largura	Número de unidades radiais	Algoritmo para localização dos centros
1	1,4713	0,8698	0,8065
2	0,4554	0,6467	1,5117
3	1,9864	0,6333	
4	0,7232	2,4865	

Tabela 6.9 – Valores médios dos mínimos quadrados dos efeitos principais, para o experimento envolvendo 15 casos de treinamento, com uso de fatoriais mistos, para vida da ferramenta. Fonte: Minitab 13.0®

Médias dos mínimos quadrados para os efeitos principais			
Nível	Algoritmo para determinação dos parâmetros de largura	Número de unidades radiais	Algoritmo para localização dos centros
1	1,43769	0,67747	0,48712
2	0,20031	0,55906	0,75761
3	0,45991	0,49393	
4	0,39155	0,75900	

Tabela 6.10 – Valores médios dos mínimos quadrados dos efeitos principais, para o experimento envolvendo 24 casos de treinamento, com uso de fatoriais mistos, para vida da ferramenta. Fonte: Minitab 13.0®

Médias dos mínimos quadrados para os efeitos principais			
Nível	Algoritmo para determinação dos parâmetros de largura	Número de unidades radiais	Algoritmo para localização dos centros
1	2,31423	0,44149	0,85085
2	0,10117	0,23035	0,73826
3	0,48669	0,23630	
4	0,27613	2,27008	

Tabela 6.11 – Valores médios dos mínimos quadrados dos efeitos principais, para o experimento envolvendo 30 casos de treinamento, com uso de fatoriais mistos, para vida da ferramenta. Fonte: Minitab 13.0®

Médias dos mínimos quadrados para os efeitos principais			
Nível	Algoritmo para determinação dos parâmetros de largura	Número de unidades radiais	Algoritmo para localização dos centros
1	4,9981	0,2968	0,3516
2	0,0527	0,1815	2,6606
3	0,7081	0,1958	
4	0,2654	5,3501	

Tabela 6.12 – Valores médios dos mínimos quadrados dos efeitos principais, para o experimento envolvendo 48 casos de treinamento, com uso de fatoriais mistos, para vida da ferramenta. Fonte: Minitab 13.0®

Médias dos mínimos quadrados para os efeitos principais			
Nível	Algoritmo para determinação dos parâmetros de largura	Número de unidades radiais	Algoritmo para localização dos centros
1	8,7505	0,1831	1,1916
2	0,0054	0,1653	3,6367
3	0,7867	0,1781	
4	0,1141	9,1300	

Tabela 6.13 – Valores médios dos mínimos quadrados dos efeitos principais, para o experimento envolvendo 60 casos de treinamento, com uso de fatoriais mistos, para vida da ferramenta. Fonte: Minitab 13.0®

Médias dos mínimos quadrados para os efeitos principais			
Nível	Algoritmo para determinação dos parâmetros de largura	Número de unidades radiais	Algoritmo para localização dos centros
1	8,3413	0,1695	1,6803
2	0,2066	0,1559	5,2282
3	5,0182	0,1706	
4	0,2510	13,3213	

Tabela 6.14 – Valores médios dos mínimos quadrados dos efeitos principais, para o experimento envolvendo 120 casos de treinamento, com uso de fatoriais mistos, para vida da ferramenta. Fonte: Minitab 13.0®

Médias dos mínimos quadrados para os efeitos principais			
Nível	Algoritmo para determinação dos parâmetros de largura	Número de unidades radiais	Algoritmo para localização dos centros
1	4,5368	0,1480	0,4988
2	0,0013	0,1485	3,3086
3	1,8367	0,1574	
4	1,2401	7,1609	

Tabela 6.15 – Valores médios dos mínimos quadrados dos efeitos principais, para o experimento envolvendo 150 casos de treinamento, com uso de fatoriais mistos, para vida da ferramenta. Fonte: Minitab 13.0®

Médias dos mínimos quadrados para os efeitos principais			
Nível	Algoritmo para determinação dos parâmetros de largura	Número de unidades radiais	Algoritmo para localização dos centros
1	77,201	0,140	21,004
2	0,000	0,144	23,022
3	6,978	0,160	
4	3,871	87,607	

Tabela 6.16 – Valores médios dos mínimos quadrados dos efeitos principais, para o experimento envolvendo 240 casos de treinamento, com uso de fatoriais mistos, para vida da ferramenta. Fonte: Minitab 13.0®

Médias dos mínimos quadrados para os efeitos principais			
Nível	Algoritmo para determinação dos parâmetros de largura	Número de unidades radiais	Algoritmo para localização dos centros
1	2,7829	0,1308	0,4675
2	0,0002	0,1344	7,2453
3	5,5790	0,1498	
4	7,0636	15,0106	

Tabela 6.17 – Valores médios dos mínimos quadrados dos efeitos principais, para o experimento envolvendo 300 casos de treinamento, com uso de fatoriais mistos, para vida da ferramenta. Fonte: Minitab 13.0®

Médias dos mínimos quadrados para os efeitos principais			
Nível	Algoritmo para determinação dos parâmetros de largura	Número de unidades radiais	Algoritmo para localização dos centros
1	11,700	0,127	9,717
2	0,000	0,138	9,644
3	8,912	0,153	
4	18,108	38,302	

Tabela 6.18 – Valores médios dos mínimos quadrados dos efeitos principais, para o experimento envolvendo 400 casos de treinamento, com uso de fatoriais mistos, para vida da ferramenta. Fonte: Minitab 13.0®

Médias dos mínimos quadrados para os efeitos principais			
Nível	Algoritmo para determinação dos parâmetros de largura	Número de unidades radiais	Algoritmo para localização dos centros
1	20,512	0,138	8,508
2	0,000	0,154	19,436
3	27,482	0,171	
4	7,894	55,426	

Tabela 6.19 – Valores médios dos mínimos quadrados dos efeitos principais, para o experimento envolvendo 500 casos de treinamento, com uso de fatoriais mistos, para vida da ferramenta. Fonte: Minitab 13.0®

Médias dos mínimos quadrados para os efeitos principais			
Nível	Algoritmo para determinação dos parâmetros de largura	Número de unidades radiais	Algoritmo para localização dos centros
1	3,0146	0,1380	0,2060
2	0,0000	0,1612	6,7582
3	5,6388	0,1819	
4	5,2749	13,4473	

6.4 Resultado dos experimentos definitivos com Fatoriais Completos

As tabelas 6.20 a 6.23 contêm os valores das médias e dos desvios padrão obtidos em cada corrida realizada durante o ciclo experimental. Mostram-se em destaque os menores valores médios da resposta Razão de Desvios Padrão obtidos para cada um dos experimentos.

Tabela 6.20 – Médias e desvios padrão dos experimentos para vida da ferramenta com fatorial completo utilizando 12, 15 e 24 casos de treinamento. Fonte: Minitab® 13.0

Corrida	Número de casos no conjunto de treinamento					
	12		15		24	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
1	0,598907	0,350963	0,295549	0,191945	0,015773	0,005841
2	0,504275	0,182452	0,384623	0,164683	0,146341	0,071015
3	0,172907	0,142566	0,035392	0,021013	0,016347	0,006039
4	0,466517	0,314463	0,301593	0,184707	0,134627	0,065694
5	0,707547	0,581037	0,214327	0,141688	0,014712	0,003631
6	0,480357	0,172603	0,42785	0,223562	0,115159	0,059105
7	0,187686	0,132692	0,037546	0,02028	0,017425	0,005043
8	0,459864	0,436267	0,210086	0,077847	0,114968	0,033418

Tabela 6.21 – Médias e desvios padrão dos experimentos para vida da ferramenta com fatorial completo utilizando 30, 48 e 60 casos de treinamento. Fonte: Minitab® 13.0

Corrida	Número de casos no conjunto de treinamento					
	30		48		60	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
1	0,013246	0,005096	0,007545	0,001631	0,008906	0,002602
2	0,121415	0,052722	0,073683	0,027712	0,10163	0,030905
3	0,018993	0,005928	0,001352	0,000638	0,00566	0,002206
4	0,117749	0,05707	0,05555	0,01949	0,064636	0,021191
5	0,011406	0,003335	0,007792	0,001839	0,009466	0,002936
6	0,088629	0,03444	0,068212	0,024747	0,085577	0,022459
7	0,018366	0,00658	0,001395	0,000673	0,005259	0,001291
8	0,100537	0,027436	0,05598	0,016682	0,057044	0,014613

Tabela 6.22 – Médias e desvios padrão dos experimentos para vida da ferramenta com fatorial completo utilizando 120, 150 e 240 casos de treinamento. Fonte: Minitab® 13.0

Corrida	Número de casos no conjunto de treinamento					
	120		150		240	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
1	0,004319	0,00151	0,000775	0,000441	0,443789	0,060575
2	0,066782	0,014724	0,057015	0,017569	0,000528	0,000124
3	0,00064	0,000213	0,000339	0,000176	0,480235	0,047114
4	0,052731	0,011753	0,048585	0,01563	0,000135	4,86E-05
5	0,004049	0,001239	0,000939	0,000502	0,378438	0,037292
6	0,057791	0,012564	0,050187	0,011231	0,000569	0,000138
7	0,000705	0,000215	0,000353	0,000162	0,407246	0,027852
8	0,048991	0,011244	0,044848	0,008662	0,000122	3,52E-05

Tabela 6.23 – Médias e desvios padrão dos experimentos para vida da ferramenta com fatorial completo utilizando 300, 400 e 500 casos de treinamento. Fonte: Minitab® 13.0

Corrida	Número de casos no conjunto de treinamento					
	300		400		500	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
1	0,000273	0,000156	7,6E-05	4,41E-05	0,516852	0,046007
2	0,071848	0,014245	0,028603	0,007248	7,54E-05	2,42E-05
3	4,84E-05	1,98E-05	3,1E-05	1,49E-05	0,594836	0,03364
4	0,067364	0,019605	0,023796	0,00572	2,72E-05	1,28E-05
5	0,000313	0,000154	6,57E-05	2,58E-05	0,430947	0,031127
6	0,059075	0,008555	0,025153	0,006213	8,48E-05	2,38E-05
7	4,49E-05	2,37E-05	3,08E-05	1,33E-05	0,530831	0,020118
8	0,051878	0,011474	0,019633	0,005148	2,66E-05	1,15E-05

Os experimentos com a utilização de fatoriais completos permitiram a determinação dos níveis dos fatores que mais influem na resposta Razão de Desvios Padrão, bem como estimar quantitativamente o efeito das interações entre os fatores. A análise preliminar dos dados obtidos não indicou evidência de que os resíduos seguissem a distribuição normal. Para resolver este problema, fez-se a transformada Box-Cox de todos os resultados obtidos. A

figura 6.7 mostra, a título de exemplo, a transformação Box-Cox dos resultados do experimento envolvendo 48 casos de treinamento. Os gráficos das transformações Box-Cox para todos os experimentos do terceiro ciclo relacionados à vida da ferramenta podem ser encontrados nos arquivos anexados a este trabalho.

Feita a transformação, partiu-se para a análise dos efeitos principais e de interação entre os fatores. Os dados usados para a realização da análise de cada experimento foram os valores resultantes da transformação Box-Cox efetuada.

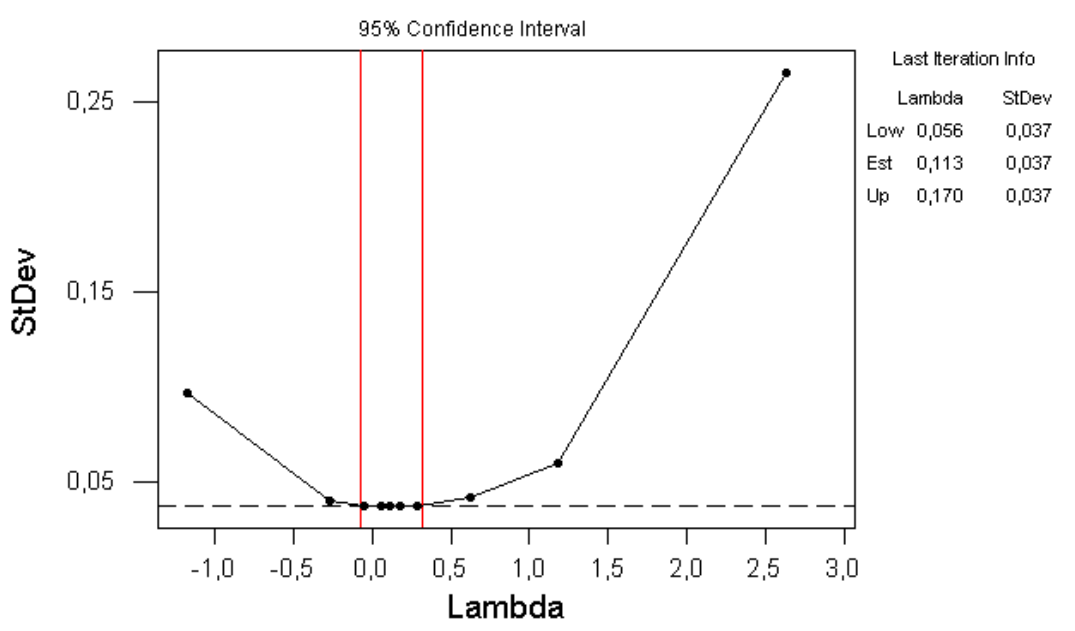


Figura 6.7 - Transformação Box-Cox para resposta Razão de Desvios Padrão, experimento para vida da ferramenta com arranjos fatoriais completos utilizando 48 casos de treinamento. Fonte: Minitab 13.0®

De acordo com as recomendações de vários autores (Montgomery, 2001; Box, Hunter & Hunter, 1978; Noordin et al., 2004), em todo projeto de experimentos deve-se realizar uma análise dos resíduos das respostas. Sempre que o arranjo escolhido utilizar réplicas dos experimentos, originar-se-ão resíduos. Segundo Montgomery (2001), os resíduos devem ser normais, independentes (não-correlacionados) e identicamente distribuídos.

Assim sendo, após a análise feita para cada experimento foram analisados os resíduos da predição da vida da ferramenta, 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 também a análise dos resíduos. Este processo foi repetido até que os resíduos apresentassem distribuição normal. As figuras 6.8 e 6.9 mostram, a título de

exemplo, o gráfico dos efeitos principais e o gráfico dos efeitos de interação obtidos do experimento envolvendo um conjunto de treinamento composto por 48 casos de treinamento.

A figura 6.10 traz o resultado da análise residual para o mesmo experimento. O teste de normalidade para o experimento em foco apresenta um valor P igual a 0,717. Não existe evidência, ao nível de significância adotado, de que os resíduos não sigam a distribuição normal. Os efeitos principais e de interação, e também a análise residual para todos os experimentos deste ciclo relacionados à vida da ferramenta podem ser encontrados nos arquivos anexados a este trabalho.

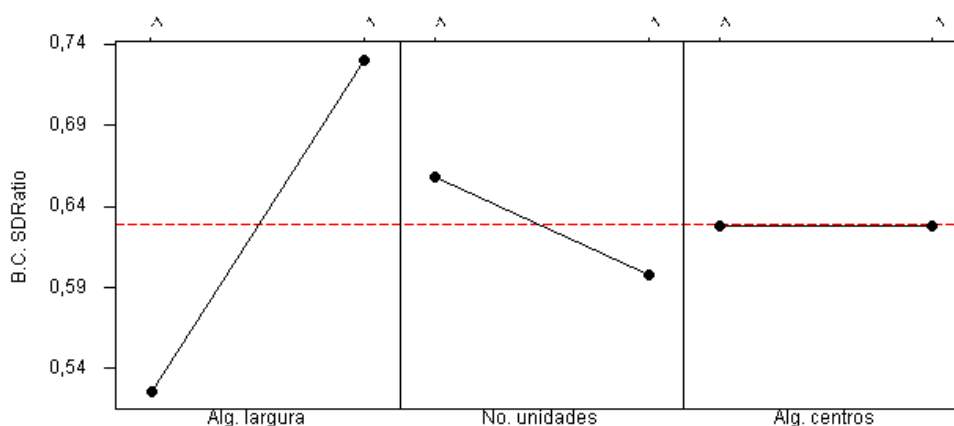


Figura 6.8 - Efeitos principais para a resposta Razão de Desvios Padrão, experimento para vida da ferramenta com arranjos fatoriais completos utilizando 48 casos de treinamento. Fonte: Minitab 13.0®

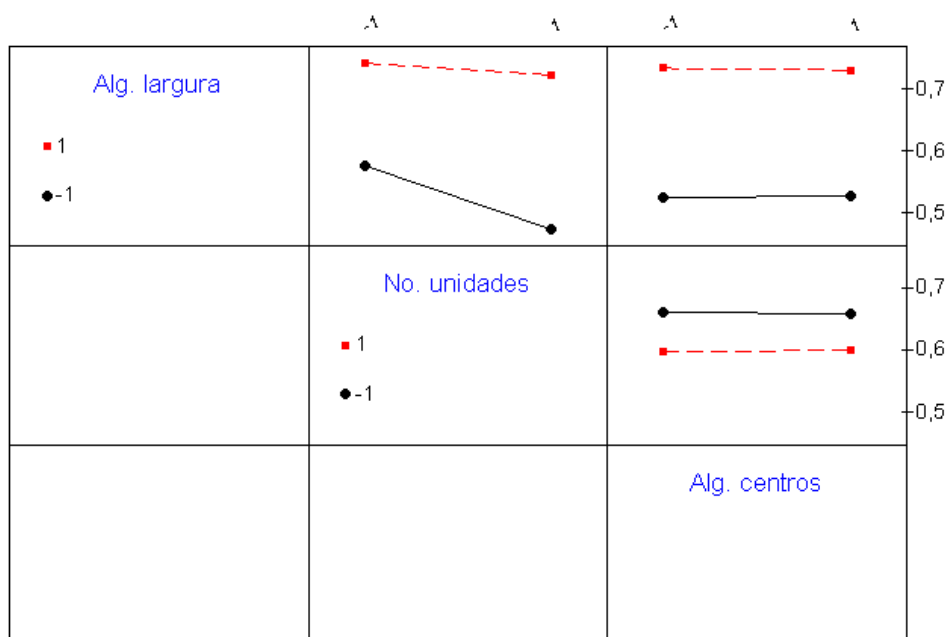


Figura 6.9 - Efeitos de interação para a resposta Razão de Desvios Padrão, experimento para vida da ferramenta com arranjos fatoriais completos utilizando 48 casos de treinamento. Fonte: Minitab 13.0®

Uma vez que os resíduos foram normalizados, foi possível utilizar os resultados da análise de variância, que foi o teste empregado na análise. O resultado dos testes permitiu estabelecer quais fatores e interações foram significativos para o desempenho da rede na tarefa de predição. As tabelas 6.24 e 6.25 mostram, respectivamente, as tabelas contendo os resultados para os testes de significância individual dos efeitos e o sumário da análise de variância (ANOVA) para o experimento envolvendo 48 casos de treinamento.

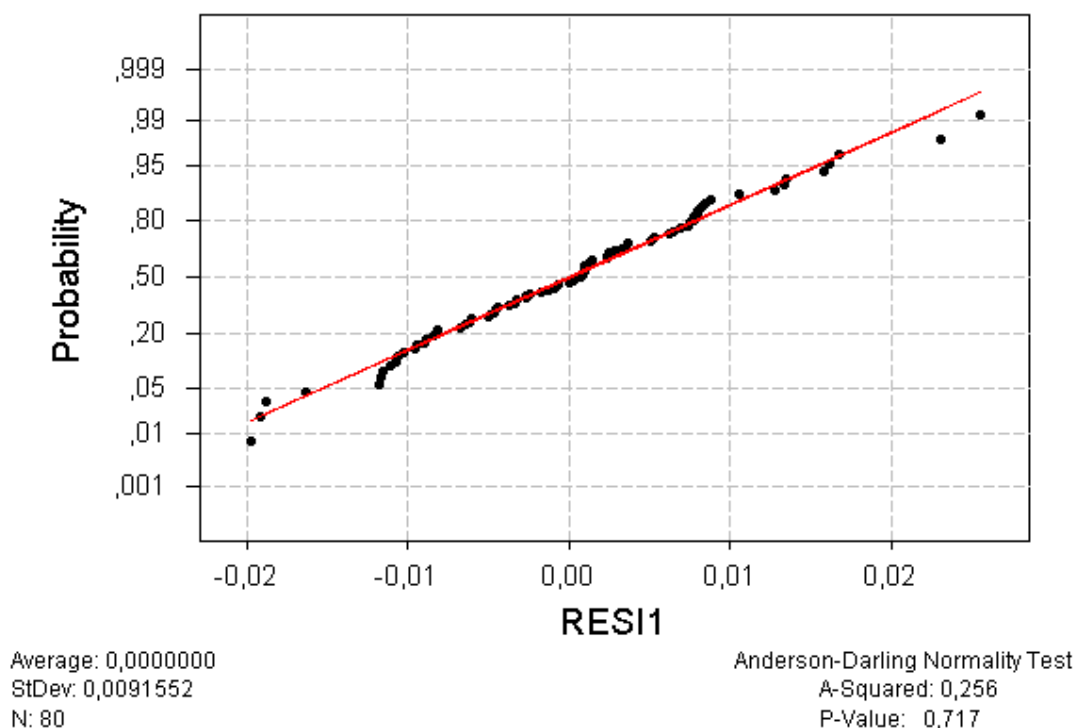


Figura 6.10 - Análise da normalidade dos resíduos, experimento para vida da ferramenta com arranjos fatoriais completos utilizando 48 casos de treinamento. Fonte: Minitab 13.0®

Os resultados da tabela 6.24 mostram quais são os efeitos significativos para o desempenho da rede na tarefa de predição. Consideram-se como significativos os fatores cujo valor P é inferior ao nível de significância adotado, que foi de 0,05. Seguindo este critério os efeitos significativos para a predição da vida da ferramenta no experimento envolvendo 48 casos de treinamento são os efeitos principais dos fatores ‘algoritmo para determinação do parâmetro de largura’ e ‘número de unidades radiais’, bem como o efeito da interação entre estes dois fatores. Em relação aos demais fatores, não há evidência estatística de influência dos mesmos sobre o desempenho da rede. As tabelas que contêm os resultados para significância individual dos efeitos bem como o sumário da análise de variância para todos os experimentos deste ciclo podem ser encontradas nos arquivos anexados a este trabalho.

Ainda que permitam conhecer os efeitos significativos para o projeto de uma rede neural que apresente bom desempenho, os coeficientes obtidos não permitem a elaboração de um modelo de rede por meio de uma equação. Isto se dá porque os parâmetros livres que permitem modelar redes neurais são seus pesos e limiares de ativação. Já os coeficientes obtidos neste estudo são referentes a algoritmos de treinamento e número de unidades radiais. Na utilização de uma rede neural para uma aplicação prática, o conhecimento dos valores absolutos dos pesos entre as sinapses e limiares de ativação não é essencial. A utilidade da determinação dos efeitos significativos reside em fornecer informações que permitam projetar redes que predigam com exatidão e precisão, sem que haja necessidade de conhecer e controlar os valores absolutos de pesos e limiares de disparo assumidos pela rede resultante. Desta forma, não foram elaborados modelos de rede com os coeficientes encontrados pela análise de variância.

Tabela 6.24 – Efeitos estimados dos coeficientes, experimento para vida ferramenta com arranjo fatorial completo utilizando 48 casos de treinamento. Fonte: Minitab 13.0®

Termo	Efeito	Coefficiente	Desvio Padrão do Coeficiente	T	Valor P
Constante		0,62852	0,001072	586,20	0,000
Alg. det. parâmetro largura	0,20561	0,10281	0,001072	95,88	0,000
No. unidades radiais	-0,06110	-0,03055	0,001072	-28,50	0,000
Alg. det. centros	-0,00035	-0,00018	0,001072	-0,16	0,870
Alg. det. parâmetro largura *No. unidades radiais	0,04132	0,02066	0,001072	19,27	0,000
Alg. det. parâmetro largura *Alg. det. centros	-0,00249	-0,00125	0,001072	-1,16	0,249
No. unidades radiais *Alg. det. centros	0,00200	0,00100	0,001072	0,93	0,354
Alg. det. parâmetro largura *No. unidades radiais *Alg. det. centros	0,00193	0,00097	0,001072	0,90	0,370

Tabela 6.25 – Análise de variância, experimento para vida da ferramenta com arranjo fatorial completo utilizando 48 casos de treinamento. Fonte: Minitab 13.0®

Fonte	Graus de liberdade	SQ Seq.	SQ Ajustada	MQ Ajustada	F	Valor P
Efeitos Principais	3	0,920186	0,920186	0,306729	3E+03	0,000
Interações de Segunda Ordem	3	0,034347	0,034347	0,011449	124,49	0,000
Interações de Terceira Ordem	1	0,000075	0,000075	0,000075	0,81	0,370
Erro Residual	72	0,006622	0,006622	0,000092		
Erro Puro	72	0,006622	0,006622	0,000092		
Total	79	0,961230				

A figura 6.11 mostra, a título de exemplo, o diagrama de Pareto dos efeitos no experimento para vida da ferramenta envolvendo 48 casos de treinamento. O valor de alfa adotado foi de 0,05. A análise do diagrama permite identificar graficamente os efeitos significativos para o desempenho da rede neural. Contém a mesma informação da tabela 6.24 expressa em forma gráfica. O diagrama de Pareto permite também obter uma impressão da importância relativa de cada efeito. Nota-se pelo exame da figura 6.11 a maior importância do fator ‘algoritmo para determinação do parâmetro de largura’ para o desempenho da rede, quando se compara seu efeito ao dos demais fatores. Os diagramas de Pareto para todos os experimentos realizados para vida da ferramenta durante este ciclo experimental podem ser encontrados nos arquivos anexados a este trabalho.

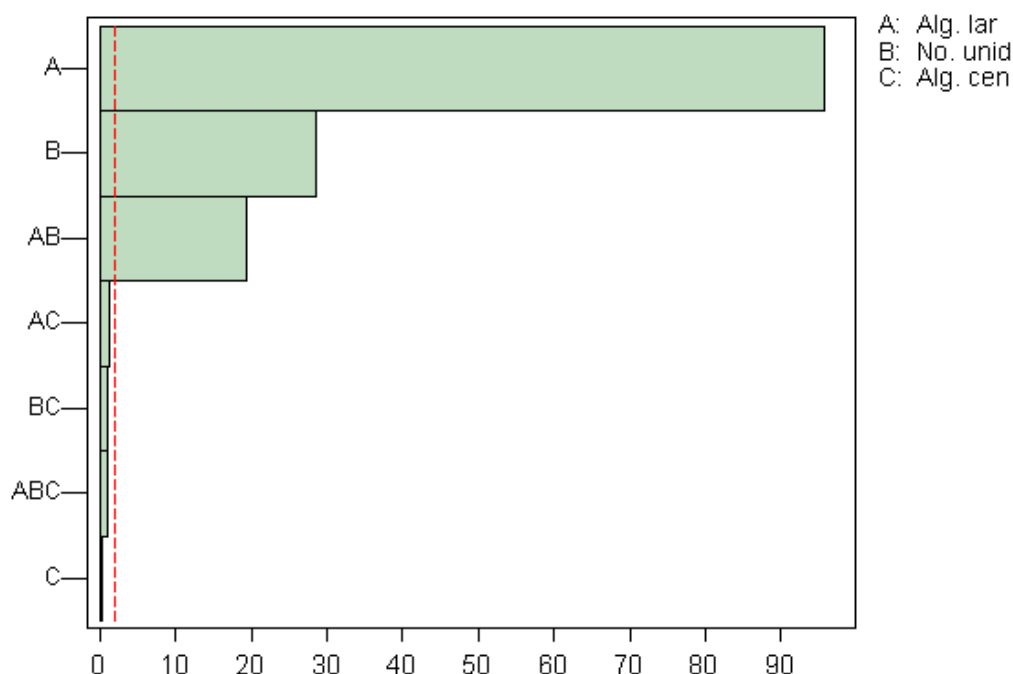


Figura 6.11 - Diagrama de Pareto ($\alpha=0,05$), experimento para vida da ferramenta com arranjos fatoriais completos utilizando 48 casos de treinamento. Fonte: Minitab 13.0®

6.5 Análise dos resultados dos experimentos definitivos com Fatoriais Completos

A análise dos resultados levou à determinação da configuração que apresenta melhores resultados na tarefa de predição da vida da ferramenta para cada tamanho do conjunto de treinamento. A tabela 6.26 indica, para cada conjunto, os níveis dos fatores que levaram ao melhor desempenho.

A tabela 6.27 traz a decodificação dos níveis da tabela 6.26 nos valores dos tratamentos correspondentes a cada nível. Nota-se que o algoritmo isotrópico com fator de escala igual a 10 apresentou os melhores resultados em todos os experimentos. Em relação ao número de unidades radiais na camada oculta, os experimentos realizados com 12 e 15 casos de treinamento indicaram melhores resultados com um número de 9 e 12 unidades, respectivamente. Isto corresponde a uma relação de $\frac{3}{4}$ entre o número de unidades e o número de casos de treinamento. Em todos os demais experimentos o número de unidades radiais que apresentou melhores resultados correspondeu a uma proporção de $\frac{1}{2}$ em relação ao número de casos de treinamento. Para o algoritmo de determinação dos centros nota-se uma alternância entre os algoritmos apontados pela análise como os mais adequados.

Tabela 6.26 – Níveis dos fatores apontados como os melhores para predição da vida da ferramenta, para os experimentos definitivos envolvendo arranjos fatoriais completos

Número de casos no conjunto de treinamento	Alg. determinação do parâmetro de largura da função radial	Número de unidades radiais	Alg. determinação de centros da função radial
12	-1	+1	-1
15	-1	+1	-1
24	-1	-1	+1
30	-1	-1	+1
48	-1	+1	-1
60	-1	+1	+1
120	-1	+1	-1
150	-1	+1	-1
240	+1	+1	+1
300	-1	+1	+1
400	-1	+1	+1
500	+1	+1	+1

Tabela 6.27 – Configurações apontadas como as melhores para predição da vida da ferramenta, para cada experimento envolvendo arranjos fatoriais completos

Número de casos no conjunto de treinamento	Alg. determinação do parâmetro de largura da função radial	Número de unidades radiais	Alg. determinação de centros da função radial
12	Alg. Isotrópico - Fator de escala = 10	9	Sub-amostragem
15	Alg. Isotrópico - Fator de escala = 10	12	Sub-amostragem
24	Alg. Isotrópico - Fator de escala = 10	12	Alg. K-Médias
30	Alg. Isotrópico - Fator de escala = 10	15	Alg. K-Médias
48	Alg. Isotrópico - Fator de escala = 10	36	Sub-amostragem
60	Alg. Isotrópico - Fator de escala = 10	30	Alg. K-Médias
120	Alg. Isotrópico - Fator de escala = 10	60	Sub-amostragem
150	Alg. Isotrópico - Fator de escala = 10	75	Sub-amostragem
240	Alg. Isotrópico - Fator de escala = 10	120	Alg. K-Médias
300	Alg. Isotrópico - Fator de escala = 10	150	Alg. K-Médias
400	Alg. Isotrópico - Fator de escala = 10	200	Alg. K-Médias
500	Alg. Isotrópico - Fator de escala = 10	250	Alg. K-Médias

A figura 6.12 mostra, a título de ilustração, o diagrama esquemático da melhor configuração de rede obtida para o experimento com 60 casos de treinamento após o final de uma corrida experimental. A figura 6.13 mostra, a título de ilustração, o gráfico da superfície de resposta da mesma configuração. As diferentes cores sobre a superfície indicam o valor da predição da vida da ferramenta em pontos formados pela combinação das variáveis independentes (ap , f) mantendo-se Vc em um valor fixo. Deve-se lembrar que os valores mostrados nos eixos das variáveis independentes estão em unidades codificadas. Os valores das predições para vida são os indicados em minutos pela legenda mostrada na figura.

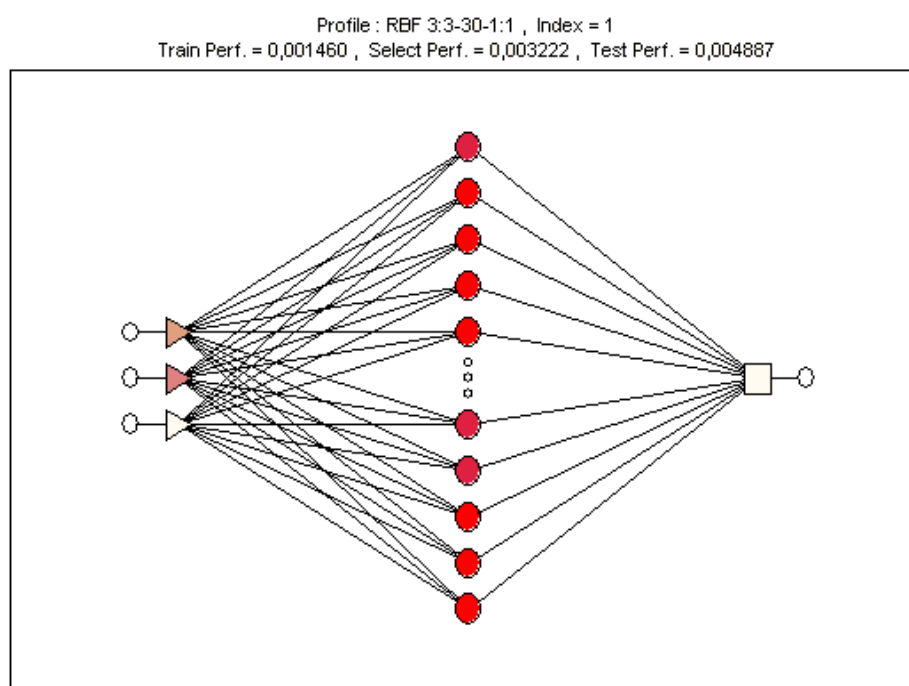


Figura 6.12 – Diagrama de rede RBF obtida durante o experimento para vida da ferramenta com a utilização de 60 casos de treinamento. Fonte: STATISTICA 7.1®

A análise dos diagramas de Pareto contidos nos arquivos anexados a este trabalho juntamente com a análise dos resultados dos valores dos mínimos quadrados, realizada no Minitab®, permite estimar a importância relativa dos fatores e interações sobre a resposta Razão de Desvios Padrão. Observa-se que para os experimentos realizados o algoritmo para determinação do parâmetro de largura da função radial foi o fator mais importante, seguido pelo número de unidades radiais. O fator menos influente nos experimentos realizados foi o algoritmo para determinação dos centros das funções radiais. Em alguns experimentos a análise excluiu este fator do rol dos fatores significativos para o desempenho da rede.

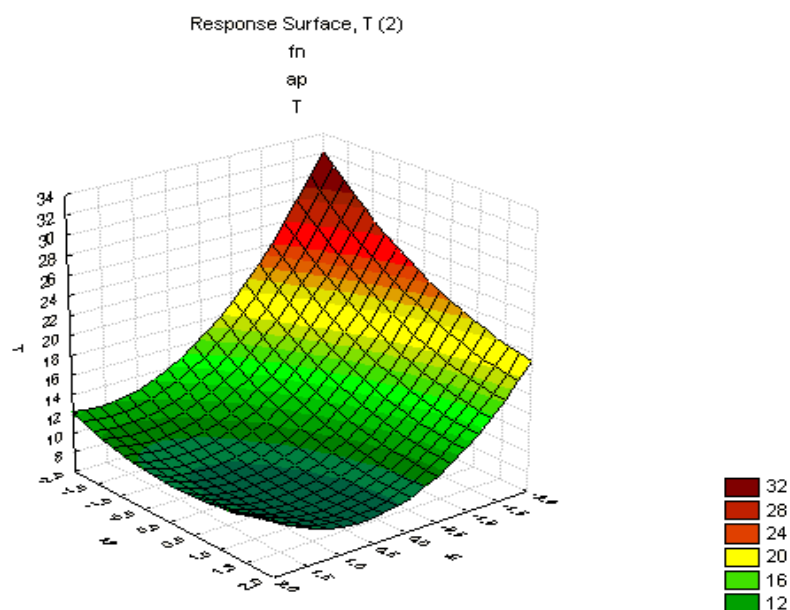


Figura 6.13 – Diagrama de superfície de erro da melhor rede obtida para vida da ferramenta com a utilização de 60 casos de treinamento. Fonte: STATISTICA 7.1®

O fator ‘algoritmo para determinação do parâmetro de largura’ é o mais influente em onze dos doze experimentos realizados. A única exceção é o experimento conduzido com o número mínimo de casos de treinamento (12 casos) no qual o fator ‘número de unidades radiais’ foi o mais influente. O fator ‘número de unidades radiais’ foi o efeito principal mais influente no experimento com 12 casos de treinamento e foi o segundo efeito principal mais importante em dez dos demais experimentos. O efeito do fator principal ‘algoritmo para determinação dos centros’ foi o menos influente dos três efeitos principais em onze dos doze experimentos. A exceção foi o experimento com 30 casos de treinamento, no qual o fator foi o segundo efeito principal mais influente.

Nota-se que os efeitos do fator ‘algoritmo para determinação do parâmetro de largura’ tornam-se dominantes para os experimentos com número de casos de treinamento igual ou superior a 15. Sua influência torna-se mais evidente com o aumento do número de casos no conjunto de treinamento. O parâmetro número de unidades radiais passa a ter uma influência menor à medida que aumenta o número de casos de treinamento.

Em relação à existência de interações, todos os doze experimentos realizados apresentaram pelo menos uma ocorrência de efeito de interação significativo para a resposta. Em diversos casos nota-se mesmo que os efeitos de interação superaram os valores de efeitos principais.

Foi feita uma análise que levou em consideração os valores obtidos por cada efeito no teste individual de significância e também o princípio da Hierarquia (Montgomery, 2001). Este princípio estabelece que, se o efeito de um termo de ordem alta é significativo, todos os termos de ordem baixa que o compõem devem ser conservados, independentemente da sua significância individual. Por exemplo, se a interação AB é significativa, mas o efeito principal A não o é, então ambos os efeitos, A e B, devem ser conservados. O resultado desta análise apontou, para cada experimento realizado para vida da ferramenta durante este ciclo experimental, quais efeitos são significativos para o desempenho da rede. A tabela 6.28 apresenta de forma sumária, quais são estes efeitos para cada número de casos no conjunto de treinamento.

Tabela 6.28 – Sumário dos efeitos significativos para vida da ferramenta, obtidos dos experimentos definitivos com fatoriais completos

Termo	Número de casos no conjunto de treinamento											
	12	15	24	30	48	60	120	150	240	300	400	500
Constante	X	X	X	X	X	X	X	X	X	X	X	X
Alg. det. parâmetro largura	X	X	X	X	X	X	X	X	X	X	X	X
No. unidades radiais	X	X	X	X	X	X	X	X	X	X	X	X
Alg. det. centros		X	X	X		X	X	X	X	X	X	X
Alg. det. parâmetro largura *No. unidades radiais	X	X	X	X	X		X	X	X	X	X	X
Alg. det. parâmetro largura *Alg. det. centros		X	X	X		X	X	X	X	X	X	X
No. unidades radiais *Alg. det. centros		X		X			X		X	X		
Alg. det. parâmetro largura *No. unidades radiais *Alg. det. centros		X							X			

A tabela 6.29 mostra os valores médios da resposta Razão de Desvios Padrão, seu desvio padrão e erro padrão da média, para as melhores configurações de rede obtidas em cada experimento. Nota-se que os valores médios da resposta, para as melhores configurações de rede obtidas, caem a cada aumento do número de casos disponível para treinamento. Deve-se lembrar que os resultados mostrados na tabela 6.29 foram obtidos com casos de teste, ou seja, casos que não tomaram parte no treinamento da rede e foram apresentados à mesma somente durante a fase de testes. Estes resultados aferem de maneira independente a capacidade de generalização das redes obtidas.

Nota-se uma queda expressiva nos valores médios de Razão de Desvios Padrão já no experimento realizado com 15 casos de treinamento e uma redução contínua naquele valor com o aumento do número de casos. Os valores obtidos são extremamente baixos, chegando mesmo à casa de 10^{-5} , o que constitui excelente resultado na tarefa de predição.

Em relação à variabilidade das redes, nota-se que o desvio padrão das melhores configurações obtidas cai com o aumento do número de casos de treinamento, com exceção do experimento realizado com 60 casos de treinamento, que apresentou um desvio padrão superior àquele da melhor configuração obtida com 48 casos de treinamento. Este resultado corrobora o resultado que havia sido obtido durante o experimento para 60 casos de treinamento com utilização de arranjos de Taguchi, no qual a variabilidade da melhor rede obtida para 60 casos também foi nitidamente superior à obtida pela melhor rede obtida para 48 casos.

Tabela 6.29 – Valores médios da resposta Razão de Desvios Padrão, desvios padrão e erros-padrão da média para as melhores redes obtidas para predição da vida da ferramenta durante o ciclo experimental definitivo com uso de fatoriais completos

Número de casos no conjunto de treinamento	Valor médio da resposta Razão de Desvios Padrão para a configuração proposta	Desvio Padrão para a resposta Razão de Desvios Padrão	Erro-padrão da média
12	0,172907	0,142566	0,824524
15	0,035392	0,021013	0,593722
24	0,014712	0,003631	0,246805
30	0,011406	0,003335	0,29239
48	0,001352	0,000638	0,471893
60	0,005259	0,001291	0,245484
120	0,00064	0,000213	0,332813
150	0,000339	0,000176	0,519174
240	0,000122	3,52E-05	0,288525
300	4,49E-05	2,37E-05	0,52784
400	3,08E-05	1,33E-05	0,431818
500	2,66E-05	1,15E-05	0,432331

Em relação à tabela 6.29, a simples observação da mesma não permite afirmar que uma resposta difere de outra. Com o propósito de comparar estatisticamente as melhores configurações obtidas, as médias e variâncias das melhores configurações de rede foram comparadas estatisticamente.

Para testar-se a igualdade das médias obtidas pelas melhores configurações, foram realizados testes de análise de variância (ANOVA) entre todos os pares possíveis entre as melhores configurações de rede. Por exemplo, a média da resposta da melhor rede obtida para 12 casos de treinamento foi comparada por meio da ANOVA à média da resposta da melhor rede obtida para 15 casos. Assim foi feito entre todas as combinações possíveis de conjuntos de treinamento. Os resultados dos testes podem ser observados na tabela 6.30, em que cada célula contém o valor P resultante de um teste de ANOVA. Na tabela, um valor P superior ao nível de significância adotado indica que não há evidência estatística de diferença entre as

variâncias e um valor P inferior a 0,05 indica forte evidência de diferença entre as variâncias. Pode-se notar que, ao nível de significância adotado existe evidência de que todas as médias obtidas pelas melhores configurações de rede para cada número de casos no conjunto de treinamento são diferentes, exceto uma. Esta exceção corresponde às redes obtidas para os conjuntos de treinamento contendo 400 e 500 casos, que não apresentaram evidência de diferença entre suas médias, como indicado pelo valor P resultante da ANOVA, igual a 0,074.

Tabela 6.30 – Valores P obtidos dos testes de ANOVA para comparação das médias das respostas das melhores configurações de rede para vida da ferramenta

Número de casos no conjunto de treinamento →												
↓	12	15	24	30	48	60	120	150	240	300	400	500
12	--	--	--	--	--	--	--	--	--	--	--	--
15	0,000	--	--	--	--	--	--	--	--	--	--	--
24	0,000	0,000	--	--	--	--	--	--	--	--	--	--
30	0,000	0,000	0,000	--	--	--	--	--	--	--	--	--
48	0,000	0,000	0,000	0,000	--	--	--	--	--	--	--	--
60	0,000	0,000	0,000	0,000	0,000	--	--	--	--	--	--	--
120	0,000	0,000	0,000	0,000	0,000	0,000	--	--	--	--	--	--
150	0,000	0,000	0,000	0,000	0,000	0,000	0,000	--	--	--	--	--
240	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	--	--	--	--
300	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	--	--	--
400	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	--	--
500	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,074	--

Para comparar as variâncias entre as configurações, foram aplicados testes de Levene para hipótese nula de variâncias iguais entre todos os pares de respostas possíveis da tabela 6.29 ao nível de significância de 0,05. Por exemplo, a variância da melhor rede obtida para 12 casos de treinamento foi comparada por meio do teste de Levene à variância da melhor rede obtida para 15 casos de treinamento. Assim foi feito para todos os pares das melhores redes. A figura 6.14 ilustra, a título de exemplo, o resultado do teste de comparação de variâncias entre as melhores configurações de rede obtidas durante os experimentos realizados para vida da ferramenta com a utilização de 48 e 60 casos de treinamento.

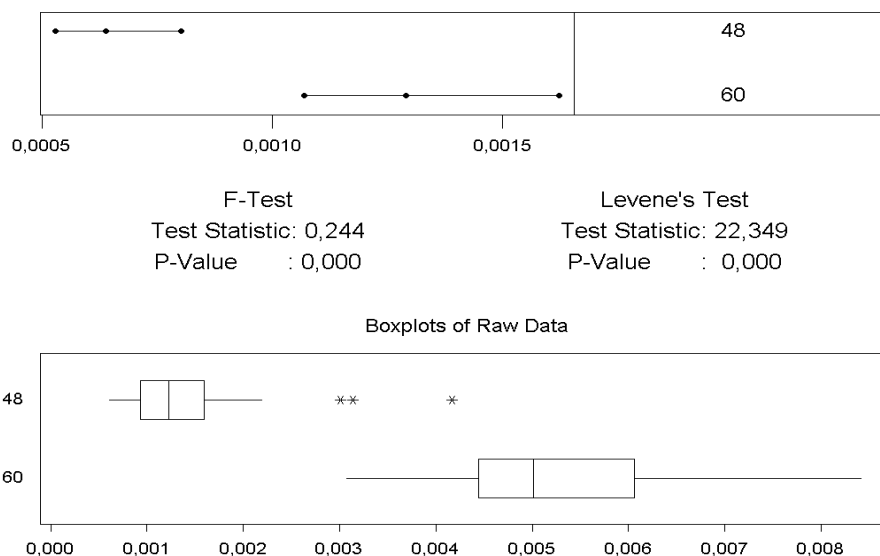


Figura 6.14 – Resultado do teste de comparação entre as variâncias das melhores configurações de rede obtidas para vida da ferramenta durante experimentos utilizando 48 e 60 casos de treinamento.

Fonte: Minitab 13.0®

Os resultados dos testes para comparação das variâncias aplicados a todos os pares podem ser observados na tabela 6.31, em que cada célula contém o valor P resultante de um teste de Levene. Na tabela, um valor P superior ao nível de significância adotado indica que não há evidência estatística de diferença entre as variâncias e um valor P inferior a 0,05 indica forte evidência de diferença entre as variâncias.

Tabela 6.31 – Valores P resultantes do teste de Levene para comparação das variâncias das melhores configurações de rede para vida da ferramenta

No. casos	Var.	Número de casos no conjunto de treinamento											
		12	15	24	30	48	60	120	150	240	300	400	
12	0,020325	--	--	--	--	--	--	--	--	--	--	--	--
15	0,000442	0,000	--	--	--	--	--	--	--	--	--	--	--
24	1,32E-05	0,000	0,000	--	--	--	--	--	--	--	--	--	--
30	1,11E-05	0,000	0,000	0,931	--	--	--	--	--	--	--	--	--
48	4E-07	0,000	0,000	0,000	0,000	--	--	--	--	--	--	--	--
60	1,7E-06	0,000	0,000	0,000	0,000	0,000	--	--	--	--	--	--	--
120	4,52E-08	0,000	0,000	0,000	0,000	0,000	0,000	--	--	--	--	--	--
150	3,11E-08	0,000	0,000	0,000	0,000	0,000	0,000	0,079	--	--	--	--	--
240	1,24E-09	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	--	--	--	--
300	5,7E-10	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,047	--	--
400	1,8E-10	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,004	--
500	1,3E-10	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,326

A análise da tabela 6.31 mostra que, ao nível de significância adotado, não há evidência de diferença entre as variâncias das melhores redes obtidas para 24 e 30 casos de treinamento. O mesmo se conclui para as melhores redes obtidas para 120 e 150 casos de

treinamento, e também para as melhores redes obtidas durante os experimentos com 400 e 500 casos de treinamento. Todos os demais resultados indicados na tabela apontam para evidência, ao nível de 0,05, de diferença entre as variâncias das melhores redes.

Nota-se da análise da tabela 6.31 que há evidência, ao nível de significância de 0,05, de diferença entre a variância das melhores redes obtidas durante os experimentos com 48 e 60 casos. Há também evidência, ao mesmo nível de significância, de diferença entre a variância das melhores redes obtidas durante os experimentos com 60 e 120 casos. Tais resultados são evidenciados pelos valores P iguais 0,000 para os teste de Levene aplicados entre aqueles pares de redes. Estes resultados comprovam que a variância da melhor rede obtida para 60 casos é superior à da melhor rede obtida para 48 casos (o número imediatamente inferior de casos de treinamento) e também superior à da melhor rede obtida para 120 casos (o número imediatamente superior de casos). Observa-se, portanto, que entre as melhores redes obtidas para cada experimento, há uma tendência de redução na variância à medida que aumenta o número de casos de treinamento disponível.

Em resumo, observa-se que as melhores redes obtidas em cada experimento apresentam bons resultados na tarefa de predição da vida da ferramenta. A figura 6.15 mostra os diagramas de Box-plot para as melhores configurações obtidas para cada tamanho do conjunto de treinamento. Os pontos indicam a média da resposta da melhor configuração de rede obtida para um conjunto de treinamento. Os asteriscos indicam *outliers*. Os limites de cada caixa indicam, a partir de baixo, o 1º, 2º e 3º quartis das observações de predição de cada rede. Na figura 6.16 foram retirados os gráficos das melhores redes obtidas para os experimentos com 12 e 15 casos, com o objetivo de aumentar a definição do gráfico para os demais experimentos.

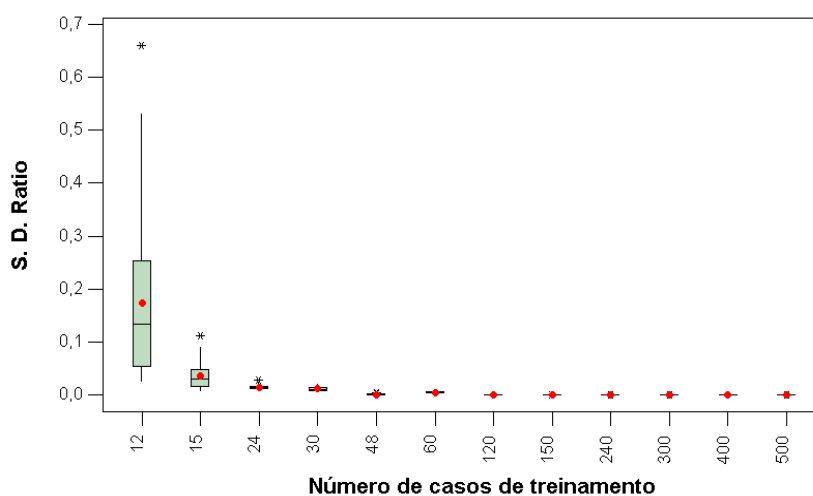


Figura 6.15 –Box-plot das melhores configurações de rede obtidas para vida da ferramenta durante os experimentos definitivos com arranjos fatoriais completos. Fonte: Minitab 13.0®

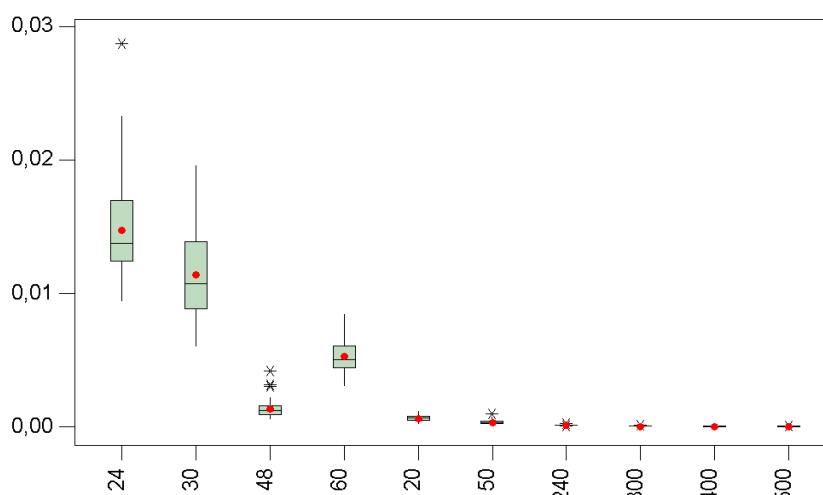


Figura 6.16–Box-plot das melhores configurações de rede de 24 a 500 casos obtidas para vida da ferramenta durante os experimentos definitivos com arranjos fatoriais completos. Fonte: Minitab 13.0®

Por fim, comparou-se o desempenho das melhores redes obtidas em cada experimento ao desempenho de um modelo de otimização linear, o algoritmo da matriz Pseudo-Inversa (Bishop, 1995; Golub e Kahan, 1965). O resultado é mostrado na figura 6.17.

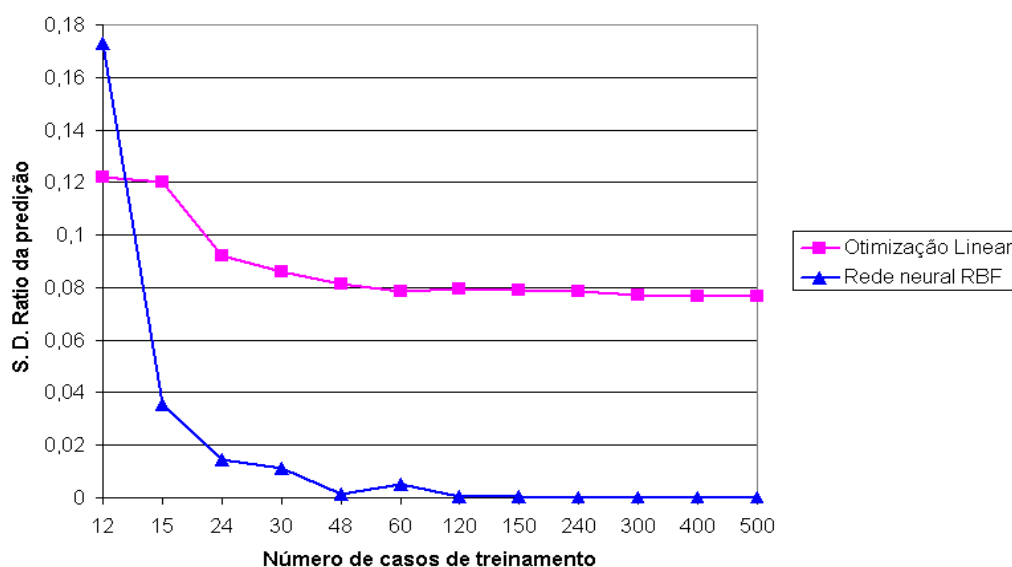


Figura 6.17 –Comparação de desempenho da predição da vida da ferramenta feita entre as melhores configurações de rede e o método da matriz Pseudo-Inversa para otimização linear. Fonte: Excel®.

A análise do gráfico indica que o desempenho do método de otimização linear é superior ao da rede neural apenas na situação em que 12 casos estão disponíveis para a modelagem. Em todos os demais experimentos realizados com um número maior de casos a rede neural apresentou desempenho muito superior ao método de otimização linear.

7 ANÁLISE DOS RESULTADOS PARA RUGOSIDADE MÉDIA (R_a)

7.1 Considerações Iniciais

Neste capítulo são apresentados os resultados obtidos para Rugosidade Média (R_a). Inicialmente são apresentados os resultados obtidos do ciclo experimental envolvendo arranjos ortogonais de Taguchi e discutidos seus resultados. A seguir são apresentados os resultados obtidos do ciclo experimental planejado com arranjos fatoriais mistos. Finalmente são apresentados os resultados obtidos durante o ciclo experimental definitivo realizado com utilização de arranjos fatoriais completos. Os resultados deste último ciclo experimental são discutidos em detalhes. A análise dos resultados foi realizada com a utilização do software estatístico Minitab®, versão 13.0.

7.2 Resultados dos experimentos com arranjos ortogonais de Taguchi

A tabela 7.1 traz as médias da resposta Razão de Desvios Padrão para todas as corridas dos oito experimentos realizados. A tabela 7.2 traz os desvios padrão associados às respostas das corridas.

Fez-se a análise prevista pelo método Taguchi para as médias da resposta, para valor da relação sinal-ruído e para o desvio padrão. A análise considera a dimensão da diferença entre o maior e o menor efeito calculado para o nível de um fator.

Para os experimentos realizados, a análise fornece os valores dos efeitos principais de cada fator sobre cada uma das grandezas sob análise (média da resposta, relação sinal-ruído e variância), bem como o *ranking* dos impactos dos fatores sobre aquelas grandezas. Estas informações estão sumarizadas nas tabelas contidas nos arquivos anexados a este trabalho.

Nota-se que, em todos os casos e para as três análises, o algoritmo para determinação do parâmetro de largura da função radial foi indicado como sendo o fator mais influente. Em relação à minimização da média da resposta e maximização da relação sinal-ruído, o número de unidades radiais foi o segundo fator mais influente todos os experimentos, exceto os realizados com conjuntos de treinamento contendo 24 e 48 casos, nos quais o segundo fator mais influente foi o algoritmo para determinação dos centros da função radial. Para a minimização dos desvios padrão, a análise de Taguchi indicou a mesma configuração para os

oito experimentos, sendo que o número de unidades radiais foi apontado como sendo o segundo fator mais influente.

Tabela 7.1 – Valores médios de Razão de Desvios Padrão obtidos durante os experimentos envolvendo arranjos ortogonais de Taguchi para Rugosidade Média (R_a)

Número da corrida	Número de casos no conjunto de treinamento							
	24	30	48	60	240	300	400	500
1	0,628802	0,601076	0,522039	0,518149	0,525391	0,499767	0,520130	0,533464
2	6,723140	31,596046	97,639313	25,554231	8,539650	136,409844	39,964778	9,638446
3	0,020493	0,017081	0,010203	0,007873	0,000161	0,000055	0,000043	0,000027
4	0,015347	0,016409	0,001517	0,159770	0,000080	0,000059	0,000083	0,000042
5	0,300682	2,059012	0,154282	11,769162	0,302916	33,274003	28,585290	0,252391
6	0,264722	0,253606	0,132365	0,112056	0,097023	0,102912	0,090886	0,080076
7	0,209856	0,146091	0,062477	0,049657	0,188020	31,535303	0,058813	1,332922
8	0,219666	0,165157	0,075796	0,058061	0,046739	0,050059	0,040693	0,036463

Tabela 7.2 – Valores do desvio padrão da resposta Razão de Desvios Padrão obtidos durante os experimentos envolvendo arranjos ortogonais de Taguchi para Rugosidade Média (R_a)

Número da corrida	Número de casos no conjunto de treinamento							
	24	30	48	60	240	300	400	500
1	0,097507	0,054823	0,056103	0,062090	0,044123	0,036352	0,038504	0,040225
2	25,610531	69,309196	284,475841	165,438803	24,592671	433,037908	116,382653	62,502353
3	0,007782	0,005116	0,001351	0,002331	0,000048	0,000032	0,000023	0,000013
4	0,001007	0,000000	0,000130	0,575948	0,000030	0,000026	0,000058	0,000035
5	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000001	0,000000
6	0,069950	0,067611	0,028554	0,022283	0,012444	0,016940	0,014007	0,009498
7	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000
8	0,074861	0,042196	0,016699	0,013782	0,008525	0,012590	0,009017	0,007559

As figuras 7.1 a 7.3 mostram, a título de exemplo, os gráficos dos efeitos principais para o experimento conduzido com o conjunto de treinamento contendo 300 casos. Cada figura é dividida em três partes. Cada parte está relacionada a um fator, indicado na parte superior da figura. A parte inferior mostra os níveis para cada fator. Por eles pode-se ter uma idéia da importância relativa de cada um dos efeitos. Quanto maior a diferença entre os valores dos efeitos principais em dado nível de um fator sobre a grandeza analisada (valores médios de Razão de Desvios Padrão, relação sinal-ruído e desvio padrão da resposta), maior a influência do mesmo. Através das figuras pode-se ver também os níveis que a análise indica serem aqueles que levarão a rede ao melhor desempenho. O conjunto dos gráficos relativos ao ciclo experimental envolvendo arranjos ortogonais de Taguchi pode ser visto nos arquivos anexados a este trabalho.

Em concordância com o que é mostrado nas figuras 7.1 a 7.3 e analisando as tabelas resultantes da análise de Taguchi, os níveis dos fatores apontados como sendo os que mais contribuem para a minimização da resposta e do desvio padrão e para a maximização da robustez do sistema são os sumarizados na tabela 7.3. Para os valores médios de Razão de Desvios Padrão encontrou-se que, exceto para um conjunto de treinamento, as combinações apontadas são as mesmas. No caso da relação sinal-ruído, em seis dos casos a configuração apontada como sendo a que leva à maior robustez é a mesma. Em se tratando da variância, a configuração apontada como sendo a que mais contribui para sua redução foi a mesma em sete dos oito experimentos.

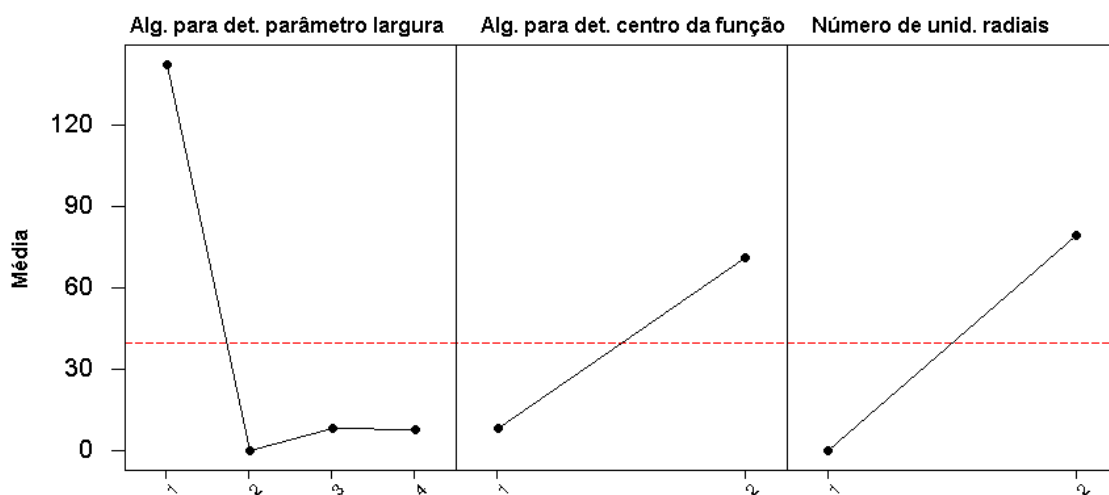


Figura 7.1 – Efeitos principais sobre a média da resposta Razão de Desvios Padrão para conjunto de treinamento de 300 casos. Fonte: Minitab 13.0®

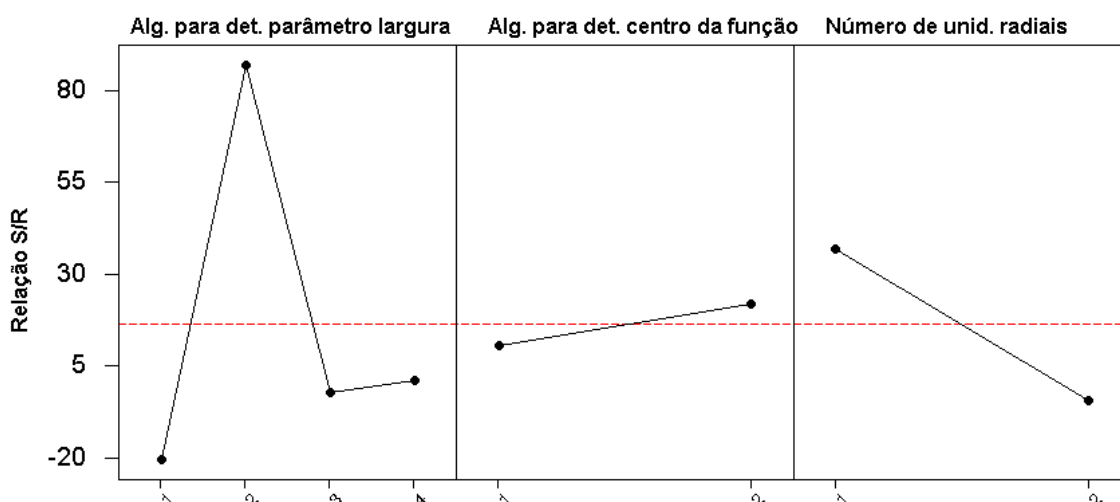


Figura 7.2 – Efeitos principais sobre a função sinal-ruído para conjunto de treinamento de 300 casos. Fonte: Minitab 13.0®

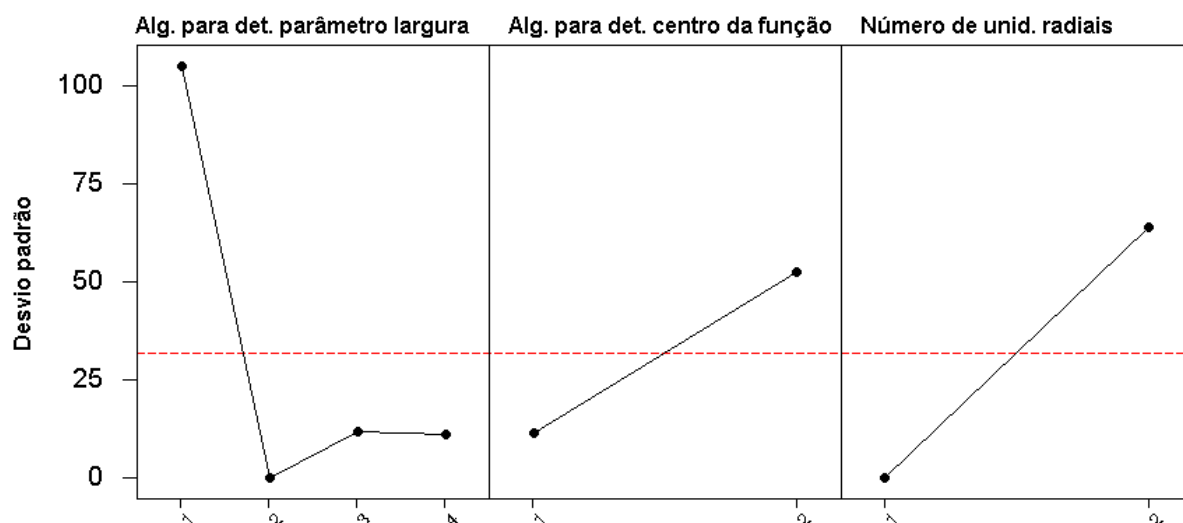


Figura 7.3 – Efeitos principais sobre o desvio padrão da resposta Razão de Desvios Padrão para conjunto de treinamento de 300 casos. Fonte: Minitab 13.0®

Tabela 7.3 – Configurações apontadas como as melhores para predição da Rugosidade Média (R_a), para cada experimento envolvendo arranjos ortogonais de Taguchi

Número de casos no conjunto de treinamento	Combinação para valores médios de Razão de Desvios Padrão	Combinação para valores da relação sinal-ruído	Combinação para valores do desvio padrão
24	2:1:1	2:1:1	2:1:1
30	2:1:1	2:1:1	2:1:1
48	2:1:1	2:1:1	2:1:1
60	4:1:1	4:1:1	4:1:1
240	2:1:1	2:1:1	2:1:1
300	2:1:1	2:2:1	2:1:1
400	2:1:1	2:1:1	2:1:1
500	2:1:1	2:1:1	2:1:1

Para cada experimento foram aplicados testes de hipótese da estatística Z (Montgomery e Runger, 2003) para inferência sobre a média de uma população. Testes da estatística t de Student e análise de variância não foram utilizados porque testes estatísticos preliminares não apresentaram evidência estatística de igualdade de variância entre as amostras, nem de as mesmas possuírem uma distribuição normal, o que seria condição para utilização daquelas técnicas. Dentro de um dado experimento, cada amostra foi comparada por meio do teste Z ao valor da média de cada uma das demais amostras ao nível de significância de 0,05. Testou-se, ao nível de significância de 0,05, a hipótese nula de a média da amostra testada ser igual ao valor da outra média. O objetivo destes testes foi o de estabelecer estatisticamente qual configuração apresentou a menor média da resposta Razão de Desvios Padrão dentre as configurações testadas.

Fazendo-se uma comparação entre os resultados obtidos através da análise de Taguchi e aqueles obtidos nos testes Z foram encontradas discrepâncias em quatro dos oito experimentos. Nos experimentos com 24, 48 e 240 casos de treinamento, a análise apontou para uma configuração que não era a que possuía a menor média da resposta entre as corridas daqueles experimentos. Percebeu-se também que, no experimento com 60 casos de treinamento, a análise indicou uma configuração que não havia sido testada durante o procedimento experimental.

Para estes casos discrepantes, fez-se nova comparação estatística com auxílio do teste Z. Comparou-se as amostras das configurações indicadas pela análise de Taguchi com as médias das corridas apontadas pelo primeiro teste Z como as mais baixas do experimento em questão. Para o caso do experimento com 60 casos de treinamento, em que a análise de Taguchi apontou para uma configuração que não fazia parte do arranjo ortogonal, montou-se a configuração indicada pela análise e realizou-se uma corrida experimental extra, com o conjunto de treinamento apropriado e com o mesmo número de repetições das demais corridas. Após isso, fez-se o segundo teste, cujos resultados são mostrados em tabela contida nos arquivos anexados a este trabalho. Testou-se, ao nível de significância de 0,05, a hipótese nula de a média da amostra indicada pela análise de Taguchi ser igual ao valor da média da configuração com melhor desempenho durante as corridas experimentais. Do resultado, nota-se que as configurações indicadas pela análise de Taguchi possuem médias de Razão de Desvios Padrão maiores (e, portanto, desempenho pior) que as médias das melhores configurações experimentais, como se verifica pelo valor P igual a zero. Tal fato pode indicar a existência de interações entre os fatores do experimento, o que não pôde ser confirmado devido ao arranjo ortogonal escolhido para os experimentos.

Assim sendo, as melhores configurações obtidas para a predição da Rugosidade Média (R_a) da superfície usinada são mostradas na tabela 7.4. Mostram-se também os ditos valores médios da resposta Razão de Desvios Padrão bem como os valores da relação sinal-ruído e desvios padrão da resposta para tais configurações.

Em relação à tabela 7.4, a simples observação da mesma não permite afirmar que uma resposta difere de outra. Com o propósito de comparar estatisticamente as melhores configurações obtidas, foram aplicados testes para comparação de médias e variâncias para todos os pares possíveis de configurações.

Tabela 7.4 – Configurações de níveis com melhor desempenho para os experimentos envolvendo arranjos ortogonais de Taguchi e respectivos valores de Razão de Desvios Padrão para predição da Rugosidade Média (R_a)

Número de casos no conjunto de treinamento	Algoritmo para determinação do parâmetro de largura da função radial	Algoritmo para determinação dos centros da função radial	Número de unidades	Valor médio da resposta Razão de Desvios Padrão	Desvio Padrão
24	Alg. Isotrópico Fator de escala = 10	K-Médias	24	0,015347	0,001007
30	Alg. Isotrópico Fator de escala = 10	K-Médias	30	0,016409	0,000000
48	Alg. Isotrópico Fator de escala = 10	K-Médias	48	0,001517	0,000130
60	Alg. Isotrópico Fator de escala = 10	Sub-amostragem	30	0,007873	0,002331
240	Alg. Isotrópico Fator de escala = 10	K-Médias	240	0,000080	0,000030
300	Alg. Isotrópico Fator de escala = 10	Sub-amostragem	150	0,000055	0,000032
400	Alg. Isotrópico Fator de escala = 10	Sub-amostragem	200	0,000043	0,000023
500	Alg. Isotrópico Fator de escala = 10	Sub-amostragem	250	0,000027	0,000013

Para comparação entre as médias dos valores da resposta Razão de Desvios Padrão obtidas pelas melhores configurações recorreu-se outra vez a um teste de inferência sobre a média de uma população (teste Z). Foram comparadas, ao nível de significância de 0,05, as médias de Razão de Desvios Padrão da melhor configuração obtida em um experimento com o valor médio obtido no experimento com o conjunto de treinamento de tamanho imediatamente inferior. Os resultados destes testes podem ser observados na tabela 7.5.

Tabela 7.5 – Valores P resultantes do teste para inferência sobre a média entre pares de amostras para as melhores configurações obtidas em cada experimento envolvendo arranjos ortogonais de Taguchi para Rugosidade Média (R_a)

Número de casos no conjunto de treinamento maior	Valores médios da resposta para o conjunto de treinamento maior	Número de casos no conjunto de treinamento menor	Valores médios da resposta para o conjunto de treinamento menor	Valor P
30	0,016409	24	0,015347	0,000
48	0,001517	30	0,016409	0,000
60	0,007873	48	0,001517	0,000
240	0,000080	60	0,007873	0,000
300	0,000055	240	0,000080	0,000
400	0,000043	300	0,000055	0,000
500	0,000027	400	0,000043	0,000

Da análise da tabela 7.5 extrai-se que, para as melhores configurações obtidas, ao nível de significância de 0,05 existe suficiente evidência de que a média da resposta Razão de Desvios Padrão aumenta quando se passa do conjunto contendo 24 casos de treinamento para o conjunto contendo 30 casos. Existe evidência de que a média da melhor configuração obtida para os conjuntos contendo 48 casos de treinamento é inferior à daquela correspondente ao conjunto contendo 30 casos.

Nota-se, porém, que para a melhor configuração obtida com o conjunto de treinamento contendo 60 casos existe suficiente evidência de que sua média seja superior àquela obtida da melhor configuração para o experimento envolvendo 48 casos. A partir desse ponto, nota-se que os valores médios de Razão de Desvios Padrão obtidos pela melhor configuração para um determinado número de casos no conjunto de treinamento apresentam suficiente evidência, ao nível de significância adotado, de que sua média é inferior à obtida por aquela do conjunto de treinamento de tamanho imediatamente inferior.

Como a melhor configuração de rede obtida para o conjunto contendo 60 casos de teste apresentou média superior à obtida pela melhor configuração obtida para 48 casos, repetiu-se o teste comparando a média obtida entre a melhor configuração de rede para 60 casos e aquela obtida pela melhor configuração obtida para 30 casos, ao nível de significância de 0,05. O resultado é mostrado em tabela contida nos arquivos anexados a este trabalho. Desta última comparação observa-se que há forte evidência, ao nível de significância adotado, de que a rede treinada com 60 casos apresenta média de Razão de Desvios Padrão inferior à rede treinada com 30 casos.

Para testar as variâncias entre as configurações, foram aplicados testes de Levene para hipótese nula de variâncias iguais entre todos os pares de respostas possíveis da tabela 7.4, ao nível de significância de 0,05. Optou-se por esse teste para comparação entre variâncias pelo fato de testes preliminares indicarem não haver evidência de que as amostras da tabela 7.4 sigam a distribuição normal. Os resultados dos testes de Levene podem ser observados na tabela 7.6, na qual um valor P superior ao nível de significância adotado indica que não há evidência estatística de diferença entre as variâncias e um valor P inferior a 0,05 indica forte evidência de diferença entre as variâncias.

A análise da tabela 7.6 indica que não há evidência de diferença na variância entre as melhores configurações obtidas para o conjunto de treinamento com 24 casos e aqueles conjuntos com 30, 48, 240, 300, 400 e 500 casos de treinamento. Fornece também evidência que a variância da melhor configuração obtida para o conjunto com trinta casos possui

variância inferior às melhores configurações obtidas para conjuntos com mais casos de treinamento.

O resultado dos testes apresenta forte evidência de que a melhor configuração para 48 casos apresenta variância menor que a da melhor configuração para o conjunto contendo 60 casos. Evidencia também que não há, ao nível de significância adotado, diferença entre a variância da melhor configuração para o conjunto de treinamento com 48 casos e a variância das melhores configurações de rede para os conjuntos contendo 240 e 300 casos. Revela ainda que a variância para a melhor configuração com 48 casos apresenta variância superior às melhores configurações obtidas para as situações com 400 e 500 casos de treinamento. O conjunto contendo 60 casos, por sua vez apresenta forte evidência de variância superior às melhores configurações de todos os conjuntos com maior número de casos.

Tabela 7.6 – Valores P resultantes do teste de Levene entre pares de amostras para as melhores configurações obtidas em cada experimento envolvendo arranjos ortogonais de Taguchi para Rugosidade Média (R_a)

	Número de casos no conjunto de treinamento	24	30	48	60	240	300	400
Número de casos no conjunto de treinamento	Variâncias observadas	1,0E-06	7,0E-19	1,7E-08	5,4E-06	9,0E-10	1,0E-09	5,3E-10
Valores P obtidos do teste de Levene								
30	7,0E-19	0,153	-	-	-	-	-	-
48	1,7E-08	0,282	0,005	-	-	-	-	-
60	5,4E-06	0,000	0,000	0,000	-	-	-	-
240	9,0E-10	0,203	0,000	0,118	0,000	-	-	-
300	1,0E-09	0,203	0,000	0,125	0,000	0,938	-	-
400	5,3E-10	0,189	0,000	0,061	0,000	0,124	0,150	-
500	1,7E-10	0,174	0,000	0,024	0,000	0,000	0,001	0,010

No que diz respeito à melhor configuração obtida para o conjunto contendo 240 casos de treinamento, o teste fornece evidência estatística de que não há diferença entre sua variância e aquela dos conjuntos contendo 300 e 400 casos de treinamento. O teste sugere que a variância daquela configuração é superior à variância observada no conjunto contendo 500 casos.

O resultado dos testes de Levene para o conjunto contendo 300 casos de treinamento fornece evidência ao nível de significância de 0,05 de que a variância da melhor configuração observada para o conjunto não é diferente da observada na melhor configuração obtida para o conjunto contendo 400 casos, porém existe evidência, ao nível de significância adotado, de

que sua variância é superior à da melhor configuração obtida para o conjunto de treinamento contendo 500 casos. Em relação ao conjunto contendo 400 casos de treinamento, o teste não evidenciou diferença significativa entre a variância da melhor rede obtida para o conjunto e aquela do conjunto contendo 500 casos, o que se nota pelo valor P obtido, igual a 0,01.

Com a exceção dos experimentos envolvendo 30 e 60 casos de treinamento, para as melhores configurações obtidas, observou-se uma tendência à redução da média de Razão de Desvios Padrão e da variância à medida que aumentou o número de casos de treinamento disponíveis para a rede. Isso sugere um melhor desempenho das redes na predição da Rugosidade Média (R_a) da superfície usinada à medida que aumenta o número de casos de treinamento disponíveis. O gráfico de box plot para as melhores configurações obtidas é mostrado na figura 7.4. Os pontos indicam a média da resposta da melhor configuração de rede obtida para um conjunto de treinamento. Os asteriscos indicam *outliers*. Os limites de cada caixa indicam, a partir de baixo, o 1º, 2º e 3º quartis das observações de predição de cada rede. Na figura pode-se visualizar que a média e a dispersão são maiores nos experimentos envolvendo 24 e 30 casos de teste, e menores nos experimentos com maior número de casos de treinamento. A exceção é o experimento com 60 casos de treinamento, no qual mesmo a melhor configuração apresentou valor médio da resposta e dispersão superiores ao observado com a melhor configuração do conjunto contendo 48 casos de treinamento.

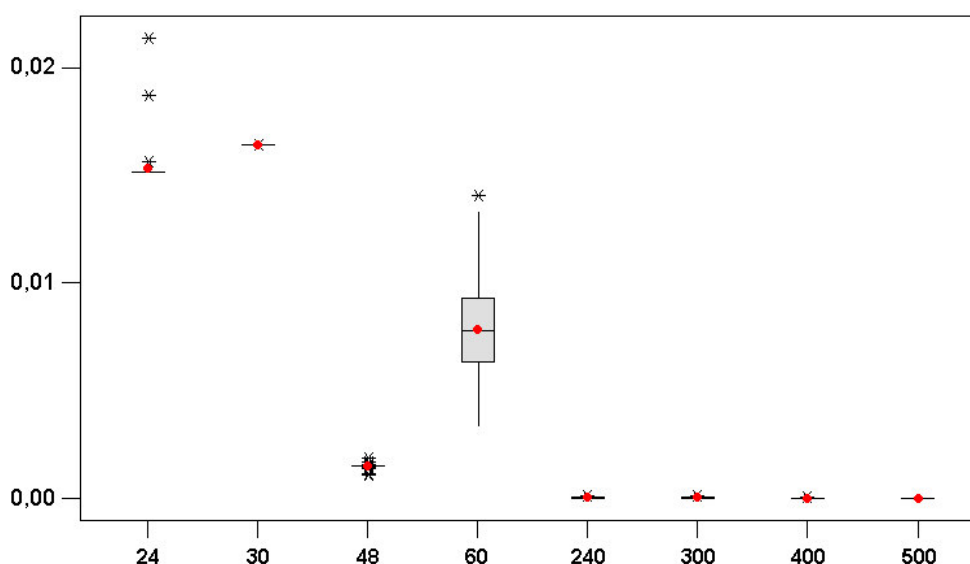


Figura 7.4 – *Box-plot* da resposta Razão de Desvios Padrão para as melhores configurações obtidas em cada experimento envolvendo arranjos ortogonais de Taguchi para Rugosidade Média (R_a). Fonte: Minitab 13.0®

Os resultados deste ciclo experimental indicam que mesmo para situações envolvendo um número pequeno de casos de treinamento (como os que envolvem apenas 24 ou 30 casos), as redes apresentam um bom desempenho na tarefa de predição para os casos de teste apresentados, como evidenciam os valores mostrados pela tabela 7.4. Isso sugere que redes RBF podem constituir-se numa alternativa válida e economicamente viável para a tarefa de predição da Rugosidade Média (R_a). Além disso, as discrepâncias entre os resultados experimentais e os resultados da análise indicam a possibilidade da existência de interações entre os níveis dos fatores, o que não pôde ser confirmado devido ao arranjo ortogonal utilizado. Os resultados também levantaram dúvidas sobre qual seria o desempenho das redes com conjuntos contendo menos que 24 casos de treinamento e também contendo números de casos entre 60 e 240 casos. Questionou-se também se outras proporções entre o número de unidades radiais e o número de casos de treinamento poderiam levar as redes a apresentar melhores resultados.

Tais resultados e as questões levantadas durante o experimento levaram ao planejamento e execução do segundo ciclo experimental com a utilização de fatoriais completos mistos, com um maior número de conjuntos de treinamento e mais níveis para o fator número de unidades radiais.

7.3 Resultados dos experimentos com Fatoriais Completos mistos

O segundo ciclo experimental teve por objetivos estimar o valor dos efeitos dos fatores sobre o desempenho das redes para predição da Rugosidade Média, estimar o efeito de possíveis interações entre os níveis dos fatores sob teste e avaliar o desempenho das redes com a utilização de mais níveis do fator número de unidades radiais.

Com efeito, os experimentos deste ciclo permitiram estimar os níveis dos fatores que mais influem na resposta Razão de Desvios Padrão, bem como estimar a existência de interação entre os fatores. As figuras 7.5 e 7.6 mostram os efeitos principais e as interações obtidas dos experimentos envolvendo um conjunto de treinamento composto por 500 casos.

Os gráficos dos efeitos principais e de interação para os demais experimentos deste ciclo com a utilização de arranjos fatoriais mistos podem ser encontrados nos arquivos anexados a este trabalho.

A tabela 7.7 contém o resultado da análise de variância realizada para o experimento envolvendo 500 casos de treinamento. As tabelas contendo o resultado da análise de variância

para os demais experimentos realizados durante este ciclo podem ser encontradas nos arquivos anexados a este trabalho.

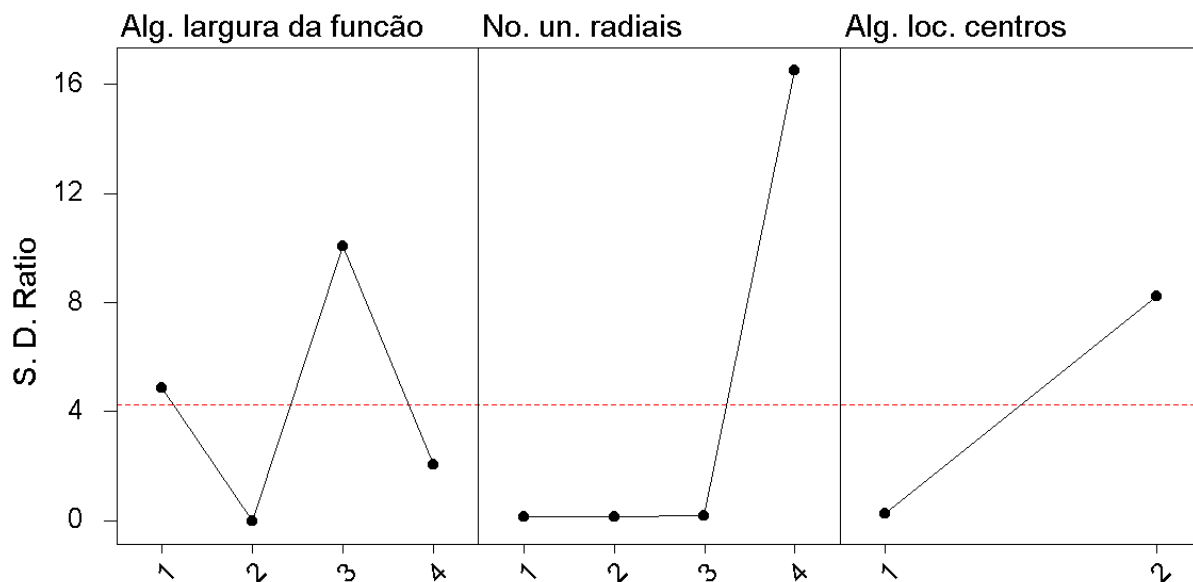


Figura 7.5 – Efeitos principais para a resposta Razão de Desvios Padrão, para experimento com arranjos fatoriais mistos envolvendo 500 casos de treinamento para Rugosidade Média (R_a). Fonte: Minitab 13.0®

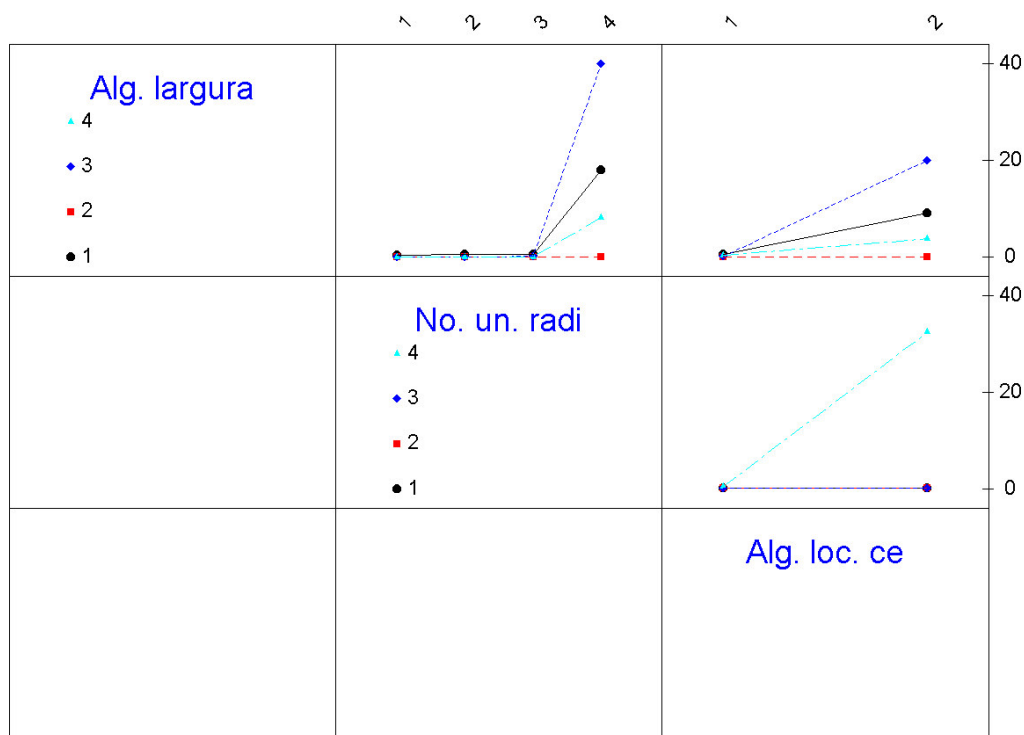


Figura 7.6 – Efeitos de interação para a resposta Razão de Desvios Padrão, para experimento com arranjos fatoriais mistos envolvendo 500 casos de treinamento para Rugosidade Média (R_a). Fonte: Minitab 13.0®

A análise da tabela 7.7 permite concluir, ao nível de significância adotado de 0,05, que todos os três fatores são significativos para a resposta Razão de Desvios Padrão, o que se infere pelo valor P observado nas linhas dedicadas aos efeitos principais. Do mesmo modo observa-se que todas as interações são significativas, o que se infere pelo valor P observado nas linhas dedicadas aos efeitos de interação, que é inferior ao nível de significância adotado.

Tabela 7.7 – Análise de variância para experimento com arranjos fatoriais mistos envolvendo 500 casos de treinamento para Rugosidade Média (R_a). Fonte: Minitab 13.0®

Fonte	Graus de liberdade	SQ Seq.	SQ Ajustada	MQ Ajustada	F	Valor P
Algoritmo para determinação do parâmetro de largura da função	3	4561,3	4561,3	1520,4	4,93	0,002
Número de unidades radiais	3	16061,9	16061,9	5354,0	17,36	0,000
Algoritmo para localização dos centros	1	5085,4	5085,4	5085,4	16,49	0,000
Algoritmo para determinação do parâmetro de largura da função *Número de unidades radiais	9	13372,2	13372,2	1485,8	4,82	0,000
Algoritmo para determinação do parâmetro de largura da função *Algoritmo para localização dos centros	3	4526,0	4526,0	1508,7	4,89	0,002
Número de unidades radiais*Algoritmo para localização dos centros	3	15318,7	15318,7	5106,2	16,56	0,000
Algoritmo para determinação do parâmetro de largura da função *Número de unidades radiais* Algoritmo para localização dos centros	9	13592,1	13592,1	1510,2	4,90	0,000
Erro	288	88816,2	88816,2	308,4		
Total	319	161333,9				

A análise dos resultados deste ciclo experimental permitiu concluir que as interações entre os níveis dos fatores são significativas. Mostrou também que existe grande diferença no desempenho das redes em função dos níveis adotados. Algumas configurações mostraram ótimos resultados, enquanto que outras apresentaram péssimos resultados. Quase todos os efeitos e interações foram significativos.

Os resultados descritos levantaram questionamentos sobre a eficácia do ciclo experimental. A grande diferença de desempenho existente entre as configurações fez com que os efeitos dos tratamentos fossem elevados. Isso levou a análise a apontar todos, ou quase todos os efeitos como significativos para o desempenho da rede. Questionou-se então se as significâncias dos efeitos poderiam estar artificialmente elevadas pelas grandes diferenças no

desempenho. Concluiu-se, neste ponto, que a presença de configurações com desempenho ruim no planejamento experimental não agregou valor à análise.

Os fatos descritos motivaram a realização de um terceiro ciclo experimental com fatoriais completos com três fatores e dois níveis. Neste terceiro ciclo, para aqueles fatores testados anteriormente em quatro níveis, foram utilizados apenas os dois níveis dos fatores que apresentaram melhores resultados durante o ciclo com fatoriais mistos.

Desta forma aqueles fatores anteriormente distribuídos em quatro níveis tiveram esse número reduzido a dois. Foram mantidos os níveis dos efeitos que apresentaram os menores valores de mínimos quadrados durante este ciclo. As tabelas de 7.8 a 7.19 trazem os valores dos mínimos quadrados dos efeitos principais de cada fator, para os doze experimentos realizados durante este ciclo. Em cada experimento destacam-se os valores das médias dos mínimos quadrados dos níveis dos fatores selecionados para o terceiro ciclo experimental.

Em suma, dois foram os principais resultados obtidos deste ciclo experimental. O primeiro foi a evidência de que interações entre os níveis dos fatores podem ser significativas para o desempenho da rede. O segundo foi a identificação dos níveis dos fatores que levam as redes a melhores desempenhos, provendo a base para planejamento e execução do ciclo experimental definitivo.

Tabela 7.8 – Valores médios dos mínimos quadrados dos efeitos principais, para o experimento envolvendo 12 casos de treinamento, com uso de fatoriais mistos, para Rugosidade Média (R_a). Fonte: Minitab 13.0®

Médias dos mínimos quadrados para os efeitos principais			
Nível	Algoritmo para determinação dos parâmetros de largura	Número de unidades radiais	Algoritmo para localização dos centros
1	0,86521	0,74018	0,55733
2	0,42586	0,56450	1,14714
3	1,41915	0,60263	
4	0,69871	1,50162	

Tabela 7.9 – Valores médios dos mínimos quadrados dos efeitos principais, para o experimento envolvendo 15 casos de treinamento, com uso de fatoriais mistos, para Rugosidade Média (R_a). Fonte: Minitab 13.0®

Médias dos mínimos quadrados para os efeitos principais			
Nível	Algoritmo para determinação dos parâmetros de largura	Número de unidades radiais	Algoritmo para localização dos centros
1	1,14632	0,59452	0,50103
2	0,26847	0,54656	0,69055
3	0,50187	0,43950	
4	0,46650	0,80259	

Tabela 7.10 – Valores médios dos mínimos quadrados dos efeitos principais, para o experimento envolvendo 24 casos de treinamento, com uso de fatoriais mistos, para Rugosidade Média (R_a). Fonte: Minitab 13.0®

Médias dos mínimos quadrados para os efeitos principais			
Nível	Algoritmo para determinação dos parâmetros de largura	Número de unidades radiais	Algoritmo para localização dos centros
1	1,27869	0,48674	0,46659
2	0,11790	0,28727	0,65432
3	0,45094	0,29066	
4	0,39430	1,17717	

Tabela 7.11 – Valores médios dos mínimos quadrados dos efeitos principais, para o experimento envolvendo 30 casos de treinamento, com uso de fatoriais mistos, para Rugosidade Média (R_a). Fonte: Minitab 13.0®

Médias dos mínimos quadrados para os efeitos principais			
Nível	Algoritmo para determinação dos parâmetros de largura	Número de unidades radiais	Algoritmo para localização dos centros
1	5,9779	0,3790	0,4095
2	0,0682	0,2597	3,2556
3	0,7055	0,2611	
4	0,5786	6,4305	

Tabela 7.12 – Valores médios dos mínimos quadrados dos efeitos principais, para o experimento envolvendo 48 casos de treinamento, com uso de fatoriais mistos, para Rugosidade Média (R_a). Fonte: Minitab 13.0®

Médias dos mínimos quadrados para os efeitos principais			
Nível	Algoritmo para determinação dos parâmetros de largura	Número de unidades radiais	Algoritmo para localização dos centros
1	16,4069	0,2144	6,2184
2	0,0065	0,1906	2,7426
3	1,4113	0,1960	
4	0,0973	17,3210	

Tabela 7.13 – Valores médios dos mínimos quadrados dos efeitos principais, para o experimento envolvendo 60 casos de treinamento, com uso de fatoriais mistos, para Rugosidade Média (R_a). Fonte: Minitab 13.0®

Médias dos mínimos quadrados para os efeitos principais			
Nível	Algoritmo para determinação dos parâmetros de largura	Número de unidades radiais	Algoritmo para localização dos centros
1	11,3350	0,2033	1,3321
2	0,0072	0,1734	9,4335
3	2,7633	0,1798	
4	7,4255	20,9746	

Tabela 7.14 – Valores médios dos mínimos quadrados dos efeitos principais, para o experimento envolvendo 120 casos de treinamento, com uso de fatoriais mistos, para Rugosidade Média (R_a).

Fonte: Minitab 13.0®

Médias dos mínimos quadrados para os efeitos principais			
Nível	Algoritmo para determinação dos parâmetros de largura	Número de unidades radiais	Algoritmo para localização dos centros
1	1,9433	0,1594	0,2595
2	0,0015	0,1502	4,6726
3	2,9205	0,1625	
4	4,9989	9,3921	

Tabela 7.15 – Valores médios dos mínimos quadrados dos efeitos principais, para o experimento envolvendo 150 casos de treinamento, com uso de fatoriais mistos, para Rugosidade Média (R_a).

Fonte: Minitab 13.0®

Médias dos mínimos quadrados para os efeitos principais			
Nível	Algoritmo para determinação dos parâmetros de largura	Número de unidades radiais	Algoritmo para localização dos centros
1	54,441	0,158	17,681
2	0,000	0,157	19,365
3	11,606	0,168	
4	8,043	73,609	

Tabela 7.16 – Valores médios dos mínimos quadrados dos efeitos principais, para o experimento envolvendo 240 casos de treinamento, com uso de fatoriais mistos, para Rugosidade Média (R_a).

Fonte: Minitab 13.0®

Médias dos mínimos quadrados para os efeitos principais			
Nível	Algoritmo para determinação dos parâmetros de largura	Número de unidades radiais	Algoritmo para localização dos centros
1	4,3558	0,1533	0,3389
2	0,0002	0,1591	3,9508
3	1,3574	0,1762	
4	2,8659	8,0908	

Tabela 7.17 – Valores médios dos mínimos quadrados dos efeitos principais, para o experimento envolvendo 300 casos de treinamento, com uso de fatoriais mistos, para Rugosidade Média (R_a).

Fonte: Minitab 13.0®

Médias dos mínimos quadrados para os efeitos principais			
Nível	Algoritmo para determinação dos parâmetros de largura	Número de unidades radiais	Algoritmo para localização dos centros
1	13,619	0,146	4,209
2	0,000	0,157	9,128
3	8,196	0,171	
4	4,859	26,200	

Tabela 7.18 – Valores médios dos mínimos quadrados dos efeitos principais, para o experimento envolvendo 400 casos de treinamento, com uso de fatoriais mistos, para Rugosidade Média (R_a).

Fonte: Minitab 13.0®

Médias dos mínimos quadrados para os efeitos principais			
Nível	Algoritmo para determinação dos parâmetros de largura	Número de unidades radiais	Algoritmo para localização dos centros
1	5,9249	0,1415	1,9504
2	0,0001	0,1553	10,4083
3	12,7245	0,1712	
4	6,0680	24,2495	

Tabela 7.19 – Valores médios dos mínimos quadrados dos efeitos principais, para o experimento envolvendo 500 casos de treinamento, com uso de fatoriais mistos, para Rugosidade Média (R_a).

Fonte: Minitab 13.0®

Médias dos mínimos quadrados para os efeitos principais			
Nível	Algoritmo para determinação dos parâmetros de largura	Número de unidades radiais	Algoritmo para localização dos centros
1	4,8586	0,1410	0,2620
2	0,0000	0,1575	8,2349
3	10,0661	0,1757	
4	2,0691	16,5196	

7.4 Resultados dos experimentos definitivos com Fatoriais Completos

As tabelas 7.20 a 7.23 contêm os valores das médias e dos desvios padrão obtidos em cada corrida realizada durante o ciclo experimental. Mostram-se em destaque os menores valores médios da resposta Razão de Desvios Padrão obtidos para cada um dos experimentos.

Tabela 7.20 – Médias e desvios padrão dos experimentos para Rugosidade Média com fatorial completo utilizando 12, 15 e 24 casos de treinamento. Fonte: Minitab® 13.0

Corrida	Número de casos no conjunto de treinamento					
	12		15		24	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
1	0,56118	0,08221	0,465004	0,162328	0,018287	0,005806
2	0,505283	0,136892	0,516523	0,118283	0,21587	0,071909
3	0,334699	0,226554	0,047284	0,018055	0,023019	0,006327
4	0,495452	0,10756	0,396984	0,050284	0,204438	0,052888
5	0,566374	0,058891	0,507818	0,124788	0,018606	0,004375
6	0,483585	0,032952	0,471427	0,044499	0,213594	0,069842
7	0,420654	0,254259	0,049875	0,022596	0,023095	0,004672
8	0,501209	0,10751	0,396884	0,038822	0,188146	0,04446

Tabela 7.21 – Médias e desvios padrão dos experimentos para Rugosidade Média com fatorial completo utilizando 30, 48 e 60 casos de treinamento. Fonte: Minitab® 13.0

Corrida	Número de casos no conjunto de treinamento					
	30		48		60	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
1	0,017624	0,004252	0,010409	0,001571	0,007115	0,002537
2	0,175173	0,047964	0,096469	0,033666	0,144514	0,047587
3	0,014064	0,002929	0,001955	0,001147	0,000842	0,000312
4	0,143526	0,053669	0,07797	0,025383	0,161889	0,049949
5	0,01677	0,004356	0,010682	0,001518	0,00758	0,002265
6	0,168517	0,042015	0,080466	0,019587	0,108651	0,022971
7	0,014755	0,004655	0,001678	0,000873	0,000815	0,000298
8	0,13675	0,03665	0,060899	0,015172	0,123486	0,035963

Tabela 7.22 – Médias e desvios padrão dos experimentos para Rugosidade Média com fatorial completo utilizando 120, 150 e 240 casos de treinamento. Fonte: Minitab® 13.0

Corrida	Número de casos no conjunto de treinamento					
	120		150		240	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
1	0,470862	0,070372	0,001011	0,000649	0,000612	0,000192
2	0,004759	0,001734	0,065564	0,019195	0,109803	0,02556
3	0,456688	0,054739	0,000362	0,000151	0,000157	5,91E-05
4	0,000569	0,000256	0,056262	0,015275	0,105278	0,014803
5	0,422963	0,039442	0,001046	0,000557	0,00061	0,000184
6	0,004608	0,001338	0,054795	0,012011	0,095503	0,012707
7	0,41766	0,038303	0,00036	0,000158	0,000143	5,25E-05
8	0,000657	0,000242	0,052164	0,010411	0,096582	0,01439

Tabela 7.23 – Médias e desvios padrão dos experimentos para Rugosidade Média com fatorial completo utilizando 300, 400 e 500 casos de treinamento. Fonte: Minitab® 13.0

Corrida	Número de casos no conjunto de treinamento					
	300		400		500	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
1	0,000269	0,000124	0,481711	0,05269	7,5E-05	3,02E-05
2	0,049898	0,013958	9,7E-05	4,31E-05	0,038271	0,011655
3	5,31E-05	2,16E-05	0,526024	0,03733	3,0E-05	1,4E-05
4	0,051391	0,013491	4,29E-05	1,94E-05	0,036978	0,008574
5	0,000248	9,94E-05	0,398326	0,020239	7,6E-05	2,21E-05
6	0,050523	0,0121	0,00012	4,63E-05	0,036125	0,007105
7	5,34E-05	2,82E-05	0,45359	0,017403	3,51E-05	1,24E-05
8	0,05137	0,011185	4,36E-05	2,1E-05	0,036336	0,007107

Os experimentos com a utilização de fatoriais completos permitiram a determinação dos níveis dos fatores que mais influem na resposta Razão de Desvios Padrão, bem como estimar quantitativamente o efeito das interações entre os fatores. A análise preliminar dos dados obtidos não indicou evidência de que os resíduos seguissem a distribuição normal. Para resolver este problema, fez-se a transformada Box-Cox de todos os resultados obtidos. A

figura 7.7 mostra, a título de exemplo, a transformação Box-Cox dos resultados do experimento envolvendo 60 casos de treinamento. Os gráficos das transformações Box-Cox para todos os experimentos do terceiro ciclo relacionados à Rugosidade Média podem ser encontrados nos arquivos anexados a este trabalho.

Feita a transformação, partiu-se para a análise dos efeitos principais e de interação entre os fatores. Os dados usados para a realização da análise de cada experimento foram os valores resultantes da transformação Box-Cox efetuada.

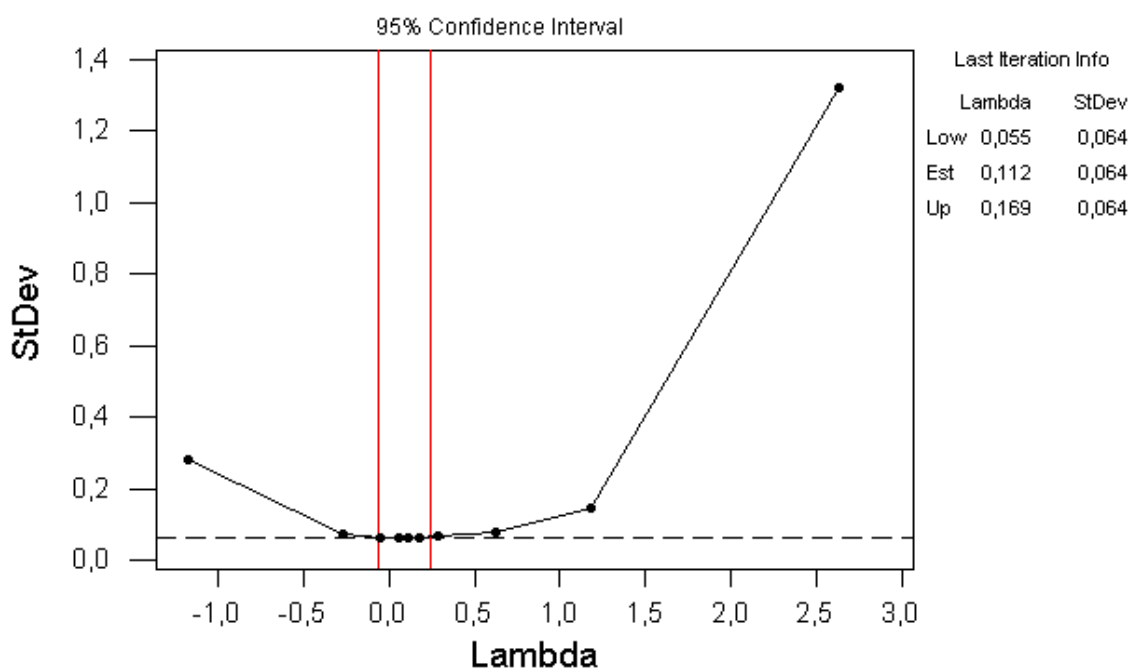


Figura 7.7 - Transformação Box-Cox para resposta Razão de Desvios Padrão, experimento para Rugosidade Média com arranjos fatoriais completos utilizando 60 casos de treinamento. Fonte: Minitab 13.0®

De acordo com as recomendações de vários autores (Montgomery, 2001; Box, Hunter & Hunter, 1978; Noordin et al., 2004), em todo projeto de experimentos deve-se realizar uma análise dos resíduos das respostas. Resíduos são as diferenças entre duas ou mais observações e a média formada por estas observações (Montgomery, 2001). Sempre que o arranjo escolhido utilizar réplicas dos experimentos, originar-se-ão resíduos. Segundo Montgomery (2001), os resíduos devem ser normais, independentes (não-correlacionados) e identicamente distribuídos.

Assim sendo, após a análise feita para cada experimento foram analisados os resíduos da predição da Rugosidade Média, 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 também a análise dos resíduos. Este processo foi repetido até que os resíduos apresentassem distribuição normal. As figuras 7.8 e 7.9 mostram, a título de exemplo, o gráfico dos efeitos principais e o gráfico dos efeitos de interação obtidos do experimento envolvendo um conjunto de treinamento composto por 60 casos de treinamento.

A figura 7.10 traz o resultado da análise residual para o mesmo experimento. O teste de normalidade para o experimento em foco apresenta um valor P igual a 0,215. Não existe evidência, ao nível de significância adotado, de que os resíduos não sigam a distribuição normal. Efeitos principais e de interação, e também a análise residual para todos os experimentos deste ciclo relacionados à Rugosidade Média podem ser encontrados nos arquivos anexados a este trabalho.

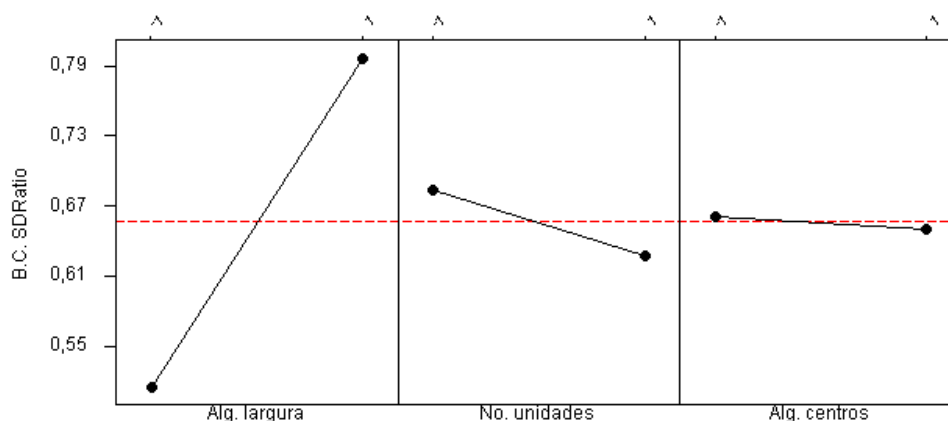


Figura 7.8 - Efeitos principais para a resposta Razão de Desvios Padrão, experimento para Rugosidade Média com arranjos fatoriais completos utilizando 60 casos de treinamento. Fonte: Minitab 13.0®

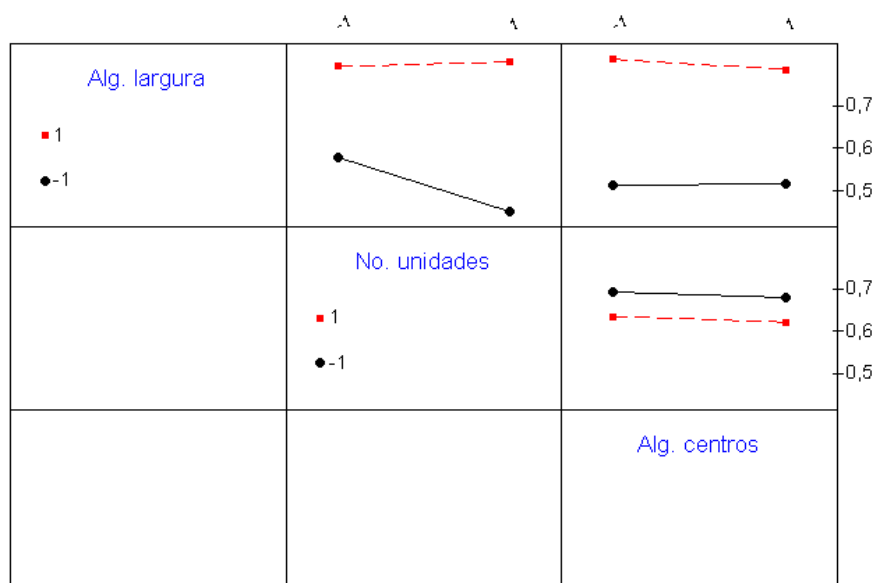


Figura 7.9 - Efeitos de interação para a resposta Razão de Desvios Padrão, experimento para Rugosidade Média com arranjos fatoriais completos utilizando 60 casos de treinamento. Fonte: Minitab 13.0®

Uma vez que os resíduos foram normalizados, foi possível utilizar os resultados da análise de variância, que foi o teste empregado na análise. O resultado dos testes permitiu estabelecer quais fatores e interações foram significativos para o desempenho da rede na tarefa de predição. As tabelas 7.24 e 7.25 mostram respectivamente as tabelas contendo os resultados para os testes de significância individual dos efeitos e o sumário da análise de variância (ANOVA) para o experimento envolvendo 60 casos de treinamento.

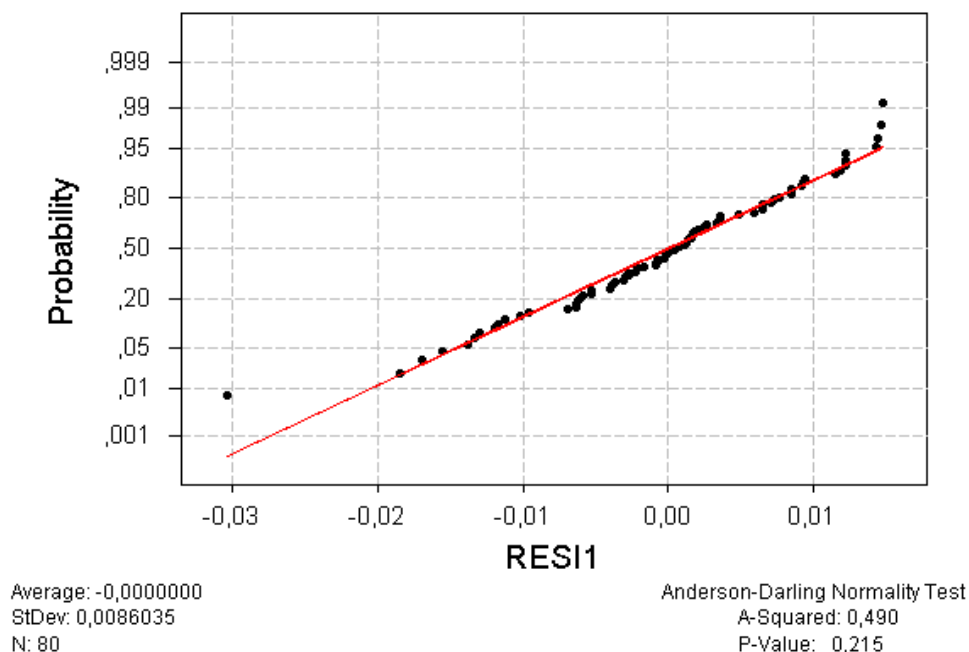


Figura 7.10 - Análise da normalidade dos resíduos, experimento para Rugosidade Média com arranjos fatoriais completos utilizando 60 casos de treinamento. Fonte: Minitab 13.0®

Os resultados da tabela 7.24 mostram quais são os efeitos significativos para o desempenho da rede na tarefa de predição. Consideram-se como significativos os fatores cujo valor P é inferior ao nível de significância adotado, que foi de 0,05. Seguindo este critério os efeitos significativos para a predição da Rugosidade Média no experimento envolvendo 60 casos de treinamento são os efeitos principais dos fatores ‘algoritmo para determinação do parâmetro de largura’, ‘número de unidades radiais’ e ‘algoritmo para determinação dos centros’, bem como o efeito da interação entre os fatores ‘algoritmo para determinação do parâmetro de largura’ e ‘número de unidades radiais’ e também o efeito da interação entre os fatores ‘algoritmo para determinação do parâmetro de largura’ e ‘algoritmo para determinação dos centros’. Em relação às demais interações, não há evidência estatística de influência das mesmas sobre o desempenho da rede. As tabelas que contêm os resultados para significância individual dos efeitos bem como o sumário da análise de variância para todos os experimentos deste ciclo podem ser encontradas nos arquivos anexados a este trabalho.

Ainda que permitam conhecer os efeitos significativos para o projeto de uma rede neural que apresente bom desempenho, os coeficientes obtidos não permitem a elaboração de um modelo de rede por meio de uma equação. Isto se dá porque os parâmetros livres que permitem modelar redes neurais são seus pesos e limiares de ativação. Já os coeficientes obtidos neste estudo são referentes a algoritmos de treinamento e número de unidades radiais. Na utilização de uma rede neural para uma aplicação prática, o conhecimento dos valores absolutos dos pesos entre as sinapses e limiares de ativação não é essencial. A utilidade da determinação dos efeitos significativos reside em fornecer informações que permitam projetar redes que predigam com exatidão e precisão, sem que haja necessidade de conhecer e controlar os valores absolutos de pesos e limiares de disparo assumidos pela rede resultante. Desta forma, não foram elaborados modelos de rede com os coeficientes encontrados pela análise de variância.

Tabela 7.24 – Efeitos estimados dos coeficientes, experimento para Rugosidade Média (R_a) com arranjo fatorial completo utilizando 48 casos de treinamento. Fonte: Minitab 13.0®

Termo	Efeito	Coefficiente	Desvio Padrão do Coeficiente	T	Valor P
Constante		0,65600	0,001008	651,07	0,000
Alg. det. parâmetro largura	0,28331	0,14166	0,001008	140,59	0,000
No. unidades radiais	-0,05728	-0,02864	0,001008	-28,43	0,000
Alg. det. Centros	-0,01175	-0,00587	0,001008	-5,83	0,000
Alg. det. parâmetro largura *No. unidades radiais	0,06766	0,03383	0,001008	33,58	0,000
Alg. det. parâmetro largura *Alg. det. centros	-0,01294	-0,00647	0,001008	-6,42	0,000
No. unidades radiais *Alg. det. centros	-0,00128	-0,00064	0,001008	-0,64	0,526
Alg. det. parâmetro largura *No. unidades radiais *Alg. det. centros	0,00173	0,00086	0,001008	0,86	0,394

Tabela 7.25 – Análise de variância, experimento para Rugosidade Média (R_a) com arranjo fatorial completo utilizando 48 casos de treinamento. Fonte: Minitab 13.0®

Fonte	Graus de liberdade	SQ Seq.	SQ Ajustada	MQ Ajustada	F	Valor P
Efeitos Principais	3	1,67368	1,67368	0,557894	7E+03	0,000
Interações de Segunda Ordem	3	0,09494	0,09494	0,031647	389,67	0,000
Interações de Terceira Ordem	1	0,00006	0,00006	0,000060	0,74	0,394
Erro Residual	72	0,00585	0,00585	0,000081		
Erro Puro	72	0,00585	0,00585	0,000081		
Total	79	1,77453				

A figura 7.11 mostra, a título de exemplo, o diagrama de Pareto dos efeitos no experimento para Rugosidade Média envolvendo 60 casos de treinamento. O valor de alfa adotado foi de 0,05. A análise do diagrama permite identificar graficamente os efeitos significativos para o desempenho da rede neural. Contém a mesma informação da tabela 7.24 expressa em forma gráfica. O diagrama de Pareto permite também obter uma impressão da importância relativa de cada efeito. Nota-se pelo exame da figura 7.11 a maior importância do fator ‘algoritmo para determinação do parâmetro de largura’ para o desempenho da rede, quando se compara seu efeito ao dos demais fatores. Os diagramas de Pareto para todos os experimentos realizados para Rugosidade Média durante este ciclo experimental podem ser encontrados nos arquivos anexados a este trabalho.

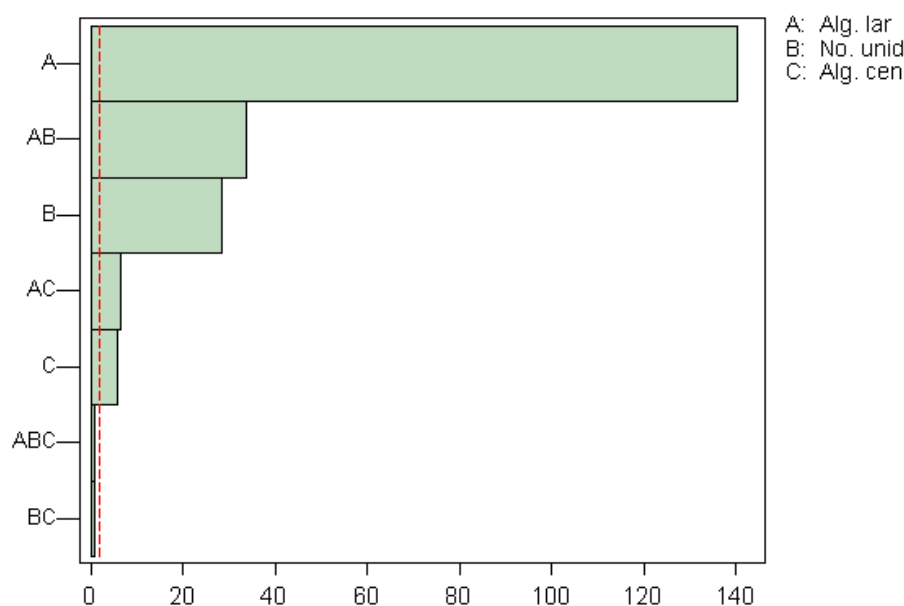


Figura 7.11 - Diagrama de Pareto ($\alpha=0,05$), experimento para Rugosidade Média com arranjos fatoriais completos utilizando 60 casos de treinamento. Fonte: Minitab 13.0®

7.5 Análise dos resultados dos experimentos definitivos com Fatoriais Completos

A análise dos resultados levou à determinação da configuração que apresenta melhores resultados na tarefa de predição da Rugosidade Média para cada tamanho do conjunto de treinamento. A tabela 7.26 indica, para cada conjunto, os níveis dos fatores que levaram ao melhor desempenho.

A tabela 7.27 traz a decodificação dos níveis da tabela 7.26 nos valores dos tratamentos correspondentes a cada nível. Nota-se que o algoritmo isotrópico com fator de escala igual a 10 apresentou os melhores resultados em todos os experimentos. Em relação ao número de unidades radiais na camada oculta, os experimentos realizados com 12, 15, 30, 48 e 60 casos de treinamento indicaram melhores resultados com um número de 9, 12, 23, 36 e 45 unidades, respectivamente. Isto corresponde a uma relação de $\frac{3}{4}$ entre o número de unidades e o número de casos de treinamento. Nos demais experimentos o número de unidades radiais que apresentou melhores resultados correspondeu a uma proporção de $\frac{1}{2}$ em relação ao número de casos de treinamento. Para o algoritmo de determinação dos centros nota-se uma alternância entre os algoritmos apontados pela análise como os mais adequados.

Tabela 7.26 – Níveis dos fatores apontados como os melhores para predição da Rugosidade Média (R_a), para os experimentos definitivos envolvendo arranjos fatoriais completos

No. casos de treinamento	Alg. Det. parâmetro de largura da função radial	No. de unidades radiais	Alg. Det. de centros da função radial
12	-1	+1	-1
15	-1	+1	-1
24	-1	-1	-1
30	-1	+1	-1
30 (alternativa)	-1	+1	+1
48	-1	+1	+1
60	-1	+1	+1
120	+1	+1	-1
150	-1	+1	+1
240	-1	+1	+1
300	-1	+1	-1
400	+1	+1	-1
500	-1	+1	-1

Tabela 7.27 – Configurações apontadas como as melhores para predição da Rugosidade Média (R_a), para cada experimento envolvendo arranjos fatoriais completos

No. casos de treinamento	Alg. Det. parâmetro de largura da função radial	No. unidades radiais	Alg. Det. de centros da função radial
12	Alg. Isotrópico - Fator de escala = 10	9	Sub-amostragem
15	Alg. Isotrópico - Fator de escala = 10	12	Sub-amostragem
24	Alg. Isotrópico - Fator de escala = 10	12	Sub-amostragem
30	Alg. Isotrópico - Fator de escala = 10	23	Sub-amostragem
30 (alternativa)	Alg. Isotrópico - Fator de escala = 10	23	Alg. K-Médias
48	Alg. Isotrópico - Fator de escala = 10	36	Alg. K-Médias
60	Alg. Isotrópico - Fator de escala = 10	45	Alg. K-Médias
120	Alg. Isotrópico - Fator de escala = 10	60	Sub-amostragem
150	Alg. Isotrópico - Fator de escala = 10	75	Alg. K-Médias
240	Alg. Isotrópico - Fator de escala = 10	120	Alg. K-Médias
300	Alg. Isotrópico - Fator de escala = 10	150	Sub-amostragem
400	Alg. Isotrópico - Fator de escala = 10	200	Sub-amostragem
500	Alg. Isotrópico - Fator de escala = 10	250	Sub-amostragem

A figura 7.12 mostra, a título de ilustração, o diagrama esquemático da melhor configuração de rede obtida para o experimento com 120 casos de treinamento após o final de uma corrida experimental. A figura 7.13 mostra, a título de ilustração, o gráfico da superfície de resposta da mesma configuração. As diferentes cores sobre a superfície indicam o valor da predição da rugosidade média (R_a) em pontos formados pela combinação das variáveis independentes (ap , f) mantendo-se V_c em um valor fixo. Deve-se lembrar que os valores mostrados nos eixos das variáveis independentes estão em unidades codificadas. Os valores das predições para rugosidade são os indicados pela legenda mostrada na figura.

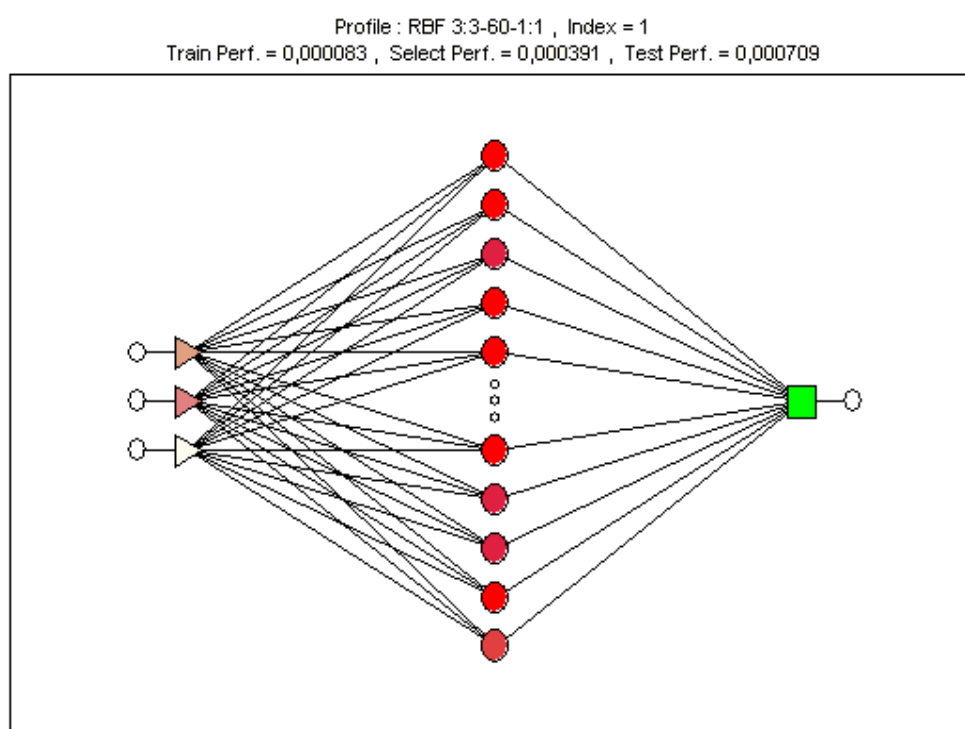


Figura 7.12 – Diagrama de rede RBF obtida para o experimento para Rugosidade Média com a utilização de 120 casos de treinamento. Fonte: STATISTICA 7.1®

A análise dos diagramas de Pareto contidos nos arquivos anexados a este trabalho juntamente com a análise dos resultados dos valores dos mínimos quadrados, realizada no Minitab®, permite estimar a importância relativa dos fatores e interações sobre a resposta Razão de Desvios Padrão. Observa-se que para os experimentos realizados o algoritmo para determinação do parâmetro de largura da função radial foi o fator mais importante. A influência dos demais fatores foi bem menor. Em alguns experimentos a análise chegou inclusive a excluir o fator ‘algoritmo para determinação dos centros’ do rol dos fatores significativos para o desempenho da rede.

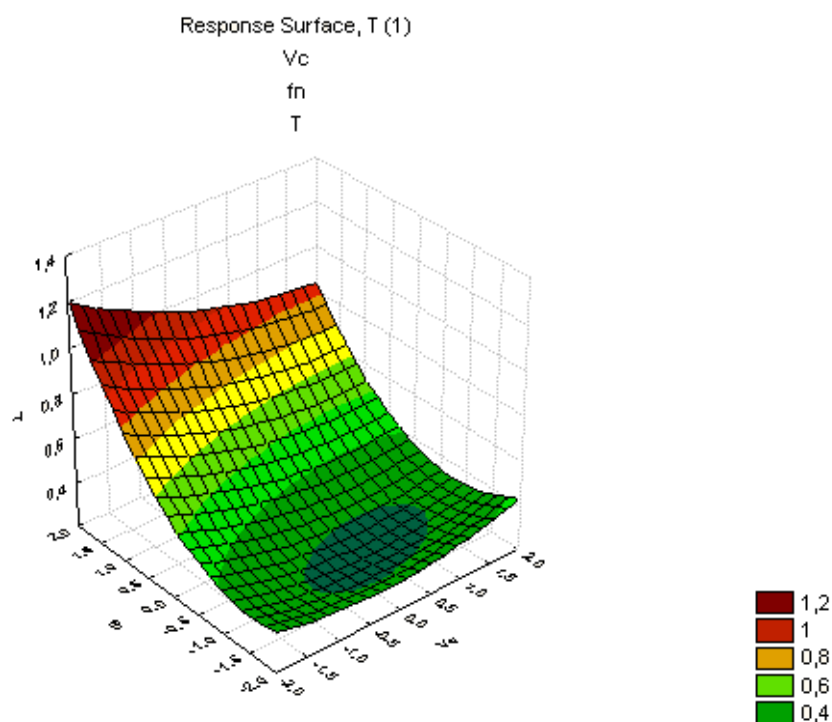


Figura 7.13 – Diagrama de superfície de erro da melhor rede obtida para Rugosidade Média com a utilização de 120 casos de treinamento. Fonte: STATISTICA 7.1®

Nota-se que o fator ‘algoritmo para determinação do parâmetro de largura’ é o mais influente em dez dos doze experimentos realizados. As exceções são os experimentos conduzidos com o número mínimo de casos de treinamento (12 casos) no qual o fator não é significativo, e no experimento com 24 casos de treinamento, no qual o efeito do fator é o terceiro mais significativo, sendo inferior aos efeitos do número de unidades radiais e à interação dupla entre o fator ‘algoritmo para determinação do parâmetro de largura’ e o fator ‘número de unidades radiais’. Nota-se que os efeitos deste fator tornam-se dominantes para os experimentos com número de casos de treinamento igual ou superior a 24.

O fator ‘número de unidades radiais’ foi o efeito principal mais influente nos experimentos com 12 e 15 casos de treinamento e foi o segundo efeito principal mais importante nos outros dez experimentos. Nos experimentos com maior número de casos de treinamento este fator apresentou influência progressivamente menor. O fator ‘algoritmo para determinação dos centros’ foi o menos influente dos efeitos principais nos doze experimentos.

Em onze dos doze experimentos realizados é possível detectar-se ao menos uma interação significativa. A exceção é o experimento para 30 casos de treinamento. Em diversos casos nota-se mesmo que a influência dos efeitos de interações supera a dos respectivos efeitos principais. É interessante notar que para o experimento conduzido com 12 casos de

treinamento o efeito apontado pela análise como o mais influente foi o de uma interação (existente entre o fator ‘algoritmo para determinação do parâmetro de largura’ e o fator ‘número de unidades radiais’).

Ainda no experimento realizado com 30 casos de treinamento a configuração que obteve o melhor resultado difere da apontada pela análise como a melhor. Porém, se inspecionarmos a tabela de análise de variância para o experimento, observamos que o fator ‘algoritmo para determinação dos centros’ não é significativo ao nível de 0,05. Desta forma, tanto a configuração apontada pela análise como a que apresentou melhores resultados podem ser escolhidas indistintamente. Ambas as configurações para o referido experimento estão incluídas nas tabelas 7.26 e 7.27, que relacionam as melhores configurações obtidas.

A mesma análise realizada para vida da ferramenta que considerou a significância do efeito de cada fator e também o princípio da Hierarquia (Montgomery, 2001) foi repetida para rugosidade da peça. O resultado desta análise apontou, para cada experimento realizado para Rugosidade Média durante este ciclo experimental, quais efeitos são significativos para o desempenho da rede. A tabela 7.28 apresenta de forma sumária, quais são estes efeitos para cada número de casos no conjunto de treinamento.

Tabela 7.28 – Sumário dos efeitos significativos para Rugosidade Média (R_a), obtidos dos experimentos definitivos com fatoriais completos

Termo	Número de casos no conjunto de treinamento											
	12	15	24	30	48	60	120	150	240	300	400	500
Constante	X	X	X	X	X	X	X	X	X	X	X	X
Alg. det. parâmetro largura	X	X	X	X	X	X	X	X	X	X	X	X
No. unidades radiais	X	X	X	X	X	X	X	X	X	X	X	X
Alg. det. centros	X	X			X	X	X		X		X	X
Alg. det. parâmetro largura *No. unidades radiais	X	X	X		X	X	X	X	X	X	X	X
Alg. det. parâmetro largura *Alg. det. centros	X	X			X	X	X				X	X
No. unidades radiais *Alg. det. centros	X						X				X	
Alg. det. parâmetro largura *No. unidades radiais *Alg. det. centros											X	

A tabela 7.29 mostra os valores médios da resposta Razão de Desvios Padrão, seu desvio padrão e erro padrão da média, para as melhores configurações de rede obtidas em cada experimento. Nota-se que os valores médios da resposta, para as melhores configurações de rede obtidas, caem a cada aumento do número de casos disponível para treinamento. Deve-se lembrar que os resultados mostrados na tabela 7.29 foram obtidos com casos de teste, ou seja, casos que não tomaram parte no treinamento da rede e foram apresentados à mesma

somente durante a fase de testes. Estes resultados aferem de maneira independente a capacidade de generalização das redes obtidas.

Pode-se notar uma queda expressiva nos valores médios da resposta Razão de Desvios Padrão já a partir de 15 casos de treinamento e uma redução contínua naquele valor com o aumento do número de casos. Os valores obtidos são extremamente baixos, chegando mesmo à casa de 10^{-5} , o que constitui excelente resultado na tarefa de predição.

Em relação à variabilidade das redes, nota-se que o desvio padrão das melhores configurações obtidas cai com o aumento do número de casos de treinamento, sem exceções. Até mesmo a melhor configuração obtida durante o experimento com 60 casos de treinamento apresentou um desvio padrão menor do que a melhor rede obtida durante o experimento com 48 casos de treinamento. Este resultado difere daquele obtido durante o correspondente experimento com arranjos ortogonais de Taguchi. A diferença entre as configurações de rede obtidas nos dois experimentos é o número de unidades radiais. No experimento com o uso de fatoriais completos, um dos níveis testados foi uma quantidade de unidades radiais correspondente a $\frac{3}{4}$ dos casos de treinamento disponíveis. Tal proporção não havia sido testada durante o ciclo experimental com arranjos ortogonais de Taguchi.

Tabela 7.29 – Valores médios da resposta Razão de Desvios Padrão, desvios padrão e erros-padrão da média para as melhores redes obtidas para predição da Rugosidade durante o ciclo experimental definitivo com uso de fatoriais completos

Número de casos no conjunto de treinamento	Valor médio da resposta Razão de Desvios Padrão para a configuração proposta	Desvio Padrão para a resposta Razão de Desvios Padrão	Erro-padrão da média
12	0,334699	0,226554	0,676889
15	0,047284	0,018055	0,381842
24	0,018287	0,005806	0,317493
30	0,014064	0,002929	0,208262
48	0,001678	0,000873	0,520262
60	0,000815	0,000298	0,365644
120	0,000569	0,000256	0,449912
150	0,00036	0,000158	0,438889
240	0,000143	5,25E-05	0,367133
300	5,31E-05	2,16E-05	0,40678
400	4,29E-05	1,94E-05	0,452214
500	3E-05	1,4E-05	0,466667

Em relação à tabela 7.29, a simples observação da mesma não permite afirmar que uma resposta difere de outra. Com o propósito de comparar estatisticamente as melhores configurações obtidas, foram aplicados testes para variância e para as médias da resposta para cada uma das configurações.

Para testar-se a igualdade das médias obtidas pelas melhores configurações, foram realizados testes de análise de variância entre todos os pares possíveis entre as melhores configurações de rede. Por exemplo, a média da resposta da melhor rede obtida para 12 casos de treinamento foi comparada por meio da ANOVA à média da resposta da melhor rede obtida para 15 casos. Assim foi feito entre todas as combinações possíveis de conjuntos de treinamento. Os resultados dos testes podem ser observados na tabela 7.30, em que cada célula contém o valor P resultante de um teste de ANOVA. Na tabela, um valor P superior ao nível de significância adotado indica que não há evidência estatística de diferença entre as médias e um valor P inferior a 0,05 indica forte evidência de diferença entre as médias. Pode-se notar que, ao nível de significância adotado existe evidência de que todas as médias obtidas pelas melhores configurações de rede para cada número de casos do conjunto de treinamento são diferentes. O único valor diferente de zero é o do teste envolvendo as melhores configurações de rede obtidas para 300 e 400 casos de treinamento, que é de 0,008. O valor P é muito inferior ao nível de significância adotado, fornecendo evidência de que, também nesse caso, as médias não são iguais.

Tabela 7.30 – Valores P obtidos dos testes de ANOVA para comparação das médias das respostas das melhores configurações de rede para Rugosidade Média

		Número de casos no conjunto de treinamento →									
↓	12	15	24	30	48	60	120	150	240	300	400
12	--	--	--	--	--	--	--	--	--	--	--
15	0,000	--	--	--	--	--	--	--	--	--	--
24	0,000	0,000	--	--	--	--	--	--	--	--	--
30	0,000	0,000	0,000	--	--	--	--	--	--	--	--
48	0,000	0,000	0,000	0,000	--	--	--	--	--	--	--
60	0,000	0,000	0,000	0,000	0,000	--	--	--	--	--	--
120	0,000	0,000	0,000	0,000	0,000	0,000	--	--	--	--	--
150	0,000	0,000	0,000	0,000	0,000	0,000	0,000	--	--	--	--
240	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	--	--	--
300	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	--	--
400	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,008
500	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000

Para comparar as variâncias entre as configurações, foram aplicados testes de Levene para hipótese nula de variâncias iguais entre todos os pares de respostas possíveis da tabela 7.29 ao nível de significância de 0,05. Por exemplo, a variância da melhor rede obtida para 12 casos de treinamento foi comparada por meio do teste de Levene à variância da melhor rede obtida para 15 casos de treinamento. Assim foi feito para todos os pares das melhores redes. A figura 7.14 ilustra, a título de exemplo, o resultado do teste de comparação de variâncias

entre as melhores configurações de rede obtidas durante os experimentos realizados para Rugosidade Média com a utilização de 60 e 120 casos de treinamento.

Os resultados dos testes para comparação das variâncias aplicados a todos os pares podem ser observados na tabela 7.31, em que cada célula contém o valor P resultante de um teste de Levene. Na tabela, um valor P superior ao nível de significância adotado indica que não há evidência estatística de diferença entre as variâncias e um valor P inferior a 0,05 indica forte evidência de diferença entre as variâncias.

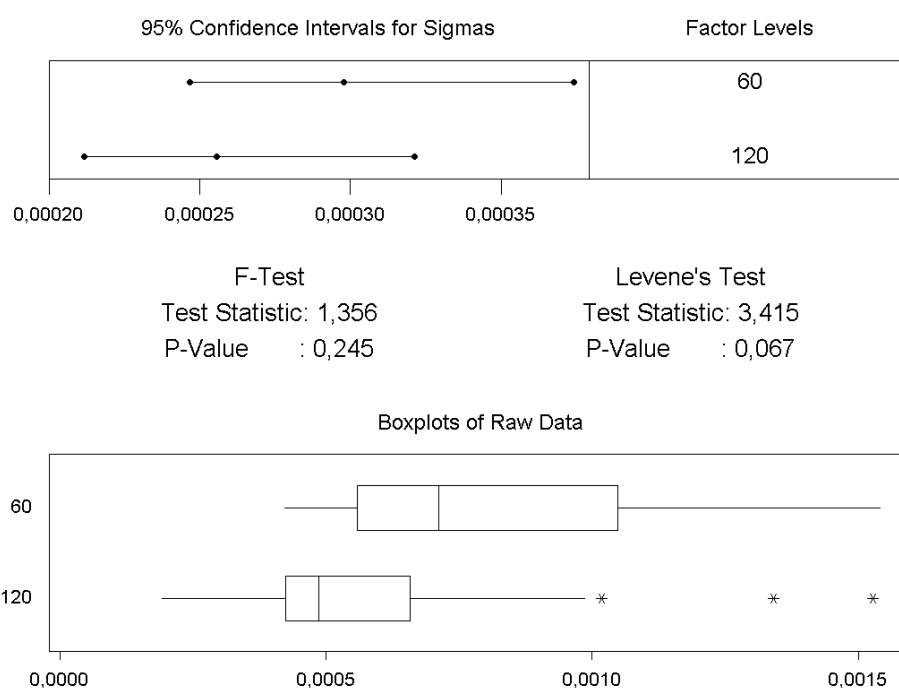


Figura 7.14 – Resultado do teste de comparação entre as variâncias das melhores configurações de rede obtidas para Rugosidade Média durante experimentos utilizando 60 e 120 casos de treinamento. Fonte: Minitab 13.0®

A análise da tabela 7.31 mostra que, ao nível de significância adotado, não há evidência de diferença entre as variâncias das melhores redes obtidas para 48 e 60 casos de treinamento. O mesmo se conclui para as melhores redes obtidas para 60 e 120 casos de treinamento, e também para as melhores redes obtidas durante os experimentos com 300 e 400 casos de treinamento. Todos os demais resultados indicados na tabela apontam para evidência, ao nível de 0,05, de diferença entre as variâncias das melhores redes.

Observa-se, portanto, que entre as melhores redes obtidas para cada experimento, há uma tendência inequívoca de redução na variância à medida que aumenta o número de casos de treinamento disponíveis.

Tabela 7.31 – Valores P resultantes do teste de Levene entre pares de amostras para as melhores configurações obtidas em cada experimento do ciclo experimental definitivo para Rugosidade Média (R_a) com o uso de fatoriais completos

No. casos	Var.	Número de casos no conjunto de treinamento											
		12	15	24	30	48	60	120	150	240	300	400	500
12	0,051327	--	--	--	--	--	--	--	--	--	--	--	--
15	0,000326	0,000	--	--	--	--	--	--	--	--	--	--	--
24	3,37E-05	0,000	0,000	--	--	--	--	--	--	--	--	--	--
30	8,6E-06	0,000	0,000	0,000	--	--	--	--	--	--	--	--	--
48	8E-07	0,000	0,000	0,000	0,000	--	--	--	--	--	--	--	--
60	8,87E-08	0,000	0,000	0,000	0,000	0,052	--	--	--	--	--	--	--
120	6,54E-08	0,000	0,000	0,000	0,000	0,010	0,067	--	--	--	--	--	--
150	2,49E-08	0,000	0,000	0,000	0,000	0,010	0,000	0,025	--	--	--	--	--
240	2,75E-09	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	--	--	--	--
300	4,7E-10	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	--	--	--
400	3,8E-10	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,341	--	--
500	2E-10	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,002	0,036	--

Em resumo, observa-se que as melhores redes obtidas em cada experimento apresentam bons resultados na tarefa de predição da Rugosidade Média. A figura 7.15 mostra os diagramas de Box-plot para as melhores configurações obtidas. Os pontos indicam a média da resposta da melhor configuração de rede obtida para um conjunto de treinamento. Os asteriscos indicam *outliers*. Os limites de cada caixa indicam, a partir de baixo, o 1º, 2º e 3º quartis das observações de predição de cada rede. Na figura 7.16 foram retirados os gráficos das melhores redes obtidas para os experimentos com 12 e 15 casos, com o objetivo de aumentar a definição do gráfico para os demais experimentos.

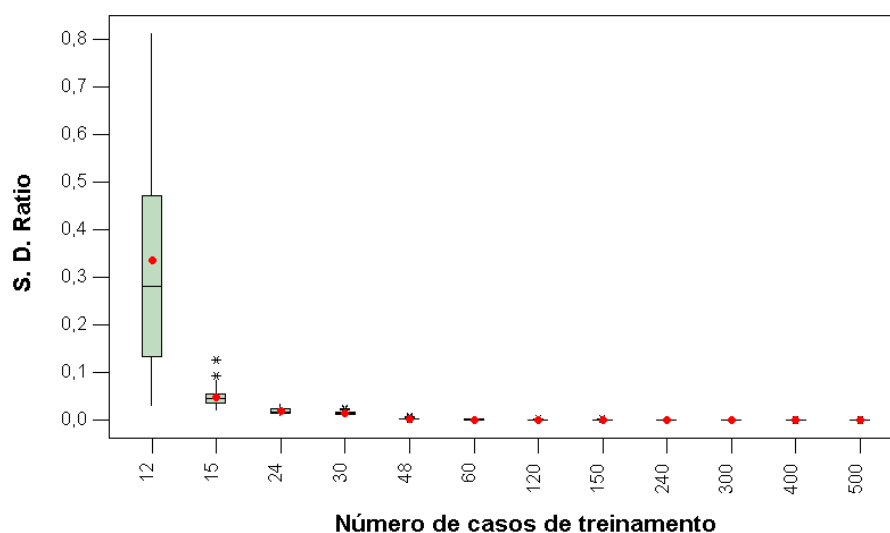


Figura 7.15 – Box-plot das melhores configurações de rede obtidas para Rugosidade Média durante os experimentos definitivos com arranjos fatoriais completos. Fonte: Minitab 13.0®

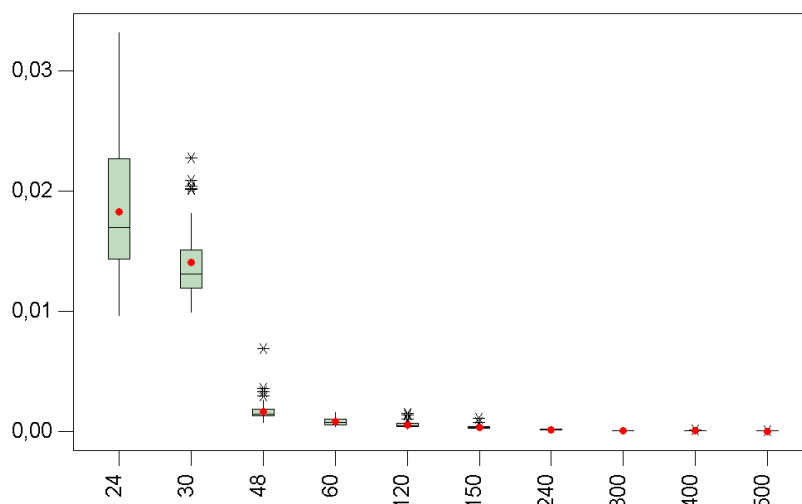


Figura 7.16– *Box-plot* das melhores configurações de rede de 24 a 500 casos obtidas para Rugosidade Média durante os experimentos definitivos com arranjos fatoriais completos. Fonte: Minitab 13.0®

Por fim, comparou-se o desempenho das melhores redes obtidas em cada experimento ao desempenho de um modelo de otimização linear, o algoritmo da matriz Pseudo-Inversa (Bishop, 1995; Golub and Kahan, 1965). O resultado é mostrado na figura 7.17.

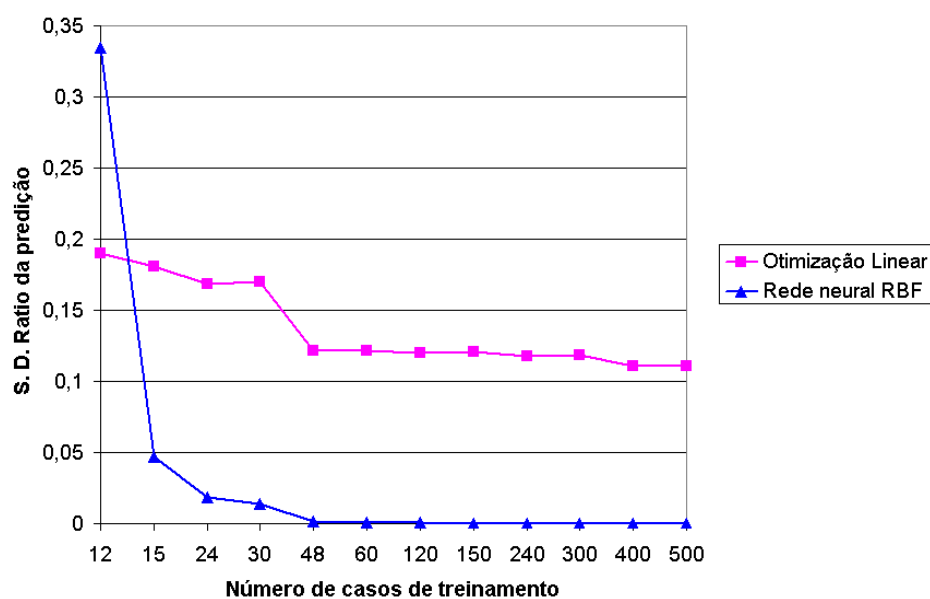


Figura 7.17 – Comparação de desempenho da predição da Rugosidade Média feita entre as melhores configurações de rede e o método da matriz Pseudo-Inversa para otimização linear. Fonte: Excel®

A análise do gráfico indica que o desempenho do método de otimização linear é superior ao da rede neural apenas na situação em que 12 casos estão disponíveis para a modelagem. Em todos os demais experimentos realizados com um número maior de casos a rede neural apresentou desempenho muito superior ao método de otimização linear.

8 CONCLUSÕES

8.1 Considerações Iniciais

O presente trabalho propôs-se a realizar um estudo sobre o desempenho e a variabilidade de redes neurais artificiais de arquitetura RBF (função de base radial) projetadas com auxílio da metodologia de projeto de experimentos (DOE) e aplicadas à tarefa de predição da vida da ferramenta e da rugosidade da peça em processo de torneamento do aço SAE/ABNT 52100, 55 HRC, com ferramenta de cerâmica mista. Os parâmetros de corte do processo de usinagem foram utilizados como entradas das redes. O valor da Razão de Desvios Padrão foi utilizado como medida de desempenho na predição. Foram utilizados casos de treinamento gerados a partir de modelos polinomiais quadráticos completos, obtidos e validados em experimento efetuado segundo a Metodologia da Superfície de Resposta. Experimentos com diferentes tamanhos do conjunto de treinamento das redes foram executados e seus resultados analisados.

As conclusões obtidas do estudo são apresentadas a seguir. As mesmas são baseadas nos resultados dos experimentos realizados durante o terceiro ciclo experimental, realizado com o uso da técnica de Fatoriais Completos.

8.2 Conclusões sobre vida da ferramenta e rugosidade da peça

Conclui-se que as redes RBF projetadas segundo a metodologia de projeto de experimentos podem apresentar alto grau de exatidão e pequena variabilidade na tarefa de predição da vida da ferramenta e da rugosidade. Para vida da ferramenta a metodologia aplicada apontou, no melhor caso, para uma rede com média da resposta igual a $2,66 \cdot 10^{-5}$ e desvio padrão igual a $1,15 \cdot 10^{-5}$. Para rugosidade a metodologia apontou, no melhor caso, para uma rede com média da resposta igual a $3,0 \cdot 10^{-5}$ e desvio padrão igual a $1,4 \cdot 10^{-5}$.

Os resultados indicam graus de exatidão e precisão elevados na predição das grandezas sob estudo. Além disso, os resultados e a análise estatística realizada permitem concluir que a exatidão das melhores redes obtidas durante os experimentos realizados segundo a metodologia DOE aumenta com o aumento do número de casos de treinamento disponíveis. Os resultados e a análise estatística permitem concluir também que a dispersão das melhores redes obtidas em cada experimento é reduzida com o aumento do número de

casos de treinamento disponíveis. Estas conclusões são válidas para predição da vida da ferramenta e para predição da rugosidade da peça.

A melhora do desempenho das redes com o aumento do número de casos é prevista pela literatura, mas a mesma não aponta métodos para obtenção de redes com bom desempenho. A metodologia de projeto de experimento tornou possível neste estudo identificar parâmetros de projeto que resultaram em redes com exatidão e precisão crescentes com o aumento do número de exemplos disponíveis, concordando assim com a teoria. Os resultados sugerem que a metodologia de projeto de experimentos pode constituir uma abordagem sistemática de projeto superior à procura de configurações por tentativa e erro, ou à estratégia de variar um parâmetro por vez.

Os resultados obtidos mostram que redes treinadas para predição da vida da ferramenta com apenas 24 casos de treinamento podem apresentar um valor de Razão de Desvios Padrão igual a 0,014712, ou seja, uma exatidão da ordem de 1%. Em relação à rugosidade os resultados fornecem evidências de que redes treinadas com apenas 24 casos de treinamento podem apresentar um valor de Razão de Desvios Padrão igual a 0,018287, o que também indica uma exatidão da ordem de 1% na tarefa de predição. Pode-se concluir também que redes RBF representam uma alternativa econômica para a predição, vez que mesmo com um número relativamente baixo de casos de treinamento foi possível obter-se bom desempenho.

Notou-se que as médias e valores dos desvios padrão das melhores redes obtidas para predição da rugosidade são maiores do que os valores correspondentes das melhores redes obtidas para predição da vida da ferramenta. Ainda que os resultados tenham sido excelentes para ambos os casos, tal observação indica um desempenho ligeiramente superior das redes obtidas para predição da vida da ferramenta na comparação com as redes obtidas para predição da rugosidade.

A análise dos resultados permite concluir que os fatores utilizados têm influência significativa no desempenho das redes. O algoritmo de determinação do parâmetro de largura da função radial foi um fator significativo em todos os experimentos para vida da ferramenta e para rugosidade, assim como o fator número de unidades radiais. Os resultados das análises indicam que o algoritmo para determinação do parâmetro de largura da função radial é o fator com maior importância relativa no desempenho das redes RBF para as tarefas propostas. A influência de tal fator se torna dominante com o aumento do número de casos de treinamento disponíveis, como se observa pelos diagramas de Pareto e pelos valores dos efeitos medidos. O segundo fator mais influente foi o número de unidades radiais.

O fator menos influente foi o algoritmo para determinação dos centros das funções radiais. A influência deste último fator foi reduzida em quase todos os experimentos, tendendo à insignificância nos experimentos com maior número de casos de treinamento. Esta constatação sugere que a abundância dos exemplos faz com que qualquer dos dois algoritmos testados possa ser usado indistintamente na busca por redes que apresentem bom desempenho na predição, desde que os casos de treinamento representem uniformemente o domínio da função modelada. Tal conclusão pode simplificar o projeto de redes RBF para as tarefas propostas.

É possível concluir que os efeitos de interação entre níveis dos fatores envolvidos nos experimentos são significativos para o desempenho das redes RBF na tarefa de predição da vida da ferramenta e da rugosidade. Em cada um dos experimentos conduzidos para vida da ferramenta e em onze dos doze experimentos conduzidos para rugosidade ao menos um efeito de interação foi apontado como significativo pela análise de variância. Em alguns casos efeitos de interação tripla foram considerados significativos. Pôde-se observar também que há casos em que os efeitos de interação são mais pronunciados que os correspondentes efeitos principais do nível do fator.

Entre os efeitos de interação observados, a existente entre o algoritmo para determinação do parâmetro de largura da função radial e o número de unidades radiais foi apontada como significativa em onze dos doze experimentos para vida da ferramenta e em onze dos doze experimentos para rugosidade. A importância relativa deste efeito também foi pronunciada, rivalizando com, e por vezes superando, o efeito principal do número de unidades radiais. A interação entre o algoritmo para determinação do parâmetro de largura da função radial e o algoritmo para determinação dos centros foi apontada como significativa em dez dos doze experimentos para vida da ferramenta e em sete dos doze experimentos para rugosidade. Seu efeito, porém, foi bem menos pronunciado do que a interação anteriormente citada. As demais interações foram apontadas como significativas em um número menor de experimentos, e sua influência relativa não foi acentuada.

Observou-se que o algoritmo Isotrópico para determinação do parâmetro de largura da função radial com fator de escala igual a 10 foi o que levou a rede aos melhores resultados em todos os experimentos realizados, tanto para vida da ferramenta quanto para rugosidade superficial.

A análise dos resultados apontou também as melhores proporções entre o número de unidades radiais e o número de casos de treinamento para cada conjunto de treinamento utilizado. Para vida da ferramenta observa-se um melhor desempenho para redes com

proporções de $\frac{3}{4}$ entre número de unidades e casos de treinamento nos experimentos com 12 e 15 casos. Em todos os demais experimentos para vida da ferramenta uma proporção de $\frac{1}{2}$ foi a que apresentou melhor desempenho.

Nos experimentos realizados para rugosidade observou-se que nos experimentos envolvendo 12, 15, 24, 48 e 60 casos a melhor proporção entre unidades e quantidade de casos foi igual a $\frac{3}{4}$ enquanto que nos demais experimento essa proporção foi igual a $\frac{1}{2}$. Observa-se que para os experimentos contendo um número de casos de treinamento igual ou superior a 120 a proporção de $\frac{1}{2}$ entre unidades radiais e casos de treinamento foi a que produziu melhores resultados para vida da ferramenta e também para rugosidade.

Os resultados da comparação entre valores de Razão de Desvios Padrão das melhores redes obtidas e curvas correspondentes ajustadas pelo algoritmo da decomposição por valor singular indicaram que em todos os experimentos, exceto aqueles com 12 casos de treinamento, as redes neurais apresentaram melhores resultados na predição, tanto para vida da ferramenta quanto para rugosidade. As redes neurais mostraram-se capazes de modelar de modo exato relações de natureza não linear existentes entre variáveis independentes e dependentes nos problemas sob estudo. Este fato sugere que as redes RBF são uma ferramenta poderosa para a tarefa de modelagem, desde que a correta configuração de rede seja identificada e implementada.

Um ponto a ser destacado é a praticidade do uso de redes RBF para tarefas de predição. O projeto das redes é simples e pode levar a redes com bom desempenho com a definição de apenas três parâmetros, como feito neste trabalho. Há inclusive casos em que o algoritmo para determinação dos centros não apresenta influência significativa, ou apresenta influência ínfima, o que reduziria a dois o número de parâmetros de projeto. Além disso, o treinamento das redes RBF é extremamente rápido. Este fato foi de grande utilidade para a execução dos experimentos. Possibilitou testar muitas configurações diferentes de redes e repetir os testes por um número elevado de vezes, com o objetivo de gerar resultados estatisticamente válidos, em um período de tempo reduzido.

Deve-se ressaltar que as conclusões obtidas neste trabalho não devem ser extrapoladas para outros tipos de redes neurais nem tampouco para dados referentes a outras operações de usinagem, outros materiais ou ferramentas.

Espera-se que os resultados e conclusões obtidos sejam uma contribuição para a melhoria do processo sob estudo. Que os mesmos possam contribuir para melhoria da produtividade na indústria por meio de um aproveitamento mais eficiente da vida das

ferramentas de corte e que a predição da rugosidade da peça em ambientes produtivos possa contribuir para a melhoria da qualidade de superfície dos produtos usinados.

8.3 Recomendações para trabalhos futuros

O presente trabalho pode ser continuado de várias maneiras. O mesmo estudo ora realizado pode ser refeito para redes de arquitetura MLP, para redes de arquitetura GRNN, ou ainda para alguma outra arquitetura, com o objetivo de identificar configurações de rede que apresentem bom desempenho nas tarefas de predição da vida e da rugosidade. Pode ser feito ainda estudo comparativo entre os desempenhos obtidos pelas melhores configurações de cada arquitetura de rede.

Pode-se também elaborar um estudo da predição da vida e da rugosidade com a mesma metodologia aqui aplicada, alterando-se, porém, o processo de usinagem sob estudo. Outros processos elegíveis para serem os processos de furação e fresamento, por exemplo.

Uma das restrições encontradas para a realização deste trabalho foi a falta de um banco de dados contendo dados sobre vida da ferramenta e sobre rugosidade que pudesse ser utilizado para formar os conjuntos de treinamento e de testes para as redes neurais. Recomenda-se que a metodologia e, ou, as configurações de rede obtidas neste estudo sejam utilizadas na predição da vida da ferramenta e da rugosidade utilizando-se dados reais, caso seja possível obtê-los. Há que se lembrar que uma base de dados para treinamento de redes RBF deve conter uma distribuição uniforme de casos de treinamento que cubra o espaço experimental.

Finalmente, recomenda-se que as redes RBF, pelo excelente desempenho e praticidade demonstrados neste trabalho, sejam consideradas como alternativas para a solução problemas de outros domínios da engenharia envolvendo a modelagem de relações não-lineares entre variáveis independentes e dependentes.

APÊNDICE A – Tópicos em Estatística e Simulação

A.1 Inferência sobre a média de uma população com variância conhecida

A média amostral \bar{X} é um estimador não tendencioso da média μ de uma população, desde que se conheça a variância σ da mesma. Desde que se apliquem as condições do teorema central do limite, a distribuição de \bar{X} é aproximadamente normal com média μ e variância σ/\sqrt{n} , onde n é o tamanho da amostra. Para testar-se a hipótese nula $\mu = \mu_0$, usa-se a estatística de teste dada pela equação (A.1) (Montgomery e Runger, 2003).

$$Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \quad (\text{A-1})$$

Rejeita-se a hipótese nula de que as médias são iguais caso o valor P resultante seja inferior ao nível de significância adotado. O teste, ainda segundo Montgomery e Runger (2003), pode ser utilizado desde que o tamanho da amostra seja superior a trinta.

A.2 Teste de Levene

O teste de Levene é utilizado para testar-se a igualdade de variâncias entre amostras. É testada, a um dado nível de significância, a hipótese nula de que as variâncias das amostras envolvidas são iguais. Tal teste é recomendado quando não há evidência de que as amostras testadas sigam a distribuição normal. A hipótese nula de que as variâncias são iguais é rejeitada caso ocorra a situação descrita pela equação (A-2) (NIST/SEMATECH, 2006).

$$W > F_{(\alpha, k-1, N-k)} \quad (\text{A-2})$$

onde F é o valor crítico superior da distribuição F com $k-1$ e $N-k$ graus de liberdade, ao nível de significância de α e W é estatística de teste dada pela equação (A-3) (NIST/SEMATECH, 2006).

$$W = \frac{(N - k) * \sum_{i=1}^k N_i (\bar{Z}_i - \bar{Z}_{..})^2}{(k - 1) * \sum_{i=1}^k \sum_{j=1}^{N_i} (\bar{Z}_{ij} - \bar{Z}_i)^2} \quad (\text{A-3})$$

onde N é o número total de leituras envolvidas na amostras sob teste, k é o número de amostras sendo comparadas, N_i é o número de leituras em cada amostra, $\bar{Z}_{..}$ é média total das amostras, \bar{Z}_i corresponde à média de uma amostra e \bar{Z}_{ij} é dado pela equação (A-4) (NIST/SEMATECH, 2006).

$$Z_{ij} = |Y_{ij} - \bar{Y}_i| \quad (\text{A-4})$$

onde Y_{ij} corresponde à leitura j da i -ésima amostra e \bar{Y}_i à média da i -ésima amostra.

A.3 Diagrama de Pareto

O Diagrama de Pareto é um histograma ordenado de maneira decrescente que permite expressar graficamente uma variação. O Diagrama de Pareto procura separar as poucas causas vitais das muitas triviais (PAIVA, 2004).

Dentro da metodologia de Projeto de Experimentos, este diagrama é usado para demonstrar quais efeitos e interações são significativos para um determinado modelo.

Quando há réplicas dos experimentos a linha de referência desenhada sobre o diagrama de Pareto é dada pela estatística $T(\alpha; df(MSe))$, onde α é o nível de significância e $df(MSe)$ representa o número de graus de liberdade do Erro Médio Quadrático. Qualquer efeito padronizado que ultrapasse esse valor é tido como significativo. O efeito padronizado é encontrado dividindo-se o valor do coeficiente do fator pelo respectivo erro padrão (SE). A figura A-1 ilustra o conceito do diagrama de Pareto.

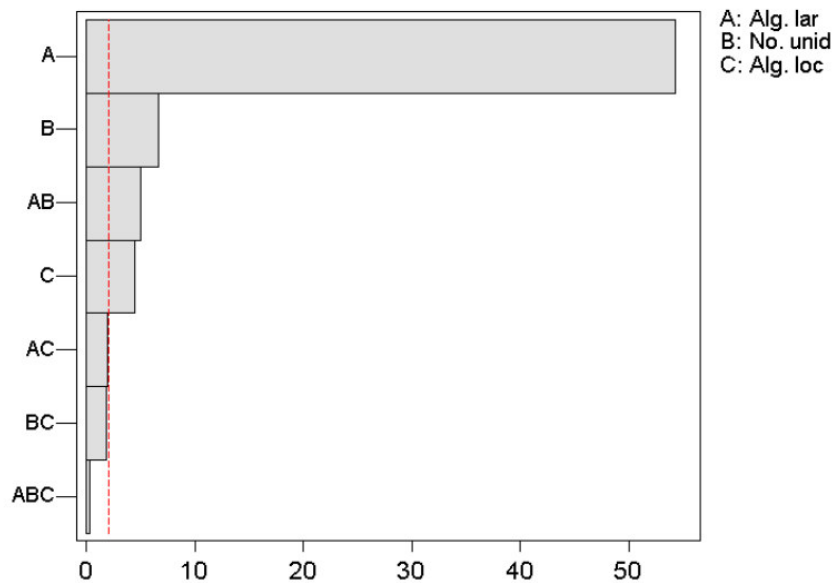


Figura A-1 – Diagrama de Pareto. Fonte: Minitab® Statistical Software Release 13.0

A.4 Transformações de Box-Cox

Ao realizar uma regressão linear de Y em função de X , uma transformação apropriada de X pode incrementar significativamente o ajuste da curva obtida. As transformações de Box e Cox (1964) constituem uma família muito útil de transformações. Tais transformações são definidas pela equação (A-5) (NIST/SEMATECH, 2006).

$$T(x) = (X^\lambda - 1) / \lambda \tag{A-5}$$

onde X é a variável transformada e λ é o parâmetro de transformação. Para $\lambda = 0$, os logaritmos naturais dos dados são usados ao invés da fórmula.

A escolha do valor apropriado de λ é feito com o auxílio do gráfico de linearidade de Box-Cox. Este gráfico mostra a correlação entre Y e os valores da transformada de X em função de valores de λ . No gráfico, λ corresponde às coordenadas do eixo horizontal e o valor da correlação entre Y e a transformada de X às coordenadas do eixo vertical. O valor de λ que corresponde à máxima correlação (ou mínima no caso de correlação negativa) no gráfico é a escolha ótima para λ . A figura A-2 ilustra o conceito da transformação de Box-Cox (NIST/SEMATECH, 2006).

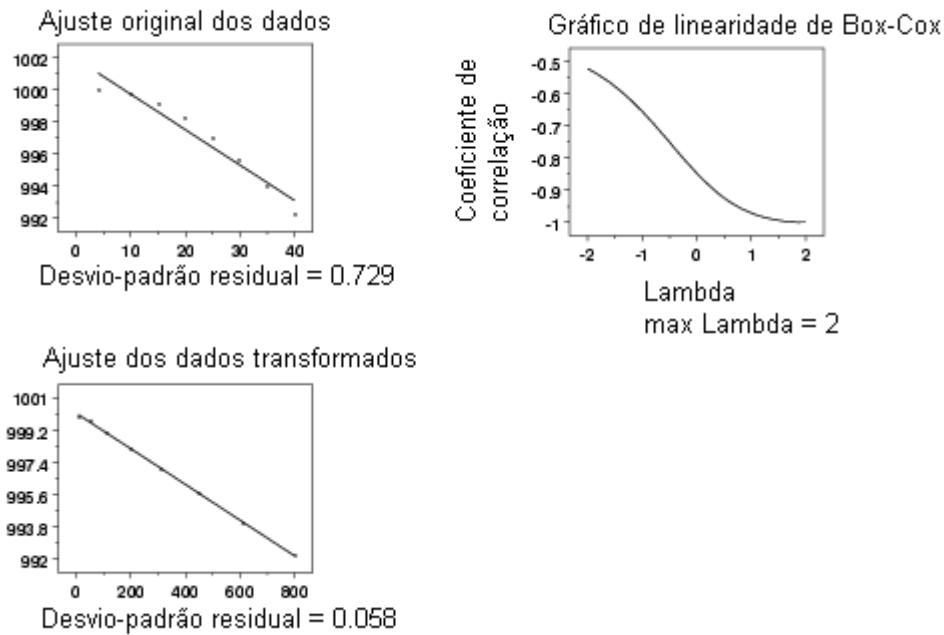


Figura A-2 – Ilustração do conceito da transformação de Box-Cox. Fonte: Adaptado de NIST/SEMATECH (2006)

A.5 Meta-modelos e simulação

Devido ao poder computacional requerido e ao grande número de parâmetros requeridos, o uso de modelagem determinística em aplicações de grande escala ou em estudos de otimização é muitas vezes inviável. Uma alternativa é o uso de meta-modelos, os quais podem ser construídos para aplicações específicas a partir de um número limitado de simulações determinísticas. Piñeros Garcet et al (2006) propuseram uma definição matemática de meta-modelos, bem como um método para sua definição e validação. Seja uma função numérica f_N como mostrado na equação (A-6) (Piñeros Garcet et al., 2006).

$$\vec{y}_N = f_N(\vec{x}_N) \tag{A-6}$$

onde \vec{y}_N é a saída numérica do modelo; \vec{x}_N é o vetor das entradas numéricas do modelo. Os elementos do vetor de entradas são $\vec{x}_N = (x_N^1; \dots; x_N^i; \dots; x_N^n)$ com $\vec{x}_N \in \mathfrak{R}^n$ e n é a dimensão de \vec{x}_N .

Suponhamos que haja alguma restrição e nem todos os elementos de \mathfrak{R}^n sejam entradas possíveis para f_N ; algumas entradas devem ser positivas, por exemplo. Assim

$\bar{x}_N \in \mathcal{X}_N$, onde \mathcal{X}_N corresponde ao conjunto de todas as possíveis entradas para o modelo numérico f_N . Dado um vetor de entrada $\bar{x}_{N,i}$, é possível transformá-lo usando uma função g como mostrado na equação (A-7) (Piñeros Garcet et al., 2006).

$$\bar{x}_{M,i} = g(\bar{x}_{N,i}) \quad (\text{A-7})$$

No processo de meta-modelagem, o propósito é o de reduzir a dimensionalidade do problema através da redução do número de variáveis, de modo que g transforme $\bar{x}_{N,i}$ em um vetor $\bar{x}_{M,i}$ de menor dimensão dado por ($\bar{x}_{M,i} \in \mathfrak{R}^m$ com $m < n$), por meio da exclusão de alguns elementos de $\bar{x}_{N,i}$. Uma outra definição da função de transformação g é dada pela equação (A-8) (Piñeros Garcet et al., 2006).

$$\bar{x}_{M,i} = (x_M^1; \dots; x_M^M) = g(\bar{x}_{N,i}) = g(x_N^1; \dots; x_N^n) \quad (\text{A-8})$$

Um meta-modelo f_M é uma simplificação de um modelo numérico f_N : que utiliza ou um sub-conjunto das entradas numéricas do modelo numérico, como na equação (A-9), ou funções de um sub-conjunto de entradas do mesmo modelo, como na equação (A-10) (Piñeros Garcet et al., 2006).

$$\bar{y}_N = f_M(x_N^1; \dots; x_N^m) + \varepsilon; \quad \text{sendo } m < n \quad (\text{A-9})$$

ou

$$\bar{y}_N = f_M(g(x_N^1; \dots; x_N^m)) + \varepsilon; \quad \text{sendo } m < n \quad (\text{A-10})$$

onde ε é um termo que inclui tanto o erro proveniente da exclusão de $n - m$ fatores como o erro de ajuste do meta-modelo. A figura (A-3) ilustra as relações entre meta-modelos e modelos numéricos.

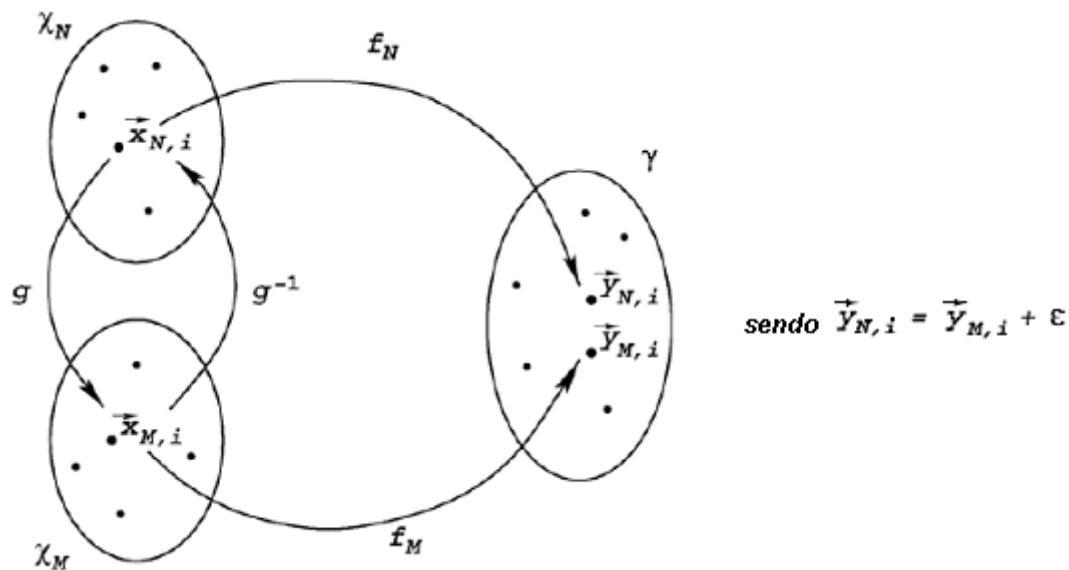


Figura A.3 – Ilustração do conceito de meta-modelos. Adaptado de Piñeros Garcet et al.(2006)

A construção de meta-modelos envolve oito etapas, listadas a seguir (Piñeros Garcet et al., 2006):

- 1) Escolher as variáveis de entrada do meta-modelo e, eventualmente, a forma da função g ;
- 2) Escolher a forma matemática de f_M ;
- 3) Criar um conjunto de calibração $\{\bar{x}_{M,1}, \dots, \bar{x}_{M,k}, \dots, \bar{x}_{M,p}\}$ de vetores de entrada que sigam algum dado projeto de experimentos;
- 4) Criar um conjunto de validação $\{\bar{x}_{M,p+1}, \dots, \bar{x}_{M,p+k}, \dots, \bar{x}_{M,r}\}$ de vetores de entrada que sigam algum dado projeto de experimentos. O objetivo do conjunto de validação é obter uma medida quantitativa da qualidade do meta-modelo escolhido;
- 5) Realizar um experimento numérico: executar f_N sobre os conjuntos de calibração e validação para obter os correspondentes valores de saída de calibração e validação $\{\bar{y}_{M,1}, \dots, \bar{y}_{M,k}, \dots, \bar{y}_{M,p}\}, \{\bar{y}_{M,p+1}, \dots, \bar{y}_{M,p+k}, \dots, \bar{y}_{M,r}\}$;
- 6) Calibrar f_M com a utilização de $\{\bar{x}_{M,1}, \dots, \bar{x}_{M,k}, \dots, \bar{x}_{M,p}\}$ e de $\{\bar{y}_{M,1}, \dots, \bar{y}_{M,k}, \dots, \bar{y}_{M,p}\}$;
- 7) Utilizando o modelo calibrado, calcular $\{\bar{y}_{M,p+1}, \dots, \bar{y}_{M,p+k}, \dots, \bar{y}_{M,r}\}$, onde $\forall i = p + 1, \dots, r, \bar{y}_{M,i} = f_M(\bar{x}_{M,i})$;
- 8) Validar o modelo f_M utilizando os indicadores do modelo calculados a partir da diferença entre $(\bar{y}_{M,i} - \bar{y}_{N,i}), \forall i = p + 1, \dots, r$.

Segundo os autores, a confiabilidade dos meta-modelos desenvolvidos segundo o método proposto depende da validade dos modelos numéricos dos quais são derivados, e é limitada ao domínio do espaço de entrada do qual o modelo numérico é derivado. São, em suma, interpoladores. (PIÑEROS GARCET et al, 2006).

REFERÊNCIAS BIBLIOGRÁFICAS

ABU-MAHFOUZ, I. Drilling wear detection and classification using vibration signals and artificial neural network. *International Journal of Machine Tools and Manufacture*, V43, no. 7, p.707-720, 2003.

AGOSTINHO, O. L.; RODRIGUES, A. C.; LIRANI, J. *Tolerâncias, Ajustes, Desvios e Análise de Dimensões*. São Paulo: Editora Edgar Blucher, LTDA., 1990, 295p.

ALAJMI, M. S.; ORABY, S. E.; ESAT, I. I. Neural network models on the prediction of tool wear in turning process: a comparison study. *Proceedings of the IASTED Conference Artificial Intelligence and Applications*, Innsbruck, Austria, p.301-309, 2005.

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. 114p. Dissertação de mestrado – UFRGS – Universidade Federal do Rio Grande do Sul.

ARENAS MIRANDA, G. A. *Uma contribuição ao processo de furação sem fluido de corte com broca de metal duro revestida com TiAlN*. 2003. 160p. Tese de Doutorado – UNICAMP – Universidade Estadual de Campinas.

AXINTE, D. A. Approach into the use of probabilistic neural networks for automated classification of tool malfunctions in broaching. *International Journal of Machine Tools and Manufacture*. In Press, Corrected Proof, Available online 2 November 2005.

BASEGROUP LAB. Neural networks. *Introduction to RBF networks*. <http://www.basegroup.ru/english.htm>, página acessada em 12 de Setembro, 2006.

BATISTA, E. A. *Desenvolvimento de um sistema especialista para otimização do processo de usinagem*. 2000. 194p. Dissertação de mestrado – UNIMEP – Universidade Metodista de Piracicaba.

BENARDOS, P. G.; VOSNIAKOS, G. C. Prediction of surface roughness in CNC in machining: a review. *International Journal of Machine Tools & Manufacture*, V43, 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*, V18, p.343-354, 2002.

BERTRAND, J. W. M.; FRANSOO, J. C. Operations management research methodologies using quantitative modeling. *International Journal of Operations & Production Management*. V22(2), p.241-264, 2002.

BISHOP, C. M. *Neural Networks for Pattern Recognition*. New York: Oxford University Press Inc., 1995, 482p.

BONIFÁCIO, M. E. R.; DINIZ, A. E. Monitoring the tool life in finish turning using vibration signals. *Revista Brasileira de Ciências Mecânicas*, v16(1), p.58-71, 1994.

BOX, G. E. P.; HUNTER, W. G.; HUNTER, J. S. *Statistics for Experimenters*. New York: John Wiley & Sons, 1978.

BOX, G. E. P.; COX, D. R. An Analysis of Transformations, *Journal of the Royal Statistical Society*, 211-243, discussion 244-252, 1964.

BRAGA, A. P.; CARVALHO, A. P. L. F.; LUDERMIR, T. B. *Redes Neurais Artificiais: Teoria e Aplicações*. Rio de Janeiro: LTC Livros Técnicos Editora S. A., 2000, 262p.

BROOMHEAD, D. S.; LOWE, D. Multivariable functional interpolation and adaptive networks. *Complex Systems*, v2, p.321-355, 1988.

BRYMAN, A. *Research Methods and Organization Studies*. Unwin Hyman, London, pp.2-50, 60-120, 1989.

CASALINO, G.; CURCIO, F.; MINUTOLO, F. M. C. Investigation on Ti6Al4V laser welding using statistical and Taguchi approaches. *Journal of Materials Processing Technology*, 167, p.422-428, 2005.

CHAO, P. Y.; HWANG, Y. D. An improved neural network model for the prediction of cutting tool life. *Journal of Intelligent Manufacturing*, V8, p.107-115, 1996.

CHIEN, W. T.; CHOU, C. Y. The predictive model for machinability of 304 stainless steel. *Journal of Materials Processing Technology*, V118, p.442-447, 2001.

CHOUDHURY, S.K.; BARTARYA, G. Role of temperature and surface finish in prediction tool wear using neural network and design of experiments. *International Journal of Machine Tools & Manufacture*, V43, p.747-753, 2003.

CHOUDHURY, I. A.; EL-BARADIE, M. A. Tool-life prediction model by design of experiments for turning high strength steel (290 BHN). *Journal of Materials Processing Technology*, V77, p.319-326, 1998.

CUS, F.; ZUPERL, U. Approach to optimization of cutting conditions by using artificial neural networks. *Journal of Materials Processing Technology*, In Press, Corrected Proof, Available online 18 January 2006.

CYBENKO, G. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems* v2, p.304-314, 1989.

DAS, S.; BANDYOPADHYAY, P. P.; CHATTOPADHYAY, A. B. Neural-networks-based tool wear monitoring in turning medium carbon steel using a coated carbide tool. *Journal of Materials Processing Technology*, V63, no.1-3, p.187-192, 1997.

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*, v116, 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*. São Paulo: Artliber, 2005, 248p.

DOS SANTOS, A. L. B. *Metodologia via redes neurais para a estimativa da rugosidade e do desgaste de ferramentas de corte no processo de fresamento frontal*. 2001. 187p. Tese de doutoramento – UFU – Universidade Federal de Uberlândia.

DURMUS, H. K.; OZKAYA, E.; MERI-Ç, C. The use of neural networks for the prediction of wear loss and surface roughness of AA 6351 aluminium alloy. *Materials and Design* 27, p.156-159, 2006.

EL-MOUNAYRI, H.; KISHAWY, H.; BRICENO, J. Optimization of CNC ball end milling: a neural network-based model. *Journal of Materials Processing Technology*, V166, no. 1, p.50-62, 2005.

ERZURUMLU, T.; OKTEM, H. Comparison of response surface model with neural network in determining the surface quality of moulded parts. *Materials and Design*. Available on-line, 21 October 2005.

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*, V45, 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*, V49, no.3-4, p.255-264, 1995.

FANG, X. D.; SAFI-JAHANSHAHI. A new algorithm for developing a reference model for predicting surface roughness in finishing machining of steels. *International Journal of Production Research*, V35 (1), p.179-197, 1997.

FERRARESI, D. *Fundamentos da usinagem dos metais*. São Paulo: Ed. Edgard Blucher Ltda., 1977, 751p.

FERREIRA, J. R. *Torneamento de Materiais Compósitos Reforçados com Fibras de Carbono*. 1999. 199p. Tese de Doutorado - Unicamp - Universidade Estadual de Campinas.

GHASEMPOOR-NOBANDGANY, A. *On-line wear estimation in turning*. 1998. 175p. Tese de Doutorado. Queen's University, Kingston, Ontario, Canadá.

GOLUB, G.; KAHAN, W. Calculating the Singular Values and Pseudo-Inverse of a Matrix. *Journal of the Society for Industrial and Applied Mathematics: Series B, Numerical Analysis*, V2, no.2, p.205-224, 1965.

HAYKIN, S. *Neural Networks - A comprehensive foundation*. New York: Mcmillan Publishing Company, 1994, 696p.

HEBB, D. O. *The Organization of Behavior*. Wiley, 1949.

HOPFIELD, J. J. Neural Networks and physical systems with emergent collective properties. *Proceedings of the National Academy of Sciences*, no.79, p.2554-2558, 1982.

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.

ISO 3685, *Tool Life Testing with Single Turning Point*, International Organization for Standardization, 1993.

KHAW, J. F. C.; LIM, B. S.; LIM, L. E. N. *Optimal design of neural networks using the Taguchi method, Neurocomputing*. V7, p.225-245, 1995.

KIM, Y. S.; YUM B. J. Robust Design of Multilayer Feedforward Neural Networks: an experimental approach. *Engineering applications of artificial intelligence*, V38, p.1-15, 2003.

KOVÁCS, Z. L. *Redes Neurais Artificiais - Fundamentos e Aplicações: um texto básico*. São Paulo: Colegium Cognition/Editora Edição Acadêmica, 1996, 173p.

LEE, J. H.; KIM, D. E.; LEE, S. J. Application of neural networks to flank wear prediction. *Mechanical Systems and Signal Processing* (1996) V10, p.265-276.

LIAO, H. A data envelopment analysis method for optimizing multi-response problem with censored data in the Taguchi method. *Computer & Industrial Engineering*, V46, p.817-835, 2004.

LIU, Q.; ALTINTAS, Y. On-line monitoring of flank wear in turning with multilayered feed-forward neural network. *International Journal of Machine Tools and Manufacture*, V39, no. 12, p.1945-1959, 1999.

MANUAL STATISTICA (Statistical Software) – Versão 7.1 – Statsoft Inc., 2005.

MANUAL MINITAB (Statistical Software) – Versão 13.1 – Minitab Inc., 2003.

MINSKY, M. L.; PAPERT, S. A. *Perceptrons: an introduction to computational geometry*. Cambridge, MA: MIT Press, 1969.

McCULLOCH, W. S.; PITTS, W. *A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics*, no. 5, p.115-133, 1943.

MONTGOMERY, D. C.; RUNGER, G. C. *Estatística Aplicada e Probabilidade para Engenheiros*. Rio de Janeiro: LTC Livros Técnicos Editora S. A., 2003, 496p.

MONTGOMERY, D. C. *Design and Analysis of Experiments*, John Wiley & Sons, 5^a ed., 2001, 699p.

NIST/SEMATECH, *e-Handbook of Statistical Methods*, <http://www.itl.nist.gov/div898/handbook/>, página acessada em 02 de Junho, 2006.

NOORDIN, M. Y.; VENKATESH, V. C.; SHARIF, S.; ELTING, S.; ABDULLAH, A. Application of response surface methodology in describing the performance of coated carbide tools when turning AISI 1045 steel. *Journal of Materials Processing Technology*, V145., 2004.

OLIVEIRA, A. J. *Otimização das condições de usinagem visando o torneamento a seco do aço ABNT 1045 em operação de desbaste*. 2003. 115p. Dissertação de Mestrado – Unicamp – Universidade Estadual de Campinas.

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*, V45, p.467-479, 2005.

ÖZEL, T.; NADGIR, A. Prediction of flank wear by using back-propagation neural network modeling when cutting hardened H-13 steel with chamfered and honed CBN tools. *International Journal of Machine Tools and Manufacture*, V42, p.287-297, 2002.

PAIVA, A. P. *Metodologia de Superfície de Resposta Multivariada – Uma proposta de otimização para processos de manufatura com múltiplas respostas correlacionadas*. 2006. 279p. Tese de Doutorado – UNIFEI – Universidade Federal de Itajubá.

PAIVA, A. P.; FERREIRA, J. R.; BALESTRASSI, P. P. Aplicação da metodologia de projeto de experimentos ao estudo dos parâmetros de usinagem de um aço SAE/ABNT 52100 utilizando ferramentas cerâmicas mistas. *COBEM – Congresso Brasileiro de Engenharia Mecânica*, Ouro Preto, 2005.

PAIVA, A. P. *Estudo da Minimização de Erro em Medições de Concentração de Emulsões por Titração Karl-Fischer utilizando-se Projeto de Experimentos*. 2004. 196p. Dissertação de Mestrado - UNIFEI – Universidade Federal de Itajubá.

PANDA, S.S.; SINGH, A.K.; CHAKRABORTY, D.; PAL, S.K. Drill wear monitoring using back-propagation neural network. *Journal of Materials Processing Technology*, In Press. Corrected Proof, Available online 20 December 2005.

PEREIRA, J. C. C. *Determinação de modelos de Vida de ferramenta e rugosidade no torneamento do aço ABNT 52100 Endurecido utilizando Metodologia de Superfície de Resposta (DOE)*. 2006. Dissertação de Mestrado - UNIFEI - Universidade Federal de Itajubá. 140p.

PETERSON, G. E.; ST. CLAIR, D. C.; AYLWARD, S. R.; BOND, W. E. Using Taguchi's method of experimental design to control errors in layered perceptrons. *IEEE Transactions on neural networks*, V6, no. 4. p.949-961, 1995.

PIÑEROS GARCET, J. D.; ORDOÑEZ, A.; ROOSEN, J.; VANCLOOSTER, M. Metamodelling: Theory, concepts and application to nitrate leaching modeling. *Ecological Modelling*. V193, no.3-4, p.629-644, 2006.

ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychology Review*, no. 65, p.386-408, 1958.

ROSS, P. *Taguchi Techniques for Quality Engineering*. New York: McGraw Hill, 1988.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. *Nature*, no. 323, p.533-536, 1986.

SANDVIK-Coromant, *Manual Técnico de Usinagem*, Sandviken, Suécia, 2005.

SANJAY, C.; NEEMA, M.L.; CHIN, C.W. Modelling of tool wear in drilling by statistical analysis and artificial neural network. *Journal of Materials Processing Technology*, V170, no. 3, p.494-500, 2005.

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*, V16, no. 4, July 2002, p.487-546, 2002.

SILVA, R. G.; REUBEN, R. L.; BAKER, K. J.; WILCOX, S. J. Tool wear monitoring of turning operations by neural network and expert system classification of a feature set generated from multiple sensor. *Mechanical Systems and Signal Processing*, 12(2), p.319-322, 1998.

SHAW, M. *Metal Cutting Principles*. New York: Oxford University Press, 2004, 651p.

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. *The International Journal of Advanced Manufacture Technology* (2006) 27:661-666.

SPECHT, D. F. A general regression neural network. *IEEE Transactions on Neural Networks*, V2, no. 6, p.568-576, 1991.

TAGUCHI, G.; ELSAYED, E. A.; HSIANG, T. *Engenharia da Qualidade em Sistemas de Produção*. São Paulo: McGraw-Hill, 1987, 235p.

TANSEL, I. N.; ARKAN, B.; BAO, W. Y.; MAHENDRAKAR, N.; SHISLER, B.; SMITH, D.; McCOOL, M. Tool wear estimation in micro-machining. Part II: neural-network-based periodic inspector for non-metals. *International Journal of Machine Tools and Manufacture*. V40, p.609-620, 2000.

TOSUN, N.; ÖZLER, L. A study of tool life in hot machining using artificial neural networks and regression analysis method. *Journal of Materials Processing Technology*. V124, no.1-2, p.99-104, 2002.

TSAI, K. M.; WANG, P. J. Predictions on surface finish in electrical discharge machining based upon neural network models. *International Journal of Machine Tools and Manufacture*. V41, no. 10, p.1385-1403, 2001.

TURK, R.; PERUS, I.; TERCELJ, M.; New starting points for the prediction of tool wear in hot forging. *International Journal of Machine Tools & Manufacture*, V44, p.1319-1331, 2004.

TRENT, E. M., WRIGHT, P. K., *Metal Cutting*. Woburn, MA: Butterworth Heinemann, 2000, 446p.

ZAWADA-TOMKIEWICZ, A. Classifying the wear of turning tools with neural networks. *Journal of Materials Processing Technology*. V109, no. 3, p.300-304, 2001.