

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA E  
INFORMÁTICA INDUSTRIAL

ANDRÉ EUGÊNIO LAZZARETTI

**CLASSIFICAÇÃO DE EVENTOS EM REDES DE DISTRIBUIÇÃO DE  
ENERGIA ELÉTRICA UTILIZANDO MODELOS NEURAIIS  
AUTÔNOMOS**

DISSERTAÇÃO DE MESTRADO

CURITIBA

2010

ANDRÉ EUGÊNIO LAZZARETTI

**CLASSIFICAÇÃO DE EVENTOS EM REDES DE DISTRIBUIÇÃO DE  
ENERGIA ELÉTRICA UTILIZANDO MODELOS NEURAIIS  
AUTÔNOMOS**

Dissertação apresentada ao Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial da Universidade Tecnológica Federal do Paraná como requisito parcial para obtenção do grau de “Mestre em Ciências” – Área de Concentração: Informática Industrial.

Orientador: Prof. Dr. Hugo Vieira Neto

Co-orientador: Dr. Rodrigo Jardim Riella

**CURITIBA**

**2010**

---

Dados Internacionais de Catalogação na Publicação

---

L432c Lazzaretti, André Eugênio  
Classificação de eventos em redes de distribuição de energia elétrica  
utilizando modelos neurais autônomos / André Eugênio Lazzaretti. – 2010.  
117 f. : il. ; 30 cm

Orientador: Hugo Vieira Neto

Co-orientador: Rodrigo Jardim Riella

Dissertação (Mestrado) – Universidade Tecnológica Federal do Paraná.  
Programa de Pós-Graduação em Engenharia Elétrica e Informática  
Industrial. Área de concentração: Informática industrial, Curitiba, 2010.

Bibliografia: p. 114-7

1. Redes elétricas. 2. Energia elétrica – Distribuição. 3. Fourier,  
Transformações de. 4. Wavelets (Matemática). 5. Redes neurais  
(Computação). 5. Engenharia elétrica – Dissertações. I. Vieira Neto, Hugo,  
orient. II. Riella, Rodrigo Jardim, co-orient. III. Universidade Tecnológica  
Federal do Paraná. Programa de Pós-Graduação em Engenharia Elétrica e  
Informática Industrial. IV. Título.

---

CDD (22. ed.) 621.3

Biblioteca Central da UTFPR

Título da Dissertação Nº 533:

**“Classificação de Eventos em Redes de Distribuição  
de Energia Elétrica Utilizando Modelos Neurais  
Autônomos”**

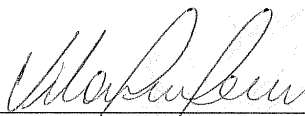
por

**André Eugênio Lazzaretti**

Esta dissertação foi apresentada como requisito parcial à obtenção do título de MESTRE EM CIÊNCIAS – Área de Concentração: Informática Industrial, pelo Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial – CPGEI – da Universidade Tecnológica Federal do Paraná – UTFPR – Campus Curitiba às 14h do dia 06 de julho de 2010. O trabalho foi aprovado pela Banca Examinadora, composta pelos professores:



Prof. Hugo Vieira Neto, Dr.  
(Presidente – UTFPR)



Prof. Vitor Hugo Ferreira, Dr.  
(UFF-RJ)



Prof. Leandro dos Santos Coelho, Dr.  
(PUC-PR)



Profª. Myriam Regattieri Delgado, Dr.  
(UTFPR)

Visto da coordenação:



Prof. Fábio Kurt Schneider, Dr.  
(Coordenador do CPGEI)

## **AGRADECIMENTOS**

Ao meu orientador Hugo Vieira Neto e ao meu co-orientador Rodrigo Jardim Riella pelo interesse e enorme apoio nessa linha de pesquisa.

Ao amigo Vitor Hugo Ferreira pelo apoio fundamental e pelas diversas discussões que impactaram diretamente na estrutura desse trabalho.

Aos vários colegas do Lactec que auxiliaram de forma significativa no desenvolvimento do trabalho, dentre os quais gostaria de citar Ivandro A. Bacca, Pedro Henrique M. dos Santos, Renata J. Bachega, Luiz Felipe R. B. Toledo, Marcelo A. Ravaglio, Fábio A. Guerra, Bruno Marchesi, Helon V. Ayala e Carlos Ribas.

Ao professor Leandro Coelho pelas discussões sobre parte da metodologia aqui apresentada.

Ao colega Julio Omeri da Copel Distribuição por nos fornecer os dados e apresentar os principais problemas da concessionária em relação ao método proposto.

Ao Lactec e a Copel Distribuição pelo incentivo e suporte na realização do trabalho.

Aos meus pais pelo apoio incondicional em todos os momentos da minha vida e à minha querida namorada Gabriela pela compreensão durante a realização deste trabalho.

## RESUMO

LAZZARETTI, A. E.. Classificação de Eventos em Redes de Distribuição de Energia Elétrica Utilizando Modelos Neurais Autônomos. 117 f. Dissertação – Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial, Universidade Tecnológica Federal do Paraná. Curitiba, 2010.

Este trabalho apresenta uma metodologia para classificação de eventos de curto-circuito e manobras em redes de distribuição de energia elétrica, com base nos registros oscilográficos de tensão na barra da subestação de distribuição. São apresentados os resultados obtidos para duas formas distintas de pré-processamento dos sinais de tensão, sendo a primeira baseada em Transformada de Fourier e a segunda em Transformada Wavelet para diferentes famílias de funções wavelet. Foram comparados três modelos neurais para o processo de classificação: *Multi-Layer Perceptron*, *Radial Basis Function* e *Support Vector Machine*. Os modelos foram treinados levando em conta uma característica autônoma de operação das redes, ou seja, a seleção automática do modelo e o controle de complexidade. Os resultados foram validados para um conjunto de simulações realizadas no programa *Alternative Transient Program*, visando a aplicação prática do método proposto em um equipamento registrador de oscilografias, desenvolvido pelo Lactec em conjunto com a Copel - Curitiba, PR, denominado *Power Quality Monitor*. Foram obtidos resultados com desempenho na ordem de 90% de acerto médio para as diferentes formas de pré-processamento e diferente modelos neurais.

**Palavras-chave:** Classificação de Eventos em Redes de Distribuição, Classificação de Falhas, Transformada de Fourier, Transformada Wavelet, Redes Neurais, Controle de Complexidade, Modelos Autônomos, Simulação de Redes de Distribuição.

## ABSTRACT

LAZZARETTI, A. E.. Event Classification in Power Distribution Networks using Autonomous Neural Models. 117 f. Dissertação – Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial, Universidade Tecnológica Federal do Paraná. Curitiba, 2010.

This work presents a method for automatic classification of faults and events related to quality of service in power distribution networks, based on oscillographies of the bar feeder voltages of the distribution substation. We present the results for two distinct pre-processing forms of the voltage signals. The first is based on the Fourier Transform and the second on the Wavelet Transform for different families of wavelet functions. We compared three neural models for the process of classification: Multi-Layer Perceptron, Radial Basis Function and Support Vector Machine. The models were trained taking into account the autonomous operation of networks, i.e. automatic model selection and control complexity. The results were validated for a set of simulations performed using the Alternative Transient Program, aimed at practical implementation of the proposed method in an oscillograph logger, developed by Lactec together with Copel, called the Power Quality Monitor. The results were obtained with performance on the order of 90% of average accuracy for the various pre-processing forms and neural models.

**Keywords:** Event Classification in Power Distribution Systems, Fault Classification, Fourier Transform, Wavelet Transform, Neural Networks, Complexity Control, Autonomous Models, Power Distribution Systems Simulation.

## LISTA DE FIGURAS

FIGURA 1 – SISTEMA ELÉTRICO DE POTÊNCIA .....	18
FIGURA 2 – CURTO-CIRCUITO MONOFÁSICO .....	23
FIGURA 3 – CURTO-CIRCUITO BIFÁSICO PARA TERRA .....	23
FIGURA 4 – CURTO-CIRCUITO BIFÁSICO .....	23
FIGURA 5 – CURTO-CIRCUITO TRIFÁSICO PARA TERRA .....	24
FIGURA 6 – CURTO-CIRCUITO TRIFÁSICO .....	24
FIGURA 7 – CURVA DO ELO FUSÍVEL .....	25
FIGURA 8 – CURVA DO RELIGADOR .....	26
FIGURA 9 – CHAVEAMENTO DE BANCO DE CAPACITORES .....	27
FIGURA 10 – ABERTURA DE UM RELIGADOR .....	28
FIGURA 11 – FECHAMENTO DE UM RELIGADOR .....	28
FIGURA 12 – METODOLOGIA .....	30
FIGURA 13 – ESPECTROS TEMPO-FREQUÊNCIA .....	42
FIGURA 14 – DECOMPOSIÇÃO DO SINAL ATRAVÉS DO BANCO DE FILTROS ..	44
FIGURA 15 – APROXIMAÇÕES E DETALHES .....	44
FIGURA 16 – DECOMPOSIÇÃO EM UM NÍVEL PARA 1024 AMOSTRAS .....	45
FIGURA 17 – RECOMPOSIÇÃO EM UM NÍVEL PARA 512 COEFICIENTES .....	45
FIGURA 18 – RECOMPOSIÇÃO EM UM NÍVEL PARA 512 COEFICIENTES .....	46
FIGURA 19 – ESTRUTURA TÍPICA DE UM NEURÔNIO .....	50
FIGURA 20 – ESTRUTURA E TREINAMENTO DE UMA MLP .....	59
FIGURA 21 – ARQUITETURA DE UMA RBF .....	62
FIGURA 22 – ESTRUTURA DE UMA REDE GWR .....	66
FIGURA 23 – RELAÇÃO DE DOMINÂNCIA ENTRE DUAS SOLUÇÕES .....	71
FIGURA 24 – SOLUÇÕES NA FRONTEIRA DE PARETO .....	71
FIGURA 25 – ESTRUTURA E TREINAMENTO DE UMA RBF .....	73
FIGURA 26 – HIPERPLANO ÓTIMO DE SEPARAÇÃO .....	74
FIGURA 27 – ERRO DE CLASSIFICAÇÃO (CASO 1) .....	75
FIGURA 28 – ERRO DE CLASSIFICAÇÃO (CASO 2) .....	75
FIGURA 29 – ESTRUTURA E TREINAMENTO DE UMA SVM .....	82
FIGURA 30 – DIAGRAMA ESQUEMÁTICO DO MODELO NO ATP DRAW .....	84
FIGURA 31 – ESQUEMÁTICO DO EQUIVALENTE ELÉTRICO .....	85
FIGURA 32 – ESQUEMÁTICO DO EQUIVALENTE ELÉTRICO NO ATP DRAW ...	85
FIGURA 33 – ESQUEMÁTICO DO TRANSFORMADOR DA SUBESTAÇÃO .....	86
FIGURA 34 – ESQUEMÁTICO DO TRANSFORMADOR DA SUBESTAÇÃO NO ATP DRAW .....	87
FIGURA 35 – ESQUEMÁTICO DO TRANSFORMADOR DE ATERRAMENTO NO ATP DRAW .....	87
FIGURA 36 – ESQUEMÁTICO DE UM BANCO DE CAPACITORES NO ATP DRAW	88
FIGURA 37 – ESQUEMÁTICO DE UM ALIMENTADOR NO ATP DRAW .....	89
FIGURA 38 – MODELO $\pi$ EQUIVALENTE .....	90
FIGURA 39 – ESQUEMÁTICO DA CARGA NO <i>ATP DRAW</i> .....	90
FIGURA 40 – LOCAL DE OCORRÊNCIA DOS EVENTOS .....	93
FIGURA 41 – POSSÍVEIS SOLUÇÕES PARA OS MODELOS DA REDE RBF .....	99
FIGURA 42 – CONVERGÊNCIA DO AG DURANTE O TREINAMENTO DA REDE SVM .....	105



## LISTA DE TABELAS

TABELA 1	– CATEGORIAS DOS FENÔMENOS ELETROMAGNÉTICOS .....	19
TABELA 2	– CURTOS-CIRCUITOS .....	22
TABELA 3	– REFERÊNCIAS .....	31
TABELA 4	– BANDAS DE FREQUÊNCIA DA DECOMPOSIÇÃO WAVELET ....	48
TABELA 5	– PARÂMETROS DE CABOS .....	90
TABELA 6	– PARÂMETROS DOS ALIMENTADORES .....	91
TABELA 7	– PARÂMETROS VARIADOS PARA O CONJUNTO 1 .....	92
TABELA 8	– PARÂMETROS VARIADOS PARA O CONJUNTO 2 .....	92
TABELA 9	– NÚMERO DE INSTÂNCIAS POR CLASSE .....	95
TABELA 10	– ACERTOS PARA A MLP .....	96
TABELA 11	– ACERTOS PARA A RBF .....	98
TABELA 12	– ACERTOS PARA A SVM .....	103

## LISTA DE SIGLAS

DEC	Duração Equivalente de Interrupção por Conjunto de Consumidores
FEC	Frequência Equivalente de Interrupção por Conjunto de Consumidores
DIC	Duração de Interrupção por Unidade Consumidor
FIC	Frequência de Interrupção por Unidade Consumidora
ONS	Operador Nacional do Sistema
PQM	<i>Power Quality Meter</i>
SEP	Sistema Elétrico de Potência
LT	Linhas de Transmissão
QEE	Qualidade de Energia Elétrica
TF	Transformada de Fourier
TW	Transformada Wavelet
RNA	Redes Neurais Artificiais
THS	Transformada-S Hiperbólica
EW	Entropia Wavelet
TWP	Transformada Wavelet Packet
TRF	Transformada Rápida de Fourier
TFJ	Transformada de Fourier Janelada
TWD	Transformada Wavelet Discreta
FIR	<i>Finite Impulse Response</i>
MLP	<i>Multi-Layer Perceptron</i>
RBF	<i>Radial Basis Function</i>
SVM	<i>Support Vector Machine</i>
DAR	Determinação Automática de Relevância
GWR	<i>Grow When Required</i>

AGMO	Algoritmos Genéticos Multi-Objetivo
AG	Algoritmo Genético
ATP	<i>Alternative Transient Program</i>
TAT	Transformador de Aterramento
BC	Bancos de Capacitores

## SUMÁRIO

<b>1 INTRODUÇÃO</b>	<b>12</b>
1.1 MOTIVAÇÃO	12
1.2 OBJETIVOS	16
1.2.1 Objetivo Geral	16
1.2.2 Objetivos Específicos	16
1.2.3 Estrutura da Dissertação	17
<b>2 EVENTOS EM SISTEMAS DE DISTRIBUIÇÃO DE ENERGIA</b>	<b>18</b>
2.1 EVENTOS DE QEE	18
2.1.1 Transitórios	20
Transitórios Impulsivos	20
Transitórios Oscilatórios	20
2.1.2 Variações de tensão de curta duração	20
Afundamento	20
Elevação	21
Interrupção	21
2.2 EVENTOS SELECIONADOS	21
2.2.1 Curto-circuito no Sistema Elétrico de Distribuição	21
2.2.2 Transitórios Oscilatórios	26
Chaveamento de Banco de Capacitores	26
Abertura da Rede de Média Tensão	27
Fechamento da Rede de Média Tensão	28
<b>3 REVISÃO DA LITERATURA SOBRE CLASSIFICAÇÃO DE EVENTOS</b>	<b>29</b>
<b>4 CONCEITOS E MÉTODOS PARA CLASSIFICAÇÃO DE EVENTOS</b>	<b>35</b>
4.1 PRÉ-PROCESSAMENTO	35
4.1.1 Análise de Fourier	35
Série de Fourier Contínua	35
Série de Fourier Discreta	36
Transformada Contínua de Fourier	37
Transformada Discreta de Fourier	37
Transformada Rápida de Fourier	38
Transformada de Fourier Janelada	38
4.1.2 Transformada Wavelet	39
Transformada Wavelet Contínua	40
Propriedades da Transformada Wavelet	40
Transformada Wavelet Discreta	41
Função de Escalonamento	43
Transformada Wavelet Discreta - Algoritmo Rápido	43
Funções Wavelet	46
Extração de Características	47
4.2 PROCESSAMENTO E REDES NEURAIS ARTIFICIAIS	49
4.2.1 Redes <i>Perceptron</i> Multi-Camadas com Inferência Bayesiana	53

Treinamento Bayesiano de MLPs .....	54
Determinação Automática de Relevância .....	57
Seleção Bayesiana de Modelos .....	58
Resumo do Treinamento de MLPs com Inferência Bayesiana .....	59
4.2.2 Redes Neurais com Função de Base Radial .....	59
Centros Fixos Seleccionados ao Acaso .....	62
Seleção Auto-Organizada dos Centros .....	63
Seleção Supervisionada dos Centros .....	63
Método de Aprendizagem Proposto para RBFs .....	64
Seleção de Entradas .....	65
Rede Neural <i>Grow When Required</i> .....	66
Algoritmos Genéticos .....	68
Algoritmos Genéticos Multi-Objetivo .....	69
Resumo do Treinamento de RBFs .....	72
4.2.3 <i>Support Vector Machines</i> .....	73
Aplicação em Padrões não Linearmente Separáveis .....	75
Seleção Automática dos Hiperparâmetros .....	79
Resumo do Treinamento de SVMs .....	81
<b>5 MODELAGEM DO PROBLEMA EM AMBIENTE DE SIMULAÇÃO .....</b>	<b>83</b>
5.1 EQUIVALENTE DA CONCESSIONÁRIA .....	84
5.2 TRANSFORMADOR DA SUBESTAÇÃO .....	85
5.3 TRANSFORMADOR DE ATERRAMENTO .....	87
5.4 BANCOS DE CAPACITORES .....	88
5.5 ALIMENTADORES E CABOS .....	89
5.6 CARGAS .....	90
5.7 GERAÇÃO AUTOMÁTICA DE EVENTOS .....	91
<b>6 RESULTADOS E DISCUSSÕES .....</b>	<b>94</b>
6.1 RESULTADOS PARA AS REDES MLP .....	94
6.1.1 Acerto Total do Classificador .....	94
6.1.2 Capacidade de Generalização .....	96
6.1.3 Estrutura Definida .....	97
6.2 RESULTADOS PARA AS REDES RBF .....	97
6.2.1 Acerto Total do Classificador .....	97
6.2.2 Capacidade de Generalização .....	98
6.2.3 Estrutura Definida .....	100
6.3 RESULTADOS PARA AS REDES SVM .....	102
6.3.1 Acerto Total do Classificador .....	102
6.3.2 Capacidade de Generalização .....	103
6.3.3 Estrutura Definida .....	104
6.4 COMPARAÇÃO GERAL .....	106
6.4.1 Acerto Total dos Classificadores .....	106
6.4.2 Capacidade de Generalização .....	107
6.4.3 Estrutura Definida .....	107
<b>7 CONCLUSÕES E TRABALHOS FUTUROS .....</b>	<b>110</b>
7.1 CONTRIBUIÇÕES DO TRABALHO .....	112
7.2 PUBLICAÇÕES REALIZADAS .....	113
<b>REFERÊNCIAS .....</b>	<b>114</b>

# 1 INTRODUÇÃO

## 1.1 MOTIVAÇÃO

A Qualidade de Energia Elétrica (QEE), segundo Dugan et al. (2002), é qualquer desvio associado com a tensão, corrente ou frequência nominais que resulte em danos ou falhas em equipamentos de consumidores de energia elétrica. O termo QEE se tornou um dos termos mais citados, tanto por consumidores (normalmente industriais) quanto por concessionárias de energia elétrica para tratar de uma série de problemas que envolvem o fornecimento adequado da energia elétrica. Apesar desses problemas não serem novos no setor elétrico, se tornaram tema de discussão recente principalmente pela abordagem corrente, que trata esses problemas de forma conjunta, ao invés de estudá-los separadamente. Além disso, alguns fatores contribuíram de forma significativa nesse estudo (DUGAN et al., 2002):

- Surgimento de cargas com comportamento mais sensível a eventos de QEE, as quais utilizam microprocessadores e equipamentos baseados em eletrônica de potência;
- A utilização de equipamentos para redução das perdas e controle de fator de potência, que aumentam a injeção de conteúdo harmônico do sistema;
- O aumento no interesse dos consumidores finais em assuntos que envolvem QEE, buscando equipamentos que reduza o impacto na QEE do sistema;
- Maior integração dos processos industriais, fazendo com que a falha em determinado ponto de uma planta industrial leve a falhas em diversos outros locais dessa planta.

Cabe salientar aqui que o termo QEE é o mais encontrado na definição dos eventos que são tratados nesse trabalho. Porém, a QEE está relacionada de forma direta com a qualidade de tensão entregue a um consumidor, uma vez que a corrente elétrica é função da carga, sob a qual não se tem controle das características. Essa corrente circulando no sistema causa distúrbios na tensão, como por exemplo, uma carga com elevado conteúdo harmônico de corrente, que provoca distorção na forma de onda de tensão quando circula pela impedância do sistema.

Outro fator fortemente relacionado com os problemas de QEE é o impacto econômico resultante dos eventos. Os equipamentos com falhas em função do mau fornecimento de energia elétrica podem resultar em grandes prejuízos financeiros ao empresário ou à concessionária fornecedora. O prejuízo ainda estende-se à parada do processo industrial, que em muitos casos possui uma complicada e demorada restauração, fator determinante em um meio industrial competitivo, onde qualquer minuto de paralisação na produção pode acarretar perdas financeiras de extrema importância. Cabe então, a utilização de novos indicadores econômicos (MELO; CAVALCANTI, 2003) que juntamente aos indicadores de grandezas elétricas, possibilitam uma melhor quantificação da qualidade de energia. Tais índices, denominados Índices dos Custos Associados, determinam os prejuízos industriais devido à qualidade de energia. São eles:

- Custos referentes à manutenção ou à compra de novos equipamentos em substituição aos danificados pela má qualidade de energia;
- Custo associado à perda de material que estava em produção durante o problema energético;
- Custo associado à perda de um material já produzido durante o processo de armazenamento;
- Custo referente à perda de matérias primas;
- Custo financeiro devido à perda de vendas durante uma interrupção;
- Custo da proteção de equipamentos contra efeitos da má qualidade de energia;
- Custo da geração própria de energia através de geradores de emergência e baterias;
- Custo referente às perdas de informações armazenadas em computadores;
- Custo referente à reposição de horas devido a uma parada de produção;
- Custo referente aos gastos utilizados para se retomar o ritmo normal da produção no caso de ocorrência de uma interrupção;

Demonstra-se assim, que o aspecto financeiro é o de maior importância nesse setor de qualidade de energia. Com isso, explica-se a razão de há pouco tempo atrás, a preocupação dos consumidores era restrita apenas à sua demanda, fator de potência e a assinatura de contratos junto à concessionária fornecedora. Hoje, promove-se uma grande evolução nos sistemas de gerenciamento de energia elétrica, os quais acrescentam a preocupação com tempo de fornecimento e de interrupção, número de interrupções, níveis de tensão, frequência, formas de onda

de tensão e corrente, efeitos transitórios e componentes harmônicos. Entretanto, esses novos índices ainda não possuem legislação nacional para o setor de distribuição de energia elétrica. Atualmente, apenas a conformidade de tensão e a continuidade de serviço são regulamentadas pela Agência Nacional de Energia Elétrica (ANEEL):

- Conformidade da Tensão: Regulamentada pela resolução número 505 (ANEEL, 2001) cujas disposições relativas à conformidade dos níveis de tensão de energia elétrica em regime permanente já se encontram de forma consolidadas e claras.
- Continuidade do Serviço: Regulamentada pela resolução número 024 (ANEEL, 2000) que estabelece as disposições relativas à continuidade da distribuição de energia elétrica às unidades consumidoras através dos indicadores Duração Equivalente de Interrupção por conjunto de Consumidores, DEC, e Frequência Equivalente de Interrupção por conjunto de Consumidores, FEC. Eles representam o tempo, em média, que cada consumidor de um conjunto ficou sem o fornecimento de energia elétrica e o número de interrupções, em média, que um consumidor de um conjunto teve durante um intervalo de tempo, respectivamente. A ANEEL define as metas DEC e FEC a cada quatro anos (nos anos da revisão tarifária) para os conjuntos de consumidores de cada concessionária de distribuição. Vale ressaltar que a partir do ano de 2010, a ANEEL utilizará os índices DIC, Duração de Interrupção por Unidade Consumidor, e o FIC, Frequência de Interrupção por Unidade Consumidora, para aplicação de multas às concessionárias que ultrapassarem os limites estabelecidos, os quais atuam de forma individual na avaliação da interrupção de energia dos consumidores.

Para outros critérios de qualidade de energia o Operador Nacional do Sistema, ONS, já publicou em Procedimentos de Rede (ONS, 2000a, 2000b, 2000c), válidos apenas para a rede básica do sistema, as definições de nomenclaturas, indicadores e padrões dos índices: Variação de Tensão de Curta Duração, Variação Momentânea de Tensão, Variação Temporária de Tensão, Interrupção Momentânea de Tensão, Afundamento Momentâneo de Tensão, Elevação Momentânea de Tensão, Interrupção Temporária de Tensão, Afundamento Temporária de Tensão, Elevação Temporária de Tensão, Harmônicos e Flutuação de Tensão. Tais resoluções e procedimentos mostram a atual regulamentação referente à qualidade de energia definida e fiscalizada pelos órgãos competentes no cenário nacional. No Brasil, ainda se encontram em desenvolvimento tais procedimentos, porém em consenso às normas internacionais. Um procedimento que merece destaque aqui é o módulo 8 dos Procedimentos de Distribuição da ANEEL (ANEEL, 2008). Esse procedimento descreve os principais limites referentes à QEE do serviço e do



produto em redes de distribuição (concessionárias e consumidores), tais como: harmônicos, flutuações de tensão, variações de tensão de curta duração, tensão em regime permanente, variações de frequência, entre outros. Vale salientar que esse procedimento ainda não se encontra em vigor, já que ainda passa por um processo de avaliação e aperfeiçoamento.

Frente à necessidade do monitoramento desses eventos, a Companhia Paranaense de Energia (Copel) desenvolveu em conjunto com o Instituto de Tecnologia para o Desenvolvimento (Lactec) localizados em Curitiba, Paraná, um sistema voltado para monitoramento contínuo da QEE para barras de subestações de distribuição (RIELLA et al., 2008). Esse sistema é composto por um *software* de supervisão e visualização remota e um equipamento de monitoramento contínuo, denominado Power Quality Meter, PQM, conectado diretamente na saída dos transformadores de potencial das barras das subestações. O equipamento efetua medições periódicas dos valores de tensão, frequência, distorção harmônica total de tensão, harmônicos individuais até o 50º harmônico e desequilíbrio de tensão, ciclo a ciclo, com resolução de 128 amostras por ciclo (frequência de amostragem de 7680 Hz). Durante os cálculos, o equipamento também implementa funções de geração de oscilografia, com disparo através de níveis mínimos e máximos, configuráveis para cada parâmetro de medição, com resolução de disparo de até um ciclo.

A função responsável por gerar oscilografias é uma ferramenta de extrema importância para pós-análise, pois fornece um embasamento em termos físicos dos eventos que geraram as oscilografias, auxiliando a mitigação de defeitos. A dificuldade encontrada nessa metodologia é o grande volume de dados gerados pelos monitores de QEE, dificultando a análise dos registros por parte dos usuários desse banco de dados (p.ex. engenheiros de proteção e manutenção). A grande quantidade de dados impede que seja feita uma identificação para cada oscilografia registrada, resultando em uma falta de classificação adequada para os vários eventos, na maioria das oscilografias armazenadas. Essa falta de informação, por sua vez, acaba dificultando o desenvolvimento de um classificador automático do conjunto de oscilografias, uma vez que não há uma forma de se desenvolver um algoritmo de aprendizagem supervisionada com base em um conjunto de dados sem classificação prévia. Adicionalmente, deve-se levar em conta a dificuldade de se identificar e diferenciar a causa de um evento com base na sua oscilografia de tensão, como p. ex., o caso de eventos de religamento de alimentadores e curtos-circuitos trifásicos, os quais apresentam formas de onda bastante semelhantes, sendo por vezes classificados incorretamente.

Dentre os diversos eventos relacionados com Qualidade de Energia Elétrica que ocorrem em redes de distribuição, a classe de maior impacto são os curtos-circuitos (faltas) no sis-

tema elétrico (MORAIS et al., 2007), sendo que cerca de 70% dessas, ocorrem no sistema de distribuição (DUGAN et al., 2002). Além disso, os eventos típicos de Qualidade de Energia Elétrica, eventos transitórios como chaveamento de banco de capacitores e eventos de manobra, ainda representam um forte impacto no que diz respeito ao ressarcimento de danos a consumidores, ou seja, estão fortemente relacionados com a queima de equipamentos.

Com base nessa realidade, um sistema de classificação automática de oscilografias forneceria uma base para distinção correta de eventos, auxiliando no processo de mitigação, manutenção e localização dos eventos, formando um sistema de auxílio à tomada de decisões relacionadas ao sistema de distribuição, por parte da concessionária.

## 1.2 OBJETIVOS

### 1.2.1 Objetivo Geral

Efetuar a classificação de eventos transitórios e curtos-circuitos em redes de distribuição de energia elétrica utilizando dados simulados de tensão trifásica na saída de subestação modelada para simulação digital, utilizando diferentes formas de pré-processamento do sinal sob análise (tensão trifásica) e classificadores baseados em modelos neurais autônomos, visando a aplicação em um sistema real.

### 1.2.2 Objetivos Específicos

Os objetivos específicos são:

- Validar os modelos dos principais elementos de uma subestação de distribuição de energia elétrica, com o intuito de utilizá-los para construção de um ambiente de simulação dos eventos mais relevantes que ocorrem em redes de distribuição, fornecendo uma base para o desenvolvimento de um sistema para geração automática de eventos simulados.
- Comparar as diferentes formas de pré-processamento das formas de onda de tensão trifásica (com base em Transformada de Fourier e Transformada Wavelet) de forma a classificar eventos transitórios e curtos-circuitos com um único classificador.
- Avaliar o desempenho de diferentes modelos neurais, do tipo *Multi-Layer Perceptron*, *Radial Basis Function* e *Support Vector Machine*, com características autônomas (principalmente com relação ao controle de complexidade e seleção de entradas), cujas funcionalidades ainda não foram avaliadas para os eventos em questão.

- Propor um método que possa ser aplicado em um sistema real, tendo em vista as diversas características apresentadas ao longo do trabalho, de forma que o resultado do processo de classificação possa servir de base para implementação de outras técnicas igualmente úteis para uma concessionária de distribuição de energia elétrica, como p. ex., a localização de curtos-circuitos.

### 1.2.3 Estrutura da Dissertação

O trabalho inicia com a fundamentação do problema abordado (Capítulos 1 e 2), passando pela definição dos principais conceitos de QEE, fundamentando a escolha dos eventos a serem classificados. Após essa fundamentação, encontra-se também uma revisão da literatura (Capítulo 3), seguida dos principais conceitos relacionados com o pré-processamento e classificação dos sinais sob análise (Capítulo 4). Em seguida, é feita a apresentação da modelagem necessária para simulação dos eventos escolhidos (Capítulo 5), detalhando cada componente da subestação de distribuição modelada em ambiente de simulação. O capítulo seguinte (6) fornece a análise dos resultados e as discussões associadas com a metodologia escolhida. Finalmente, é apresentada a conclusão (Capítulo 7), apontando eventuais trabalhos futuros do presente tema.

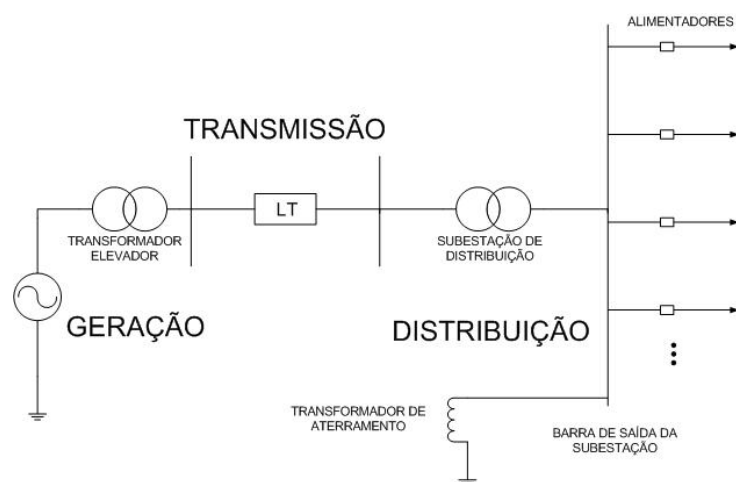
## 2 EVENTOS EM SISTEMAS DE DISTRIBUIÇÃO DE ENERGIA

Nesse capítulo pretende-se apresentar os principais eventos em redes de distribuição, visando a modelagem dos componentes principais para a simulação digital desses eventos.

### 2.1 EVENTOS DE QEE

Em função do termo QEE ser aplicado em uma grande quantidade de fenômenos eletromagnéticos no Sistema Elétrico de Potência, SEP, cabe nesse ponto expor os principais termos associados aos eventos de QEE, sendo que esses termos formam a base para os eventos estudados nesse trabalho.

A tabela 1 apresenta as principais categorias dos fenômenos eletromagnéticos que ocorrem ao longo do SEP (figura 1), composto por geração, linhas de transmissão, LT, e distribuição, bem como suas principais características (duração, magnitude e conteúdo de frequência típico) (DUGAN et al., 2002).



**Figura 1: Sistema Elétrico de Potência Resumido no caso da Copel.**

A divisão dos eventos conforme suas características é importante pois auxilia a análise dos resultados de medições, descrevendo de forma mais adequada a ocorrência de eventos de QEE,

Tabela 1: Categorias dos fenômenos eletromagnéticos.

<b>Categoria</b>	<b>Conteúdo Espectral Típico</b>	<b>Duração Típica</b>	<b>Amplitude de Tensão Típica</b>
<b>Transitórios</b>			
Impulsivos			
Nanosegundo	5 ns	< 50 ns	
Microsegundo	1 us	50 ns — 1 ms	
Milisegundo	0,1 ms	> 1 ms	
Oscilatórios			
Baixa Frequência	<5 kHz	3 — 50 ms	0,4 p.u.
Média Frequência	5 — 500 kHz	20 us	0,4 p.u.
Alta Frequência	0,5 — 5 MHz	5 us	0,4 p.u.
<b>Variações de Tensão de Curta Duração</b>			
Instantânea			
Afundamento de Tensão		0,5 — 30 ciclos	0,1 — 0,9 p.u.
Elevação de Tensão		0,5 — 30 ciclos	1,1 — 1,8 p.u.
Momentânea			
Interrupção		0,5 ciclos — 3 s	< 0,1 p.u.
Afundamento de Tensão		30 ciclos — 3 s	0,1 — 0,9 p.u.
Elevação de Tensão		30 ciclos — 3 s	1,1 — 1,4 p.u.
Temporária			
Interrupção		3 s — 1 minuto	< 0,1 pu
Afundamento de Tensão		3 s — 1 minuto	0,1 — 0,9 p.u.
Elevação de Tensão		3 s — 1 minuto	1,1 — 1,2 p.u.
<b>Variações de Tensão de Longa Duração</b>			
Subtensão Sustentada		> 1 minuto	0,8 — 0,9 p.u.
Sobretensão Sustentada		> 1 minuto	1,1 — 1,2 p.u.
Interrupção Sustentada		> 1 minuto	0,0 p.u.
<b>Distorção da Forma de Onda</b>			
Harmônicos	0 — 100°	regime permanente	0 — 20 %
Inter-harmônicos	0 — 6 kHz	regime permanente	0 — 2 %
Ruído	faixa ampla	regime permanente	0 — 1 %
Nível CC		regime permanente	0 — 0,1 %
<b>Flutuação de Tensão</b>	< 25Hz	intermitente	0,1 — 7 %
<b>Desequilíbrio de Tensão</b>		regime permanente	0,5 — 2 %
<b>Variação de Frequência</b>		< 10 s	

Fonte: (DUGAN et al., 2002)

além de formar uma base de referência para o processo de classificação desses eventos no SEP. A seguir, são descritos os tópicos mais relevantes da tabela 1, no escopo desse trabalho.

### 2.1.1 Transitórios

São fenômenos que possuem pouca duração no tempo, porém importantes para a avaliação da qualidade de energia no sistema. Dentro de uma grande quantidade de efeitos causadores de transitórios, pode-se destacar: chaveamento de banco de capacitores, curtos-circuitos e descargas atmosféricas. Os fenômenos transitórios podem ser divididos em transitórios impulsivos e transitórios oscilatórios.

#### Transitórios Impulsivos

São efeitos normalmente causados por descargas atmosféricas que ao atingir sistemas elétricos apresentam impulsos unidirecionais em polaridade (positivos ou negativos) de elevada magnitude e frequência. Esses eventos podem elevar o potencial de terra em determinadas regiões, prejudicando equipamentos sensíveis a esse tipo de fenômeno (DELMONT-FILHO, 2003).

#### Transitórios Oscilatórios

São transitórios causados pelo chaveamento de circuitos indutivos ou capacitivos, como manobra de linhas de transmissão ou distribuição, corte de carga indutiva, inserção de bancos de capacitores e a energização de transformadores. O resultado é um efeito na tensão e na corrente com magnitude e frequência de oscilação distinta do padrão do sistema elétrico (DUGAN et al., 2002).

### 2.1.2 Variações de tensão de curta duração

São variações de tensão com o tempo estabelecido entre 0,5 ciclo e 1 minuto. A faixa de tempo pode ser dividida em instantânea, momentânea e temporária. Existem três tipos definidos de variação de tensão: elevação de tensão, afundamento de tensão e interrupção de tensão, todos distintos entre si através da magnitude da tensão durante o efeito verificado (DUGAN et al., 2002).

#### Afundamento

É o efeito caracterizado pela diminuição do valor eficaz da tensão em um sistema, deixando um valor residual entre 0,1 e 0,9 p.u. Geralmente é provocado por um curto-circuito ou pelo funcionamento de cargas específicas que exigem grandes correntes de partida, tais como as

presentes em motores de grande porte. Os equipamentos eletrônicos como computadores, inversores de frequência e de controladores lógicos voltados para automação de processo, podem não suportar afundamentos de tensão e serem desligados (DUGAN et al., 2002).

### Elevação

Um curto-circuito fase-terra pode causar elevação de tensão em suas fases sãs, devido ao desbalanceamento do sistema no ponto de falta, aumentando o valor da magnitude da tensão entre 1,1 a 1,8 p.u. Outro fator responsável pela elevação de tensão é o desligamento de grandes cargas (DUGAN et al., 2002).

### Interrupção

O efeito denominado interrupção é descrito como um decréscimo da tensão a um valor inferior a 0,1 p.u. da tensão nominal, durante um tempo não superior a 1 minuto. É causada por faltas severas no sistema, falhas na rede e problemas relacionados ao controle dos sistemas elétricos (DUGAN et al., 2002).

## 2.2 EVENTOS SELECIONADOS

Tendo em vista a base conceitual exposta na seção anterior, juntamente com a atual regulamentação do setor elétrico, foram selecionados nesse trabalho os principais eventos para a concessionária em termos de QEE, com o intuito de reproduzi-los em ambiente de simulação, gerando-se assim um banco de dados de eventos devidamente classificados. Esses eventos correspondem a curtos-circuitos monofásicos, bifásicos, bifásicos para a terra, trifásicos e trifásicos para a terra, chaveamento de bancos de capacitores e manobras nos religadores da subestação de distribuição (abertura e fechamento). A seguir esses eventos são descritos em detalhes.

### 2.2.1 Curto-circuito no Sistema Elétrico de Distribuição

Muito embora os sistemas de distribuição sejam projetados de forma a se obter uma instalação de boa qualidade, associada à utilização de materiais adequados aos diversos locais a que essa rede atende, o sistema como um todo estará sujeito a eventos de curto-circuito de natureza aleatória, desde a geração até a distribuição (KINDERMANN, 2007). Dentre as diversas causas desses eventos pode-se citar:

- Problemas de Isolação - Associados com a qualidade, projeto ou até mesmo o envelhecimento dos materiais utilizados na isolação dos condutores do sistema.
- Problemas Mecânicos - Resultantes da ação da natureza, como por exemplo, vento, árvores próximas aos condutores e acúmulo de neve.
- Problemas Elétricos - Provenientes de surtos gerados por manobras na rede ou descargas atmosféricas diretas ou indiretas.
- Problemas de Natureza Térmica - O aquecimento nos componentes do SEP pode provocar sobrecargas que, quando prolongadas, levam à redução da vida útil desses componentes.
- Problemas de Manutenção - Provenientes da falta de treinamento, inspeção, material ou controle de qualidade das equipes de manutenção.
- Atos de Vandalismo.

Em um SEP típico (Figura 1), Dugan et al. (2002) mostram que 31% das faltas que causam falhas de equipamentos em um consumidor industrial, ocorrem no sistema de transmissão, sendo o restante nas redes de distribuição. Desses 69% restantes, que também provocam falhas em equipamentos, 23% ocorrem no próprio alimentador onde os equipamentos estão instalados, enquanto que 46% ocorrem em alimentadores paralelos ao alimentador dos equipamentos danificados. Isso demonstra a importância de se avaliar os eventos de curto-circuito em redes de distribuição (um dos focos do presente trabalho), uma vez que esses estão associados de forma significativa com danos provocados a consumidores.

Um curto-circuito é definido pelo número de fases envolvidas e a presença do terra durante a falta. Ou seja, um curto-circuito pode ser monofásico, bifásico ou trifásico sendo que o único curto-circuito que envolve obrigatoriamente o terra é o curto-circuito monofásico. A tabela 2 mostra a ocorrência percentual desses eventos.

**Tabela 2: Curtos-Circuitos.**

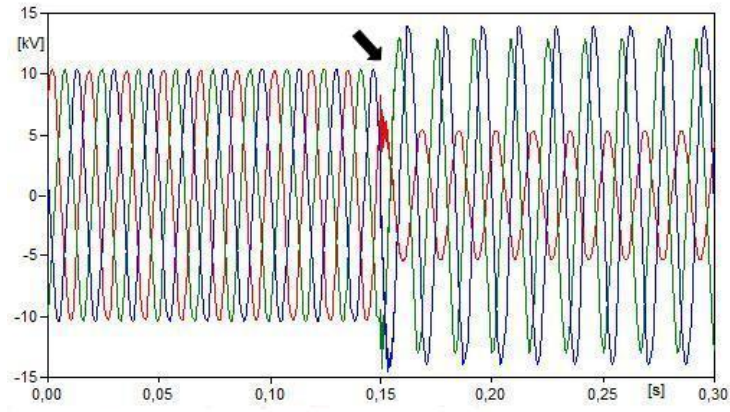
Tipos de Curto Circuito	Ocorrência (%)
Trifásico	6
Bifásico	15
Bifásico para a Terra	16
Monofásico para a Terra	63

**Fonte: (KINDERMANN, 2007)**

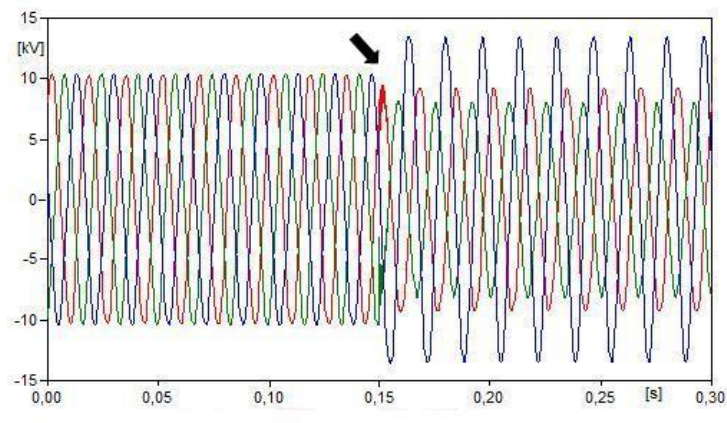
Nas figuras 2, 3, 4, 5 e 6, estão representadas as formas de onda de tensão na saída da subestação de um sistema de distribuição 13,8 kV para um curto-circuito monofásico, bifásico



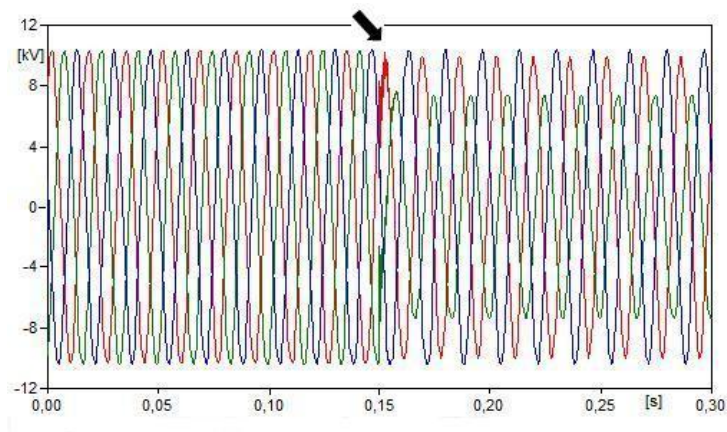
para terra, bifásico, trifásico para terra e trifásico, respectivamente, sendo que as setas indicam o instante de ocorrência da falta. Essas formas de onda foram obtidas através da simulação digital considerando a resistência de falta de 20 ohms.



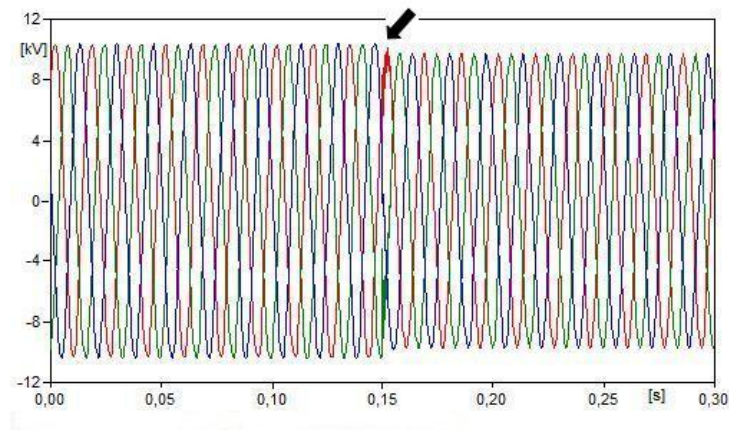
**Figura 2: Curto-Circuito Monofásico em um Alimentador.**



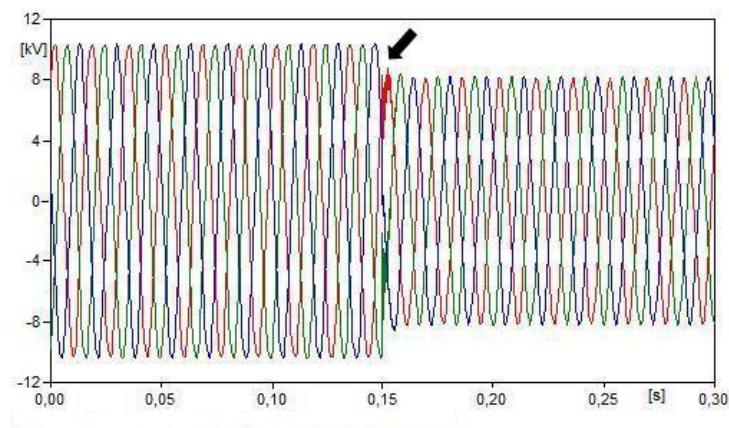
**Figura 3: Curto-Circuito Bifásico para a Terra em um Alimentador.**



**Figura 4: Curto-Circuito Bifásico em um Alimentador.**



**Figura 5: Curto-Circuito Trifásico para a Terra em um Alimentador.**

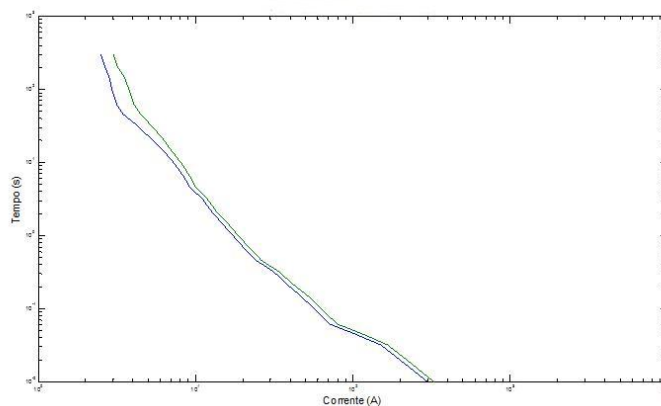


**Figura 6: Curto-Circuito Trifásico em um Alimentador.**

Nesses eventos de curto-circuito são observados diversos fenômenos de qualidade de energia elétrica, como afundamentos e elevações de tensão e transitórios oscilatórios. Esses fenômenos caracterizam um curto-circuito, sendo que essas características serão de fundamental importância no decorrer do processo de classificação desses eventos. Com relação à duração, um curto-circuito pode ser classificado como permanente ou temporário. O curto-circuito permanente está associado com eventos irreversíveis mesmo após a atuação da proteção, enquanto que no curto-circuito temporário o sistema pode ser restabelecido após a atuação da proteção. Um curto-circuito temporário está normalmente associado a uma sobretensão oscilatória ou impulsiva que provoca o rompimento da isolamento (através do ar ionizado) em isoladores da rede de distribuição (normalmente sob condição não ideal de operação), sendo que essa condição é extinta quando ocorre a atuação da proteção (KINDERMANN, 2007). Além disso, é comum o contato dos condutores da rede de média tensão com galhos de árvores, pássaros ou até mesmo condições anormais de umidade e vento, que levam a um curto-circuito temporário.

A porcentagem de ocorrência de um curto-circuito temporário é aproximadamente 96% (4% apenas para o curto-circuito permanente) (KINDERMANN, 2007), fazendo com que o restabelecimento do sistema seja de fundamental importância, após a ocorrência desses eventos. Para tanto, as subestações de distribuição são equipadas com religadores automáticos instalados nas saídas dos alimentadores da rede de distribuição, capazes de efetuar o restabelecimento do sistema, após um curto-circuito temporário. Além do religador, ao longo do alimentador são instalados elos-fusíveis que possuem a capacidade de extinguir a falta, porém sem permitir o religamento automático.

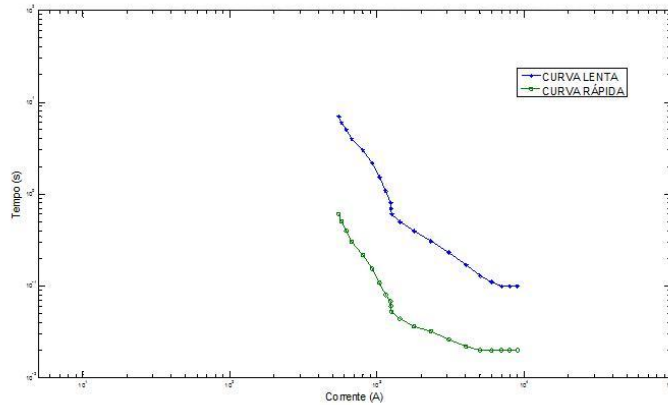
O elo fusível é um dos elementos mais encontrados em alimentadores de redes de distribuição, uma vez que ele é normalmente instalado nos ramais e transformadores desses alimentadores. A principal função do elo fusível é operar em faltas permanentes e isolar a seção com a falta do restante do circuito. Um elo fusível, por ser normalmente constituído de prata ou estanho, possui uma característica de atuação (rompimento em função da temperatura) que varia de acordo com o tempo em que uma dada corrente circula através dos seus terminais. Essa característica responde a uma curva denominada tempo-corrente inversa, sendo que nessa curva o tempo de atuação decai, à medida que a corrente aumenta. A figura 7 mostra uma curva tempo-corrente inversa típica para um elo fusível do tipo 1H (elo fusível com corrente nominal de 1A). A atuação do elo fusível normalmente se dá através de duas curvas, uma que representa a atuação mínima (curva azul na figura 7) e outra o tempo de atuação máximo (curva verde na figura 7).



**Figura 7: Curva de um Elo Fusível do tipo 1H.**

Já o religador da subestação possui duas curvas de atuação tempo-corrente inversa. Uma delas, denominada de curva lenta, é a curva de atuação para regime permanente, ou seja, quando ocorre um curto-circuito, o religador atua de acordo com a curva lenta. Uma vez extinto o curto-circuito (abertura do religador), ocorre o religamento e o religador passa a operar em outra curva

tempo-corrente inversa, denominada curva rápida. Com isso, garante-se uma operação mais rápida do religador, caso o curto-circuito permaneça após o religamento. Essas curvas estão representadas na figura 8.



**Figura 8: Curva de um Religador da Subestação.**

A análise das figuras 7 e 8 será importante posteriormente, para a justificativa do tempo de análise (janela do sinal) selecionada para classificação das formas de onda (eventos).

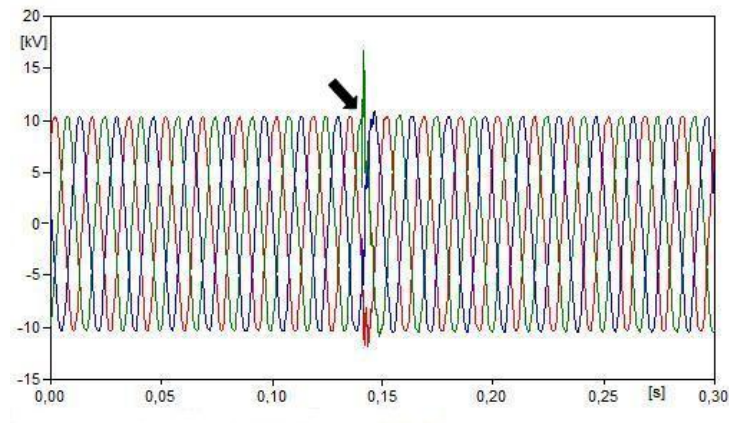
### 2.2.2 Transitórios Oscilatórios

As duas principais fontes de sobretensões transitórias em SEP são descargas atmosféricas e manobras em chaves ao longo do sistema, principalmente chaveamento em bancos de capacitores. Como mostrado anteriormente, as sobretensões transitórias podem ser geradas com diferentes conteúdos de frequência (tabela 1). A seguir, serão discutidos alguns aspectos dos principais causadores de sobretensões transitórias que estão dentro do escopo do presente trabalho.

#### Chaveamento de Banco de Capacitores

Os bancos de capacitores são normalmente utilizados em redes de distribuição com a finalidade de se corrigir o fator de potência, reduzindo perdas e quedas de tensão ao longo da rede, sendo que, em geral, a instalação de bancos de capacitores é a solução de mais baixo custo e com desempenho adequado para tal finalidade. A grande desvantagem são as sobretensões transitórias decorrentes do chaveamento desses bancos. Normalmente são observados transitórios de tensão de 1,0 a 2,0 p.u. da tensão nominal, resultando em problemas de isolamento e o mau funcionamento de equipamentos susceptíveis a tais variações de tensão. Outro aspecto importante no chaveamento de bancos de capacitores é o conteúdo de frequência associado com o

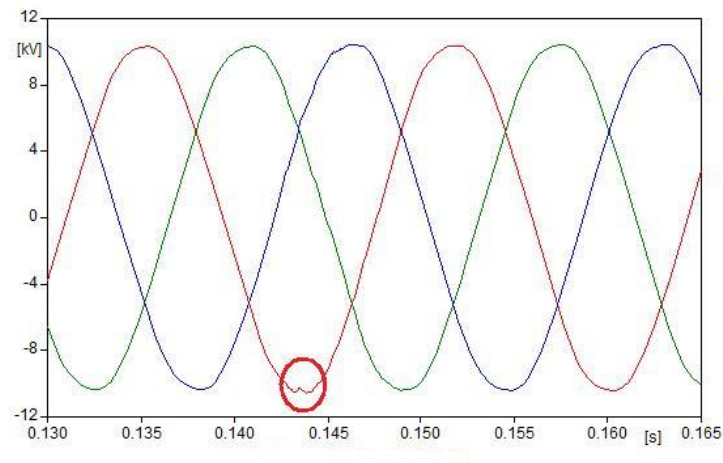
chaveamento. Supondo que o chaveamento de um banco de capacitores instalado em uma rede de distribuição ocorre quando a tensão no ponto onde o banco está instalado está no seu valor máximo e levando em conta que o capacitor não pode mudar instantaneamente o valor da tensão nos seus terminais, a tendência é a tensão no ponto da instalação do banco ir a zero e subir à medida que os capacitores forem carregando. Porém, em função do restante do sistema possuir uma indutância relevante, a tensão nos capacitores se eleva acima do nominal e oscila na frequência natural do sistema (DUGAN et al., 2002). A duração típica desses eventos é de 0,5 a 3 ciclos com frequência característica entre 300 e 1000 Hz. A figura 9 mostra a tensão na saída de uma subestação de distribuição para a simulação de um chaveamento de banco de capacitores trifásico, com potência de 600 kVAr instalado a uma distância de 5 km da subestação em questão. A seta indica o instante de ocorrência do chaveamento.



**Figura 9: Chaveamento de Banco de Capacitores.**

#### Abertura da Rede de Média Tensão

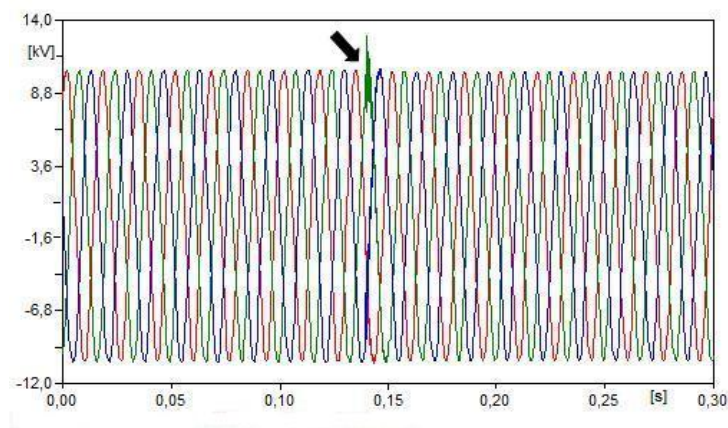
A abertura de um alimentador em redes de distribuição é um evento que não provoca sobretensões transitórias tão severas quanto os demais transitórios oscilatórios, porém é um fenômeno bastante comum em redes de distribuição e em alguns casos, em função da sua propagação, está associado com a queima de equipamentos em consumidores de baixa tensão. A figura 10 mostra a tensão na saída de uma subestação de distribuição para a abertura trifásica de um alimentador dessa subestação. Observa-se que após a ocorrência do evento (aproximadamente 0,145 s), a tensão nas três fases sofre uma pequena elevação (indicada com um círculo), em função da redução da corrente demandada pelo sistema de distribuição.



**Figura 10: Abertura de um Religador na Saída da Subestação.**

#### Fechamento da Rede de Média Tensão

O fechamento do religador de um alimentador de uma rede de distribuição pode provocar sobretensões transitórias com magnitudes de até 2 p.u. da tensão nominal, conteúdo de frequência superior ao de um chaveamento de banco de capacitores e duração típica de 0,5 ciclos (DUGAN et al., 2002). A figura 11 mostra a tensão na saída de uma subestação de distribuição para um religamento trifásico de um alimentador dessa subestação, com uma seta indicando o instante de religamento..



**Figura 11: Fechamento de um Religador na Saída da Subestação.**

### 3 REVISÃO DA LITERATURA SOBRE CLASSIFICAÇÃO DE EVENTOS

Pretende-se nesse capítulo apresentar os principais trabalhos sobre a classificação de eventos relacionados com esse trabalho. Basicamente, a classificação de eventos em um SEP pode ser dividida de acordo com:

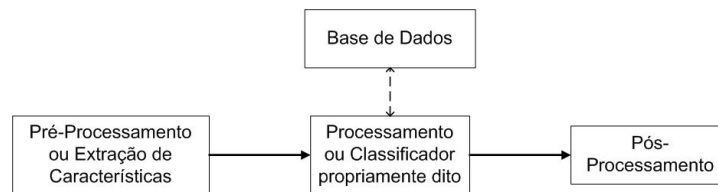
- O local de ocorrência ao longo do SEP - A classificação de eventos pode ser realizada com dados provenientes da Geração, Transmissão ou Distribuição de Energia, de acordo com o objetivo do classificador.
- Dados utilizados na classificação - Podem ser utilizados dados de tensão, corrente ou ambos para o processo de classificação. Eventualmente podem ser adicionadas informações de outra natureza, como p. ex. estado dos disjuntores à jusante ou à montante do local em que se registra o evento.
- Tipos de eventos a serem classificados - A escolha dos eventos está fortemente relacionada com a aplicação do classificador. De forma geral, existem classificadores voltados à classificação de eventos de qualidade de energia (afundamentos, elevações, transitórios, etc), classificadores para curtos-circuitos e classificadores que agrupam duas classes de eventos: curto-circuito e transitórios.

Além disso, a metodologia utilizada nesse trabalho para classificação dos eventos já vem sendo bastante discutida e aplicada com sucesso nos diversos trabalhos dessa área (a seguir apresentados), sendo que a mesma consiste basicamente em três estágios (figura 12):

- Pré-Processamento ou Extração de Características - Nesse estágio são extraídas as características dos sinais em análise, normalmente através do uso de uma transformação dos sinais no domínio do tempo para o domínio da frequência, via Transformada de Fourier, TF, ou para o domínio tempo-frequência, via Transformada Wavelet, TW.
- Processamento ou Classificação propriamente dita - O processamento é feito frequentemente por redes neurais artificiais (RNAs), onde são extraídas informações preliminares

acerca do banco de dados de eventos, utilizando os dados já processados.

- Pós Processamento - O pós-processamento consiste em se efetuar o processo de decisão, ao qual se aplica um grau de confiança ao resultado do estágio anterior (normalmente usando sistemas *fuzzy*), para cada classe do problema. Vale ressaltar que em alguns casos, em função da característica de saída do estágio de processamento, não é necessária a aplicação de um estágio de pós-processamento.



**Figura 12: Metodologia Normalmente Utilizada.**

A tabela 3 mostra, a partir das características mencionadas, alguns dos principais trabalhos relacionados com a classificação de eventos, principalmente os que abordam curto-circuito e eventos transitórios ao longo do SEP.

Através da análise dos trabalhos expostos na tabela 3, algumas observações podem ser feitas. A primeira delas diz respeito ao estágio de ocorrência do evento ao longo do SEP. Embora não tenham sido discutidos aqui, a grande maioria dos trabalhos sobre classificação de eventos se concentra em linhas de transmissão de energia. Uma possível justificativa para esse fato é que, historicamente, os esforços dispendidos para a automação do SEP se concentraram inicialmente na geração e transmissão.

No caso da transmissão em particular, houve uma preocupação intensa com o monitoramento das linhas, já que a partir dos dados obtidos desse monitoramento seria possível dispor de uma ferramenta que auxiliasse na mitigação de defeitos em linhas de transmissão. Cabe ressaltar que eventos que possam ocorrer nessas linhas (principalmente curtos-circuitos permanentes) podem resultar em desligamentos de grande proporção no SEP, uma vez que estas podem estar conectando pontos críticos do sistema.

Uma outra observação que pode ser feita acerca da tabela 3 diz respeito ao tipo de eventos classificados. Como mencionado anteriormente, uma possível separação para os eventos que ocorrem no SEP aborda duas classes gerais de eventos: curto-circuito e transitórios. De todos os trabalhos selecionados, a grande maioria leva em conta apenas os curtos-circuitos, dentre os quais destacam-se: (YOUSSEF, 2001; KASHYAP; SHENOY, 2003; DAG; UCAK, 2004; DOLA; CHOWDHURY, 2005; GAFOOR; RAO, 2006; AGUILERA; ORDUNA; RATTA, 2006;



**Tabela 3: Referências Bibliográficas Classificadas de Acordo com suas Características.**

Autor	Ano	Estágio de Ocorrência	Tipo de Eventos	Dados Utilizados	Pré-Proc.	Proc.	Pós-Proc.
Youssef	2001	Trans.	Faltas	Corrente	TW	Próprio	Não
Kashyap e Shenoy	2003	Trans.	Faltas	Tensão	TW	RNA	Não
Dag e Ucak	2004	Distr.	Faltas	Tensão e Corrente	TW	RNA	Não
Dola e Chowdhury	2005	Distr.	Faltas	Tensão e Corrente	Filtro	Vários	Não
Zheng-You et al.	2006	Trans.	Faltas e Transitórios	Corrente	TW	RNA	Não
Gafoor e Rao	2006	Trans.	Faltas	Corrente	TW	RNA	Não
Silva et al.	2006	Trans.	Faltas e Transitórios	Tensão e Corrente	TW	RNA	Não
Aguilera et al.	2006	Distr.	Faltas	Tensão	TW	Próprio	Não
Zadeh et al.	2006	Distr.	Faltas	Tensão e Corrente	TF	RNA	Não
Samantaray et al.	2006	Trans.	Faltas	Tensão e Corrente	THS	RNA	Não
Zhang et al.	2006	Trans.	Faltas	Corrente	EW	RNA	<i>Fuzzy</i>
Zhao et al.	2006	Geração	Faltas	Corrente	TWP	RNA	Não
Das et al.	2006	Trans.	Faltas	Corrente	TW e TF	Próprio	Não
Morais	2007	Trans.	Faltas	Tensão e Corrente	TW	Vários	Não
Davari et al.	2007	Distr.	Faltas	Campo Elétrico e Magnético	DFT	Próprio	Não
Malathi e Marimuthu	2008	Trans.	Faltas	Corrente	TW	RNA	Não
Florez et al.	2008	Distr.	Faltas	Tensão e Corrente	Próprio	RNA	Não
Ngaopitakkul et al.	2008	Trans.	Faltas Simultâneas	Corrente	TW	Próprio	Não
Pirouti et al.	2008	Distr.	Faltas	Tensão	TF	RNA	Não
Salim et al.	2008	Distr.	Faltas	Tensão e Corrente	TW	RNA	Não
Ravikumar et al.	2008	Trans.	Faltas	Tensão e Corrente	Est. de Fasores	RNA	Não
Raddy e Mohanta	2008	Trans.	Faltas	Corrente	TW	RNA	<i>Fuzzy</i>

KHORASHADI; AGHAEBRAHIMI, 2006; SAMANTARAY; DASH; PANDA, 2006; ZHANG; ZHENGYOU; QIAN, 2006; ZHAO et al., 2006; DAS; SINGH; SINHA, 2006; MORAIS et al., 2007; DAVARI; SHAHRTASH, 2007; MALATHI; MARIMUTHU, 2008; FLOREZ; MORALES; LONDONO, 2008), além de (NGAOPITAKKUL; PONGCHAISSRIKUL; KUNAKORN, 2008; PIROUTI; FATIH; SADIK, 2008; SALIM et al., 2008; RAVIKUMAR; THUKARAM; KHINCHA, 2008; RADDY; MOHANTA, 2008).

Apenas dois trabalhos (ZHENG-YOU; XIAOQING; GUOMING, 2006; SILVA; SOUZA; BRITO, 2006) abordam a classificação de faltas e transitórios de forma conjunta. Em Zheng-You, Xiaoqing e Guoming (2006) a análise conjunta no processo de classificação visa a identificação de faltas do tipo monofásicas e eventos transitórios como abertura de disjuntor, chaveamento de banco de capacitores e descargas atmosféricas. Já em (SILVA; SOUZA; BRITO, 2006) a análise conjunta visa a separação de eventos correspondentes à faltas (monofásicas, bifásicas e trifásicas) de eventos transitórios, porém sem identificar o tipo de evento transitório ocorrido. Além disso, ambos trabalhos estão voltados para sistemas de transmissão.

Com relação aos dados de entrada dos classificadores, os trabalhos utilizam em geral dados de tensão, corrente ou ambas as grandezas para o processo de classificação. Normalmente quando se procura classificar faltas, utilizam-se dados de corrente e, eventualmente dados de tensão, pois são as correntes (nas três fases) que caracterizam um curto-circuito. Já para transitórios, utilizam-se dados de tensão, uma vez que transitórios estão associados com fenômenos de QEE, e a QEE por sua vez, está relacionada com a forma de onda de tensão. No entanto, a escolha dos dados depende muito da sua disponibilidade desses dados. Em um sistema de distribuição, alvo do presente trabalho, nem sempre se dispõe de dados de tensão e corrente. Um exemplo dessa realidade é o trabalho realizado por Davari e Shahrtash (2007), que utiliza medidas de campos eletromagnéticos (visando obter uma estimativa para o potencial elétrico e corrente elétrica) para o processo de classificação.

Já para o pré-processamento, observa-se o uso expressivo da TW. Como pode ser encontrado em grande parte dos trabalhos publicados nessa área, existem alguns aspectos nessa transformação que justificam seu uso, como p. ex.:

- Representação da TW em função da soma de diversas funções wavelet com diferentes escalas e localizações (característica tempo-frequência). Essa representação permite que sejam acopladas técnicas de classificação e detecção de eventos no mesmo processo de decomposição wavelet.
- Melhor caracterização dos transitórios envolvidos, resultado das funções wavelet uti-

lizadas na transformação.

- Possibilidade de se aplicar técnicas de extração de características que comprimem boa parte da informação contida na resposta da TW, reduzindo a complexidade dos classificadores do estágio de processamento.

Muito embora essas características já estejam validadas na literatura para os problemas de classificação em questão (DAG; UCAK, 2004), ainda existem algumas questões a serem discutidas no uso da TW. Uma delas diz respeito à escolha da função wavelet (chamada wavelet-mãe). Existem diversas famílias de funções wavelet que podem ser aplicadas no intuito de pré-processar os sinais de análise visando a classificação de eventos. No entanto, não se encontra um método para a seleção da função wavelet mais apropriada para o pré-processamento, principalmente devido a particularidades existentes em cada aplicação. Em geral, a metodologia consiste em se investigar as diversas famílias visando encontrar a que forneça, em última instância, um melhor desempenho na classificação. Além disso, como mostram Das, Singh e Sinha (2006), em determinados casos, é possível que a TF possa apresentar uma boa caracterização de eventos para o processo de classificação, superando a TW e evidenciando que a eficiência do pré-processamento é extremamente dependente da aplicação de uma metodologia definitiva para a escolha do algoritmo mais adequado sendo ainda é uma questão em aberto.

No estágio de processamento observa-se que a maior parte da literatura utiliza RNA ou comparativos entre RNA e outros métodos. Alguns trabalhos (YOUSSEF, 2001; AGUILERA; ORDUNA; RATTA, 2006; DAS; SINGH; SINHA, 2006; DAVARI; SHAHRTASH, 2007; NGAOPITAKKUL; PONGCHAIKRIKUL; KUNAKORN, 2008) utilizam um método de processamento próprio, extraindo informações da análise dos eventos a serem classificados. Existe ainda um outro grupo que utiliza outros métodos de classificação como o método baseado em árvores de decisão encontrado em (MORAIS et al., 2007). Uma característica comum a todos esses métodos é a forte dependência de algum conhecimento prévio dos problemas ou até mesmo uma custosa análise dos eventos a serem classificados, o que demanda conhecimentos específicos do problema. No caso das RNA existem algumas questões ainda pouco exploradas para esse tipo de classificação, como o controle de complexidade e seleção de entradas. Essas duas características (principalmente o controle de complexidade) conferem ao modelo neural uma característica autônoma, reduzindo a dependência de conhecimentos específicos dos eventos. Esses modelos já foram aplicados com sucesso em problemas de previsão de carga (FERREIRA; SILVA, 2007).

Por último, no estágio de pós-processamento, podem ser utilizadas técnicas que complementam o estágio de processamento, como p. ex. sistemas *fuzzy*. No entanto, como apre-

sentado anteriormente, a utilização desse estágio está relacionada com as características do processamento realizado, sendo que em grande parte dos casos, não é necessária a utilização do pós-processamento.

Sendo assim, pretende-se nesse trabalho apresentar um método de classificação conjunta de faltas e eventos transitórios para sistemas de distribuição, com base na medida de tensão trifásica na saída de uma subestação de distribuição (com base nos registros do PQM), comparando diversas formas de pré-processamento (via TW e TF) e redes neurais com características autônomas (com pouca intervenção do usuário), as quais não necessitam de um estágio de pós-processamento.

Nos próximos tópicos serão apresentados os demais aspectos teóricos envolvidos com a metodologia aqui descrita, descrevendo os métodos de pré-processamento utilizados e as diversas estruturas de RNAs testadas, bem como demais técnicas associadas a esses tópicos, utilizadas ao longo do desenvolvimento desse trabalho.

## 4 CONCEITOS E MÉTODOS PARA CLASSIFICAÇÃO DE EVENTOS

### 4.1 PRÉ-PROCESSAMENTO

Nesse primeiro tópico serão descritos os métodos de pré-processamento utilizados nesse trabalho, os quais têm por objetivo extrair as informações contidas nos sinais sob análise (tensão trifásica na barra de uma subestação de distribuição), iniciando pela análise de Fourier em uma abordagem normalmente utilizada na literatura, visando à apresentação dos principais aspectos da transformada wavelet.

#### 4.1.1 Análise de Fourier

As ferramentas da análise de Fourier podem ser divididas em dois grupos, as séries e as transformadas, tanto para o domínio do tempo discreto quanto para o domínio do tempo contínuo. De forma geral, tanto as séries quanto as transformadas, decompõem uma função (ou sinal) em uma somatória de exponenciais complexas harmonicamente e infinitesimalmente relacionadas, respectivamente. Esta decomposição é de grande valia, pois as exponenciais complexas são auto-vetores de sistemas lineares invariantes no tempo (OPPENHEIM; WILLSKY, 1997).

#### Série de Fourier Contínua

Uma série de Fourier contínua decompõe uma função em uma somatória ponderada de senos e cossenos, sendo aplicada em funções periódicas. Uma função periódica pode ser representada pela série trigonométrica de Fourier mostrada na equação (1).

$$f(t) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} (a_n \cos n\omega_0 t + b_n \text{sen} n\omega_0 t) \quad (1)$$

Sendo que  $a_0$ ,  $a_n$  e  $b_n$  representam os coeficientes de Fourier e  $\omega_0$  a frequência angular fundamental do sinal no domínio do tempo  $t$ . Para que a série de Fourier exista é necessário

que algumas condições sejam respeitadas. Essas condições são conhecidas como Condições de Dirichlet (OPPENHEIM; WILLISKY, 1997), são elas:

- A função deve ter um número finito de descontinuidades em um período.
- A função deve ter um número finito de máximos e mínimos num período.
- A função deve absolutamente integrável em um período.

Para se determinar a série de Fourier de uma função é necessário se determinar os valores de  $a_n$  e  $b_n$  da equação (1). Esses coeficientes são dados pelas equações (2) e (3) (OPPENHEIM; WILLISKY, 1997).

$$a_n = \frac{2}{T} \int_{-T/2}^{T/2} f(t) \cos(n\omega_0 t) dt \text{ com } n = 0, 1, 2, \dots, \quad (2)$$

$$b_n = \frac{2}{T} \int_{-T/2}^{T/2} f(t) \text{sen}(n\omega_0 t) dt \text{ com } n = 1, 2, 3, \dots, \quad (3)$$

Onde  $T$  é o período associado à função. A série de Fourier pode ser dada também, na sua forma complexa, como mostra a equação (4).

$$f(t) = \sum_{n=-\infty}^{\infty} c_n e^{jn\omega_0 t}, \quad (4)$$

Sendo o coeficiente  $c_n$  determinado pela equação (5):

$$c_n = \frac{1}{T} \int_{-T/2}^{T/2} f(t) e^{-jn\omega_0 t} dt \text{ com } n = 0, \pm 1, \pm 2, \dots \quad (5)$$

A partir dos coeficientes determinados é possível a obtenção dos espectros de frequência discretos que caracterizam a resposta da série contínua de Fourier.

#### Série de Fourier Discreta

Assim como sinais contínuos, sinais discretos podem ser representados por uma somatória de exponenciais complexas, como na equação (4). A principal diferença entre a série contínua e a discreta é que um sinal discreto é representado por uma somatória finita. Esta característica

tem como consequência a não existência de questões matemáticas relacionadas à convergência da série discreta (OPPENHEIM; WILLSKY, 1997). A série de Fourier de tempo discreto é dado pela equação (6).

$$x[n] = \sum_{k=0}^{N-1} a_k e^{jk \frac{2\pi}{N} n}, \quad (6)$$

Sendo  $N$  o período associado à função e o coeficiente  $a_k$  determinado pela equação (7).

$$a_k = \frac{1}{N} \sum_{n=0}^{N-1} x[n] e^{-jk \frac{2\pi}{N} n} \quad (7)$$

### Transformada Contínua de Fourier

Para que exista a série de Fourier de um sinal, este deve ser periódico. Como mostra HSU (1973), a tentativa de decompor sinais não periódicos manipulando as expressões (1), (2) e (3) realizada por Fourier, mostrou-se que um sinal não periódico é a somatória de exponenciais complexas infinitesimalmente relacionadas. Esse resultado leva em conta que o período de um sinal não periódico é infinito. Sendo assim, obtém-se que a transformada de Fourier de um sinal é dado pela equação (8):

$$X(j\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt \quad (8)$$

A partir da função  $X(j\omega)$  obtida em (8), é possível restaurar a função original através da equação (9):

$$x(t) = \int_{-\infty}^{\infty} X(j\omega) e^{j\omega t} d\omega, \quad (9)$$

Onde  $\omega$  é a frequência angular do sinal (função) de análise. As equações (8) e (9) são conhecidas como transformada de Fourier e transformada inversa de Fourier, respectivamente.

### Transformada Discreta de Fourier

A transformada discreta de Fourier é obtida a partir da série discreta de Fourier de modo semelhante à transformada contínua de Fourier. O par de transformadas discretas é dado pelas equações (10) e (11):

$$X(e^{j\Omega}) = \sum_{n=-\infty}^{\infty} x[n]e^{-j\Omega n} \quad (10)$$

e

$$x[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\Omega})e^{j\Omega n} d\Omega, \quad (11)$$

sendo  $x[n]$  o sinal no domínio do tempo discreto,  $X(e^{j\Omega})$  o sinal discreto transformado para o domínio da frequência e  $\Omega$  a frequência angular.

Uma última observação relevante está relacionada com propriedades de periodicidade das representações de Fourier. De forma geral, representações contínuas em um determinado domínio (tempo ou frequência) são não periódicas no outro domínio. O contrário também é válido, representações discretas em um determinado domínio (tempo ou frequência) são periódicas no outro domínio (HAYKIN; VEEN, 2001).

#### Transformada Rápida de Fourier

Em 1962, um algoritmo chamado Transformada Rápida de Fourier, TRF, foi apresentado. Sua história é longa e suas origens podem ser encontradas em algumas anotações de Gauss, como mostra (OPPENHEIM; WILLSKY, 1997). Seu ressurgimento foi de grande valia para o desenvolvimento moderno da análise de Fourier, principalmente no que se refere ao processamento digital de sinais, tornando computacionalmente viável a implementação de diversas aplicações que utilizam a transformada e a série discreta de Fourier. O cálculo dos coeficientes da transformada discreta de Fourier requer  $N^2$  multiplicações complexas e adições (MALLAT, 1999), mas através da reorganização dos cálculos na TRF, a ordem de complexidade do algoritmo é reduzido a  $N \log_2 N$ , sendo  $N$  o comprimento do sinal. Devido ao baixo custo computacional, a TRF é também usada para efetuar convoluções discretas. Nesse trabalho foi utilizada a implementação presente na *Signal Processing Toolbox* do *software* Matlab com mais detalhes disponíveis em Mathworks (2009a).

#### Transformada de Fourier Janelada

A transformada de Fourier janelada, TFJ, inicialmente proposta por Dennis Gabor (OPPENHEIM; WILLSKY, 1997), consiste em aplicar uma janela de observação do sinal, deslocada no tempo, de forma a se obter o mapeamento do sinal sob análise em um gráfico bidimensional de tempo e frequência. Assim, inclui-se o termo indicador da variação temporal na equação



responsável pelo cálculo da transformada de Fourier. Esse termo também pode ser aplicado no caso da transformada discreta de Fourier, como mostra a equação (12),

$$WDFT(k, m) = \sum_n x[n]w[n - m]e^{\left(\frac{-j2\pi kn}{N}\right)}, \quad (12)$$

sendo que a função janela pode ser representada pela equação (13).

$$w(n - m) = \begin{cases} 1, & \text{se } 0 \leq (n - m) \leq (N - 1) \\ 0, & \text{caso contrario} \end{cases} \quad (13)$$

O objetivo de uma decomposição tempo-frequência de um sinal é separar esse sinal em várias partes (função janela) e analisá-las separadamente, definindo assim, componentes de frequência existentes em diferentes faixas de tempo permitindo assim, o tratamento de sinais não estacionários.

Nesse ponto, uma questão importante que deve ser ressaltada, diz respeito à precisão com que se pode localizar a função em análise no domínio do tempo e da frequência. Segundo (MALLAT, 1999), essa precisão está associada com o princípio da incerteza, que afirma não ser possível determinar com precisão a localização no tempo e na frequência simultaneamente. Sendo assim, ao se reduzir a função janela, maior é a resolução temporal, mas menor é a resolução em frequência, e vice-versa.

No caso da TFJ, a escolha da janela é fortemente dependente desse princípio e pode, por vezes, comprometer a análise através dessa transformação. Portanto, evidencia-se assim, uma dificuldade em se utilizar a TFJ, que é a de adaptar uma função janela para englobar as componentes de frequência existentes no sinal, com uma boa resolução temporal, uma vez que a função janela é fixa ao longo da decomposição.

#### 4.1.2 Transformada Wavelet

Tendo em vista a dificuldade exposta para a TFF, um grupo de pesquisadores franceses (MALLAT, 1999) introduziu uma nova transformada, denominada Transformada Wavelet, cujo tamanho da janela é variável, permitindo que sinais com conteúdo bastante diversificado em frequência pudessem ser analisados com boa resolução temporal em cada faixa de frequência (DELMONT-FILHO, 2007). Essa transformação utiliza uma função de translação e escala, onde a função janela é deslocada e escalonada, cobrindo todo o espectro desejado. A resposta dessa transformada é um conjunto de representações tempo-frequência, todas com diferentes resoluções (análise multiresolução).

## Transformada Wavelet Contínua

A análise wavelet de um sinal  $x(t)$  pode ser dada através da transformada wavelet contínua e é definida como mostra a equação (14):

$$TWC(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t) \psi \left( \frac{t-b}{a} \right) dt, \quad (14)$$

Sendo  $a$  o fator de escala,  $b$  o fator de translação e  $\Psi$  denominada wavelet-mãe (análoga aos senos e cossenos da série de Fourier). O termo  $a^{-\frac{1}{2}}$  é utilizado para normalização dos coeficientes da transformada diante das diferentes escalas. Esses coeficientes representam o quanto o sinal em análise e a wavelet-mãe transladada e escalonada se correlacionam. O fator de escala e o fator de translação são análogos à frequência e ao tempo, respectivamente. Pode-se definir também, a transformada wavelet inversa, que efetua a reconstrução do sinal a partir dos coeficientes da transformada wavelet contínua, como mostra a equação (15).

$$x(t) = \frac{1}{\sqrt{a}} \int \int TWC(a, b) \psi \left( \frac{t-b}{a} \right) da db. \quad (15)$$

## Propriedades da Transformada Wavelet

Antes da apresentação das propriedades envolvidas com a transformação, deve-se levar em conta a energia do sinal em análise, ou seja, é necessário que o mesmo tenha energia finita para que seja possível cobrir todo o espectro de frequência através da transformada wavelet. Baseado nisso, pode-se mostrar as duas propriedades que dão à transformada wavelet sua característica: a admissibilidade e a regularidade. A admissibilidade está relacionada com a possibilidade de reconstrução do sinal sem distorção, desde que sua transformada de Fourier,  $\Psi(\omega)$ , seja nula para  $\omega = 0$ , ou seja:

$$|\Psi(\omega)|^2|_{\omega=0} = 0. \quad (16)$$

Uma consequência dessa propriedade, é que a média do sinal da wavelet-mãe calculada no tempo, deve ser zero:

$$\int \psi(t) dt = 0. \quad (17)$$

Já a condição de regularidade, mostra que a wavelet-mãe deve possuir rápido decaimento

para zero na bordas, caracterizando a duração da wavelet-mãe. Para se determinar a regularidade de uma função wavelet é necessário estimar o número de momentos nulos associado (SOUZA et al., 2007). Uma função wavelet  $\Psi(t)$  possui  $p$  momentos nulos se  $\int_{-\infty}^{\infty} t^k \psi(t) dt = 0$ , para  $0 \leq k < p$ . Quanto maior o número de momentos nulos envolvidos, maior a regularidade (ou suavidade) da função wavelet.

Existem ainda outras questões relevantes relacionadas com as propriedades das funções wavelet, como:

- Simetria associada à função;
- Suporte compacto, ou seja, se a maioria da energia da função wavelet está restrita a um intervalo finito;
- Possibilidade de formação de um sistema ortogonal;
- Complexidade na formulação matemática envolvida.

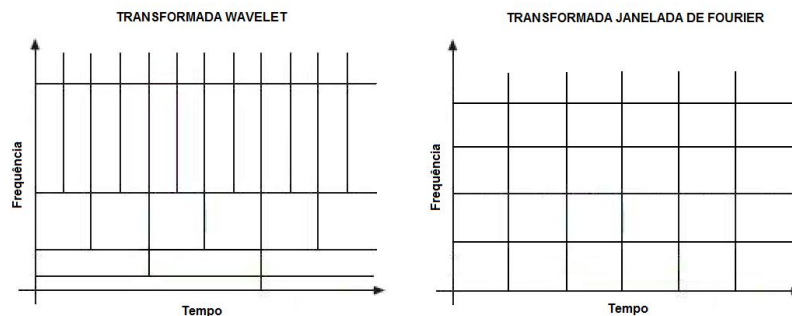
No decorrer da apresentação das funções wavelet utilizadas nesse trabalho, essas propriedades serão retomadas, complementando algumas de suas principais características.

### Transformada Wavelet Discreta

Uma vez estabelecido o conceito da transformada wavelet, deve-se procurar uma forma prática para sua implementação, já que a aplicação direta da transformada wavelet contínua apresenta alguns inconvenientes. O primeiro deles é que, durante o cálculo da TWC, as funções resultantes do escalonamento e translação da wavelet-mãe perdem a característica de ortogonalidade (MALLAT, 1999). Essa característica de ortogonalidade está relacionada com a possibilidade de reconstrução do sinal, ou seja, se a transformação utilizar bases ortogonais, existe a garantia da reconstrução exata do sinal através da transformada wavelet inversa. A perda de ortogonalidade resulta em coeficientes com elevada redundância. Uma forma de se evitar essa redundância é, ao invés de escalonar e transladar a função wavelet-mãe continuamente, dividir o escalonamento e translação em faixas, discretizando a faixa do cálculo da transformada. Além dessa divisão, é necessária a ortogonalidade da transformação, que deve ser feita através da escolha de uma wavelet-mãe ortogonal, para o caso discreto. Essa escolha elimina o problema da redundância relacionado com a transformação. A equação (18) mostra o cálculo dessa transformada:

$$TWD(m, k) = \frac{1}{\sqrt{a_0^m}} \sum_n x(n) \psi \left( \frac{k - nb_0 a_0^m}{a_0^m} \right), \quad (18)$$

Sendo  $a$  o fator de escala,  $b$  o fator de translação (ambos função de um parâmetro inteiro  $m$ ) e  $k$  a amostra do sinal de entrada em questão. Normalmente utiliza-se  $a_0$  como sendo 2 e  $b_0$  como sendo 1, originando assim, uma amostragem diádica tanto para o eixo das frequências como para o eixo do tempo (contribuindo para formação de uma base ortonormal para o espaço  $L^2(\mathfrak{R})$ )<sup>1</sup>. A figura 13 mostra gráfico tempo-frequência, resultado da aplicação da transformada wavelet discreta, TWD, em comparação com o gráfico de tempo-frequência gerado pela TJF, assinalando novamente a questão da representação da TW.



**Figura 13: Espectros Tempo-Frequência.**

**Fonte: (DELMONT-FILHO, 2007)**

Outro aspecto a se observar é com relação ao espectro gerado pela translação e escalonamento da wavelet-mãe. Da teoria de Fourier tem-se que uma compressão no tempo é equivalente a uma dilatação na frequência e deslocamento do espectro. Com isso, tem-se que uma compressão de fator dois no tempo, provoca uma dilatação de dois na frequência e, consequentemente, um deslocamento dos componentes de frequência (espectro). Essa propriedade pode ser aplicada para o caso da transformada wavelet, onde a wavelet-mãe é escalonada no tempo. Assim, chega-se ao espectro que pode ser coberto através do escalonamento e translação da wavelet-mãe. Para se ter uma boa cobertura do espectro, deve-se fazer com que cada passo do escalonamento sobreponha-se aos seus adjacentes (MALLAT, 1999).

Essa abordagem implica em se observar a wavelet-mãe como um filtro sendo que a série de wavelets geradas pelo escalonamento da wavelet-mãe podem ser vistas como um banco de filtros. Essa interpretação será útil mais adiante, quando for apresentado o algoritmo rápido da transformada wavelet discreta.

<sup>1</sup>O espaço  $L^2(\mathfrak{R})$  é definido como todas funções mensuráveis de quadrado integrável sobre  $\mathfrak{R}$ .

## Função de Escalonamento

A implementação da transformada wavelet discreta apresenta inconvenientes. O primeiro deles, com relação à redundância, é resolvido com a escolha adequada da wavelet-mãe. Outro inconveniente está relacionado com o número de dilatações e compressões necessárias para cobrir todo o espectro. A cada nova dilatação da wavelet-mãe no tempo, por exemplo, uma nova parte do espectro é coberta, deixando ainda uma parte a ser atendida pelas próximas dilatações no tempo. Isso significa que, para se ter todo o espectro atendido, seriam necessárias infinitas dilatações no tempo. Uma solução para este problema foi proposta por (MALLAT, 1999), onde é introduzida uma função denominada função de escalonamento, responsável por cobrir as frequências de ordem mais baixa.

A principal conclusão dessa técnica é que pode-se combinar o resultado do espectro da wavelet-mãe com o resultado do espectro da função de escalonamento, cobrindo todo o espectro desejado sem necessitar de um número infinito de dilatações ou compressões para se cobrir o espectro. A função de escalonamento pode ser vista como um filtro passa-baixas, com largura de faixa ajustável dependendo da análise escolhida. A condição para a escolha da largura de faixa desse filtro é que o sinal possa ser reconstruído a partir dos coeficientes do resultado fornecido pela wavelet-mãe e função de escalonamento. Isso implica que a largura de faixa desta função de escalonamento não pode introduzir perda de informação na transformação.

## Transformada Wavelet Discreta - Algoritmo Rápido

A última questão com relação à implementação da transformada wavelet discreta é a forma como deve ser feita a transformada utilizando banco de filtros. Essa metodologia irá resultar na construção do algoritmo rápido para o cálculo da transformada e consiste em dividir o espectro do sinal de análise em duas sub-bandas de mesmo comprimento, a sub-banda das baixas frequências e a sub-banda das altas frequências. Essa divisão pode prosseguir até se chegar à decomposição desejada ou até o número de divisões permitidas pelo comprimento do sinal de análise (número de amostras). A figura 14, mostra a abordagem em questão.

As saídas dos filtros passa-altas (PA), como  $4N$  e  $2N$ , são denominadas de detalhes, já que nelas estão contidas as componentes de alta frequência do sinal. As saídas dos filtros passa-baixas (PB) são denominadas aproximações e nelas estão as componentes de frequência responsáveis por caracterizar o sinal, como representado na figura 15.

Os filtros utilizados nesse cálculo possuem uma característica especial, pois são os mesmos para todos os níveis (passa-altas e passa-baixas) e estão correlacionados pela expressão (19):

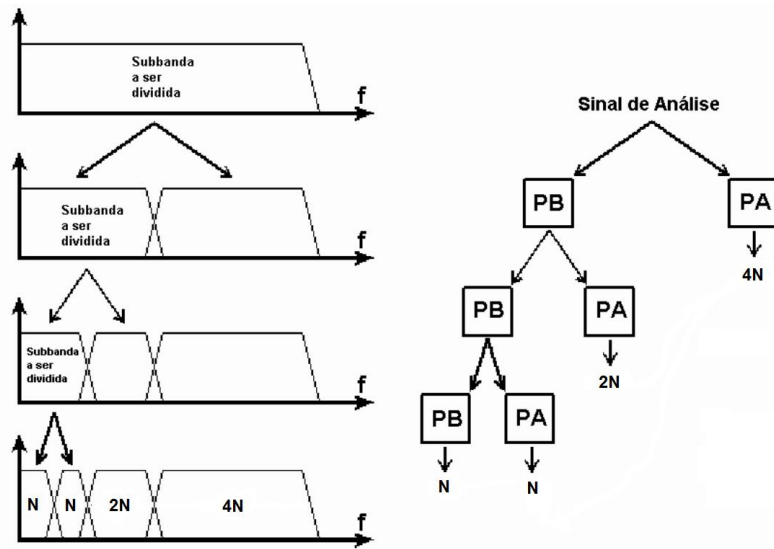


Figura 14: Decomposição do sinal através do banco de filtros.

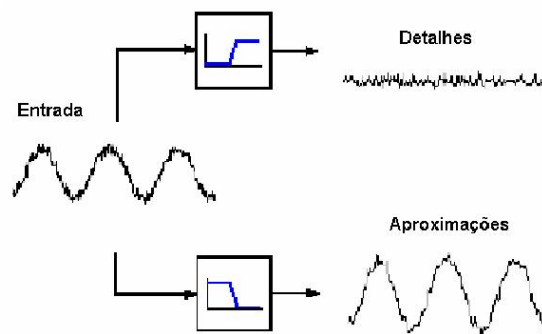


Figura 15: Aproximações e Detalhes.

Fonte: (MATHWORKS, 2009b)

$$h[L-1-n] = (-1)^n l(n), \quad (19)$$

Sendo  $L$  o comprimento do filtro e  $h$  e  $l$  os filtros passa-altas e passa-baixas, respectivamente. Esses tipos de filtros são comumente usados em processamento de sinais e são conhecidos como *quadrature mirror filters* (filtros em quadratura espelhada). No caso de funções wavelet não simétricas, os filtros de decomposição e recomposição são diferentes e podem ser gerados a partir da função de escalonamento (filtro passa-baixas) e a partir da wavelet-mãe (filtro passa-altas).

Na prática, utiliza-se uma subamostragem de fator dois na saída de cada filtro, de forma a se otimizar os sinais, pois com o espectro dividido ao meio, as amostras tornam-se redundantes. Essa subamostragem insere erro na análise, que devem ser levados em conta no processo de

decomposição através da transformada em questão. A reconstrução utiliza a mesma ideia, mas ao invés de subamostrar, ela insere amostras (com valor zero) no sinal, para se obter no final, um sinal com o mesmo número de amostras do sinal inicialmente decomposto. Essa metodologia fica mais clara visualizando-se as figuras 16 e 17:

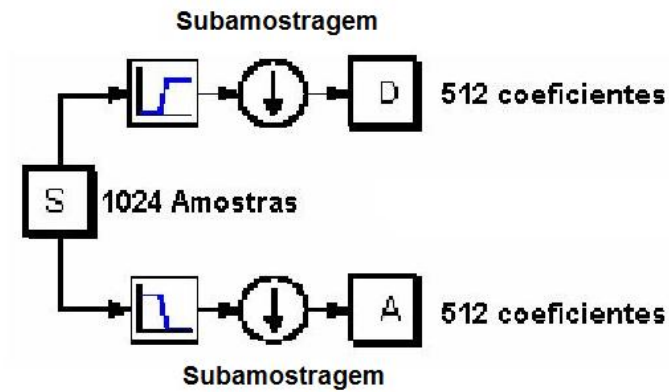


Figura 16: Decomposição em um nível para 1024 Amostras.

Fonte: (MATHWORKS, 2009b)

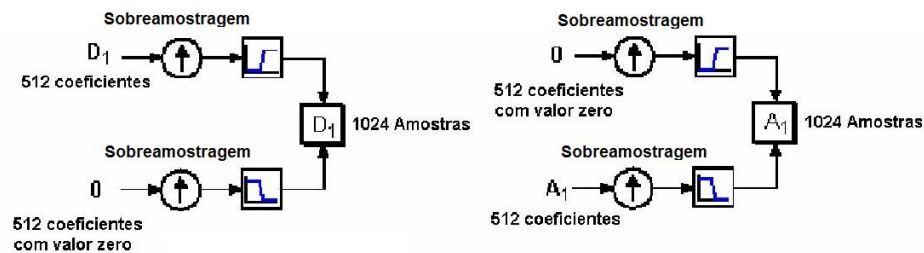
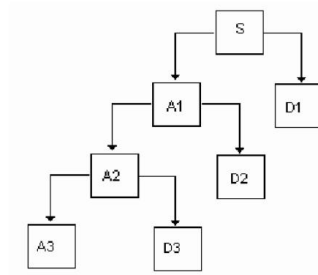


Figura 17: Reconstrução em um nível para 512 coeficientes.

Fonte: (MATHWORKS, 2009b)

Em uma decomposição de um único nível, a reconstrução pode ser dada pela soma direta das aproximações e detalhes do primeiro nível, ou seja,  $A_1 + D_1$ . Para vários níveis, a reconstrução deve conter a soma dos termos de aproximação e detalhe do último nível e os termos dos detalhes dos níveis anteriores, como na figura 18.

Na figura 18, a reconstrução do sinal pode ser dada pela soma de  $A_3 + D_3 + D_2 + D_1$ , uma vez que esses sinais tenham sido sobreamostrados e filtrados pelos filtros de reconstrução. Uma outra forma de se observar a característica do banco de filtros na análise wavelet discreta é através da comparação da equação (20), que calcula a transformada wavelet discreta com a equação (21), relacionada com a convolução discreta de um filtro de resposta finita ao impulso, FIR.



**Figura 18: Recomposição em um nível para 512 coeficientes.**

Fonte: (MATHWORKS, 2009b)

$$TWD(m, k) = \frac{1}{\sqrt{a_0^m}} \sum_n x(n) \Psi(a_0^{-m}k - nb_0) \quad (20)$$

$$y(n) = \frac{1}{c} \sum x(k) h(n - k), \quad (21)$$

Sendo,  $h(n - k)$  a resposta ao impulso do filtro do tipo FIR. Assim sendo, pode-se inferir que o termo  $\Psi(a_0^{-m}k - nb_0)$  é a resposta ao impulso do filtro da equação da transformada wavelet discreta, caracterizando uma implementação prática para essa transformada. Nesse trabalho foi utilizado o algoritmo proposto por Mallat (1999) implementado e validado na Wavelet *toolbox* do software Matlab, como mostra (MATHWORKS, 2009b).

### Funções Wavelet

Atualmente, existem diversas funções wavelet disponíveis na literatura, sendo seu uso fortemente dependente da aplicação. Em função da aplicação, é possível estabelecer regras para construção de novas funções wavelets, atendendo às restrições e necessidades da aplicação, bem como a complexidade envolvida na formulação dessas regras (FARIA, 1997). Este trabalho não tem por objetivo construir uma nova função wavelet para o processamento em questão, mas sim, testar algumas das wavelets mais utilizadas para análise de sinais em classificação de eventos ao longo do SEP, visando a comparação do desempenho do processo de classificação como um todo. Foram utilizadas funções wavelets ortogonais (das famílias Haar, Daubechies, Symmlets e Coiflets) e wavelets biortogonais:

- Haar: É a wavelet mais simples no espaço  $L^2(\mathfrak{R})$ . Em função de não ser continuamente diferenciável, essa wavelet possui poucas aplicações atualmente (DELMONT-FILHO, 2007).



- Daubechies: Introduzidas inicialmente pela pesquisadora Ingrid Daubechies (DAUBECHIES, 1992), essas wave-lets possuem uma característica bastante relevante relacionada com sua ortonormalidade e a capacidade de gerar suporte compacto. O suporte compacto está associado com a ordem da wavelet, ou seja, quanto maior a ordem do filtro, maior a compactação do suporte. O tamanho do filtro de uma wavelet Daubechies é o dobro do número de momentos nulos dessa wavelet. Quando a wavelet de Daubechies possui apenas um momento nulo, obtém-se a wavelet de Haar. As wavelets de Daubechies são representadas pela notação  $dbN$ , sendo  $N$  a ordem associada com o filtro. Neste trabalho foram testadas as wavelets:  $db1$  (Haar),  $db4$ ,  $db8$ ,  $db12$  e  $db15$ .
- Symmlets: São funções wavelet bastante parecidas com as Daubechies, também desenvolvida por Ingrid Daubechies (DAUBECHIES, 1992). Possuem algumas vantagens quando comparadas com as wavelets Daubechies, principalmente por apresentar maior simetria. As wavelets de Daubechies são representadas pela notação  $symN$ , sendo  $N$  a ordem associada com o filtro. Nesse trabalho foram utilizadas  $sym2$ ,  $sym4$  e  $sym8$ .
- Coiflets: São wavelets caracterizadas por possuírem semelhança maior entre o sinal original e o transformado. Essas wavelets possuem momentos nulos tanto para a wavelet mãe, quanto para a função escala e são caracterizadas por apresentar uma característica quase simétrica. A representação é feita de acordo com a ordem da wavelet (análogo a wavelet Daubechies). As wavelets de Daubechies são representadas pela notação  $coifN$ , sendo  $N$  a ordem associada com o filtro. Nesse trabalho foram utilizadas  $coif1$ ,  $coif3$  e  $coif5$ .
- Wavelets Biortogonais: Foram desenvolvidas para possuir simetria e suporte compacto, sendo esse o principal diferencial sobre as demais famílias de wavelets. As wavelets de Daubechies são representadas pela notação  $biorN1.N2$ , sendo  $N1$  a ordem associada com o filtro de decomposição e  $N2$  a ordem associada com o filtro de reconstrução. Neste trabalho foram utilizadas  $bior1.1$ ,  $bior1.5$ ,  $bior2.2$ ,  $bior2.8$ ,  $bior3.1$ ,  $bior3.9$ ,  $bior4.4$ ,  $bior5.5$  e  $bior6.8$ .

Mais detalhes sobre as famílias wavelet e a obtenção dos filtros de decomposição e reconstrução podem ser obtidos em (DAUBECHIES, 1992).

#### Extração de Características

Com objetivo de extrair o máximo de características dos sinais em análise, sem perda de informação relevante, foi aplicado um pré-processamento utilizando o cálculo da energia contida nas várias sub-bandas no caso da TW e nas diversas raias do espectro no caso da TRF.

O objetivo principal do cálculo da energia por sub-banda ou raia (como descrito na maioria dos trabalhos que utilizam TW e TRF) está relacionado com uma melhor caracterização do conteúdo de frequência dos eventos, além de uma redução da dimensionalidade para o sistema de classificação no caso da TW, uma vez que os sinais de tensão trifásica na saída da subestação foram modelados de acordo com a aquisição feita pelo PQM, cuja taxa de amostragem é de 7680 Hz, resultando em 128 amostras por ciclo de 60 Hz. Essa quantidade de amostras pode levar a uma complexidade computacional elevada principalmente no caso da TW, de forma que uma redução da dimensionalidade se torna uma tarefa necessária para o processo de classificação como um todo.

Para o pré-processamento baseado em TRF, os sinais de tensão das três fases da barra da subestação foram decompostos pela transformada de Fourier e para o cálculo de energia, foi considerado até o quinquagésimo harmônico para cada uma das três fases (de acordo com a amostragem), totalizando um vetor de entrada com 150 valores correspondentes à magnitude das raias da TRF.

Já para o pré-processamento baseado em TW, os sinais de tensão das três fases foram decompostos em dez níveis, sendo obtidos os dez sinais detalhe mais o sinal aproximação para cada fase, totalizando 33 sinais no domínio wavelet. Em função da taxa de amostragem do sinal ser 7680 Hz, optou-se por utilizar dez níveis da transformada, já que com esse número de níveis é possível cobrir grande parte do conteúdo de frequência do sinal, com uma resolução adequada. De acordo com a taxa de amostragem utilizada, a divisão em frequência para a decomposição fica como mostrado na tabela 4.

**Tabela 4: Bandas de Frequência da Decomposição Wavelet.**

Nível da Transformada	Banda de Frequência (Hz)
Sinal Detalhe 1	3840-1920
Sinal Detalhe 2	1920-960
Sinal Detalhe 3	960-480
Sinal Detalhe 4	480-240
Sinal Detalhe 5	240-120
Sinal Detalhe 6	120-60
Sinal Detalhe 7	60-30
Sinal Detalhe 8	30-15
Sinal Detalhe 9	15-7,5
Sinal Detalhe 10	7,5-3,75
Sinal Aproximação 10	0-3,75

Dessa forma, o vetor de entrada para o classificador no caso de um pré-processamento wavelet, possui 33 valores (11 coeficientes para cada uma das três fases).

Depois da decomposição (Wavelet ou Fourier) é realizado o cálculo de energia em cada um dos níveis da TW ou das raias da TRF. Esse cálculo é feito com base no conteúdo de energia pré-falta e durante a falta, ou seja, é obtida a razão de energia para cada um dos níveis, entre o primeiro ciclo após o instante de ocorrência da falta e o ciclo imediatamente anterior a esse mesmo instante. A escolha de um ciclo evita que ocorra a interferência da atuação da proteção localizada ao longo do sistema de distribuição na metodologia utilizada, já que em todos os casos de falta simulados a proteção levaria, no mínimo, mais de um ciclo para atuar, como mostram as figuras 7 e 8. A razão entre essas energias é dada pela expressão (22):

$$Energia_{sub-banda} = \frac{1 + En_{DE}}{1 + En_{AE}} - 1, \quad (22)$$

sendo que  $E_{DE}$  representa a energia do primeiro ciclo durante a ocorrência do evento e  $E_{AE}$  a energia do ciclo antes da ocorrência do evento.

A formulação acima é uma modificação da relação de energia direta  $E_{DE}/E_{AE}$ , de modo que as divisões por zero são evitadas. Foi assumido que a energia é normalizada, de modo que  $E \in [-1, 1]$ . Usando esta técnica, foi possível identificar com precisão em que sub-bandas TWD houve um aumento ou diminuição de energia que caracterizam os eventos. Esta propriedade não foi observada quando se utiliza a relação de energia direta  $E_{DE}/E_{AE}$ , uma vez que essa relação é sensível a ruído. Para a utilização da equação (22), foi adotado como critério de normalização o somatório das energias em todos os níveis como sendo o valor unitário em todo o conjunto de dados.

Com isso, concluem-se os principais aspectos teóricos envolvidos com o pré-processamento, sendo os próximos tópicos dedicados ao segundo estágio da classificação: processamento.

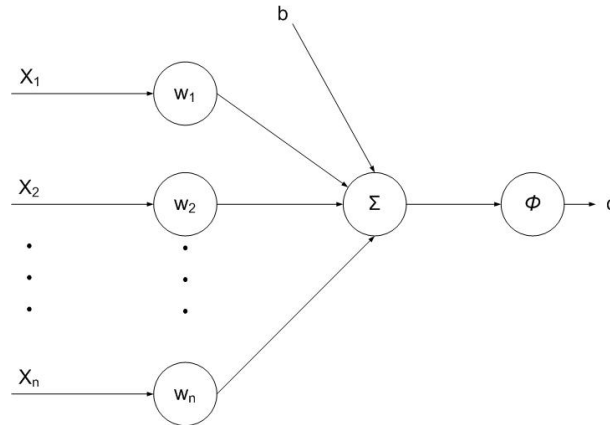
## 4.2 PROCESSAMENTO E REDES NEURAIIS ARTIFICIAIS

Uma vez determinados os vetores de entrada, gerados pelo estágio de pré-processamento, faz-se necessária a apresentação dos principais aspectos teóricos envolvidos com o estágio de processamento, formando a base teórica relacionada com a metodologia proposta.

Como o método de processamento escolhido é baseado em diferentes modelos de redes neurais, cabe aqui a apresentação de alguns conceitos fundamentais relacionados a elas, sendo que grande parte desses conceitos fundamentam a escolha de um processamento baseado em RNA para a classificação em questão.

Uma rede neural artificial, segundo Haykin (2001), consiste em um processador de sinais

paralelamente distribuído, formado por unidades de processamento simples (neurônio), que possuem a capacidade de armazenar conhecimento (pesos sinápticos) utilizando um algoritmo de aprendizado, tornando seu conhecimento disponível para uso futuro. A figura 19 mostra a estrutura típica de um neurônio artificial e suas conexões.



**Figura 19: Estrutura típica de um neurônio.**

**Fonte: (FERREIRA, 2008)**

A saída  $c$  pode ser representada pela equação (23). Nessa equação, a saída é função de uma combinação do conjunto de pesos sinápticos  $\underline{w} \in \mathfrak{R}^n$ ,  $\underline{w} = [w_1 \dots w_n]^t$ , uma camada de entrada  $\underline{x} \in \mathfrak{R}^n$ ,  $\underline{x} = [x_1 \dots x_n]^t$ , um *bias* associado  $b \in \mathfrak{R}$  e uma função  $\varphi(\cdot) : \mathfrak{R} \rightarrow \mathfrak{R}$ , denominada função de ativação.

$$c = \varphi \left( \sum_{i=1}^n w_i x_i + b \right). \quad (23)$$

Usualmente, os neurônios são dispostos em camadas e a interligação entre as camadas define a arquitetura do modelo neural. Existem duas estruturas comumente utilizadas: redes alimentadas adiante (*feedforward*) e redes recorrentes. Em uma rede alimentada adiante, as camadas são conectadas de forma consecutiva e adjacente, sendo que o sinal flui da entrada para a saída em sentido único. Já nas redes recorrentes, existem laços de realimentação, ao longo das camadas de neurônios. Nessa dissertação serão comparadas três diferentes redes neurais alimentadas adiante: Uma rede *Multi-Layer Perceptron*, MLP, uma rede *Radial Basis Function*, RBF, e uma rede do tipo *Support Vector Machine*, SVM, em função da grande maioria dos trabalhos dessa área abordarem redes alimentadas adiante com essas estruturas.

Dentre algumas das vantagens da utilização de redes neurais, pode-se destacar:

- A possibilidade de representação de modelos não lineares.

- O mapeamento entrada-saída, implícito à definição de rede neural, o qual permite que sejam realizados mapeamentos de conjuntos entrada-saída sem a necessidade do desenvolvimento de modelos matemáticos associados aos conjuntos.
- A elevada capacidade de adaptação a mudanças nas condições do ambiente para o qual a rede foi projetada.
- A simplicidade da estrutura que uma rede possui, quando comparada à complexidade dos modelos gerados.

Vale ressaltar que o processo de aprendizagem, em linhas gerais, está associado com a estimação dos parâmetros da rede neural. Para tanto, faz-se necessária a definição de um conjunto de dados (denominado conjunto de treinamento) que será utilizado no processo de aprendizado da rede neural. Em algumas abordagens, o conjunto de treinamento é dividido em dois subconjuntos, o conjunto de estimação e o conjunto de validação. O conjunto de estimação é utilizado no processo de estimação dos parâmetros da rede, sendo que esse processo pode ser interrompido periodicamente, a fim de se avaliar o desempenho da rede através do conjunto de validação, com o intuito de se evitar o ajuste excessivo dos dados (esse assunto será retomado nos próximos tópicos desse trabalho). Além disso, utiliza-se também, um conjunto de teste com a finalidade de se avaliar o desempenho da rede para padrões ainda não vistos, sendo que esse desempenho possui grande importância no desenvolvimento dos modelos apresentados nesse trabalho.

Uma característica bastante importante, que está relacionada com uma das principais características das redes utilizadas nesse trabalho, está relacionada com o teorema da aproximação universal (BISHOP, 1995). Esse teorema afirma que uma rede neural com estrutura alimentada adiante, pode aproximar com precisão arbitrária qualquer função contínua, desde que contenha ao menos uma camada intermediária (também denominada camada oculta) de neurônios com função de ativação contínua  $\varphi$ , não-constante, limitada e uma saída linear, representando a função de aproximação (FERREIRA, 2008) <sup>2</sup>.

Com isso, pode-se dizer que a saída  $y$  de uma rede neural contendo  $m$  neurônios na camada oculta pode ser dada pela equação (24).

$$y = \sum_{j=1}^m w_j \varphi \left( \sum_{i=1}^n w_{ji} x_i + b_j \right) + b. \quad (24)$$

---

<sup>2</sup> Isso fundamenta a escolha das redes desse trabalho (MLP, RBF e SVM) com uma única camada oculta.

Nessa equação,  $w_{ji}$  representa o conjunto de pesos que ligam as entradas ao  $j$ -ésimo neurônio da camada escondida, sendo  $b_j$  o *bias* desse neurônio. Já  $w_j$  representa os pesos que ligam os neurônios da camada oculta ao neurônio linear da saída, sendo  $b$  o *bias* do neurônio de saída.

Uma consequência relevante da aproximação universal, diz respeito ao principal objetivo da construção de um modelo baseado na aprendizagem, que é o desenvolvimento de uma máquina que seja capaz de representar um modelo estatístico do processo gerador dos dados disponíveis e não uma representação exata desse conjunto de dados (BISHOP, 1995). Essa característica implica em um modelo com resultados satisfatórios tanto para o conjunto de dados disponíveis, quanto para o um conjunto de dados a serem apresentados ao modelo. Essa capacidade é está relacionada com o controle de complexidade (ou regularização) do modelo, sendo normalmente denominada capacidade de generalização.

Um modelo com boa capacidade de generalização tende a evitar o ajuste excessivo dos seus parâmetros, na presença de dados ruidosos, ou seja, a generalização evita que a máquina de aprendizagem “decore” padrões existentes nos dados de entrada (nesse caso, o ruído) permitindo que o modelo tenha bom desempenho frente a padrões não utilizados no estágio de aprendizado.

Uma vez que o elevado grau de não-linearidade está relacionado diretamente com os dados de entrada, uma seleção adequada dos parâmetros de entrada de uma rede neural (principalmente a eliminação variáveis irrelevantes para o modelo) pode contribuir com uma melhor capacidade de generalização. Além disso, com o pré-processamento utilizado nesse trabalho, é possível se determinar, através da seleção de entradas, quais as bandas (ou raias no caso da TRF) de frequências são mais relevantes para o classificador, auxiliando em última instância, na análise dos transitórios envolvidos com os eventos escolhidos nesse trabalho.

Portanto, uma análise conjunta do controle de complexidade e seleção de entradas poderá fornecer um método de treinamento bastante eficiente e com grande capacidade de generalização. Essa análise conjunta caracteriza um modelo neural autônomo, ainda pouco utilizado nas várias aplicações de RNAs.

Os aspectos teóricos relacionados com a seleção de entradas e com a regularização utilizados serão apresentados ao longo de cada um dos três modelos neurais avaliados nesse trabalho. Maiores detalhes sobre esses métodos podem ser vistos em (BISHOP, 1995) e (FERREIRA, 2008). Vale ressaltar que não se encontram aplicações para classificação de eventos em SEP utilizando métodos que abordam controle de complexidade e seleção de entradas na literatura.

As aplicações apresentadas aqui para as redes *Perceptron Multi-Camadas* e Máquinas de

Vetor Suporte, encontram fundamento em (BISHOP, 1995) e (VAPNIK; CHAPELLE, 2000), cujas técnicas foram testadas e validadas em (FERREIRA; SILVA, 2007) e (FERREIRA, 2008) para problemas de previsão de carga elétrica. Um detalhe importante é que o principal diferencial deste trabalho está no desenvolvimento de redes neurais com funções de base radial utilizando conceitos de autonomia associados ao processo de treinamento, permitindo a comparação entre esses três métodos. No entanto, a fundamentação teórica e a metodologia serão apresentados na sua sequência usual na literatura, visando facilitar o entendimento dos conceitos de treinamento autônomo.

#### 4.2.1 Redes *Perceptron* Multi-Camadas com Inferência Bayesiana

O primeiro modelo neural avaliado para classificação de eventos em redes de distribuição é baseado em uma arquitetura bastante usual de redes alimentadas adiante: *Multi-Layer Perceptrons*. Redes do tipo MLP têm sido aplicadas com sucesso em diversos problemas nas mais diversas áreas. Redes MLP possuem um algoritmo de treinamento supervisionado muito popular, conhecido como *algoritmo de retropropagação do erro (error back-propagation)* que utiliza como base a aprendizagem por correção de erro.

Nessa dissertação, no entanto, será apresentada uma rede MLP que utiliza um conceito de inferência bayesiana associado<sup>3</sup>, uma vez que o algoritmo de retropropagação pode ser visto como um caso particular da aplicação de técnicas bayesianas no treinamento de redes MLP. Além disso, pode-se destacar outras vantagens na utilização de redes MLP com inferência bayesiana (BISHOP, 1995):

- O controle de complexidade (regularização) apresenta uma interpretação natural dentro dessa abordagem, apresentada ao longo do desenvolvimento teórico deste modelo.
- O método permite a utilização de técnicas que realizam a determinação automática da relevância das entradas, sendo que através dessa técnica é possível se realizar a seleção de entradas, contribuindo com a capacidade de generalização do modelo.
- Através do cálculo da probabilidade *a posteriori* de cada estrutura (conhecido como cálculo da evidência) é possível se efetuar a comparação entre diferentes modelos (principalmente com relação à complexidade), utilizando somente os dados de treinamento.

Com base no método que utiliza o cálculo da probabilidade *a posteriori* de cada estrutura, os três níveis hierárquicos de inferência são explorados: iniciando pela estimação dos parâmetros

---

<sup>3</sup>A aplicação de inferência bayesiana ao desenvolvimento de MLPs foi proposta originalmente por David J.C. Mackay em 1992 (BISHOP, 1995).

(pesos sinápticos e *bias*), seguido pelo cálculo dos hiperparâmetros que surgem ao longo do desenvolvimento teórico do processo de treinamento (sua análise permite o desenvolvimento do método encapsulado de seleção de entradas) e finalizando com a escolha do modelo mais provável à luz dos dados (FERREIRA, 2008). Nos tópicos seguintes serão apresentados alguns aspectos relativos aos três níveis hierárquicos envolvidos no processo de treinamento.

### Treinamento Bayesiano de MLPs

Uma vez definida a estrutura a ser utilizada (número de camadas ocultas, número de neurônios por camada e tipo de função de ativação de cada neurônio) e selecionado o conjunto de pares entrada-saída  $D = \{X, Y\}$ ,  $X = \{x_1, \dots, x_n\}$ ,  $Y = \{d_1, \dots, d_n\}$ , o objetivo do treinamento do modelo sob o ponto de vista da inferência bayesiana consiste na determinação do vetor de parâmetros  $\underline{w}$  que maximize a probabilidade *a posteriori*  $p(\underline{w}|X, Y)$  dada pela equação (25). Nesse ponto, vale ressaltar que no presente trabalho a notação  $\underline{x}$  representa um vetor coluna e  $\underline{A}$  representa uma matriz, tal que:

$$p(\underline{w}|X, Y) = \frac{p(Y|\underline{w}, X)p(\underline{w}|X)}{p(Y|X)}. \quad (25)$$

Para se determinar a probabilidade *a posteriori*  $p(\underline{w}|X, Y)$ , é necessário se determinar a probabilidade *a priori* do vetor de parâmetros  $\underline{w}$ ,  $p(\underline{w}|X)$ , a função de verossimilhança  $p(Y|\underline{w}, X)$ , relacionada com a distribuição de probabilidade de  $x_i$  pertencer a uma dada classe e o fator de normalização  $p(Y|X)$ .

Para determinação da função de verossimilhança  $p(Y|\underline{w}, X)$ , é necessária a abordagem apresentada em (BISHOP, 1995), na qual se busca uma boa representação da função de erro para dados de saída que utilizam representação binária, ou seja, uma forma de se modelar a probabilidade *a posteriori* de um determinado padrão pertencer a uma dada classe, no caso de um problema de duas classes. Com isso, pode-se representar a função de verossimilhança pela equação (26).

$$p(Y|\underline{w}, X) = \prod_{i=1}^N [f(x_i, \underline{w})]^{d_i} [1 - f(x_i, \underline{w})]^{1-d_i}. \quad (26)$$

Usualmente, procura-se representar a função de verossimilhança (equação 27) pelo seu logaritmo negativo, resultando na função de erro de entropia cruzada, representada pela equação (28).



$$p(Y|\underline{w}, X) = \exp(-G(Y|\underline{w}, X)) \quad (27)$$

$$G(Y|\underline{w}, X) = -\sum_{i=1}^N d_i \ln[f(\underline{x}_i, \underline{w})] + (1 - d_i) \ln[1 - f(\underline{x}_i, \underline{w})]. \quad (28)$$

Para utilização correta da equação (27) é necessário que  $d_i \in [0, 1]$ . Quando se utiliza uma função de ativação como mostra a equação (29)<sup>4</sup>, a função  $f(\underline{x}_i, \underline{w})$ , que representa a saída da rede MLP, pode ser representada pela equação (30).

$$g(a) = \frac{1}{1 + \exp(-a)} \quad (29)$$

$$f(\underline{x}_i, \underline{w}) = \frac{1}{1 + \exp\left\{-\sum_{i=1}^m w_i \phi_i\left[\sum_{j=1}^N w_{ij} x_j + b_i\right]\right\}}. \quad (30)$$

Na equação (30),  $\phi_i(a) : \Re \rightarrow \Re$  representa a função de ativação do  $i$ -ésimo neurônio da camada oculta;  $b_i$  o *bias* deste neurônio;  $m$  o número de neurônios na camada oculta; e  $\underline{w}$  o vetor incluindo todos os parâmetros do modelo, i.e., pesos sinápticos e *bias*.

Já para  $p(\underline{w}|X)$ , o fator fundamental relacionado com sua escolha está associado com a função específica de cada parâmetro ao longo da estrutura da rede. Ou seja, espera-se que grupos diferentes de pesos, por exemplo, pesos ligando uma determinada entrada ao modelo e pesos conectando neurônios da camada oculta à saída, apresentem comportamentos distintos ao longo do processo de estimação. Pode-se definir essa função específica dos parâmetros como um conhecimento prévio acerca do problema, sendo assim, uma escolha baseada na utilização de diferentes distribuições gaussianas, todas com vetor médio nulo e diferentes matrizes de covariância para diferentes grupos de pesos, atende esse objetivo (FERREIRA, 2008). Desta forma, a distribuição  $p(\underline{w}|X)$  pode ser representada pela equação (31).

$$p(\underline{w}|X) = \prod_{i=1}^g p(\underline{w}_i) = \frac{1}{\prod_{i=1}^g \left(\frac{2\pi}{\alpha_i}\right)^{\frac{M_i}{2}}} e^{-\frac{1}{2} \sum_{i=1}^g \alpha_i \|\underline{w}_i\|^2}, \quad (31)$$

Sendo que  $\underline{w}_i$  representa o grupo contendo  $M_i$  parâmetros,  $g$  o número de grupos nos quais os pesos e *bias* são agrupados, com  $\alpha_i$  respondendo pelo hiperparâmetro dado pelo inverso da

<sup>4</sup>Em (BISHOP, 1995) demonstra-se que para saídas que representam uma probabilidade, recomenda-se a utilização dessa função.

variância da distribuição gaussiana com vetor média nulo utilizada para representação *a priori* de  $\underline{w}_i$  e  $g$  relacionado com o número de grupos de parâmetros.

De posse das distribuições  $p(Y|\underline{w}, X)$  e  $p(\underline{w}|X)$ , a probabilidade *a posteriori*  $p(\underline{w}|X, Y)$  é dada pela equação (32).

$$p(\underline{w}|X, Y) = \frac{1}{Z_S} e^{-S(\underline{w})}, \quad (32)$$

Sendo o funcional  $S(\underline{w})$  representado pela equação (33).

$$S(\underline{w}) = - \sum_{i=1}^N d_i \ln f(\underline{x}_i, \underline{w}) + (1 - d_i) \ln [1 - f(\underline{x}_i, \underline{w})] + \frac{1}{2} \sum_{i=1}^g \left( \alpha_i \sum_{j=1}^{M_i} w_{ij}^2 \right), \quad (33)$$

Onde  $Z_S = \int e^{-S(\underline{w})} d\underline{w}$  corresponde a um fator de normalização associado. O funcional  $S(\underline{w})$  contém basicamente duas parcelas. A primeira parcela está relacionada com o risco empírico do processo de treinamento. A definição de risco empírico está associada com o método de otimização desenvolvido para determinação do conjunto de pesos de uma rede neural. Com a minimização deste risco (através de um método baseado na retro-propagação do erro, p. ex.) é possível se estimar o valor ótimo para o conjunto de pesos com base nos dados disponíveis. Sendo assim, pode-se dizer que a parcela relacionada ao risco empírico é responsável pelo ajuste do modelo aos dados de treinamento, não levando em conta um possível ajuste excessivo dos dados na presença de padrões de entrada ruidosos, e não garantindo assim, uma boa generalização. No entanto, a segunda parcela apresentada na equação (33),  $\frac{1}{2} \sum_{i=1}^g \left( \alpha_i \sum_{j=1}^{M_i} w_{ij}^2 \right)$ , procura estabelecer uma contrapartida entre o ajuste aos dados de treinamento e a capacidade de generalização através da inserção de conhecimento prévio proveniente da probabilidade *a priori*  $p(\underline{w}|X)$ .

Essa inserção de conhecimento prévio encontra fundamento na teoria da regularização, cuja principal característica é, através da inserção de um conhecimento prévio, obter suavidade do mapeamento resultante do treinamento do modelo (BISHOP, 1995). Essa suavidade está relacionada com a presença de padrões de treinamento ruidosos. Espera-se que, através de um mapeamento suave, padrões de entrada similares (diferenciados apenas pela presença de ruído) produzam saídas bastantes similares, permitindo assim, a capacidade de generalização para o modelo. A teoria da regularização pode ser representada pela equação (34).

$$\min_{\underline{w}} R[f(\underline{x}, \underline{w})] = E_S[f(\underline{x}, \underline{w})] + \lambda E_C[f(\underline{x}, \underline{w})], \quad (34)$$

Sendo o funcional regularizador representado pela equação (35).

$$E_C[f(\underline{x}, \underline{w})] = \|Pf(\underline{x}, \underline{w})\|^2 \quad (35)$$

Na equação (35),  $P$  é um operador diferencial responsável pela restrição de suavidade,  $\|\cdot\|$  é uma norma definida no espaço de  $Pf(\underline{x}, \underline{w})$  e  $\lambda$  é o parâmetro de regularização, que define o equilíbrio entre o ajuste dos dados de treinamento e o controle de complexidade.

Sendo assim, pode-se afirmar que a parcela em questão está associada com o funcional regularizador  $E_C[f(\underline{x}, \underline{w})]$ . Esse funcional é conhecido como decaimento de pesos, responsável pelo mapeamento suave da teoria da regularização. Nesse caso, o hiperparâmetro  $\underline{\alpha}_i$  pode ser visto como uma estimativa para o parâmetro de regularização da equação (34). Vale ressaltar que atualmente a estimativa do parâmetro de regularização consiste em uma das grandes dificuldades na utilização da teoria da regularização e, frente a isso, o presente método fornece uma forma automática para seleção desse parâmetro, sem a necessidade de um conjunto de validação, uma vez que a estimativa de  $\underline{\alpha}_i$  é um subproduto do processo de treinamento.

Portanto, para maximizar  $p(\underline{w}|X, Y)$ , ou seja, se obter o conjunto ótimo para os parâmetros da rede (e indiretamente o hiperparâmetro  $\underline{\alpha}_i$ ), é necessário minimizar  $S(\underline{w})$  em relação às duas parcelas: minimização do erro de treinamento e aplicação de um regularizador baseado no decaimento de pesos, constituindo assim, um dos estágios fundamentais do processo de treinamento bayesiano de MLPs.

#### Determinação Automática de Relevância

Além do procedimento analítico para estimação de parâmetros e hiperparâmetros do modelo, esse método apresenta uma técnica para seleção automática de atributos (entradas) da rede MLP. Essa técnica está associada com a relevância dos grupos de pesos ligados às entradas do modelo e por isso é conhecido como Determinação Automática de Relevância, DAR.

A DAR consiste em particionar o vetor  $\underline{w}$  em  $n + 3$  grupos, de forma que os  $n$  primeiros grupos representam os pesos que ligam cada uma das entradas à camada oculta (figura 20). Os demais grupos são: os *biases* dos neurônios na camada intermediária, os pesos que ligam a camada oculta à saída e o *bias* do neurônio de saída. Com esse agrupamento é possível se determinar a relevância de cada entrada através da análise dos  $n$  hiperparâmetros  $\alpha_i$ . Observando-se a equação (33) é possível notar que quanto maior o valor de  $\underline{\alpha}_i$ , menor deve se a magnitude de  $\underline{w}_i$ , no contexto da minimização de  $S(\underline{w})$ . Dessa forma, pode-se inferir que quanto maior o valor do hiperparâmetro  $\alpha$ , menor é a contribuição da entrada associada.

No intuito de se estabelecer um limiar para determinação da relevância das entradas, foi

utilizado o método de inserção de variáveis de entrada de natureza aleatória, sem correlação com a saída, como sugere Ferreira (2008). Nesse método é inserida uma variável de prova aleatória (uma nova entrada do modelo) no início do processo de treinamento. A variável de prova  $x_p$  é gerada a partir de uma distribuição uniforme definida no conjunto de ocorrências  $C = \{x \in N | x = 0, 1, \dots, k\}$  com  $k$  possibilidades de ocorrência, dada pela equação (36):

$$p(x_p) = \begin{cases} \frac{1}{k}, & \text{se } x_p \in C \\ 0, & \text{se } x_p \notin C \end{cases} \quad (36)$$

Finalizado o treinamento, utiliza-se do valor estimado  $\underline{\alpha}_0$  dessa variável para se estabelecer o limiar de relevância das demais entradas, ou seja, variáveis de entrada que possuem um valor de  $\underline{\alpha}$  maior que o valor estimado para  $\underline{\alpha}_0$ , podem ser descartadas (consideradas irrelevantes para o modelo), mantendo assim apenas as variáveis de entrada consideradas relevantes para o modelo final.

#### Seleção Bayesiana de Modelos

De posse das estimativas de  $\underline{w}$  e  $\underline{\alpha}_i$ , pode-se utilizar da maximização da evidência para se avaliar diferentes modelos neurais, nesse caso, MLPs com diferentes números de neurônios na camada oculta. A maximização da evidência está associada com o modelo mais provável à luz dos dados  $H_h$ , cujo logaritmo é dado pela expressão (39) (FERREIRA, 2008).

$$\ln p(Y|H_h) = -S(\underline{w}) - \frac{1}{2} \ln |\underline{A}(\underline{w})| + \frac{1}{2} \sum_{i=1}^g M_i \alpha_i + 2 \ln m + \ln m! \quad (38)$$

$$+ \frac{1}{2} \sum_{i=1}^g \ln \left( \frac{2}{\gamma_i} \right) + \frac{1}{2} \ln \left( \frac{2}{N - \gamma} \right)$$

Sendo que  $\gamma_i$  representa o número efetivo de parâmetros estimados para o  $i$ -ésimo grupo de pesos  $\underline{w}_i^* = [w_{i1}^*, \dots, w_{iM_i}^*]^t$  e  $\gamma$  o número efetivo de parâmetros estimados do modelo, dados pela equação (40).

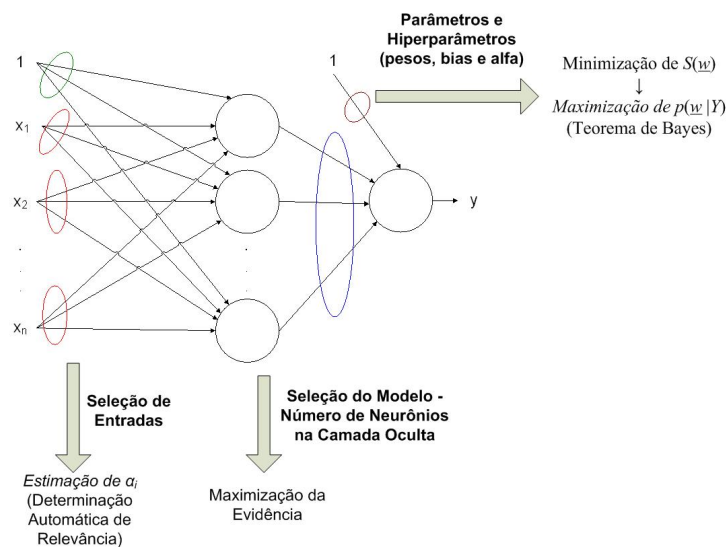
$$\gamma_i = \alpha_i \sum_{j=1}^{M_i} (w_{ij}^*)^2; \gamma = \sum_{i=1}^g \gamma_i \quad (40)$$

Maiores detalhes teóricos sobre a maximização da evidência podem ser encontrados em (BISHOP, 1995), com aplicações em (FERREIRA; SILVA, 2007) e (FERREIRA, 2008).

## Resumo do Treinamento de MLPs com Inferência Bayesiana

A figura 20 mostra a estrutura de uma MLP e os principais tópicos referentes ao treinamento bayesiano. Um detalhe importante é que esta rede contém um total de quinze saídas na sua última camada, sendo que cada saída (neurônio) está associado a uma classe do processo de classificação. Em linhas gerais, o algoritmo pode ser resumido em:

1. Inserção de uma variável de prova de natureza aleatória.
2. Estimação dos parâmetros e hiperparâmetros utilizando a minimização de  $S(\underline{w})$ .
3. Seleção de entradas, com base no valor do hiperparâmetro  $\underline{\alpha}_0$  referente à variável de entrada aleatória.
4. Redefinição dos parâmetros e hiperparâmetros para o modelo com as entradas selecionadas utilizando a minimização de  $S(\underline{w})$ .
5. Definição do modelo com a maior evidência, dado por  $\ln p(Y|H_h)$ .



**Figura 20: Estrutura e Treinamento de uma rede MLP.**

### 4.2.2 Redes Neurais com Função de Base Radial

As redes do tipo *Radial Basis Function* (RBF) objetivam abordar o treinamento através de um problema de ajuste de curva em um espaço de alta dimensionalidade. Nesse sentido, pode-se afirmar que as funções de base radial (presentes na camada oculta de uma RBF) constituem uma base para os padrões de entrada, visando o mapeamento não linear para um espaço de alta

dimensionalidade. A estrutura de uma rede RBF pode ser representada, basicamente, por três camadas fundamentais. A camada de entrada, que efetua a conexão da rede com o ambiente para o qual a rede foi projetada, uma segunda camada (denominada camada oculta), responsável pela transformação não linear do espaço de entrada para o espaço de maior dimensão (espaço da camada oculta) e uma última camada, a camada de saída, que efetua uma operação linear, atribuindo a resposta ao padrão de entrada (HAYKIN, 2001).

O mapeamento para um espaço de maior dimensionalidade encontra fundamento no teorema de Cover, sobre a separabilidade de padrões. Esse teorema afirma que um problema de classificação fortemente não linear tem maior probabilidade de ser linearmente separável em um espaço de elevada dimensionalidade do que em um espaço com dimensionalidade mais baixa (BISHOP, 1995). Isso fundamenta o processo de treinamento das RBFs e possui relevância ao longo dos aspectos teóricos apresentados nos próximos tópicos.

O desenvolvimento das redes RBF encontra fundamento na teoria da regularização, sendo que por vezes, essas redes são denominadas redes de regularização, de acordo com a equação (34).

O principal objetivo do treinamento de uma rede, do ponto de vista da teoria da regularização, é minimizar as duas parcelas do funcional de Tikhonov  $\min_{\underline{w}} R[f(\underline{x}, \underline{w})]$ : parcela do erro de treinamento e do termo de regularização. Uma possível interpretação para essa minimização conjunta, guarda analogia com o dilema *bias*-variância. Nesse dilema, o desempenho da rede pode ser decomposto em duas parcelas conflitantes: *bias* e variância, as quais estão relacionadas com o ajuste aos dados disponíveis e com o nível de flexibilidade da função estimada, respectivamente (FERREIRA, 2008). Modelos excessivamente ajustados aos dados de treinamento irão apresentar *bias* reduzido, porém elevada variância em virtude do grau de complexidade fornecido. No caso da teoria da regularização, uma escolha ótima para o parâmetro de regularização  $\lambda$  pode fornecer um equilíbrio satisfatório entre o *bias* e a variância (mantendo a generalização adequada), através da inserção necessária e suficiente de conhecimento prévio (HAYKIN, 2001).

Uma possível solução para a equação (34), faz uso do conceito de funções de Green<sup>5</sup>  $G(\underline{x}, \underline{x}_i)$ , de acordo com a equação (41). Nessa equação, a solução de (34), passa a ser dada em função da expansão de  $F_\lambda(\underline{x})$ , em termos da função de Green (nesse caso dada como uma função de base radial) e do vetor de pesos  $\underline{w}$  dado pela solução da equação (42).

---

<sup>5</sup>Uma função de Green desempenha o papel de um operador diferencial linear necessário para a diferenciação do funcional de Tikhonov, durante sua minimização.

$$F_\lambda(\underline{x}) = \sum_{i=1}^N w_i G(\|\underline{x} - \underline{x}_i\|) \quad (41)$$

$$\underline{w} = \left( \underline{G} + \lambda \underline{I} \right)^{-1} \underline{d}. \quad (42)$$

Na equação (42),  $\underline{G}$  representa a matriz de Green, responsável por aplicar o operador diferencial às diversas combinações dos  $\underline{x}_i$  (denominados centros da expansão de  $F_\lambda(\underline{x})$ ) e os padrões de entrada  $\underline{x}$ ,  $\underline{I}$  responde pela matriz identidade e  $\underline{d}$  é o vetor de padrões de saída do conjunto de treinamento. No caso particular de  $\lambda = 0$ , obtém-se a solução  $\underline{w} = \underline{G}^+ \underline{d}$ , na qual a matriz  $\underline{G}^+$  é denominada pseudo-inversa, dada pela equação (43).

$$\underline{G}^+ = \left( \underline{G}^T \underline{G} \right)^{-1} \underline{G}^T \quad (43)$$

Uma função de Green normalmente utilizada em aplicações de redes RBF, que atua como operador diferencial linear sujeito às restrições de invariância à translação e rotação (definida positivamente para todo  $i$ ) necessárias para sua utilização prática <sup>6</sup>, é a função gaussiana multivariada dada pela equação (44).

$$G(\underline{x}, \underline{x}_i) = \exp \left( -\frac{1}{2\sigma_i^2} \|\underline{x} - \underline{x}_i\|^2 \right) \quad (44)$$

Com isso, pode-se estabelecer que a arquitetura básica de uma RBF obedece a arquitetura exposta pela figura 21, sendo que a saída da rede assim representada pode ser dada em função das  $m_1$  funções de base radial e pesos lineares  $\underline{w}_j$ , como mostra a equação (45).

$$y = \sum_{i=1}^{m_1} w_i \phi_i(\underline{x}), \quad (45)$$

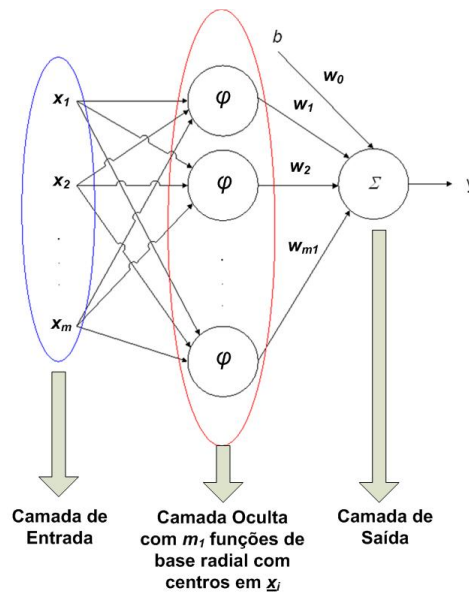
Sendo que a função de base radial  $\phi(\cdot)$  tem como entrada, basicamente, a medida da distância euclidiana entre um padrão de entrada e o centro da função de base  $\underline{t}_i$  (a ser determinado), como demonstra a equação (46).

$$\phi_i(\underline{x}) = G(\|\underline{x} - \underline{t}_i\|) \quad (46)$$

As principais características de uma rede RBF formulada de acordo com a equação (45) podem ser descritas como:

---

<sup>6</sup>Uma função que atende a essas restrições pode ser denominada função de base radial.



**Figura 21: Arquitetura de uma RBF.**

- Aproximadores universais, já que podem representar qualquer função contínua multivariada através de um número suficiente de unidades ocultas.
- A existência da propriedade da melhor aproximação, que implica em afirmar que sempre haverá uma escolha de coeficientes ótima que garante à rede uma aproximação igualmente ótima.

Nos próximos tópicos serão apresentadas as principais técnicas de aprendizagem aplicadas em redes RBF. Entende-se que a aprendizagem da RBF consiste na estimação dos pesos lineares da saída, a otimização da camada oculta (definição da localização dos centros  $t_i$ ) e seleção dos espalhamentos ou larguras  $\sigma_i$  da rede. Serão descritas aqui três abordagens básicas para essa finalidade (Centros Fixos Selecionados ao Acaso, Seleção Auto-Organizada dos Centros e Seleção Supervisionada dos Centros), as quais formam a base da escolha do método de treinamento proposto para redes RBF nesta dissertação.

#### Centros Fixos Selecionados ao Acaso

Nessa abordagem, a localização dos centros é escolhida aleatoriamente utilizando como base o conjunto de treinamento, levando em conta a suposição de que os dados de treinamento estejam distribuídos de forma representativa para o problema considerado (HAYKIN, 2001). Nesse método utilizam-se funções gaussianas isotrópicas<sup>7</sup> como funções de base, cujo desvio padrão é fixo e função da distância máxima entre os centros escolhidos e o número máximo

<sup>7</sup>Neste caso, funções circularmente simétricas.



de unidades ocultas. Com isso, o processo de aprendizagem consiste somente na estimação dos pesos da camada de saída da rede. Para tanto, normalmente se utiliza o cálculo da pseudo-inversa dado pela expressão (43).

Esse método, como mostra (HAYKIN, 2001), não apresenta preocupação com o uso da regularização, uma vez que a escolha dos centros é extremamente dependente dos dados disponíveis para treinamento e pode levar ao ajuste excessivo do modelo, tornando o resultado do método dependente da escolha e disponibilidade de dados de treinamento.

### Seleção Auto-Organizada dos Centros

Esse método, um dos mais utilizados no treinamento de redes RBF, consiste em aplicar um método de treinamento híbrido, composto por dois estágios (HAYKIN, 2001):

1. Estágio de aprendizagem auto-organizada, cujo objetivo é estimar a localização adequada para os centros das funções de base radial.
2. Estágio de aprendizagem supervisionada, responsável por estimar os pesos lineares da camada de saída.

Para o primeiro estágio é necessária a utilização de um método de agrupamento (*clustering*) que particione os dados de entrada em subgrupos de forma homogênea. Para tanto, existem diversos métodos de agrupamento, como k-médias, métodos de agrupamento *fuzzy* e métodos baseados em algoritmos de agrupamento automáticos (HAYKIN; VEEN, 2001). Já para o segundo estágio, pode-se utilizar um método baseado em mínimos quadrados para a estimação dos pesos lineares de saída.

Um problema encontrado nessa abordagem está relacionado com a capacidade de generalização. Segundo Haykin (2001), redes RBF que utilizam esse método de treinamento não obtêm resultados de generalização tão bons quanto redes MLP treinadas apenas com o algoritmo de retropropagação, sendo necessária portanto, a inclusão de um estágio ao processo de treinamento, que possa melhorar a capacidade de generalização da RBF.

### Seleção Supervisionada dos Centros

Nesse último método de aprendizagem, todos os parâmetros livres da rede (pesos, posições e espalhamentos dos centros) sofrem um processo de aprendizagem supervisionado, através de um método de correção de erro via descida de encosta. Para isso, faz-se necessária a definição da

função custo (ou objetivo), normalmente dada pelo erro de treinamento. Com isso, atualizam-se os valores dos parâmetros no método de descida de encosta, ponderando essa atualização através do erro de classificação para dados de treinamento. Através dessa modificação (algoritmo completamente supervisionado), garante-se que uma rede RBF possa generalizar tão bem quanto uma rede MLP treinada com retropropagação do erro (HAYKIN, 2001).

#### Método de Aprendizagem Proposto para RBFs

Tendo em vista os três métodos de treinamento acima detalhados, optou-se nesse trabalho, por adotar um método que utiliza características provenientes dos três processos de aprendizagem, ou seja, o método proposto consiste basicamente em uma Seleção Auto-Organizada dos Centros, porém durante o estágio de aprendizagem supervisionada, são reajustados os centros definidos no primeiro estágio e os pesos lineares. Essa metodologia pode ser melhor compreendida da seguinte forma:

1. No primeiro estágio, ou estágio de aprendizagem auto-organizada, utiliza-se um método de clusterização baseado em uma rede neural do tipo *Grow When Required*, GWR, (MARS-LAND; SHAPIRO; NEHMZOW, 2002), cujo produto é a definição automática do número de centros, bem como suas respectivas localizações.
2. No segundo estágio, ou estágio de aprendizagem supervisionada, a localização dos centros proveniente do primeiro estágio é utilizada como ponto de partida para um método de otimização que utiliza algoritmo genético multi-objetivo, AGMO, (KONAKA; COIT; SMITH, 2006) para a localização dos centros e larguras, além da estimação dos pesos lineares realizada pelo método da pseudo-inversa.

A utilização da pseudo-inversa é justificada pela sua simplicidade e bons resultados, uma vez que pode ser vista como um estimador linear derivado do método de mínimos quadrados (HAYKIN, 2001), no entanto, a utilização desse estimador leva em conta que o parâmetro de regularização é nulo, desconsiderando assim, a parcela associada com a generalização da rede. Para contornar esse fato, uma abordagem baseada na modificação da função custo (ou objetivo) do método de seleção supervisionada dos centros foi adotada, como mostra (DEMUTH; BEALE, 2002). Nessa abordagem, a função que inicialmente levava em conta apenas o erro de treinamento, passa a levar em conta a complexidade da rede (regularização), através da inserção de um termo de controle de complexidade, que guarda analogia com o decaimento de pesos, como mostra a equação (47), do erro de classificação da rede.

$$\varepsilon = \frac{1}{N} \sum_{j=1}^N [d_i - f(x_i)]^2 + \frac{1}{N} \sum_{j=1}^N w_j^2, \quad (47)$$

sendo  $f(x_i)$  a saída da rede. Como comentado anteriormente, o decaimento de pesos está associado com mapeamento suave da teoria da regularização, garantindo assim, que durante o processo de treinamento se mantenha a preocupação com a generalização do modelo. Para tanto, utiliza-se o AGMO, de forma que sejam garantidas as duas minimizações: erro de treinamento e complexidade da rede.

Maiores detalhes sobre o algoritmo de treinamento adotado para as redes RBF nesse trabalho serão apresentadas no último tópico dessa seção. Nos tópicos seguintes serão discutidos aspectos teóricos envolvidos com a metodologia proposta, buscando justificar as escolhas apresentadas.

### Seleção de Entradas

Uma observação importante a ser detalhada inicialmente diz respeito à escolha da função de base. Neste trabalho, uma variação da função gaussiana representada na equação (44) foi selecionada, de forma que possa ser inserido na rede RBF um método de seleção de entradas análogo ao da MLP, como mostra a equação (48).

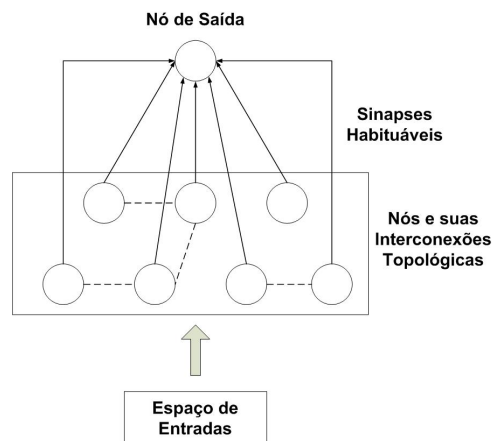
$$G(x_i, x_j) = e^{-\sum_{l=1}^n \frac{1}{2} \left( \frac{\sigma_l x_{il} - \sigma_l x_{jl}}{\sigma_l} \right)^2}. \quad (48)$$

Associa-se o vetor  $\sigma_l$  a cada uma das entradas da rede, permitindo que, através da metodologia baseada na inserção de uma entrada de natureza aleatória descorrelacionada com a saída da rede proposta para redes MLP, seja possível se estabelecer um método para seleção automática de entradas em redes RBF. Como será descrito adiante, a escolha ótima dos parâmetros  $\sigma_l$  poderá, através da comparação dos valores de  $\sigma_l$  com o valor  $\sigma_0$  (determinado para a variável aleatória inserida), determinar quais as variáveis relevantes do modelo, da mesma forma que se tratavam  $\alpha_0$  e  $\underline{\alpha}$  no caso das redes MLP.

Vale ressaltar que as larguras  $\sigma_i$  associadas com os centros das funções de base foram mantidas constantes e unitárias durante a primeira otimização de  $\underline{\sigma}_l$  visando a utilização dessa variável na seleção de entradas (ver último tópico dessa seção). Já na avaliação do modelo final (com as estradas já selecionadas), são otimizadas as duas variáveis  $\sigma_i$  e  $\underline{\sigma}_l$ , otimizando a escolha dos parâmetros da rede.

## Rede Neural *Grow When Required*

As redes do tipo *Grow When Required* (GWR) foram inicialmente propostas por Marsland, Shapiro e Nehmzow (2002). Seus princípios encontram fundamento nos mapas auto-organizáveis de Kohonen (HAYKIN, 2001). Nessa topologia, a rede é composta por neurônios que representam os centros dos agrupamentos (*clusters*) do espaço de entrada, sendo que a sua localização (representada pelos vetores de pesos da rede) é um indicador das características intrínsecas das entradas, uma vez que a apresentação de uma nova entrada irá provocar diferentes níveis de ativação para cada neurônio da rede. A figura 25 mostra a estrutura típica de uma GWR. Nesse estrutura estão representados os três principais componentes da rede GWR: camada de entrada, camada de agrupamento (neurônios) e a camada de saída (composta por um único neurônio). A conexão entre os neurônios da camada de agrupamento e a saída é feita pelas chamadas sinapses habituáveis, as quais estão sujeitas a um modelo de habituação cuja principal função é redução da eficiência sináptica de um determinado neurônio, em função de entradas repetidamente apresentadas (VIEIRA-NETO, 2006).



**Figura 22: Estrutura de uma rede GWR.**

O processo de aprendizagem da rede se dá pelo método do vencedor-leva-tudo (do inglês *winner-take-all*), ou seja, apenas um neurônio é ativado na camada intermediária da rede, inibindo a ativação dos demais, porém utilizando a informação topológica da ativação para adaptar os neurônios vizinhos ao vencedor. Além disso, a rede GWR possui uma característica muito particular, que está associada com sua capacidade de inserir neurônios através de seu modelo de habituação e a ativação do neurônio vencedor, o que fundamenta parte de sua escolha como método de clusterização para redes RBF. Esse modelo de habituação, principal característica da rede GWR, é dado por uma equação diferencial de primeira ordem, que controla a taxa com que o modelo se habitua a padrões repetidamente apresentados, bem como a taxa com que o padrão volta a ser considerado como novidade, a medida que deixa de ser apresentado a rede.

O algoritmo da rede GWR pode ser resumido pela sequência representada a seguir, sendo que maiores detalhes da implementação dessa rede podem ser encontrados em (MARSLAND; SHAPIRO; NEHMZOW, 2002; VIEIRA-NETO, 2006):

1. Inicialize a rede com dois neurônios escolhidos aleatoriamente do espaço de entrada.
2. Para cada neurônio da rede, calcule a distância euclidiana do vetor de entradas.
3. Selecione os dois melhores neurônios,  $s$  e  $t$ , com base no cálculo da distância euclidiana acima mencionada.
4. Crie uma conexão entre os dois neurônios  $s$  e  $t$ , caso ela já exista atribua para essa conexão a idade zero, a qual será utilizada para definir se essa conexão deve ou não ser desfeita futuramente.
5. Se os critérios de inserção de neurônio forem atendidos (critério de habituação e ativação do neurônio) deve-se inserir um novo neurônio  $r$ , cujo vetor de pesos é função do vetor de pesos de  $s$  e do vetor de entradas. Além disso deve-se remover a ligação entre  $s$  e  $t$  e criar uma ligação entre  $r$  e  $t$  e também entre  $s$  e  $r$ .
6. Se nenhum neurônio foi inserido, atualize as posições (pesos) do neurônio vencedor e dos que estão conectados a ele (neurônios vizinhos) de acordo com a taxa de aprendizagem  $\varepsilon$  previamente ajustada.
7. Incremente em um, a idade dos neurônios com conexão final em  $s$ .
8. Reduzir o contador de frequência de ativação do neurônio  $s$ , de acordo com a taxa com que o modelo se habitua a padrões repetidamente apresentados e a taxa com que o padrão volta a ser considerado como novidade, de acordo com seus valores iniciais.
9. Verificar se existe alguma conexão ou neurônio a ser excluído, ou seja, neurônio sem nenhuma conexão ou conexão com idade máxima atingida.
10. Se houverem mais entradas a serem avaliadas, retorne ao início, desde que nenhum critério de parada seja atingido.

A rede GWR utilizada nesse trabalho tem por base a implementação sugerida por (MARSLAND; SHAPIRO; NEHMZOW, 2002), sendo que esse autor disponibiliza um pacote de *scripts* em Matlab para sua aplicação.

## Algoritmos Genéticos

Um algoritmo genético, AG, possui como principal característica a inspiração em processos biológicos baseados em teorias evolutivas, em particular a Teoria da Seleção Natural. Este tipo de algoritmo é voltado para a solução de problemas de otimização através da manipulação de uma população de indivíduos, visando preservar os indivíduos com as melhores características genéticas dessa população, sendo que do ponto de vista de otimização, esses indivíduos representam as regiões mais promissoras do espaço de busca (GOLDBERG, 1989). No intuito de se apresentar os principais operadores e o funcionamento geral de um AG, faz-se necessária a apresentação de uma terminologia particular desse método, a saber:

- Cromossomo - Representa a estrutura de dados que codifica o problema. A codificação é um estágio importante de um AG pois é através dela que se garantirá uma boa representação do problema e, conseqüentemente, a convergência para uma solução satisfatória para o problema através dos operadores genéticos. Em geral, a codificação das variáveis do problema pode ser real ou binária.
- Genótipo - Informação existente na solução.
- Fenótipo - É a decodificação de um indivíduo, representado por um genótipo.
- Indivíduo - Uma possível solução do espaço de busca.
- População - Conjunto de possíveis soluções (indivíduos) no espaço de busca.
- Geração - Uma iteração do AG, que pode resultar em uma nova população.

A ideia básica de um AG é gerar uma população inicial formada por uma população aleatória com distribuição uniforme de indivíduos, sendo que durante o processo evolutivo, cada indivíduo é avaliado por um parâmetro de aptidão (denominado *fitness*), o qual reflete a sua capacidade de resolver o problema. Esse parâmetro de aptidão está associado com a função objetivo do problema, ou seja, a função que se busca otimizar. A função objetivo é dependente do problema em particular, sendo que essa deve ser formulada *a priori*, levando em conta o conjunto de variáveis da otimização (nesse caso, representadas pelos indivíduos da população). Depois de definida a aptidão de toda a população, a parcela dos indivíduos considerados mais aptos é mantida, enquanto que os indivíduos restantes são eliminados através do método de seleção. Os indivíduos mantidos pela seleção podem ter suas características modificadas por operadores genéticos, gerando descendentes para a próxima geração. Este processo é repetido

até que um conjunto de soluções satisfatórias seja encontrado (GOLDBERG, 1989). Dessa forma pode-se dizer que um AG convencional responde pela seguinte sequência de operações:

1. Inicie os parâmetros de controle do AG.
2. Crie uma população inicial de forma aleatória com distribuição uniforme.
3. Avalie do *fitness* dos indivíduos dessa população.
4. Caso o número de gerações seja alcançado ou outro critério de parada seja atingido, o AG finaliza sua busca. Caso contrário, aplique operadores genéticos visando obter melhores soluções e retorne ao passo anterior.

No intuito de garantir a convergência do algoritmo, faz-se uso dos operadores genéticos. Existem, basicamente, três operadores genéticos utilizados: seleção, cruzamento e mutação (GOLDBERG, 1989). A seleção é responsável por escolher os indivíduos mais aptos que seguirão no processo iterativo do AG, sendo que existem diversos métodos disponíveis na literatura para tal finalidade. O operador de cruzamento realiza a recombinação genética entre dois indivíduos da população, gerando novos indivíduos que herdam as características dos indivíduos geradores (com boa aptidão), mantendo assim o processo evolutivo do algoritmo. Por se tratar de um operador dominante, a taxa de cruzamento é normalmente escolhida entre 70% e 100% da população. O último operador, operador de mutação, é responsável por introduzir diversidade genética na população através da alteração de um ou mais componentes de indivíduos selecionados. Normalmente utilizam-se taxas de mutação entre 0,1% e 10% (GOLDBERG, 1989).

O AG convencional não foi aplicado diretamente nas redes RBF desse trabalho. No entanto, a sua formulação é importante na definição do AGMO, que foi utilizado diretamente no treinamento das redes RBF. A abordagem convencional do AG será retomada adiante, na seção referente às redes do tipo *Support Vector Machine*.

### Algoritmos Genéticos Multi-Objetivo

Tendo em vista o caráter multi-funcional apresentado na equação (47), faz-se necessária a introdução do conceito de otimização multi-objetivo, dado um processo de decisão que envolve  $K$  objetivos, dos quais o responsável pela decisão não possui uma preferência bem definida, dentre os diversos objetivos. Em casos reais, os objetivos são normalmente conflitantes entre si, uma vez que uma otimização voltada para apenas um dos objetivos não fornece uma solução

adequada aos demais e, em função disso, uma solução multi-objetivo perfeita para todos os objetivos se torna inviável. Uma solução razoável para um problema multi-objetivo consiste em investigar um conjunto de soluções, de forma que cada solução atendo os objetivos em um nível considerado aceitável, sem ser dominado por qualquer outra solução (KONAKA; COIT; SMITH, 2006). Um problema multi-objetivo pode ser formulado<sup>8</sup> pela equação (49) (SIERRA; COELLO, 2006).

$$\text{minimizar } f(\underline{x}) := [f_1(\underline{x}), f_2(\underline{x}), \dots, f_k(\underline{x})], \quad (49)$$

Sujeita às restrições impostas pelas equações (50) e (51).

$$\underline{g}_i(\underline{x}) \leq 0, \text{ para } i = 1, 2, \dots, m, \quad (50)$$

$$\underline{h}_i(\underline{x}) = 0, \text{ para } i = 1, 2, \dots, p, \quad (51)$$

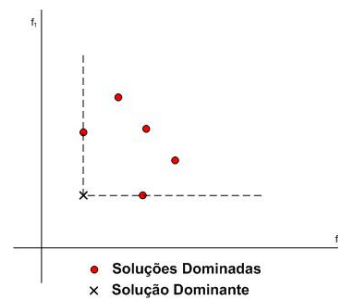
Sendo que  $\underline{x} = [x_1, x_2, \dots, x_n]^t$  representa o vetor de variáveis de decisão,  $f_i$ , para  $i = 1, \dots, k$  são as funções objetivos do problema e  $i = 1, \dots, m$ ,  $j = 1, \dots, p$  são as funções de restrição do problema. Com isso, pode-se definir algumas das nomenclaturas mais utilizadas em otimização multi-objetivo (SIERRA; COELLO, 2006):

- Uma solução  $\underline{x}$  é dita dominante sobre uma solução  $\underline{y}$ , se  $f_i(\underline{x}) \leq f_i(\underline{y})$  e  $\underline{x} \neq \underline{y}$  para todo  $\underline{x}, \underline{y}$  e  $i$ . A figura 23 demonstra essa relação de dominância.
- Uma solução  $\underline{x}$  é dita não-nominada no conjunto de soluções, se não existe uma outra solução  $\underline{x}'$  tal que  $f(\underline{x}) < f(\underline{x}')$ .
- Um conjunto de soluções  $\underline{x}^*$  é chamada de solução ótima de Pareto, se ela for formada por soluções não-dominadas no espaço de possíveis soluções  $\Upsilon$ .
- O conjunto ótimo de Pareto  $P^*$  pode ser definido como  $P^* = \{\underline{x} \in \Upsilon\}$ , de forma que  $\underline{x}$  corresponde à solução ótima de Pareto.
- Uma fronteira de Pareto pode ser definida como  $PF^* = \{f(\underline{x}) \in \mathfrak{R}^k | \underline{x} \in P^*\}$ , como mostra a figura 24, para um problema multi-objetivo com duas funções objetivo.

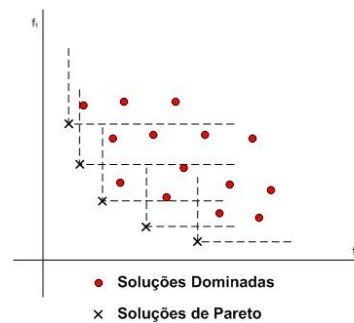
Sendo assim, um problema de otimização multi-objetivo pode ser dado através da determinação do conjunto de Pareto, visto através da Fronteira de Pareto, respeitando as restrições dadas

<sup>8</sup>Considerando, por conveniência, problemas de minimização.





**Figura 23: Relação de Dominância entre duas Soluções.**



**Figura 24: Soluções na Fronteira de Pareto.**

pelas equações (50) e (51). O método escolhido nesse trabalho para efetuar a busca pela solução de Pareto está baseada em algoritmos genéticos multi-objetivo (AGMO). Essa escolha é fundamentada na capacidade de um AG pesquisar simultaneamente diferentes regiões de um espaço de soluções, tornando possível encontrar um conjunto diversificado de soluções para problemas difíceis com características não-convexas, descontínuas, e multi-modais no espaços de soluções (KONAKA; COIT; SMITH, 2006). O operador de cruzamento de um AG, p. ex., pode fornecer uma boa exploração de novas soluções associando objetivos diferentes, visando a obtenção de soluções não-dominadas em partes inexploradas da fronteira de Pareto.

Muitas das aplicações multi-objetivo utilizam métodos baseados em AG, como mostra (KONAKA; COIT; SMITH, 2006). Para tanto, foram desenvolvidas diversas variações do primeiro algoritmo genético com propriedades multi-objetivo denominado de *Vector Evaluated Genetic Algorithm*. Nesse trabalho, optou-se pela utilização do algoritmo baseado no conceito de dominância<sup>9</sup> denominado *Fast Non-Dominated Sorting Genetic Algorithm (NSGA-II)* desenvolvido por Deb et al. (2002), cujas principais vantagens são a eficiência já comprovada por diversos testes realizados e a quantidade de parâmetros a serem especificados é bastante reduzida.

<sup>9</sup>As soluções no primeiro *front*, representado pelas soluções não-dominadas, são melhores que as soluções nos demais *fronts* (demais soluções).

O funcionamento do algoritmo NSGA-II é fundamentado em dois mecanismos relacionados com o processo de seleção: *Fast Non-Dominated Sorting* e a Distância de Crowding. O funcionamento do algoritmo pode ser detalhado como:

1. Defina a População Inicial Aleatoriamente.
2. Aplique a relação de dominância do método *Fast Non-Dominated Sorting* cuja fundamentação foi descrita no início desse tópico, para toda a população, separando as possíveis soluções em *fronts*, de forma que no primeiro *front* estão as soluções não-dominadas, sendo que indivíduos do *front N* são soluções melhores (do ponto de vista de otimização multi-objetivo) que os indivíduos do *front N + 1*.
3. Aplique o operador de diversidade baseado na distância de *Crowding*, ordenando cada indivíduo da população de acordo com sua distância em relação aos pontos vizinhos e garantindo que haja um melhor espalhamento das soluções na fronteira de Pareto.
4. Utilize um método de seleção *Seleção por Torneio Binário*<sup>10</sup> com base no cálculo da distância de *Crowding* para a escolha dos indivíduos pais, os quais estarão sujeitos a cruzamento e mutação para geração da próxima população.
5. Repita o processo até que algum critério de parada seja atingido (p. ex., número máximo de gerações), selecionando o indivíduo com melhor compromisso entre as duas funções objetivo (nesse caso em específico).

Maiores detalhes sobre o método *Fast Non-Dominated Sorting* e a Distância de Crowding podem ser encontrados em (DEB et al., 2002).

#### Resumo do Treinamento de RBFs

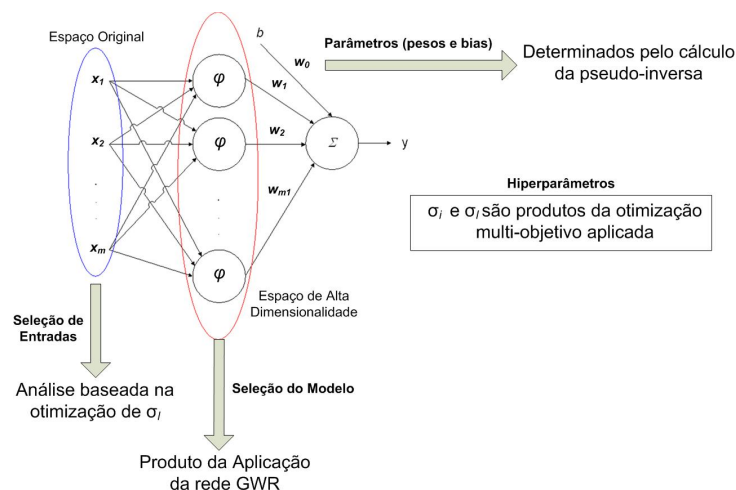
A figura 25 mostra a estrutura de uma RBF e os principais tópicos referentes ao treinamento adotado. Um detalhe importante é que esta rede contém um total de quinze saídas na sua última camada, sendo que cada saída (neurônio) está associado a uma classe do processo de classificação. As principais etapas do algoritmo podem ser resumidas em:

1. Inserção de uma variável de entrada de natureza aleatória.

---

<sup>10</sup>Nesse método dois indivíduos são selecionados aleatoriamente e seu fitness é comparado. O indivíduo com melhor fitness é selecionado como um dos pais utilizados no cruzamento.

2. Estimação do número de centros, bem como suas respectivas localizações através da rede GWR.
3. Otimização via AGMO da localização dos centros da rede GWR e do parâmetro  $\underline{\sigma}_l$ , mantendo os valores  $\sigma_i$  constantes e unitários. Nessa etapa os pesos lineares são calculados utilizando a pseudo-inversa.
4. Seleção de entradas, com base no valor do parâmetro  $\underline{\sigma}_0$  determinado pela otimização anterior.
5. Estimação do número de centros, bem como suas respectivas localizações para o novo modelo, através da rede GWR.
6. Otimização via AGMO da localização dos centros da rede e dos parâmetros  $\underline{\sigma}_l$  e  $\sigma_i$  do modelo final, mantendo o cálculo dos pesos lineares pelo método da pseudo-inversa.



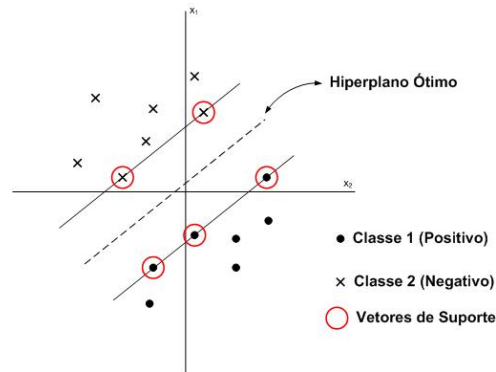
**Figura 25: Estrutura e Treinamento de uma RBF.**

#### 4.2.3 Support Vector Machines

Uma rede do tipo *Support Vector Machine* (SVM), inicialmente desenvolvida para solução de problemas de classificação desenvolvida por V. Vapnick (VAPNIK, 1998) consiste basicamente em um classificador, cuja característica fundamental está associada ao conceito de hiperplano ótimo, baseado na maximização da margem de separação  $\rho$ . O objetivo da construção da margem de separação (superfície de decisão) é encontrar um hiperplano que obtenha a máxima separação entre as classes do problema<sup>11</sup> visando a solução de problemas cuja quantidade de

<sup>11</sup>O conceito de SVMs foi desenvolvido para problemas de duas classes, mais comumente mencionados como separação entre exemplos positivos e negativos.

dados disponíveis é reduzida e pouco, ou até mesmo nenhum conhecimento prévio pode ser utilizado. O conceito de hiperplano ótimo de separação pode ser melhor compreendido, através da análise da figura 26, na qual estão representados padrões linearmente separáveis, divididos pelo hiperplano ótimo de separação.



**Figura 26: Hiperplano Ótimo de Separação.**

Na figura 26 está representado um conceito fundamental na determinação do hiperplano ótimo: vetores de suporte. Os vetores de suporte são os pontos situados mais próximos da superfície de decisão, sendo que sua localização dá suporte à construção da margem de separação e, conseqüentemente, a definição do hiperplano ótimo de separação. Esse conceito será retomado mais adiante, na formulação matemática de uma SVM.

As redes SVM foram desenvolvidas com base em um novo paradigma da área de aprendizado de máquina, conhecido como aprendizado estatístico (FERREIRA, 2008). Dentro desse conceito, pode-se dizer que uma rede SVM é uma implementação do método de minimização estrutural do risco. Esse método mostra que o erro de generalização de um modelo de aprendizagem é limitado por duas parcelas, uma referente ao erro para o conjunto de treinamento (risco empírico) e outro termo associado com uma medida de complexidade conhecida como dimensão de Vapnik e Chervonenkis, popularmente denominada dimensão VC.

O conceito da dimensão VC guarda analogia com o dilema *bias*-variância, cujo conceito está associado ao desempenho do modelo para novos dados (conjunto de teste). Da mesma forma que naquele dilema, pode-se afirmar que modelos com elevada dimensão VC, apesar de ajustarem de forma satisfatória os dados de treinamento, apresentarão reduzida capacidade de generalização (FERREIRA, 2008). No caso de uma SVM, durante o processo de treinamento o termo associado com o risco empírico é nulo e o termo associado à dimensão VC é minimizado, garantindo assim boa capacidade de generalização para padrões linearmente separáveis.

Apesar de não possuir expressão analítica geral para modelos não-lineares, a dimensão VC de hiperplanos  $F(\underline{x}_j, \underline{w})$  com margem de separação  $\rho$  é limitada superiormente pela equação

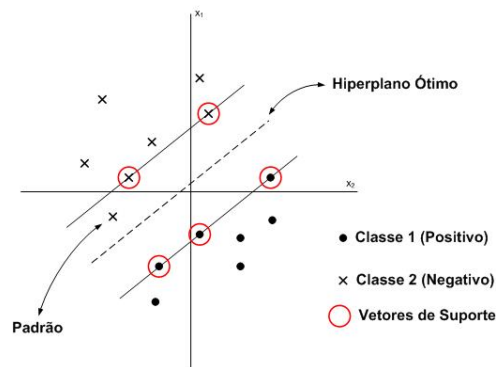
(52) (HAYKIN; VEEN, 2001).

$$VC[F(\underline{x}, \underline{W})] \leq \frac{R^2}{\rho^2}, \quad (52)$$

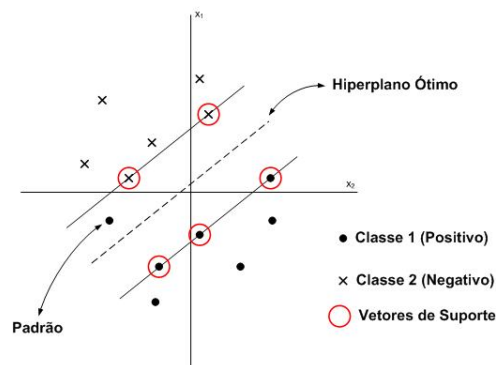
sendo que  $R$  corresponde ao raio da menor hipersfera que engloba a imagem dos padrões de treinamento no espaço de características<sup>12</sup>.

### Aplicação em Padrões não Linearmente Separáveis

O conceito de hiperplano ótimo de separação pode ser expandido para problemas de classificação de padrões não-linearmente separáveis. Nesse caso não é possível construir um hiperplano ótimo de separação sem a existência de erros de classificação. Essa impossibilidade pode se manifestar de duas formas (HAYKIN, 2001), como ilustram as figuras 27 e 28.



**Figura 27: Erro de Classificação (Caso 1).**



**Figura 28: Erro de Classificação (Caso 2).**

Na figura 27 o padrão responsável pelo erro se encontra dentro da região de separação, porém do lado correto da superfície de decisão, já na figura 28 o padrão se encontra no lado

<sup>12</sup>Espaço associado com o mapeamento realizado durante o treinamento de SVMs.

errado da superfície de decisão e no interior da região de separação, configurando uma classificação incorreta. Para contornar esse problema, utiliza-se um mapeamento do espaço original de representação para um espaço de dimensão elevada (espaço de características), aonde a probabilidade do problema ser linearmente separável é elevada. Dessa forma, as SVMs podem ser vistas como classificadores lineares aplicados a um espaço de representação expandido, de dimensão maior que o espaço de representação original do problema, com o mapeamento obtido de forma intrínseca ao processo de treinamento (FERREIRA, 2008). Matematicamente, a saída de uma SVM pode ser dada pela equação (53), em função do *bias*  $b$  e do conjunto de parâmetros ( $\underline{W}$ ) do modelo representado pela equação (54), além do mapeamento não-linear  $\underline{\varphi}(\underline{x}) : \mathcal{R}^n \rightarrow \mathcal{R}^N$  das entradas no espaço de características mostrado na equação (55).

$$f(\underline{x}, \underline{W}, b) = \text{sign} \left[ \underline{W}^t \underline{\varphi}(\underline{x}) + b \right] \quad (53)$$

$$\underline{W} = \left[ W_1 \quad W_2 \quad \dots \quad W_N \right]^t \quad (54)$$

$$\underline{\varphi}(\underline{x}) = \left[ \varphi_1(\underline{x}) \quad \varphi_2(\underline{x}) \quad \dots \quad \varphi_N(\underline{x}) \right]^t \quad (55)$$

Na equação (53),  $\text{sign}[a]$  representa a função sinal (equação 56).

$$\begin{cases} \text{sign}[a] = 1, \text{ para } a \geq 0 \\ \text{sign}[a] = -1, a < 0 \end{cases} \quad (56)$$

Uma vez realizado o mapeamento não-linear para um espaço de maior dimensionalidade, faz-se necessária a construção do hiperplano ótimo nesse novo espaço, configurando assim, os dois principais estágios responsáveis pelo treinamento de uma rede SVM. Para tanto, deve-se considerar primeiramente o conceito de variáveis escalares não negativas  $\xi_i|_{i=1}^N$ , que formam a base para essa construção. No caso de padrões linearmente separáveis, a construção do hiperplano ótimo de separação está sujeita à restrição imposta pela equação (57), cujo objetivo é determinar o valor ótimo para os pesos  $\underline{w}$  e *bias*  $b$  dados os padrões  $\underline{x}_i$ , levando em conta que o vetor de pesos deve minimizar a equação (58).

$$d_i(\underline{w}\underline{x}_i + b) \geq 1 \text{ para } i = 1, 2, \dots, N \quad (57)$$

$$\Phi(\underline{w}) = \frac{1}{2} \underline{w}^T \underline{w} \quad (58)$$

Já para o caso não-linearmente separável, faz-se necessária a inclusão das variáveis escalares não negativas, denominadas variáveis soltas, cujo objetivo é mensurar o desvio de um determinado padrão da condição ideal de separabilidade de padrões, como mostra a equação (59), relacionada à restrição da construção do hiperplano de separação.

$$d_i \left[ \underline{W}^t \underline{\varphi}(x) + b \right] \geq 1 - \xi_i, \text{ para } i = 1, 2, \dots, N \quad (59)$$

Para  $0 \leq \xi_i \leq 1$ , o determinado padrão se encontra dentro da região de separação, porém do lado correto do hiperplano ótimo (figura 27). Já para  $\xi_i > 1$ , o determinado padrão está do lado incorreto do hiperplano (figura 28)<sup>13</sup>. Dessa forma, o problema passa a ser encontrar um hiperplano ótimo para o qual o erro de classificação é minimizado. Como mostra Haykin (2001), um modo de se obter esse hiperplano é através do problema de otimização restrita exposto na equação (60).

$$\Phi(\underline{w}, \xi) = \frac{1}{2} \underline{W}^t \underline{W} + C \sum_{i=1}^N \xi_i \quad (60)$$

Sujeito a restrição imposta pela equação (61).

$$\begin{cases} d_i \left[ \underline{W}^t \underline{\varphi}(x) + b \right] \geq 1 - \xi_i \\ \xi_i \geq 0 \end{cases}, i = 1, 2, \dots, N \quad (61)$$

Na equação (60), o primeiro termo da função objetivo é responsável pelo controle de complexidade do modelo por meio da maximização da margem de separação  $\rho$ . Esse termo está associado à dimensão VC, de forma que sua minimização garantirá ao modelo boa capacidade de generalização. No segundo termo, pode-se afirmar que  $\sum_{i=1}^N \xi_i$  funciona como um limite superior para o erro do conjunto de teste, visto que para dados de teste corretamente classificados,  $\xi_i$  é igual a zero (caso linearmente separável). Nesse termo, o hiperparâmetro  $C$  é responsável pelo equilíbrio entre a complexidade do modelo e o ajuste dos dados de treinamento, sendo desta forma denominado parâmetro de regularização (HAYKIN, 2001).

O problema de otimização formulado na equação (60) pode ser resolvido utilizando o

<sup>13</sup>Nesse ponto vale uma primeira definição matemática para vetores de suporte, ou seja, são os vetores que satisfazem a equação (59), mesmo se  $\xi_i > 0$

método dos multiplicadores de Lagrange, resultando na formulação apresentada na equação (62) e sujeita às restrições impostas pela equação (63).

$$\max_{\underline{\alpha}} \Psi(\underline{\alpha}) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N d_i d_j K(\underline{x}_i, \underline{x}_j) \alpha_i \alpha_j \quad (62)$$

$$\begin{aligned} 0 &\leq \alpha_i \leq C, i = 1, 2, \dots, N \\ \sum_{i=1}^N \alpha_i d_i &= 0, \end{aligned} \quad (63)$$

sendo que,  $\underline{\alpha}$  representa o conjunto de multiplicadores de Lagrange,  $K(\underline{x}_i, \underline{x}_j)$  é conhecido como o núcleo (*kernel*) do produto interno no espaço de características e  $d$  a saída desejada da rede. Como citado anteriormente, o primeiro estágio do processo de treinamento de uma SVM está relacionado com o mapeamento do espaço de entrada para um espaço de maior dimensionalidade, de forma que um problema não-linearmente separável no espaço original se torne separável nesse novo espaço, de alta dimensionalidade. Esse mapeamento é análogo ao efetuado nas redes RBF e está fundamentado no teorema de Cover. Segundo esse teorema, para que o problema passe a ser linearmente separável nesse novo espaço, são necessárias duas condições. Primeiro que o mapeamento (transformação) deve ser não linear e segundo, que a dimensionalidade do novo espaço deve ser suficientemente alta para que a probabilidade do problema se tornar linear seja igualmente alta.

No entanto, esse mapeamento não aborda a questão relacionada com a construção do hiperplano ótimo, fazendo com que seja necessário um segundo estágio de treinamento relacionado com a construção do hiperplano e garantindo assim, a capacidade de generalização da rede. A questão fundamental relacionada com esse estágio está associada com o fato de que, uma vez realizado o mapeamento, os vetores utilizados na construção do hiperplano pertencem ao novo espaço (espaço de características), tornando necessária a inclusão do conceito do núcleo do produto interno. Vale ressaltar que essa construção está baseada na minimização estrutural do risco, garantindo assim, a minimização da dimensão VC e, conseqüentemente, uma boa capacidade de generalização (HAYKIN, 2001).

O núcleo do produto interno pode ser definido, como mostra a equação (64), com base no transformação não-linear representada por  $\varphi(\cdot)$ .

$$K(\underline{x}_i, \underline{x}_j) = [\varphi(\underline{x}_i)]^t \varphi(\underline{x}_j) \quad (64)$$

Através da utilização da equação (64), é possível se estabelecer a construção do hiperplano



ótimo de separação no espaço de características, sem ter que se considerar esse espaço de forma explícita, o que é de grande valia para o estabelecimento da margem de separação e projeto ótimo do hiperplano. Existem diversos tipos de kernel  $K(\underline{x}_i, \underline{x}_j)$  que podem ser utilizados na construção de uma rede SVM. A escolha do kernel está fundamentada no teorema de Mercer, cuja fundamentação está relacionada com o fato de que se um *kernel* candidato é realmente um núcleo de produto interno em algum espaço, e portanto admissível para o uso em SVMs. Neste trabalho, será utilizado o kernel gaussiano modificado (que atende as restrições do teorema de Mercer) para que seja possível a utilização de um método de seleção de entradas análogo ao da RBF, como sugere (FERREIRA, 2008). O *kernel* utilizado para a SVM é o mesmo adotado na RBF, representado pela equação (65).

$$K(\underline{x}_i, \underline{x}_j) = e^{-\sum_{l=1}^n \sigma_l^2 (x_{il} - x_{jl})^2} \quad (65)$$

Sendo que  $\sigma_l^2, l = 1, 2, \dots, n$  representam os hiperparâmetros do kernel. O método de seleção de entradas adotado para a SVM é o mesmo utilizado nas redes RBF apresentadas no tópico relacionado com seleção de entradas para a RBF. Assim como nas redes RBF, a escolha do kernel está associada fundamentalmente à possibilidade de se aplicar um método automático para seleção de entradas.

Um último detalhe a ser ressaltado é que, no ponto ótimo da equação (62) nem todos  $\alpha_i^*$  são não-nulos. Os vetores para os quais  $\alpha_i^*$  é diferente de zero são chamados vetores de suporte, os quais definem a superfície de decisão da SVM pela expressão (66), formando assim a definição geral para vetores de suporte.

$$f(\underline{x}, \underline{W}, b) = \text{sign} \left[ \sum_{i=1}^{NS} \alpha_i d_i K(\underline{x}_i, \underline{x}) + b \right], \quad (66)$$

sendo que NS representa o número de vetores suporte, na equação (66).

### Seleção Automática dos Hiperparâmetros

O último estágio do processo que confere à rede SVM sua característica autônoma, está relacionado com a determinação automática dos hiperparâmetros: constante de regularização  $C$  e hiperparâmetros do kernel  $\sigma_l^2$ . Para tanto, optou-se por utilizar nesse trabalho, uma técnica proposta em (FERREIRA, 2008), cuja abordagem está baseada na minimização de um limite superior do erro de generalização estimado via validação cruzada única (*leave-one-out*). O procedimento de validação cruzada única é um método de reamostragem, cuja estimativa para

a capacidade de generalização do modelo pode ser considerada quase não-tendenciosa (FERREIRA, 2008). Esse método busca avaliar a capacidade de generalização para uma base de dados contendo  $N$  pares entrada-saída, através do desempenho do modelo para uma cardinalidade  $N - 1$ , ou seja, busca-se através da retirada de um padrão entrada-saída a cada iteração, avaliar o desempenho de generalização utilizando o padrão entrada-saída retirado. Com o intuito de utilizar todos os padrões para validação, o procedimento é repetido  $N$  vezes, sendo o erro de generalização determinado pela média aritmética entre os erros de cada iteração.

Fica evidente que a utilização de um método que tem por base a iteração descrita acima necessita de um esforço computacional elevado, principalmente para problemas com diversos padrões entrada-saída. No entanto, como a estimação via validação cruzada única apresenta resultados bastante relevantes, foram desenvolvidos ao longo dos últimos anos, diversos limites analíticos que utilizam como base essa estimação (FERREIRA, 2008). Um deles, está baseado em um limite superior da estimativa associada com a relação *raio/margem* (erro de generalização), o qual utiliza o conceito de extensão dos vetores suporte (*span of support vectors*). Nesse método, adota-se uma variação do *kernel*, a qual visa agrupar o hiperparâmetro  $C$  ao *kernel*, fazendo com que esse hiperparâmetro passe a ser visto como mais um parâmetro do *kernel*, como demonstra (CHAPELLE et al., 2002). O limite superior  $T[f(\underline{x}, \underline{W}, b)]$  pode ser representado pela expressão (67).

$$T[f(\underline{x}, \underline{W}, b)] = \sum_{i=1}^{NS} \alpha_i S_i^2, \quad (67)$$

Sendo que  $\alpha_i$  representam os multiplicadores de Lagrange associados ao vetor suporte  $\underline{x}_i$  e  $S_i$  representa a extensão do  $i$ -ésimo vetor suporte, dada pela equação (68).

$$S_i^2 = \frac{1}{\left(\underline{\tilde{K}}^{-1}\right)_{ii}}, \quad (68)$$

Onde  $\left(\underline{\tilde{K}}^{-1}\right)_{ii}$  representa o  $i$ -ésimo elemento da diagonal da inversa da matriz  $\underline{\tilde{K}}$  dada pela expressão (69).

$$\underline{\tilde{K}} = \begin{bmatrix} \underline{K}_{NS} & \underline{u} \\ \underline{u}^t & 0 \end{bmatrix} \quad (69)$$

Sendo  $\underline{K}_{NS}$ , a matriz com o núcleo do produto interno entre todos os vetores suporte e  $\underline{u} \in \Re^{NS}$  um vetor unitário. Maiores detalhes sobre a extensão de vetores suporte podem ser encontrados em (VAPNIK; CHAPELLE, 2000), sendo que grande parte do embasamento teórico desse tema

foge ao escopo desse trabalho. Nesse momento, valem apenas algumas ressalvas acerca do limite superior do erro de generalização utilizado.

A primeira ressalva está relacionada com as propriedades geométricas do limite do erro de generalização. Da definição da extensão dos vetores de suporte (VAPNIK; CHAPELLE, 2000), obtém-se que o desempenho de generalização da SVM está associado com a distribuição dos vetores de suporte no espaço de características. Essa distribuição implica em se obter um mapeamento concentrado no espaço de características, ou seja, além de se maximizar a superfície de separação, é necessário minimizar o raio da menor hipersfera que contém a imagem de todos os padrões no espaço de características, visando garantir boa generalização através da minimização da dimensão VC, como demonstra a equação (52). Com isso, garante-se boa generalização através da escolha ótima do parâmetros  $C$  e  $\sigma_f^2$ , vistos agora como parâmetros do *kernel* e conseqüentemente relacionados com o mapeamento concentrado no espaço de características.

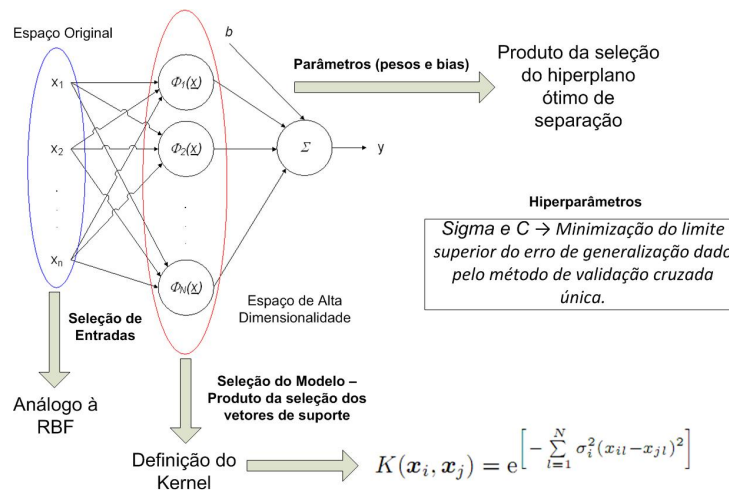
A segunda ressalva está associada com a característica multimodal de  $T[f(\underline{x}, \underline{W}, b)]$ . Em função disso, optou-se neste trabalho, por utilizar algoritmos genéticos (GOLDBERG, 1989) para minimização desse funcional em função de  $C$  e  $\sigma_f^2$ , detalhado no capítulo dedicado aos resultados desse trabalho.

Com isso, apresentam-se todos os tópicos associados com o treinamento autônomo de uma rede SVM. O tópico a seguir, resume as etapas do treinamento, apontando na estrutura de uma SVM os estágios mais relevantes.

#### Resumo do Treinamento de SVMs

A figura 29 mostra a estrutura de uma SVM e os principais tópicos referentes ao seu treinamento. De forma geral, o algoritmo pode ser resumido em:

1. Inserção de uma variável de prova de natureza aleatória.
2. Definição do conjunto inicial dos hiperparâmetros de forma aleatória com distribuição uniforme.
3. Cálculo dos multiplicadores de Lagrange em função dos hiperparâmetros.
4. Minimização do limite superior do erro de generalização, em função dos hiperparâmetros e multiplicadores de Lagrange.
5. Seleção das entradas mais relevantes para o modelo.



**Figura 29: Estrutura e Treinamento de uma SVM.**

6. Redefinição os multiplicadores de Lagrange em função dos hiperparâmetros.
7. Minimização do limite superior do erro de generalização, em função dos novos hiperparâmetros e multiplicadores de Lagrange.

## 5 MODELAGEM DO PROBLEMA EM AMBIENTE DE SIMULAÇÃO

Como citado anteriormente, existe uma grande dificuldade para a utilização das oscilografias geradas pelo PQM, uma vez que não se encontram disponíveis as identificações adequadas para essas oscilografias, impossibilitando a implementação de um algoritmo de aprendizagem com base nesses dados. Sendo assim, no intuito de se obter as formas de onda que caracterizam os eventos selecionados nesse trabalho, da forma mais próxima possível das formas de onda registradas pelo PQM, optou-se por realizar simulações digitais no ambiente do *software Alternative Transient Program*, ATP, (EMTP, 1995). O ATP é um sistema universal para a simulação digital de transitórios eletromagnéticos, bem como eventos de natureza eletromecânica. Seu principal objetivo é fornecer subsídios (modelos básicos de elementos do SEP) para confecção de redes complexas e sistemas de controle de estrutura arbitrária em ambiente de simulação. Dentre suas principais características pode-se destacar (EMTP, 1995):

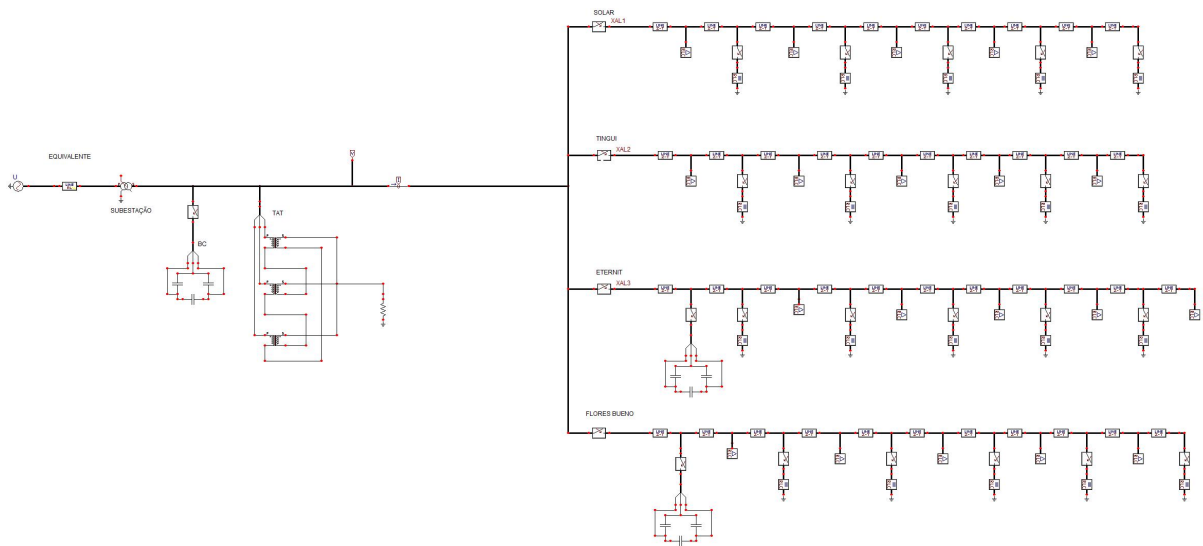
- Capacidade de resolver as equações diferenciais dos componentes do sistema no domínio do tempo, utilizando um método baseado na integração trapezoidal.
- Possibilidade da definição de condições iniciais diferentes de zero, as quais podem ser determinadas automaticamente por uma solução de estado estacionário fasorial ou podem ser inseridas pelo usuário.
- Distúrbios simétricos ou não simétricos, tais como faltas, surtos de raios e quaisquer tipos de manobras podem ser gerados de forma bastante confiável.

Para tanto, o ATP faz uso de uma biblioteca que contém diversos elementos básicos de um SEP, como p. ex., elementos resistivos, indutivos e capacitivos concentrados, linhas de transmissão e distribuição com cabos com parâmetros ajustáveis (concentrados ou distribuídos), resistências e indutâncias não-lineares, transformadores com saturação e histerese e interruptores comuns, com abertura controlada por tempo ou valor de tensão.

Com base nessas características, optou-se por construir um padrão de rede de distribuição de forma reduzida, com o intuito de reproduzir uma subestação de distribuição real com seus

principais componentes e alimentadores, visando obter as formas de onda dos eventos escolhidos no trabalho para posterior desenvolvimento do método de classificação. A subestação escolhida para a modelagem foi a subestação do Atuba da Copel Distribuição localizada na cidade de Curitiba, no estado do Paraná. A justificativa para a escolha dessa subestação foi a facilidade na obtenção dos dados para modelagem, uma vez que já se conheciam os parâmetros necessários para modelagem de todos os seus componentes, resultado de trabalhos anteriores realizados pela própria Copel em conjunto com o Lactec.

O modelo foi construído no software *ATP Draw* que consiste em uma interface gráfica para o ATP. Nessa interface foram inseridos alguns dos principais componentes da subestação Atuba, a saber: equivalente da concessionária, transformador da subestação, transformador de aterramento( TAT), bancos de capacitores( BC), quatro alimentadores de uma barra (denominados Solar, Tingui, Eternit e Flores Bueno) e cargas ao longo dos alimentadores. A figura 30 mostra um diagrama esquemático contendo todos esses elementos representados no *ATP Draw*.



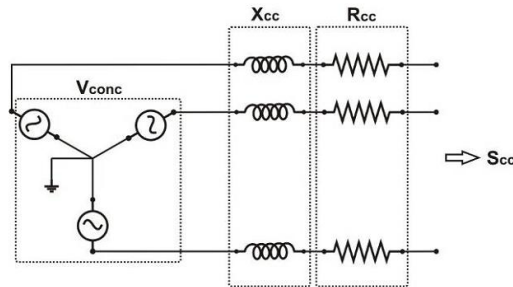
**Figura 30: Diagrama Esquemático do Modelo no ATP Draw.**

A seguir, serão descritos os detalhes da modelagem de cada um desses elementos, bem como a forma como foram gerados os eventos para o método de classificação.

### 5.1 EQUIVALENTE DA CONCESSIONÁRIA

A representação do equivalente elétrico para uma concessionária (equivalente do sistema elétrico anterior ao ponto que se objetiva simular) foi baseada na recomendação de Oliveira e

Colaboradores (2000), a qual representa o equivalente através de uma fonte trifásica em série com uma impedância RL, como mostra a figura 31.



**Figura 31: Esquemático do Equivalente Elétrico.**

Na figura 31,  $V_{conc}$  representa a tensão da concessionária em V,  $S_{cc}$  a potência de curto circuito em VA necessária para a estimação dos seus respectivos valores de reatância indutiva de curto-circuito  $X_{cc}$  e resistência de curto-circuito  $R_{cc}$ , ambas dadas em  $\Omega$  (formando a impedância série).

No caso do ATP, o equivalente foi representado por uma fonte trifásica ideal (69 kV, tensão que chega à subestação) ligada em série com um modelo simétrico e acoplado de linha como mostra a figura 32, cujos parâmetros fornecidos ao modelo são de resistência e reatância indutiva de sequência positiva e sequência zero. Esse dados foram obtidos junto à Copel e possuem os seguintes valores:

- Resistência de Sequência Zero:  $1\Omega$ .
- Resistência de Sequência Positiva:  $1\Omega$ .
- Reatância Indutiva de Sequência Zero:  $3,9\Omega$ .
- Reatância Indutiva de Sequência Positiva:  $6,6\Omega$ .

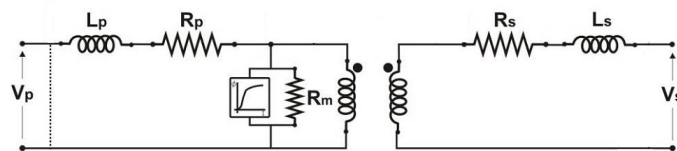


**Figura 32: Esquemático do Equivalente Elétrico no ATP Draw.**

## 5.2 TRANSFORMADOR DA SUBESTAÇÃO

O modelo do transformador trifásico, elemento responsável por transformar os níveis de tensão de 69 kV para 13,8 kV, pode ser representado pelo esquema unifilar da figura 33, como mostra Toro (1999), sendo que para tanto, devem ser especificados os seguintes parâmetros:

- $U(V)$  - Tensão nominal dos enrolamentos primário e secundário.
- $R(\Omega)$  - Resistência dos enrolamentos primário e secundário.
- $Lm(H)$  - Indutância de dispersão dos enrolamentos primário e secundário.
- $I(A)$  - Valor de pico da corrente no joelho da curva de magnetização.
- $F(Wb - T)$  - Valor de pico do fluxo no joelho da curva de magnetização.
- $Rm(\Omega)$  - Resistência de magnetização (associada com as Perdas em Vazio).
- Ligação do Transformador.

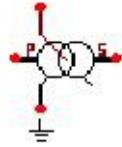


**Figura 33: Esquemático do Transformador da Subestação.**

Nesse modelo, em função da dificuldade encontrada para obtenção dos dados de saturação do transformador, não foram consideradas as variáveis associadas com a curva de magnetização do transformador (característica não-linear da modelagem). Sendo assim, os dados utilizados para a representação do transformador da subestação (figura 34) foram:

- Tensão nominal do enrolamento primário - 69 kV entre fases.
- Tensão nominal do enrolamento secundário - 13,8 kV entre fases.
- Resistência do enrolamento primário - 0,0144 $\Omega$ .
- Resistência do enrolamento secundário - 0,45 $\Omega$ .
- Indutância de dispersão do enrolamento primário - 0,36mH.
- Indutância de dispersão do enrolamento secundário - 11,21mH.
- Resistência de magnetização - 10<sup>15</sup> $\Omega$ .
- Ligação do Transformador - Estrela no lado de alta e Delta no lado de Baixa tensão.





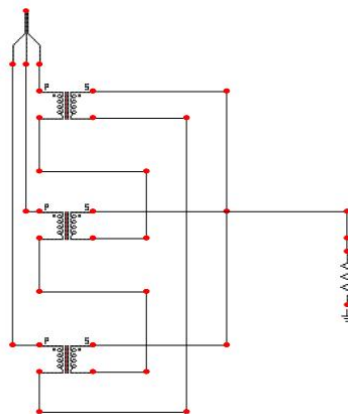
**Figura 34: Esquemático do Transformador da Subestação no *ATP Draw*.**

### 5.3 TRANSFORMADOR DE ATERRAMENTO

Um transformador de aterramento (TAT) tem por objetivo prover uma ligação à terra para circuitos que não possuem tal conexão, sendo nesse caso em particular, circuitos com ligação em delta (secundário do transformador da subestação). A utilização de um transformador de aterramento permite a circulação de correntes de sequência zero (TORO, 1999) na ocorrência de curtos-circuitos que envolvam a terra, permitindo a implementação de métodos de proteção contra esse tipo de evento. Além disso, o TAT promove uma limitação de sobretensões que possam vir a ocorrer nas fases não faltosas.

O TAT mais comumente utilizado em redes de distribuição da Copel possui uma ligação bastante particular, constituída por um enrolamento primário ligado em *zig-zag* sem a existência de um enrolamento secundário, configurando assim um autotransformador de dois enrolamentos com relação unitária entre eles.

Uma possível representação desse elemento no ATP, como sugere Mason (2000), consiste em utilizar três transformadores monofásicos com relação de transformação unitária, interligados em série e aterrados por uma resistência de 20 ohms, como mostra a figura 35, implementada no *ATP Draw*.



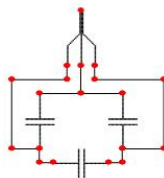
**Figura 35: Esquemático do Transformador de Aterramento no *ATP Draw*.**

Nessa representação, é necessário especificar a tensão nominal dos enrolamentos primário e secundário, resistência dos enrolamentos primário e secundário, indutância de dispersão dos enrolamentos primário e secundário e a resistência de magnetização, de cada transformador de modo a atender às especificações do TAT. No caso da Copel, para um TAT que suporte 2940 Ampéres durante 10 segundos (TAT existente na subestação Atuba), deve-se utilizar os seguintes parâmetros para cada transformador monofásico:

- Tensão nominal do enrolamento primário - 13,8 kV entre fases.
- Tensão nominal do enrolamento secundário - 13,8 kV entre fases.
- Resistência do enrolamento primário - 0,5Ω.
- Resistência do enrolamento secundário - 0,5Ω.
- Indutância de dispersão do enrolamento primário - 3,95Ω.
- Indutância de dispersão do enrolamento secundário - 3,95Ω.
- Resistência de magnetização -  $10^{11}\Omega$ .

#### 5.4 BANCOS DE CAPACITORES

No modelo de distribuição utilizado existem três diferentes bancos de capacitores. O primeiro deles está instalado na subestação e os outros dois ao longo de dois alimentadores (Eternit e Flores Bueno). A especificação desses bancos é feita apenas pela sua potência reativa trifásica, uma vez que os valores das capacitâncias são retiradas a partir desse valor. A ligação dos bancos é em delta (padrão utilizado pela Copel), conforme mostra a figura 36.



**Figura 36: Esquemático de um Banco de Capacitores no ATP Draw.**

As potências reativas dos bancos utilizados são:

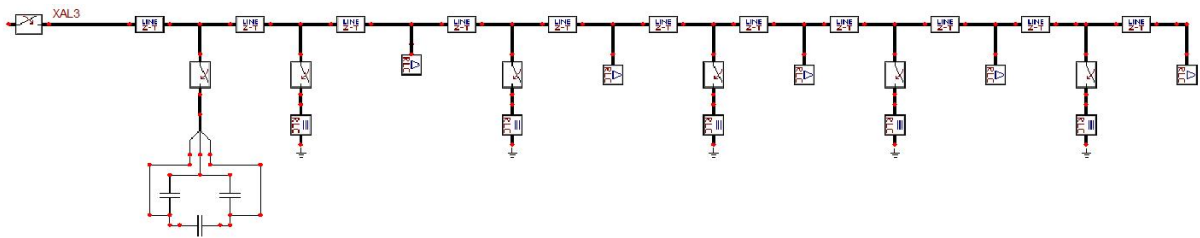
- Banco de Capacitores da Subestação - 4800kVAR.
- Banco de Capacitores do Alimentador Eternit - 600kVAR.

- Banco de Capacitores do Alimentador Flores Bueno - 600kVAR.

## 5.5 ALIMENTADORES E CABOS

Em função da representação completa de um alimentador de distribuição ser extremamente complexa, dada sua extensão e diversos ramais, optou-se aqui por representar apenas o tronco principal do alimentador, sendo esse seccionado em diversos trechos, de forma que pudessem ser inseridos em cada trecho um componente para geração de eventos (como p. ex., um chave) e uma carga equivalente naquele ponto, como mostra a figura 37.

Em cada trecho foi selecionado um tipo de cabo, tendo em vista a configuração original do tronco do alimentador. Dentre os cabos utilizados, pode-se citar: CA 336MMC, CAA 1/0 AWG, CAA 2 AWG, CAA 4 AWG e XLPE 70.



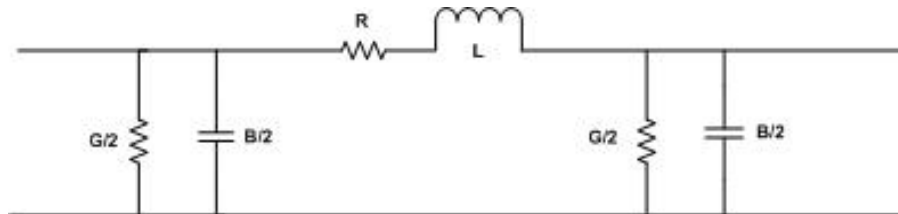
**Figura 37: Esquemático de um Alimentador no ATP Draw.**

Para a representação destes cabos, foi utilizado o componente trifásico para linhas com parâmetros distribuídos e constantes (Modelo de Clarke) (OLIVEIRA et al., 2000). Este modelo assume que os parâmetros elétricos são constantes, ou seja, não se alteram em função da frequência.

Esse método divide a linha em elementos de circuitos  $\pi$  infinitesimais colocados em cascata, a partir dos parâmetros elétricos fornecidos pelo usuário, de sequência positiva e também de sequência zero ou nula. Em seguida, calcula-se um modelo  $\pi$  equivalente (figura 38) para ambas as sequências. Este método é apropriado para simulação de eventos transitórios e em casos de sistemas desequilibrados. Os parâmetros elétricos de entrada para esse modelo são:

- Resistência de sequência positiva por quilômetro  $R/I+$ .
- Resistência de sequência zero por quilômetro  $R/I0$ .
- Indutância de sequência positiva por quilômetro ou sua admitância  $A+$ .
- Indutância de sequência positiva por quilômetro ou sua admitância  $A0$ .

- Capacitância de sequência positiva por quilômetro ou sua susceptância  $B+$ .
- Capacitância de sequência zero por quilômetro ou sua susceptância  $B0$ .
- Comprimento da linha em quilômetros.



**Figura 38: Modelo  $\Pi$  Equivalente.**

A tabela 5 mostra os cabos em função dos seus parâmetros, obtidos junto ao banco de dados da Copel.

**Tabela 5: Parâmetros de Cabos.**

Cabo	$R/I+$ ( $\Omega/km$ )	$R/I0$ ( $\Omega/km$ )	$A+$ ( $mH/km$ )	$A0$ ( $mH/km$ )	$B+$ ( $\mu F/km$ )	$B0$ ( $\mu F/km$ )
CAA 336 AWG	0,1900	0,3680	1,0366	4,92	0,01	0,004
CAA 1/0 AWG	0,6960	0,8737	1,3746	5,2548	0,01	0,004
CAA 2 AWG	1,050	1,228	1,3894	5,2700	0,01	0,004
CAA 4 AWG	1,597	1,7747	1,3793	5,2601	0,01	0,004
XLPE 70	0,5683	0,7461	0,1436	2,5000	0,01	0,004

## 5.6 CARGAS

As cargas foram dispostas em modelos equivalentes distribuídos ao longo dos alimentadores (no total de cinco por alimentador), com ligação em delta (figura 39). Essas cargas são compostas por elementos resistivos, indutivos e capacitivos, de forma que seus valores são ajustados de acordo com a potência ativa e reativa provenientes da curva de carga de cada alimentador. A tabela 6 mostra o comprimento aproximado do tronco principal de cada alimentador, bem como sua potência ativa e reativa média.



**Figura 39: Esquemático da Carga no ATP Draw.**

**Tabela 6: Parâmetros dos Alimentadores.**

Alimentador	Comprimento ( <i>km</i> )	Potência Ativa ( <i>kW</i> )	Potência Reativa ( <i>kvar</i> )
Flores Bueno	21,9	3555	579
Solar	31,5	5397	949
Eternit	34,8	2000	100
Tingui	12,9	2600	120

As curvas de carga obtidas foram seccionadas em nove valores, os quais representam a carga diária de um alimentador em termos de sua potência ativa e reativa, de forma que esses nove valores pudessem ser utilizados na geração dos diversos eventos, caracterizando a ocorrência em diferentes estágios de carregamento.

### 5.7 GERAÇÃO AUTOMÁTICA DE EVENTOS

Uma vez determinados os parâmetros envolvidos na modelagem, é possível se estabelecer um método para geração automática dos eventos para o processo de classificação, levando em conta os parâmetros a serem variados nas simulações.

Nesse trabalho, foram variados os instantes e o local de ocorrência dos eventos, resistências de falta envolvidas (no caso dos curtos-circuitos), os pontos escolhidos nas curvas de carga dos alimentadores (sendo simulados os eventos para diferentes valores de carregamento) e por último, foi considerada uma distorção harmônica típica dessa subestação, aproximando o modelo da operação real do sistema.

Essa variação foi feita de forma automática através de um programa desenvolvido em Matlab, considerando um arquivo base em formato *.atp*, no qual poderiam ser alteradas todas as variáveis envolvidas na geração dos eventos. Além disso, esse programa permite a geração dos resultados de simulação, executando a simulação de cada arquivo editado em formato *.atp*, resultando em um arquivo com formato manipulável em ambiente Matlab, que foi o software escolhido para aplicação do processo de classificação.

Com isso, pode-se obter todos os arquivos necessários para a simulação com um único arquivo base, gerados de forma automática e com flexibilidade no ajuste das variáveis a serem manipuladas na simulação como mostram as tabelas 7 e 8.

Adicionalmente, pode-se observar na figura 40 a numeração correspondente aos locais de ocorrência dos eventos (pontos de 1 a 20). Para cada conjunto obteve-se um total de 6579 formas de onda, sendo 441 casos (instâncias) para cada evento de curto-circuito e 432 casos

**Tabela 7: Parâmetros Variados para o Conjunto 1.**

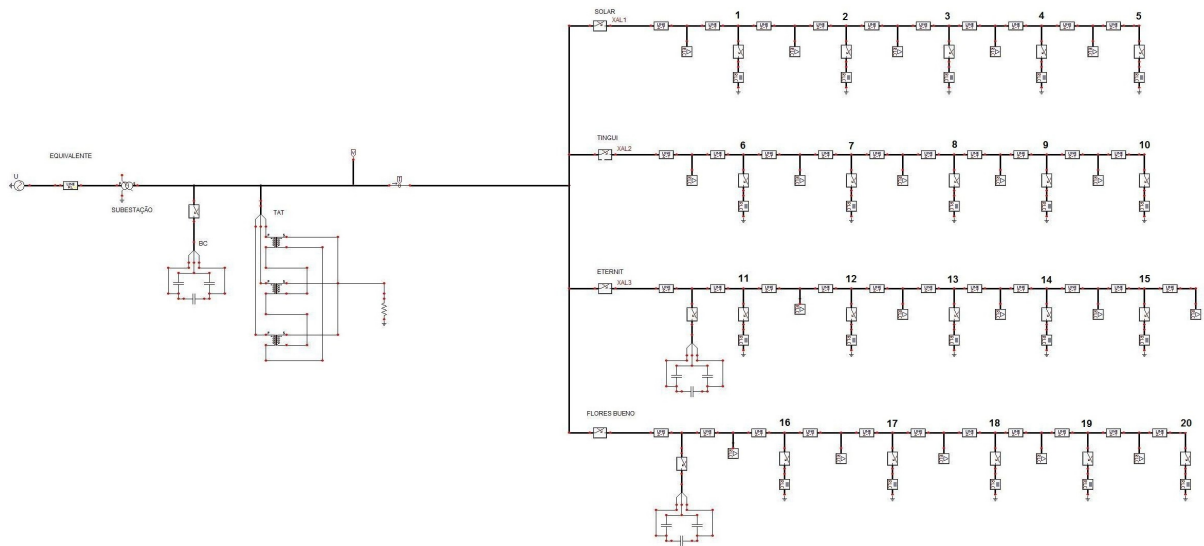
Evento	Resistência de Falta [ $\Omega$ ]	Local de Ocorrência	Ângulo de Incidência (Graus)	Pontos da Curva de Carga
Curto-circuito Monofásico	5; 300; 2000	Pontos Ímpares	0; 45; 90	1; 3; 7
Curto-circuito Bifásicos	5; 300; 2000	Pontos Ímpares	0; 45; 90	1; 3; 7
Curto-circuito Bifásicos com Terra	5; 300; 2000	Pontos Ímpares	0; 45; 90	1; 3; 7
Curto-circuito Trifásicos	5; 300; 2000	Pontos Ímpares	0; 45; 90	1; 3; 7
Curto-circuito Trifásicos com Terra	5; 300; 2000	Pontos Ímpares	0; 45; 90	1; 3; 7
Chaveamento de Bancos de Capacitores	Não há	Os três Bancos	0; 45; 90	1; 3; 7
Abertura nos Religadores	Não há	Todos os religadores	0; 45; 90	1; 3; 7
Fechamento nos Religadores	Não há	Todos os religadores	0; 45; 90	1; 3; 7

**Tabela 8: Parâmetros Variados para o Conjunto 2.**

Evento	Resistência de Falta [ $\Omega$ ]	Local de Ocorrência	Ângulo de Incidência (Graus)	Pontos da Curva de Carga
Curto-circuito Monofásico	10; 600; 1000	Pontos Pares	10; 60; 130	2; 4; 9
Curto-circuito Bifásicos	10; 600; 1000	Pontos Pares	10; 60; 130	2; 4; 9
Curto-circuito Bifásicos com Terra	10; 600; 1000	Pontos Pares	10; 60; 130	2; 4; 9
Curto-circuito Trifásicos	10; 600; 1000	Pontos Pares	10; 60; 130	2; 4; 9
Curto-circuito Trifásicos com Terra	10; 600; 1000	Pontos Pares	10; 60; 130	2; 4; 9
Chaveamento de Bancos de Capacitores	Não há	Os três Bancos	10; 60; 130	2; 4; 9
Abertura nos Religadores	Não há	Todos os religadores	10; 60; 130	2; 4; 9
Fechamento nos Religadores	Não há	Todos os religadores	10; 60; 130	2; 4; 9

para os demais eventos, incluindo uma classe de eventos sem problema aparente, denominada classe normal. Os conjuntos um e dois são utilizados como base para a obtenção dos conjuntos de treinamento e teste dos classificadores apresentados neste trabalho. Com isso, forma-se toda

a base de dados necessária para o desenvolvimento dos classificadores.



**Figura 40: Local de Ocorrência dos Eventos.**

## 6 RESULTADOS E DISCUSSÕES

Nesse capítulo serão discutidos os principais resultados obtidos para os classificadores testados, utilizando diferentes formas de pré-processamento dos sinais de análise. Na análise dos resultados serão levados em consideração os seguintes aspectos: acerto total do conjunto de treinamento, acerto total para o conjunto de teste, capacidade de generalização do classificador, estrutura neural automaticamente definida e seleção de entradas.

Como base de comparação entre os classificadores e seus diferentes pré-processamentos, foram gerados os conjuntos de treinamento e de teste a partir dos arquivos simulados apresentados no capítulo anterior, retirando aleatoriamente padrões das diferentes classes do problema. Com isso, é possível realizar a avaliação de todos os classificadores utilizando os mesmos conjuntos de treinamento e teste. A seleção aleatória dos padrões foi escolhida visando garantir uma melhor forma de se avaliar a capacidade de generalização dos modelos, uma vez que um número diferente de instâncias por classe reflete uma realidade do sistema prático, pois normalmente não se dispõe de um número de oscilografias iguais por classe. O resultado final dessa seleção, em relação ao número de instâncias por classe, está representado na tabela 9.

### 6.1 RESULTADOS PARA AS REDES MLP

Os resultados do primeiro classificador, baseado em uma rede neural MLP estão representados na tabela 10.

#### 6.1.1 Acerto Total do Classificador

Para o conjunto de treinamento das redes MLP, o acerto médio foi de aproximadamente 91% com o maior acerto percentual para a função wavelet *sym2*, sendo esse acerto de 97% para o conjunto em questão. Já para o conjunto de teste, essa função wavelet juntamente com as funções *db8* e *coif5*, obtiveram o maior desempenho, ultrapassando 96% de acerto. Nesse conjunto, o acerto médio foi de 90%.



**Tabela 9: Número de Instâncias por Classe.**

Classe	Instâncias Treinamento	Instâncias de teste
Abertura Trifásica de um Alimentador	434	425
Chaveamento de Banco de Capacitores	414	450
Curto Circuito Bifásico entre as fases A e B	480	396
Curto Circuito Bifásico entre as fases B e C	394	486
Curto Circuito Bifásico entre as fases C e A	435	440
Curto Circuito Trifásico	427	409
Curto Circuito Trifásico para a terra	329	296
Curto Circuito Bifásico para a terra entre as fases A e B	427	454
Curto Circuito Bifásico para a terra entre as fases B e C	433	447
Curto Circuito Bifásico para a terra entre as fases C e A	438	442
Curto Circuito monofásico para a terra na fase A	460	422
Curto Circuito monofásico para a terra na fase B	439	438
Curto Circuito monofásico para a terra na fase C	455	423
Sinal sem defeito	427	437
Religamento de um alimentador trifásico	441	409

A família de funções wavelet que obteve o melhor desempenho, tanto para o conjunto de treino quanto para o conjunto de teste, foi a *Symlet* com acerto médio acima de 94%. O desempenho para as demais formas de pré-processamento foi bastante próximo, exceto para o pré-processamento baseado em TRF, que obteve um acerto inferior ao acerto médio das funções wavelet para o conjunto de teste (75% da TRF contra 90% das wavelets).

Uma possível justificativa para esse fato encontra fundamento na diferença entre as energias das raias da TRF e das energias dos diversos níveis da TWD. No caso da TWD, os transitórios envolvidos com os eventos simulados são bem caracterizados nos diversos níveis da TWD, já na TRF, essa informação encontra-se espalhada nos harmônicos de ordem próxima as frequências geradas nos eventos. Essa conclusão se tornará mais evidente no tópico referente à seleção de entradas das redes RBF e SVM.

**Tabela 10: Acertos para a MLP.**

Pré Processamento	Acerto Treino(%)	Acerto Teste (%)	Neurônios Ocultos	Entradas Seleccionadas
<i>db1</i>	87	87	19	33/33
<i>db4</i>	83	83	11	33/33
<b><i>db8</i></b>	<b>96</b>	<b>96</b>	<b>20</b>	<b>33/33</b>
<i>db12</i>	64	65	16	33/33
<i>db15</i>	95	95	18	33/33
<i>coif1</i>	93	92	13	33/33
<i>coif3</i>	93	93	8	33/33
<b><i>coif5</i></b>	<b>96</b>	<b>96</b>	<b>17</b>	<b>33/33</b>
<b><i>sym2</i></b>	<b>97</b>	<b>96</b>	<b>20</b>	<b>33/33</b>
<i>sym4</i>	95	94	17	33/33
<i>sym8</i>	87	86	19	33/33
<i>bior1.1</i>	95	93	19	33/33
<i>bior1.5</i>	78	77	19	33/33
<i>bior2.2</i>	87	87	20	33/33
<i>bior2.8</i>	91	90	18	33/33
<i>bior3.1</i>	89	88	19	33/33
<i>bior3.9</i>	89	93	18	33/33
<i>bior4.4</i>	93	93	19	33/33
<i>bior5.5</i>	89	88	20	33/33
<i>bior6.8</i>	91	90	19	33/33
<i>Fourier</i>	79	74	10	192/192

### 6.1.2 Capacidade de Generalização

A análise comparativa dos acertos de treinamento e teste fornece um forte indicativo da capacidade de generalização do classificador. Para a MLP, o desempenho do conjunto de teste difere em aproximadamente 1% do desempenho do conjunto de treinamento para o pré-processamento baseado em TWD. Para a TRF, essa diferença fica em 5%. Esses resultados indicam que a rede MLP possui boa capacidade de generalização, uma vez que seu desempenho para padrões não contidos no conjunto de treinamento foi bastante próximo ao desempenho de treinamento. Além disso, pode-se destacar que a separação realizada para os conjuntos de treino e teste foi bem sucedida, pois a informação armazenada nos parâmetros da rede foi suficiente para a rede obter bom desempenho para os padrões de teste.

No caso da MLP, a aplicação do regularizador baseado no decaimento de pesos garante que durante o processo de estimação dos parâmetros e hiperparâmetros da rede, mantenha-se um compromisso entre a minimização do erro de treinamento e o controle de complexidade da rede, nesse caso fundamentada pelo mapeamento suave resultante da aplicação da teoria da regularização.

### 6.1.3 Estrutura Definida

A estrutura definida na camada intermediária para as redes MLP é produto da aplicação da seleção *bayesiana* de modelos através da maximização da evidência. Foram testadas redes variando o número de neurônios na camada oculta, sendo que a estrutura selecionada é a que apresenta a maior evidência, ou seja, aquela mais provável à luz dos dados. Todos os pré-processamentos foram avaliados variando de 1 a 25 neurônios na camada oculta, sendo que em nenhum caso foi selecionado um modelo com mais de 20 neurônios.

Um detalhe interessante é que modelos com diferentes pré-processamentos resultaram em modelos diferentes com um desempenho muito próximo, como p. ex. a função *coif5* e a função *sym2*. Nesse caso, a função *coif5* conseguiu um resultado de 96% de acerto para o conjunto de teste com 17 neurônios na camada oculta. Já a função *sym2* obteve o mesmo acerto com 20 neurônios ocultos. Isso demonstra a forte dependência do pré-processamento na caracterização dos sinais nas diversas bandas de frequência, além de confirmar a eficiência do método escolhido para a seleção do modelo que representa um bom desempenho de generalização para modelos mais simples. De modo geral, a família *coiflet* resultou em um menor número de neurônios.

Já com relação à seleção de entradas, o método de determinação automática de relevância resultou em uma seleção de todas as entradas para todas as formas de pré-processamento avaliadas, ou seja, em nenhum caso foi encontrada uma variável de entrada irrelevante para o método. A análise comparativa entre os  $\alpha_i$  obtidos para as variáveis de entrada do problema e o  $\alpha_0$  da variável de natureza aleatória inserida, mostra que para o classificador, o nível de irrelevância associada à entrada de natureza aleatória não é suficiente para eliminar demais entradas. De um modo geral, todas as entradas para as redes MLP são relevantes na classificação, indicando também a adequação do método de pré-processamento.

## 6.2 RESULTADOS PARA AS REDES RBF

Um resumo dos classificadores baseados em redes RBF pode ser representado pela tabela 11.

### 6.2.1 Acerto Total do Classificador

O acerto médio para o conjunto de treinamento das redes RBF nas diversas formas de pré-processamento foi de aproximadamente 79%, sendo que as funções wavelet *db4* e *db12* obtiveram o maior acerto percentual, ultrapassando 92% de acerto. As mesmas funções obtiveram

**Tabela 11: Acertos para a RBF.**

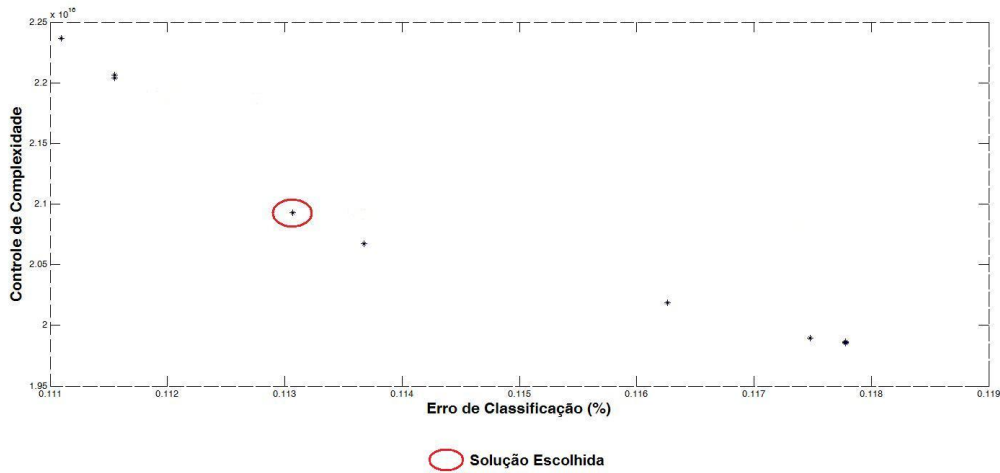
Pré Processamento	Acerto Treino(%)	Acerto Teste (%)	Neurônios Ocultos	Entradas Seleccionadas
<i>db1</i>	63	62	223	14/33
<b><i>db4</i></b>	<b>92</b>	<b>90</b>	<b>610</b>	<b>33/33</b>
<i>db8</i>	89	87	532	27/33
<b><i>db12</i></b>	<b>92</b>	<b>91</b>	<b>536</b>	<b>29/33</b>
<i>db15</i>	79	77	393	16/33
<i>coif1</i>	78	76	352	18/33
<i>coif3</i>	81	80	337	18/33
<i>coif5</i>	76	75	309	18/33
<i>sym2</i>	86	83	463	25/33
<i>sym4</i>	84	83	347	18/33
<i>sym8</i>	78	77	381	18/33
<i>bior1.1</i>	74	73	252	18/33
<i>bior1.5</i>	87	85	493	29/33
<i>bior2.2</i>	70	69	350	18/33
<i>bior2.8</i>	78	77	324	18/33
<i>bior3.1</i>	70	69	409	18/33
<i>bior3.9</i>	82	80	273	18/33
<i>bior4.4</i>	88	85	473	26/33
<i>bior5.5</i>	73	72	344	18/33
<i>bior6.8</i>	86	85	419	21/33
<i>Fourier</i>	46	42	436	31/192

o maior desempenho para o conjunto de teste, superando 90% de classificações corretas. Para esse conjunto, o acerto médio foi de 77%. A família de funções wavelet que obteve o melhor desempenho foi a *Daubechies*, com acerto médio acima de 81%, sendo que as demais famílias obtiveram desempenho muito próximo ao da média. A única forma de pré-processamento que resultou em um acerto excessivamente inferior ao acerto médio foi o método baseado em TRF, cujo desempenho foi menor que 50% para os conjuntos de treino e teste. A mesma justificativa utilizada para MLP pode ser repetida aqui, com relação às energias das raias da TRF em relação às energias dos diversos níveis da TWD.

### 6.2.2 Capacidade de Generalização

No caso da RBF, observa-se que o desempenho para o conjunto de teste é bastante próximo ao do conjunto de treinamento para todas as formas de pré-processamento. Essa observação comprova o bom desempenho das redes para padrões não inseridos durante o treinamento. Além disso a otimização multi-objetivo realizada durante o treinamento, inclui uma parcela responsável, mesmo que indiretamente, pelo controle de complexidade. Essa otimização pode

ser vista através da análise da figura 41, que contém o resultado da fronteira de Pareto para o pré-processamento utilizando a função *db12*.



**Figura 41: Possíveis Soluções para os Modelos da Rede RBF.**

Nessa figura, evidencia-se a minimização conjunta do erro de treinamento e da complexidade do modelo, sendo o resultado final escolhido com base na solução com a maior média harmônica, levando em conta o compromisso entre os dois objetivos da minimização. A solução final está ressaltada na figura 41. Para essa otimização durante o treinamento das RBFs, foram utilizados os seguinte parâmetros para o AGMO:

- Função Objetivo - Dada pela equação (47), na qual visa-se minimizar-se as duas parcelas  $f_1 + \omega f_2$ , sendo  $\omega$  o fator de normalização relacionada com a parcela de controle de complexidade.
- Espaço de busca com população inicial da localização dos centros definida em  $[-1, 1]$ . Os hiperparâmetros  $\sigma_l$  definidos em  $[\log(10^{-5}), \log(10^6)]$ , como sugere (FERREIRA, 2008) e  $\sigma_i$  definida em  $[10^{-10}, 100]$ .
- Indivíduos - Inicialmente formados por  $\sigma_l$  e a localização dos centros, e posteriormente por  $\sigma_i$ ,  $\sigma_l$  e a localização dos centros, todos utilizando codificação em números reais.
- Número de Gerações - 20. Em função da rápida convergência do algoritmo, o número de gerações pode ser reduzido para 20.
- Número da População - 20, sugerido por (FERREIRA, 2008).
- Cruzamento - Em um único ponto dos dois indivíduos pais, com uma taxa de cruzamento de 90%, como sugere (DEB et al., 2002).

- Mutação - Utilizada a taxa de  $1/n$ , sendo  $n$  o número de variáveis de decisão do problema, de acordo com (DEB et al., 2002).
- Escolha da melhor solução - Dado pela solução não-dominada com melhor média harmônica, como mostra (SIERRA; COELLO, 2006).

### 6.2.3 Estrutura Definida

No caso da RBF, a seleção da estrutura é feita através da aplicação de uma rede neural GWR. O produto dessa rede é a definição automática do número de neurônios na camada oculta da RBF, bem como suas respectivas localizações. Nota-se que para a RBF o número de neurônios escolhidos na camada oculta é relativamente alto, sendo em média aproximadamente 380 neurônios ocultos. Essa quantidade está diretamente relacionada com a natureza da localização dos centros no algoritmo da rede GWR.

Após uma série de testes com a base de dados dessa dissertação, além das sugestões propostas por (MARSLAND; SHAPIRO; NEHMZOW, 2002), obtiveram-se os seguintes valores para os parâmetros da rede GWR:

- Número máximo de iterações do Algoritmo - 50;
- Valor de Comparação para o Critério de Habituação - 0,1;
- Valor de Comparação para o Critério de Ativação do neurônio - 0,7;
- Taxa de Aprendizagem do neurônio vencedor - 0,1;
- Taxa de Aprendizagem dos neurônios vizinhos - 0,01;
- Taxa com que o modelo se habitua a padrões repetidamente apresentados - 0,3;
- Taxa com que o padrão passa a ser considerado como novidade - 0,1.

Sendo os critério de parada dados por:

- Número Máximo de neurônios - 2000;
- Número Máximo de Conexões - Quatro vezes maior que o número de neurônios;
- Número Máximo de Vizinhos - 500;
- Idade Máxima da Conexão - 100.

O número de neurônios inseridos ao longo do treinamento da GWR pode ser visto como o número de novidades encontradas ao longo da apresentação de novos padrões de entrada. Essas novidades podem ser interpretadas como sendo padrões que não se encontram próximos a nenhum centro já estabelecido pela GWR, configurando um novo centro para a rede. Cada novo centro caracteriza uma nova região no espaço dos centros da GWR, a qual responde por uma determinada característica contida nos padrões de entrada. Outra observação relevante diz respeito ao número de neurônios e ao número de entradas selecionadas. À medida que o número de entradas selecionadas aumenta, o número de centros aumenta de forma bastante linear, reforçando a característica da rede neural GWR como detector de novidades. Em outras palavras, pode-se afirmar que quanto mais entradas forem selecionadas, mais características serão identificadas - relevantes ou não - nas diversas bandas de frequência, resultando em um número maior novidades detectadas nos padrões de entrada, durante o processo de treinamento da GWR.

O método de seleção de entradas adotado para as redes RBF mostrou-se bastante eficiente em relação ao número de entradas selecionadas, resultando em média uma seleção de 18 entradas para o classificador. No entanto, a interpretação associada às bandas de frequência possui uma característica bastante particular para cada forma de pré-processamento dos sinais.

Analisando inicialmente as funções wavelet com melhor desempenho: *db4* e *db12*. No caso da *db4*, o método de seleção de entradas não encontrou variáveis irrelevantes para o modelo, levando em conta todas as 11 bandas de frequência para as três fases dos sinais de tensão. Já para a função *db12*, os sinais detalhe *D2* da fase B, *D5*, *D8* e *D10* da fase C foram considerados irrelevantes para o processo de classificação. Um detalhe importante é que a função *db12* obteve um desempenho ligeiramente superior ao da função *db4* para o conjunto de teste. Essa seleção realizada para a função *db12* possui uma característica bastante específica em relação as bandas de frequência escolhidas. Nesse caso, nenhum nível de energia da TWD correspondente à fase A foi eliminado, ao contrário da fase C da qual foram retirados três diferentes bandas de frequência. Nesse caso, os níveis eliminados não fornecem uma caracterização específica dos eventos simulados, já que a seleção de entradas se encontra basicamente na Fase C e os eventos foram gerados de forma equilibrada em relação às três fases.

Já para a função *coif5*, nota-se que dentre todas as entradas selecionadas, o sinal detalhe *D10* e o sinal aproximação *A1* podem ser eliminados da decomposição para todas as fases do sinal de tensão, uma vez que foram considerados irrelevantes para o classificador com o pré-processamento em questão. Essa eliminação implica em não considerar a banda de frequência de 0 à 7 Hz, pois em função desse pré-processamento a banda de frequência em questão não

caracteriza os eventos a serem classificados.

Uma característica comum a maioria das funções utilizadas diz respeito a não eliminação dos sinais detalhe de alta frequência. Nesses sinais encontram-se as principais características transitórias dos eventos estudados, principalmente nos dois primeiros níveis de decomposição. Além disso, pode-se citar a eliminação do sinal detalhe *D10* em grande parte das funções wavelet. Acredita-se que a frequência contida nesse nível da TWD não está associada a nenhuma característica específica dos eventos simulados, pois se encontra em uma faixa de frequência inferior à frequência fundamental dos sinais de tensão (60 Hz), faixa essa que conforme sugere a literatura (DUGAN et al., 2002), não caracteriza eventos transitórios ao longo do SEP. Pode-se destacar por último, o desempenho do método de seleção de entradas para o pré-processamento baseado em TRF. Nesse caso, dos 192 valores utilizados para o pré-processamento, apenas 31 foram selecionados, sendo eliminados principalmente os harmônicos de ordem elevada, os quais não representam de forma significativa os transitórios envolvidos nos diversos eventos simulados.

### 6.3 RESULTADOS PARA AS REDES SVM

A tabela 12 exprime os resultados obtidos para as redes SVM.

#### 6.3.1 Acerto Total do Classificador

No caso da SVM, o desempenho das funções wavelet foi bastante próximo, tanto para o conjunto de treino quanto para o conjunto de teste, sendo o acerto médio desses conjuntos 96% e 95%, respectivamente. Dentre as diversas funções wavelet destacam-se *db1* e *coif1* com acerto superior a 98% para o conjunto de treinamento e as funções *db1* e *bior2.8* com acerto de 96% para o conjunto de teste.

Com relação ao pré-processamento baseado em TRF, o comportamento observado para as demais redes avaliadas foi repetido no caso da SVM, ou seja, o desempenho para esse pré-processamento foi inferior ao desempenho obtido para as wavelets. Nesse caso o acerto para o conjunto de teste foi de 66%, ao passo que para as wavelets foi obtido 95% para esse mesmo conjunto. Com esse resultado comprova-se que a caracterização em frequência proposta pela TRF dificulta a caracterização dos sinais envolvidos com os eventos simulados independentemente do método de classificação utilizado.



**Tabela 12: Acertos para a SVM.**

Pré Processamento	Acerto Treino(%)	Acerto Teste (%)	Número Médio de Vetores Suporte $\pm$ Desvio Padrão	Número Médio de Entradas Seleccionadas $\pm$ Desvio Padrão
<i>db1</i>	<b>98</b>	<b>97</b>	<b>74</b> $\pm$ 122	<b>25/33</b> $\pm$ 6
<i>db4</i>	<b>97</b>	<b>96</b>	<b>80</b> $\pm$ 106	<b>23/33</b> $\pm$ 7
<i>db8</i>	<b>96</b>	<b>96</b>	<b>78</b> $\pm$ 103	<b>24/33</b> $\pm$ 7
<i>db12</i>	95	94	96 $\pm$ 114	24/33 $\pm$ 6
<i>db15</i>	92	90	102 $\pm$ 130	24/33 $\pm$ 6
<i>coif1</i>	<b>98</b>	<b>96</b>	<b>74</b> $\pm$ 81	<b>23/33</b> $\pm$ 7
<i>coif3</i>	96	94	91 $\pm$ 112	25/33 $\pm$ 6
<i>coif5</i>	<b>96</b>	<b>95</b>	<b>65</b> $\pm$ 92	<b>25/33</b> $\pm$ 6
<i>sym2</i>	97	94	80 $\pm$ 105	25/33 $\pm$ 6
<i>sym4</i>	97	96	78 $\pm$ 101	23/33 $\pm$ 6
<i>sym8</i>	96	95	80 $\pm$ 106	24/33 $\pm$ 6
<i>bior1.1</i>	97	95	81 $\pm$ 127	23/33 $\pm$ 7
<i>bior1.5</i>	96	93	78 $\pm$ 112	22/33 $\pm$ 7
<i>bior2.2</i>	97	95	85 $\pm$ 117	25/33 $\pm$ 7
<b><i>bior2.8</i></b>	<b>97</b>	<b>97</b>	<b>94</b> $\pm$ 120	<b>24/33</b> $\pm$ 6
<i>bior3.1</i>	97	94	88 $\pm$ 117	24/33 $\pm$ 6
<i>bior3.9</i>	95	94	78 $\pm$ 109	24/33 $\pm$ 7
<i>bior4.4</i>	96	94	77 $\pm$ 101	24/33 $\pm$ 7
<i>bior5.5</i>	96	95	83 $\pm$ 115	25/33 $\pm$ 6
<i>bior6.8</i>	96	95	87 $\pm$ 114	25/33 $\pm$ 7
<i>Fourier</i>	75	66	303 $\pm$ 219	135/192 $\pm$ 39

### 6.3.2 Capacidade de Generalização

Uma vez que a rede SVM tem por base o método de minimização estrutural do risco, garante-se que durante o treinamento do modelo, será mantido o compromisso entre a minimização do erro de treinamento e sua capacidade de generalização, em função do termo associado com a dimensão VC presente na estimação dos parâmetros da rede. Esse resultado fica evidente através da comparação entre os desempenhos obtidos para o treinamento e teste da rede.

Além disso, no método de treinamento utilizado leva em conta a minimização do limite superior do erro de generalização estimado via validação cruzada única, sendo que um dos produtos desse método é uma estimativa para o parâmetro  $C$ . Essa variável pode ser vista como um parâmetro de regularização para essas redes, mantendo equilíbrio entre o ajuste aos dados de treinamento e a complexidade do modelo. Como citado anteriormente esse limite é minimizado através do uso de AG, cujas parametrizações utilizadas para essa minimização foram:

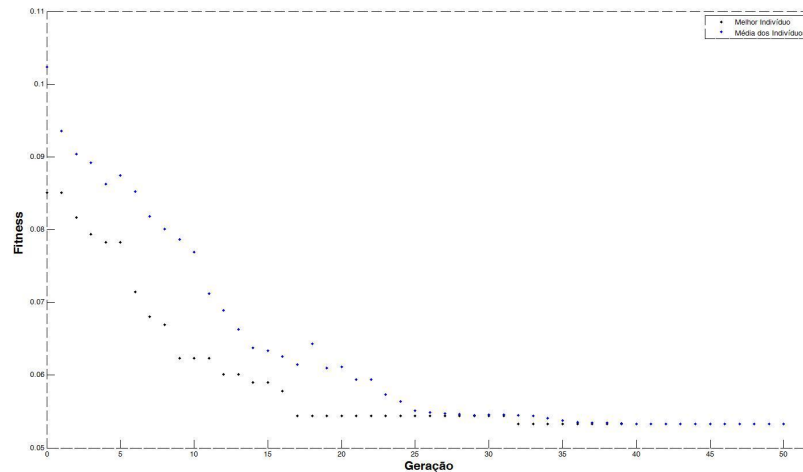
Para tanto, o algoritmo genético foi parametrizado com as seguintes características:

- Função Objetivo - Dada por  $T[f(\underline{x}, \underline{W}, b)]$ , a qual visa-se minimizar.
- Indivíduos - Formados por  $C$  e  $\sigma_f^2$ , utilizando codificação em números reais.
- Número de Gerações - 30. Em função da rápida convergência do algoritmo como mostra a figura 42, o número de gerações pode ser reduzido para 30.
- Número de Indivíduos da População - 20.
- População inicial definida com valores aleatórios entre os limites sugeridos por (FERREIRA, 2008) para  $C$  e  $\sigma_f^2$ , da mesma forma que as redes RBF, apresentadas anteriormente.
- Seleção - Baseada no método da amostragem universal estocástica. Esse método deriva de um método de seleção conhecido como método da roleta, no qual os indivíduos são organizados por diferentes setores formando uma roleta, de forma que a seleção dos indivíduos é feita pelo sorteio aleatório nessa roleta, sendo retirado um indivíduo para cada rodada da roleta. No caso particular da amostragem universal estocástica, são estabelecidos diversos ponteiros igualmente espaçados na roleta, de forma que esses ponteiros representem as diversas rodadas do método da roleta convencional.
- Cruzamento - Em um único ponto dos dois indivíduos pais, com uma taxa de mutação de 80%, de forma que se mantenha um compromisso entre a diversidade genética e a convergência do algoritmo.
- Mutação - Utilizada a taxa de 5%, de forma a se evitar uma rápida convergência para mínimos locais, sem que a busca tenha uma característica aleatória.

### 6.3.3 Estrutura Definida

Na tabela 12 estão representados os valores médios e o desvio padrão tanto para as entradas selecionadas, quanto para o número de vetores suporte em função do treinamento do modelo SVM ser realizado em pares de classes, dificultando a apresentação desses resultados com maiores detalhes neste trabalho, devido ao grande número de combinações possíveis entre as classes. Nesse sentido, o desvio padrão fornece uma ideia da dispersão dos valores em torno do valor médio.

No caso dos vetores suporte, foram selecionados em média 80 vetores suporte para as diversas redes SVM, nas suas diversas combinações (aos pares). Em função desses valores serem



**Figura 42: Convergência do AG durante o Treinamento da rede SVM.**

aplicados aos pares de classes, sua interpretação é bastante particular para cada caso. No entanto, pode-se verificar que a complexidade dentre os vários modelos com diferentes formas de pré-processamento é bastante próxima, ou seja, o número de vetores suporte selecionados para os vários modelos e seus diversos pares de classe foi bastante próximo da média geral de vetores suporte selecionados. A exceção é o pré-processamento realizado com TRF que obteve uma seleção média de 303 vetores suporte. Nesse caso, o número de vetores suporte variou de forma bastante significativa em relação à média para os diversos pares de classes avaliados, caracterizando um relação bastante específica entre cada par de classes.

Já com relação ao número de entradas selecionadas, observa-se que o método de seleção se mostrou eficiente, determinando em média 24 entradas como relevantes para o classificador. Ao contrário da interpretação aplicada às redes RBF, no caso da SVM uma análise sobre as principais bandas de frequência selecionadas se torna inviável, pois como o treinamento é feito aos pares, as bandas de frequência relevantes são válidas apenas para o par sob análise. No entanto, observou-se que as bandas rejeitadas nos diversos pares de classes se concentram nas bandas de mais baixa frequência, ou seja, os níveis mais elevados da TWD, como verificado nas redes RBF. Isso demonstra novamente que a caracterização dos eventos se concentra nos primeiros níveis da transformada.

No caso da TRF, a seleção de entradas não demonstrou a mesma caracterização com relação aos eventos classificados. Para esse pré-processamento, cada par de classe teve uma seleção específica de entradas apresentando pouca ou até mesmo nenhuma correlação entre os pares avaliados.

## 6.4 COMPARAÇÃO GERAL

Com os resultados expostos acima, pode-se comparar o desempenho das diversas redes testadas, de forma geral, apontando as vantagens e desvantagens dos métodos, como mostram os tópicos seguintes.

### 6.4.1 Acerto Total dos Classificadores

Analisando o acerto médio de todas as formas de pré-processamento para os três tipos de redes testadas, observa-se que a SVM obteve desempenho superior às demais redes. Seu acerto médio para o conjunto de teste foi de 94%, enquanto que as redes MLP obtiveram 90% e as redes RBF 77%.

Além disso, as redes SVM apresentaram desempenho muito próximo para as várias formas de pré-processamento baseadas em TWD, com um desvio bastante baixo em relação ao valor médio, indicando que sua aplicação é menos dependente das características das funções wavelet utilizadas em comparação com as redes RBF e MLP.

Com base nos resultados expostos, pode-se afirmar que um pré-processamento baseado em TWD possui melhores características do que o baseado em TRF, para o problema de classificação exposto nesse trabalho. Além disso, as famílias *Daubechies*, *Coiflet* e *Symlet* mostraram-se as mais promissoras das funções wavelet avaliadas para todos do modelos neurais testados.

Com o intuito de se avaliar as matrizes de confusão<sup>1</sup> geradas pelos classificadores, foram escolhidas três métricas baseadas em testes estatísticos: o teste  $\chi^2$ , o  $V$  de Cramer e uma medida baseada na entropia e no coeficiente de incerteza associados com o processo de classificação, como sugere (NEHMZOW, 2006).

O primeiro método, o teste  $\chi^2$ , visa fornecer uma medida que subsidia a análise da significância da associação entre a saída desejada e a saída real desse classificador. Esse teste no entanto, é fortemente dependente número de amostras existentes na matriz de confusão, fazendo com que a correlação entre a saída desejada e a saída real se torne dependente do número de amostras disponíveis. Uma forma de se contornar essa dependência é através da reparametrização realizada pelo método do  $V$  de Cramer, cujo resultado torna a análise independente do número de amostras. Com isso, pode-se afirmar que método do  $V$  de Cramer representa uma forma de se determinar se existe ou não, uma associação significativa entre as

---

<sup>1</sup>São matrizes responsáveis por associar a saída do classificador com as classes do problema. Na sua diagonal principal encontram-se as classes corretamente classificadas, enquanto que o restante da matriz apresenta os erros cometidos pelo classificador para cada classe.

linhas e as colunas da matriz de confusão, permitindo a comparação de matrizes com diferentes número de amostras (para um mesmo classificador). Já as medidas baseadas em entropia atribuem um coeficiente relacionado com a qualidade da associação entre a resposta desejada e a resposta real, por vezes denominada de força da associação.

Depois de efetuada a avaliação dos classificadores por esses testes, notou-se que os seus resultados, principalmente para o teste  $V$  de Cramer e as medidas baseadas em entropia, demonstraram um comportamento muito próximo ao acerto dos classificadores para o conjunto de teste. Esse resultado fundamenta a análise através dos acertos percentuais dos classificadores, ou seja, os testes estatísticos efetuados fundamentam a comparação entre os diversos classificadores através do acerto médio para o conjunto de teste.

Finalmente, vale ressaltar que o treinamento e a avaliação das redes foram repetidos diversas vezes para algumas formas de pré-processamento, visando verificar a repetibilidade dos resultados. Para todos os modelos, o acerto se manteve muito próximo ao longo das repetições realizadas, confirmando os resultados expostos nas tabelas 10, 11 e 12.

#### 6.4.2 Capacidade de Generalização

Em todos os modelos avaliados, os métodos de controle de complexidade escolhidos se mostraram bastante eficientes, garantido aos modelos boa capacidade de generalização. De forma geral, os três modelos testados apresentaram resultados próximos em relação à generalização, não apresentando limitadores para suas aplicações.

#### 6.4.3 Estrutura Definida

Com relação à seleção da estrutura, cabe salientar as características autônomas utilizadas nos três modelos avaliados. Para os três modelos foi implementada uma técnica particular para a definição da estrutura, objetivando a menor (ou a menos impactante) intervenção do usuário no treinamento das redes. Nos três casos obteve-se sucesso na definição automática da estrutura, porém cada modelo apresentou particularidades que devem ser levadas em conta na avaliação global dos classificadores.

No caso da MLP é necessária a definição do número mínimo e máximo de neurônios a serem avaliados na camada oculta da rede. Em alguns testes iniciais, observou-se que o número de neurônios máximo escolhido não foi suficiente para a definição do modelo com a maior evidência à luz dos dados. Nesses casos foi necessária a alteração do valor máximo de neurônios, visando obter o melhor modelo com base nos dados de treinamento. Isso demons-

tra que existe a necessidade da intervenção do usuário na definição do modelo. No entanto, verificou-se que mesmo nos casos em que o número máximo de neurônios teve de ser alterado, o desempenho da rede com um número de neurônios maior pouco alterou em relação à rede inicialmente avaliada, mostrando que a dependência em relação a esse parâmetro é menos crítica no contexto geral do treinamento da MLP.

Para as redes RBF, os parâmetros previamente selecionados para a rede GWR impactam fortemente no resultado gerado por essa rede, alterando de forma significativa o treinamento da rede RBF como um todo. Os testes realizados mostraram que alguns parâmetros possuem influência maior que outros sobre o resultado gerado, em especial o número de épocas escolhido. Além disso, os parâmetros do AGMO também influenciam o resultado do treinamento, mesmo que de forma menos impactante.

Isso vale também para a SVM, que utiliza um AG para a minimização do limite superior do erro de generalização. Nesse caso, a influência dos parâmetros é comparável ao AGMO da RBF, que para a SVM em particular é função da natureza da função a ser minimizada. Um detalhe importante é que, no estágio responsável por essa minimização, foi avaliada uma técnica baseada em *Otimização por Enxame de Partículas* (SIERRA; COELLO, 2006), com o intuito de comparar seu desempenho com o desempenho do AG. Os resultados obtidos para os dois métodos foram muito próximos, optando-se assim, por apresentar os resultados com base no AG inicialmente desenvolvido.

Em resumo, cada um dos modelos avaliados apresenta em pelo menos um estágio do processo de treinamento, a dependência de uma seleção prévia de um conjunto de inicial de parâmetros. No caso especial da RBF, essa realidade representa o maior inconveniente sob o aspecto do treinamento autônomo, principalmente no estágio não-supervisionado do treinamento. Atualmente, a dependência dos parâmetros está sendo avaliada através de uma análise mais aprofundada das técnicas utilizadas, principalmente em relação à GWR, além da possibilidade de se utilizar técnicas alternativas para os problemas expostos, mantendo o foco nas propriedades autônomas desejadas. No entanto, essa é uma questão ainda em aberto, tanto para a RBF, quanto para os demais modelos testados.

Com relação à seleção de entradas, a RBF apresentou uma estimativa bastante promissora com relação às bandas de frequência mais relevantes para a classificação. Essa resposta fornece uma caracterização bastante importante dos eventos analisados, além de ressaltar que a TWD promove uma divisão de frequência capaz de caracterizar muito bem os sinais sob análise. Os resultados iniciais dessa técnica poderão servir de base para um estudo mais aprofundado das características dos eventos em redes de distribuição, facilitando na compreensão, identificação

e localização desses eventos. No caso da MLP essa análise não pode ser realizada em função do método não ter considerado variáveis irrelevantes durante o treinamento. O mesmo acontece para a SVM, porém nesse caso a análise fica inviabilizada devido ao treinamento dessas redes ser feito aos pares.

De forma geral, a SVM apresentou os melhores e mais promissores resultados, tanto para o desempenho das redes, quanto para as características autônomas avaliadas. A característica principal do método de treinamento dessas redes fundamenta o conceito de autonomia exposto no decorrer desse trabalho, relacionado com o controle de complexidade e seleção do modelo.

## 7 CONCLUSÕES E TRABALHOS FUTUROS

Como foi discutido no início do trabalho, em função do crescente uso de dispositivos eletrônicos de potência e controles industriais, aumentou-se a preocupação, principalmente por parte das concessionárias de energia, com eventos relacionados com Qualidade de Energia Elétrica. Esses eventos alteram significativamente a forma de onda de tensão e/ou corrente resultantes de curto-circuito ou má operação de equipamentos instalados em consumidores de energia. Essa alteração nas formas de onda pode, por vezes, resultar em danos a esses consumidores, forçando as concessionárias a monitorar o fornecimento dessa energia para que essa seja entregue de forma adequada nos vários estágios do sistema elétrico de potência.

Uma consequência bastante impactante para a concessionária nesse sentido é que, em função da grande quantidade e diversidade de eventos que ocorrem na rede, os registros armazenados não são devidamente classificados e analisados, dificultando a utilização de forma adequada dos bancos de dados gerados, como foi mostrado no caso dos registros do PQM. Com isso, faz-se necessário o desenvolvimento de um método de classificação automática para os dados gerados (oscilografias) pelo PQM.

Apesar dos métodos de classificação de eventos, principalmente de QEE, estarem bastante difundidos na literatura, a particularidade de cada sistema de distribuição implica em um desenvolvimento específico do classificador em questão. Nesse trabalho, a grande dificuldade foi associar eventos de naturezas distintas, como eventos de curto-circuito e eventos relacionados com manobras na rede, no mesmo classificador. O desempenho dos métodos de classificação sugere que tal associação é possível, mesmo para uma grande quantidade de classes de eventos, como as classes avaliadas nesse trabalho. Além disso, destaca-se que esse bom desempenho em termos de classificação de eventos de natureza distinta foi obtido apenas com oscilografias de tensão, diferentemente de outros trabalhos que, no caso de classificação de faltas, utilizam dados de corrente ou de tensão e corrente. Na continuação desse trabalho, pretende-se adicionar novas classes de eventos associadas com a particularidade de cada alimentador, levando em conta uma metodologia para tratar essas diferentes particularidades.



O modelo de simulação utilizado para geração de oscilografias devidamente identificadas, visando a estimação e avaliação dos classificadores com aprendizado supervisionado, apresentou resultados bastante promissores, principalmente por levar em conta uma característica típica dos sistemas de distribuição: conteúdo harmônico dos alimentadores. Além disso, os parâmetros variados forneceram ao método de classificação um grande conjunto de reais possibilidades quanto à ocorrência dos eventos, auxiliando na capacidade de generalização dos classificadores. Ao longo do processo de construção e validação do modelo, observou-se que o nível de detalhamento dos componentes do modelo influi de forma bastante expressiva no processo de classificação como um todo. Com o intuito de aumentar o nível de detalhamento, estão sendo feitas modificações, principalmente com relação aos alimentadores, incluindo nesses os transformadores de distribuição com a carga localizada em cada um dos transformadores, ao invés de representá-la de forma equivalente como foi apresentado nesse trabalho. Espera-se com isso, que o modelo possa representar de forma mais próxima um sistema de distribuição real, fazendo com que se maximize o desempenho do classificador treinado com dados simulados frente aos dados reais.

No estágio de pré-processamento foi sugerido que a forma como se dá a divisão em frequência dos sinais de tensão dos eventos simulados, caracteriza de forma expressiva o desempenho de classificação dos modelos neurais avaliados. Em função disso, a TWD apresentou melhores resultados que TRF, sugerindo que para a classificação em questão, deve-se optar por métodos que levem em conta as características mais proeminentes dos sinais de entrada. Adicionalmente, foi testada uma grande quantidade de funções wavelets, sendo essas comparadas entre si para todos os modelos neurais. Essa comparação forneceu uma base para análise das características de cada família de funções wavelet, permitindo selecionar as melhores famílias para o problema de classificação abordado, as quais correspondem as famílias: *daubechies*, *coiflet* e *symlet*. Atualmente, estão sendo avaliadas outras características presentes na transformada wavelet, bem como outras formas de pré-processamento, a fim de se procurar diferentes formas de caracterizar o sinal, resultando em um número menor de entradas para o classificador, além de comparar com os resultados já obtidos.

Com relação aos modelos neurais, esse trabalho mostra a possibilidade de se trabalhar com o conceito de autonomia junto ao processo de treinamento das redes neurais. A autonomia aqui apresentada, está relacionada com o controle de complexidade e seleção automática da estrutura do modelo, a saber, a seleção de entradas e a definição da camada oculta das redes. Esse método de treinamento é pouco empregado em problemas de classificação, sendo que para a classificação de eventos em redes de distribuição trata-se de uma metodologia bastante original. Além disso, o desempenho geral dos classificadores foi satisfatório, principalmente para

redes SVM, as quais apresentaram o maior acerto global para o conjunto de teste. Em cada uma das redes avaliadas ainda existem questões a serem otimizadas com relação à autonomia, no entanto, ressalva-se que esse é o primeiro passo na avaliação desse tipo de classificador e, no contexto do trabalho, uma operação autônoma é de grande valia tendo em vista a possibilidade do método de classificação desenvolvido, poder ser aplicado nas várias subestações de distribuição da Copel. Vale ressaltar finalmente que, a validação completa da classificação proposta será dada quando essa for aplicada com sucesso aos dados gerados pelo PQM, sendo essa uma das etapas vindouras do trabalho.

Com relação aos objetivos listados inicialmente, ficou demonstrado no capítulo 5 a validação dos modelos dos principais elementos de uma subestação de distribuição de energia. Adicionalmente, nos capítulos 5 e 6, evidencio-se a comparação de diferentes formas de pré-processamento das formas de onda de tensão trifásica de forma a se classificar eventos transitórios e curtos-circuitos com um único classificador, a avaliação do desempenho de diferentes modelos neurais com características autônomas e a elaboração de um método que possa ser aplicado em um sistema real, tendo em vista as diversas características apresentadas ao longo do trabalho.

A classificação automática de oscilografias provenientes dos PQMs pode resultar em ganho enorme na pós-análise de eventos da rede distribuição, uma vez que tal função não está disponível para essas oscilografias. Além disso, a classificação de eventos correspondentes a faltas no sistema de distribuição pode facilitar ou até mesmo permitir que seja feita a localização dos defeitos, com base nas oscilografias registradas pelos PQMs durante as faltas, uma vez que os dados classificados podem ser utilizados para o emprego de técnicas com base em registros oscilográficos para localização de faltas em redes de distribuição.

## 7.1 CONTRIBUIÇÕES DO TRABALHO

Dentre as principais contribuições deste trabalho, destacam-se as relacionadas com os modelos neurais avaliados, sendo elas:

- Aplicação e desenvolvimento de redes RBF com seleção automática de entradas e controle de complexidade para classificação dos eventos apresentados ao longo do trabalho.
- Adaptação de redes MLP com treinamento *bayesiano* e redes SVM com seleção automática de entradas e controle de complexidade utilizadas em problemas de previsão de carga, para o problema de classificação apresentado.

## 7.2 PUBLICAÇÕES REALIZADAS

Ao longo do desenvolvimento desse trabalho foram publicados dois artigos em congressos utilizando como base os resultados parciais obtidos na dissertação. Abaixo, seguem os detalhes dessas publicações.

- André Eugênio Lazzaretti, Vitor H. Ferreira, Hugo Vieira Neto, Rodrigo J. Riella and Julio Omori, *Classification of Events in Distribution Networks using Autonomous Neural Models. In Proceedings of ISAP 2009: 15th International Conference on Intelligent System Applications to Power Systems*, Curitiba, Brazil, 2009.
- Vitor Ferreira, André Eugênio Lazzaretti, Hugo Vieira Neto, Rodrigo Riella, Julio Omori, *Classificação de Eventos em Redes de Distribuição de Energia Utilizando Transformada Wavelet e Modelos Neurais Autônomos (in Portuguese)*. In *Anais do IX Congresso Brasileiro de Redes Neurais / Inteligência Computacional* e selecionado para a revista *Learning and Nonlinear Models*, Ouro Preto, Brazil, 2009. ISSN 21771200.

## REFERÊNCIAS

- AGUILERA, C.; ORDUNA, E.; RATTA, G. Fault detection, classification and faulted phase selection approach based on high-frequency voltage signal applied to a series-compensated line. **IEEE Proceedings on Generation, Transmission and Distribution**, v. 153, n. 4, p. 469–475, July 2006.
- ANEEL. **Resolução N° 024**. [S.l.], 2000.
- ANEEL. **Resolução N° 505**. [S.l.], 2001.
- ANEEL. **Procedimentos de Distribuição, Módulo 8**. [S.l.], 2008.
- BISHOP, C. M. **Neural Networks for Pattern Recognition**. Oxford: Oxford University Press, 1995.
- CHAPELLE, O. et al. Choosing multiple parameters for support vector machines. **Machine Learning**, v. 46, n. 1-3, p. 131–159, January 2002.
- DAG, O.; UCAK, C. Fault classification for power distribution systems via a combined wavelet-neural approach. In: **Proceedings of the 2004 International Conference on Power Systems Technology**. [S.l.: s.n.], 2004. v. 2, p. 1309–1314.
- DAS, D.; SINGH, N.; SINHA, A. A comparison of Fourier transform and wavelet transform methods for detection and classification of faults on transmission lines. In: **IEEE Power India Conference**. [S.l.: s.n.], 2006.
- DAUBECHIES, I. **Ten Lectures on Wavelets**. [S.l.]: SIAM, Philadelphia, 1992.
- DAVARI, S. A.; SHAHRTASH, S. M. Fault classification in distribution power lines based on electromagnetic field measurement and Fourier transform. In: **IEEE PES Conference and Exposition**. [S.l.: s.n.], 2007.
- DEB, K. et al. A fast and elitist multi-objective genetic algorithm: NSGA-II. **IEEE Transactions on Evolutionary Computation**, v. 6, p. 182–197, 2002.
- DELMONT-FILHO, O. **Utilização da transformada wavelet para caracterização de distúrbios na qualidade de energia elétrica**. Dissertação (Mestrado) — Escola de Engenharia de São Carlos, USP, 2003.
- DELMONT-FILHO, O. **Um algoritmo para detecção, localização e classificação de Distúrbios na Qualidade de Energia Elétrica Utilizando a Transformada Wavelet**. Tese (Doutorado) — Escola de Engenharia de São Carlos, USP, 2007.
- DEMUTH, H.; BEALE, M. **Neural Network Toolbox**. [S.l.], 2002.
- DOLA, H.; CHOWDHURY, B. Data mining for distribution system fault classification. In: **Proceedings of the 37th Annual North American Power Symposium**. [S.l.: s.n.], 2005. p. 457–462.

DUGAN, R. C. et al. **Electrical Power System Quality**. 2nd. ed. New York: McGraw-Hill, 2002.

EMTP. **Alternative Transients Program (ATP) Rule Book**. [S.l.], 1995.

FARIA, R. R. A. **Aplicação de Wavelets na Análise de Gestos Musicais em Timbres de Instrumentos Acústicos Tradicionais**. Dissertação (Mestrado) — Universidade de São Paulo (USP), 1997.

FERREIRA, V. H. **Desenvolvimento de Modelos Neurais Autônomos para Previsão de Carga Elétrica**. Tese (Doutorado) — COPPE/UFRJ, 2008.

FERREIRA, V. H.; SILVA, A. P. Alves da. Toward estimating autonomous neural network load forecasters. **IEEE Transactions on Power Systems**, v. 22, n. 4, p. 1554–1562, November 2007.

FLOREZ, J. J. M.; MORALES, G.; LONDONO, S. P. Classification methodology and feature to assist fault location in power distribution systems. **Rev. Fac. Ing. Univ. Antioquia**, v. 44, p. 83–96, 2008.

GAFOOR, S. A.; RAO, P. V. R. Wavelet based fault detection, classification and location in transmission lines. In: **Proceedings of the First International Power and Energy Conference**. [S.l.: s.n.], 2006. p. 114–118.

GOLDBERG, D. E. **Genetic Algorithms in Search, Optimization and Machine Learning**. Reading, MA: Addison-Wesley Professional, 1989.

HAYKIN, S. **Redes Neurais: Princípios e Prática**. [S.l.]: Bookman, 2001.

HAYKIN, S.; VEEN, B. V. **Sinais e Sistemas**. [S.l.]: Porto Alegre Bookman, 2001. 668 p.

HSU, H. P. **Análise de Fourier**. [S.l.]: Rio de Janeiro: LTC- Livros Técnicos e Científicos, 1973.

KASHYAP, K. H.; SHENOY, U. J. Classification of power system faults using wavelet transforms and probabilistic neural networks. In: **Proceedings of the 2003 International Symposium on Circuits and Systems**. [S.l.: s.n.], 2003. p. 423–426.

KHORASHADI, H.; AGHAEBRAHIMI, M. R. A novel approach to fault classification and fault location for medium voltage cables based on artificial neural network. **International Journal of Computational Intelligence**, v. 2, n. 2, p. 90–93, Spring 2006.

KINDERMANN, G. **Curto-Circuito**. [S.l.]: Universidade Federal de Santa Catarina, 2007.

KONAKA, A.; COIT, D. W.; SMITH, A. E. Multi-objective optimization using genetic algorithms: A tutorial. **Elsevier Science**, v. 15, p. 992–1007, 2006.

MALATHI, V.; MARIMUTHU, N. Multi-class support vector machine approach for fault classification in power transmission line. In: **IEEE International Conference on Sustainable Energy Technologies**. [S.l.: s.n.], 2008. p. 66–71.

MALLAT, S. **A Wavelet Tour of Signal Processing**. 2nd edition. ed. [S.l.]: Academic Press, 1999.

- MARSLAND, S.; SHAPIRO, J.; NEHMZOW, U. A self-organising network that grows when required. **Elsevier Science**, v. 15, p. 1041–1058, 2002.
- MASON, R. **The Art and Science of Protective Relaying**. [S.l.]: General Electric, 2000.
- MATHWORKS. **Signal Processing Toolbox Users Guide - The Mathworks Inc.** [S.l.], 2009.
- MATHWORKS. **Wavelet Toolbox User's Guide - The Mathworks Inc.** [S.l.], 2009.
- MELO, M. O. B. C.; CAVALCANTI, G. A. Avaliação do impacto da qualidade de energia elétrica no mercado e na produção industrial: análise e metodologia. In: **Proceedings of the V SBQEE**. [S.l.: s.n.], 2003.
- MORAIS, J. et al. Data mining applied to the electric power industry: Classification of short-circuit faults in transmission lines. In: **Seventh International Conference on Intelligent Systems Design and Applications**. [S.l.: s.n.], 2007. p. 943–948.
- NEHMZOW, U. **Scientific Methods in Mobile Robotics**. [S.l.]: Springer, 2006.
- NGAOPITAKKUL, A.; PONGCHAISSRIKUL, W.; KUNAKORN, A. Analysis of characteristics of simultaneous faults in electrical power systems using wavelet transform. In: **Proceedings of the IEEE International Conference on Sustainable Energy Technologies**. [S.l.: s.n.], 2008. p. 249–252.
- OLIVEIRA, C. C. B. d. et al. **Introdução a Sistemas Elétricos de Potência**. [S.l.]: Bulcher, 2000.
- ONS. **Procedimentos de rede, Módulo 2, Submódulo 2.2: Padrões de desempenho da rede básica**. [S.l.], 2000.
- ONS. **Procedimentos de rede, Módulo 2, submódulo 2.8: Gerência dos indicadores de desempenho da rede básica**. [S.l.], 2000.
- ONS. **Procedimentos de rede, Módulo 3, submódulo 3.8: Requisitos mínimos para conexão à rede básica**. [S.l.], 2000.
- OPPENHEIM, A. V.; WILLSKY, A. S. **Signals & Systems**. [S.l.]: NAWAB, 1997.
- PIROUTI, M.; FATIH, A. A.; SADIK, I. B. Fault identification and classification for short medium voltage underground cable based on artificial neural networks. **Electrical Engineering**, v. 59, p. 272–276, 2008.
- RADDY, M. J.; MOHANTA, D. K. Adaptive-neuro-fuzzy inference system approach for transmission line fault classification and location incorporating effects of power swings. **IET Gener. Transm. Distr.**, v. 2, p. 235–244, 2008.
- RAVIKUMAR, B.; THUKARAM, D.; KHINCHA, H. P. Application of support vector machines for fault diagnosis in power transmission lines. **IET Generation, Transmission and Distribution**, v. 2, p. 119–130, 2008.
- RIELLA, R. J. et al. Desenvolvimento de um sistema de monitoramento contínuo da qualidade da energia elétrica para subestações de distribuição. In: **Anais do XVIII Seminário Nacional de Distribuição de Energia Elétrica – SENDI**. Olinda: [s.n.], 2008.

- SALIM, R. et al. Hybrid fault diagnosis scheme implementation for power distribution systems automation. **IEEE Transactions on Power Delivery**, v. 23, p. 1846–1856, 2008.
- SAMANTARAY, S. R.; DASH, P. K.; PANDA, G. Fault classification and location using HS-transform and radial basis function neural network. **Electric Power Systems Research**, v. 1, p. 1–9, 2006.
- SIERRA, M. R.; COELLO, C. A. C. Multi-objective particle swarm optimizers: A survey of the state-of-the-art. **International Journal of Computational Intelligence Research**, v. 2, p. 287–308, 2006.
- SILVA, K. M.; SOUZA, B. A.; BRITO, N. S. D. Fault detection and classification in transmission lines based on wavelet transform and ANN. **IEEE Transactions on Power Delivery**, v. 21, n. 4, p. 2058–2063, October 2006.
- SOUZA, E. et al. Comparação das bases de wavelets ortonormais e biortogonais: Implementação, vantagens e desvantagens no posicionamento com GPS. **TEMA Tend. Mat. Apl. Comput.**, v. 8, p. 149–158, 2007.
- TORO, V. D. **Fundamentos de Maquinas Eletricas**. [S.l.]: Editora LTC, 1999.
- VAPNIK, V.; CHAPELLE, O. Bounds on error expectation for support vector machines. **Neural Computation**, v. 12, n. 9, p. 2013–2036, September 2000.
- VAPNIK, V. N. **Statistical Learning Theory**. New York: Wiley-Interscience, 1998.
- VIEIRA-NETO, H. **Visual Novelty Detection for Autonomous Inspection Robots**. Tese (Doutorado) — University of Essex, England, 2006.
- YOUSSEF, O. A. S. Fault classification based on wavelet transforms. In: **Transmission and Distribution Conference and Exposition**. [S.l.: s.n.], 2001. p. 531–536.
- ZHANG, B.; ZHENGYOU, H.; QIAN, Q. Application of wavelet entropy and adaptive nerve-fuzzy inference to fault classification. In: **Proceedings of the International Conference on Power System Technology**. [S.l.: s.n.], 2006. p. 1–6.
- ZHAO, R. et al. A new machine learning method based on PCA and SVM. In: **Proceedings of the International Conference on Computational Intelligence and Security**. [S.l.: s.n.], 2006. p. 187–190.
- ZHENG-YOU, H.; XIAOQING, C.; GUOMING, L. Wavelet entropy measure definition and its application for transmission line fault detection and identification. In: **Proceedings of the 2006 International Conference on Power System Technology**. [S.l.: s.n.], 2006. p. 1–6.