

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

ANDRÉ EUGÊNIO LAZZARETTI

**SEGMENTAÇÃO, CLASSIFICAÇÃO E DETECÇÃO DE NOVAS  
CLASSES DE EVENTOS EM OSCILOGRAFIAS DE REDES DE  
DISTRIBUIÇÃO DE ENERGIA ELÉTRICA**

TESE DE DOUTORADO

CURITIBA

2015

ANDRÉ EUGÊNIO LAZZARETTI

**SEGMENTAÇÃO, CLASSIFICAÇÃO E DETECÇÃO DE NOVAS  
CLASSES DE EVENTOS EM OSCILOGRAFIAS DE REDES DE  
DISTRIBUIÇÃO DE ENERGIA ELÉTRICA**

Tese de Doutorado 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 “Doutor em Ciências” – Área de Concentração: Engenharia da Computação.

Orientador: Prof. Dr. Hugo Vieira Neto

Co-orientador: Prof. Dr. Vitor Hugo Ferreira

**CURITIBA**

**2015**

---

**Dados Internacionais de Catalogação na Publicação**

---

L432s Lazzaretti, André Eugênio  
2015 Segmentação, classificação e detecção de novas classes de eventos em oscilografias de redes de distribuição de energia elétrica / André Eugênio Lazzaretti.-- 2015.  
220 f.: il.; 30 cm

Texto em português, com resumo em inglês.

Tese (Doutorado) - Universidade Tecnológica Federal do Paraná. Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial, Curitiba, 2015.  
Bibliografia: f. 207-212.

1. Energia elétrica - Distribuição. 2. Energia elétrica - Transmissão. 3. Localização de falha elétrica. 4. Transitórios (Eletricidade). 5. Osciladores elétricos. 6. Wavelets (Matemática). 7. Sistemas de energia elétrica - Estabilidade. 8. Métodos de simulação. 9. Engenharia elétrica - Teses. I. Vieira Neto, Hugo, orient. II. Ferreira, Vitor Hugo, coorient. III. Universidade Tecnológica Federal do Paraná - Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial. IV. Título.

CDD 22 -- 621.3

---

**Biblioteca Central da UTFPR, Câmpus Curitiba**

Título da Tese Nº. \_\_\_\_\_

# **Segmentação, Classificação e Detecção de Novas Classes de Eventos em Oscilografias de Redes de Distribuição de Energia Elétrica**

por

## **André Eugênio Lazzaretti**

**Orientador:** Prof. Dr. Hugo Vieira Neto  
**Coorientador:** Prof. Dr. Vitor Hugo Ferreira

Esta tese foi apresentada como requisito parcial à obtenção do grau de DOUTOR EM CIÊNCIAS – Área de Concentração: Engenharia de Computação, pelo Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial – CPGEI – da Universidade Tecnológica Federal do Paraná – UTFPR, às 14h00 do dia 27 de fevereiro de 2015. O trabalho foi aprovado pela Banca Examinadora, composta pelos doutores:

---

Prof. Dr. Vitor Hugo Ferreira  
(Presidente – UFF)

---

Prof. Dr. Alexandre Rasi Aoki  
(LACTEC/UFPR)

---

Dr. Germano Lambert Torres  
(PS Soluções)

---

Prof. Dr. Alceu de Souza Britto Júnior  
(PUCPR)

---

Prof.<sup>a</sup> Dr.<sup>a</sup> Lúcia Valéria Ramos de Arruda  
(UTFPR)

Visto da Coordenação:

---

Prof. Emilio Carlos Gomes Wille, Dr.  
(Coordenador do CPGEI)

## AGRADECIMENTOS

Ao meu orientador Hugo Vieira Neto e ao meu coorientador Vitor Hugo Ferreira, pelo interesse, paciência e enorme apoio científico nesta linha de pesquisa. Tenho certeza que ao longo dos últimos seis anos nos tornamos grandes amigos e acredito que esse foi o resultado mais importante obtido ao longo desta caminhada. Valeu Hugo! Valeu Vitor!

Ao prof. Dr. David M. J. Tax, por me receber na Universidade Tecnológica de Delft (TU Delft) na Holanda e pelas diversas discussões que impactaram diretamente na estrutura deste trabalho, bem como futuros desenvolvimentos nesta linha de pesquisa.

Aos vários colegas do Lactec que auxiliaram de forma significativa no desenvolvimento do trabalho, dentre os quais gostaria de citar: Luiz Felipe R. B. Toledo, Fábio A. Guerra, Marcio R. Sans, Mateus D. Teixeira, Henry L. L. Salamanca, Diogo B. Dahlke, Edemir L. Kowalski, Signie L. F. Santos, Victor S. Borges, Jefferson Chapieski, Ricardo C. Scholz, Mariana C. Coelho e Eunelson J. Silva Júnior.

Ao colega Cleverson L. S. Pinto da Copel Distribuição, por nos fornecer os dados e apresentar os principais problemas da concessionária em relação ao método proposto, bem como discussões ao longo do desenvolvimento que facilitaram a identificação das metodologias mais adequadas ao problema.

Ao Lactec e a Copel Distribuição, pelo incentivo e suporte na realização do trabalho. Gostaria de ressaltar que parte deste desenvolvimento foi financiado pelo programa de Pesquisa e Desenvolvimento da Agência Nacional de Energia Elétrica, dentro de dois projetos representados pelos códigos *PD-2866-0256/2011* e *PD-2866-023/2006*.

À CAPES pelo apoio financeiro durante o estágio sanduíche na Holanda.

Aos demais membros da banca avaliadora, Germano L. Torres, Alexandre R. Aoki, Alceu S. Britto Júnior e Lúcia Valéria R. Arruda pelas sugestões e contribuições na revisão do documento.

Aos meus pais e minha irmã, pelo apoio incondicional em todos os momentos da minha vida e a minha querida namorada (companheira, amiga e inúmeras outras atribuições que não cabem neste espaço) Gabriela da Silva Carlessi, pela compreensão durante a realização deste trabalho e por estar presente nos momentos mais importantes dessa jornada.

## RESUMO

LAZZARETTI, André Eugênio. SEGMENTAÇÃO, CLASSIFICAÇÃO E DETECÇÃO DE NOVAS CLASSES DE EVENTOS EM OSCILOGRAFIAS DE REDES DE DISTRIBUIÇÃO DE ENERGIA ELÉTRICA. 220 f. Tese de Doutorado – Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial, Universidade Tecnológica Federal do Paraná. Curitiba, 2015.

Este trabalho apresenta novas abordagens para duas das etapas fundamentais relacionadas com análise automática de oscilografias de redes de distribuição: a detecção dos instantes transitórios e a sua classificação. Para comparação e validação dos métodos são utilizadas duas bases de dados, sendo uma delas formada por dados simulados no aplicativo *Alternative Transient Program* e outra contendo dados reais de oscilógrafos instalados em uma rede de distribuição de energia elétrica. Os dados reais apresentam um conjunto de eventos relevante para as análises aqui propostas, principalmente por conter uma gama variada de eventos, incluindo transitórios decorrentes de descargas atmosféricas. Com relação à detecção de transitórios (segmentação de oscilografias), foram testados os métodos atualmente propostos na literatura, os quais contemplam Filtro de Kalman, Transformada *Wavelet* Discreta e Modelos Autorregressivos, além de serem propostas duas novas técnicas baseadas no Operador de Energia de Teager e Representação de Dados Utilizando Vetores Suporte. Demonstra-se que, tanto para dados simulados quanto para dados reais, o método de detecção baseado na Representação de Dados utilizando Vetores Suporte aponta para um melhor desempenho global no processo de detecção. Com relação à classificação automática de oscilografias, propõe-se uma nova abordagem incluindo um estágio dedicado à detecção de padrões não inseridos no aprendizado prévio do classificador, denominados de novidades, além da própria classificação multiclasse normalmente empregada para diferenciar múltiplas classes conhecidas *a priori*. São testadas abordagens utilizando a detecção de novidades e classificação multiclasse em estágios simultâneos e subsequentes, com base nos classificadores *X-Means*, *K-Vizinhos-Mais-Próximos* e Representação de Dados Utilizando Vetores Suporte com diferentes formulações, além do próprio classificador multiclasse baseado em Máquinas de Vetor Suporte. Adicionalmente, é proposto um tratamento aos padrões considerados como novidades, com o intuito de fornecer informações ao especialista sobre as similaridades existentes entre os padrões desse conjunto. Para realizar esse processo, optou-se por utilizar modelos de agrupamento automático. Os resultados finais, principalmente para a base de dados incluindo eventos reais, mostram que é possível obter um desempenho de classificação relevante (acima de 80%) para cada um dos estágios do processo de classificação proposto, o qual inclui a detecção de novidades, a classificação multiclasse e o processamento de padrões classificados como novidades (agrupamento automático).

**Palavras-chave:** Análise de Oscilografias, Classificação de Oscilografias, Detecção de Novidades, Detecção de Novas Classes, Segmentação de Oscilografias.

## ABSTRACT

LAZZARETTI, André Eugênio. SEGMENTATION, CLASSIFICATION AND IDENTIFICATION OF NEW CLASSES OF EVENTS IN ELECTRICAL POWER DISTRIBUTION NETWORKS OSCILLOGRAPHIES. 220 f. Tese de Doutorado – Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial, Universidade Tecnológica Federal do Paraná. Curitiba, 2015.

This work presents new approaches for two of the fundamental steps in automatic waveform analysis in electrical distribution systems: transient time detection and its classification. Two datasets were used to compare and validate the proposed methods. The first is composed by simulated waveforms, by using the Alternative Transient Program, while the second is formed by real data from a monitoring system developed for overhead distribution power lines. The real data present a set of relevant events for the analysis proposed here, mainly due to the variety of events, including lightning-related transients. Regarding transient detection (waveform segmentation), the experiments involve usual segmentation methods, such as Kalman filtering, standard Discrete Wavelet Transform, and autoregressive models, besides two new techniques based on the Teager Energy Operator and Support Vector Data Description. The results obtained on both simulated and real world data demonstrate that the method based on Support Vector Data Description outperforms other methods in the transient identification task. Regarding the automatic waveform classification, a new approach including the detection of classes not defined in the training stage (called novelties) is presented. Also, the classifier is able to discriminate among multiple known classes, normally defined as multi-class classification. Two different approaches are compared, by using multi-class classification and novelty detection in two subsequent stages and in a simultaneous way. The following classifiers were assessed: X-Means,  $K$ -Nearest-Neighbors, and Support Vector Data Description with different formulations, besides the Support Vector Machine for multi-class classification. Furthermore, a technique for the post-processing of the novelties is presented, in order to provide some useful information for the experts, regarding possible similarities in the novelty set. To accomplish this task, automatic clustering methods were used. The final results, especially for the dataset with real examples, show that it is possible to obtain a relevant classification performance (above 80%) in each one of the three stages of the classification process: multi-class classification, novelty detection, and the post-processing applied to the novelties (automatic clustering).

**Keywords:** Novelty Detection, New Class Identification, Waveform Analysis, Waveform segmentation, Waveform classification.

## LISTA DE FIGURAS

FIGURA 1	– Visão geral do sistema elétrico de potência .....	17
FIGURA 2	– Reconhecimento de padrões com conjunto aberto .....	21
FIGURA 3	– Metodologia proposta e contribuições do trabalho .....	25
FIGURA 4	– Detalhamento da metodologia de classificação .....	26
FIGURA 5	– Detalhamento da nomenclatura da metodologia de classificação .....	27
FIGURA 6	– Resumo dos principais eventos discutidos .....	29
FIGURA 7	– Exemplos de interrupção de energia .....	31
FIGURA 8	– Exemplos de restabelecimento de energia .....	31
FIGURA 9	– Diferença no local de medição .....	32
FIGURA 10	– Origem das variações de tensão .....	32
FIGURA 11	– Curto-circuito não envolvendo o terra .....	33
FIGURA 12	– Curto-circuito envolvendo o terra .....	34
FIGURA 13	– Exemplo do diagrama fasorial de um curto-circuito bifásico para o terra	35
FIGURA 14	– Comparação entre dois curtos-circuitos de origem distinta .....	36
FIGURA 15	– Partida de um motor de indução de 800cv/4,16kV .....	36
FIGURA 16	– Energização de transformador .....	37
FIGURA 17	– Forma de onda resultante de um fechamento de um banco de capacitores	38
FIGURA 18	– Evolução do processo de descarga atmosférica .....	39
FIGURA 19	– Forma de onda da corrente de descarga atmosférica .....	39
FIGURA 20	– Forma de onda resultante de uma descarga atmosférica indireta .....	40
FIGURA 21	– Oscilografia simulada com diferentes estados .....	41
FIGURA 22	– Segmentação e disparo de uma oscilografia .....	42
FIGURA 23	– Cálculo dos resíduos para modelos autorregressivos .....	61
FIGURA 24	– Exemplo do disparo para modelos autorregressivos .....	62
FIGURA 25	– Exemplo de segmentação para modelos autorregressivos .....	63
FIGURA 26	– Exemplo de segmentação para modelos utilizando filtro de Kalman ....	66
FIGURA 27	– Exemplo de segmentação para modelos utilizando Operador de Energia de Teager .....	67
FIGURA 28	– Espectros Tempo-Frequência .....	70
FIGURA 29	– Decomposição do sinal através do banco de filtros .....	72
FIGURA 30	– Aproximações e detalhes .....	72
FIGURA 31	– Decomposição em um nível para 1024 amostras .....	73
FIGURA 32	– Recomposição em um nível para 512 coeficientes .....	73
FIGURA 33	– Recomposição em um nível para 512 coeficientes .....	73
FIGURA 34	– Exemplo de decomposição utilizando a TWPD .....	75
FIGURA 35	– Exemplo de detecção utilizando a TWD .....	76
FIGURA 36	– Hiperplano ótimo de separação .....	80
FIGURA 37	– Erro de classificação .....	81
FIGURA 38	– Extensão dos vetores suporte .....	86
FIGURA 39	– Exemplo de aplicação do modelo <i>K</i> -Médias com três centros .....	92
FIGURA 40	– Primeira etapa do método <i>X</i> -Médias .....	93
FIGURA 41	– Segunda etapa do método <i>X</i> -Médias .....	93



FIGURA 42	– Resumo do processo de seleção dos centros durante a execução do $X$ -Médias .....	94
FIGURA 43	– Modelo de representação dos dados utilizando uma hiperesfera .....	96
FIGURA 44	– Valores dos multiplicadores de Lagrange para diferentes entradas .....	98
FIGURA 45	– Variação dos parâmetros para o modelo RDVS .....	100
FIGURA 46	– Definição automática dos parâmetros para o modelo RDVS .....	101
FIGURA 47	– Composição dos diversos níveis da TWD com o modelo RDVS .....	103
FIGURA 48	– Visão geral da metodologia proposta para o cálculo de distância .....	105
FIGURA 49	– Exemplo de segmentação para modelos utilizando RDVS .....	106
FIGURA 50	– Formulação RDVS com exemplos negativos e otimização da margem ..	109
FIGURA 51	– Detecção de novidades e classificação multiclasse simultâneos .....	111
FIGURA 52	– Classificação multiclasse com um classificador uma-classe por classe ..	112
FIGURA 53	– Exemplo de uma análise ROC .....	115
FIGURA 54	– Exemplo de uma análise ROC para detecção de variações abruptas .....	116
FIGURA 55	– Exemplo de aplicação da métrica de validação externa baseada na matriz de confusão dos grupos .....	119
FIGURA 56	– Diagrama esquemático do modelo de simulação .....	122
FIGURA 57	– Esquemático do modelo equivalente elétrico da concessionária .....	122
FIGURA 58	– Esquemático do transformador da subestação .....	123
FIGURA 59	– Detalhe do sistema de monitoramento .....	134
FIGURA 60	– Metodologia de segmentação com base na estimação espectral .....	137
FIGURA 61	– Metodologia de segmentação com o Operador de Energia de Teager ...	138
FIGURA 62	– Metodologia de segmentação com base na TWD .....	139
FIGURA 63	– Metodologia de segmentação com base em RDVS .....	141
FIGURA 64	– Detalhe da análise ROC para diferentes limiares .....	142
FIGURA 65	– Metodologia de treinamento dos modelos aplicada aos dados simulados	143
FIGURA 66	– Metodologia de teste dos modelos aplicada aos dados simulados .....	145
FIGURA 67	– Metodologia de treinamento dos modelos aplicada aos dados reais .....	149
FIGURA 68	– Metodologia para agrupamento automático aplicada aos dados simulados	154
FIGURA 69	– Metodologia para agrupamento automático aplicada aos dados reais ...	155
FIGURA 70	– Análise ROC das 170 oscilografias simuladas .....	158
FIGURA 71	– Detalhe do método TWD+RVDS para descargas atmosféricas subseqüentes .....	162
FIGURA 72	– Detalhe do método de detecção TWD+RVDS para abertura subsequente a um curto-circuito .....	163
FIGURA 73	– Detalhamento da codificação apresentada para o pré-processamento ...	168
FIGURA 74	– Validação cruzada para comparação final .....	182
FIGURA 75	– Análise ROC para as cinco classes conhecidas .....	187
FIGURA 76	– Visualização em duas dimensões das cinco classes conhecidas previamente e as novidades utilizando o método <i>t-Distributed Stochastic Neighbor Embedding</i> .....	188
FIGURA 77	– Comparação de desempenho para diferentes métodos de agrupamento para as três classes de curto-circuito monofásico .....	192
FIGURA 78	– Comparação de desempenho para diferentes métodos de agrupamento para as todas as classes de curto-circuito .....	194
FIGURA 79	– Comparação de desempenho para diferentes arranjos de classes utilizando o método $K$ -Médias .....	196
FIGURA 80	– Comparação de desempenho para diferentes números de padrões por	

	classe e diferentes classes utilizando o método X-Médias .....	197
FIGURA 81 – Resultados de classificação, detecção de novidades e identificação de	classes .....	203

## LISTA DE TABELAS

TABELA 1	– Tipos de curto-circuito .....	33
TABELA 2	– Tabela de representação ROC (matriz de confusão) para um problema de duas classes .....	114
TABELA 3	– Matriz de confusão para um problema de quatro classes .....	117
TABELA 4	– Matriz de confusão para o problema de agrupamento .....	119
TABELA 5	– Tabela de parâmetros de cabos .....	127
TABELA 6	– Tabela de transformadores de distribuição .....	128
TABELA 7	– Eventos registrados pelo sistema de monitoramento .....	135
TABELA 8	– Desempenho do método de segmentação para dados simulados .....	158
TABELA 9	– Desempenho do método de segmentação para Apucarana-1 .....	160
TABELA 10	– Desempenho do método de segmentação para São Mateus-1 .....	163
TABELA 11	– Desempenho do método de segmentação para São Mateus-2 .....	164
TABELA 12	– Variação da ordem do Filtro de Kalman .....	166
TABELA 13	– Variação da ordem do modelo autorregressivo .....	166
TABELA 14	– Variação das funções <i>wavelet</i> para o modelo RDVS .....	166
TABELA 15	– Variação dos parâmetros para o modelo RDVS .....	167
TABELA 16	– Modelo RDVS com parâmetros ajustados automaticamente .....	167
TABELA 17	– Desempenho global (%) para diferentes pré-processamentos (abordagem DNG-CMC) .....	169
TABELA 18	– Desempenho global (%) para diferentes pré-processamentos (abordagem DNI-CBS) .....	169
TABELA 19	– Desempenho individual (%) do detector de novidades para a abordagem DNG-CMC (dados simulados) .....	170
TABELA 20	– Desempenho individual (%) do classificador multiclasse com base em MVS (dados simulados) .....	171
TABELA 21	– Desempenho global (%) para a abordagem DNG-CMC (dados simulados) .....	171
TABELA 22	– Resultado da classificação multiclasse para um arranjo contendo classes de curto-circuito monofásico e bifásico .....	172
TABELA 23	– Resultado da classificação multiclasse para um arranjo contendo classes de curto-circuito monofásico e bifásico incluindo detecção de novidades ..	173
TABELA 24	– Desempenho individual do detector de novidades para a abordagem DNI-CBS (dados simulados) .....	175
TABELA 25	– Desempenho global da abordagem DNI-CBS (dados simulados) .....	175
TABELA 26	– Desempenho individual (%) do detector de novidades para a abordagem DNG-CMC utilizando dados reais e simulados .....	179
TABELA 27	– Desempenho individual do classificador MVS utilizando dados reais e simulados .....	179
TABELA 28	– Desempenho global (%) para a abordagem DNG-CMC utilizando dados reais e simulados .....	179
TABELA 29	– Desempenho individual (%) do detector de novidades para a abordagem DNI-CBS utilizando dados reais e simulados .....	180

TABELA 30 – Desempenho global (%) da abordagem DNI-CBS utilizando dados reais e simulados .....	180
TABELA 31 – Matriz de confusão para a classificação multiclasse isolada com o modelo MVS e o pré-processamento <i>245.760-twd-db20</i> .....	183
TABELA 32 – Matriz de confusão para a classificação multiclasse isolada com o modelo MVS e o pré-processamento <i>245.760-twd-db20</i> com módulo e fase ..	183
TABELA 33 – Detecção de novidades e classificação multiclasse com a abordagem DNG-CMC .....	184
TABELA 34 – Detecção de novidades e classificação multiclasse com a abordagem DNI-CBS .....	184
TABELA 35 – Detecção de novidades e classificação multiclasse com a abordagem DNI-CBS com modelos baseados na RDVS .....	185
TABELA 36 – Matriz de confusão para o modelo RDVS-M .....	185
TABELA 37 – Desempenho por classe para o modelo RDVS-M .....	186
TABELA 38 – Resultado para o método RPEM com 200 padrões divididos em três grupos .....	195
TABELA 39 – Desempenho (%) do método de agrupamento automático <i>X-Médias</i> para diferentes arranjos de classes (dados simulados) .....	198
TABELA 40 – Desempenho (%) do método de agrupamento automático <i>X-Médias</i> para dados reais .....	199
TABELA 41 – Matriz de confusão para o pré-processamento <i>245.760-twd-db20</i> (com módulo e fase) para o agrupamento automático através do <i>X-Médias</i> .....	199

## LISTA DE SIGLAS

SEP	Sistema Elétrico de Potência
ANEEL	Agência Nacional de Energia Elétrica
QEE	Qualidade de Energia Elétrica
ATP	<i>Alternative Transient Program</i>
AR	Autorregressivo
RDVS	Representação de Dados com Vetores Suporte
OET	Operador de Energia de Teager
MVS	Máquina de Vetor Suporte
TWD	Transformada <i>Wavelet</i> Discreta
MUSIC	<i>Multiple Signal Classification</i>
ESPRIT	<i>Estimation of Signal Parameters via Rotational Invariance</i>
FK	Filtro de Kalman
TWPD	Transformada <i>Wavelet Packet</i> Discreta
KVMP	<i>K-Vizinhos-Mais-Próximos</i>
BIC	<i>Bayesian Information Criterion</i>
AG	Algoritmo Genético
ROC	<i>Receiver Operating Characteristic</i>
AAC	Área Abaixo da Curva
TAT	Transformador de Aterramento
BC	Banco de Capacitores
RINDAT	Rede Integrada Nacional de Detecção de Descargas Atmosféricas

## SUMÁRIO

<b>1 INTRODUÇÃO</b>	<b>16</b>
1.1 CENÁRIO ESCOLHIDO	18
1.2 OBJETIVO GERAL	23
1.3 OBJETIVOS ESPECÍFICOS	23
1.4 METODOLOGIA PROPOSTA	24
1.5 ORGANIZAÇÃO DO TEXTO	28
<b>2 EVENTOS ANALISADOS</b>	<b>29</b>
2.1 INTERRUPÇÕES	29
2.2 VARIAÇÕES DE TENSÃO	32
2.2.1 Curto-Circuito	32
2.2.2 Partida de Cargas Especiais	35
2.2.3 Energização de Transformadores	36
2.3 TRANSITÓRIOS	37
2.4 OBSERVAÇÕES FINAIS	40
<b>3 REVISÃO BIBLIOGRÁFICA</b>	<b>42</b>
3.1 DETECÇÃO DE EVENTOS	42
3.2 CLASSIFICAÇÃO DE EVENTOS	47
3.2.1 Classificação de Eventos em SEP	47
3.2.2 Detecção de Novidades	50
3.2.3 Detecção de Novidades e Classificação Multiclasse	51
3.2.4 Identificação de Novas Classes	52
3.2.5 Agrupamento Automático	53
3.2.6 Classificação Multiclasse e Identificação de Novas Classes	54
3.2.7 Identificação de Novas Classes em Fluxo Contínuo de Dados	56
<b>4 FUNDAMENTAÇÃO TEÓRICA</b>	<b>59</b>
4.1 MODELOS AUTORREGRESSIVOS	60
4.2 FILTRO DE KALMAN	63
4.3 OPERADOR DE ENERGIA DE TEAGER	65
4.3.1 Determinação Automática do Limiar para o OET	67
4.4 PRÉ-PROCESSAMENTO	68
4.4.1 Transformada <i>Wavelet</i>	68
4.4.2 Transformada <i>Wavelet Packet</i> Discreta	75
4.4.3 Aplicação para Detecção de Eventos	75
4.5 EXTRAÇÃO DE CARACTERÍSTICAS PARA CLASSIFICAÇÃO	77
4.6 MÓDULO E FASE DO SINAL	78
4.7 CLASSIFICADORES SUPERVISIONADOS MULTICLASSE	79
4.7.1 Máquina de Vetor Suporte	79
4.7.2 Aplicação em Padrões não Linearmente Separáveis	80
4.7.3 Seleção Automática dos Hiperparâmetros	85
4.8 CLASSIFICADORES UMA-CLASSE	87
4.8.1 <i>K</i> -Vizinhos-Mais-Próximos	89

4.8.2	<i>K</i> -Médias	90
4.8.3	<i>X</i> -Médias	90
4.8.4	Detecção de Novidades com o <i>K</i> -Médias e <i>X</i> -Médias	95
4.8.5	Representação de Dados com Vetores Suporte	95
4.8.6	Determinação Automática dos Parâmetros de Treinamento	99
4.8.7	Representação de Dados com Vetores Suporte para Detecção de Eventos	102
4.8.8	RDVS com Exemplos Negativos	106
4.8.9	RDVS com Exemplos Negativos e Otimização da Margem	109
4.8.10	Abordagem para Classificação Multiclasse	111
4.9	ANÁLISE ESTATÍSTICA	113
4.9.1	Análise ROC	113
4.9.2	Matriz de Confusão para mais de Duas Classes	116
4.9.3	Análise de Desempenho dos Métodos de Agrupamento	117
4.9.4	Comparação Estatística entre Desempenhos	119
<b>5</b>	<b>METODOLOGIA</b>	<b>121</b>
5.1	SISTEMA SIMULADO	121
5.1.1	Modelo Equivalente da Concessionária	122
5.1.2	Transformador da Subestação	123
5.1.3	Transformador de Aterramento	124
5.1.4	Bancos de Capacitores	125
5.1.5	Motores de Indução	125
5.1.6	Linhas Aéreas	126
5.1.7	Transformadores de Distribuição e Cargas	127
5.1.8	Eventos Gerados para os Métodos de Segmentação	129
5.1.9	Eventos Gerados para Classificação e Identificação de Novas Classes	130
5.2	SISTEMA DE MONITORAMENTO (DADOS REAIS)	132
5.3	SEGMENTAÇÃO DE EVENTOS	136
5.3.1	Modelos de Estimção Espectral	136
5.3.2	Modelos Baseados no Operador de Energia de Teager	137
5.3.3	Modelos Baseados Apenas na Transformada <i>Wavelet</i>	138
5.3.4	Modelos Baseados em TWD+RDVS	139
5.3.5	Análise Estatística para Segmentação de Eventos	140
5.4	CLASSIFICAÇÃO DE EVENTOS E DETECÇÃO DE NOVIDADES	142
5.4.1	Dados Simulados	143
5.4.2	Parâmetros Ajustáveis dos Dados Simulados	146
5.4.3	Dados Reais	148
5.4.4	Parâmetros Ajustáveis dos Dados Reais	151
5.5	AGRUPAMENTO AUTOMÁTICO E IDENTIFICAÇÃO DE NOVAS CLASSES	152
5.5.1	Dados Simulados	153
5.5.2	Parâmetros Ajustáveis dos Dados Simulados	154
5.5.3	Dados Reais	155
5.5.4	Parâmetros Ajustáveis dos Dados Reais	155
5.6	PACOTES DE FUNÇÕES UTILIZADAS	156
<b>6</b>	<b>RESULTADOS</b>	<b>157</b>
6.1	SEGMENTAÇÃO DE EVENTOS	157
6.1.1	Base de Dados Simulados	157
6.1.2	Base de Dados Reais	159
6.1.3	Definição dos Parâmetros	165

6.2 CLASSIFICAÇÃO E DETECÇÃO DE NOVIDADES EM DADOS SIMULADOS	167
6.2.1 Análise Inicial dos Parâmetros para os Dados Simulados	168
6.2.2 Abordagem DNG-CMC	170
6.2.3 Desempenho Individual do Detector de Novidades	170
6.2.4 Desempenho Individual do Classificador Multiclasse	171
6.2.5 Desempenho Global	171
6.2.6 Abordagem DNI-CBS	174
6.2.7 Desempenho Individual do Detector de Novidades	174
6.2.8 Desempenho Global	175
6.2.9 Conclusões Iniciais para os Dados Simulados	175
6.3 CLASSIFICAÇÃO E DETECÇÃO DE NOVIDADES EM DADOS REAIS	177
6.3.1 Análise Inicial dos Parâmetros para os Dados Reais	178
6.3.2 Abordagem DNG-CMC	178
6.3.3 Abordagem DNI-CBS	179
6.3.4 Análise de Melhoria de Desempenho para os Dados Reais	180
6.3.5 Conclusões para os Dados Reais	189
6.4 IDENTIFICAÇÃO DE NOVAS CLASSES	190
6.4.1 Seleção do Método de Agrupamento	191
6.4.2 Resultados para os Dados Simulados	197
6.4.3 Resultados para os Dados Reais	198
<b>7 CONCLUSÕES</b>	<b>201</b>
7.1 TRABALHOS FUTUROS	205
<b>REFERÊNCIAS</b>	<b>207</b>
<b>Apêndice A – PUBLICAÇÕES RELACIONADAS</b>	<b>219</b>



## 1 INTRODUÇÃO

A análise de oscilografias<sup>1</sup> consiste em uma ferramenta de suma importância para a distinção correta de eventos que alteram as formas de onda de tensão e corrente nos diversos estágios do sistema elétrico de potência. Essa análise fornece uma base para o processo de mitigação, manutenção e localização de falhas, formando um sistema de auxílio à tomada de decisões por parte das concessionárias de energia (KINDERMANN, 2007).

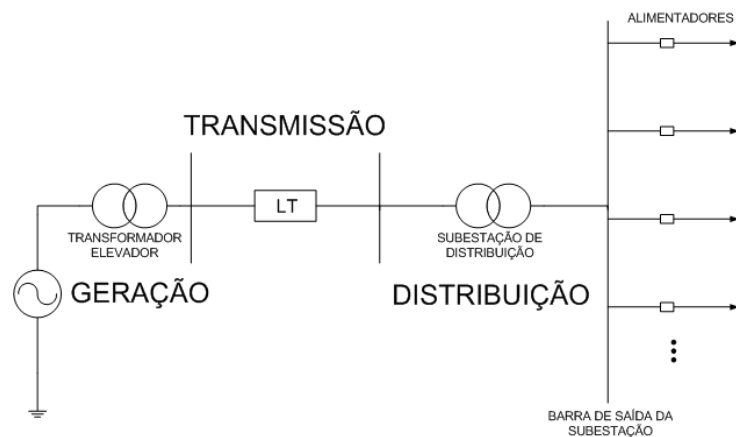
Em um sistema elétrico de potência (SEP) existem diversos fenômenos responsáveis por alterações nas formas de onda de tensão e corrente tanto na geração, quanto nas linhas de transmissão (LT) e nas redes de distribuição, dentre os quais destacam-se os curtos-circuitos. Em um SEP típico (Figura 1), Dugan e colaboradores (2002) mostram que 31% das faltas que causam falhas de equipamentos de um consumidor industrial ou danos aos consumidores ocorrem no sistema de transmissão, sendo o restante nas redes de distribuição. Dos 69% restantes, que também provocam falhas ou danos, 23% ocorrem no próprio alimentador onde os equipamentos ou consumidores estão instalados, enquanto 46% ocorrem em alimentadores paralelos ao alimentador dos equipamentos danificados.

Uma das principais causas dos curtos-circuitos está associada aos transitórios decorrentes de descargas atmosféricas (DUGAN et al., 2002). Pode-se afirmar que as faltas produzidas em decorrência das descargas atmosféricas são, para sistemas de distribuição aéreos, a causa preponderante das interrupções e dos desligamentos, afetando, com isso, os índices de continuidade. Em geral, estima-se que cerca de 35% dos desligamentos são atribuídos a descargas atmosféricas para as concessionárias distribuidoras de energia (BOLLEN; GU, 2006).

Além de afetar os índices de desempenho do sistema, as descargas atmosféricas provocam danos ou, no mínimo, uma redução considerável da vida útil nos equipamentos da concessionária e dos consumidores. Atualmente, as concessionárias de distribuição apresentam taxa de falhas de transformadores de distribuição da ordem de 1,5% ao ano, sendo atribuída a cifra de 30% para as falhas causadas pelos efeitos diretos ou indiretos das descargas atmosféricas (LAZ-

---

<sup>1</sup>Oscilografias podem ser definidas como um registro da forma de onda de tensão ou corrente amostradas a uma taxa constante em um determinado intervalo de tempo (DUGAN et al., 2002).



**Figura 1: Visão geral do sistema elétrico de potência. Convém ressaltar que esse formato do SEP ainda é realizado em grande parte do Brasil. Com o advento das redes inteligentes e a popularização da geração distribuída, é possível que esse cenário seja alterado principalmente na distribuição, em função da inserção de unidades geradoras em unidades consumidoras.**

ZARETTI et al., 2011). Fora isso, existe o problema da compensação que a empresa deve pagar aos consumidores que reclamam por danos ocorridos em seus equipamentos.

No caso particular das oscilografias de tensão em um sistema de distribuição, objeto de pesquisa deste trabalho, existe uma gama de eventos que também apresenta impacto relevante em relação às falhas de equipamentos ou danos a consumidores, uma vez que a sua ocorrência altera as formas de onda de tensão e corrente na rede e, em função disso, sua correta identificação se faz necessária. Dentre esses eventos é possível destacar (DUGAN et al., 2002):

- Transitórios decorrentes de chaveamentos e manobras, como a abertura e fechamento de chaves ao longo do alimentador e a manobra em bancos de capacitores com potência significativa instalados na rede ou nas subestações, cujas sobretensões podem chegar a níveis críticos do ponto de vista de suportabilidade de equipamentos.
- Partida de cargas significativas em um alimentador, como grandes motores ou um grande bloco de carga (indústria), que provocam afundamentos de tensão de longa duração.

Além do aspecto técnico, existe um aspecto financeiro que vem promovendo 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. No Brasil, encontra-se em vigor uma série de procedimentos voltados à conformidade desses índices, em consenso às normas internacionais. Um procedimento que merece destaque aqui é o módulo 8 dos Procedimentos de Distribuição da Agência Nacional de Energia

Elétrica, ANEEL (ANEEL, 2008). Esse procedimento descreve os principais limites referentes à qualidade do serviço e do produto em redes de distribuição, cuja extrapolação pode incorrer em multas por parte dos órgãos reguladores das concessionárias de energia.

Frente à restrição severa em relação aos limites a serem atendidos, as concessionárias de energia vêm desenvolvendo uma série de medidas visando permitir um amplo monitoramento das suas redes e identificação das variações de forma de onda que possam ocorrer, buscando classificar os eventos ocorridos. Esse processo de classificação é fundamental e existem diversos benefícios que podem ser apontados na aplicação do mesmo. O primeiro deles pode ser associado à identificação de causa do evento (XU et al., 2007b). A construção de uma base de dados de eventos devidamente classificados, em especial de curtos-circuitos, pode viabilizar a análise do agente causador da oscilografia registrada, ou seja, as características transitórias dos curtos-circuitos podem apontar para a causa desse curto-circuito, distinguindo faltas geradas por problemas mecânicos (KINDERMANN, 2007) de faltas ocasionadas por elementos do sistema de distribuição (por exemplo, falhas em transformadores), ou faltas relacionadas com descargas atmosféricas.

Com relação à correta identificação de descargas atmosféricas, é possível determinar os níveis típicos das sobretensões ocorridas, bem como auxiliar na mitigação de eventos associados a esse fenômeno. Já para os demais eventos, pode-se prever um auxílio na redução dos impactos de manobras, além de subsidiar decisões relacionadas com instalações críticas do ponto de vista de qualidade de energia elétrica (QEE)<sup>2</sup>.

## 1.1 CENÁRIO ESCOLHIDO

Com as necessidades e os benefícios de uma correta identificação dos eventos em redes de distribuição devidamente apontados, optou-se neste trabalho por utilizar como base para testes e validação da metodologia, o projeto proposto pela Companhia Paranaense de Energia (COPEL) em conjunto com o Instituto de Tecnologia para o Desenvolvimento (Lactec), dentro do programa de Pesquisa e Desenvolvimento da ANEEL, intitulado *Instalação Piloto para Avaliação dos Níveis de Sobretensão Atmosférica no Sistema da COPEL Distribuição*. O objetivo desse projeto foi o desenvolvimento de um sistema automático para a medição e registro de fenômenos eletromagnéticos rápidos em redes de distribuição de energia em média e baixa

---

<sup>2</sup>Qualidade de Energia Elétrica, segundo Dugan et al. (2002), diz respeito a qualquer desvio associado à tensão, corrente ou frequência nominais que resulte em danos ou falhas em equipamentos de consumidores de energia elétrica. O termo QEE tornou-se 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.

tensão e sua utilização no monitoramento do sistema de distribuição da COPEL, em especial alimentadores de grande extensão, situados em locais com grande incidência de descargas atmosféricas.

Durante a primeira fase do referido projeto (LAZZARETTI et al., 2011) foram instalados quatro protótipos do sistema de monitoramento, sendo dois deles em Apucarana-PR e dois em São Mateus do Sul-PR. Em aproximadamente um ano de operação desse sistema na rede de distribuição, observou-se a ocorrência de 149 variações significativas (variações abruptas ou transitórias) nas formas de onda, sendo que aproximadamente 57% dessas variações estão relacionadas às descargas atmosféricas, 17% a eventos de curto-circuito, 8% a manobras no alimentador e 18% a outros eventos, dentre os quais destacam-se os eventos cuja caracterização foi inviabilizada em função da complexidade na análise e a falta de informação sobre as condições que geraram o registro do evento, como será descrito posteriormente.

De um modo geral, a grande dificuldade encontrada na classificação desse último grupo de eventos é a própria complexidade na análise de cada forma de onda, considerando a possibilidade de ocorrência de eventos simultâneos, sequenciais e até mesmo eventos ocorridos em outros estágios do SEP, cujos reflexos são registrados pelos sistemas de monitoramento. Adicionalmente, em alguns casos não foram obtidas informações da concessionária sobre os eventos ocorridos e estado de conexão da rede de distribuição no momento do registro dos eventos, dificultando a interpretação dos fenômenos envolvidos.

Para um especialista realizar a classificação de cada forma de onda registrada no período foram necessárias em média duas horas de análise. Essa análise considerava duas etapas distintas: a segmentação da forma de onda em diferentes instantes transitórios de análise e a classificação propriamente dita. A classificação considerou que todas as informações necessárias para se obter uma caracterização confiável dos eventos já estavam disponíveis no momento da análise, já que, para efetuar a classificação, foram utilizados outros dados além da própria forma de onda, tais como:

- Localização e caracterização de descargas atmosféricas ocorridas próximas aos sistemas de monitoramento;
- Sistema de Ocorrências da COPEL Distribuição, que fornecia um provável agente causador do evento (em especial para desligamentos);
- Simulações auxiliares no *Alternative Transient Program*.

Esse cenário sugere que o desenvolvimento de um classificador automático de oscilo-

grafias com identificação automática dos instantes transitórios forneceria uma grande economia no tempo aplicado para correta identificação dos eventos. Seria importante também que esse classificador fosse capaz de identificar cada evento de acordo com a forma de onda registrada, sem a necessidade de outras bases de dados. Essa característica permitiria que o processo de classificação pudesse ser feito em tempo real<sup>3</sup>, já que alguns dados utilizados na classificação levam mais de um dia para serem disponibilizados ao especialista.

Adicionalmente, ao longo do tempo, novas classes de eventos podem ser armazenadas pelos registradores. Nesse contexto, pode-se considerar o exemplo de uma subestação de médio porte que esteja operando com carga bastante reduzida em relação à sua capacidade nominal e sem a necessidade de bancos de capacitores para o controle de fator de potência e queda de tensão nesse sistema. É de se esperar que, à medida que novas cargas vão sendo inseridas no sistema, exista a necessidade da instalação de bancos de capacitores na rede, principalmente em função de características específicas provenientes de um novo grupo de consumidores. A partir do momento em que é feita a instalação desse banco, os registradores já instalados nessa rede passam a ser susceptíveis às manobras nesse banco, podendo efetuar o armazenamento das formas de onda decorrentes dessas manobras<sup>4</sup>.

O conjunto de registros obtidos por esses sistemas reflete um cenário interessante e comum a todo registrador que venha a ser utilizado para classificação de eventos em redes de distribuição. Uma classificação automática de eventos para um registrador deve considerar a identificação (classificação) das classes de eventos já registradas, bem como identificar novas classes de eventos que possam surgir ao longo do tempo em que o registrador se encontra instalado, ou mesmo registros espúrios que possam ser observados no período. Esse processo de identificação de novas classes é normalmente denominado de detecção de novidades (HODGE; AUSTIN, 2004). No contexto deste trabalho, uma forma de onda pode ser caracterizada como novidade nos seguintes cenários:

- Falta de registros prévios para caracterização desses eventos. Nesse caso, o especialista pode saber da existência do evento, tendo até já simulado o evento computacionalmente. Porém, não existem registros reais da ocorrência desse evento<sup>5</sup>, como demonstrado em Lazzaretti et al. (2011);

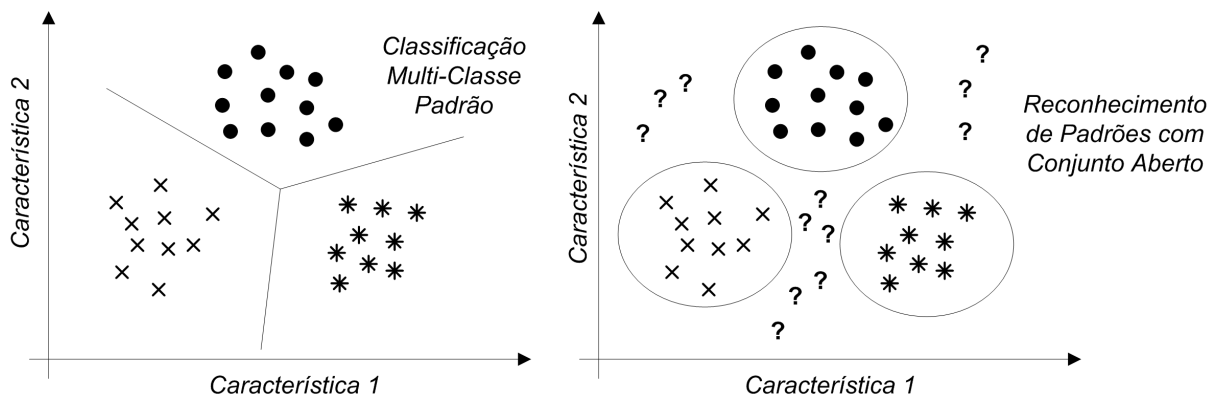
<sup>3</sup>Tempo Real no contexto deste trabalho deve ser compreendido como o tempo necessário para processar a classificação automática do evento, depois de armazenado o registro da forma de onda.

<sup>4</sup>Nesse caso, considerando que não existam informações prévias sobre o comportamento do chaveamento de um banco de capacitores em um sistema de distribuição, para o especialista responsável por classificar as formas de onda.

<sup>5</sup>Sabe-se que determinados eventos ainda não possuem uma modelagem consolidada na literatura, a exemplo de tensões induzidas por descargas atmosféricas indiretas (BOLLEN; GU, 2006).

- Dificuldade na interpretação do fenômeno. Podem existir eventos, tais como os descritos no fim do Capítulo 2, que possuem uma identificação e caracterização bastante complexa. Esses registros incluem eventos simultâneos e subsequentes. Também é possível incluir nesse conjunto, eventos que ocorrem em outros estágios do SEP, principalmente na transmissão;
- Desconhecimento com relação à ocorrência de um evento com a forma de onda registrada. O exemplo do chaveamento do banco de capacitores, colocado aqui de forma didática, ilustra esse cenário. No entanto, com o conjunto de mudanças que estão sendo feitas no setor de distribuição, é de se esperar que possam existir um conjunto grande de eventos transitórios desconhecidos atualmente, principalmente relacionados com entrada (ou saída) de unidades geradoras distribuídas no sistema.

Esse tipo de classificação é normalmente verificado em problemas de reconhecimento de padrões com conjunto aberto (SCHEIRER et al., 2013), onde verifica-se um conhecimento incompleto do problema durante o estágio de treinamento e novas classes podem aparecer durante a fase de teste, conforme ilustra a Figura 2.



**Figura 2: Reconhecimento de padrões com conjunto aberto. Em abordagens usuais de classificação multiclasse, os exemplos de treinamento e teste estão relacionados com classes conhecidas *a priori*. Em problemas de reconhecimento de padrões com conjunto aberto, existe um conjunto de classes conhecidas, porém existe a possibilidade de inúmeras classes desconhecidas previamente.**

Uma vez identificada uma forma de onda ou um conjunto de oscilografias como novidade, o problema passa a ser como proceder com esses padrões. A proposta deste trabalho é permitir que algumas informações possam ser retiradas desse conjunto de dados, fornecendo um procedimento de auxílio na identificação de novas classes, facilitando a inclusão desse conhecimento no classificador proposto.

A dificuldade desse procedimento é que, em princípio, não existirão informações disponíveis sobre os eventos caracterizados como novidades, uma vez que o especialista res-

ponsável pela rotulação (identificação) dos padrões não terá conhecimento consolidado sobre esses padrões.

Nesse sentido, a utilização de métodos de aprendizado supervisionado, ou mesmo semisupervisionado (SCHÖLKOPF; SMOLA, 2001), é inviabilizada no contexto proposto. Uma alternativa para prover um auxílio ao especialista no processo de identificação e, até mesmo, caracterizar novas classes presentes nos dados classificados como novidades, é aplicar modelos de agrupamento automático (FILIPPONE et al., 2008). Esses modelos, caracterizados por serem fundamentalmente não-supervisionados, permitem que sejam identificados e agrupados eventos com certo grau de similaridade, dentro do conjunto de padrões definidos como novidades.

Voltando ao exemplo da subestação de médio porte operando sem banco de capacitores citado anteriormente, é de se esperar que à medida que eventos transitórios relacionados à manobra desse banco sejam continuamente registrados (o evento considerado como novidade seja continuamente apresentado), além de identificar a existência dessa novidade<sup>6</sup>, é necessário identificar que todos os eventos de manobra do banco de capacitores possuem características comuns, e que esses eventos devem ser agrupados de acordo com essas características, facilitando o processo de identificação dessa classe de eventos por parte do especialista.

Nesse ponto fica claro que o processo de agrupamento automático é apenas uma parte do processo de identificação de classes. Ou seja, os padrões agrupados por similaridade dão um indício de que um determinado conjunto de eventos pode formar uma classe. No entanto, é necessário que o especialista analise essa informação e, através do uso de outras fontes de dados, confirme que aquele conjunto de eventos corresponda efetivamente a uma classe. No caso da manobra do banco de capacitores, o especialista poderia receber a informação dos horários de manobra do banco e, verificando que todos os eventos registrados naquele horário fazem parte do mesmo grupo de acordo com o agrupamento automático aplicado, inferir que aquele conjunto de eventos faz parte de uma classe denominada *Manobra de Banco de Capacitores*. Com isso, à medida que novas classes são identificadas e confirmadas por especialistas, será possível incorporar esse conhecimento ao classificador, conferindo um aspecto dinâmico ao processo de aprendizagem na metodologia aqui proposta.

Convém ressaltar novamente que esse exemplo dos transitórios decorrentes de manobras em banco de capacitores pode ser estendido para diferentes cenários, principalmente para as novas redes de distribuição inteligentes (SIMARD et al., 2009). Nessas redes, a quantidade de novos eventos transitórios a serem registrados irá depender fortemente da complexidade que

---

<sup>6</sup>Considerando que, nesse caso, a manobra de um banco de capacitores é caracterizado como uma novidade por um dos três motivos apresentados: falta de registros prévio para sua caracterização, dificuldade na interpretação do fenômeno ou desconhecimento com relação à ocorrência desse evento com a forma de onda registrada.

essa rede irá ter, em função da conexão com geração distribuída, novas formas de operação do sistema, ou mesmo novas cargas conectadas no sistema. Nesse sentido, a detecção de novidades e a identificação de novas classes podem se tornar parte fundamental do processo de classificação, sendo indispensáveis para correta identificação de eventos transitórios em redes de distribuição. A possibilidade de implantação de registradores oscilográficos com capacidade de classificar registros de forma automática e adaptativa, somada à concentração da informação em centros de operação (DAVIDSON et al., 2006), fornece uma autonomia para os operadores e responsáveis pela tomada de decisões nesses centros, de forma que o tempo dispendido na mitigação e manutenção da falha sejam reduzidos, contribuindo para a melhoria dos índices de continuidade da concessionária (ANEEL, 2008).

Com esse cenário exposto, pode-se apresentar os objetivos do trabalho.

## 1.2 OBJETIVO GERAL

Desenvolver e validar um método que permita a detecção e a classificação de instantes transitórios em oscilografias. A classificação deverá ser capaz de identificar e tratar eventos caracterizados como novidades (registros não conhecidos *a priori*), visando conferir ao método uma característica automática e dinâmica no estágio de classificação.

## 1.3 OBJETIVOS ESPECÍFICOS

- Modelar e simular, utilizando o software *Alternative Transient Program* (ATP), um sistema de distribuição de energia reduzido, considerando as principais características existentes nesse sistema, de forma que esse modelo sirva de base para validação inicial da metodologia proposta. Vale ressaltar que o modelo utilizado aqui se baseia nos desenvolvimentos inicialmente propostos por Lazzaretti (2009).
- Levantar os principais eventos a serem simulados, de forma que o modelo possa ser ajustado para atender às necessidades dos vários eventos.
- Desenvolver um método de detecção de eventos (ou segmentação da oscilografia) que possa ser aplicado para caracterização do início e duração dos transitórios entre eventos de uma oscilografia, facilitando em um primeiro momento, o processo de classificação.
- Analisar como os sinais serão processados, visando à escolha da melhor e mais compacta forma de extrair as características dos sinais sob análise, permitindo que o classificador obtenha desempenho adequado para o problema em questão.



- Pesquisar, avaliar e propor técnicas de detecção de novidades que possam ser empregadas em conjunto com a classificação supervisionada, fornecendo a flexibilidade de identificar novos padrões pelo classificador.
- Pesquisar e avaliar técnicas de agrupamento automático, visando fornecer um método que auxilie o especialista a identificar novas classes presentes nos padrões classificados como novidades.
- Aplicar o método desenvolvido em um primeiro momento nos dados simulados e, em seguida, nos dados reais, verificando seu desempenho dentro do escopo apresentado.

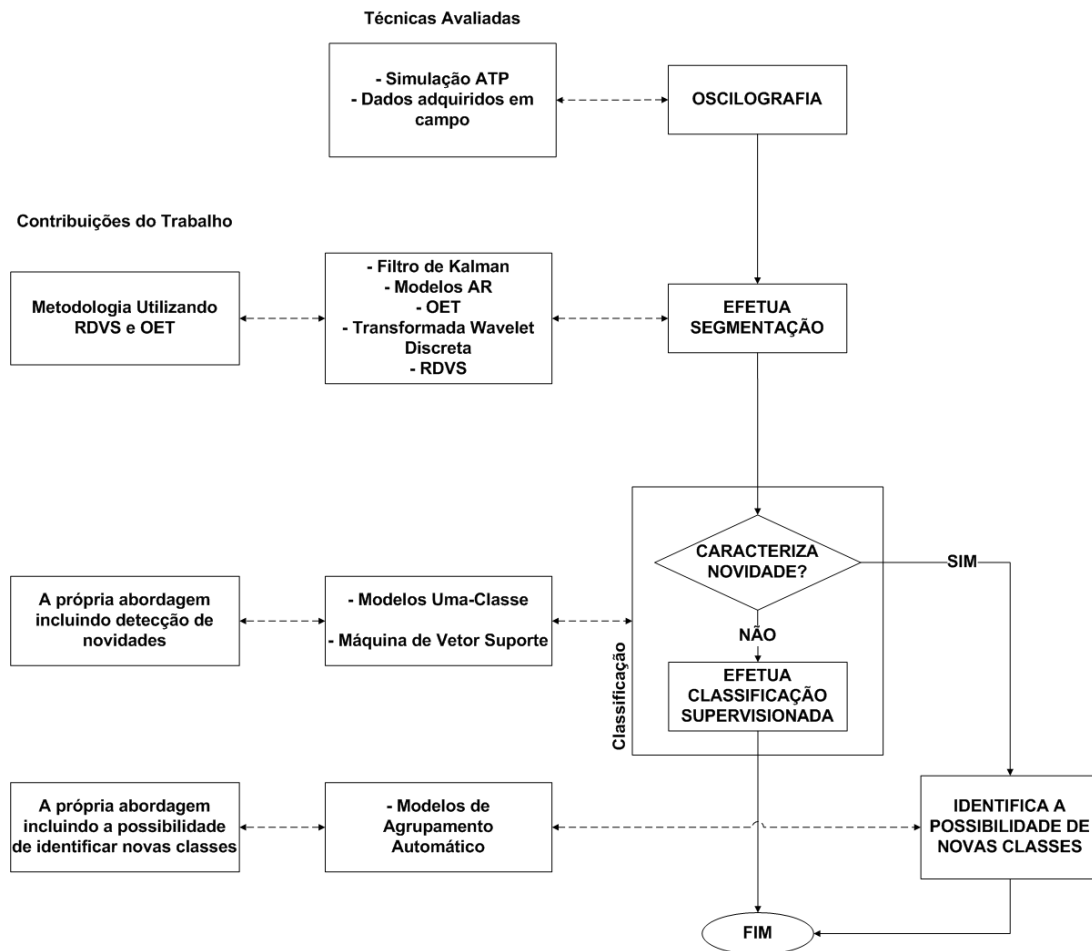
#### 1.4 METODOLOGIA PROPOSTA

Com o intuito de facilitar a compreensão da metodologia proposta, bem como as principais contribuições do trabalho, optou-se por apresentar um resumo inicial de cada módulo da metodologia separadamente. A Figura 3 mostra um fluxograma da metodologia aplicada neste trabalho.

O primeiro estágio desse fluxograma está associado à obtenção do oscilograma. Conforme comentado anteriormente, os dados utilizados neste trabalho foram gerados por um conjunto de simulações realizadas no ATP, bem como dados reais registrados em campo, cujos resultados são apresentados em Lazzaretti et al. (2011). O grande diferencial nesse estágio é a utilização de dados reais para validação da metodologia, que ainda é pouco observada para classificação de eventos em redes de distribuição, conforme apresenta o Capítulo 3. Além disso, ressalta-se a inclusão da análise de transitórios decorrentes de descargas atmosféricas, assunto que também é muito pouco explorado atualmente e ainda não possui métodos de simulação validados.

Na sequência, propõe-se a avaliação da oscilografia, buscando identificar instantes transitórios com alteração significativa na forma de onda e, por isso, devem ser posteriormente classificados. Essa etapa é conhecida como segmentação ou detecção dos eventos. Nesse estágio foram avaliados métodos baseados nos resíduos mais proeminentes de modelos de estimação espectral<sup>7</sup>, como o Filtro de Kalman e modelos autorregressivos (AR), técnica de reconhecimento de padrões baseada em Representação de Dados Utilizando Vetores Suporte (RDVS), Transformada *Wavelet* Discreta e o Operador de Energia de Teager (OET). A

<sup>7</sup>Neste trabalho, os modelos autorregressivos e baseados em Filtro de Kalman serão denominados como modelos de estimação espectral, pois são utilizados com essa finalidade na sua formulação teórica, apresentada em Bollen e Gu (2006). No entanto, é importante ressaltar que esses modelos operam fundamentalmente no domínio temporal.



**Figura 3: Metodologia proposta e contribuições do trabalho.**

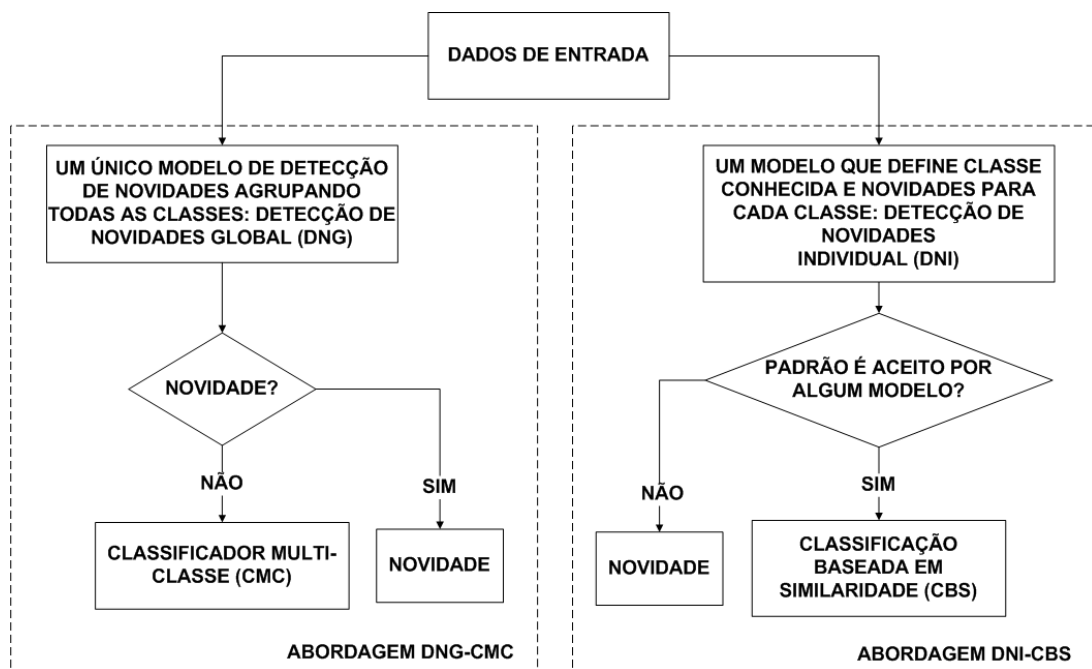
contribuição nesse estágio é a metodologia de carácter original desenvolvida para detecção de eventos através da RDVS. Esse modelo, quando combinado com um pré-processamento baseado em Transformada *Wavelet*, promove uma representação que permite a detecção de eventos com grande desempenho, quando comparada com as técnicas atualmente empregadas. Além disso, o método baseado em OET também demonstra que, através de uma técnica bastante simples, é possível realizar a detecção de instantes transitórios de sinais de forma relativamente eficaz.

Em seguida, inicia-se o processo de classificação, sendo esse dividido em duas etapas principais: classificação multiclasse e detecção de novidades. Na classificação multiclasse foi avaliado o classificador multiclasse baseado em Máquina de Vetor Suporte (MVS), apresentado em (LAZZARETTI, 2009). Naquele trabalho, demonstrou-se que esses modelos possuem grande desempenho quando aplicados na classificação de eventos em redes de distribuição.

Com relação à detecção de novidades, optou-se por avaliar o desempenho de classificadores uma-classe. Caso o evento detectado contenha características não relacionadas com as

características dos eventos previamente selecionados no modelo de aprendizagem supervisionado (conhecimento *a priori*), faz-se necessária essa identificação para posterior análise. Essa identificação caracteriza uma detecção de novidades. Para tanto, os seguintes classificadores uma-classe foram avaliados: Agrupamento X-Médias (que utiliza fundamentalmente o modelo *K*-Médias), Agrupamento utilizando os *K*-Vizinhos Mais Próximos e a própria Representação de Dados Utilizando Vetores Suporte com diferentes formulações.

Pode-se resumir o processo global de classificação (bloco assinalado na Figura 3 – detecção de novidades e classificação multiclasse) utilizando o modelo MVS e classificadores uma-classe em dois cenários distintos, conforme mostra a Figura 4.



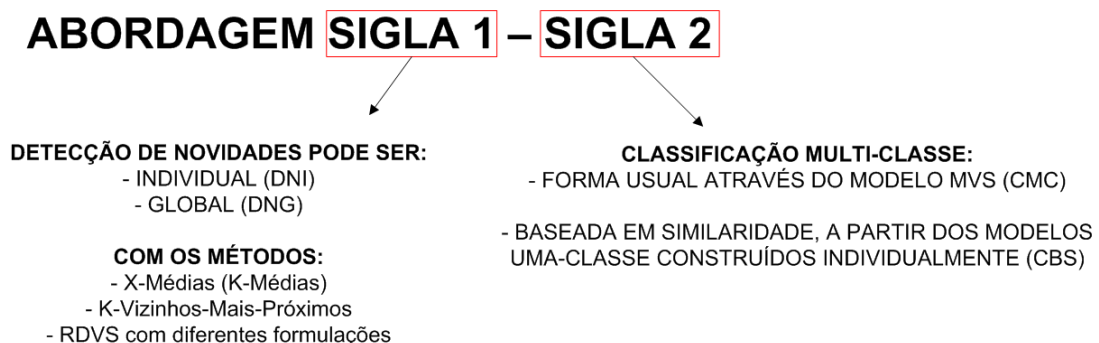
**Figura 4: Detalhamento da metodologia de classificação.**

No primeiro cenário (DNG-CMC), o detector de novidades é construído agrupando todas as classes conhecidas do problema. Sobre o mesmo conjunto de classes também é modelado o classificador multiclasse. No estágio de avaliação, resumidamente demonstrado na esquerda da Figura 4, um padrão é inicialmente comparado ao detector de novidades global; caso não configure uma novidade, é utilizado o classificador multiclasse para definir a classe que esse padrão pertence.

De forma similar, o detector de novidades poderia ser construído para cada classe conhecida do problema, ao invés de ser aplicado para todas as classes. Com isso, a detecção de novidades seria realizada individualmente, de forma simultânea (ou conjunta) à classificação multiclasse que, neste caso, é feita através de uma medida de similaridade entre o padrão a ser classificado e cada um dos modelos uma-classe estimados. É importante ressaltar que nessa

abordagem se utiliza apenas os modelos uma-classe, sendo esses aplicados à detecção de novidades e à classificação multiclasse. Essa modificação resulta na abordagem denominada DNI-CBS, representada à direita da Figura 4.

Na Figura 5 são descritas as nomenclaturas utilizadas em cada uma das duas abordagens de classificação. De um modo geral, a primeira sigla demonstra a forma como é feita a detecção de novidades, podendo ser realizada de forma global, agrupando todas as classes para definir o modelo uma-classe, ou aplicada individualmente para cada classe conhecida do problema. A segunda sigla corresponde à forma como é feita a classificação multiclasse, que neste trabalho será feita utilizando apenas o modelo MVS.



**Figura 5: Detalhamento da nomenclatura da metodologia de classificação.**

Essas abordagens serão mais bem detalhadas ao longo do desenvolvimento deste documento, focando em como é construída a etapa de treinamento e a etapa de teste de cada abordagem proposta.

Vale salientar que, para classificação multiclasse e detecção de novidades, é necessário existir um conjunto mínimo de dados previamente rotulados para sua utilização. As vantagens e desvantagens dessa necessidade serão discutidas ao longo do trabalho.

A última etapa da metodologia está associada com a identificação de novas classes. Conforme citado anteriormente, a proposta deste trabalho é utilizar o agrupamento automático como parte do processo de identificação de classes. Uma vez aplicado o agrupamento automático, é possível indicar quais eventos são similares entre si, fornecendo ao especialista informações sobre a possibilidade de existência de novas classes em uma distribuição de dados previamente desconhecida.

Uma questão importante é que o agrupamento deve ter uma característica automática, principalmente com relação ao número de grupos e suas características, evitando ao máximo a interferência do especialista nesse processo, uma vez que o resultado do agrupamento automático dado pelo número de grupos e suas características é o que representa o conhecimento

que estará disponível para o especialista sobre os padrões definidos como novidades.

No instante em que o agrupamento é aplicado, pode existir mais do que apenas uma classe no conjunto de dados caracterizados como novidades. Nesse caso, vale ressaltar que a definição do número de grupos e suas características não é um indicativo sobre o número de classes existentes, pois é possível (e bastante provável) que existam classes formadas por mais de um grupo. É importante esclarecer que a separação de padrões (classificação) é muito diferente do agrupamento de padrões proposto. Por essa razão, o agrupamento automático é apenas um indicativo inicial da existência de novas classes e necessita da intervenção do especialista para confirmação das classes existentes.

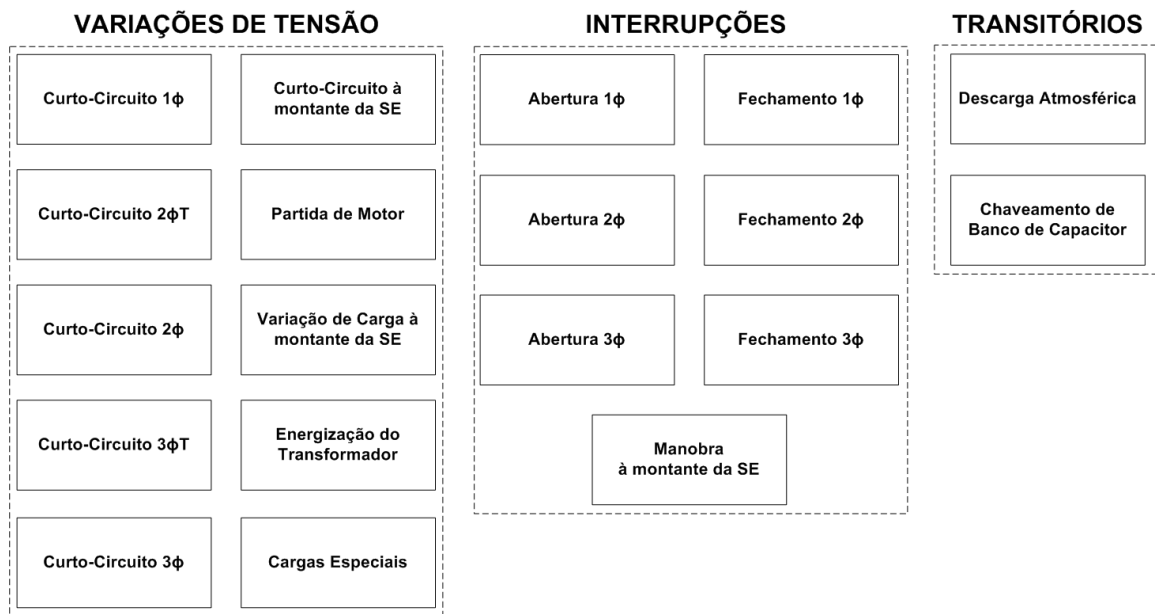
Destaca-se que a originalidade do estágio de classificação baseia-se em uma abordagem conceitual para classificação de eventos com base em oscilografias, tendo em vista a possibilidade de incluir a detecção de novidades como parte do processo de classificação, já que não foram encontrados trabalhos realizados dentro dessa perspectiva em análise automática de oscilografias em SEP. Além disso, é proposta uma forma de processar os eventos caracterizados como novidades, visando facilitar o processo de identificação de novas classes no contexto de classificação de oscilografias.

## 1.5 ORGANIZAÇÃO DO TEXTO

No capítulo 2 serão apresentados os conceitos que definem cada uma das classes de eventos analisados neste trabalho. Serão apresentadas formas de onda típicas e a dificuldade na detecção e caracterização de cada evento, tendo em vista os aspectos dinâmicos de operação de um sistema elétrico de distribuição. No capítulo 3 será apresentada uma revisão da literatura para os métodos de detecção e classificação atualmente utilizados em SEP. Também serão discutidas abordagens utilizadas em outras áreas de pesquisa, principalmente processamento de imagens, cujos métodos de classificação incluem o conceito de detecção de novidades e identificação de classes. No capítulo 4 apresentar-se-á o referencial teórico para todos os estágios da metodologia proposta. No capítulo 5, a metodologia do trabalho será segmentada em cada um dos seus estágios formadores, sendo discutidos os detalhes de cada um desses estágios de forma bastante detalhada. No capítulo 6 serão discutidos os resultados obtidos e, finalmente, no capítulo 7 serão apresentadas as conclusões, detalhando uma proposta dos métodos a serem utilizados em cada etapa, contribuições da tese e sugestões para trabalhos futuros.

## 2 EVENTOS ANALISADOS

Ao longo do Sistema Elétrico de Potência, pode-se agrupar as oscilografias de tensão em três grandes classes: interrupções, variações de tensão e transitórios (BOLLEN; GU, 2006). A Figura 6 resume os principais eventos de acordo com essa divisão, em redes de distribuição. A seguir, serão descritos os principais eventos associados a cada uma dessas três classes, demonstrando casos reais e exemplos simulados digitalmente, caracterizando o impacto causado por cada classe de eventos no sistema de distribuição utilizado como base neste trabalho.



**Figura 6:** Resumo dos principais eventos discutidos. Nesse resumo,  $T$  corresponde ao terra, SE à subestação e  $\phi$  à fase.

### 2.1 INTERRUPÇÕES

Quando se tratam interrupções em registros oscilográficos, em especial no caso de eventos de QEE, normalmente se utiliza como critério de detecção uma magnitude de tensão nula ou próxima de zero (até 10% do valor nominal da tensão) (BOLLEN, 1999). Uma outra definição possível para esse termo baseia-se na inexistência de uma conexão galvânica entre o

consumidor e o restante do sistema elétrico, ou seja, a abertura de um dispositivo de interrupção, como um elo fusível por exemplo, provoca a interrupção de fornecimento do ponto de vista do consumidor.

Em geral, as interrupções são classificadas de acordo com a sua duração e causa. No caso particular deste trabalho, a causa das interrupções é de maior interesse, em função das características do sistema de classificação que se deseja desenvolver. Sendo assim, é possível apontar as seguintes causas:

- Interrupções Forçadas:

Abertura de um disjuntor ou elo fusível em função de um curto-circuito;

Abertura inadvertida de um disjuntor ou outro dispositivo de manobra.

- Interrupções Planejadas:

Desligamento de parte do sistema, visando, por exemplo, à manutenção do trecho desligado.

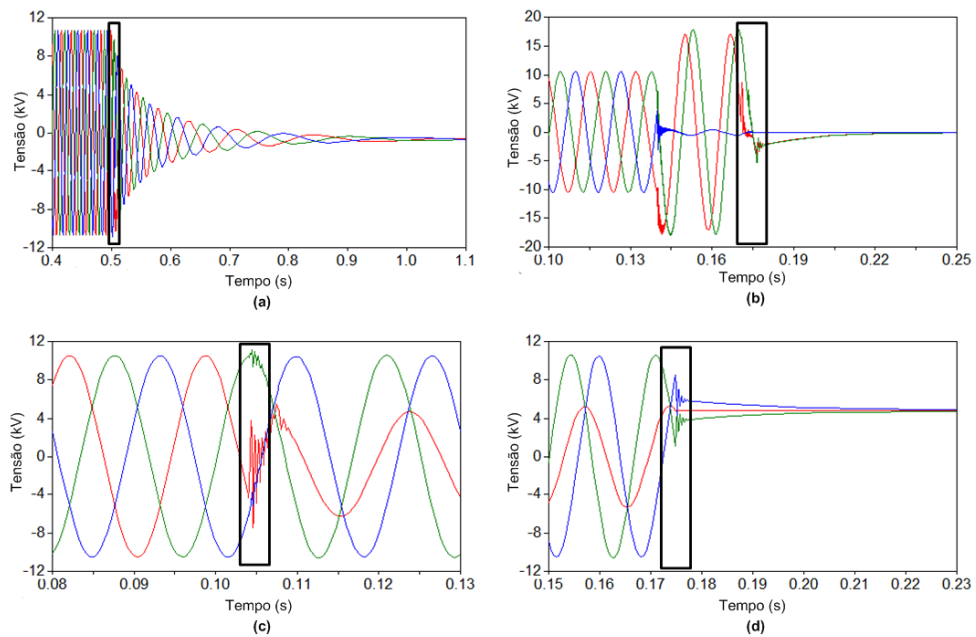
Na Figura 7 observam-se quatro eventos distintos associados a interrupções de energia gerados por simulação. É importante notar que, além da alteração das características em regime permanente ou regime quase-estacionário<sup>1</sup>, uma abertura pode provocar transitórios, os quais auxiliam na caracterização do evento. Esses transitórios estão assinalados com retângulos na Figura 7.

Uma vez finalizada a interrupção, é necessário o restabelecimento do sistema desenergizado, seja de forma automática (atuação do religador da subestação), seja de forma manual, através da atuação de chaves ou disjuntores. De forma análoga aos eventos de abertura, os eventos de restabelecimento (religamento) são caracterizados por um regime transitório precedente ao regime quase-estacionário, dada a configuração de religamento das chaves ou disjuntores. Na Figura 8 observam-se dois casos de religamento, sendo um deles monofásico e outro trifásico, gerados por simulação.

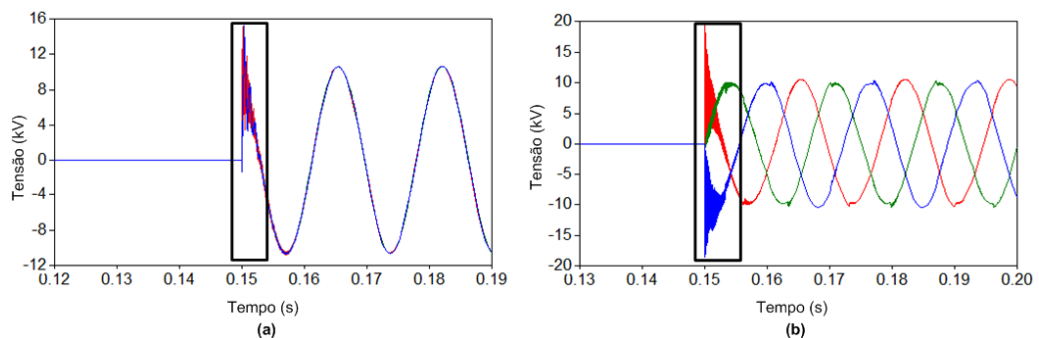
Tomando como base os eventos apresentados na Figura 7, fica claro que a configuração prévia do sistema influencia de forma direta nas características transitórias e de regime, dos eventos discutidos até então. Além disso, se esses registros tivessem sido coletados à montante do local do evento, obter-se-ia uma característica bastante distinta nas formas de onda. Para

---

<sup>1</sup>Nesse contexto, entende-se por regime permanente (regime quase-estacionário) as características de magnitude e fase da frequência fundamental do sistema, nos ciclos subsequentes à ocorrência do transitório que caracteriza o evento.



**Figura 7: Exemplos de interrupção de energia. Todas as formas de onda correspondem a medidas à jusante do local do evento. (a) Interrupção trifásica em um sistema com motores de potência significativa como carga. (b) Interrupção provocada por um curto-circuito monofásico. (c) Abertura planejada de uma das fases do sistema. (d) Abertura planejada da segunda fase de um sistema em abertura monofásica prévia.**

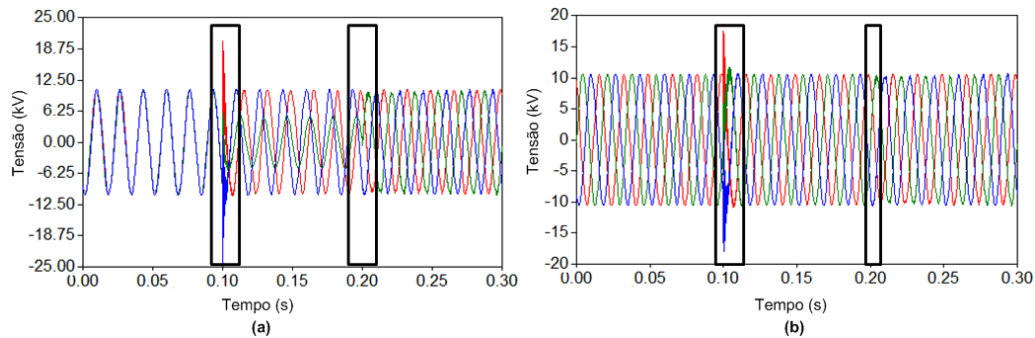


**Figura 8: Exemplos de restabelecimento de energia. Todas as formas de onda foram registradas à jusante do local do evento. (a) Religamento monofásico em um sistema previamente desenergizado. (b) Religamento Trifásico.**

ilustrar este fato, a Figura 9 demonstra o caso de dois religamentos monofásicos em um sistema energizado apenas por uma fase, registrado na subestação e ao longo do alimentador.

Desse modo, pode-se afirmar que uma correta classificação para eventos de abertura ou religamento (manobras no alimentador) depende primeiramente do local de registro do evento. Uma vez estabelecido o local, nota-se que, para eventos ocorridos à montante do ponto de medição, existe uma dependência do estado anterior do sistema para caracterização dos eventos em regime permanente, ao contrário do que se observa para eventos ocorridos à jusante, sendo esses, caracterizados basicamente pelos transitórios observados.

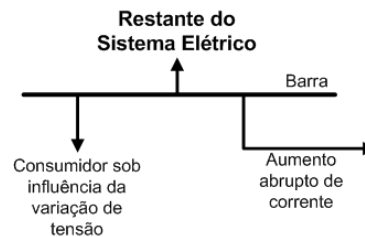




**Figura 9: Diferença no local de medição. Religamento monofásico em um sistema energizado por apenas uma fase, (a) monitorando ao longo do alimentador e (b) na subestação.**

## 2.2 VARIAÇÕES DE TENSÃO

As variações de tensão estão associadas a um aumento significativo na corrente demandada por um determinado circuito, cujas consequências estão diretamente associadas a afundamentos ou elevações de tensão em outros pontos que possuem o mesmo ponto de acoplamento, conforme pode ser observado na Figura 10.



**Figura 10: Origem das variações de tensão. Uma variação abrupta de corrente em um determinado trecho do sistema de distribuição promove uma variação de tensão nos consumidores que compartilham o mesmo barramento de tensão.**

Fonte: Adaptado de (BOLLEN, 1999)

Em geral, as variações de corrente responsáveis por causar as variações de tensão aqui expostas estão relacionadas a curtos-circuitos, partida de cargas especiais, como motores de grande potência, e energização de transformadores.

### 2.2.1 CURTO-CIRCUITO

Um curto-circuito é definido principalmente 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 1 mostra a ocorrência percentual desses eventos no SEP do Brasil (KIN-

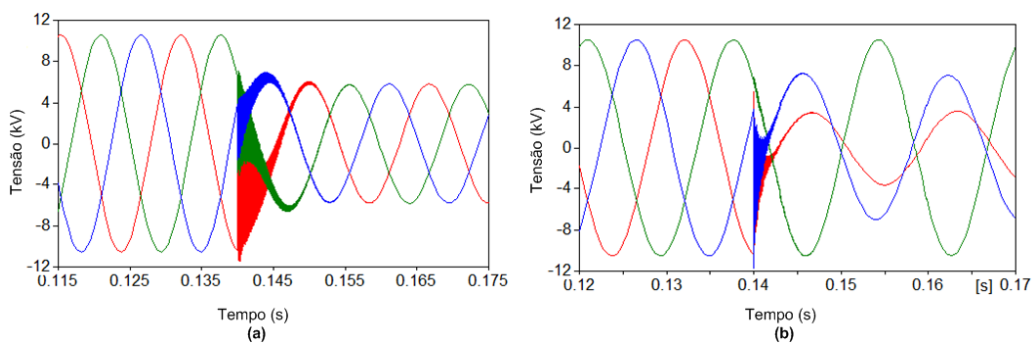
DERMANN, 2007).

**Tabela 1: Tipos de curto-circuito.**

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

Fonte: (KINDERMANN, 2007)

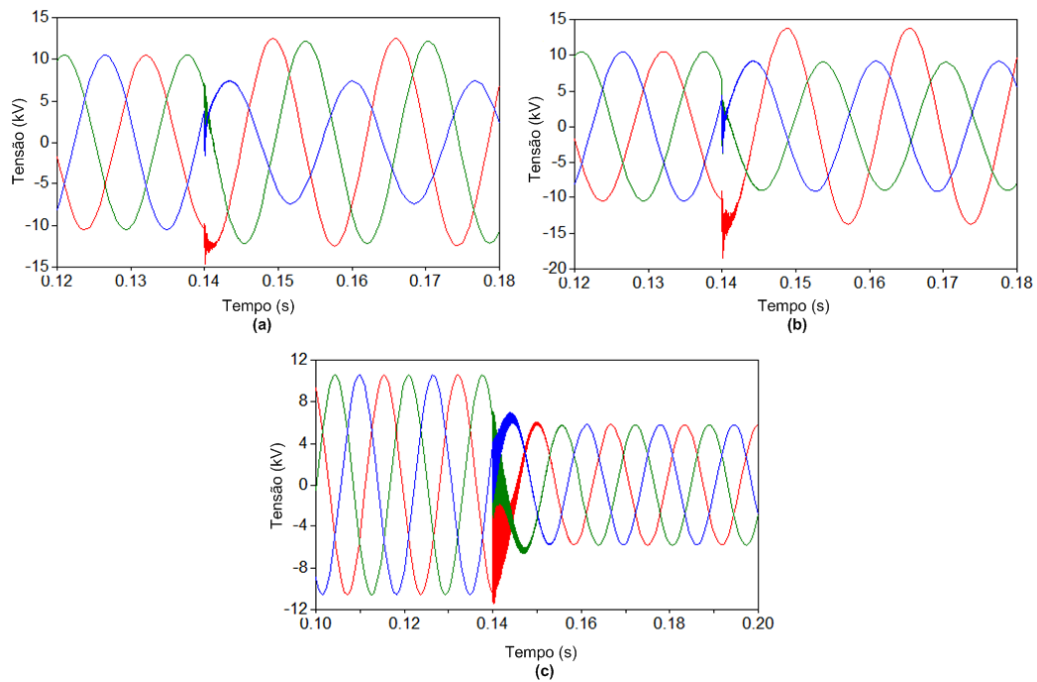
Na Figura 11 estão representados dois casos de curto-circuito que não envolvem o terra (gerados por simulação). Observa-se, no caso da Figura 11b, que a fase que não está em falta apresenta pequenas variações, tanto em regime quase-estacionário, quanto em regime transitório. Já na Figura 12, que ilustra curtos-circuitos que envolvem o terra (gerados por simulação), observa-se que a(s) fase(s) não envolvida(s) com a falta, apresenta(m) uma elevação de tensão característica de sistemas ligados em delta na média tensão e reator de aterramento para circulação dessa corrente de falta (MASON, 2000). Em todos os casos, as fases envolvidas no curto-circuito, ou seja, as fases que fornecem a potência demandada pelo curto-circuito, apresentam um afundamento de tensão.



**Figura 11: Curto-circuito trifásico (a) e bifásico (b), ambos não envolvendo o terra.**

É importante observar que as formas de onda de tensão registradas durante um curto-circuito são impostas para todos os consumidores conectados no mesmo barramento, apresentando pequenas variações transitórias e de regime ao longo do alimentador monitorado. Quando observados desse modo, pode-se afirmar que os curtos-circuitos são menos dependentes do ponto de monitoramento (registro) da oscilografia (MASON, 2000).

Além dos critérios acima mencionados, é possível apontar outros fatores relevantes na caracterização de uma falta, como a resistência de falta, os valores de magnitude e fase da frequência fundamental do sinal (60 Hz) para o sistema em falta e o elemento causador da falta, como galhos de árvore, contato de animais, descargas atmosféricas, entre outros.

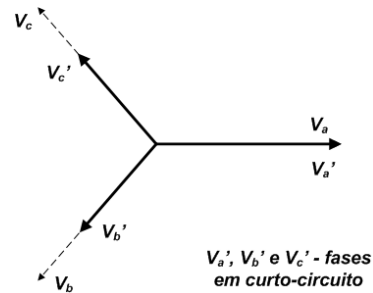


**Figura 12: Curto-circuito (a) monofásico, (b) bifásico e (c) trifásico, todos envolvendo o terra.**

Com relação à resistência de falta, a característica mais proeminente está associada ao nível do afundamento de tensão resultante da corrente demandada pelo curto-circuito nas fases envolvidas neste evento. Quanto maior a corrente solicitada, mais severo o afundamento de tensão, podendo chegar a valores inferiores a 0,1 p.u. Em alguns casos, a resistência de falta é bastante elevada, levando a correntes de curto-circuito não significativas para a atuação da proteção. Essas faltas são comumente denominadas faltas de alta impedância e sua identificação tem sido alvo de diversos estudos, principalmente em sistemas de distribuição (SAMANTARAY; DASH, 2008; SARLAK; SHAHRTASH, 2011; MICHALIK et al., 2006; ZAMORA et al., 2007), uma vez que, dependendo do local do registro do evento, não se observam alterações nas formas de onda de tensão e corrente que possam caracterizar esse evento. Seu impacto é significativo do ponto de vista de segurança, uma vez que nesse tipo de evento o cabo energizado pode permanecer em contato com o solo sem o desligamento do sistema através da proteção.

Já com relação aos valores de magnitude e fase da frequência fundamental do sinal, pode-se observar um padrão associado a cada caso específico de curto-circuito, que normalmente se denomina critério de classificação ABC (BOLLEN, 1999). Esse critério fornece uma característica da falta de acordo com as variações dos fasores de tensão durante o evento, o arranjo de ligação do transformador e a conexão da carga. Ao todo são sete diferentes diagramas fasoriais que, quando combinados com a ligação do transformador e da carga, fornecem uma classificação do curto-circuito. Para ilustrar esse fato, pode-se observar o exemplo apresentado na Figura 13, na qual se considera um curto-circuito bifásico envolvendo o terra, com

um transformador estrela-estrela e uma carga ligada em estrela.



**Figura 13: Exemplo do diagrama fasorial de um curto-circuito bifásico para o terra. Esse arranjo fasorial é classificado como tipo *E*, dentro da classificação dos sete arranjos distintos (*A*, *B*, *C*, *D*, *E*, *F* e *G*). Observa-se um afundamento nas duas fases envolvidas na falta (*B* e *C*), enquanto na fase não associada ao evento, não se observam variações em relação ao estado pré-falta (diagrama fasorial em tracejado).**

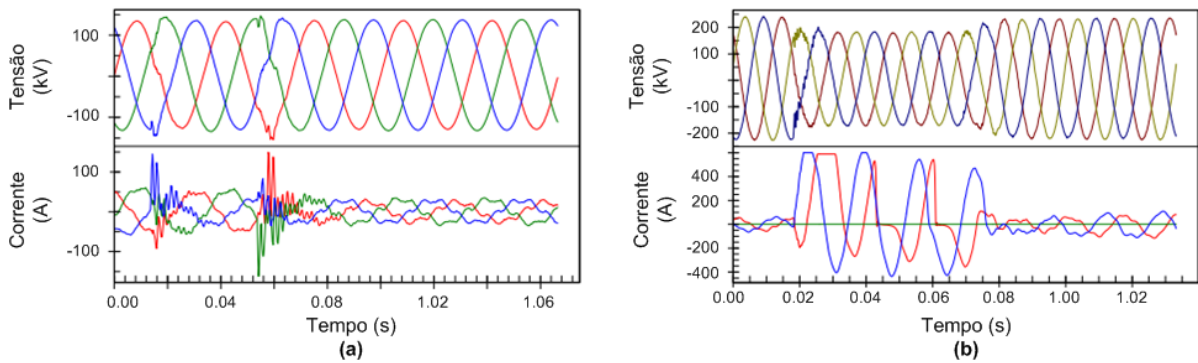
Por último, o elemento causador da falta pode levar a diferentes características nas formas de onda, como demonstram Kulkarni et al. (2010), Kulkarni e Santoso (2010), Nunez et al. (2010), Xu et al. (2007b), Xu e Chow (2006), Xu et al. (2007a). O contato dos condutores energizados com galhos de árvore em função da ação do vento, por exemplo, pode provocar sucessivos curtos-circuitos, envolvendo uma ou mais fases, resultando em abertura do dispositivo de proteção, ou até mesmo, sendo autoextintas. Já um curto-circuito decorrente de uma descarga atmosférica pode se manifestar de diferentes formas, sendo as principais delas, faltas provocadas pelo *flashover*<sup>2</sup> no isolador da rede ou a ruptura dos dielétricos dos equipamentos instalados na rede, sob influência da sobretensão de origem atmosférica. Cada um desses efeitos gera uma impressão distinta nos transitórios associados à falta, dificultando sua caracterização imediata. Na Figura 14 observa-se que o comportamento de tensão e corrente pode ser bastante distinto, mesmo para eventos de mesma natureza (curto-circuito).

### 2.2.2 PARTIDA DE CARGAS ESPECIAIS

A carga responsável pelo maior impacto em termos de QEE no sistema elétrico, principalmente de distribuição, é a partida de motores de indução de grande potência. Esse tipo de evento é caracterizado por um afundamento trifásico seguido por um retorno lento à normalidade, como mostra a Figura 15, gerada via simulação.

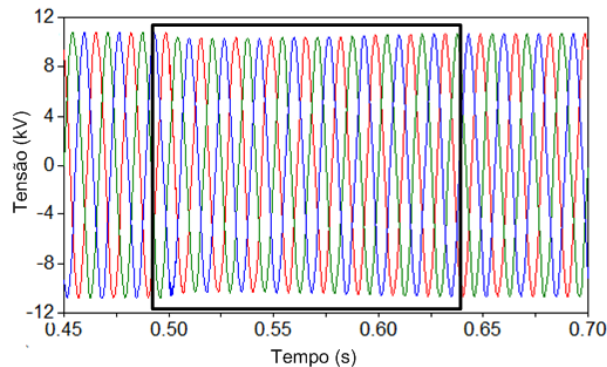
Além da potência e o fator de carga do motor, existem outras características que contribuem com a severidade e a duração desse tipo de evento, como o momento de inércia da carga (TORO, 1999). Observa-se que o parâmetro mais relevante associado a esse evento é

<sup>2</sup>O *flashover* é o mecanismo de falha no isolamento que pode ser associado à incidência de uma descarga atmosférica em um condutor energizado da linha de distribuição (VISACRO-FILHO, 2005).



**Figura 14: Comparação entre dois curtos-circuitos de origem distinta. Em (a) observa-se uma falta decorrente do contato de condutores de um sistema de transmissão com galhos de árvore e em (b) um curto-circuito decorrente de descarga atmosférica nesse mesmo sistema.**

**Fonte: (KULKARNI et al., 2010)**



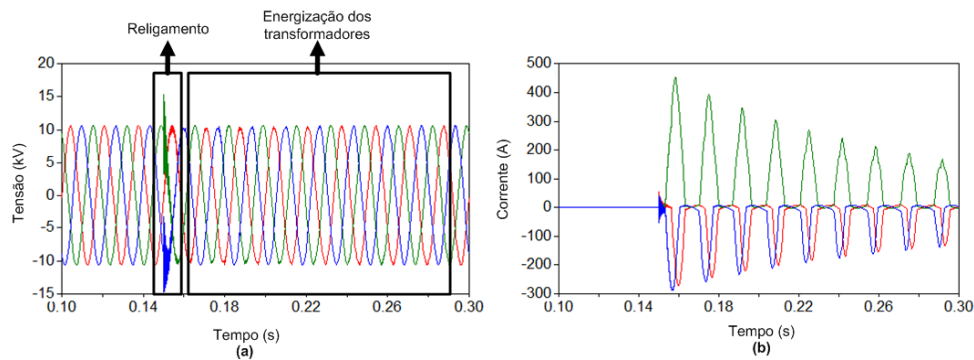
**Figura 15: Partida de um motor de indução de 800cv/4,16kV.**

a sua longa duração, que pode provocar afundamentos de tensão com duração superior em comparação aos eventos anteriormente apresentados.

### 2.2.3 ENERGIZAÇÃO DE TRANSFORMADORES

A energização de transformadores possui uma característica muito parecida com a partida de grandes motores, pois também apresenta um afundamento seguido de um lento retorno à normalidade. Além disso, normalmente se observa a presença de harmônicos de ordem ímpar na forma de onda, resultante da corrente de energização do transformador. A Figura 16 demonstra as formas de onda decorrentes da energização de um conjunto de transformadores, após um religamento trifásico.

A energização de transformadores está relacionada à ocorrência de outros eventos, como o próprio religamento trifásico acima mencionado, o que por vezes dificulta a sua caracterização.



**Figura 16: Energização de transformador. Tensão (a) e corrente (b) resultantes da energização de transformador.**

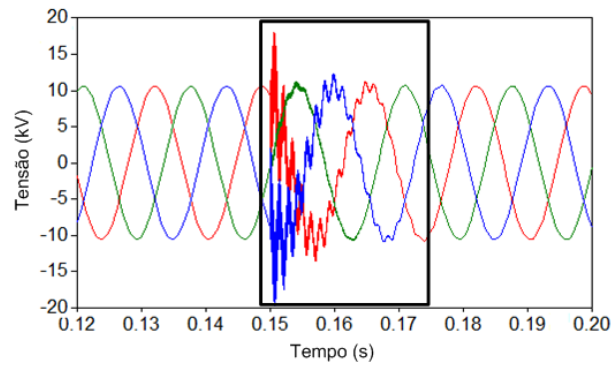
### 2.3 TRANSITÓRIOS

A definição de fenômenos transitórios em sistemas elétricos possui estreita relação com a duração do evento, sendo muitas vezes caracterizados como eventos com duração inferior a um ciclo de 60Hz. Esse termo baseia-se no termo transitório normalmente empregado na análise de circuitos elétricos (GREENWOOD, 1991), uma vez que representa um fenômeno de curta duração, normalmente verificado na transição entre dois estados quase-estacionários da forma de onda.

É possível dividir os transitórios de acordo com a direcionalidade dos eventos, ou seja, no caso dos transitórios denominados impulsivos tem-se um aspecto unidirecional na forma de onda, enquanto os transitórios denominados oscilatórios variam em torno do novo estado estável do sistema. Dentro dos diversos fenômenos transitórios observados para sistemas de distribuição, existem basicamente dois grandes grupos: transitórios provenientes de descargas atmosféricas e transitórios decorrentes de manobra em bancos de capacitores.

No caso dos transitórios decorrentes de chaveamento de banco de capacitores, destaca-se que, do ponto de vista de QEE, os de maior impacto estão associados ao fechamento, ou à energização desses bancos. A Figura 17 demonstra um exemplo simulado de fechamento trifásico de um banco de capacitores de 600kvar em 13,8kV. Os transitórios observados são função, principalmente, do valor da capacitância por fase do banco, sendo que esses valores normalmente geram transitórios com conteúdo de frequência entre 500Hz e 1200Hz (BOLLEN, 1999).

Já as descargas atmosféricas compreendem um dos fenômenos mais complexos e de difícil caracterização na natureza (MACGORMAN; RUST, 1998). Seu completo conhecimento ainda é um tema em aberto para ciência. No entanto, existem diversas aproximações que podem ser adotadas, que sob determinados aspectos do ponto de vista de engenharia, podem auxiliar na



**Figura 17: Forma de onda resultante de um fechamento de um banco de capacitores.**

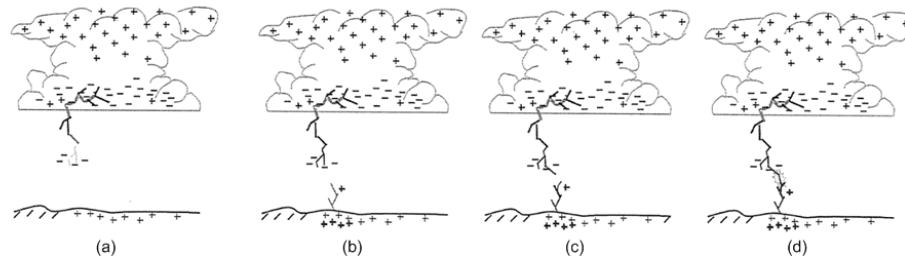
caracterização dos seus efeitos, principalmente em redes de energia elétrica (COORAY, 2003).

Uma descarga atmosférica é caracterizada basicamente por dois fenômenos principais bastante distintos: a formação do canal precursor da descarga e o fluxo de corrente de retorno após o canal constituído (VISACRO-FILHO, 2005). Esses fenômenos podem ser caracterizados sequencialmente, tendo como principal elemento na sua formação a existência de nuvens com elevado acúmulo de cargas negativas na sua base.

Inicialmente, tem-se que o acúmulo de cargas na base de uma nuvem provoca um aumento significativo do campo elétrico entre a nuvem e o solo, dada a diferença de potencial existente entre esses dois elementos. Em alguns pontos, no interior da base da nuvem, o valor do campo elétrico local pode ultrapassar o valor da rigidez dielétrica do ar, promovendo uma descarga elétrica, cuja consequência é a formação de um canal ionizado de plasma. Esse canal, por ser de natureza condutiva, recebe uma grande quantidade de cargas negativas das regiões vizinhas, permitindo que o campo elétrico local seja novamente capaz de romper a rigidez dielétrica do ar, repetindo o processo de descarga (UMAN, 2001). Esse processo pode se repetir para regiões externas à base da nuvem, de forma descendente, estabelecendo o canal precursor da descarga.

Em função da configuração de cargas elétricas negativas apresentada na nuvem, no solo ocorre um acúmulo superficial de cargas positivas. À medida que ocorre a evolução do canal precursor da descarga, pode-se observar um aumento correspondente da densidade local das cargas no solo, de modo que o campo elétrico nesse local também possa provocar uma ruptura da rigidez dielétrica, resultando em um processo de descargas ascendentes, similar ao processo descendente. Em um determinado ponto do processo de descarga ascendente e descendente, pode ocorrer a interligação entre esses dois canais, caso a distância entre eles seja inferior a um limite associado a quantidade de carga acumulada no canal descendente. Essa interligação (conexão) entre os canais promove a circulação de uma corrente elétrica de alto valor a partir

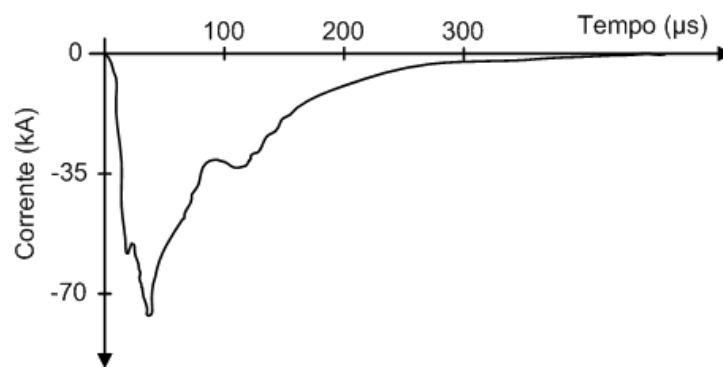
do ponto de conexão, denominada corrente de retorno (VISACRO-FILHO, 2005). A Figura 18 ilustra o processo descrito até aqui.



**Figura 18: Evolução do processo de descarga atmosférica a partir da formação do canal descendente (a), seguido pela indução e formação do canal ascendente (b e c) e finalizando com a conexão entre os canais, gerando a corrente de retorno.**

**Fonte: (VISACRO-FILHO, 2005)**

As descargas atmosféricas podem ser classificadas como negativas ou positivas, de acordo com o fluxo de cargas estabelecido, variando principalmente com as características geográficas do local de ocorrência da descarga. A Figura 19 mostra um exemplo de corrente de descarga negativa, sendo que esse tipo de descarga é a mais comumente encontrada em regiões nas quais se encontram os sistemas de distribuição discutidos neste trabalho.



**Figura 19: Forma de onda da corrente de descarga atmosférica negativa.**

**Fonte: (VISACRO-FILHO, 2005)**

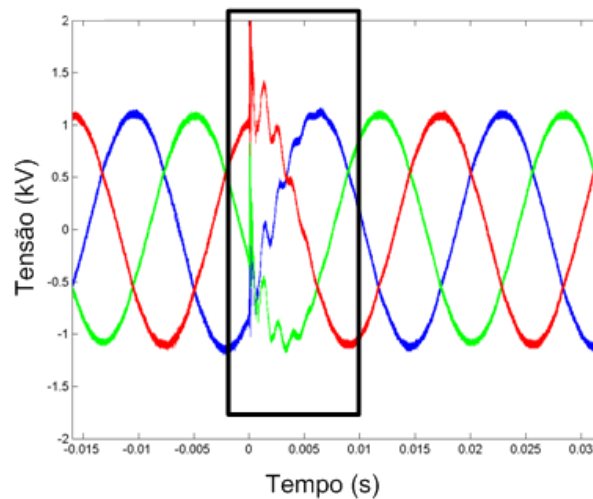
Em redes de distribuição, o maior impacto causado pelas descargas atmosféricas está associado aos fenômenos de tensão induzida para descargas ditas indiretas (PAOLONE et al., 2009). O canal precursor da descarga e, em especial, a corrente de retorno são responsáveis por gerar campos eletromagnéticos de intensidade capaz de estabelecer tensões induzidas em linhas aéreas próximas ao local de ocorrência da descarga.

Para uma compreensão mais ampla das características apresentadas nas tensões induzidas nas linhas é necessário o detalhamento de três fatores principais: os modelos de corrente



de retorno, os campos eletromagnéticos gerados pela corrente de retorno e o cálculo do acoplamento entre o campo eletromagnético da corrente de retorno e as redes de distribuição sob influência desse campo, normalmente denominadas de linhas iluminadas (VISACRO-FILHO, 2005). Um detalhamento maior desses fatores foge ao escopo do presente trabalho. Em Cooray (2003) é possível encontrar um resumo desses aspectos, principalmente sobre os efeitos de descargas atmosféricas em redes de distribuição.

A forma de onda típica resultante de uma descarga atmosférica indireta pode ser observada na Figura 20. Para essa forma de onda, estima-se que o provável local de incidência da descarga atmosférica foi a um quilômetro da linha sob supervisão. Destaca-se que, mesmo para distâncias dessa magnitude, ainda se observam sobretensões significativas nas formas de onda.



**Figura 20: Forma de onda resultante de uma descarga atmosférica indireta em um sistema de monitoramento real.**

**Fonte: (LAZZARETTI et al., 2011)**

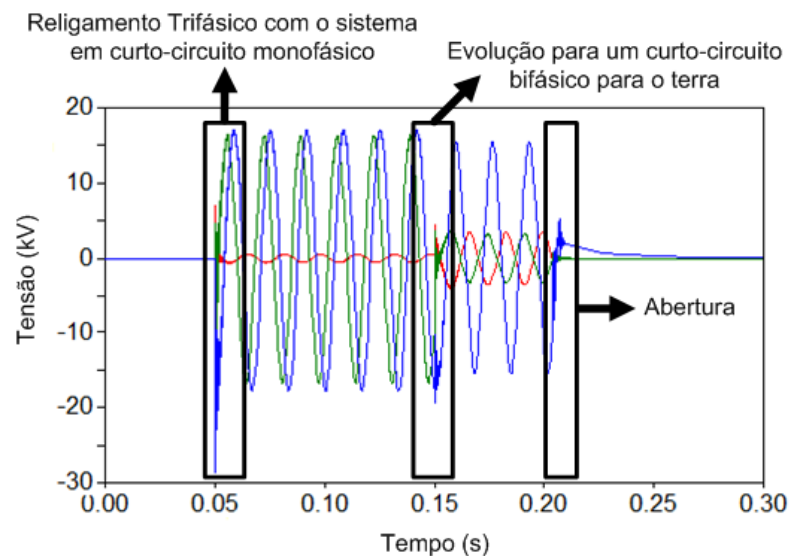
## 2.4 OBSERVAÇÕES FINAIS

Tanto para os eventos simulados, quanto para os eventos registrados pelos sistemas instalados em campo, as classes escolhidas para análise estão representadas na Figura 6. No entanto, é importante assinalar que essas classes não formam um conjunto fechado. Existe a possibilidade de existirem registros de classes não listadas e discutidas aqui, principalmente em redes de distribuição inteligentes, caracterizando um cenário dinâmico em termos de classes a serem registradas por sistemas de monitoramento em redes de distribuição.

Outra consideração importante a ser ressaltada está relacionada à possibilidade de

ocorrência de eventos em outros estágios do SEP, cujas alterações nas formas de onda podem ser refletidas no sistema de distribuição monitorado. Esse cenário justifica a inclusão dos eventos de curto-circuito e variações de carga à montante da subestação que configuram os eventos mais frequentemente observados nessas redes.

Adicionalmente, é importante salientar que, até o presente momento, todos os eventos foram apresentados isoladamente dentro de uma oscilografia. Nesse sentido, assinala-se que a principal dificuldade na análise de oscilografias contendo uma gama tão diversificada de eventos reside no fato de que todos esses eventos podem ocorrer de forma individual (como apresentado), sequencial ou simultânea. Ou seja, pode-se observar, por exemplo, um religamento trifásico de um religador de uma subestação de distribuição, sem que haja a extinção de um curto-circuito monofásico previamente estabelecido. Além disso, esse curto-circuito monofásico pode evoluir para um curto-circuito bifásico para o terra, provocando novamente a abertura trifásica do sistema, conforme exemplifica a Figura 21, onde se observa a ocorrência de um evento isolado (curto-circuito bifásico para o terra), sequencial a um evento simultâneo, envolvendo o religamento e o curto-circuito monofásico sustentado.



**Figura 21: Oscilografia simulada com diferentes estados.**

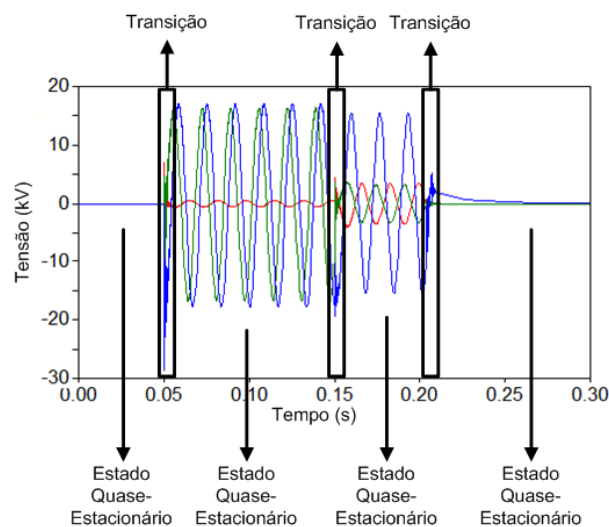
Pode-se interpretar cada evento da Figura 6 como um estado particular do sistema, sendo que pode haver uma mudança entre esses estados, ou até mesmo a ocorrência de mais de um estado simultaneamente. Adicionalmente, um estado prévio pode influenciar, mesmo que indiretamente, as características de forma de onda do estado subsequente, formando um cenário bastante complexo para a análise e correta identificação de cada evento. Esses detalhes são contemplados na metodologia proposta neste trabalho, sendo que eventos simultâneos e subsequentes serão considerados como novidades em todas as análises feitas neste trabalho.

### 3 REVISÃO BIBLIOGRÁFICA

#### 3.1 DETECÇÃO DE EVENTOS

Como foi discutido no Capítulo 2, a alteração que ocorre na forma de onda devido à transição entre dois eventos é de suma importância para caracterização de cada ocorrência em um registro oscilográfico. Basicamente, o processo de detecção envolve dois estágios distintos: a identificação dos instantes de tempo em que ocorrem as transições (disparo) e a segmentação da forma de onda em trechos disjuntos, separando esses trechos em estados quase-estacionários, permitindo uma janela de análise adaptada para cada classe de evento (BOLLEN; GU, 2006).

Na Figura 22 observa-se, através da marcação de retângulos, os instantes de transição entre estados quase-estacionários, os quais caracterizam o processo de disparo. Entre cada processo de transição, encontram-se os segmentos das oscilografias. A análise conjunta desses dois parâmetros fornece a análise completa de cada evento, muito embora seja notório que grande parte dos eventos apresente suas principais características no regime quase estacionário no qual se encontram.



**Figura 22: Segmentação e disparo de uma oscilografia.**

Existem diversos métodos para realizar os estágios de disparo e segmentação de oscilo-

grafias, os quais normalmente são tratados de maneira conjunta. De um modo geral, as técnicas são agrupadas de acordo com a natureza do método aplicado aos sinais. Os principais grupos compreendem (BOLLEN; GU, 2006):

- Detecção de variações através da análise do sinal no domínio do tempo;
- Métodos de filtragem;
- Transformações tempo-frequência, como a Transformada Janelada de Fourier e Transformada *Wavelet*;
- Resíduos mais proeminentes de modelos baseados em estimação espectral;
- Técnicas de reconhecimento de padrões.

A partir dos grupos acima listados, iniciou-se uma busca na literatura tendo em vista a aplicação dessas técnicas para detecção de variações abruptas em sinais, principalmente de sinais relacionados a oscilografias em sistemas de potência (SANTOSO et al., 2009; UKIL; ZIVANOVIC, 2006b).

De todos os métodos observados, os mais simples e mais utilizados estão associados à análise do sinal no domínio do tempo. Nesse sentido, busca-se através da extração de características, como o cálculo do valor eficaz do sinal, determinar os instantes de início e fim de um evento. Em Bollen et al. (2007) observa-se que o desempenho desses métodos é adequado para eventos relacionados com variações significativas nos valores eficazes das formas de onda em regime quase estacionário. No entanto, para transitórios e variações menos significativas, o desempenho desses métodos pode comprometer a precisão na localização temporal dos eventos e, até mesmo, a detecção da transição entre eventos.

Uma primeira alternativa para contornar as dificuldades apresentadas pelos métodos de análise temporal é a utilização de filtros para uma melhor caracterização das transições existentes entre os segmentos de uma forma de onda (UKIL; ZIVANOVIC, 2005a). Wiot (2004) propõe um procedimento de filtragem do sinal no domínio da frequência, cuja principal função é tornar mais proeminente as transições entre estados, mesmo com a presença de ruído. Além disso, o filtro proposto é de baixo custo computacional, o que torna a metodologia mais robusta do que os métodos no domínio do tempo.

Ainda nesse cenário, uma das grandes inovações observadas na detecção de eventos é a utilização de transformações do domínio do tempo para o domínio tempo-frequência. Em especial, destaca-se o uso da Transformada *Wavelet* Discreta (TWD) (BOLLEN; GU, 2006).

Nesse tipo de transformação, as informações temporais existentes nos sinais são preservadas em diferentes bandas de frequência, nas quais podem estar diferentes informações de transição entre estados. Essa característica é de fundamental importância para os eventos observados em SEP, uma vez que existe uma gama bastante variada de eventos e transições, conforme exposto no capítulo anterior para sistemas de distribuição.

Em Costa et al. (2006, 2010), a característica de preservação de informações em diferentes bandas de frequência é validada para oscilografias obtidas de registradores instalados em um sistema de transmissão. Naqueles trabalhos são segmentados os trechos de pré-falta, falta, eliminação da falta, pós-falta e ruídos existentes no estágio pós-falta, nos casos de curtos-circuitos e demais eventos de manobra observados no sistema. Através de um conjunto de regras aplicadas à energia dos coeficientes dos diversos níveis da TWD, obteve-se um sucesso de detecção em 1690 registros, de um total de 1696, com erro médio de 0,3 ms para início dos eventos e 1,4 ms para o final.

No caso particular desses métodos, merecem destaque aqui os trabalhos propostos pelo pesquisador Abhisek Ukil e demais colaboradores (UKIL; ZIVANOVIC, 2005a, 2005b, 2006a; UKIL; IVANOVI, 2007; UKIL; ZIVANOVIC, 2008). A metodologia empregada por esse grupo de pesquisadores consiste em aplicar uma decomposição *Wavelet* nos sinais sob análise e, a partir das transformações obtidas, estabelecer a segmentação da oscilografia de forma automática. Para tanto, utiliza-se um procedimento de cálculo do limiar universal apresentado por Donoho e Johnstone (1994), seguido de uma filtragem de suavização. Com isso, garante-se que todo o processo de detecção seja feito sem a necessidade de definição de limiares empíricos, o que evita a forte dependência de uma análise posterior, normalmente realizada de forma supervisionada pelos operadores do sistema de análise de oscilografias (COSTA et al., 2010).

O desempenho dessa metodologia é bastante relevante para os eventos de diferente natureza registrados ao longo do SEP. Em Ukil e Zivanovic (2008), propõe-se uma alteração da função *Wavelet* de Haar, cuja consequência é um aumento no desempenho de detecção, chegando a um acerto de 99% para um grupo de sinais, cujo acerto era de 80% utilizando o mesmo procedimento com a *Wavelet* de Haar tradicional (MALLAT, 1999).

Muito embora os procedimentos de detecção com base em TWD possuam uma complexidade computacional maior quando comparados aos demais métodos, Gencer et al. (2010) demonstram que é possível realizar a segmentação de oscilografias com base na TWD em tempo real, utilizando um processador digital de sinais dsPIC modelo 30F4011, operando em 120 MHz.

A grande dificuldade encontrada para os métodos de detecção baseados em TWD é a

combinação dos diversos níveis da TWD para determinar a segmentação completa da oscilografia. Normalmente, utiliza-se apenas o primeiro nível dessa transformada para realizar o processo de segmentação (SILVA et al., 2006; UKIL; ZIVANOVIC, 2008). No entanto, um transitório decorrente de um evento no SEP contém diferentes conteúdos de frequência. Esse conteúdo de frequência está presente em diferentes níveis da TWD. A composição desses níveis ainda é pouco explorada na literatura e é um dos temas centrais do método de segmentação proposto no presente trabalho.

Com relação aos métodos de estimação espectral, destaca-se sua larga aplicação para detecção de variações abruptas em sinais (GUSTAFSSON, 2000). Dentre os diversos métodos existentes, os mais utilizados são os baseados em modelos autorregressivos, filtros de Kalman, *Multiple Signal Classification* (MUSIC) e *Estimation of Signal Parameters via Rotational Invariance* (ESPRIT). Uma característica comum a esses métodos é que todos realizam a estimação das características do sinal (espectro) e através do resíduo calculado amostra a amostra, é possível estabelecer as transições, uma vez que se esperam resíduos de mais alta magnitude em pontos onde existem transições súbitas, já que o erro de estimação é maior nesses casos.

Para aplicações em SEP, encontra-se ainda na literatura a utilização de métodos baseados em filtro de Kalman (MORETO; ROLIM, 2010a, 2010b; LE et al., 2010; STYVAKTAKIS et al., 2001; STYVAKTAKIS, 2002; STYVAKTAKIS et al., 2002; UKIL; ZIVANOVIC, 2007), modelos AR (SANTAMAR et al., 1999; BOLLEN et al., 2007) e outros métodos de estimação de pseudoespectro, como o método de Prony (LOBOS et al., 2006).

Dentre esses trabalhos, destacam-se os que são baseados no filtro de Kalman, que fundamentam-se nos trabalhos de Dash et al. (1999), Kennedy et al. (2003), além de Nishiyama (1997) e Nishiyama (2007). Nesses trabalhos são apresentados os fundamentos teóricos para a estimação do espectro dos sinais considerando uma abordagem linear ou não-linear (Filtro de Kalman Estendido). A abordagem não linear permite que sejam utilizadas técnicas de detecção de eventos com decaimento exponencial, de variação lenta, normalmente observados em sistemas de geração, como demonstram Moreto e Rolim (2010a). Em (UKIL; ZIVANOVIC, 2007) são empregados diversos filtros de Kalman em paralelo, somados à metodologia de determinação de limiar proposta por Donoho e Johnstone (1994), para estimar os instantes de tempo em que ocorrem transições relacionadas a pré-falta, falta, operação do disjuntor e religamento, de forma automática. Os resultados desses trabalhos demonstram que as técnicas baseadas em estimação espectral possuem desempenho bastante próximo e, em alguns casos até superior, às técnicas baseadas na transformada *Wavelet*.

Um outro grupo de técnicas, ainda pouco explorado na detecção de eventos em SEP

(ABDEL-GALIL, 2003), fundamenta-se nos métodos baseados em reconhecimento de padrões. Uma possível explicação para esse fato é que grande parte dos algoritmos de reconhecimento de padrões possuem um estágio dedicado ao aprendizado supervisionado dos padrões. No caso da detecção de eventos esse estágio é de certa forma indesejável (DESOBRY et al., 2005), pois exige conhecimento prévio de grande parte dos sinais a serem segmentados, além das diferentes características observadas nos estados quase-estacionários.

Em contrapartida, existe um outro grupo de técnicas de reconhecimento de padrões que utiliza, em partes, conceitos de aprendizagem não-supervisionada (SCHÖLKOPF; SMOLA, 2001). Em especial, destaca-se a utilização de métodos de detecção de novidades, demonstradas em (DAVY et al., 2006; DAVY; GODSILL, 2002; DESOBRY; DAVY, 2003; MA; PERKINS, 2003; DESOBRY et al., 2005). Nesses trabalhos, são aplicadas Máquinas de Vetor Suporte para problemas de uma-classe, em uma abordagem baseada na razão de Fisher, para sinais das mais variadas aplicações. Os resultados desses métodos demonstram sua comparabilidade com métodos baseados na estimação do espectro, tanto em termos de desempenho, quanto em termos de complexidade computacional. Uma vantagem interessante desse tipo de abordagem é que é possível a sua aplicação no domínio do tempo-frequência, aproveitando todas as vantagens apresentadas para esses métodos, sendo uma possível forma de contornar a limitação da composição de diferentes níveis da TWD no processo de segmentação.

No caso particular do trabalho apresentado por Desobry et al. (2005) é feita uma segmentação para sinais de áudio. Esses sinais possuem uma característica temporal com diversas variações abruptas que dificultam a detecção dos eventos de interesse. No entanto, a robustez do método permite que os sinais sejam detectados com acerto superior aos métodos tradicionais (BASSEVILLE; NIKIFOROV, 2006). A maior dificuldade encontrada foi a determinação do limiar, que teve de ser feita de forma supervisionada, pois o método não incluía a possibilidade de definição automática desse parâmetro.

Um último procedimento que merece destaque foi utilizado em segmentação de objetos em imagens (KAISER, 1993; TAVARES et al., 2011). Esse método é denominado Operador de Energia de Teager e sua principal característica é a redução da energia fundamental do sinal (função envelope), permitindo que as demais componentes de frequência possam ser caracterizadas. Essa característica é fundamental na detecção de eventos e, dada a simplicidade do método, convém apresentar e comparar seu desempenho para sinais da rede elétrica.

### 3.2 CLASSIFICAÇÃO DE EVENTOS

Conforme apresentado anteriormente, o processo de classificação presente neste trabalho envolve, em linhas gerais, duas etapas principais:

- Classificação multiclasse;
- Detecção de novidades.

Além disso, pode-se incluir a identificação de possíveis novas classes como um tópico derivado do processo de detecção de novidades.

Em grande parte dos trabalhos apresentados na literatura, esses tópicos são tratados de forma independente, de acordo com a natureza do problema para o qual se busca a correta identificação das classes. Desse modo, optou-se por apresentar a revisão da literatura de forma individual, para cada um dos tópicos acima apresentados.

Com relação à classificação multiclasse, optou-se por focar a revisão da literatura apenas para formas de onda do SEP, por se tratar de uma metodologia mais usual para classificação desses eventos. Já para os demais estágios do processo de classificação, são discutidos trabalhos nas mais variadas áreas, incluindo principalmente, problemas envolvendo processamento de imagens.

As seções a seguir detalham os principais trabalhos relacionados a cada um dos tópicos envolvidos com o processo de classificação.

#### 3.2.1 CLASSIFICAÇÃO DE EVENTOS EM SEP

Para o levantamento bibliográfico dos classificadores com base em oscilografias no SEP, foi feita inicialmente uma busca pelos critérios a serem considerados na construção dos mesmos. Nesse aspecto, foram estabelecidos os critérios de local de ocorrência do evento, natureza dos eventos classificados, dados utilizados na classificação, pré-processamento aplicado e a técnica de classificação utilizada.

Com relação ao local de ocorrência dos eventos, o desenvolvimento dos classificadores é voltado para os diferentes estágios do sistema elétrico de potência, desde os sistemas de geração (GUO et al., 2005; TAN et al., 2007), passando pela transmissão (ADU, 2002; MAHANTY; GUPTA, 2007; VASILIC; KEZUNOVIC, 2005; SILVA et al., 2006; REDDY; MOHANTA, 2007; ZHANG; KEZUNOVIC, 2007), (DASH et al., 2007; SAMANTARAY;



DASH, 2008; RAVIKUMAR et al., 2008; DONG et al., 2009; DEMIR, 2010a; MALATHI et al., 2010) e para redes de distribuição (STYVAKTAKIS, 2002; CHIEN et al., 2002; GEREK et al., 2006; GEREK; ECE, 2006; FAIZ et al., 2007). Nota-se que grande parte dos trabalhos foca no desenvolvimento dos classificadores para sistemas de transmissão, uma vez que esse estágio do sistema de potência necessita de uma rápida mitigação dos eventos, principalmente os que envolvem curtos-circuitos. Além disso, classificadores desenvolvidos para sistemas de transmissão possuem diversas aplicações voltadas para problemas de proteção das linhas (KE-ERTHIPALA, 1998; AGUILERA et al., 2006; MELO et al., 2007). A identificação correta de um evento de curto-circuito permite a análise de eventuais falhas de atuação dos relés de proteção, além de incorporar conhecimento acerca das características do evento, tais como o tipo de curto-circuito e as fases envolvidas.

Já com relação à natureza dos eventos, observa-se que cada estágio do sistema de potência possui características específicas com relação à concepção dos classificadores. No caso dos sistemas de geração (GUO et al., 2005; TAN et al., 2007), o foco são os problemas mecânicos nos geradores, em função dos curtos-circuitos, além da identificação de faltas no interior desses geradores. Já para as linhas de transmissão e redes de distribuição, a classe de maior impacto são os curtos-circuitos (MORAIS et al., 2007), conforme mencionado no capítulo anterior. Para o caso particular dos sistemas de distribuição, 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 devido à queima de equipamentos (DUGAN et al., 2002). Considerando essa característica, uma classificação conjunta de eventos com naturezas distintas como mostram Styvaktakis (2002), Gerek et al. (2006), Gerek e Ece (2006), Silva et al. (2006), Faiz et al. (2007), Demir (2010a), faz-se necessária, já que os registradores poderão armazenar oscilografias de diferentes causas, como curtos-circuitos ou transitórios de manobra. Em alguns casos, em função das características do sistema de registro de oscilografia, é possível se estabelecer correlação com a causa do evento causador, como mostra Faiz et al. (2007), cujo objetivo da classificação é realizar a separação de eventos de curto-circuito, partida de motores, energização de transformadores e faltas simultâneas à energização de transformadores. Em Styvaktakis (2002) é apresentada uma metodologia parecida, porém não são tratados eventos simultâneos na rede.

Em termos de metodologia de classificação, o processo é normalmente realizado em dois estágios: pré-processamento das entradas e a classificação propriamente dita. No pré-processamento, é feita a extração de características dos sinais sob análise visando reduzir a dimensionalidade do espaço de entrada para o estágio seguinte. Para tanto, utiliza-se normalmente uma transformação da forma de onda de entrada (tensão ou corrente) do domínio do

tempo para outro domínio, podendo ser para a frequência via Transformada de Fourier ou para o domínio tempo-frequência via transformada *wavelet* (GEREK et al., 2006). Para a extração de características, podem ser aplicadas relações como o cálculo de energia do sinal nesse novo domínio (LAZZARETTI et al., 2009) de modo a extrair o máximo de informações relevantes para o classificador, mantendo o compromisso com a redução de dimensionalidade para o próximo estágio (DONG et al., 2009).

Além desses métodos, existem outras abordagens adotadas para o pré-processamento não relacionadas diretamente com as transformações para o domínio da frequência ou do tempo-frequência. Em (GUO et al., 2005) foi utilizada programação genética como extrator de características, de forma que essas características possam ser dadas através de expressões analíticas geradas automaticamente, o que confere a esse método uma certa autonomia no que diz respeito à extração de características. Já em Faiz et al. (2007), foi utilizado o método de Prony para determinação das características modais dos sinais de entrada, tais como frequência, amplitude, amortecimento e deslocamento de fase. Em Gerek et al. (2006), utilizam-se estatísticas de ordem elevada procurando garantir que sejam mantidas informações relevantes para o processo de classificação, além de permitirem de forma conjunta a detecção de eventos. Já Dash e Chilukuri (2004) demonstram a possibilidade de se utilizar um sistema híbrido de transformada *wavelet* com correção de fase (Transformada-S) e filtros de Kalman com o objetivo de caracterizar sinais relacionados a eventos de Qualidade de Energia. Em (KEERTHIPALA, 1998; JANIK; LOBOS, 2006) são extraídas características de componentes simétricas e fasoriais dos sinais de entrada para a classificação.

No estágio de classificação, grande parte dos trabalhos utiliza redes neurais artificiais (CHIEN et al., 2002; MAHANTY; GUPTA, 2004; SILVA et al., 2006; SAMANTARAY; DASH, 2008; RAVIKUMAR et al., 2008; DEMIR, 2010a; MALATHI et al., 2010), sistemas Fuzzy (MAHANTY; GUPTA, 2007; REDDY; MOHANTA, 2007) ou modelos híbridos (VASILIC; KEZUNOVIC, 2005; ZHANG; KEZUNOVIC, 2007; TAN et al., 2007). Além desses, ainda existem trabalhos que realizam a classificação através de uma base de regras extraídas da análise dos sinais pré-processados (DONG et al., 2009). Uma característica comum desses trabalhos é que a definição do modelo final do classificador é fortemente dependente de parâmetros pré-selecionados no estágio de construção do classificador, principalmente com relação à estrutura no caso dos modelos neurais, isto é, número de neurônios nas camadas ocultas.

Em Bollen et al. (2007) e Demir (2010a) são desenvolvidas técnicas que otimizam a definição dos parâmetros de treinamento dos modelos do tipo Máquina de Vetor Suporte, conferindo ao processo de treinamento uma característica autônoma. No entanto, o método

utilizado (via validação cruzada única) é caracterizado por ser computacionalmente intensivo e não aborda questões relacionadas à seleção da camada de entrada das redes avaliadas. Em Demir (2010b) também utiliza-se um modelo do tipo Máquina de Vetor Suporte e são consideradas questões relacionadas à seleção de entradas, parâmetros do *kernel* e capacidade de generalização da rede de forma independente, visando à classificação de fenômenos de Qualidade de Energia.

Em Lazzaretti et al. (2009) são abordados conceitos de autonomia aplicados ao treinamento de modelos para classificação de eventos em redes de distribuição, com base nos conceitos expostos em Ferreira e Silva (2007), principalmente com relação à estrutura e capacidade de generalização do modelo de forma conjunta. Foram comparados três modelos para o processo de classificação: *Perceptron* Multicamadas, Funções de Base Radial e Máquina de Vetor Suporte, sendo que a seleção automática do modelo fornecia como resultado a definição das camadas de entrada e oculta dos modelos, considerando o controle de complexidade durante o processo de treinamento.

### 3.2.2 DETECÇÃO DE NOVIDADES

Analisando a característica de todos os classificadores desenvolvidos para análise automática de oscilografias em SEP, observa-se um ponto comum entre eles: a utilização de aprendizado supervisionado para o estágio de treinamento dos modelos, independentemente do método de reconhecimento de padrões adotado, voltado para uma classificação multiclasse.

Conforme demonstrado anteriormente, no cenário proposto neste trabalho observa-se que, para aplicação em registradores instalados em campo, é necessário considerar uma característica dinâmica dos registros armazenados. Ou seja, existe a possibilidade de novos eventos serem registrados, sem que esses tenham sido previamente inseridos na base de conhecimento do classificador multiclasse inicialmente definido.

Nesse sentido, é importante que o modelo a ser utilizado considere a possibilidade de descartar em um primeiro momento, eventos que não fazem parte do seu conjunto de classes conhecidas *a priori*. Para tanto, algumas abordagens são apresentadas na literatura, principalmente para aplicações envolvendo processamento de imagens. Observou-se que nenhuma dessas abordagens foi aplicada para análise de oscilografias em SEP, indicando que o estágio dedicado à detecção de novidades incluído neste trabalho tem potencial de ser uma contribuição para análise de eventos em SEP.

Novidades, no contexto de classificação de padrões, podem ser vistas como um padrão ou um conjunto de padrões inesperados, quando comparados a um conjunto de padrões consi-

derados normais. Para realizar a detecção de novidades, existem diferentes técnicas utilizando abordagens estatísticas (paramétricas e não-paramétricas) (MARKOU, 2003a) e técnicas baseadas em outros modelos de aprendizado de máquina, como redes neurais artificiais (MARKOU, 2003b). A seguir, foram selecionados alguns dos trabalhos que utilizam essas técnicas, os quais têm estreita relação com a abordagem escolhida para detecção de novidades neste trabalho.

O primeiro conjunto de trabalhos que considera a possibilidade de descartar eventos com grande probabilidade de não pertencerem ao grupo de classes previamente definidas são os classificadores que incluem a opção de rejeição<sup>1</sup> (HERBEI; WEGKAMP, 2005). No trabalho proposto por Ilonen et al. (2006), é desenvolvido um classificador baseado em Mistura de Gaussianas que contém um método de estimação da confiabilidade de uma dada classificação.

Já em Pan et al. (2008), é proposto um estimador não-paramétrico baseado em Janelas de Parzen, cuja característica também pode fornecer a opção de rejeição de novidades, análogo ao apresentado em Ilonen et al. (2006). Nesse sentido, muitos classificadores probabilísticos, ou mesmo baseados em distâncias, podem fornecer uma análise incluindo a opção de rejeição de novidades. Para tanto, é necessário definir um limiar para que um dado padrão seja considerado como pertencente a uma classe, ou caracterizado como novidade (IOANNIS, 2009).

Em Vieira-Neto (2006) é proposta uma metodologia para exploração e inspeção de ambientes por robôs, sinalizando objetos que não fazem parte do padrão de normalidade previamente definido. O processo de aprendizado pode ser considerado dinâmico, uma vez que padrões considerados como novidades passam a ser considerados como normais, à medida que são apresentados repetidamente ao classificador. Nesse trabalho são apresentadas e comparadas duas técnicas de detecção de novidades, sendo uma delas baseada na rede neural do tipo *Grow-When-Required* e outra técnica, utilizando Análise das Componentes Principais de forma incremental, ambas com características desejáveis para operação da metodologia em tempo real.

### 3.2.3 DETECÇÃO DE NOVIDADES E CLASSIFICAÇÃO MULTICLASSE

Os métodos apresentados até aqui são dedicados ao processo de classificação multi-classe e detecção de novidades de forma individual, não considerando explicitamente a possibilidade dessas duas tarefas serem executadas em conjunto. Nesse sentido, optou-se por apresentar demais trabalhos que considerem essa classificação conjunta.

<sup>1</sup>Opção de rejeição consiste em evitar a atribuição de uma classe a um dado padrão, em função da sua probabilidade de saída. Ou seja, caso o valor da probabilidade esteja abaixo de um dado valor de referência de incerteza, rejeita-se a caracterização desse padrão. Em alguns casos pode-se utilizar essa definição para caracterizar os padrões como novidades (BISHOP, 1995).

No trabalho proposto por Gao e Li (2011) é desenvolvido um classificador utilizando um Processo Gaussiano com capacidade de identificar padrões de entrada do tipo novidade. Além disso, esse classificador é capaz de reconhecer a classe associada a um novo padrão, caso esse não seja considerado uma novidade. Para tanto, utiliza a matriz de covariância, decorrente da modelagem através do Processo Gaussiano. Com esse critério, um novo padrão é considerado novidade se possui uma alta variância em relação aos dados das classes caracterizadas como normais. A grande vantagem dessa abordagem é o fato de não necessitar realizar a detecção de novidades e a classificação multiclasse em estágios separados, uma vez que a partir da análise da previsão do processo Gaussiano, é possível realizar essa classificação de forma conjunta. Esse modelo é comparado a outros métodos apresentados na literatura (SCHÖLKOPF; SMOLA, 2001) e apresenta desempenho superior a esses métodos para problemas de reconhecimento de impressões digitais e navegação autônoma em robótica móvel.

Uma segunda abordagem compreende a utilização de classificadores uma-classe (detectores de novidade) para cada classe conhecida do problema (HAO et al., 2007; JINGLONG et al., 2012). Nessa abordagem, um novo padrão é inicialmente testado em cada um dos classificadores uma-classe modelados e, caso não pertença a nenhuma das classes modeladas, pode ser considerado como novidade. Existem casos em que os padrões são aceitos por mais de um modelo, sendo necessária uma medida de similaridade para verificar qual o modelo uma-classe é mais provável para o padrão em questão. O trabalho proposto por Oliveira et al. (2008) apresenta uma rede neural probabilística que fornece a capacidade de detectar novidades, podendo ser utilizado no contexto multiclasse, desde que utilizado para cada classe do problema. Essa abordagem é exemplificada nos trabalhos de Golub et al. (1999) e Spinosa e Carvalho (2005).

### 3.2.4 IDENTIFICAÇÃO DE NOVAS CLASSES

A identificação de novas classes, ou a identificação de novas distribuições de dados que indiquem a existência de novas classes, é uma tarefa estritamente não-supervisionada e realizada, geralmente, através de modelos de agrupamento (SPINOSA, 2008; MILLER; BROWNING, 2003).

Os modelos de agrupamento (*clustering*) fornecem uma possível caracterização dos dados, agrupando padrões através da sua similaridade. Com isso, é possível utilizar a informação de um dado grupo para indicar a existência de um novo conceito, ou classe, dentro de um problema de classificação. Vale ressaltar que nessa modelagem, uma classe não é necessariamente formada por apenas um grupo (*cluster*), podendo inclusive, ser representada por inúmeros grupos. No entanto, a identificação de um dado agrupamento pode ser um indício de que existe uma

classe, independentemente de quantos grupos formam essa classe, auxiliando o especialista no processo de identificação de novas classes.

Nesse cenário, os métodos de agrupamento automático não-supervisionado são particularmente interessantes, pois não necessitam da definição do número de grupos *a priori*. Sendo assim, optou-se por apresentar uma revisão dos métodos de agrupamento automático que poderiam ser utilizados como parte do processo de identificação de novas classes.

### 3.2.5 AGRUPAMENTO AUTOMÁTICO

Dentre os diversos métodos de agrupamento automático atualmente utilizados para problemas de aprendizado de máquina, destacam-se basicamente dois grupos: modelos probabilísticos e modelos baseados em *kernel* (JAIN, 2010; FILIPPONE et al., 2008). Nos métodos probabilísticos, busca-se uma representação dos dados através da sua função densidade de probabilidade. Em geral, o número de funções a serem utilizadas (normalmente Gaussianas) é selecionado de forma automática, bem como o seus respectivos parâmetros (CHEUNG, 2005, 2003; KRISTAN et al., 2008; ZENG; CHEUNG, 2009).

Um modelo de agrupamento automático dessa natureza que merece destaque é o modelo proposto por (CHEUNG, 2005), denominado *Rival Penalized Expectation Maximization*. Esse modelo é baseado na Maximização da Expectativa para determinação dos parâmetros das funções Gaussianas que caracterizam a função densidade de probabilidade à luz dos dados. No entanto, o autor propõe uma modificação na formulação convencional, visando à seleção automática do modelo. Essa modificação consiste em ponderar a função de verossimilhança, de forma que seja possível realizar uma competição entre cada componente individual da mistura de Gaussianas, atualizando os parâmetros da função Gaussiana vencedora e de todas as demais funções que formam a distribuição. A grande vantagem desse modelo, na sua versão atual (ZHAO et al., 2010), é que não é necessário especificar parâmetros *a priori*. Apenas se insere o intervalo de grupos que serão considerados para realizar a modelagem. Ao final, é possível descartar os grupos que não contribuem para o resultado, uma vez que todos esses grupos são vazios (não contêm padrões associados a eles).

Já os métodos baseados em *Kernel* utilizam o mapeamento (implicitamente definido) para o espaço de características através de um *kernel*, sendo que o processo de agrupamento ocorre nesse novo espaço (DAS et al., 2008; CASTRO; CHALUP, 2002; LEE; LEE, 2006, 2005; WANG et al., 2011). Esses métodos possuem grande flexibilidade para determinação automática do número de grupos. No entanto, ainda representam forte dependência dos hiperparâmetros associados ao processo de treinamento dos modelos.

Um método de agrupamento que guarda relação com os modelos baseados em *Kernel* e merece destaque dentro da possibilidade de ser utilizado no contexto de agrupamento automático é conhecido como método espectral (FILIPPONE et al., 2008). Esses modelos são baseados na teoria de grafos e a ideia básica é construir um grafo ponderado utilizando os dados de entrada, de tal forma que cada nó represente o padrão de entrada e as conexões ponderadas representem a similaridade entre cada par de padrões de entrada. Com isso, o problema de agrupamento pode ser visto como um problema de partição de grafos, que pode ser resolvido através do uso da teoria espectral de grafos. Pode-se demonstrar que os métodos espectrais e os métodos baseados em *Kernel* possuem a mesma fundamentação matemática (FILIPPONE et al., 2008).

Dentro dos modelos de agrupamento espectral, existe uma proposta que aborda a possibilidade de realizar o agrupamento de forma automática. Esse modelo é conhecido como *Self-Tuning Spectral Clustering* (ZELNIK-MANOR; PERONA, 2004). Esse método propõe realizar um processo de escalonamento local para calcular a similaridade entre os pares de padrões de entrada. Além disso, é incluída uma análise da estrutura dos autovetores da matriz de similaridade para inferir o número de grupos do agrupamento. Esse modelo foi testado para problemas de segmentação em imagens e observou-se que o desempenho pode ser comparável com demais métodos de agrupamento automático.

Um último modelo que merece destaque aqui, é o modelo *X-Médias*. Esse algoritmo utiliza como base o agrupamento *K-Médias*. No entanto, ele insere a possibilidade de realizar a determinação automática e dinâmica do número de grupos (PELLEG; MOORE, 2000). No modelo *X-Médias*, os centros são determinados considerando o critério de seleção *Bayesian Information Criterion*. Dada sua simplicidade e rápida convergência, esse modelo pode ser um bom candidato para ser utilizado como parte do processo de identificação de classes.

### 3.2.6 CLASSIFICAÇÃO MULTICLASSE E IDENTIFICAÇÃO DE NOVAS CLASSES

Existe um último conjunto de trabalhos que apresenta uma modelagem mais ampla do problema, realizando a classificação multiclasse em conjunto com a detecção de novidades, além de permitir a inclusão de novas classes a partir das novidades identificadas, caracterizando um aprendizado contínuo para o modelo. Em Miller e Browning (2003) é aplicado um modelo baseado em Mistura de Gaussianas visando realizar todas as etapas para identificação de novas galáxias a partir de imagens adquiridas do espaço. Um método muito parecido foi proposto por Spinosa (2008), onde se utiliza um agrupamento *K-Médias* que permite a inclusão de novas classes ao longo da avaliação do classificador, além de realizar a expansão das classes

previamente conhecidas, tornando o processo mais abrangente e acompanhando alterações que possam existir nos padrões de entrada, tanto para a inclusão de novidades (novos conceitos), quanto para classes já conhecidas.

Através da combinação de classificadores uma-classe (TAX; DUIN, 2001), também é possível fazer com que um classificador multiclasse se expanda à medida que novas classes vão sendo identificadas, como mostram Tax e Duin (2008). Nesse trabalho é apresentado um modelo de normalização das saídas dos classificadores uma-classe para que esses possam ser combinados independentemente de sua característica de modelagem, seja ela probabilística ou com base em distâncias. Desse modo, pode-se utilizar um conjunto de classificadores uma-classe para realizar a classificação multiclasse e detecção de novidades, com a possibilidade de inserir um novo classificador uma-classe para uma nova classe, mantendo um desempenho superior ao obtido caso fosse utilizado apenas um modelo por classe.

Outro trabalho que merece destaque dentro dessa última abordagem é o trabalho apresentado por Escalera et al. (2011). Nessa metodologia os autores propõem uma alteração da codificação de saída das classes, visando utilizar os classificadores em um cenário de aprendizado *on-line*. O algoritmo é implementado em duas camadas, sendo uma delas realizada previamente, com base no conhecimento disponível inicialmente, e a segunda voltada para o aprendizado *on-line*. A vantagem desse algoritmo é que ele permite utilizar diferentes classificadores multiclasse como base, uma vez que o foco dos autores é o processo de codificação das saídas das classes.

Ainda nesse contexto, o trabalho proposto por Hempstalk e Frank (2008) apresenta uma comparação entre uma abordagem multiclasse, classificação binária e detecção de novidades para realizar a classificação multiclasse, considerando a possibilidade de novas classes surgirem ao longo do tempo. Os autores discutem a influência de se conhecer previamente alguns padrões de classes consideradas como novidades e qual o impacto desse conhecimento na escolha da abordagem de classificação. Adicionalmente, é apresentada uma forma de realizar uma comparação estatística justa entre todas as abordagens. Como conclusão, observa-se que existe uma forte dependência das características da base de dados na escolha da metodologia, uma vez que os resultados foram avaliados para bases de dados de diferentes naturezas.

Trabalhos mais recentes, como os métodos apresentados por Scheirer et al. (2013) e Bodesheim et al. (2013), procuram definir o problema de uma maneira mais ampla, formalizando a detecção de novidades e classificação multiclasse em um mesmo cenário. Ressalta-se que essas metodologias recentes, que resultam nos melhores desempenhos para problemas de visão computacional (SCHEIRER et al., 2013, 2014; COSTA et al., 2014), derivam da metodo-



logia presente em Scheirer et al. (2013), a qual contém considerável correlação com os modelos apresentados em Hao et al. (2007) e Wu e Ye (2009).

### 3.2.7 IDENTIFICAÇÃO DE NOVAS CLASSES EM FLUXO CONTÍNUO DE DADOS

Atualmente, existe um grande esforço por parte dos pesquisadores da área de aprendizado de máquina para aplicar técnicas de reconhecimento de padrões, principalmente envolvendo agrupamento automático, em dados gerados de forma contínua em ambientes dinâmicos. Esse cenário é comumente denominado de fluxo contínuo de dados (do inglês, *data stream*). Nesse tipo de aplicação, o acesso aos dados é bastante restrito e esparso, além de apresentar limitações de memória e de tempo de processamento. Uma das grandes dificuldades desse tipo de análise de dados é a dinâmica existente na característica das classes, bem como o surgimento de novas classes ao longo do tempo. Em função desse motivo, optou-se por apresentar um detalhamento dos trabalhos que abordam esse cenário na literatura.

Em Masud e Gao (2009) é apresentado um método para detecção de novas classes em um problema de classificação multiclasse, denominado *MineClass*. Além disso, os autores incluíram a possibilidade de alteração das características das classes ao longo do tempo. Essa alteração é denominada, em aplicações de fluxo contínuo de dados, de mudança de conceito (do inglês, *concept-drifting*). O método proposto naquele trabalho é completamente não-paramétrico, não necessitando assumir distribuições *a priori* sobre os dados. O método foi aplicado para diferentes bases de dados e os resultados comparados com a abordagem proposta por Spinosa (2008), mostrando que o *MineClass* apresenta resultado superior na maior parte das bases de dados utilizadas como referência.

Já em Masud et al. (2010a), é proposta uma continuação do trabalho apresentado em Masud e Gao (2009). A principal diferença é a inclusão de um módulo responsável por identificar novas características presentes nos dados de entrada. Para realizar essa identificação, o método denominado *DXMiner* emprega seleção de características durante o processo de treinamento dos modelos, identificando a mudança nos atributos que caracterizam os dados de forma dinâmica. O método foi testado em bases de dados de imagens e mostrou-se superior aos demais métodos aplicados em fluxo contínuo de dados, principalmente por fazer a consideração sobre a mudança das características dos dados de entrada, algo ainda não inserido nos demais métodos utilizados como base de comparação.

Em Masud et al. (2010b) são discutidos os mesmos aspectos apresentados em Masud et al. (2010a). A principal diferença é que naquele, são apresentados diferentes métodos para detecção de novas classes. Basicamente, são comparadas duas formas de identificar novida-

des. A primeira delas considera a diferenciação entre classes normais e novidades através de comparação com limiares, de forma parecida com o que propõem Tax e Duin (2008), porém incluindo uma característica adaptativa do limiar, de acordo com a variação dos conceitos das classes e a presença de novos atributos nos dados de entrada. Já a segunda abordagem, utiliza o coeficiente discreto de Gini. Essas duas formas de identificar novidades são utilizadas em um ambiente no qual várias classes podem surgir no mesmo intervalo de tempo, como imagens de satélite e categorização de assuntos na rede social *Twitter*.

Ainda nesse mesmo contexto, Al-Khateeb et al. (2012) propõem o tratamento de classes que surgem em um dado momento, desaparecem posteriormente e, depois de um certo tempo, voltam a surgir. Essas classes são denominadas de classes recorrentes em aplicação de fluxo contínuo de dados. Esse tipo de tratamento é importante, pois nesse cenário é possível que uma classe recorrente seja classificada como uma nova classe, em função de mudanças nas características das classes ao longo do tempo. Esse é um problema típico de bases de dados que envolvem categorização de assuntos em redes sociais. O método proposto se mostrou bastante eficaz na detecção de classes recorrentes, minimizando sua identificação como novidades.

Uma abordagem diferente para detecção de novidades é proposta em Zweig et al. (2012). Naquele trabalho, a identificação de novas classes é feita em dois níveis do processo de classificação. Inicialmente, um classificador mais geral é aplicado ao novo padrão. Se esse padrão for aceito por esse primeiro classificador, um segundo classificador mais específico é aplicado em outro nível do processo. Caso esse segundo classificador rejeite o padrão, considera-se que o padrão é uma novidade. Em ambos os classificadores o processo de detecção de novidades é feito através da comparação com limiares definidos empiricamente. O método foi aplicado em bases de dados de imagens e observou-se que o desempenho de detecção de novidades é bastante superior, quando comparado a um processo de detecção de novidades em apenas um estágio. No entanto, observou-se que para o processo como um todo, incluindo a classificação multiclasse, o desempenho global ainda é relativamente inferior, quando comparado a outros métodos.

O trabalho proposto por Farid et al. (2013) apresenta uma metodologia que contempla a classificação multiclasse, detecção de novidades e identificação de novas classes, utilizando métodos convencionais de agrupamento e classificação supervisionada. O modelo final é composto por cinco diferentes módulos, os quais incluem: ponderação das entradas, construção da árvore de decisão, agrupamento de novas classes, classificador adaptativo para mudanças de conceitos e classificação multiclasse incluindo detecção de novidades. Os resultados foram comparados com métodos convencionais de detecção de novidades, como os métodos *K-*

Vizinhos-Mais-Próximos e Mistura de Gaussianas, para diferentes bases de dados e mostraram-se superiores em todos os casos analisados.

Em resumo, as metodologias utilizadas nesses trabalhos formam a base escolhida para o presente trabalho, que tem por objetivo a classificação multiclasse em conjunto com a detecção de novidades e identificação de novas classes, promovendo um processo de classificação mais robusto e adaptado ao dinamismo existente no registro de eventos em redes de distribuição.

## 4 FUNDAMENTAÇÃO TEÓRICA

Os aspectos teóricos serão apresentados de acordo com a metodologia proposta para o trabalho. Para tanto, optou-se por organizar as seções com o intuito de facilitar a compreensão sequencial de cada módulo dentro da metodologia. Nesse sentido, o presente capítulo será apresentado na seguinte ordem:

- Serão inicialmente apresentados os métodos utilizados para detecção de instantes transitórios com base nos modelos autorregressivos, filtro de Kalman e Operador de Energia de Teager. Como o modelo de Representação de Dados Utilizando Vetores Suporte utiliza dados pré-processados para efetuar o processo de detecção, além de ser utilizado também para classificação, optou-se por incluir posteriormente sua fundamentação teórica na seção dedicada aos classificadores uma-classe (detectores de novidades).
- Na sequência serão discutidas as formas de pré-processar o sinal, tanto do ponto de vista de detecção, quanto de classificação. O foco dessa seção é a transformada *wavelet* discreta na sua versão convencional e na versão *packet*. Serão apresentados também os detalhes da utilização dessa transformada para detecção de eventos e a extração de características utilizada para classificação multiclasse, detecção de novidades e agrupamento automático.
- Definidos os métodos de pré-processamento e cálculos associados, serão apresentados os detalhes teóricos dos seguintes classificadores uma-classe no contexto de detecção de novidades: *K*-Médias, *X*-Médias, *K*-Vizinhos-Mais-Próximos e Representação de Dados Utilizando Vetores Suporte. Nessa seção também serão detalhadas as duas abordagens escolhidas para detecção de novidades e classificação multiclasse: DNG-CMC e DNI-CBS. Adicionalmente, será apresentado o modelo RDVS para segmentação e classificação de eventos (principalmente na abordagem DNI-CBS) e a utilização do método *X*-Médias como identificador de classes. Dentro da metodologia DNG-CMC utiliza-se um classificador baseado em Máquinas de Vetor Suporte para a classificação multiclasse. Como a motivação e os aspectos teóricos desse modelo envolvem muitos detalhes, optou-se por

inserir uma seção dedicada a essa etapa da metodologia.

- Finalmente, são detalhados os fundamentos estatísticos para análise e comparação de resultados de detecção, classificação e identificação de novas classes.

#### 4.1 MODELOS AUTORREGRESSIVOS

O modelo autorregressivo de um sinal discreto  $x(n)$  pode ser representado como uma combinação ponderada das amostras anteriores do sinal  $x(n-i)$ ,  $i = 1, 2, \dots, N$ , juntamente com um ruído branco  $w(n)$ , como demonstra a Equação (1) (BOLLEN, 1999):

$$x(n) = - \sum_{i=1}^N a_i x(n-i) + b_0 w(n), \quad (1)$$

sendo  $N$  a ordem do modelo estimado e  $b_0$  uma ponderação do ruído. A função de transferência desse modelo  $H_{AR}(z)$ , cuja entrada é o próprio ruído branco  $w(n)$ , pode ser representada pela Equação (2):

$$H_{AR}(z) = \frac{b_0}{1 + \sum_{i=1}^N a_i z^{-i}}. \quad (2)$$

A estimação dos parâmetros  $a_i$  pode ser dada através da resolução das equações normais, representadas na Equação (3):

$$\sum_{m=1}^N a_m r_x(k, m) = -r_x(k, 0) \quad k = 1, \dots, N, \quad (3)$$

sendo que  $r_x(k, m) = E[x(n-k)x(n-m)]$  é a autocorrelação de  $x(n)$  e o operador  $E(\cdot)$  representa o cálculo da esperança matemática (AGUIRRE, 2007). Uma vez determinados os valores de  $a_i$ , pode-se estimar  $b_0$  através do seguinte conjunto de equações:

$$|b_0|^2 = r_x(0, 0) + \sum_{m=1}^N a_m r_x(0, m). \quad (4)$$

A partir dessas exposições, a representação do espectro de magnitude do sinal  $A(e^{j\omega})$  pode ser dada por:

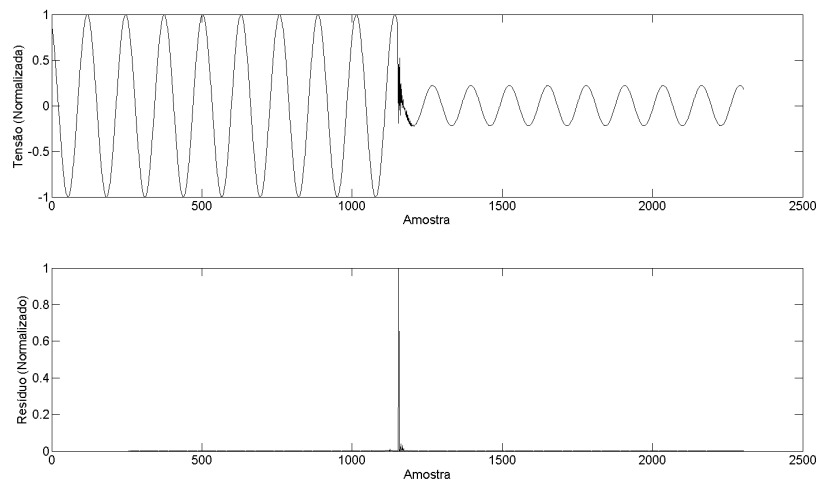
$$A(e^{j\omega}) = |H_{AR}e^{j\omega}| = \left| \frac{b_0}{1 + \sum_{i=1}^N a_i e^{-j\omega_i}} \right|. \quad (5)$$

Isso implica que cada um dos coeficientes  $a_i$  está diretamente associado à magnitude dos harmônicos  $\omega_i = 2\pi f_i$  do sinal, sendo  $f_i$  a frequência do  $i$ -ésimo harmônico.

O último estágio dessa estimação, e o mais importante do ponto de vista de detecção de eventos, é o cálculo dos resíduos do modelo. No caso de modelos autorregressivos, os resíduos são calculados através da Equação (6):

$$e(n) = \hat{x}(n) + \sum_{i=1}^N a_i \hat{x}(n-i). \quad (6)$$

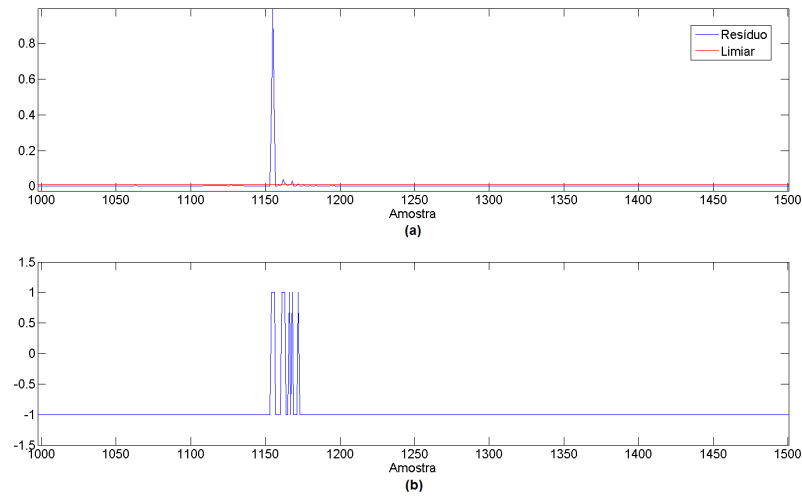
Para exemplificar, na Figura 23 está representado o valor normalizado do resíduo calculado para uma fase de um evento relacionado a um curto-circuito monofásico. Observa-se que nos instantes em que ocorre a transição entre o estado pré-falta e o estado em falta, os resíduos apresentam elevada magnitude. Com isso, tem-se os instantes de tempo em que ocorrem as variações nos sinais (disparo), fazendo com que seja possível a segmentação dos eventos do sinal.



**Figura 23: Cálculo dos resíduos para modelos autorregressivos.**

Para realizar a segmentação, é necessária a aplicação de um limiar de comparação que forneça a separação entre trechos em transição ou em regime quase-estacionário. No exemplo da Figura 23, poder-se-ia determinar um limiar no valor de 0,01 para realizar a segmentação. Com isso, instantes de tempo cujo valor do resíduo é superior ao valor do limiar poderiam ser considerados como transições entre estados (recebendo valor 1, por exemplo) e, os instantes de tempo com o resíduo inferior ao limiar, seriam caracterizados por estados quase-estacionários, ou mesmo, o sinal em regime permanente (recebendo valor -1, por exemplo). Essa operação está representada na Figura 24.

Uma possível forma de se estabelecer o cálculo automático do limiar baseia-se na proposta de Ukil e Zivanovic (2005a, 2005b, 2006a), Ukil e Ivanovi (2007), Ukil e Zivanovic (2008), cujos aspectos teóricos são apresentados em Donoho e Johnstone (1994). A ideia



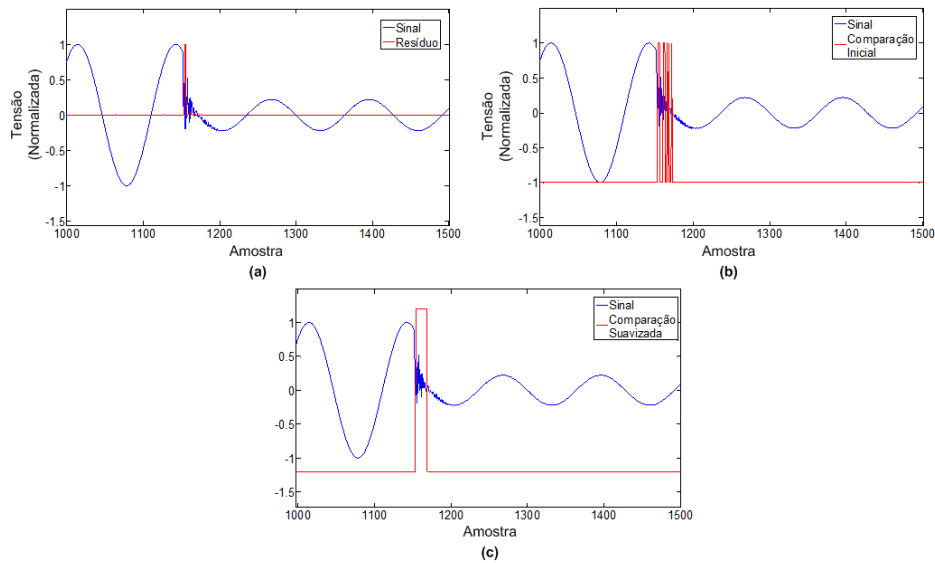
**Figura 24: Exemplo do disparo para modelos autorregressivos. (a) Comparação entre o resíduo e o limiar e (b) atribuição dos instantes transitórios.**

principal desta metodologia é a aplicação do conceito de um limiar universal, cujo cálculo é fundamentado no desvio médio absoluto do sinal. Sua proposição inicial era voltada para redução de ruído de um sinal no domínio *Wavelet*. Porém, sua aplicação pode ser bastante útil para determinação automática de limiares em resíduos de modelos de estimação espectral (UKIL; ZIVANOVIC, 2007). Sua fundamentação teórica considera que é possível utilizar esse limiar no cenário em que o ruído a ser descartado possui uma distribuição normal com média nula e variância  $\sigma_r$  ( $N(0, \sigma_r)$ ). Sendo assim, a determinação deste limiar é dada por:

$$\text{limiar} = \frac{\bar{a}}{0,6725} \sqrt{2 \ln(N)}, \quad (7)$$

sendo  $N$  o comprimento do sinal,  $a = |x - \bar{x}|$ ,  $\bar{x}$  o valor mediano do resíduo  $x$  e  $\bar{a}$  o valor mediano de  $a$ .

Como pode ser observado na Figura 24, existem instantes de tempo cujo valor do resíduo calculado é inferior ao valor limiar, muito embora esses instantes pertençam ao transitório de mudança de estado. Esse fenômeno faz com que seja necessário mais um estágio para a segmentação do sinal sob análise, conforme demonstram Ukil e Zivanovic (2006a), de forma a suavizar a resposta do método de detecção. Basicamente, deve-se considerar que os picos muito próximos devem ser agrupados, de forma que os instantes de tempo que possuem o resíduo inferior ao limiar, mas estão próximos de instantes de tempo cujo resíduo é superior ao limiar, devem ser agrupados e considerados como parte do transitório do sinal. Com isso, obtém-se todo o processo de segmentação, conforme exposto na Figura 25.



**Figura 25: Exemplo de segmentação para modelos autorregressivos. Cálculo dos Resíduos do sinal (a), seguido da limiarização através do cálculo automático do limiar (b). Suavização (c) da limiarização apresentada em (b).**

## 4.2 FILTRO DE KALMAN

O filtro de Kalman (FK) é um tipo especial de filtro, cuja principal característica é a estimação recursiva ótima dos estados e parâmetros de um modelo (KALMAN, 1960). O filtro de Kalman realiza essa estimação através de um processo de predição-correção no qual são representadas as relações entre as medidas e os estados, bem como as relações entre os estados atuais e anteriores (AGUIRRE, 2007; MORETO; ROLIM, 2010a). Para tanto, faz-se necessária a representação do modelo em equações de estado, conforme demonstra a sua formulação básica, dada pelas equações de estado (8) e observação (9):

$$\underline{x}(n) = \underline{A}(n-1)\underline{x}(n-1) + \underline{w}(n), \quad (8)$$

$$\underline{z}(n) = \underline{C}(n)\underline{x}(n) + \underline{v}(n), \quad (9)$$

sendo  $\underline{x}(n)$  o vetor de variáveis de estado,  $\underline{A}(n)$  a matriz responsável pela transição entre dois estados subsequentes,  $\underline{C}(n)$  a matriz de observação que relaciona os estados  $\underline{x}(n)$  com as medidas (observações)  $\underline{z}(n)$ ,  $\underline{w}(n)$  é o ruído do modelo com média nula e matriz de covariância dada por  $E[\underline{w}(n)\underline{w}^T(n)] = \underline{Q}$  e  $\underline{v}(n)$  o ruído de medição com média nula e matriz de covariância dada por  $E[\underline{v}(n)\underline{v}^T(n)] = \underline{Q}$ . Ambos os ruídos são considerados ruído branco.

Antes de se aplicar o filtro de Kalman, é necessário definir as variáveis de estado  $\underline{x}(n)$  de acordo com o problema a ser modelado. Uma vez definidas essas variáveis, as equações de estado (8) e observação (9) podem ser determinadas assumindo as matrizes  $\underline{A}(n)$  e  $\underline{C}(n)$  com



valores constantes. Para o problema de detecção de eventos com base em estimação espectral (BOLLEN; GU, 2006), pode-se assumir a seguinte formulação para o vetor de entradas (medidas)  $\underline{z}(n)$ :

$$z(n) = \sum_{k=1}^K s_k(n) + v(n), \quad (10)$$

sendo o conteúdo harmônico representado por  $s_k = \underline{A}_k e^{jn\omega_k}, k = 1, \dots, K$ , para um total de  $K$  harmônicos a serem estimados. Sendo assim, pode-se definir o vetor de estados  $\underline{x}(n)$  como:

$$\underline{x}(n) = [s_{1,r}(n) \ s_{1,i}(n) \ \dots \ s_{K,r}(n) \ s_{K,i}(n)]^T, \quad (11)$$

sendo  $s_{1,r}(n)$  e  $s_{1,i}(n)$  as partes real e imaginária de  $s_r(n)$  no instante  $n$ , respectivamente, representados pelas equações (12) e (13):

$$s_{1,r}(n) = \text{Re}[s_k(n)] = a_k \cos(n\omega_k + \varphi_k), \quad (12)$$

$$s_{k,i}(n) = \text{Im}[s_k(n)] = a_k \text{sen}(n\omega_k + \varphi_k). \quad (13)$$

Nesta equação,  $a_k$  e  $\varphi_k$  representam a magnitude e fase do  $k$ -ésimo harmônico, respectivamente. Considerando as variáveis  $s_{k,r}(n+1)$  e  $s_{k,i}(n+1)$  no instante  $n+1$ , é possível estabelecer as seguintes relações:

$$s_{k,r}(n+1) = a_k \cos[(n+1)\omega_k + \varphi_k] = s_{k,r} \cos(\omega_k) - s_{k,i} \text{sen}(\omega_k), \quad (14)$$

$$s_{k,i}(n+1) = a_k \text{sen}[(n+1)\omega_k + \varphi_k] = s_{k,r} \text{sen}(\omega_k) + s_{k,i} \cos(\omega_k). \quad (15)$$

Combinando as equações (11), (14) e (15), obtém-se a matriz de transição entre estados  $\underline{A}(n)$ , como mostra a Equação (16):

$$\begin{bmatrix} \cos(\omega_1) & -\text{sen}(\omega_1) & 0 & \dots & 0 & 0 & 0 \\ \text{sen}(\omega_1) & \cos(\omega_1) & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & \cos(\omega_K) & -\text{sen}(\omega_K) \\ 0 & 0 & 0 & \dots & 0 & \text{sen}(\omega_K) & \cos(\omega_K) \end{bmatrix}, \quad (16)$$

sendo  $\omega_k$  um parâmetro fornecido *a priori*, relacionado ao conteúdo harmônico do sistema, uma vez que  $\omega_k = k\omega_0$  e  $k = 1, 2, \dots, K$ ,  $\omega_0 = 2\pi f_0/f_s$ , com  $f_s$  respondendo pela frequência de amostragem escolhida e  $f_0$  a frequência fundamental.

Através da representação de  $\underline{z}(n)$  como um somatório ponderado de senos e cossenos

e um ruído branco  $v(n)$ ,  $\underline{\underline{C}}(n)$  torna-se um vetor de comprimento  $2K$ :

$$\underline{\underline{C}}(n) = [ 1 \ 0 \ \dots \ 1 \ 0 ]. \quad (17)$$

Além disso, os vetores  $\underline{v}(n) = [ v(n) \ 0 \ \dots \ 0 \ 0 ]^T$  e  $\underline{w}(n) = [ w(n) \ 0 \ \dots \ 0 ]^T$  são ambos de comprimento  $2K$ .

Depois de definida a modelagem do problema, o cálculo do filtro de Kalman é feito de maneira recursiva, utilizando uma estimativa inicial do vetor de estados e da matriz de covariância  $\hat{\underline{\underline{P}}}(n-1)$ . Com esses valores calcula-se o ganho do filtro  $\underline{\underline{K}}(n)$  para o instante  $n$ :

$$\underline{\underline{K}}(n) = \hat{\underline{\underline{P}}}(n-1)[\underline{\underline{C}}(n)]^{*T} [[\underline{\underline{C}}(n)]^{*T}\hat{\underline{\underline{P}}}(n-1)[\underline{\underline{C}}(n)]^{*T} + \underline{\underline{Q}}]^{-1}, \quad (18)$$

sendo  $\underline{\underline{Q}}$  a covariância associada ao ruído de medição e a operação  $*$  o conjugado complexo. Com o ganho, pode-se atualizar a matriz de covariância e o vetor de estados, utilizando também a própria medida  $\underline{z}(n)$  para correção:

$$\hat{\underline{\underline{P}}}(n) = \hat{\underline{\underline{P}}}(n-1)(I - \underline{\underline{K}}(n)\underline{\underline{C}}(n)), \quad (19)$$

$$\hat{\underline{x}}(n) = \hat{\underline{x}}(n-1) + \underline{\underline{K}}(n)[\underline{z}(n) - \underline{\underline{C}}(n)\hat{\underline{x}}(n-1)], \quad (20)$$

sendo o termo  $[\underline{z}(n) - \underline{\underline{C}}(n)\hat{\underline{x}}(n-1)]$  o próprio resíduo de predição. O último estágio do processo é a atualização do vetor de estados e de sua matriz de covariância:

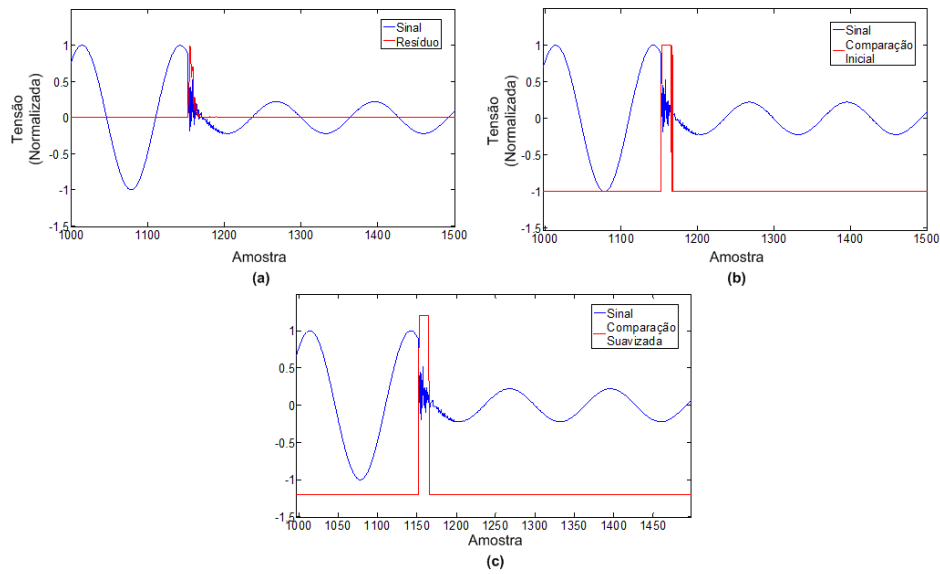
$$\underline{x}(n+1) = \underline{\underline{A}}(n)\underline{x}(n), \quad (21)$$

$$\hat{\underline{\underline{P}}}(n+1) = \underline{\underline{A}}(n)\hat{\underline{\underline{P}}}(n)[\underline{\underline{A}}(n)]^{*T} + \underline{\underline{Q}}_w. \quad (22)$$

Esse processo é repetido para todas as amostras do sinal de entrada. A partir dos resíduos estimados a cada amostra, pode-se empregar a mesma metodologia de comparação com o limiar automaticamente calculado e suavização apresentados para os modelos autorregressivos. Um exemplo dessa metodologia utilizando os resíduos calculados a partir de uma estimação realizada com Filtro de Kalman está apresentada na Figura 26.

#### 4.3 OPERADOR DE ENERGIA DE TEAGER

O Operador de Energia de Teager é um operador não-linear, cuja finalidade inicial era de extrair a energia do processo mecânico associado a um sinal composto por uma única frequência (KAISER, 1993; TAVARES et al., 2011). A utilização desse operador para detecção de anormalidades em sinais temporais fundamenta-se no trabalho proposto por Kaiser (1993),



**Figura 26: Exemplo de segmentação para modelos utilizando filtro de Kalman. Cálculo dos resíduos do sinal (a), seguido da limiarização através do cálculo automático do limiar (b). Suavização (c) da limiarização apresentada em (b).**

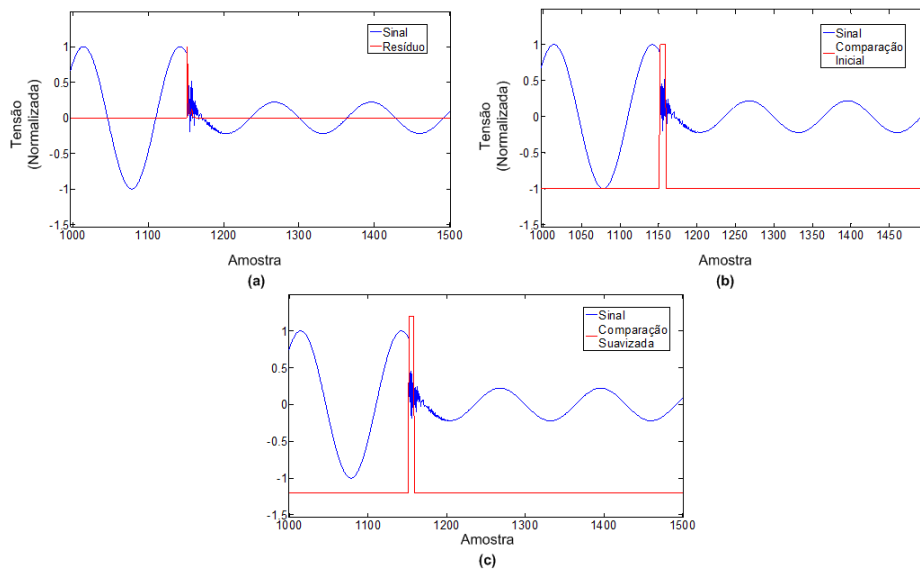
uma vez que quando aplicado a sinais não-estacionários, é possível obter o realce dos instantes de variações abruptas do sinal.

A Equação (23) mostra o cálculo desse operador para um sinal de entrada  $x(n)$ . Fica clara a simplicidade desse método, uma vez que esse operador necessita apenas das informações de amostra atual, anterior e posterior para sua estimação:

$$\Psi[x(n)] = [x(n)]^2 - x(n+1)x(n-1). \quad (23)$$

A partir do cálculo de  $\Psi[x(n)]$  para cada amostra do sinal, pode-se empregar a mesma metodologia de comparação com o limiar e suavização apresentados para os modelos autorregressivos e Filtro de Kalman. Um exemplo dessa metodologia está apresentado na Figura 27.

A grande diferença para um modelo baseado no Operador de Energia de Teager, em relação aos modelos autorregressivos e com filtro de Kalman, é que não é possível aplicar a mesma metodologia para determinação automática do limiar através do limiar universal (DONOHO; JOHNSTONE, 1994). As restrições teóricas impostas na aplicação desse cálculo não são garantidas para o Operador de Energia de Teager. Desse modo, a determinação automática do limiar segue uma proposta desenvolvida para esse modelo no presente trabalho, detalhada na seção a seguir.



**Figura 27: Exemplo de segmentação para modelos utilizando operador de Energia de Teager. Cálculo de  $\Psi[x(n)]$  (a), seguido da limiarização através da determinação empírica do limiar (b). Suavização (c) da limiarização apresentada em (b).**

#### 4.3.1 DETERMINAÇÃO AUTOMÁTICA DO LIMIAR PARA O OET

Para realizar o cálculo automático do limiar no caso do OET, optou-se por uma abordagem empírica, com base nas oscilografias normais (oscilografias sem a presença de estados transitórios). O processo de determinação do limiar inicia com a seleção de um número fixo de oscilografias normais, as quais servirão de base para a estimação do limiar. O ideal é que esteja disponível um número variado de oscilografias para essa estimação. No caso dos dados reais desse trabalho, foi utilizado um conjunto com 50 oscilografias normais.

A partir desse conjunto, é realizado o cálculo da energia pelo OET para cada uma dessas oscilografias. Em seguida, os valores calculados de energia são ordenados em ordem crescente utilizando todas as 50 oscilografias. O valor de energia localizado na posição correspondente a 99% do comprimento total do vetor ordenado, é definido como o limiar de detecção. Isso seria análogo a um teste de hipótese para uma distribuição normal com nível de significância de 1%, porém sem restrições com relação à distribuição dos dados. Ou seja, o valor do limiar escolhido seria correspondente a um determinado valor de energia, cuja probabilidade de ocorrência é igual a 1%. Isso indica que o limiar escolhido estaria no limite superior do ruído existente nas oscilografias normais – limite de caracterização de um estado quase-estacionário e um estado transitório.

Com essa escolha, assume-se um limiar constante para todas as análises feitas através do modelo baseado no OET, diferentemente da abordagem aplicada para os modelos baseados

em estimação espectral, com base no limiar universal (DONOHO; JOHNSTONE, 1994). Além disso, para a proposta de limiarização automática aqui apresentada, é necessário existir um conjunto prévio de oscilografias normais para a determinação desse limiar. No capítulo dedicado aos resultados, serão discutidos os impactos dessas duas restrições.

#### 4.4 PRÉ-PROCESSAMENTO

##### 4.4.1 TRANSFORMADA WAVELET

No início dos anos 1980, um grupo de pesquisadores franceses introduziu uma nova transformada tempo-frequência, denominada Transformada *Wavelet* (MALLAT, 1999), cuja principal característica é o tamanho variável da janela, permitindo que sinais com conteúdo bastante diversificado em frequência pudessem ser analisados com boa resolução temporal em cada faixa de frequência (DELMONT, 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, sendo normalmente denominada de análise multirresolução. Matematicamente, a análise *wavelet* de um sinal contínuo  $x(t)$  pode ser dada através da Equação (24):

$$TWC(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t) \psi \left( \frac{t-b}{a} \right) dt, \quad (24)$$

sendo  $a$  o fator de escala,  $b$  o fator de translação e  $\Psi$  a *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 (25):

$$x(t) = \frac{1}{\sqrt{a}} \int \int TWC(a, b) \psi \left( \frac{t-b}{a} \right) da db. \quad (25)$$

Uma restrição fundamental da transformada *wavelet* é que o sinal sob análise deve ter energia finita para que seja possível cobrir todo seu espectro de frequência. 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:

$$|\Psi(\omega)|^2|_{\omega=0} = 0. \quad (26)$$

Uma consequência dessa propriedade é que as *wavelets* devem possuir um espectro semelhante a um espectro de filtro banda passante. Isso é fundamental para a construção do algoritmo rápido que será detalhado posteriormente. Além disso, a média da função *wavelet*-mãe calculada no tempo deve ser zero, conferindo a essa função, o aspecto de onda:

$$\int_{-\infty}^{\infty} \psi(t) dt = 0. \quad (27)$$

Já a condição de regularidade, mostra que a *wavelet*-mãe deve possuir rápido decaimento para zero na bordas, caracterizando a duração finita da *wavelet*-mãe. Para se determinar a regularidade de uma função *wavelet* é necessário estimar o número de momentos nulos associados (DAUBECHIES, 1992). 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 da função, a qual está associada com o deslocamento de fase durante a análise *wavelet*;
- 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.

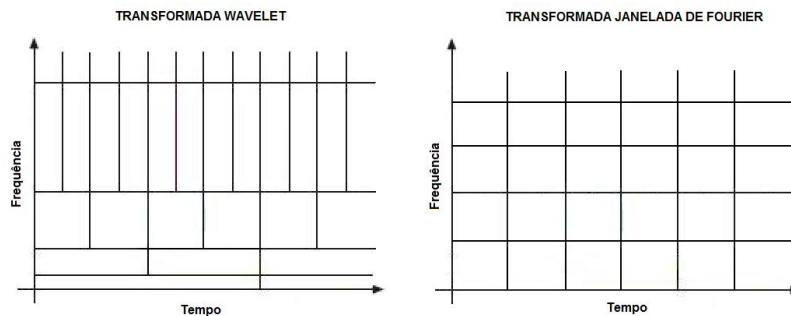
No decorrer da apresentação das funções *wavelet* essas questões serão retomadas, complementando algumas de suas principais características.

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 seu cálculo, 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 reduz o problema da redundância relacionado à transformação. A Equação (28) mostra o cálculo dessa transformada:

$$TWD(m, k) = \frac{1}{\sqrt{a_0^m}} \sum_n x(n) \psi \left( \frac{n - kb_0 a_0^m}{a_0^m} \right), \quad (28)$$

sendo  $a_0$  o fator de escala,  $b_0$  o fator de translação (ambos função de um parâmetro de escala  $m$  e do parâmetro inteiro  $k$ ) e  $n$  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. A Figura 28 mostra o gráfico tempo-frequência, resultado da aplicação da transformada *wavelet* discreta, em comparação com o gráfico de tempo-frequência gerado pela transformada Janelada de Fourier, assinalando a diferença entre ambas as representações.



**Figura 28: Espectros Tempo-Frequência.**

**Fonte: (DELMONT, 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.

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 ao número de dilatações e compressões necessárias para cobrir todo o espectro. A cada nova dilatação da *wavelet*-mãe no tempo, 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 esse 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 se pode 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. 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 dessa função de escalonamento não pode introduzir perda de informação na transformação.

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 29, mostra a abordagem em questão.

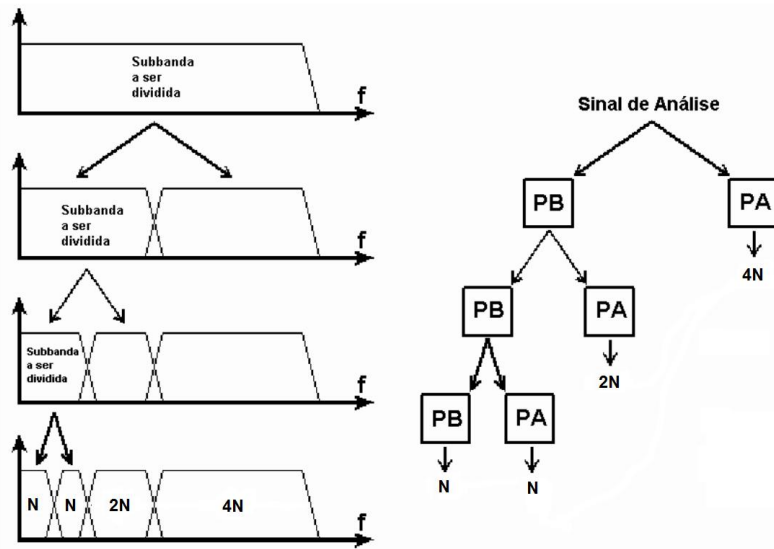
As saídas dos filtros passa-altas (PA), com  $4N$ ,  $2N$  e  $N$ , são denominadas de detalhes, já que nelas estão contidas as componentes de alta frequência do sinal. As saídas dos filtros passa-baixas (PB) são denominadas aproximações e nelas estão as componentes de frequência responsáveis por caracterizar o sinal, como representado na Figura 30.

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 Equação (29):

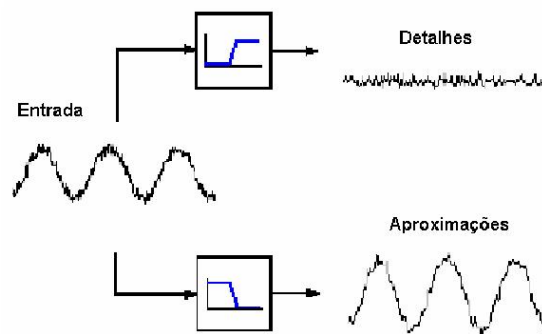
$$h[L-1-n] = (-1)^n l[n], \quad (29)$$

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





**Figura 29: Decomposição do sinal através do banco de filtros.**

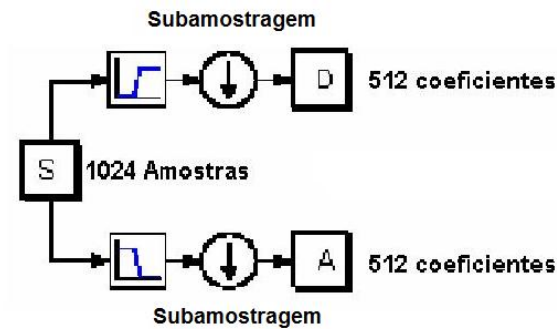


**Figura 30: Aproximações e detalhes.**

Fonte: (MATHWORKS, 2009)

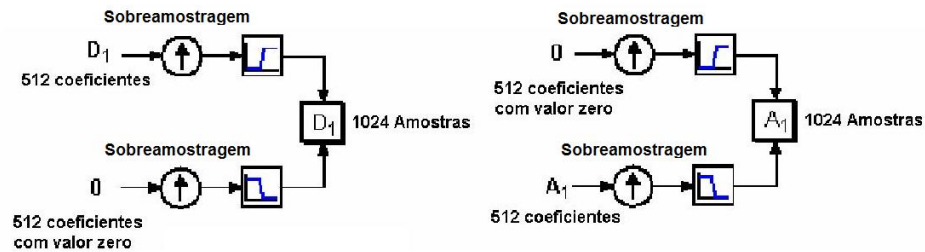
como *quadrature mirror filters* (filtros em quadratura espelhada). No caso de funções *wavelet* não simétricas, os filtros de decomposição e recomposição são diferentes e podem ser gerados a partir da função de escalonamento (filtro passa-baixas) e a partir da *wavelet*-mãe (filtro passa-altas).

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 (transformada inversa) utiliza a mesma ideia, mas ao invés de subamostrar, ela insere amostras (com valor zero) no sinal, para se obter no fim, um sinal com o mesmo número de amostras do sinal inicialmente decomposto. Essa metodologia fica mais clara nas Figuras 31 e 32:



**Figura 31: Decomposição em um nível para 1024 amostras.**

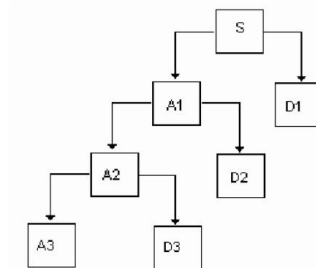
Fonte: (MATHWORKS, 2009)



**Figura 32: Recomposição em um nível para 512 coeficientes.**

Fonte: (MATHWORKS, 2009)

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 33.



**Figura 33: Reconstrução em um nível para 512 coeficientes.**

Fonte: (MATHWORKS, 2009)

Na figura 33, 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 (30), que calcula a transformada *wavelet* discreta com a Equação (31), relacionada com a convolução discreta de um filtro de resposta finita ao impulso:

$$TWD(m, k) = \frac{1}{\sqrt{a_0^m}} \sum_n x(n) \Psi(a_0^{-m}n - kb_0), \quad (30)$$

$$y(n) = \frac{1}{c} \sum x(k)h(n - k), \quad (31)$$

sendo  $h(\cdot)$  a resposta ao impulso do filtro. Assim sendo, pode-se inferir que o termo  $\Psi(a_0^{-m}n - kb_0)$  é a resposta ao impulso do filtro da equação da transformada *wavelet* discreta, caracterizando uma implementação prática para essa transformada. Neste trabalho foi utilizado o algoritmo proposto por Mallat (1999).

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). As funções *wavelet* descritas a seguir são as principais funções utilizadas em aplicações de SEP (LAZZARETTI et al., 2013):

- Haar: É a *wavelet* mais simples e, em função de não ser continuamente diferenciável, essa *wavelet* possui poucas aplicações atualmente (DELMONT, 2007).
- Daubechies: Introduzidas inicialmente pela pesquisadora Ingrid Daubechies (DAUBECHIES, 1992), essas *wavelets* possuem uma característica bastante relevante relacionada com sua ortonormalidade e a capacidade de gerar suporte compacto. O suporte compacto está associado à ordem da *wavelet*, ou seja, quanto maior a ordem do filtro (fator  $Q$ ), 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.
- 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.
- 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,

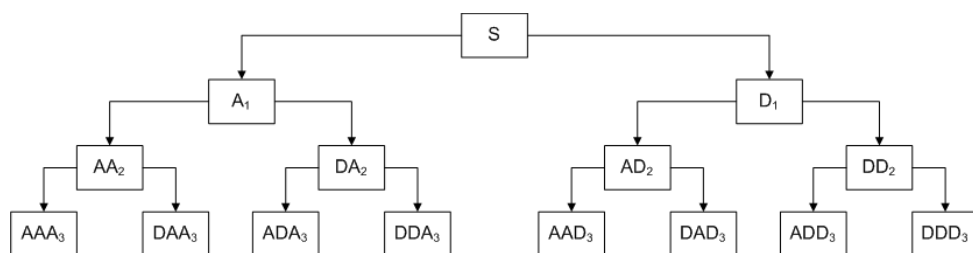
quanto para a função escala e são caracterizadas por apresentar uma característica quase simétrica.

- *Wavelets* Biortogonais: Foram desenvolvidas para possuir simetria e suporte compacto, sendo esse o principal diferencial sobre as demais famílias de *wavelets*.

Neste trabalho foram utilizadas apenas as *wavelets* de Daubechies (*dbN*), pois como demonstram Lazzaretti et al. (2013) e Delmont (2007), essa família de *wavelets* apresenta os melhores desempenhos para classificação de formas de onda de tensão em sistemas elétricos de distribuição.

#### 4.4.2 TRANSFORMADA WAVELET PACKET DISCRETA

Segundo Mallat (1999), a transformada *Wavelet Packet* discreta TWPD divide o sinal de análise em diferentes faixas de frequência, através de um algoritmo associado a um banco de filtros, bastante parecido ao processo aplicado pela TWD. A diferença desse algoritmo é a sua velocidade de processamento maior, além de outras características que resultam em uma análise mais elaborada do sinal. No caso da TWD, apenas a saída do filtro passa-baixas é decomposta nos próximos níveis. Já na TWPD, as saídas dos filtros passa-altas e passa-baixas são decompostas nos próximos níveis, como mostra a Figura 34. Desse modo, no final de uma decomposição com  $N$  níveis, obtém-se  $2^N$  sinais de no domínio *Wavelet*. Neste trabalho foram utilizados apenas os últimos níveis da TWPD, como os níveis  $AAA_3$ ,  $DAA_3$ ,  $ADA_3$ ,  $DDA_3$ ,  $AAD_3$ ,  $DAD_3$ ,  $ADD_3$  e  $DDD_3$  representados na Figura 34



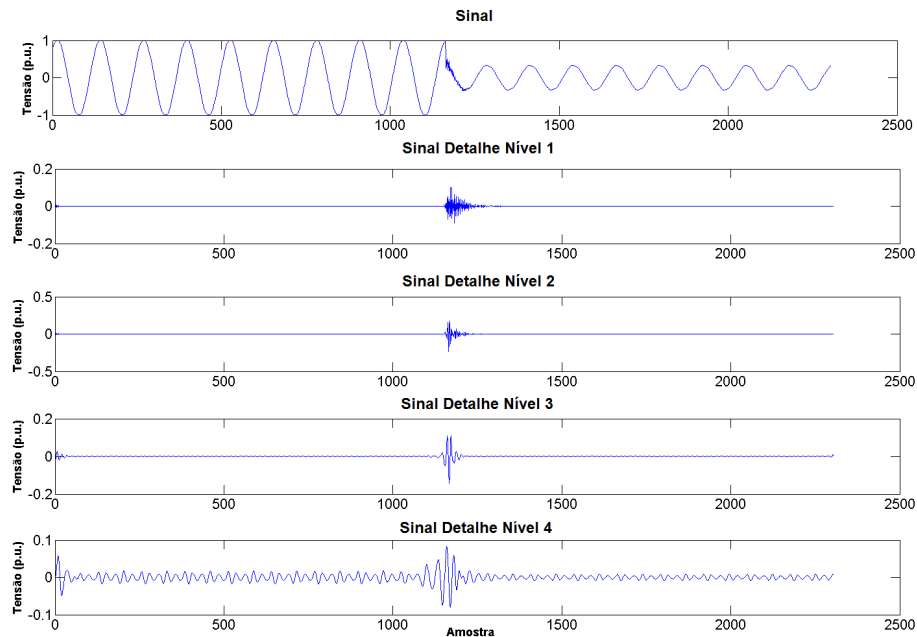
**Figura 34: Exemplo de decomposição utilizando a TWPD.**

**Fonte: (MATHWORKS, 2009)**

#### 4.4.3 APLICAÇÃO PARA DETECÇÃO DE EVENTOS

A utilização da TWD está relacionada com a sua característica de localização temporal para diferentes bandas de frequência, o que facilita o processo de identificação das transições

abruptas do sinal. Desse modo, as transições existentes na oscilografia podem ser mais bem caracterizadas de acordo com seu conteúdo de frequência nos diferentes níveis da TWD. Como exemplo da vantagem desse procedimento, observa-se na Figura 35, uma decomposição em cinco níveis da TWD utilizando a *wavelet*-mãe Daubechies-20, para um sinal com frequência de amostragem de 7680 Hz.



**Figura 35: Exemplo de detecção utilizando a TWD.**

No período de transição do evento, nota-se que os coeficientes ao longo dos níveis apresentam variações significativas, correspondentes à variação de frequência correspondente. Para que seja possível essa caracterização, é necessário selecionar a *wavelet*-mãe de forma adequada à necessidade de detecção de eventos. As *wavelets*-mãe da família Daubechies são normalmente as *wavelets* selecionadas para detecção de eventos, dado seu suporte compacto<sup>1</sup>, elevado número de momentos nulos<sup>2</sup> e o fato de possuírem filtros de fase mínima (UKIL; ZIVANOVIC, 2007).

Uma das grandes dificuldades verificadas na utilização da TWD para detecção de eventos, é a composição dos diferentes níveis dessa transformada para realizar a segmentação como um todo. Em geral, aplica-se apenas o primeiro nível como referência para limiarização do sinal de entrada (SILVA et al., 2006; UKIL; ZIVANOVIC, 2008). No presente trabalho é apresentada uma forma de compor os diferentes níveis da TWD (Seção 4.8.7), maximizando a possibilidade

<sup>1</sup> O suporte compacto garante que grande parte da energia de uma função *wavelet* esteja restrita em um intervalo finito, permitindo a localização espacial dessa transformada.

<sup>2</sup> O número de momentos nulos define, indiretamente, a suavidade de uma função *wavelet*.

de identificar um instante transitório, o qual pode ser caracterizado em diferentes níveis da TWD, com mostra a Figura 35.

#### 4.5 EXTRAÇÃO DE CARACTERÍSTICAS PARA CLASSIFICAÇÃO

Para os modelos de classificação multiclasse, detecção de novidades e agrupamento automático, é necessário extrair do pré-processamento *wavelet* o máximo de informações dos sinais em análise, sem perda de informação relevante. Para tanto, é necessária a extração de características com base nas várias sub-bandas da TWD ou da TWPD. O objetivo principal do cálculo por sub-banda ou raia está relacionado a uma melhor caracterização do conteúdo de frequência dos eventos, além de uma redução da dimensionalidade para o sistema de classificação (LAZZARETTI et al., 2009). Em cada sub-banda, foi aplicado o cálculo da energia dos sinais decompostos no domínio *wavelet*, a qual compreende a forma de extração de características mais usual para classificação de eventos (DEMIR, 2010b). No presente trabalho foi utilizada a energia absoluta ( $E_i$ ) de cada nível  $i$  da TWD, relacionado com o seu respectivo sinal detalhe  $s$  com  $N$  amostras, representada na Equação (32):

$$E_i = \sum_{j=1}^N s_{ij}^2. \quad (32)$$

Após o cálculo de energia, a extração de características é feita com base na energia calculada para um ciclo de evento normal e o ciclo durante o evento em análise. Ou seja, é obtida a razão das energias calculadas entre o primeiro ciclo após o instante de ocorrência do evento e um ciclo normal previamente estabelecido. A escolha de um ciclo permite, por exemplo, que a caracterização de um curto-circuito possa ser feita de acordo com seu módulo e fase durante a falta, como normalmente se realiza na literatura (DELMONT, 2007). A razão entre as energias calculadas para cada sub-banda  $i$  da TWD ou da TWPD é dada pela expressão (33):

$$x_i = \frac{E_i^{DOE}}{E_i^N}, \quad (33)$$

sendo que  $E_i^{DOE}$  representa a característica calculada para o primeiro ciclo durante a ocorrência do evento (DOE) e  $E_i^N$  a característica calculada para o ciclo normal pré-estabelecido, ambas para cada nível  $i$ , do sinal detalhe  $s$ .

Antes de serem aplicadas aos classificadores, todas as características são normalizadas em um intervalo  $[-1, 1]$ . Para os padrões caracterizados como novidades, utiliza-se como referência os valores máximos e mínimos obtidos para as características calculadas para as classes

conhecidas *a priori*.

O vetor de características  $\underline{x}$  para um dado evento, possui um número de características calculadas em função do processo de decomposição utilizado. No caso da TWD, o número total de características é dado pela expressão (34). Já para TWPD, a Equação (35) representa o comprimento total do vetor de entrada:

$$N_{caract-twd} = (N_{niveis} + 1) \times N_{fases}, \quad (34)$$

$$N_{caract-twpd} = 2^{N_{niveis}} \times N_{fases}, \quad (35)$$

sendo  $N_{niveis}$  o número de níveis da decomposição (TWD ou TWPD) e  $N_{fases}$  o número de fases do sistema. Por exemplo, se for calculada uma TWD em dez níveis de um sistema trifásico, obtém-se um total de 33 características. Para uma TWPD com 4 níveis de decomposição e o mesmo sistema trifásico, obtém-se um total de 48 características.

#### 4.6 MÓDULO E FASE DO SINAL

Conforme demonstrado no Capítulo 3, observa-se que uma das variáveis que define um curto-circuito é o fasor (módulo e fase) associado à frequência fundamental (60 Hz) do sistema trifásico de tensão no momento da falta. Desse modo, optou-se por incluir a estimativa fasorial, principalmente para os dados reais utilizados neste trabalho, de modo a obter uma melhor caracterização dos eventos de curto-circuito e manobra.

Existem diversas formas de realizar essa estimação, sendo que grande parte delas fundamentam-se nas técnicas realizadas por unidades de medição fasorial (AYDIN, 2000). Essas unidades de medição fasorial são utilizadas nas mais diversas funcionalidades, incluindo monitoramento, algoritmos avançados de proteção digital, controle e até mesmo modelagem para sistemas de transmissão e geração (CARVALHO et al., 2009). Para tanto, diferentes técnicas vêm sendo aplicadas, tais como Transformada Discreta de Fourier, algoritmos com correções dinâmicas, algoritmos baseados em mínimos quadrados e Filtros de Kalman (DUQUE et al., 2007).

Neste trabalho optou-se por utilizar o filtro de Kalman, uma vez que esse já havia sido utilizado para detecção de eventos e a mesma formulação pode ser adotada para estimação fasorial. O resultado da aplicação desse algoritmo é o cálculo do módulo e da fase (componente fundamental 60 Hz) do primeiro ciclo após a ocorrência do evento, fornecendo uma melhor caracterização principalmente para eventos de curto-circuito, tal qual está abordado na Seção 4.2. Vale ressaltar que os cálculos de módulo e fase também são referenciados a um

sinal normal, de modo a obter a diferença de fase e de módulo no sinal durante a ocorrência do evento. No caso particular da estimação de fase, considera-se apenas a mudança de fase logo após a ocorrência do evento.

#### 4.7 CLASSIFICADORES SUPERVISIONADOS MULTICLASSE

O classificador multiclasse escolhido neste trabalho foi o modelo do tipo Máquina de Vetor Suporte apresentado em Lazzaretti et al. (2009). Nesse trabalho ficou evidenciada a vantagem desse tipo de classificador, principalmente pelo seu desempenho e autonomia no processo de treinamento. Autonomia pode ser entendida aqui como a capacidade do modelo estimar os parâmetros e hiperparâmetros (que surgem ao longo do desenvolvimento teórico do modelo) durante o estágio de treinamento. Isso confere ao método uma menor dependência na seleção dos parâmetros nesse estágio, permitindo um processo autônomo em relação a essa seleção. A seguir, são apresentados os detalhes teóricos desse método.

##### 4.7.1 MÁQUINA DE VETOR SUPORTE

Um modelo do tipo Máquina de Vetor Suporte, inicialmente desenvolvido para solução de problemas de classificação (VAPNIK, 1998), consiste basicamente em um classificador cuja característica fundamental está associada ao conceito de hiperplano ótimo, baseado na maximização da margem de separação  $\rho$  entre as classes do problema. 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>3</sup>, visando à solução de problemas cuja quantidade de dados disponíveis é reduzida e pouco, ou até mesmo nenhum conhecimento prévio pode ser utilizado. O conceito de hiperplano ótimo de separação pode ser mais bem compreendido, através da análise da Figura 36, na qual estão representados padrões linearmente separáveis, divididos pelo hiperplano ótimo de separação.

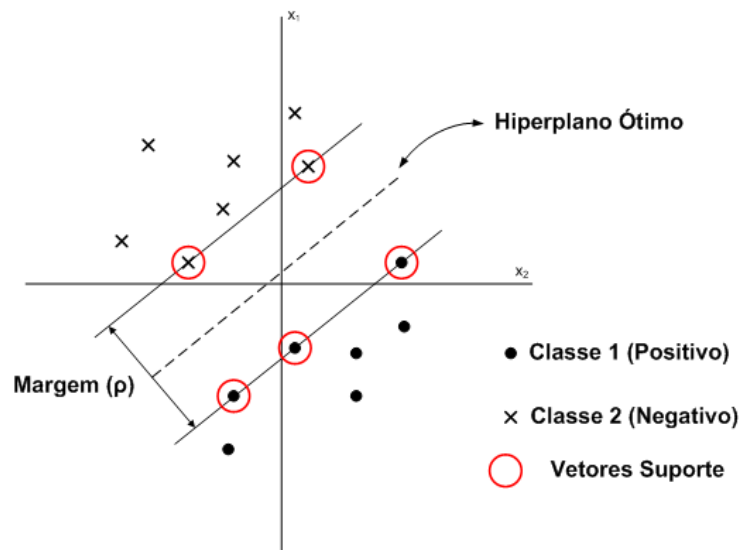
Na Figura 36 está representado um conceito fundamental na determinação do hiperplano ótimo: vetores suporte. Os vetores suporte são os pontos situados mais próximos da superfície de decisão, sendo que sua localização dá suporte à construção da margem de separação  $\rho$  e, conseqüentemente, à definição do hiperplano ótimo de separação. Esse conceito será retomado mais adiante, na formulação matemática de uma MVS.

No caso de padrões linearmente separáveis, a construção do hiperplano ótimo de

---

<sup>3</sup>O conceito de MVS foi desenvolvido para problemas de duas classes, mais comumente mencionados como separação entre exemplos positivos e negativos.





**Figura 36: Hiperplano ótimo de separação.**

separação está sujeita à restrição imposta pela Equação (36), cujo objetivo é determinar o valor ótimo para os parâmetros  $\underline{w}$  e *bias*  $b$ , dado o conjunto de treinamento com  $N$  exemplos composto pelos pares entrada  $\underline{x}_i$  e saída  $y_i$ , tal que  $\underline{x}_i \in \mathfrak{R}^n$  com e  $y_i \in \{-1, 1\}$ , considerando que o vetor de parâmetros deve minimizar a Equação (37), relacionada diretamente com a maximização da margem de separação  $\rho$ :

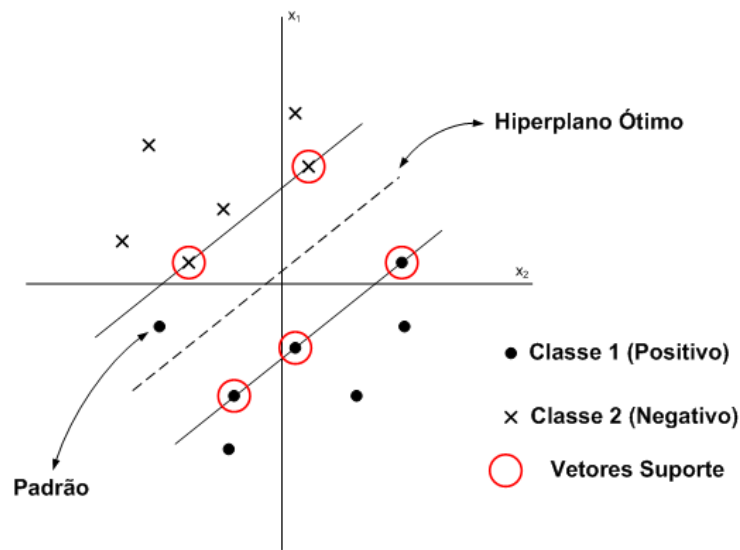
$$y_i(\underline{w}\underline{x}_i + b) \geq 1, i = 1, 2, \dots, N, \quad (36)$$

$$\Phi(\underline{w}) = \frac{1}{2}\underline{w}^T \underline{w}. \quad (37)$$

#### 4.7.2 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 (do ponto de vista de separação linear de classes). A Figura 37 mostra um exemplo em que o padrão se encontra no lado errado da superfície de decisão e no interior da região de separação, configurando uma classificação incorreta.

É possível utilizar a ideia da separação máxima entre classes no caso não-linearmente separável apresentado na Figura 37. Para tanto, deve-se considerar primeiramente o conceito de variáveis escalares não negativas  $\xi_i$ ,  $i=1, \dots, N$ , que formam a base para essa construção, permitindo erros de classificação no treinamento do modelo. Essas variáveis medem o desvio de um determinado padrão da condição ideal de separabilidade de padrões. Desse modo, o problema passa a ser encontrar um hiperplano ótimo para o qual o erro de classificação é minimizado.



**Figura 37: Erro de classificação.**

Como mostra Haykin (2001), um modo de se obter esse hiperplano é através do problema de otimização restrita exposto na Equação (38):

$$\Phi(\underline{w}, \xi) = \frac{1}{2} \underline{w}^T \underline{w} + C \sum_{i=1}^N \xi_i, \quad (38)$$

sujeito à restrição imposta pela Equação (39):

$$\begin{cases} y_i [\underline{w}^T \underline{\varphi}(x_i) + b] \geq 1 - \xi_i \\ \xi_i \geq 0 \end{cases}, i = 1, 2, \dots, N. \quad (39)$$

Para  $0 \leq \xi_i \leq 1$ , o padrão se encontra dentro da região de separação, porém do lado correto do hiperplano ótimo. Já para  $\xi_i > 1$ , o padrão está do lado incorreto do hiperplano (Figura 37). Na Equação (38), 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 desse modo, denominado parâmetro de regularização (HAYKIN, 2001).

O problema de otimização formulado na Equação (38) pode ser resolvido utilizando o método dos multiplicadores de Lagrange, resultando na formulação apresentada na Equação

(40) e sujeita às restrições impostas pela Equação (41):

$$\max_{\underline{\alpha}} \Psi(\underline{\alpha}) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N y_i y_j (x_i \cdot x_j) \alpha_i \alpha_j, \quad (40)$$

$$\begin{aligned} 0 &\leq \alpha_i \leq C, i = 1, 2, \dots, N, \\ \sum_{i=1}^N \alpha_i y_i &= 0, \end{aligned} \quad (41)$$

sendo que,  $\underline{\alpha}$  representa o conjunto de multiplicadores de Lagrange,  $(x_i \cdot x_j)$  o produto interno entre os vetores  $x_i$  e  $x_j$  e  $y_i$  a saída desejada do modelo para cada padrão de entrada.

Até o momento, mesmo para padrões não-linearmente separáveis, o classificador utilizado é linear. Para contornar esse problema, construindo um classificador não-linear, utiliza-se um mapeamento do espaço original de representação para um espaço de dimensão elevada (espaço de características), onde a probabilidade do problema ser linearmente separável é elevada. Desse modo, as MVS 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).

Esse mapeamento baseia-se no teorema de Cover (VAPNIK, 1998). 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.

Matematicamente, a saída de uma MVS pode ser dada pela Equação (42), em função do *bias*  $b$  e do conjunto de parâmetros  $\underline{w}$  do modelo representado pela Equação (43), além do mapeamento não-linear  $\underline{\varphi}(\underline{x}) : \mathfrak{R}^n \rightarrow \mathfrak{R}^N$  das entradas no espaço de características mostrado na Equação (44):

$$f(\underline{x}, \underline{w}, b) = \text{sinal} \left[ \underline{w}^T \underline{\varphi}(\underline{x}) + b \right], \quad (42)$$

$$\underline{w} = \left[ w_1 \quad w_2 \quad \dots \quad w_N \right]^T, \quad (43)$$

$$\underline{\varphi}(\underline{x}) = \left[ \varphi_1(\underline{x}) \quad \varphi_2(\underline{x}) \quad \dots \quad \varphi_N(\underline{x}) \right]^T. \quad (44)$$

Na Equação (42),  $\text{sinal}[a]$  representa a função sinal (Equação 45):

$$\begin{cases} \text{sinal}[a] = 1, \forall a \geq 0, \\ \text{sinal}[a] = -1, \forall a < 0. \end{cases} \quad (45)$$

O mapeamento apresentado não aborda a questão relacionada com a construção do hiperplano ótimo, fazendo com que seja necessário um segundo estágio de treinamento relacionado à construção do hiperplano e garantindo assim, a capacidade de generalização do modelo. 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. O núcleo do produto interno pode ser definido, como mostra a Equação (46), com base na transformação não-linear representada por  $\varphi(\cdot)$ :

$$K(\underline{x}_i, \underline{x}_j) = \varphi(\underline{x}_i) \cdot \varphi(\underline{x}_j). \quad (46)$$

Através da utilização da Equação (46), é possível 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. Com isso, a solução do modelo MVS pode ser dado pelo seguinte problema de otimização:

$$\max_{\underline{\alpha}} \Psi(\underline{\alpha}) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N y_i y_j K(\underline{x}_i, \underline{x}_j) \alpha_i \alpha_j, \quad (47)$$

s.a.:

$$\begin{aligned} 0 &\leq \alpha_i \leq C, i = 1, 2, \dots, N, \\ \sum_{i=1}^N \alpha_i y_i &= 0. \end{aligned} \quad (48)$$

Existem diversos tipos de *kernel*  $K(\underline{x}_i, \underline{x}_j)$  que podem ser utilizados na construção de um modelo MVS. 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 MVS. Neste trabalho será utilizado o *kernel* gaussiano modificado (que atende às restrições do teorema de Mercer) como sugere Ferreira (2008). O *kernel* utilizado está representado pela Equação (49):

$$K(\underline{x}, \underline{z}) = e^{-\sum_{l=1}^n \sigma_l^2 (x_l - z_l)^2}, \quad (49)$$

sendo que  $\sigma_l$  para  $l = 1, 2, \dots, n$  representam os hiperparâmetros do kernel. Usualmente utiliza-se  $\sigma = \sigma_1 = \dots = \sigma_n$  e procura-se estimar, por exemplo, através de validação cruzada, o valor de  $\sigma$ . No *kernel* aqui apresentado, procura-se estimar um valor de  $\sigma_l$  para cada entrada  $l$ . A vantagem dessa abordagem é que ela permite adaptar a escala (ponderação) de cada entrada do modelo, ajustando a métrica de acordo com o saída desejada. Além disso, a análise individual

dos valores de  $\sigma_l$  pode auxiliar no processo de seleção de entradas do modelo (FERREIRA, 2008). Trabalhos recentes (XU et al., 2011) demonstram que essa abordagem fornece resultados globais superiores à abordagem com o hiperparâmetro fixo para todas as entradas ( $\sigma = \sigma_1 = \dots = \sigma_n$ ).

No ponto ótimo da Equação (47) nem todos  $\alpha_i^*$  são não-nulos. Os vetores para os quais  $\alpha_i^*$  é diferente de zero são chamados vetores suporte, os quais definem a superfície de decisão da MVS pela expressão (50), formando assim a definição geral para vetores suporte:

$$f(\underline{x}, \underline{w}, b) = \text{sinal} \left[ \sum_{i=1}^{NS} \alpha_i y_i K(\underline{x}_i, \underline{x}) + b \right], \quad (50)$$

sendo que NS representa o número de vetores suporte.

O modelo MVS foi desenvolvido com base em um paradigma da área de aprendizado de máquina, conhecido como aprendizado estatístico (FERREIRA, 2008). Dentro desse conceito, pode-se dizer que um modelo MVS 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 a uma medida de complexidade conhecida como dimensão de Vapnik e Chervonenkis, popularmente denominada dimensão VC.

O conceito da dimensão VC possui analogia com o dilema *bias*-variância, cujo conceito está associado à complexidade da estrutura do modelo e, indiretamente, ao desempenho do modelo para novos dados (conjunto de teste). Da mesma forma que naquele dilema, é possível 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 MVS, durante o processo de treinamento o termo associado ao 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  $F(\underline{x}, \underline{w})$  com margem de separação  $\rho$  é limitada superiormente pela Equação (51) (HAYKIN, 2001):

$$VC[F(\underline{x}, \underline{w})] \leq \frac{R^2}{\rho^2}, \quad (51)$$

sendo que  $R$  corresponde ao raio da menor hiperesfera que engloba a imagem dos padrões de treinamento no espaço de características<sup>4</sup>.

<sup>4</sup>Espaço associado ao mapeamento realizado durante o treinamento de MVS.

### 4.7.3 SELEÇÃO AUTOMÁTICA DOS HIPERPARÂMETROS

O último estágio do processo está relacionado à determinação automática dos hiperparâmetros: constante de regularização  $C$  e hiperparâmetros do kernel  $\sigma_l^2$ . Para tanto, optou-se por utilizar neste 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 (do inglês, *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*). Esse limite superior  $T[f(\underline{x}, \underline{w}, b)]$  pode ser representado pela expressão (52):

$$T[f(\underline{x}, \underline{w}, b)] = \sum_{i=1}^{NS} \alpha_i S_i^2, \quad (52)$$

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 mínima distância euclidiana entre  $\underline{\phi}(\underline{x}_i)$  e qualquer uma das combinações lineares (restritas) possíveis de pontos no espaço de características associados com os demais vetores suporte:

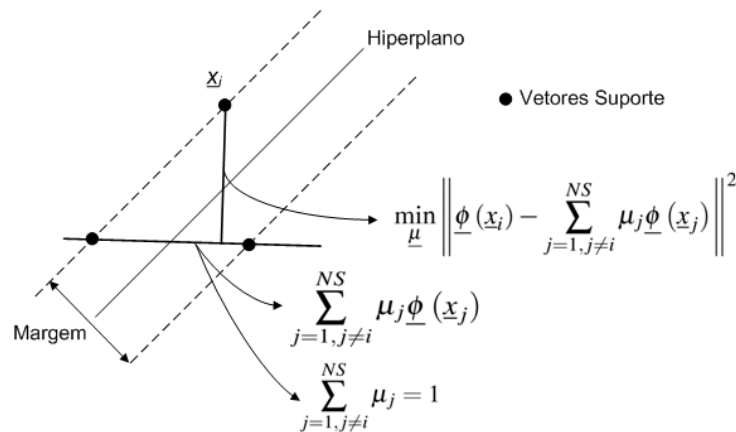
$$S_i^2 = \min_{\underline{\mu}} \left\| \underline{\phi}(\underline{x}_i) - \sum_{j=1, j \neq i}^{NS} \mu_j \underline{\phi}(\underline{x}_j) \right\|^2, \quad (53)$$

s.a.:

$$\underline{x}_j \in \Omega, \quad \sum_{j=1, j \neq i}^{NS} \mu_j = 1, \quad (54)$$

sendo  $\Omega$  o conjunto dos  $NS$  vetores suporte. A interpretação geométrica do limite acima apresentado está diretamente relacionada com a relação *raio/margem* e, conseqüentemente, à dimensão VC de hiperplanos com margem de separação  $\rho$  – Equação (51). Esse critério demonstra que maximizar somente a margem de separação  $\rho$  não garante boa capacidade de generalização, sendo necessária a obtenção de mapeamentos concentrados no espaço de características, ou seja, minimizar o raio da menor hiperesfera que contém a imagem de todos os padrões no espaço de características.

A minimização imposta para a determinação de  $S_i$  através dos índices  $\underline{\mu}$  tem como consequência a concentração dos vetores suporte no espaço de características, como demonstra a Figura 38. Sendo assim, a escolha dos hiperparâmetros ( $\sigma_l$  e  $C$ ) através da minimização desse limite considera a maximização da margem de separação com a maior concentração dos padrões no espaço de características (minimização do raio da hiperesfera). Em Chapelle et al. (2002), demonstra-se que o limite em questão fornece o menor limite superior para o erro de generalização (por validação cruzada única) em comparação a outros limites, tais como: número de vetores suporte, limite de Jaakkola-Haussler e limite de Opper-Winther. Sendo assim, pode-se afirmar que o limite em questão é o mais conservativo dos limites existentes para estimação do erro de generalização, sendo essa a justificativa para sua escolha em detrimento aos demais existentes.



**Figura 38: Extensão dos vetores suporte.** Nessa figura demonstra-se que a minimização da distância do vetor suporte  $x_i$  em relação à combinação linear restrita entre os demais vetores suporte leva à concentração desses padrões no espaço de características, minimizando o raio da menor hiperesfera que contém a imagem de todos os padrões nesse espaço.

Como o mapeamento  $\underline{\phi}(x)$  não é explicitamente definido, é necessário definir  $S_i$  diretamente no espaço de características. Para tanto, Chapelle et al. (2002) propõem o uso da seguinte relação:

$$S_i^2 = \frac{1}{\left( \underline{\tilde{K}}^{-1} \right)_{ii}}, \quad (55)$$

onde  $(\underline{\tilde{K}}^{-1})_{ii}$  representa o  $i$ -ésimo elemento da diagonal da inversa da matriz  $\underline{\tilde{K}}$  dada pela Equação (56):

$$\underline{\tilde{K}} = \begin{bmatrix} \underline{K}_{NS} & \underline{u} \\ \underline{u}^T & 0 \end{bmatrix}, \quad (56)$$

sendo  $\underline{K}_{NS}$ , a matriz com o núcleo do produto interno entre todos os  $NS$  vetores suporte e  $\underline{u} \in \mathfrak{R}^{NS}$  um vetor unitário.

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 demonstram Chapelle et al. (2002). Matematicamente, o *kernel* passa a ser representado por:

$$\widehat{K}(\underline{x}_i, \underline{x}_j) = K(\underline{x}_i, \underline{x}_j) + \frac{\delta(\underline{x}_i, \underline{x}_j)}{C}, \quad (57)$$

sendo que  $\delta(\cdot)$  representa a função delta de Kronecker:

$$\delta(\underline{x}_i, \underline{x}_j) = \begin{cases} 1, & \underline{x}_i = \underline{x}_j \\ 0, & \underline{x}_i \neq \underline{x}_j. \end{cases} \quad (58)$$

Segundo Chapelle et al. (2002), a penalização quadrática de erros no treinamento do modelo MVS equivale à estimação desses modelos considerando os dados linearmente separáveis, porém utilizando o *kernel*  $\widehat{K}(\cdot)$ .

O limite  $T[f(\underline{x}, \underline{w}, b)]$  possui uma característica multimodal, conforme sugere Ferreira (2008). Em função disso, optou-se, neste trabalho, por utilizar algoritmos genéticos (GOLDBERG, 1989) para minimização desse funcional em função de  $C$  e  $\sigma_l^2$ . Maiores detalhes desse processo de otimização serão descritos na Subseção 4.8.5, a qual realiza um procedimento similar para estimação dos parâmetros.

#### 4.8 CLASSIFICADORES UMA-CLASSE

Um classificador uma-classe no contexto do presente trabalho, também conhecido como detector de novidades, consiste em parametrizar um classificador com base em um conjunto de dados previamente conhecidos, organizados de acordo com um agrupamento definido, de modo a permitir a identificação de novos padrões não pertencentes a esse conjunto definido previamente (HODGE; AUSTIN, 2004). Esses padrões não conhecidos previamente são definidos como novidades.

Padrões não conhecidos previamente têm sua origem definida em diferentes cenários,



dentre os quais se destacam: erros no sistema de instrumentação, mudanças de comportamento do sistema sob supervisão, erros humanos, surgimento de novas características e fraudes (HODGE; AUSTIN, 2004). Nesse sentido, podem ser destacados diversos trabalhos adotando essa abordagem, tais como (TAX; DUIN, 2001):

- Análise de imagens de satélite;
- Análise de padrões de risco em *e-mails*;
- Avaliação de desempenho de redes de computadores;
- Detecção de acessos não autorizados em redes de computadores;
- Detecção de fraudes;
- Detecção de padrões não rotulados em classificação multiclases;
- Processamento de imagens.

Existem basicamente três abordagens para construção de classificadores uma-classe (MARKOU, 2003a, 2003b). A primeira delas está relacionada com modelos baseados na estimação da função densidade de probabilidades associada aos padrões de entrada. A partir dessa estimativa é possível estabelecer se um novo padrão é novidade ou não, com base na análise de limiares aplicados à função densidade de probabilidade. Nessa abordagem destacam-se os modelos baseados em Misturas de Gaussianas e estimadores utilizando Janelas de Parzen (TAX; DUIN, 2001). Os modelos baseados em Mistura de Gaussianas são métodos paramétricos, enquanto Estimadores de Parzen são não-paramétricos.

O segundo grupo de modelos compreende métodos construídos a partir de limites ou fronteiras impostas sobre o conjunto de dados, não assumindo que sua distribuição seja conhecida. Desse modo, busca-se otimizar a fronteira fechada em torno dos dados. Nesse cenário destacam-se os métodos de Representação de Dados Utilizando Vetores Suporte e *K*-Vizinhos-Mais-Próximos. Já o terceiro grupo de modelos está relacionado aos métodos de agrupamento aplicados aos dados. A partir dessa representação, é possível definir se um dado padrão é novidade ou não com base na distância em que esse padrão se encontra dos agrupamentos previamente estabelecidos. Dentro desse grupo destaca-se o agrupamento *K*-Médias, como sendo o método mais utilizado no contexto de detecção de novidades (SPINOSA, 2008).

A seguir serão descritos os aspectos teóricos de cada um dos métodos utilizados neste trabalho, visando sua utilização no contexto de classificação de padrões ou segmentação de sinais, de acordo com sua aplicação ao longo do trabalho.

#### 4.8.1 $K$ -VIZINHOS-MAIS-PRÓXIMOS

Antes de detalhar o modelo  $K$ -Vizinhos-Mais-Próximos, é importante destacar aspectos teóricos associados ao estimador de Parzen. Um estimador com base em Janelas de Parzen consiste em determinar a função densidade de probabilidade através do cálculo do número de padrões de entrada localizados no interior de uma dada região pré-definida com volume constante (SPINOSA, 2008). Essa região pode ser, por exemplo, um hipercubo de aresta  $h$  e volume  $h^l$  para as  $l$  dimensões do problema, sendo que a densidade estimada é função do número de padrões contidos em cada hipercubo e o seu volume, para todo o espaço de entrada.

Normalmente, ao invés de se utilizar uma região baseada em hipercubo, utiliza-se uma função mais suave, evitando a existência de descontinuidades presentes na estimação feita por hipercubos. Uma função normalmente utilizada para essa finalidade é a função Gaussiana. Com isso, a estimação da função densidade de probabilidade  $\hat{p}(\underline{x})$  através de Janelas de Parzen com funções Gaussianas, pode ser dada pela Equação (59):

$$\hat{p}(\underline{x}) = \frac{1}{N} \sum_{i=1}^N \frac{1}{(2\pi)^{\frac{l}{2}} h^l} \exp\left(-\frac{(\underline{x} - x_i)^T (\underline{x} - x_i)}{2h^2}\right), \quad (59)$$

sendo  $h$  o parâmetro que controla a suavidade das funções Gaussianas e  $N$  o número total de padrões de entrada. Em resumo, pode-se afirmar que a função densidade de probabilidade desconhecida é estimada através de  $N$  funções Gaussianas centradas em cada um dos padrões de entrada disponíveis (THEODORIDIS; KONSTANTINOS, 2009). Nesse ponto, pode-se observar a semelhança entre o estimador de Parzen e o método Mistura de Gaussianas com relação à estimação da função densidade de probabilidade através da soma ponderada de Gaussianas. No entanto, vale ressaltar que, no caso da Mistura de Gaussianas, o número de Gaussianas utilizadas não coincide com o número de padrões de entrada, sendo um valor definido *a priori*.

Em um estimador de Parzen, a função densidade de probabilidade é calculada mantendo o volume em torno dos dados  $\underline{x}$  constante e o número de padrões no interior do volume, variando aleatoriamente para todo o espaço de entrada. Já para o método dos  $K$ -Vizinhos-Mais-Próximos (KVMP), o número de padrões  $k$  é mantido fixo e o volume  $V(\underline{x})$  é ajustado para que sejam incluídos os  $k$  padrões. Desse modo, em regiões com baixa densidade de padrões o volume será elevado. Já para regiões densamente ocupadas por padrões, o volume será reduzido (THEODORIDIS; KONSTANTINOS, 2009). O estimador pode ser dado conforme mostra a Equação (60):

$$\hat{p}(\underline{x}) = \frac{k}{NV(\underline{x})}, \quad (60)$$

sendo  $N$  o número total de padrões e o volume  $V(\underline{x})$  correspondente ao volume da hiperesfera

contendo os  $k$  padrões da representação.

Para detecção de novidades, efetua-se o cálculo do  $k$ -ésimo vizinho mais próximo de cada entrada  $x_i$  e, em seguida, utiliza-se a média dos  $N_R$  vizinhos mais distantes para calcular o limiar e definir, posteriormente, se um novo padrão é novidade ou se faz parte do conjunto de dados normais (TAX, 2001). O parâmetro  $N_R$ , assim como  $k$ , deve ser definido *a priori*.

#### 4.8.2 K-MÉDIAS

O objetivo de apresentar o método  $K$ -Médias é que ele é parte fundamental do algoritmo de agrupamento utilizado tanto para detecção de novidades, quanto para identificação de novas classes: o modelo  $X$ -Médias. O  $K$ -Médias consiste em um agrupamento realizado sobre os padrões de entrada  $x$  em  $K$  grupos, sendo  $K$  um parâmetro definido *a priori*. O algoritmo é executado em duas etapas principais, as quais guardam relação com o método de Maximização de Expectativa (BISHOP, 1995). Na primeira etapa, cada padrão de entrada é atribuído ao agrupamento mais próximo que esse se encontra, sendo a medida de proximidade representada pela distância euclidiana entre o padrão  $x_n$  e o centro do agrupamento  $\underline{\mu}_k$ . Desse modo, pode-se definir a variável responsável pelo rótulo de cada padrão  $r_{nk} \in [0, 1]$ , onde  $k = 1, \dots, K$ , como sendo  $r_{nk} = 1$  se  $x_n$  pertence ao agrupamento  $k$  e  $r_{nk} = 0$  caso contrário. Na segunda etapa, é realizado o cálculo dos novos centros definidos pelo primeiro estágio do algoritmo, através da seguinte equação:

$$\underline{\mu}_k = \frac{\sum_{n=1}^N r_{nk} x_n}{\sum_{n=1}^N r_{nk}}. \quad (61)$$

O processo se repete até que nenhuma nova alteração seja verificada nos agrupamentos, ou se um determinado número de iterações tenha ocorrido. No fim do algoritmo, cada padrão de entrada está associado a um dos agrupamentos definidos.

#### 4.8.3 X-MÉDIAS

O modelo de agrupamento  $X$ -Médias foi construído com o intuito de incorporar ao algoritmo  $K$ -Médias, três novas abordagens que visam, principalmente, acelerar a execução do  $K$ -Médias dentro de um contexto de determinação automática e dinâmica do número de grupos (PELLEG; MOORE, 2000). As contribuições do  $X$ -Médias podem ser resumidas em:

- O  $K$ -Médias pode ser considerado um algoritmo relativamente lento, no que diz respeito ao tempo que leva para finalizar cada iteração. Nesse contexto, o modelo  $X$ -Médias

propõe a utilização de um procedimento para aceleração de cada iteração, conforme exposto em Pelleg e Moore (1999);

- No caso do  $K$ -Médias, é necessário informar previamente o número de grupos ( $K$ ) a serem escolhidos pelo algoritmo. Já no modelo  $X$ -Médias, os centros são determinados de forma automática, considerando o critério de seleção de modelo denominado *Bayesian Information Criterion* (BIC);
- Quando  $K$  é mantido fixo ao longo da execução do algoritmo, pode-se demonstrar (de forma empírica) que o resultado tende a um valor ótimo local inferior ao valor obtido quando  $K$  é alterado dinamicamente (PELLEG; MOORE, 2000). Nesse sentido, o modelo  $X$ -Médias apresenta um procedimento para alterar o número de grupos dinamicamente, também utilizando o critério BIC de seleção de modelo.

De um modo geral, o  $X$ -Médias inicia através da especificação do intervalo dado por um número mínimo ( $K_{min}$ ) e um número máximo ( $K_{max}$ ) de grupos, dentro do que o usuário acredita ser razoável para a distribuição de dados em questão. O algoritmo inicia com o limite inferior do número de centros e, ao longo do processo, insere novos centros nas regiões do espaço que indicam a necessidade de tal inserção. Durante o processo, o índice BIC de cada conjunto de centros é estimado e, no fim do processo, o conjunto com o melhor índice é escolhido como a estimativa resultante do  $X$ -Médias.

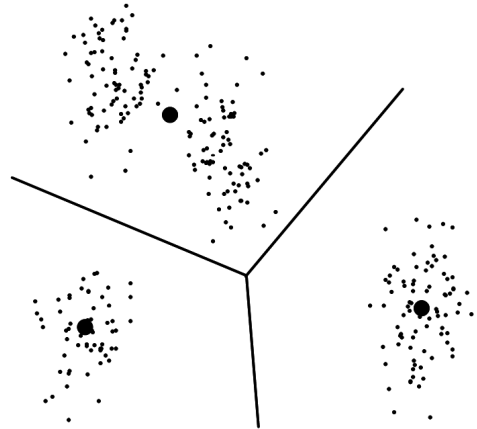
O processo alterna duas etapas principais, até que o número de centros estimado  $K$  seja superior ao limite superior definido pelo usuário (critério de parada). Essas etapas consistem em:

- Estimar os parâmetros: consiste em aplicar o algoritmo  $K$ -Médias até que sua convergência seja atingida. Nessa etapa também pode ser incluído o processo de aceleração de convergência proposto por Pelleg e Moore (1999), conforme será exposto posteriormente;
- Estimar a estrutura: essa etapa determina se novos centros devem ser inseridos e, caso seja necessária essa inserção, qual a região do espaço dos dados de entrada receberá os novos centros.

O critério de inserção de novos centros é uma das grandes contribuições do modelo  $X$ -Médias, uma vez que está diretamente associado à característica automática desse agrupamento.

Para compreender esse critério, optou-se por apresentar um exemplo da sua forma de operação, destacando cada etapa desse processo.

Inicialmente, pode-se considerar a solução do  $K$ -Médias com três centros apresentada na Figura 39.



**Figura 39: Exemplo de aplicação do modelo  $K$ -Médias com três centros.**

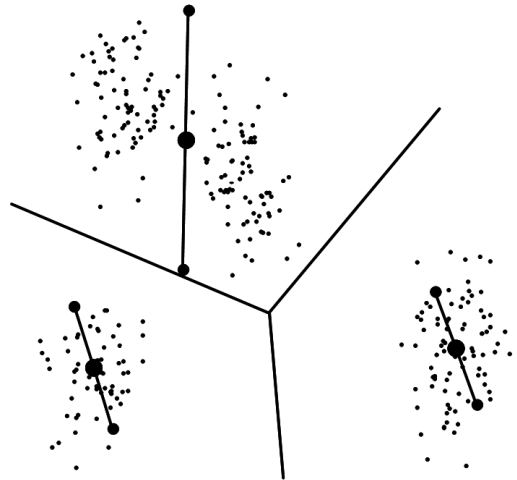
**Fonte: (PELLEG; MOORE, 2000)**

O processo de estimação da estrutura inicia dividindo cada centro em dois novos centros, denominados de centroides-filhos. Esses novos centroides são inseridos de forma que sua distância mútua seja proporcional ao tamanho da região contemplada por esses centros, em uma direção oposta ao longo de um vetor aleatoriamente determinado na região, conforme demonstra a Figura 40.

Em seguida, o  $K$ -Médias é executado novamente, para cada região assinalada na Figura 40 utilizando os dois novos centros inseridos para cada região. A Figura 41 mostra o resultado desse processo.

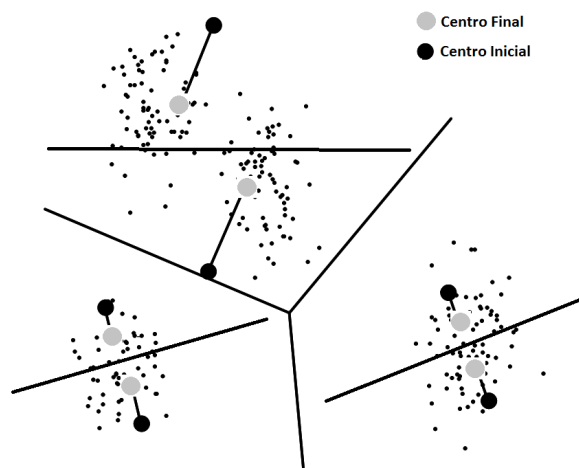
Nesse ponto, aplica-se o teste de seleção do modelo para cada par de centroides-filhos determinados. O objetivo desse teste é indicar se o aumento de número de centros promove uma melhoria na representação de uma dada região. Caso não seja observado esse incremento, do ponto de vista do critério BIC, o número de centros não é alterado para a região sob análise, conforme exemplifica a Figura 42.

Esse processo de estimação dos parâmetros e estimação da estrutura se repete até que o limite superior do número de centros fornecido pelo usuário seja atingido.



**Figura 40: Primeira etapa do método X-Médias.**

**Fonte: (PELLEG; MOORE, 2000)**

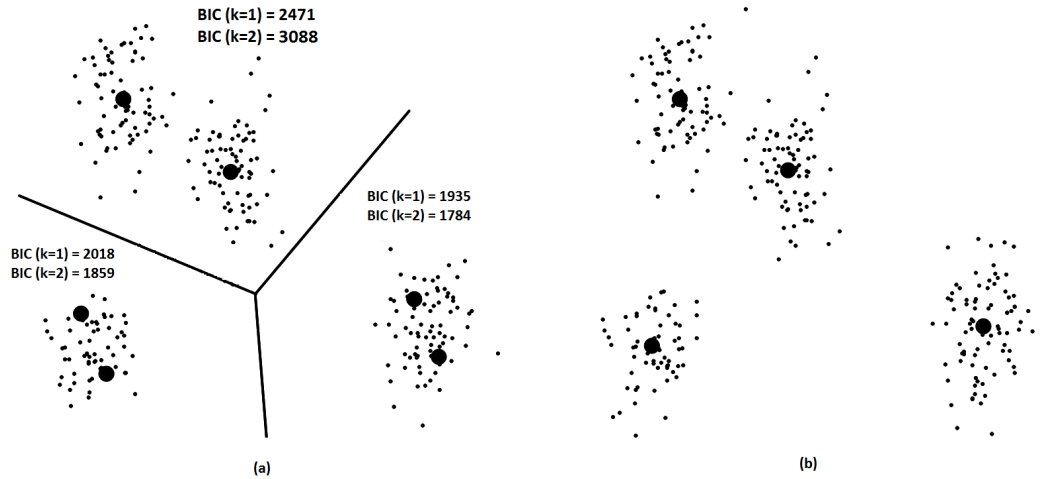


**Figura 41: Segunda etapa do método X-Médias**

**Fonte: (PELLEG; MOORE, 2000)**

Com relação ao critério de seleção de modelo, BIC, alguns apontamentos adicionais são necessários. Considera-se, inicialmente, que os dados de entrada são representados pelo conjunto  $\underline{D}$  e existe uma família de modelos  $M_j$ , que para o caso do X-Médias correspondem a modelos com diferentes grupos  $K$ .

O critério de seleção desse índice está fundamentado na probabilidade *a posteriori*  $p(M_j|\underline{D})$ . No caso do agrupamento  $K$ -Médias e, conseqüentemente, para o X-Médias, os modelos podem ser considerados como distribuições normais (Gaussianas esféricas). Desse modo,



**Figura 42: Resumo do processo de seleção dos centros durante a execução do X-Médias. Em (a) estão os resultados do índice BIC para cada caso, considerando apenas um centro e com dois centros, em cada uma das três regiões do espaço de entrada. Já em (b), está representado o resultado do processo de seleção, mantendo o resultado com melhor índice para cada caso (um centro ou dois centros).**

**Fonte: (PELLEG; MOORE, 2000)**

o índice BIC de um modelo  $M_j$  pode ser dado pela Equação (62):

$$BIC(M_j) = \hat{l}_j(\underline{D}) - \frac{m_j}{2} \log R, \quad (62)$$

sendo  $\hat{l}_j$  o logaritmo da função de verossimilhança dos dados de acordo com o  $j$ -ésimo modelo,  $m_j$  o número de parâmetros em  $M_j$  e  $R$  o número de exemplos presentes em  $\underline{D}$ .

Fixando  $1 \leq n \leq K$  e considerando apenas o conjunto de dados  $\underline{D}_n$  associados ao centroide  $n$ , a estimativa de máximo para a função de verossimilhança resulta em:

$$\hat{l}(\underline{D}_n) = -\frac{R_n}{2} \log(2\pi) - \frac{R_n M}{2} \log(\hat{\sigma}^2) - \frac{R_n - K}{2} + R_n \log R_n - R_n \log R, \quad (63)$$

sendo  $M$  a dimensionalidade dos dados de entrada e  $\hat{\sigma}^2$  a estimativa de máxima verossimilhança para a variância.

Vale ressaltar que esse índice é utilizado em dois momentos distintos ao longo da execução do X-Médias. O primeiro deles é no momento em que ocorre a definição sobre a inserção de novos centros, durante a etapa de estimativa da estrutura, conforme discutido anteriormente. Já o segundo, consiste em aplicar esse teste para cada iteração do algoritmo X-Médias, ao fim de cada etapa de estimativa da estrutura, visando definir a escolha do melhor modelo de acordo com o índice aqui apresentado.

A última questão referente à utilização do  $X$ -Médias de acordo com sua formulação original (PELLEG; MOORE, 2000) está associada com o processo de aceleração de convergência proposto por Pelleg e Moore (1999).

Na formulação original do  $K$ -Médias, uma das etapas do algoritmo consiste em determinar qual o centro associado a cada padrão do espaço de entrada. A partir dessa informação, é possível determinar, em uma segunda etapa, a localização dos novos centros, através do centro de massa dos dados correspondentes.

Analisando esse procedimento, fica claro que um subconjunto de padrões pertencentes a um dado grupo é tão informativo quanto uma análise feita de forma individual para cada padrão, desde que existam, principalmente, informações sobre a quantidade de padrões existentes nesse subconjunto. Desse modo, é possível economizar esforço computacional no cálculo da pertinência de cada padrão em relação aos centros, já que é possível calcular a pertinência equivalente de todo o subconjunto de padrões associados a um mesmo centro.

Um modelo com essas características, que foi incorporado ao  $X$ -Médias com o intuito de determinar esses subconjuntos, é a *kd-tree*. Esse modelo utiliza uma estrutura hierárquica, fornecendo um hipercubo em torno dos dados de um subconjunto de padrões de entrada pertencentes a um mesmo grupo. Em Pelleg e Moore (1999), demonstra-se que é possível acelerar em um fator de 170 a execução do algoritmo em relação ao  $K$ -Médias convencional, para dados astrofísicos de dimensão cinco, mantendo o mesmo desempenho (qualidade) do agrupamento.

Esse processo de aceleração não foi utilizado no presente trabalho, uma vez que a dimensionalidade dos dados é consideravelmente inferior à apresentada em Pelleg e Moore (1999). No entanto, naquele trabalho é possível obter um conjunto de informações detalhadas dessa implementação no modelo de agrupamento  $X$ -Médias.

#### 4.8.4 DETECÇÃO DE NOVIDADES COM O $K$ -MÉDIAS E $X$ -MÉDIAS

Para detecção de novidades, efetua-se o cálculo das distâncias entre cada padrão e os centros dos agrupamentos modelados e, em seguida, utiliza-se a média dos  $N_R$  valores mais distantes para calcular o limiar e definir, posteriormente, se um novo padrão é novidade ou faz parte do conjunto de dados normais (TAX, 2001). O parâmetro  $N_R$  deve ser definido *a priori*.

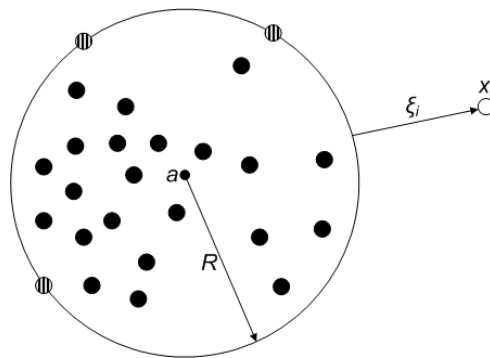
#### 4.8.5 REPRESENTAÇÃO DE DADOS COM VETORES SUPORTE

A principal característica desse modelo é a representação dos dados em um novo espaço, com elevada dimensionalidade (espaço de características), sem a necessidade de grande



esforço computacional adicional, permitindo que nesse novo espaço possam ser definidos descritores mais flexíveis aos dados de entrada (TAX, 2001), análogo ao modelo máquinas de vetor suporte.

Para realizar essa representação, existem alguns aspectos que devem ser considerados. O primeiro deles baseia-se no modelo de representação dos dados nesse novo espaço. Assume-se que para um dado conjunto de dados  $\underline{x}$ , existe uma superfície fechada ao seu redor (hiperesfera), caracterizada por um centro  $\underline{a}$  e raio  $R$ , como demonstra a Figura 43.



**Figura 43: Modelo de representação dos dados utilizando uma hipersfera. Nessa representação, três dos dados de entrada estão situados na margem da hipersfera (vetores suporte) e um objeto  $\underline{x}_i$  está situado fora da margem e possui  $\xi_i > 0$ .**

Fonte: (TAX, 2001)

No caso particular em que todos os padrões de entrada (treinamento) estão localizados no interior da hipersfera, tem-se que o risco empírico é nulo. Desse modo, pode-se definir que o risco estrutural é (VAPNIK, 1998):

$$\varepsilon(R, \underline{a}) = R^2, \quad (64)$$

que deve ser minimizado sujeito às seguintes restrições:

$$\|\underline{x}_i - \underline{a}\|^2 \leq R^2, \quad \forall i. \quad (65)$$

Quando se considera a possibilidade de existirem novidades nos padrões de treinamento, faz-se necessária a inclusão de um termo que permita tal ocorrência, penalizando os padrões mais distantes da margem da hipersfera e controlando o compromisso entre os riscos estrutural e empírico, uma vez que essa inclusão faz com que o erro de treinamento seja diferente de zero. Desse modo, o problema de minimização passa a ser:

$$\varepsilon(R, \underline{a}, \underline{\xi}) = R^2 + C \sum_i \xi_i, \quad (66)$$

agrupando o máximo de padrões no interior da hipersfera:

$$\|\underline{x}_i - \underline{a}\|^2 \leq R^2 + \xi_i, \xi_i \geq 0, \forall i, \quad (67)$$

sendo  $C$  o parâmetro responsável pelo controle estrutural do modelo, mantendo o compromisso entre o volume da hipersfera e o risco empírico.

Os parâmetros  $R$ ,  $\underline{a}$  e  $\underline{\xi}$  podem ser obtidos através da otimização do funcional (68), cujas restrições foram incluídas utilizando multiplicadores de Lagrange  $\alpha_i$  e  $\gamma_i$ :

$$L(R, \underline{a}, \underline{\xi}, \underline{\alpha}, \underline{\gamma}) = R^2 + C \sum_i \xi_i - \sum_i \alpha_i (R^2 + \xi_i - (\underline{x}_i \cdot \underline{x}_i - 2\underline{a} \cdot \underline{x}_i + \underline{a} \cdot \underline{a})) - \sum_i \gamma_i \xi_i. \quad (68)$$

Visando adequar o problema acima exposto para utilização de programação quadrática (SCHÖLKOPF; SMOLA, 2001), são calculadas as derivadas parciais do funcional (68) em relação à  $R$ ,  $\underline{a}$  e  $\underline{\xi}$  e igualando-as a zero:

$$\begin{aligned} \frac{\partial L}{\partial R} = 0 &: \sum_i \alpha_i = 1, \\ \frac{\partial L}{\partial \underline{a}} = 0 &: \underline{a} = \sum_i \alpha_i \cdot \underline{x}_i, \\ \frac{\partial L}{\partial \xi_i} = 0 &: \gamma_i = C - \alpha_i, \forall i. \end{aligned} \quad (69)$$

Readequando as restrições do problema, considerando  $\alpha_i = C - \gamma_i$  e  $\alpha_i \geq 0, \gamma_i \geq 0, \forall i$ , o resultado corresponde à seguinte formulação final:

$$L(R, \underline{a}, \underline{\xi}, \underline{\alpha}, \underline{\gamma}) = \sum_i \alpha_i (\underline{x}_i \cdot \underline{x}_i) - \sum_{i,j} \alpha_i \alpha_j (\underline{x}_i \cdot \underline{x}_j), \quad (70)$$

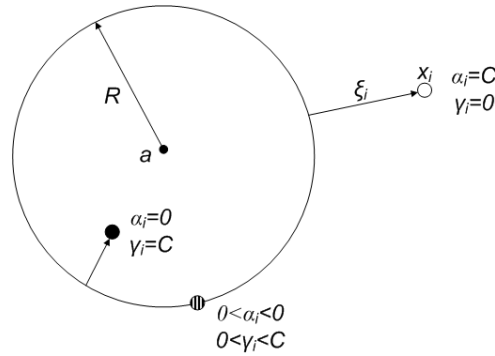
sujeita às seguintes restrições:

$$\begin{aligned} \sum_i \alpha_i &= 1 \\ 0 &\leq \alpha_i \leq C, \forall i. \end{aligned} \quad (71)$$

O problema apresentado é a representação dual do problema de minimização inicial, ou seja, é resolvido através da maximização do funcional  $L$  em relação às restrições impostas, através programação quadrática com restrições (HAO et al., 2007).

Através da análise das restrições dos multiplicadores de Lagrange, é possível se estabelecer a localização de um dado padrão, em relação às fronteiras estabelecidas pela hipersfera. Ou seja, um padrão pode estar localizado no interior, na margem ou configurar uma novidade

em relação à superfície da hiperesfera, conforme demonstra a Figura 44.



**Figura 44: Valores dos multiplicadores de Lagrange para diferentes entradas.**

Fonte: (TAX, 2001)

Os pontos situados na margem da hiperesfera, por possuírem  $\alpha_i$  diferente de zero, são caracterizados como vetores suporte, uma vez que são responsáveis por definir as fronteiras da hiperesfera, considerando que os demais padrões podem ser descartados ( $\alpha_i = 0$ ). Para verificar se um dado ponto  $\underline{z}$  encontra-se no interior na hiperesfera, pode-se calcular sua distância em relação ao centro da hiperesfera<sup>5</sup>  $\underline{a}$  e, através da sua comparação com o valor do raio da hiperesfera  $R^2$ , sabe-se a localização do ponto:

$$\|\underline{z} - \underline{a}\|^2 = (\underline{z} \cdot \underline{z}) - 2 \sum_i \alpha_i (\underline{z} \cdot \underline{x}_i) + \sum_{i,j} \alpha_i \alpha_j (\underline{x}_i \cdot \underline{x}_j) \leq R^2. \quad (72)$$

Por definição, o raio  $R^2$  corresponde à distância do centro da hiperesfera em relação a um vetor suporte  $\underline{x}_k$  situado na margem:

$$R^2 = (\underline{x}_k \cdot \underline{x}_k) - 2 \sum_i \alpha_i (\underline{x}_i \cdot \underline{x}_k) + \sum_{i,j} \alpha_i \alpha_j (\underline{x}_i \cdot \underline{x}_j). \quad (73)$$

A função  $f(\underline{z}; \underline{\alpha}, R)$  que resume a descrição dos dados apresentada até aqui, pode ser representada por:

$$f(\underline{z}; \underline{\alpha}, R) = I \left( (\underline{z} \cdot \underline{z}) - 2 \sum_i \alpha_i (\underline{z} \cdot \underline{x}_i) + \sum_{i,j} \alpha_i \alpha_j (\underline{x}_i \cdot \underline{x}_j) \leq R^2 \right), \quad (74)$$

sendo  $I(A)$  uma função que retorna 1, se  $A$  for verdadeiro e 0, caso contrário.

Para realizar o desenvolvimento da hiperesfera no espaço de características, é necessário efetuar o mapeamento ( $\Phi$ ) dos padrões de treinamento  $\underline{x}$  nesse novo espaço, de forma

<sup>5</sup>Determinado através da derivada parcial de  $L(R, \underline{a}, \underline{\xi}, \underline{\alpha}, \underline{\gamma})$  em relação à  $\underline{a}$ , resultando em  $\underline{a} = \sum_i \alpha_i \underline{x}_i$ .

que possa ser aplicada a metodologia descrita até aqui sobre os padrões mapeados  $\underline{x}^*$ . Esse mapeamento permite que a construção da hipersfera seja feita de forma mais flexível em relação à distribuição dos dados. Sendo assim, as Equações (70) e (74) podem ser reescritas como:

$$L = \sum_i \alpha_i \Phi(\underline{x}_i) \cdot \Phi(\underline{x}_i) - \sum_{i,j} \alpha_i \alpha_j \Phi(\underline{x}_i) \cdot \Phi(\underline{x}_j), \quad (75)$$

$$f(\underline{z}; \underline{\alpha}, R) = I \left( \Phi(\underline{z}) \cdot \Phi(\underline{z}) - 2 \sum_i \alpha_i \Phi(\underline{z}) \cdot \Phi(\underline{x}_i) + \sum_{i,j} \alpha_i \alpha_j \Phi(\underline{x}_i) \cdot \Phi(\underline{x}_j) \leq R^2 \right). \quad (76)$$

Como o mapeamento é aplicado apenas para os vetores sobre os quais é calculado o produto interno, pode-se utilizar a representação através do *Kernel* de Mercer (VAPNIK, 1998), não necessitando da representação do mapeamento explicitamente, uma vez que  $K(\underline{x}_i, \underline{x}_j) = \Phi(\underline{x}_i) \cdot \Phi(\underline{x}_j)$ . Essa observação resulta na seguinte modificação:

$$L = \sum_i \alpha_i K(\underline{x}_i, \underline{x}_i) - \sum_{i,j} \alpha_i \alpha_j K(\underline{x}_i, \underline{x}_j), \quad (77)$$

$$f(\underline{z}; \underline{\alpha}, R) = I \left( K(\underline{z}, \underline{z}) - 2 \sum_i \alpha_i K(\underline{z}, \underline{x}_i) + \sum_{i,j} \alpha_i \alpha_j K(\underline{x}_i, \underline{x}_j) \leq R^2 \right). \quad (78)$$

Utilizando o *kernel* gaussiano, o Lagrangeano é dado por:

$$L = \sum_i \alpha_i \exp \left( \frac{-\|\underline{x}_i - \underline{x}_i\|^2}{\sigma^2} \right) - \sum_{i,j} \alpha_i \alpha_j \exp \left( \frac{-\|\underline{x}_i - \underline{x}_j\|^2}{\sigma^2} \right), \quad (79)$$

sendo  $\sigma$  o parâmetro do *kernel* gaussiano. Nessa representação, o padrão de teste  $\underline{z}$  é classificado como novidade se:

$$\sum_i \alpha_i \exp \left( \frac{-\|\underline{z} - \underline{x}_i\|^2}{\sigma^2} \right) < \frac{1}{2} \left( 1 + \sum_{i,j} \alpha_i \alpha_j \exp \left( \frac{-\|\underline{x}_i - \underline{x}_j\|^2}{\sigma^2} \right) - R^2 \right). \quad (80)$$

#### 4.8.6 DETERMINAÇÃO AUTOMÁTICA DOS PARÂMETROS DE TREINAMENTO

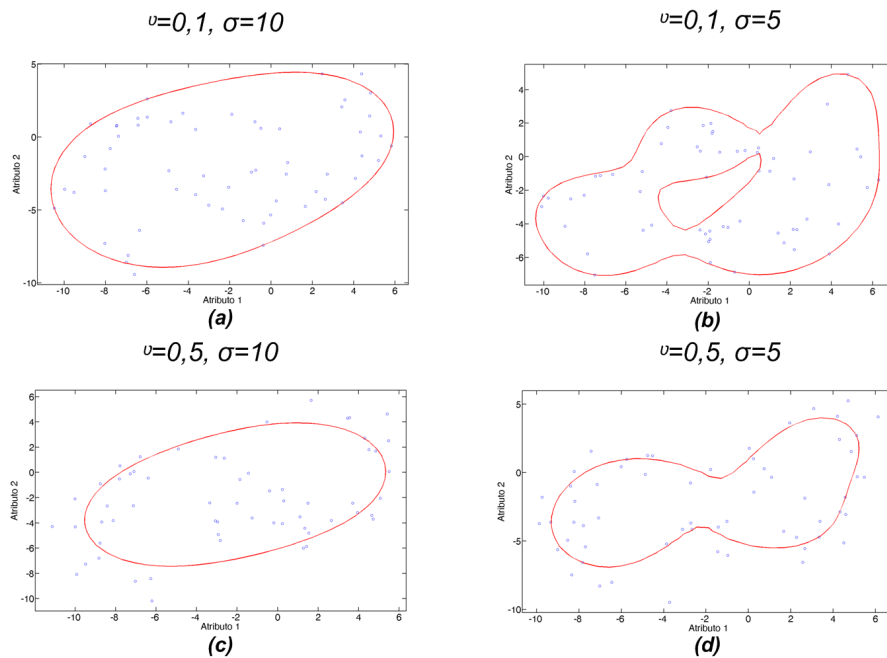
A definição automática dos parâmetros  $\sigma$  e  $C$  está diretamente associada à superfície de decisão gerada pelo modelo com base na Representação de Dados por Vetores Suporte. No caso do parâmetro do *kernel* Gaussiano  $\sigma$ , pode-se afirmar que quanto menor o valor adotado, mais detalhes são mantidos na representação, ao passo que quanto maior o valor escolhido para esse parâmetro, mais a representação tende para a solução da própria hipersfera no espaço de entrada (DUIN, 2001).

Já o parâmetro  $C$  está indiretamente relacionado à quantidade de dados a ser incluída

no interior da representação, ou seja, quantos padrões de entrada serão considerados como novidades e quantos serão considerados como normais. Uma forma de facilitar essa interpretação é utilizar a seguinte relação:

$$v = \frac{1}{NC}, \quad (81)$$

sendo  $N$  o número de padrões de entrada. Schölkopf e Smola (2001) demonstraram que  $v$  fornece um limite superior para a fração dos dados a serem rejeitados pela representação (novidades), sendo que, quanto maior o valor selecionado para esse parâmetro (ou analogamente menor o valor de  $C$ ), mais dados de entrada são representados como novidades, conforme pode ser observado na Figura 45.



**Figura 45: Variação dos parâmetros para o modelo RDVS.**

No caso de dados de entrada com dois atributos, a escolha desses parâmetros pode ser feita de forma visual, ao contrário do que ocorre para espaços de elevada dimensionalidade, cuja visualização é impossibilitada. Nesses casos, é preciso utilizar outras formas de realizar a escolha tendo por base uma boa qualidade na representação (boa capacidade de generalização do modelo). Para tanto, é necessário minimizar as taxas de falsos negativos e falsos positivos, sendo os exemplos positivos os padrões normais e os exemplos negativos as novidades, nesse caso.

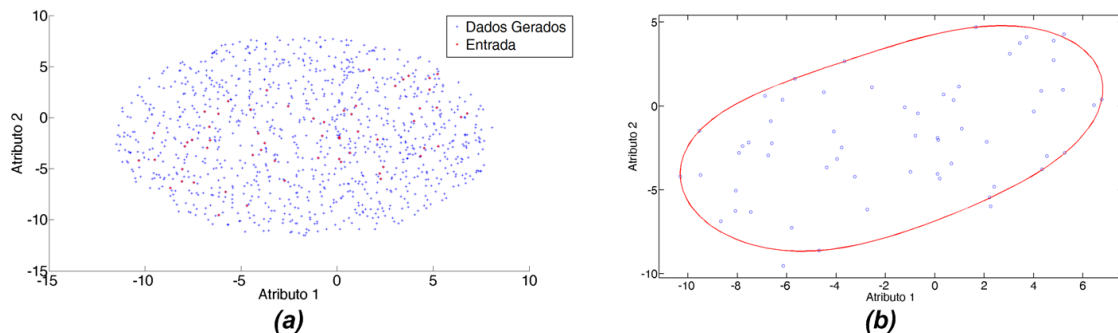
Quando se dispõe apenas dos dados normais, ou seja, não se conhecem as novidades para o modelo, a taxa de falsos negativos pode ser estimada através do número de vetores

suportes determinados pelo modelo (CHAPELLE et al., 2002), ou seja:

$$Falsos\ Negativos = \frac{No.\ de\ Vetores\ Suporte}{N}. \quad (82)$$

O problema de minimizar apenas a parcela relacionada com os falsos negativos é que a solução para o modelo poderia ser a de aceitar todo o espaço de características como padrão normal, não permitindo que novidades sejam identificadas.

Para estimar a taxa de falsos positivos é necessário ter conhecimento sobre uma fração das novidades. Na falta desse conhecimento, é possível assumir uma distribuição uniforme para as novidades, na região do espaço de características na qual se tem interesse, de forma a permitir uma estimativa da taxa de falsos positivos considerando essa distribuição. Nesse sentido, a proposta utilizada nesse trabalho baseia-se no desenvolvimento proposto por Duin (2001). Naquele trabalho, utilizam-se dados gerados artificialmente, segundo uma distribuição hipersférica uniforme, contendo uma fração de dados normais e uma fração de novidades, conforme demonstra a Figura 46.



**Figura 46: Definição automática dos parâmetros para o modelo RDVS.**

Desse modo, é possível realizar a estimação das taxas de falsos negativos e falsos positivos, estimando os parâmetros  $\sigma$  e  $C$  através da minimização do seguinte funcional:

$$\Lambda(\sigma, C) = \lambda \left( \frac{No.\ de\ Vetores\ Suporte}{N} \right) + (1 - \lambda) \left( \frac{No.\ de\ Falsos\ Positivos}{No.\ de\ novidades} \right). \quad (83)$$

Para minimização desse funcional utilizou-se o valor de 0,5 para  $\lambda$  de forma a garantir o mesmo peso para ambas as parcelas, falsos positivos e falsos negativos. Para solução do problema de minimização utilizou-se um algoritmo genético<sup>6</sup> (AG). A sua principal característica é emular processos biológicos baseados em teorias evolutivas, em particular a Teoria da Seleção

<sup>6</sup>Muito embora Duin (2001) não demonstre a necessidade de utilização de um AG para minimização desse funcional, verificou-se, na prática, desempenhos superiores de minimização utilizando um AG em relação a outros métodos de otimização lineares.

Natural. Esse tipo de algoritmo é voltado para a solução de problemas de otimização através da manipulação de uma população de indivíduos, preservando 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).

A ideia básica de um AG é gerar uma população inicial formada por um grupo aleatório 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 à função objetivo do problema. Depois de definida a aptidão de toda a população, a parcela dos indivíduos considerados mais aptos é mantida, enquanto 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. Esse processo é repetido até que um conjunto de soluções satisfatórias seja encontrado (GOLDBERG, 1989).

No intuito de garantir a convergência do algoritmo, aplicam-se 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 são utilizadas taxas de mutação entre 0,1% e 10%.

#### 4.8.7 REPRESENTAÇÃO DE DADOS COM VETORES SUPORTE PARA DETECÇÃO DE EVENTOS

Conforme apresentado anteriormente, métodos de detecção (segmentação) de oscilografias baseados na TWD normalmente utilizam apenas um nível da transformada para o processo de detecção como um todo. No entanto, o conteúdo de frequência do transitório pode estar distribuído em diferentes níveis da TWD e, em função disso, é importante utilizar uma decomposição em mais de um nível da TWD para maximizar a possibilidade de identificar as transições abruptas nas oscilografias.

Para apresentar a metodologia proposta neste trabalho, que visa combinar os diferentes

níveis da TWD, assume-se uma decomposição *wavelet* em dois níveis, conforme mostra a Figura 47. Nessa figura, *Característica 1* representa o primeiro nível de decomposição, enquanto *Característica 2* representa o segundo nível. Essa é apenas uma representação esquemática do método proposto e não corresponde a uma decomposição de um sinal real.

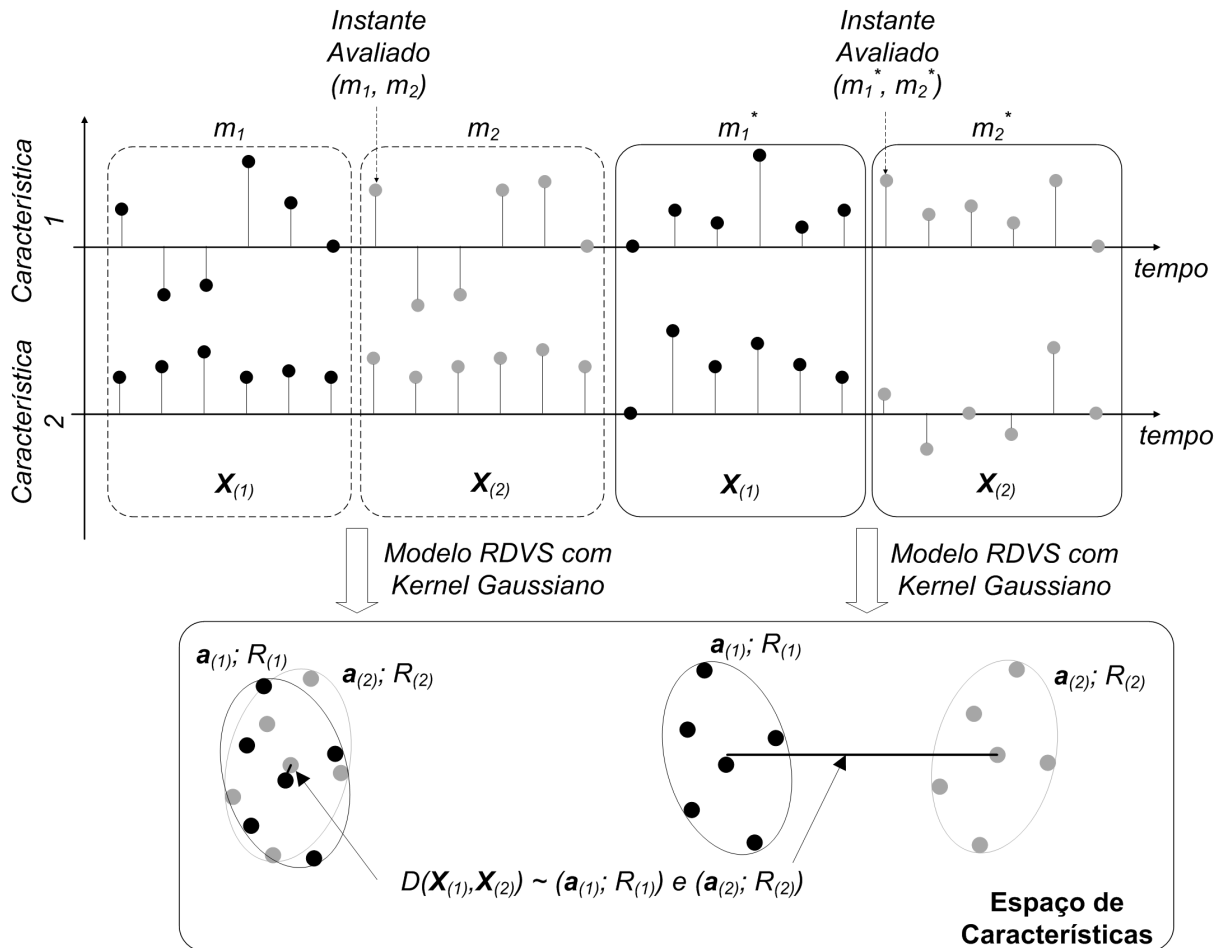


Figura 47: Composição dos diversos níveis da TWD com o modelo RDVS.

É possível utilizar o modelo RDVS (basicamente o raio e o centro da hiperesfera determinada) para combinar os diversos níveis da TWD. Para tanto, em um determinado instante de tempo  $t$ , são extraídas duas janelas para análise dos dados, gerando as sequências  $\mathbf{X}_{(1)} = (\mathbf{x}_{t-m_1}, \dots, \mathbf{x}_{t-1})$  e  $\mathbf{X}_{(2)} = (\mathbf{x}_t, \dots, \mathbf{x}_{t+m_2-1})$ , cujas dimensões são  $(2, m_1)$  e  $(2, m_2)$ , respectivamente (DESOBRY et al., 2005). Supondo também, que sobre cada uma dessas sequências, é aplicada uma Representação de Dados com Vetores Suporte, de modo que os dados de entrada desse modelo ( $\mathbf{X}_{(1)}$  e  $\mathbf{X}_{(2)}$ ) possam ser representados através dos vetores suporte das hiperesferas com raios  $R_{(1)}$ ,  $R_{(2)}$  e centros  $\mathbf{a}_{(1)}$ ,  $\mathbf{a}_{(2)}$ , como mostra a Figura 47.

Nessa figura são apresentados dois instantes de tempo distintos, nos quais é feita a análise proposta. O primeiro instante de tempo é calculado para as janelas  $m_1$  e  $m_2$ , conforme



assinalado na Figura 47. Nesse instante de tempo, toma-se a Representação de Dados com Vetores Suporte para cada janela e o seu resultado (apresentado aqui de forma didática), refletindo o comportamento dos padrões de entrada em um novo domínio, cuja representação está associada a uma hipersfera de raio e centro determinados pelo algoritmo.

O mesmo procedimento pode ser adotado para um instante de tempo posterior, conforme demonstram as janelas  $m_1^*$  e  $m_2^*$ . Nesse caso, observa-se que o mapeamento gera uma representação bastante dissimilar para as duas janelas, já que as hipersferas encontram-se bastante separadas no espaço de características. Quando se observa a característica dos atributos, percebe-se que nesse último caso existe um comportamento dissimilar entre as duas janelas sob análise, que é automaticamente refletido para o espaço de alta dimensionalidade em que se encontram as hipersferas.

É possível utilizar o método descrito para determinar variações abruptas em sinais elétricos (oscilografias). Formalmente, a distância entre as hipersferas pode ser calculada pela seguinte Equação:

$$D(\mathbf{X}_{(1)} - \mathbf{X}_{(2)}) = \frac{\|\mathbf{a}_{(1)} - \mathbf{a}_{(2)}\|^2}{R_{(1)}^2 + R_{(2)}^2}. \quad (84)$$

Essa medida de distância (dissimilaridade)  $D(\mathbf{X}_{(1)}, \mathbf{X}_{(2)})$  contém duas partes principais. A primeira está relacionada à própria distância entre as hipersferas no espaço de características. Tal distância é dada pela seguinte relação:

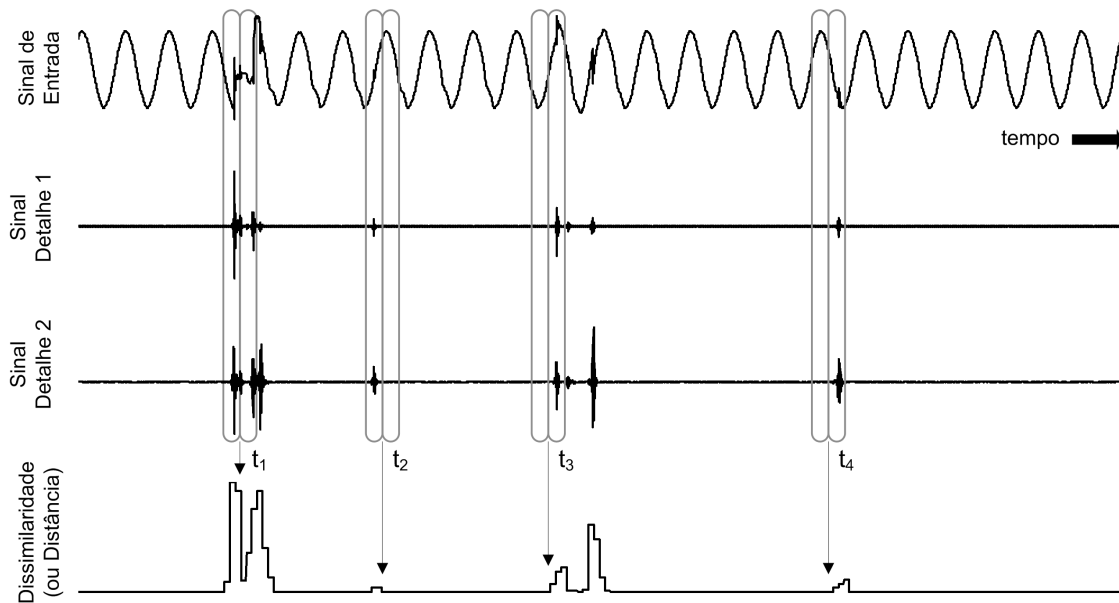
$$\begin{aligned} \|\underline{a}_{(1)} - \underline{a}_{(2)}\|^2 = & \sum_{i,j} \alpha_i^{(1)} \alpha_j^{(1)} K(\underline{x}_i^{(1)}, \underline{x}_j^{(1)}) + \sum_{i,j} \alpha_i^{(2)} \alpha_j^{(2)} K(\underline{x}_i^{(2)}, \underline{x}_j^{(2)}) \\ & - 2 \sum_{i,j} \alpha_i^{(1)} \alpha_j^{(2)} K(\underline{x}_i^{(1)}, \underline{x}_j^{(2)}). \end{aligned} \quad (85)$$

O segundo termo está relacionado à normalização dessa medida, uma vez que o raio da hipersfera está diretamente relacionado ao volume ocupado pelos dados no espaço de características. O raio  $R_{(k)}^2$ , associado à janela de análise  $m_{(k)}$ , é dado por:

$$\|\underline{a}_{(k)} - \Phi(\underline{x}_s^{(k)})\|^2 = K(\underline{x}_s^{(k)}, \underline{x}_s^{(k)}) - 2 \sum_i \alpha_i^{(k)} K(\underline{x}_i^{(k)}, \underline{x}_s^{(k)}) + \sum_{i,j} \alpha_i^{(k)} \alpha_j^{(k)} K(\underline{x}_i^{(k)}, \underline{x}_j^{(k)}). \quad (86)$$

As hipersferas são atualizadas para cada instante de tempo  $t$ , considerando as janelas sob análise, resultando em um valor para  $D(\mathbf{X}_{(1)}, \mathbf{X}_{(2)})$  proporcional à distância euclidiana entre os centros das hipersferas no espaço de elevada dimensionalidade. Quanto maior a diferença entre as amostras subsequentes, maior a distância nesse espaço e, conseqüentemente, maior a probabilidade do instante de tempo em questão estar associado a uma variação abrupta do sinal,

como mostra a Figura 48.

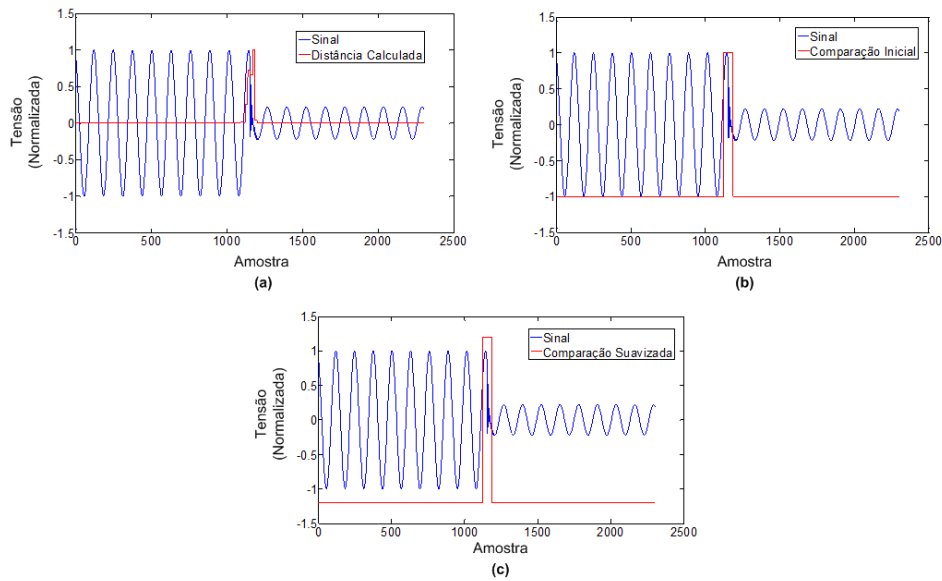


**Figura 48: Visão geral da metodologia proposta para o cálculo de distância (dissimilaridade).**

Um exemplo dessa aplicação está representada na Figura 49, sendo o sinal de entrada decomposto em quatro níveis da TWD, e as janelas de observação  $m_1$  e  $m_2$  iguais a  $1/8$  de ciclo para uma frequência de amostragem de 7680 Hz. Com a distância calculada, pode-se empregar a mesma metodologia de comparação com um limiar e a suavização apresentados para os modelos anteriores, através da comparação de  $D(\mathbf{X}_{(1)}, \mathbf{X}_{(2)})$  com o dado limiar empiricamente determinado.

Nesse tipo de representação é possível extrair as vantagens que outros mapeamentos fornecem para detecção de eventos. No caso apresentado, utiliza-se da vantagem fornecida pela aplicação da TWD, da qual são extraídos coeficientes proporcionais às variações associadas com os transitórios a serem detectados no sinal. Desse modo, pode-se, através do pré-processamento do sinal com base na TWD, realizar uma representação mais robusta para problemas de detecção de eventos, uma vez que nessa metodologia é possível maximizar as distâncias entre as hiperesferas cujos atributos de entrada são os coeficientes da TWD nos seus diferentes níveis.

A mesma observação apresentada para o Operador de Energia de Teager, em relação ao cálculo automático do limiar através do limiar universal, vale para os modelos RDVS. Ou seja, não se garante as restrições impostas para o cálculo desse limiar na modelagem RDVS. Sendo assim, optou-se por utilizar a mesma proposta descrita na Seção 4.3.1 para determinação automática do limiar do modelo RDVS, porém, nesse caso, utilizam-se as distâncias entre as hiperesferas calculadas para as oscilografias normais, como referência de determinação do limiar.



**Figura 49: Exemplo de segmentação para modelos utilizando RDVS. Cálculo de  $D(\mathbf{X}_{(1)}, \mathbf{X}_{(2)})$  (a) seguido da limiarização através da comparação com limiar empírico (b). Suavização (c) limiarização da suavização apresentada em (b). É interessante observar que, para esse modelo, não há necessidade de suavização, uma vez que a saída da limiarização não apresenta variações como nos modelos AR e FK. Além disso, pode-se afirmar que  $D(\mathbf{X}_{(1)}, \mathbf{X}_{(2)})$  é análoga aos resíduos estimados dos modelos AR e FK.**

É importante ressaltar que métodos de detecção de variações abruptas em sinais utilizam metodologias similares para estimar a dissimilaridade entre duas janelas de análise subsequentes (DESOBRY et al., 2005). Normalmente, métodos de estimação da função densidade de probabilidade são aplicados em cada janela e as funções resultantes são comparadas através de medidas de dissimilaridade entre as distribuições. No entanto, para o problema apresentado no presente trabalho, onde se verificam poucos exemplos por janela e dimensionalidade alta, métodos baseados em distribuições de probabilidade não são os mais adequados.

Adicionalmente, o método aqui proposto é similar ao trabalho apresentado por Desobry et al. (2005). No entanto, o método do presente trabalho é apresentado somente em termos das hipersferas no espaço de características, ao contrário do método proposto em Desobry et al. (2005), que considera a distância entre os hiperplanos de separação de cada janela, ambos referenciados à origem do espaço de características.

#### 4.8.8 RDVS COM EXEMPLOS NEGATIVOS

No caso particular do modelo RDVS, é possível incluir na formulação, os exemplos das demais classes durante o treinamento do modelo (quando disponíveis). Nesse caso, a formulação deve considerar a inclusão dos padrões de uma dada classe no interior da hi-

resfera, enquanto os padrões das demais classes devem se localizar no exterior da hipersfera (TAX; DUIN, 2001). A diferença dessa abordagem de treinamento em relação a um classificador do tipo MVS estimado dentro de um processo um-contra-todos é que, no caso dos modelos baseados em RDVS, existe uma restrição teórica que impõe a superfície fechada em torno da classe sob treinamento (hiperesfera). Esse tipo de característica é fundamental quando se trata de um aprendizado dinâmico, tal qual é exposto neste trabalho, uma vez que outras classes podem surgir em outras regiões do espaço de características.

Considerando o conjunto de treinamento composto pelos pares  $(\underline{x}_1, \dots, \underline{x}_N)$ , tal que  $(\underline{x}_1, y_1), \dots, (\underline{x}_N, y_N)$  com  $N$  exemplos  $K$  classes conhecidas previamente, sendo  $\underline{x}_i \in \mathbb{R}^n$  e  $y_j \in \{1, \dots, K\}$ . Para facilitar a formulação, será considerado que, para cada classe  $k$ , os rótulos da classe em questão serão  $y_k = +1$  e para as demais classes  $l$ , tal que  $l \neq k$ ,  $y_l = -1$ . Com isso, a formulação com base na RDVS passa a ser dada pela seguinte equação:

$$\varepsilon(R_k, \underline{a}_k, \underline{\xi}) = R_k^2 + C_1 \sum_{i:y_i=k} \xi_i + C_2 \sum_{l:y_l \neq k} \xi_l, \quad (87)$$

em conjunto com as restrições expostas a seguir:

$$\begin{aligned} \|\underline{x}_i - \underline{a}_k\|^2 &\leq R_k^2 + \xi_i, \forall i : y_i = k, \\ \|\underline{x}_l - \underline{a}_k\|^2 &\geq R_k^2 - \xi_l, \forall l : y_l \neq k, \\ \xi_i &\geq 0, \xi_l \geq 0, \forall i, l. \end{aligned} \quad (88)$$

É importante observar que a diferença fundamental em relação à formulação original do modelo RDVS é a inserção da penalização para padrões das classes  $l$ , tal que  $l \neq k$ , quando esses padrões estão localizados no interior da hiperesfera. Para tanto, também se faz necessário inserir um parâmetro de regularização  $C_2$ , responsável pela penalização dos padrões das classes  $l$ , tal que  $l \neq k$ , inseridos no interior da hiperesfera. As restrições podem ser incorporadas à formulação inicial, através da introdução dos multiplicadores de Lagrange  $\alpha_i, \alpha_l, \gamma_i, \gamma_l$ :

$$\begin{aligned} L(R_k, \underline{a}_k, \underline{\xi}, \underline{\alpha}, \underline{\gamma}) &= R_k^2 + C_1 \sum_{i:y_i=k} \xi_i + C_2 \sum_{l:y_l \neq k} \xi_l \\ &\quad - \sum_{i:y_i=k} \gamma_i \xi_i - \sum_{l:y_l \neq k} \gamma_l \xi_l \\ &\quad - \sum_{i:y_i=k} \alpha_i (R_k^2 + \xi_i - (\underline{x}_i \cdot \underline{x}_i - 2\underline{a}_k \cdot \underline{x}_i + \underline{a}_k \cdot \underline{a}_k)) \\ &\quad - \sum_{l:y_l \neq k} \alpha_l ((\underline{x}_l \cdot \underline{x}_l - 2\underline{a}_k \cdot \underline{x}_l + \underline{a}_k \cdot \underline{a}_k) - R_k^2 + \xi_l), \end{aligned} \quad (89)$$

sendo:

$$\alpha_i \geq 0, \alpha_l \geq 0, \gamma_i \geq 0, \gamma_l \geq 0. \quad (90)$$

Para resolver a formulação acima, basta tomar as derivadas parciais de  $L$  com relação à  $R_k$ ,  $\underline{a}_k$ ,  $\xi_i$  e  $\xi_j$ . Com isso, as seguintes relações são obtidas:

$$\begin{aligned} \sum_{i:y_i=k} \alpha_i - \sum_{l:y_l \neq k} \alpha_l &= 1, \\ \underline{a}_k &= \sum_{i:y_i=k} \alpha_i \underline{x}_i - \sum_{l:y_l \neq k} \alpha_l \underline{x}_l, \\ 0 &\leq \alpha_i \leq C_1, \forall i : y_i = k, \\ 0 &\leq \alpha_l \leq C_2, \forall l : y_l \neq k. \end{aligned} \quad (91)$$

E finalmente, o funcional  $L$  passa a ser dado por:

$$\begin{aligned} L &= \sum_{i:y_i=k} \alpha_i K(\underline{x}_i, \underline{x}_i) - \sum_{l:y_l \neq k} \alpha_l K(\underline{x}_l, \underline{x}_l) - \sum_{i,j:y_i, y_j = k} \alpha_i \alpha_j K(\underline{x}_i, \underline{x}_j) \\ &+ 2 \sum_{i:y_i=k, l:y_l \neq k} \alpha_i \alpha_j K(\underline{x}_l, \underline{x}_j) - \sum_{l,m:y_l, y_m \neq k} \alpha_l \alpha_m K(\underline{x}_l, \underline{x}_m), \end{aligned} \quad (92)$$

sujeito às restrições:

$$\begin{aligned} \sum_{i:y_i=k} \alpha_i - \sum_{l:y_l \neq k} \alpha_l &= 1, \\ 0 &\leq \alpha_i \leq C_1, \forall i : y_i = k, \\ 0 &\leq \alpha_l \leq C_2, \forall l : y_l \neq k. \end{aligned} \quad (93)$$

Novamente, o problema acima é a representação dual da minimização inicialmente proposta e pode ser resolvido através de programação quadrática com restrições.

É importante observar que o conceito de *kernel* pode ser novamente aplicado para representação do produto interno, sem a necessidade explícita da definição do mapeamento para o espaço de características, da mesma forma como foi apresentado na formulação padrão RDVS e de MVS.

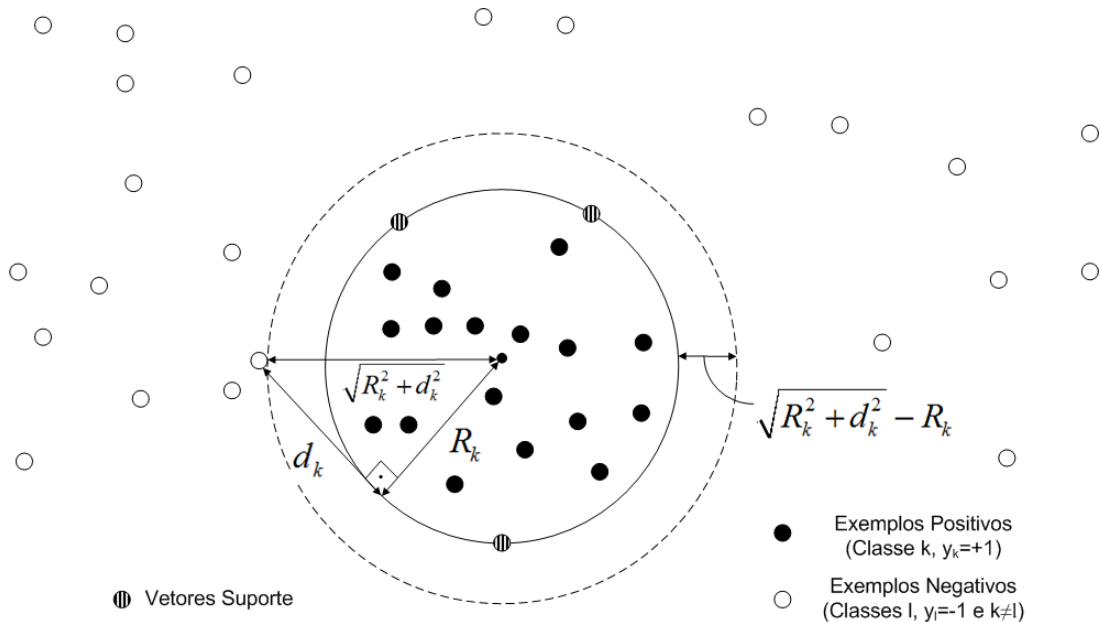
Essa otimização pode ser aplicada para cada classe conhecida do problema, de modo a obter uma hipersfera para cada classe, contendo no seu interior os exemplos da classe em questão e no exterior exemplos das demais classes. É importante destacar que a vantagem dessa abordagem na formulação apresentada para a RDVS com exemplos negativos é a própria utilização dos exemplos negativos na construção da superfície fechada de decisão. Como será demonstrado posteriormente, os resultados finais são consideravelmente alterados com essa

inclusão da formulação.

#### 4.8.9 RDVS COM EXEMPLOS NEGATIVOS E OTIMIZAÇÃO DA MARGEM

Até o momento, a inclusão dos exemplos negativos apenas acrescenta um termo de penalização para exemplos negativos no interior da hiperesfera. No entanto, não existem discussões explícitas sobre a separação máxima entre os padrões positivos e negativos, conforme é feito no modelo MVS (FERREIRA; SILVA, 2007). Uma possibilidade, nesse caso, seria incluir na formulação da RDVS com Exemplos Negativos um termo associado à tal separação, de modo a manter a penalização em conjunto com o conceito de separação máxima entre os padrões, conforme expõe Hao et al. (2007). Com isso, exemplos negativos estarão o mais distante possível da hiperesfera.

Conforme mostra a Figura 50, a margem é definida como a distância do exemplo negativo mais próximo da superfície de decisão da hiperesfera. Para adequar a formulação a esse requisito, será introduzido um novo fator de margem,  $d_k$ .



**Figura 50: Formulação RDVS com exemplos negativos e otimização da margem.**

Fonte: (HAO et al., 2007)

Matematicamente, a hiperesfera  $S_k$  definida para a classe  $k$ , caracterizada pelo centro  $a_k$  e raio  $R_k$ , pode ser encontrada através da minimização do seguinte funcional:

$$\varepsilon \left( R_k, a_k, \underline{\xi}, d_k \right) = R_k^2 + C_1 \sum_{i:y_i=k} \xi_i + C_2 \sum_{l:y_l \neq k} \xi_l - M d_k^2, \quad (94)$$

sujeito às restrições:

$$\begin{aligned}
\|\underline{x}_i - \underline{a}_k\|^2 &\leq R_k^2 + \xi_i, \forall i : y_i = k, \\
\|\underline{x}_l - \underline{a}_k\|^2 &\geq R_k^2 - \xi_l + d_k^2, \forall l : y_l \neq k, \\
\xi_i &\geq 0, \xi_l \geq 0, \forall i, l.
\end{aligned} \tag{95}$$

Em termos gerais, a principal diferença em relação à formulação anterior, é a inclusão do termo relacionado à otimização da margem de separação  $Md_k^2$ . Na Figura 50, está representada a relação entre o fator de margem  $d_k$  e a margem propriamente dita. A largura da margem é dada por  $\sqrt{R_k^2 + d_k^2} - R_k$ . Para maximizar esse termo, basta maximizar o fator de margem  $d_k$ , uma vez que a minimização de  $R_k$  está inserida em outro termo da formulação. O parâmetro  $M$  é responsável por regular o compromisso entre essa relação. A minimização exposta pode ser resolvida novamente através dos multiplicadores de Lagrange, resultando na seguinte representação:

$$\begin{aligned}
L(R_k, \underline{a}_k, d_k, \underline{\xi}, \underline{\alpha}, \gamma) &= R_k^2 - Md_k^2 + C_1 \sum_{i:y_i=k} \xi_i + C_2 \sum_{l:y_l \neq k} \xi_l \\
&\quad - \sum_{i:y_i=k} \gamma_i \xi_i - \sum_{l:y_l \neq k} \gamma_l \xi_l \\
&\quad - \sum_{i:y_i=k} \alpha_i (R_k^2 + \xi_i - (\underline{x}_i \cdot \underline{x}_i - 2\underline{a}_k \cdot \underline{x}_i + \underline{a}_k \cdot \underline{a}_k)) \\
&\quad - \sum_{l:y_l \neq k} \alpha_l ((\underline{x}_l \cdot \underline{x}_l - 2\underline{a}_k \cdot \underline{x}_l + \underline{a}_k \cdot \underline{a}_k) - R_k^2 - d_k^2 + \xi_l).
\end{aligned} \tag{96}$$

Aplicando a derivada parcial de  $L$  em relação à  $R_k$ ,  $\underline{a}_k$ ,  $d_k$ ,  $\xi_i$  e  $\xi_l$ , chega-se ao seguinte problema de maximização (representação dual):

$$\begin{aligned}
L &= \sum_{i:y_i=k} \alpha_i K(\underline{x}_i, \underline{x}_i) - \sum_{l:y_l \neq k} \alpha_l K(\underline{x}_l, \underline{x}_l) - \sum_{i,j:y_i,y_j=k} \alpha_i \alpha_j K(\underline{x}_i, \underline{x}_j) \\
&\quad + 2 \sum_{i:y_i=k, l:y_l \neq k} \alpha_i \alpha_j K(\underline{x}_l, \underline{x}_j) - \sum_{l,m:y_l,y_m \neq k} \alpha_l \alpha_m K(\underline{x}_l, \underline{x}_m),
\end{aligned} \tag{97}$$

sujeito às restrições:

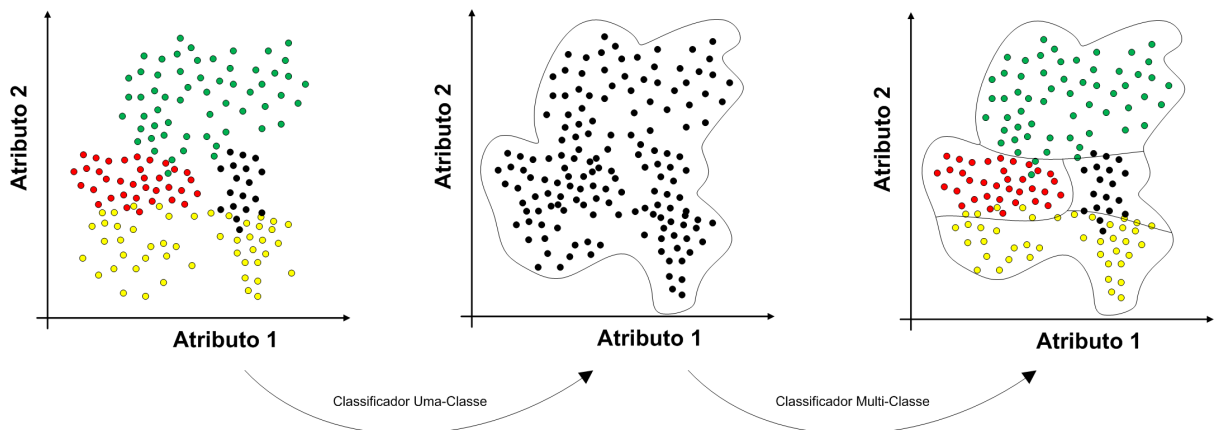
$$\begin{aligned}
\sum_{i:y_i=k} \alpha_i &= 1 + M, \\
\sum_{l:y_l \neq k} \alpha_l &= M, \\
0 &\leq \alpha_i \leq C_1, \forall i : y_i = k, \\
0 &\leq \alpha_l \leq C_2, \forall l : y_l \neq k.
\end{aligned} \tag{98}$$

Os demais parâmetros do modelo ( $R_k$ ,  $\underline{a}_k$  e  $d_k$ ) podem ser obtidos através da aplicação das condições de Karush-Kuhn-Tucker, conforme exposto por Hao et al. (2007):

$$\begin{aligned} \underline{a}_k &= \sum_{i:y_i=k} \alpha_i \varphi(\underline{x}_i) - \sum_{l:y_l \neq k} \alpha_l \varphi(\underline{x}_l), \\ R_k^2 &= \|\varphi(\underline{x}_i) - \underline{a}_k\|^2, \forall i: 0 < \alpha_i < C_1, \\ d_k^2 &= \|\varphi(\underline{x}_l) - \underline{a}_k\|^2 - R_k^2, \forall l: 0 < \alpha_l < C_2. \end{aligned} \quad (99)$$

#### 4.8.10 ABORDAGEM PARA CLASSIFICAÇÃO MULTICLASSE

É possível utilizar os classificadores uma-classe apresentados anteriormente para realizar uma classificação multiclasse (HAO et al., 2007). Existem basicamente duas abordagens para tanto. A primeira delas consiste em realizar o processo de classificação multiclasse e detecção de novidades em dois estágios (TAX; DUIN, 2001). Na fase de treinamento, todas as classes conhecidas do problema são agrupadas e sobre elas é construído um classificador uma-classe. Com isso, define-se uma fronteira sobre os padrões conhecidos, permitindo a detecção de novidades. Para essas mesmas classes, é realizado o treinamento de um classificador multiclasse que define, em linhas gerais, as fronteiras de decisão entre as classes (no presente trabalho, esse classificador é baseado em MVS). A Figura 51 representa um resumo desse procedimento.



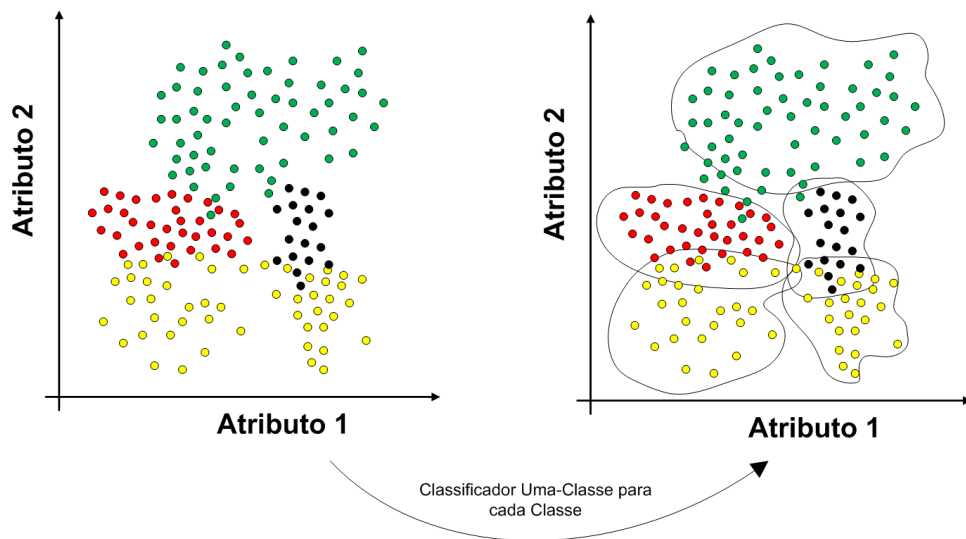
**Figura 51: Detecção de novidades e classificação multiclasse simultâneos.** Na fase de treinamento dos modelos, as classes são agrupadas para a construção do detector de novidades e, em seguida, são aplicadas a um classificador multiclasse (MVS). O resultado final está representado didaticamente aqui, através das superfícies de separação entre as classes (classificador multiclasse) e a superfície que engloba todas as classes (detector de novidades).

Na fase de avaliação, ou de teste, um novo padrão é primeiramente aplicado ao detector de novidades. Pertencendo a uma das classes conhecidas do problema, define-se sua classe através do classificador multiclasse construído. Caso contrário, o novo padrão é considerado



uma novidade. Essa é a abordagem denominada DNG-CMC apresentada no início do trabalho.

A construção de um classificador uma-classe por classe fornece uma superfície fechada para cada classe conhecida do problema. Nesse sentido, é possível realizar a identificação de novidades e a classificação multiclasse de forma conjunta, dando origem à segunda abordagem denominada de DNI-CBS, como exemplifica a Figura 52. É importante ressaltar que, quando um modelo uma-classe é construído para cada classe conhecida do problema, os padrões pertencentes às demais classes são considerados como novidades.



**Figura 52:** Classificação multiclasse com um classificador uma-classe por classe. Na fase de treinamento dos modelos, as classes são utilizadas formadas por superfícies fechadas. O resultado final está representado didaticamente aqui, através das superfícies de separação que englobam cada classe.

Na fase de teste dessa última abordagem verifica-se, para cada um dos modelos uma-classe definidos, se o novo padrão pertence ou não a essa representação. No final desse processo, alguns padrões irão pertencer a apenas um dos modelos, outros pertencerão a mais de um dos modelos e um terceiro conjunto de padrões não pertencerá a nenhum dos modelos ajustados, sendo esses, caracterizados como novidades.

No caso dos padrões pertencentes a mais de um modelo, principalmente aqueles situados na região de sobreposição entre classes no espaço dos atributos, é necessário um pós-processamento para verificar qual a classe mais provável a que esse padrão pertença. Esse pós-processamento é baseado na análise de similaridade entre o padrão e o modelo  $S_k$  que esse padrão é comparado (HAO et al., 2007). No caso dos modelos baseados em estimação da função densidade de probabilidade (Mistura de Gaussianas e estimadores de Parzen), a medida de similaridade está relacionada com a probabilidade *a posteriori*, ou seja, o padrão é atribuído ao modelo com maior probabilidade *a posteriori*. Essa metodologia também é adotada para o

modelo com base nos  $K$ -Vizinhos-Mais-Próximos.

O modelo  $K$ -Médias adota como medida de similaridade a distância euclidiana entre o padrão e o centro de cada modelo, de forma que o padrão é atribuído ao centro do modelo mais próximo. Já o modelo de RDVS considera a distância euclidiana entre o padrão de teste  $\underline{z}$  no espaço de características e o centro  $\underline{a}_k$  do modelo:

$$\begin{aligned} sim(\underline{z}, S_k) &= -\|\varphi(\underline{z}) - \underline{a}_k\|^2, \\ sim(\underline{z}, S_k) &= -K(\underline{z}, \underline{z}) + 2 \sum_i \alpha_i K(\underline{z}, \underline{x}_i) - \sum_{i,j} \alpha_i \alpha_j K(\underline{x}_i, \underline{x}_j). \end{aligned} \quad (100)$$

Já para os modelos RDVS com exemplos negativos (e otimização da margem), pode-se utilizar o mesmo procedimento para os  $k$  modelos estimados para realizar a classificação através da medida de similaridade entre o padrão de teste  $\underline{z}$  e cada modelo  $S_k$  estimado:

$$\begin{aligned} sim(\underline{z}, S_k) &= -\|\varphi(\underline{z}) - \underline{a}_k\|^2, \\ sim(\underline{z}, S_k) &= -K(\underline{z}, \underline{z}) + 2 \sum_i \alpha_i K(\underline{z}, \underline{x}_i) - 2 \sum_l \alpha_l K(\underline{z}, \underline{x}_l) \\ &\quad - \sum_{i,j} \alpha_i \alpha_j K(\underline{x}_i, \underline{x}_j) + 2 \sum_{i,l} \alpha_i \alpha_l K(\underline{x}_i, \underline{x}_l) - \sum_{l,m} \alpha_l \alpha_m K(\underline{x}_l, \underline{x}_m). \end{aligned} \quad (101)$$

O valor de  $sim(\underline{z}, S_k)$  é inversamente proporcional à distância entre  $\varphi(\underline{z})$  e  $\underline{a}_k$ . Sendo assim, a classe atribuída está relacionada ao modelo  $k$  com o maior valor de similaridade  $sim(\underline{z}, S_k)$ . É importante destacar que a utilização direta dessa medida de similaridade está associada com a normalização das saídas de cada modelo  $k$ , conforme exposto por Tax e Duin (2008).

## 4.9 ANÁLISE ESTATÍSTICA

A seguir serão apresentadas as ferramentas utilizadas para comparação estatística dos resultados, tanto para os métodos de segmentação (detecção), quanto para os de classificação de eventos.

### 4.9.1 ANÁLISE ROC

A análise ROC, do inglês *Receiver Operating Characteristic*, foi inicialmente proposta para análise e escolha de diferentes métodos de detecção de sinais, visando avaliar a qualidade de transmissão de um sinal em um canal com ruído (PRATI et al., 2008). Sua utilização em métodos de classificação baseia-se na possibilidade de avaliar cenários nos quais existe uma

grande disparidade entre número de exemplos por classe. Por ser fundamentalmente baseado no erro (acerto) de classificação, pode fornecer também uma comparação para diferentes relações custo/benefício para os diferentes erros (acertos) entre diversos métodos de classificação (FAWCETT, 2006).

Em um problema de classificação binária, onde se tem uma classe dita positiva e outra negativa, uma maneira usual de se determinar o desempenho global do classificador é através da matriz de confusão (BISHOP, 1995). Nessa matriz, os exemplos positivos classificados como positivos são denominados verdadeiros positivos. Já os eventos negativos classificados como positivos são chamados de falsos positivos. A nomenclatura dessa tabela está resumida na sua própria representação apresentada na Tabela 2.

**Tabela 2: Tabela de representação ROC (matriz de confusão) para um problema de duas classes.**

–	Valor Real (Positivo)	Valor Real (Negativo)
Valor Estimado (Positivo)	Verdadeiro Positivo (VP)	Falso Positivo (FP)
Valor Estimado (Negativo)	Falso Negativo (FN)	Verdadeiro Negativo (VN)

Na Tabela 2, a diagonal principal representa as decisões corretas do classificador, ao contrário dos demais elementos, que estão associados com o erro (confusão) dentre as classes. Com base nesses conceitos, pode-se definir a taxa de verdadeiros positivos e de falsos positivos como:

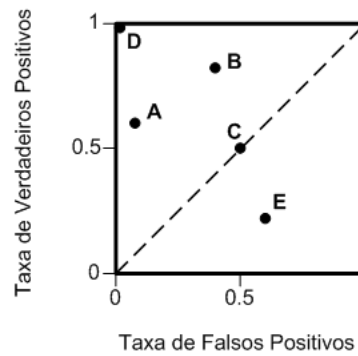
$$\text{Taxa de Verdadeiros Positivos(TVP)} = \frac{VP}{VP + FP}, \quad (102)$$

$$\text{Taxa de Falsos Positivos(TFP)} = \frac{FP}{VN + FN}. \quad (103)$$

Essas taxas definem a base do gráfico ROC, apresentado na Figura 53, no qual se avalia o compromisso entre os benefícios associados com a taxa de verdadeiros positivos e custos provenientes da taxa de falsos positivos, para diferentes classificadores discretos.

No gráfico ROC existem alguns pontos que merecem destaque. O ponto de coordenadas (0,0) representa uma estratégia de não atribuir uma classificação positiva, evitando totalmente erros do tipo falso positivos, muito embora não efetue classificações do tipo verdadeiro positivo. Uma estratégia totalmente oposta pode ser obtida com um classificador binário localizado em (1,1).

O ponto D, de coordenadas (0,1), representa uma classificação perfeita. Com isso, define-se que um ponto no espaço ROC possui um melhor desempenho quanto mais alta sua taxa de verdadeiros positivos e mais baixa a taxa de falsos positivos, ou seja, mais próximo de D. Classificadores situados na parte esquerda do gráfico podem ser considerados como conservadores, ao passo que classificadores situados na parte direita podem ser considerados como



**Figura 53: Exemplo de uma análise ROC.**

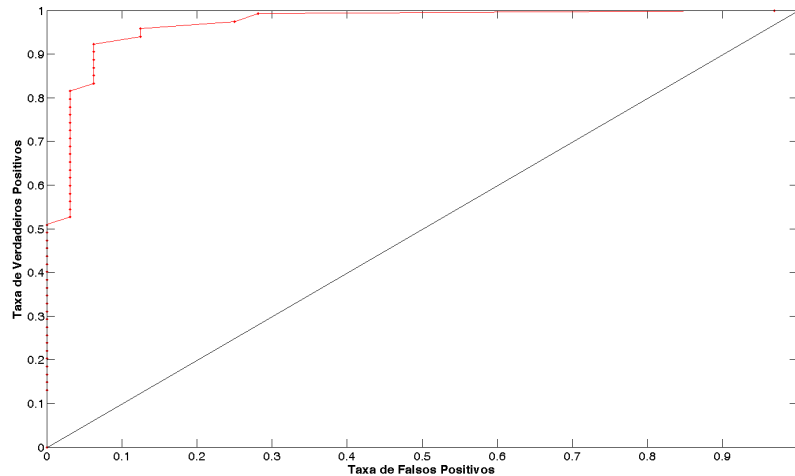
**Fonte: (PRATI et al., 2008)**

mais tolerantes, apresentando taxa de falsos positivos mais elevada. Nesse sentido, o ponto A pode ser dito mais conservador que B.

Os pontos que possuem a taxa de verdadeiros positivos igual à taxa de falsos positivos, como o exemplo do ponto C no gráfico, possuem desempenho caracterizado como aleatório, uma vez que a probabilidade de classificar um exemplo positivo de forma correta é a mesma que classificar um exemplo negativo de forma incorreta. Essa região da curva funciona como um limiar para comparação de um dado classificador binário em relação a um classificador aleatório, ou seja, pontos localizados no triângulo inferior direito do gráfico (ponto E, p. ex.) possuem desempenho inferior a um classificador de natureza aleatória para a dicotomia em questão.

Conforme apresentado anteriormente, cada ponto na curva ROC representa um classificador discreto, ou seja, um classificador que fornece apenas a classe escolhida (positiva ou negativa) para um dado exemplo. Nesse sentido, a análise ROC também pode ser estendida para classificadores cujas saídas estão relacionadas com probabilidades (HANLEY et al., 1982), ou até mesmo distâncias (TAX, 2001; DESOBRY et al., 2005). Através da definição de diferentes limiares para a saída do classificador, obtêm-se diferentes classificadores discretos na curva ROC, ou seja, se a saída do classificador é maior que o limiar definido, o exemplo é classificado como positivo, caso contrário, é negativo. Tomando por base o resíduo calculado para o evento da Figura 26 e variando-se o limiar de comparação entre 50 pontos selecionados no intervalo  $[0,1]$ , obtêm-se a curva ROC da Figura 54:

No caso particular dos problemas de detecção de eventos, a definição de diferentes limiares através da análise ROC está diretamente associada às saídas geradas pelos modelos, tanto para os baseados em resíduos (BOLLEN; GU, 2006), quanto os modelos baseados em distâncias (DESOBRY et al., 2005). Desse modo, pode-se avaliar o desempenho do classifi-



**Figura 54: Exemplo de uma análise ROC para detecção de variações abruptas.**

cador para diferentes limiares adotados, permitindo indiretamente a escolha do melhor limiar para uma dada análise, cujo padrão-ouro<sup>7</sup> é fornecido. Nesse sentido, a análise do desempenho do detector de variações abruptas é relacionada com o próprio mapeamento gerado pelo detector, seja ele probabilístico, ou com base em distâncias, independentemente do limiar escolhido, transformando o problema de detecção em  $N$  problemas de classificação binária, para cada um dos  $N$  limiares escolhidos para análise. No caso particular do procedimento de suavização proposto nas seções anteriores, o resultado do processo de classificação é um classificador binário determinado para um limiar automaticamente determinado, gerando portanto, apenas um ponto na curva ROC.

Uma forma de efetuar a análise e comparação objetiva e quantitativa de desempenho de classificadores através da curva ROC para diferentes limiares é a utilização da Área Abaixo da Curva (AAC). Nessa análise, é calculada a área abaixo da curva ROC (por exemplo da Figura 54) e quanto mais próxima do valor unitário, melhor o desempenho do classificador, já que o valor unitário da área corresponde aos pontos mais próximos das extremidades superior e esquerda da curva, com máxima taxa de verdadeiros positivos e mínima taxa de falsos positivos.

#### 4.9.2 MATRIZ DE CONFUSÃO PARA MAIS DE DUAS CLASSES

A matriz de confusão apresentada na Tabela 2 pode ser estendida para problemas de classificação supervisionada envolvendo mais de duas classes. Um exemplo desse tipo de matriz está representado na Tabela 3, para um problema de quatro classes.

<sup>7</sup>Padrão-ouro, do inglês *ground-truth*, é o rótulo atribuído a um dado padrão por um especialista. Esse rótulo normalmente é utilizado como referência em um problema de aprendizado supervisionado.

**Tabela 3: Matriz de confusão para um problema de quatro classes.**

–	Valor Verdadeiro (Classe 1)	Valor Verdadeiro (Classe 2)	Valor Verdadeiro (Classe 3)	Valor Verdadeiro (Novidades)
Estimada Classe 1	Acertos Classe 1	–	–	–
Estimada Classe 2	–	Acertos Classe 2	–	–
Estimada Classe 3	–	–	Acertos Classe 3	–
Estimada Novidades	–	–	–	Acertos Novidades

Através da análise da diagonal principal pode-se observar o acerto do classificador para cada classe. Os demais elementos da matriz indicam erros de classificação (confusão) entre classes. O acerto para uma classe  $i$  pode ser definido como mostra a seguinte equação:

$$Acerto_{classe(i)} = \frac{M_{ii}}{\sum_{j=1}^K M_{ij}}, \quad (104)$$

sendo  $M_{ij}$  o elemento da matriz de confusão na linha  $i$  e coluna  $j$ . O acerto por classe será importante na análise dos classificadores multiclasse, bem como na análise do desempenho conjunto com o detector de novidades. Nesse último, pode-se incluir a novidade como uma classe na matriz de confusão, permitindo que a análise seja feita toda através da matriz de confusão.

#### 4.9.3 ANÁLISE DE DESEMPENHO DOS MÉTODOS DE AGRUPAMENTO

Para avaliar o desempenho de um método de agrupamento existem diferentes métricas, conhecidas como índices de validação de agrupamento (THEODORIDIS; KONSTANTINOS, 2009). Existem basicamente três abordagens distintas para essa avaliação. A primeira delas considera a informação externa sobre os padrões, como um dado comportamento esperado do método de agrupamento. É possível também, utilizar alguma informação do padrão-ouro dos dados para verificar a capacidade do agrupamento em confirmar uma dada estrutura pré-estabelecida. Esse conjunto de métricas é denominado de critério de validação externa.

Um segundo conjunto de métricas aborda a própria estrutura formada pelos grupos. Ou seja, utilizam-se informações sobre as características dos grupos, através de índices como a matriz de proximidade espacial, ou matriz de vizinhança, com o intuito de avaliar a similaridade entre os padrões de um mesmo grupo e a dissimilaridade entre padrões de grupos diferentes (THEODORIDIS; KONSTANTINOS, 2009). Essa medida é conhecida como critério de

validação interna.

O terceiro e último conjunto considera a comparação entre os grupos formados por um dado agrupamento com diferentes parâmetros ou diferentes métodos de agrupamento aplicados a um mesmo conjunto de dados. Desse modo, tem-se uma medida relativa do agrupamento, considerando que uma das possibilidades de agrupamento pode ser considerada como a referência. Esse conjunto de métricas é conhecido como critério de validação relativa.

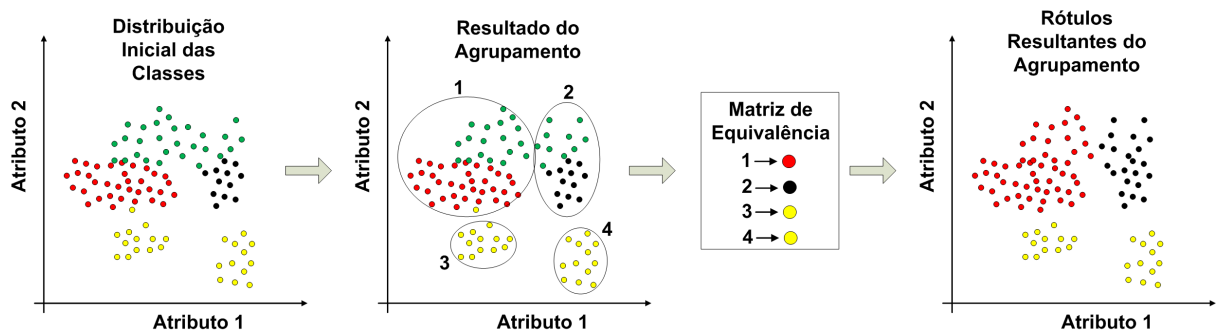
Dentre todas essas possibilidades, optou-se por utilizar neste trabalho uma métrica de validação externa baseada na matriz de confusão dos grupos. A primeira justificativa para essa utilização é que com essa métrica, é possível utilizar o padrão-ouro dos dados como referência externa, com o intuito de verificar a capacidade do método de agrupamento em colocar eventos de uma mesma classe em um mesmo grupo. Assim, espera-se que para um novo conjunto de dados, o método de agrupamento seja capaz de manter eventos de uma nova classe em um (ou eventualmente mais de um) grupo, indicando através do agrupamento a possibilidade de existência dessa nova classe.

A segunda justificativa é que a análise de desempenho através da matriz de confusão é utilizada para os métodos de classificação e detecção de novidades. Desse modo, a interpretação dos resultados fica facilitada, no caso dos métodos de agrupamento, uma vez que ela segue a mesma ideia apresentada para os modelos de classificação, utilizando como base de comparação de desempenho o acerto médio por classe.

É importante ressaltar que não estão sendo feitas restrições até o presente momento sobre o número máximo de grupos formados pelo agrupamento. Como essa questão está mais relacionada com o método de agrupamento automático do que com a métrica propriamente dita, optou-se por discutir essa questão posteriormente, na apresentação dos resultados.

Para demonstrar o funcionamento da métrica de validação externa baseada na matriz de confusão dos grupos, será apresentado um exemplo fictício ilustrando todas as etapas envolvidas até a determinação da matriz de confusão e, posteriormente, a determinação do acerto de classificação. Esse exemplo está resumido na Figura 55.

Inicialmente, tem-se uma distribuição de classes cujos rótulos do padrão-ouro estão representados pelas cores amarela, vermelha, preta e verde. Para essa distribuição de classes, é aplicado o algoritmo de agrupamento. Observa-se que o agrupamento gera quatro grupos, sendo que os grupos 1 e 2 contêm eventos de diferentes classes e os grupos 3 e 4, contêm apenas uma classe. Para realizar a associação entre o resultado do agrupamento e o rótulo atribuído das classes, é necessário verificar a matriz de equivalência entre os grupos definidos e os rótulos



**Figura 55: Exemplo de aplicação da métrica de validação externa baseada na matriz de confusão dos grupos.**

do padrão-ouro. Essa matriz de equivalência é construída através da determinação da classe majoritária em cada grupo. Ou seja, em cada grupo verifica-se qual a classe que contém mais padrões. A partir dessa definição, indica-se que o grupo receberá o rótulo da classe majoritária, realizando assim, o último estágio do processo, o qual contempla a atribuição dos rótulos pelo método de agrupamento.

A matriz de confusão resultante da Figura 55 está representada na Tabela 4. O acerto médio global, considerando as médias dos acertos por classe representado na Equação (104), é de 74%.

**Tabela 4: Matriz de confusão para o problema de agrupamento.**

–	Valor Verdadeiro (Vermelho)	Valor Verdadeiro (Verde)	Valor Verdadeiro (Preto)	Valor Verdadeiro (Amarelo)
Estimada Vermelho	34	18	0	1
Estimada Verde	0	0	0	0
Estimada Preto	0	10	12	0
Estimada Amarelo	0	0	0	24

É importante verificar que os rótulos do padrão-ouro são utilizados apenas como parte da avaliação do agrupamento, não sendo necessários para a definição dos grupos por parte do método de agrupamento.

#### 4.9.4 COMPARAÇÃO ESTATÍSTICA ENTRE DESEMPENHOS

Como grande parte dos desempenhos dos métodos de detecção e classificação avaliados neste trabalho são apresentados em termos de distribuições, optou-se por aplicar testes



estatísticos para realizar a comparação entre os métodos, de forma que seja possível avaliar se os resultados obtidos provêm de uma mesma distribuição, indicando se existe diferença de desempenho do ponto de vista estatístico.

Na literatura, existem diferentes métodos que podem ser aplicados para essa finalidade (NEHMZOW, 2006). A grande diferença entre eles é a característica das distribuições sob comparação. Se ambas as distribuições são Normais (ou Gaussianas), os testes são denominados paramétricos, pois consideram a comparação da média e do desvio padrão de cada distribuição. O teste mais utilizado para essa comparação é o *teste-t* (CASELLA; BERGER, 2010). Caso as distribuições não sejam Normais, comumente se utilizam modelos não-paramétricos, dentre os quais se destaca o *teste-U* ou teste de Wilcoxon.

Todos esses testes consideram que a hipótese nula é formulada considerando que duas distribuições distintas possuem a mesma média, no caso dos testes paramétricos, ou a mesma mediana no caso dos testes não-paramétricos, considerando o nível de significância utilizado. Sendo assim, se a hipótese nula é aceita para um dado nível de significância, as distribuições são consideradas equivalentes e, portanto, os desempenhos obtidos pelos métodos de detecção e classificação podem ser considerados iguais, do ponto de vista estatístico. Maiores detalhes sobre como opera o teste de Wilcoxon podem ser obtidas em Nehmzow (2006).

A maior parte dos resultados comparados neste trabalho possuem distribuições *não-Gaussianas*, muito embora em alguns casos as distribuições sejam próximas às distribuições Normais. Sendo assim, optou-se por utilizar o teste de Wilcoxon (NEHMZOW, 2006) com nível de significância de 5%. Um detalhe importante a ser ressaltado nesse contexto é que, mesmo as distribuições sendo *não-Gaussianas*, optou-se por apresentar os resultados em termos de valor médio e seu respectivo desvio padrão, uma vez que a análise desses parâmetros é mais intuitiva na comparação de resultados.

## 5 METODOLOGIA

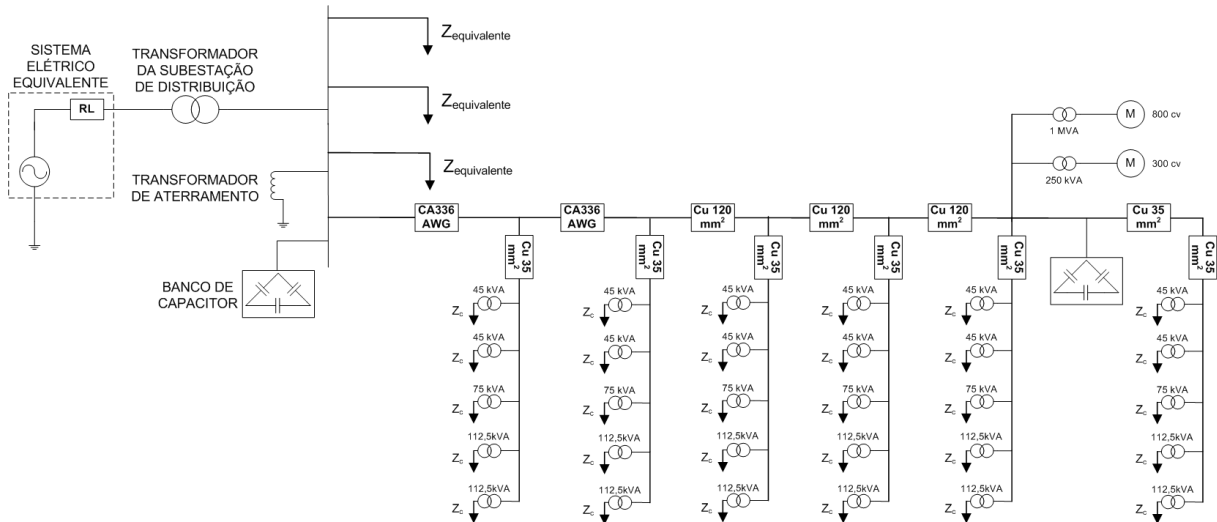
Para apresentar a metodologia empregada neste trabalho, optou-se por descrever inicialmente o sistema de simulação modelado no ATP e utilizado para gerar os eventos necessários para testar os métodos de segmentação e classificação. Em seguida, serão discutidos os eventos reais coletados em campo. Por último serão descritas as metodologias utilizadas para segmentação, classificação incluindo detecção de novidades e o agrupamento automático como parte do processo de identificação de novas classes.

### 5.1 SISTEMA SIMULADO

Com o intuito de testar e validar os modelos de segmentação e classificação, optou-se por realizar simulações digitais no ambiente do aplicativo ATP (EMTP, 1995). O ATP é um sistema amplamente utilizado para a simulação de transitórios eletromagnéticos, bem como eventos de natureza eletromecânica. Seu principal objetivo é fornecer subsídios para confecção de redes complexas e sistemas de controle de estrutura arbitrária em ambiente de simulação. Para tanto, o ATP faz uso de uma biblioteca que contém diversos elementos básicos de um SEP, como os elementos  $R$ ,  $L$  e  $C$  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, adaptou-se um padrão de rede de distribuição desenvolvido em (LAZZARETTI et al., 2009), com o objetivo de reproduzir uma subestação de distribuição típica, com seus principais componentes e alimentadores, visando obter as formas de onda dos eventos escolhidos no trabalho. O modelo foi construído no aplicativo *ATP Draw* que consiste em uma interface gráfica para o ATP. Nessa interface foram inseridos alguns dos principais componentes dessa subestação, a saber: equivalente da concessionária, transformador da subestação, transformador de aterramento (TAT), bancos de capacitores (BC), quatro alimentadores de uma barra, sendo que em três deles foi representado apenas o equivalente das

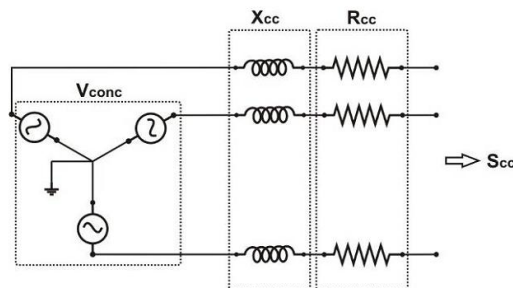
cargas e no quarto foi detalhado o tronco principal e os ramais principais. A Figura 56 mostra um diagrama esquemático contendo todos esses elementos representados. Na sequência, serão descritos os detalhes da modelagem de cada um desses elementos.



**Figura 56: Diagrama esquemático do modelo de simulação.**

### 5.1.1 MODELO EQUIVALENTE DA CONCESSIONÁRIA

A representação do equivalente elétrico para uma concessionária (equivalente do sistema elétrico anterior ao ponto que se objetiva simular) foi baseado 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 57.



**Figura 57: Esquema do equivalente elétrico da concessionária.**

Na Figura 57,  $V_{conc}$  representa a tensão da concessionária em Volts,  $S_{cc}$  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

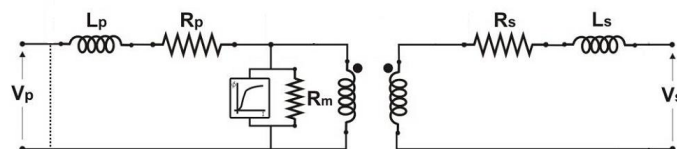
ideal (69 kV, tensão que chega à subestação) ligada em série com um modelo simétrico e acoplado de linha, cujos parâmetros fornecidos ao modelo são de resistência e reatância indutiva de sequência positiva e sequência zero. Esses dados possuem os seguintes valores:

- 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:  $6,6\Omega$ .

### 5.1.2 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 58, como mostra Toro (1999), sendo que, para o transformador utilizado, devem ser especificados os seguintes parâmetros:

- $V_p[V]$  e  $V_s[V]$  - Tensão nominal dos enrolamentos primário e secundário.
- $R_p[\Omega]$  e  $R_s[\Omega]$  - Resistência dos enrolamentos primário e secundário.
- $L_p[H]$  e  $L_s[H]$  - 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 magnético no joelho da curva de magnetização.
- $R_m[\Omega]$  - Resistência de magnetização (associada às Perdas em Vazio).
- Ligação do Transformador.



**Figura 58: Esquemático do transformador da subestação.**

Nesse modelo, em função da dificuldade encontrada para obtenção dos dados relacionados à saturação do transformador, não foram consideradas as variáveis associadas à curva

de magnetização do transformador (característica não-linear da modelagem). Sendo assim, os dados utilizados para a representação do transformador da subestação foram:

- Tensão nominal do enrolamento primário - 69kV entre fases.
- Tensão nominal do enrolamento secundário - 13,8kV entre fases.
- Resistência do enrolamento primário - 0,0014 $\Omega$ .
- Resistência do enrolamento secundário - 0,45 $\Omega$ .
- Indutância de dispersão do enrolamento primário - 0,6mH.
- Indutância de dispersão do enrolamento secundário - 11,1mH.
- Resistência de magnetização - 10<sup>15</sup> $\Omega$ .
- Ligação do Transformador - estrela no lado de alta e delta no lado de baixa tensão.

### 5.1.3 TRANSFORMADOR DE ATERRAMENTO

Um transformador de aterramento (TAT) tem por objetivo prover uma ligação ao 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 o 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 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.

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 escolhido, para um TAT que suporte 2940

Ampéres durante 10 segundos (TAT existente em grande parte das subestações de 13,8kV), deve-se utilizar os seguintes parâmetros para cada transformador monofásico:

- Tensão nominal do enrolamento primário - 13,8kV entre fases.
- Tensão nominal do enrolamento secundário - 13,8kV entre fases.
- Resistência do enrolamento primário - 0,5 $\Omega$ .
- Resistência do enrolamento secundário - 0,5 $\Omega$ .
- Indutância de dispersão do enrolamento primário - 3,95 $\Omega$ .
- Indutância de dispersão do enrolamento secundário - 3,95 $\Omega$ .
- Resistência de magnetização - 10<sup>11</sup> $\Omega$ .

#### 5.1.4 BANCOS DE CAPACITORES

No modelo de distribuição utilizado existem dois diferentes bancos de capacitores. O primeiro deles está instalado na subestação e o outro ao longo do alimentador detalhado na Figura 56. 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. As potências reativas dos bancos utilizados são:

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

#### 5.1.5 MOTORES DE INDUÇÃO

A partida de motores de indução de elevada potência foi escolhida para simular o impacto de uma das cargas mais utilizadas em processos industriais. Para tanto, optou-se por uma representação desses motores utilizando a transformada de Park, ou Transformada  $dq0$ , caracterizada por ser uma transformação linear, cujo objetivo é simplificar as equações de máquina nessa representação (CAD, 2000).

A transformada de Park representa o motor como uma máquina bifásica hipotética com elementos fixos (estator) e pseudoestacionários (rotor). Um dos eixos coincide com a linha central dos pólos magnéticos do rotor e é chamado de eixo direto (eixo  $d$ ). O outro se

situa ao longo do eixo neutro magnético e é chamado de eixo em quadratura (eixo  $q$ ). Nessa representação mantém-se a potência mecânica, o torque, a velocidade e o número de pólos. A modelagem desse dispositivo neste trabalho segue as recomendações apresentadas em Ferreira (2007). Com isso, foram obtidos os modelos de dois motores diferentes, de 300cv/220V e 800cv/4,16kV, cujas características hipotéticas<sup>1</sup> de placa estão representadas a seguir.

- Fator de Potência a plena carga: 0,87;
- Rendimento a plena carga: 93,5%;
- Rotação nominal a plena carga: 1780rpm;
- Relação entre corrente de partida e nominal: 8;
- Número de Pólos: 4;
- Momento de inércia do motor e carga:  $1\text{kg.m}^2$  para o motor de 300cv e  $4\text{kg.m}^2$  para o motor de 800cv;
- Frequência nominal: 60Hz;
- Perdas rotacionais: 3%;
- Porcentagem da velocidade inicial em relação à nominal: 0%;
- Porcentagem da carga que varia com o quadrado da velocidade: 100%;
- Potência nominal da carga: 300cv ou 800cv, conforme o motor.

#### 5.1.6 LINHAS AÉREAS

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, os transformadores de distribuição. Em cada trecho foi selecionado um tipo de cabo, tendo em vista a configuração típica 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.

---

<sup>1</sup>Esses modelos baseiam-se nos dados apresentados em Ferreira (2007), que foram adaptados para o presente trabalho, sem perder sua relação com motores de indução reais.

Para a representação desses cabos, foi utilizado o componente trifásico para linhas com parâmetros distribuídos e constantes (Modelo de Clarke) (OLIVEIRA et al., 2000). Esse 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<sup>2</sup> 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 (GREENWOOD, 1991) para ambas as sequências. Esse 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 zero 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.

A Tabela 5 mostra os cabos em função dos seus parâmetros.

**Tabela 5: Tabela de parâmetros de cabos.**

Cabo	$R/I+$ (Ohm/km)	$R/I0$ (Ohm/km)	$A+$ (mH/km)	$A0$ (mH/km)	$B+$ ( $\mu$ F/km)	$B0$ ( $\mu$ F/km)
CAA 336 AWG	0,19	0,36	1,03	4,92	0,01	0,004
Cu 120mm <sup>2</sup>	0,69	0,87	1,37	5,25	0,01	0,004
Cu 35mm <sup>2</sup>	1,05	1,22	1,38	5,27	0,01	0,004

### 5.1.7 TRANSFORMADORES DE DISTRIBUIÇÃO E CARGAS

O modelo dos transformadores de distribuição são baseados no modelo representado no esquema unifilar da Figura 58. No entanto, para os transformadores de distribuição foram incluídas as capacitâncias primárias e secundárias para o terra, bem como as capacitâncias entre



**Tabela 6: Tabela dos transformadores de distribuição.**

Dados - Transformador	45 kVA	75 kVA	112,5 kVA
Tensão nominal do enrolamento primário	13,8 kV	13,8 kV	13,8 kV
Tensão nominal do enrolamento secundário	220/127 V	220/127 V	220/127 V
Resistência do enrolamento primário (Ohms)	98,5	51,8	31,3
Resistência do enrolamento secundário (Ohms)	0,0091	0,0048	0,0029
Indutância de dispersão enrolamento primário (mH)	179,7	107,8	71,9
Indutância de dispersão enrolamento secundário (mH)	0,0998	0,06	0,04
Resistência de magnetização (Ohms)	920900	614000	460500
Ligação do Transformador	Delta Estrela	Delta Estrela	Delta Estrela
Capacitância Primário e Secundário [ $\mu$ F]	0,0003	0,0005	0,0007
Capacitância Primário e Terra [ $\mu$ F]	0,0001	0,0001	0,0002
Capacitância Secundário e Terra [ $\mu$ F]	0,0003	0,0003	0,0004

primário e secundário, conforme mostra Borghetti et al. (2009). A Tabela 6 resume os dados para os três modelos de transformador utilizados.

As cargas foram dispostas em modelos equivalentes distribuídos ao longo dos secundários dos transformadores de distribuição, com ligação em estrela. Essas cargas são compostas por elementos resistivos, indutivos e capacitivos, de forma que seus valores são ajustados de acordo com as potências ativa e reativa provenientes da curva de carga média do alimentador, que nesse caso corresponde a 3600kW e 600kVAR (valor médio de carga de um alimentador). Para os demais alimentadores, optou-se por incluir uma carga equivalente em delta, também correspondente a carga médias desses alimentadores: 4000kW/500kVAR, 4000kW/600kVAR e 2800kW/200kVAR.

<sup>2</sup>Modelo normalmente utilizado para representar o circuito equivalente por unidade de comprimento de uma linha de transmissão (GREENWOOD, 1991).

### 5.1.8 EVENTOS GERADOS PARA OS MÉTODOS DE SEGMENTAÇÃO

Com o intuito de se avaliar os métodos de segmentação, optou-se por desenvolver um conjunto de simulações para diferentes eventos isolados, sequenciais e até mesmo simultâneos. É importante que os métodos possam ser testados e validados inicialmente em um ambiente de simulação, pois nesse ambiente pode-se controlar o tipo de evento a ser simulado e suas características, o número de eventos por oscilografia e os instantes de ocorrência. Uma vez validados os métodos em um ambiente de simulação, pode-se estender a metodologia para os eventos registrados pelo sistema de monitoramento, tendo conhecimento prévio sobre os principais fatores de influência no desempenho de detecção.

Para geração dos eventos descritos a seguir, tanto para segmentação quanto para a classificação, foram inseridos elementos no modelo ATP que permitissem a simulação dos mesmos. No caso dos curtos-circuitos ao longo do alimentador, foi inserida uma chave na metade do comprimento total do alimentador, com resistências ligadas em série, permitindo a simulação de uma falta entre fases ou entre fases e o terra. Para os eventos de abertura e fechamento, uma chave foi inserida no mesmo ponto, permitindo as manobras no alimentador. Adicionalmente, é importante salientar que todos os eventos descritos foram gerados de forma automática, variando os parâmetros de interesse, para gerar um conjunto de simulações a partir de um único arquivo base no ATP. Para tanto, foi utilizada a proposta apresentada por Su et al. (2008).

Para o primeiro conjunto de simulações, foram utilizados os seguintes eventos: curto-circuito monofásico, curto-circuito bifásico para o terra, curto-circuito bifásico, curto-circuito trifásico, chaveamento de banco de capacitores, partida de motores de grande potência, abertura de alimentador e religamento de alimentador. Para cada evento foram variados os instantes de ocorrência, resistência de falta, local de medição, número de eventos por oscilografia e o intervalo entre a ocorrência de cada evento, no caso dos eventos subsequentes.

Com relação ao instante de ocorrência, foram selecionados dois pontos distintos na forma de onda para todos os eventos, sendo o primeiro na passagem por zero do sinal de tensão e o segundo no valor máximo positivo do sinal. Já para a resistência de falta, também foram selecionados dois valores: 5 ohms e 300 ohms. Para o local de medição optou-se pelo registro feito na barra de saída da subestação e um ponto exatamente na metade do comprimento do alimentador (depois da chave de manobra). Já quanto ao número de eventos por oscilografia foi adotada uma sequência normalmente verificada na prática. Por exemplo, no caso de um curto-circuito, esse pode ser seguido de uma abertura e um religamento do alimentador. Desse modo, a seguinte sequência foi adotada:

- Curtos-circuitos monofásicos podem ser: isolados, seguidos de uma abertura, seguidos de uma abertura e religamento, seguidos de curto-circuito bifásico para o terra e seguidos de curto-circuito trifásico para o terra;
- Curtos-circuitos bifásicos para terra podem ser: isolados, seguidos de uma abertura, seguidos de uma abertura e religamento e seguidos de curto-circuito trifásico para o terra;
- Curtos-circuitos bifásicos podem ser: isolados, seguidos de uma abertura e seguidos de uma segunda abertura;
- Chaveamentos de bancos de capacitor podem ser: isolados, seguidos de um curto-circuito monofásico, seguidos de curto-circuito bifásico para o terra e seguidos de curto-circuito trifásico para o terra.

Os intervalos de ocorrência entre eventos foram selecionados como 1/4 de ciclo e 1,5 ciclo, de modo a permitir avaliar a segmentação para eventos subsequentes muito próximos ou mesmo bastante separados temporalmente. Utilizando todas essas características, foram geradas 170 formas de onda trifásicas para análise. Para os métodos de detecção, optou-se por analisar apenas uma das fases, já que esse método deve ser aplicado individualmente nas três formas de onda do sistema trifásico. Para geração do padrão-ouro, optou-se por caracterizar os transitórios entre os estados quase-estacionários de cada oscilografia de forma visual, seguindo a metodologia apresentada inicialmente neste trabalho.

#### 5.1.9 EVENTOS GERADOS PARA CLASSIFICAÇÃO E IDENTIFICAÇÃO DE NOVAS CLASSES

Para se avaliar o desempenho da classificação e detecção de novidades, uma metodologia bastante similar foi adotada. Inicialmente foram definidas as classes que seriam abordadas para o problema. Para tanto, optou-se por trabalhar com eventos que ocorrem no próprio alimentador sob análise, além de eventos que ocorrem à montante da subestação. No total, são 29 classes listadas a seguir:

- Curto-circuito monofásico nas fases A, B e C;
- Curto-circuito bifásico para terra nas fases A/B, B/C e C/A;
- Curto-circuito bifásico nas fases nas fases A/B, B/C e C/A;
- Curto-circuito trifásico para terra;

- Curto-circuito trifásico;
- Abertura monofásica nas fases A, B e C;
- Abertura bifásica nas fases A/B, B/C e C/A;
- Abertura trifásica;
- Fechamento monofásico nas fases A, B e C;
- Fechamento bifásico nas fases A/B, B/C e C/A;
- Fechamento trifásico;
- Fechamento de banco de capacitor;
- Curto-circuito monofásico à montante da subestação;
- Curto-circuito bifásico para o terra à montante da subestação;
- Partida de um motor de elevada potência.

Para os eventos gerados de forma isolada na oscilografia (individualmente), foram variados os instantes de ocorrência, resistência de falta e o local de medição. Com relação ao instante de ocorrência foram selecionados, inicialmente, três pontos distintos na forma de onda para todos os eventos, sendo o primeiro na passagem por zero do sinal de tensão (referência na fase do evento), o segundo no valor máximo positivo do sinal e o terceiro um ponto intermediário entre os dois primeiros (ângulo de fase  $\pi/4$ ). Além desses valores, também foram considerados os instantes  $\pi/16$ ,  $\pi/8$ ,  $3\pi/16$ ,  $5\pi/16$ ,  $3\pi/8$  e  $7\pi/16$ . Para os eventos de manobra (abertura e fechamento) foram incluídos outros quatro instantes de ocorrência do evento nos ângulos de fase  $3\pi/4$ ,  $3\pi/2$ ,  $9\pi/8$  e  $7\pi/4$ .

Já para a resistência de falta, foram selecionados os valores:  $0,5\Omega$ ,  $1\Omega$ ,  $2\Omega$ ,  $5\Omega$ ,  $12\Omega$ ,  $22\Omega$ ,  $80\Omega$ ,  $150\Omega$  e  $300\Omega$ . Para o local de medição optou-se pelo registro feito na barra de saída da subestação e um ponto exatamente na metade do comprimento do alimentador (depois da chave de manobra). Com todos esses parâmetros variados, obteve-se um total de 18 eventos por classe para os curtos-circuitos e 10 eventos por classe para manobras. Do total de 621 eventos, 414 foram utilizados para treinamento dos modelos e o restante para a fase de testes.

Além disso, também foram incluídos eventos simultâneos e subsequentes com intervalo de  $1/4$  de ciclo entre a ocorrência de cada evento. Para geração dos eventos subsequentes, as seguintes sequências foram adotadas:

- Fechamento (monofásico ou bifásico) seguido de fechamento (bifásico ou trifásico);
- Curto-circuito (monofásico, bifásico ou trifásico) seguido de abertura (monofásica, bifásica ou trifásica);
- Abertura (monofásica ou bifásica) seguida de abertura (bifásica ou trifásica);
- Curto-circuito (monofásico ou bifásico) seguido de curto-circuito (bifásico ou trifásico);
- Chaveamento em banco de capacitor seguido de curto-circuito (monofásico, bifásico ou trifásico).

Já para os eventos simultâneos, adotaram-se as seguintes ocorrências:

- Chaveamento de banco de capacitor simultâneo a um curto-circuito (monofásico, bifásico ou trifásico);
- Religamento (monofásico, bifásico ou trifásico) simultâneo a um curto-circuito (monofásico, bifásico ou trifásico).

Os parâmetros variados foram os mesmos adotados para os eventos isolados. No entanto, as resistências de falta escolhidas foram  $0,1\Omega$ ,  $10\Omega$  e  $200\Omega$ . No total, foram obtidos 128 eventos subsequentes e 60 simultâneos.

É importante ressaltar que os eventos subsequentes e simultâneos aqui apresentados irão configurar apenas classes definidas como novidades. Ou seja, não será adotada nenhuma metodologia de classificação multirrótulo para esses eventos, uma vez que eventos subsequentes inferiores a um ciclo também poderiam ser classificados como simultâneos no contexto deste trabalho, já que a extração de características opera em um ciclo após a ocorrência do evento.

## 5.2 SISTEMA DE MONITORAMENTO (DADOS REAIS)

O sistema de monitoramento utilizado neste trabalho é resultado de um projeto de Pesquisa e Desenvolvimento proposto pela Copel Distribuição e executado pelos Institutos Lactec, intitulado *Instalação Piloto para Avaliação dos Níveis de Sobretensão Atmosférica no Sistema da COPEL Distribuição*. Os registros armazenados forneceram a possibilidade de teste dos métodos de detecção e classificação de eventos. Esse sistema permite o registro simultâneo de transitórios decorrentes de sobretensões atmosféricas e demais eventos, em todas as fases da rede de média tensão e em todas as fases e no neutro da rede secundária.

Por ser voltado para registro de eventos de diferentes naturezas, os oscilógrafos digitais possuem velocidade e resolução adequadas para o registro simultâneo de pelo menos sete sinais de sobretensão na rede de distribuição, com capacidade de memória suficiente para o registro completo desses eventos, assim como para o armazenamento de eventos sucessivos por um período de pelo menos 30 dias. Além disso, permitem a transferência dos registros armazenados para um microcomputador ou permitem a análise dos dados coletados remotamente através de uma rede de computadores. Para tanto, foram utilizados oscilógrafos digitais da Yokogawa, modelo SL-1000, com 16 canais de entrada, taxa de amostragem de até 100 MS/s por canal e banda de medição de 100 MHz.

Estimando-se sobretensões atmosféricas da ordem de até 200 kV na rede de média tensão e de até 30 kV na rede da baixa tensão, foi necessária a aquisição e a construção de divisores de tensão capacitivos especiais, para operação contínua em redes de distribuição energizada de classe 15 kV ou 36,2 kV, com banda de passagem plana de frequência industrial até aproximadamente 5 MHz. Para tanto, foram utilizados dois tipos de sistemas de monitoramento similares, a saber:

- Divisor capacitivo da Ross Engineering, com tensão máxima de 200 kV (dois sistemas trifásicos);
- Divisor capacitivo com unidade primária da Maxwell e unidade secundária fabricada pelo LACTEC, com tensão máxima de 350 kV (2 sistemas trifásicos).

Analogamente, para a rede secundária, foram utilizados:

- Divisor capacitivo da Ross Engineering, com tensão máxima de 60 kV (um sistema trifásico);
- Ponta de prova de alta tensão da Tektronix, com tensão máxima de 40 kV (três sistemas trifásicos).

Visando garantir a operação dos sistemas de monitoramento, foi prevista a instalação de *no-break*, com autonomia de pelo menos uma hora. Além disso, foi prevista a instalação de modem GPRS (do inglês, *General Packet Radio Service*), com antena externa, para possibilitar o acesso remoto aos registros de eventos armazenados. Em complemento, foi prevista a utilização de transformador isolador, com blindagem eletrostática, para proteger o circuito de alimentação dos equipamentos empregados contra surtos conduzidos, garantindo isolamento de 20 kV. A Figura 59 mostra detalhe do sistema de monitoramento desenvolvido.



**Figura 59: Detalhe do sistema de monitoramento.**

Os divisores de tensão construídos pelo LACTEC foram calibrados até o nível máximo de 350 kV. Além disso, verificou-se sua linearidade, nas duas polaridades, com desempenho satisfatório. O valor médio da banda de passagem (3dB) dos divisores de tensão construídos pelo LACTEC é 1,65 MHz. Mesmo que a banda de passagem dos divisores de tensão construídos pelo LACTEC seja inferior aos 5 MHz originalmente propostos, tais dispositivos ainda são perfeitamente aceitáveis para a medição e o registro de eventos com elevado conteúdo de frequência, como os surtos provenientes de descargas atmosféricas na rede de distribuição. O mesmo procedimento foi aplicado aos outros divisores de tensão, para validar os sistemas propostos para o monitoramento de sobretensões devido a descargas atmosféricas.

Para instalação dos quatro sistemas de monitoramento, foram selecionados quatro diferentes alimentadores, sendo dois em São Mateus do Sul/PR e outros dois em Apucarana/PR. A escolha desses locais está associada à principal finalidade deste projeto, relacionada ao monitoramento do efeito das descargas atmosféricas em redes de distribuição. Nesse sentido, esses locais apresentavam uma série de vantagens, uma vez que possuem índice cerâmico<sup>3</sup> elevado, sendo 5,82 raios/km<sup>2</sup>/ano em São Mateus do Sul e 5,17 raios/km<sup>2</sup>/ano em Apucarana.

Depois de efetuado um período de monitoramento de aproximadamente um ano, os registros foram preparados para as análises aqui propostas. Para obter o padrão-ouro dos instantes de transição entre eventos e a classificação correspondente, todos os sinais registrados pelos quatro sistemas de monitoramento foram segmentados em até cinco regiões principais, que correspondem às variações mais significativas observadas nas formas de onda, guardando as informações de localização e duração de cada transitório (formação do padrão-ouro). Para cada segmento, fez-se sua caracterização em termos do evento causador.

Para essa classificação, foram obtidas informações das descargas atmosféricas na base

<sup>3</sup>Este índice indica o número de descargas atmosféricas em uma determinada região por ano (UMAN, 2001).

de dados da Rede Integrada Nacional de Detecção de Descargas Atmosféricas (RINDAT) e do Sistema de Ocorrências da Copel Distribuição, além de simular digitalmente os principais trechos de cada alimentador no aplicativo *ATPDraw* e realizar a análise do conteúdo de frequência presente nos sinais. Com essa abordagem, foi possível separar os eventos nas seguintes classes principais: descarga atmosférica, curto-circuito monofásico e religamento trifásico.

Vale ressaltar que os eventos podem ocorrer de forma sequencial, individual e até mesmo simultânea. Uma vez ocorrido o evento, este pode ser seguido de um retorno à normalidade. Por exemplo, um curto-circuito pode durar alguns ciclos e ser extinto pela atuação do sistema de proteção. Adicionalmente, foram observados eventos que apresentam característica bastante complexa na análise da forma de onda, fazendo com que os dados utilizados para sua classificação ainda não sejam suficientes para uma caracterização criteriosa.

Uma vez estabelecidos os critérios para classificação dos eventos registrados pelos sistemas de monitoramento, foi determinada a quantidade de registros para cada classe, através da análise da causa predominante de ocorrência de cada evento, conforme mostrado na Tabela 7. Nesta tabela, demais eventos correspondem aos eventos de retorno à normalidade, curtos-circuitos bifásicos, eventos simultâneos ou subsequentes com intervalo inferior a um ciclo e eventos que não possuem uma classificação precisa, conforme mencionado anteriormente, caracterizando a classe das novidades utilizada na fase de teste dos classificadores.

**Tabela 7: Eventos registrados pelo sistema de monitoramento.**

Classe	Apucarana-1	Apucarana-2	São Mateus-1	São Mateus-2
Descarga Atmosférica	45	1	24	15
Curto-Circuito Monofásico	18	0	2	6
Religamento Trifásico	1	0	8	3
Demais Eventos	18	1	5	2
Total	82	2	39	26

Dos 149 eventos registrados, com variações significativas nas formas de onda, observa-se que 57% deles estão relacionados a descargas atmosféricas, 17% a eventos de curto-circuito, 8% a religamentos trifásicos e 18% a outros eventos. Assim, durante o período em que os quatro sistemas de monitoramento foram avaliados na rede de distribuição, observou-se predominância de sobretensões de origem atmosférica, mesmo quando se considera cada sistema de monitoramento individualmente.



### 5.3 SEGMENTAÇÃO DE EVENTOS

A metodologia adotada para realizar a segmentação dos sinais está associada diretamente com o método escolhido. No caso dos métodos de estimação espectral, modelos autor-regressivos e modelos baseados em filtro de Kalman, o processo de detecção é bastante similar, podendo ser detalhados de forma conjunta. Já os métodos baseados em Representação de Dados com Vetores Suporte e Operador de Energia de Teager possuem características particulares, que devem ser tratadas isoladamente.

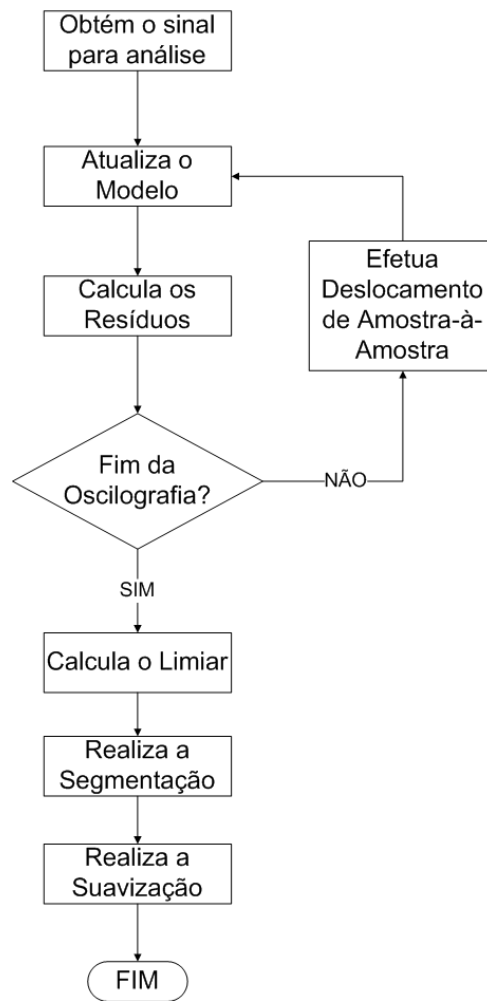
#### 5.3.1 MODELOS DE ESTIMAÇÃO ESPECTRAL

Para esses modelos, inicia-se definindo os parâmetros a serem utilizados na modelagem, como a frequência de amostragem e a ordem do modelo a ser estimado, visando realizar a estimação em todo o sinal. A frequência de amostragem utilizada, tanto para os eventos simulados quanto para os eventos monitorados, foi de 983.040 Hz, o que corresponde a 16.384 amostras/ciclo. No caso particular dos eventos monitorados, foram realizados testes com diferentes valores de frequência de amostragem, pois como será detalhado no próximo capítulo, valores elevados de frequência de amostragem elevam de forma significativa o tempo de processamento dedicado à segmentação e apresentam resultados inferiores no processo como um todo.

Com relação à ordem do modelo estimado, foram consideradas inicialmente as recomendações propostas por Bollen e Gu (2006). Naquele trabalho, os modelos são calculados utilizando os vinte e cinco primeiros harmônicos do sinal. No entanto, observou-se que para alguns eventos, esse valor leva ao surgimento de resíduos mais proeminentes não associados a transitórios, já que o conteúdo harmônico verificado em grande parte dos registros possui uma característica fortemente não-estacionária, dificultando o processo de segmentação para os transitórios de interesse. Sendo assim, optou-se por estimar os modelos utilizando apenas os vinte primeiros harmônicos para os modelos com FK e os dez primeiros harmônicos para os modelos AR. Os detalhes dessa escolha serão apresentados no Capítulo 6.

A visão geral do processo de detecção baseado em FK e modelos AR está resumido no fluxograma da Figura 60.

Depois de percorrido todo o sinal, é possível, através dos processos de limiarização e suavização, obter uma segmentação automática da forma de onda. Com relação à suavização, existem alguns aspectos que devem ser detalhados. Conforme foi observado na Figura 25, existem algumas amostras que possuem resíduo inferior ao limiar, porém estão localizadas entre



**Figura 60: Metodologia de segmentação com base na estimação espectral.**

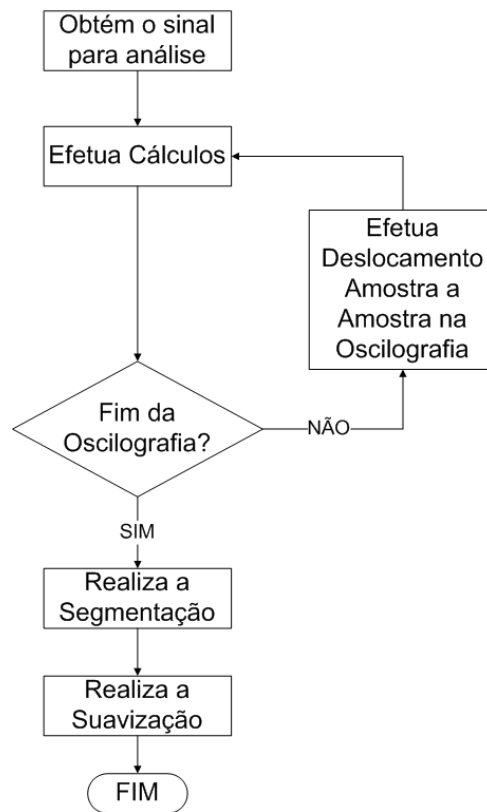
amostras cujo resíduo é superior ao limiar. Isso implica que, apesar do sinal estimado estar bastante próximo ao sinal original para essas amostras, é necessário classificá-las como pertencentes ao transitório que se deseja segmentar, pois as suas vizinhanças anteriores e posteriores indicam que o transitório ainda está ocorrendo.

Sendo assim, o processo de suavização consiste em verificar se para uma dada amostra, cujo limiar é inferior ao calculado, existe algum vizinho posterior cujo resíduo supera o limiar. Essa busca é feita para as amostras posteriores em até 1/8 de ciclo. Esse processo pode ser entendido como a eliminação da histerese existente quando se efetua a comparação do limiar com o resíduo calculado para cada amostra, conforme recomendam Ukil e Zivanovic (2006a).

### 5.3.2 MODELOS BASEADOS NO OPERADOR DE ENERGIA DE TEAGER

A metodologia empregada para os modelos baseados no Operador de Energia de Teager é similar à metodologia utilizada na Subseção 5.3.1. A diferença nesse caso é que não é

necessário especificar a ordem do modelo, pois o método é aplicado amostra a amostra, para todo o sinal, sendo utilizado para cada passo do algoritmo apenas as amostras imediatamente anterior e posterior. Além disso, a determinação do limiar é feita conforme apresentado na Seção 4.3.1. Os demais estágios (segmentação e suavização) seguem a mesma ideia dos modelos de estimação espectral. Um resumo desse método está apresentado na Figura 61.

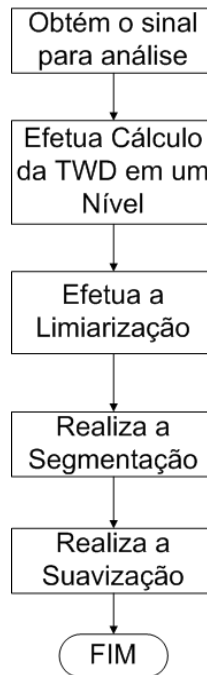


**Figura 61: Metodologia de segmentação com o Operador de Energia de Teager.**

### 5.3.3 MODELOS BASEADOS APENAS NA TRANSFORMADA *WAVELET*

Para realizar a segmentação através da TWD, optou-se por seguir a metodologia proposta em Silva et al. (2006) e Ukil e Zivanovic (2008). Essa metodologia consiste em calcular a TWD em um nível de decomposição e, a partir dos coeficientes do sinal detalhe no domínio *wavelet*, realizar a limiarização e a suavização utilizando a proposta do limiar universal (DONOHO; JOHNSTONE, 1994). A Figura 62 mostra um resumo dessa metodologia.

Com relação aos parâmetros necessários para esse modelo, tais como frequência de amostragem do sinal e a *wavelet*-mãe, optou-se por utilizar os mesmos valores selecionados para o modelo TWD+RDVS, apresentados na próxima seção.



**Figura 62: Metodologia de segmentação com base na TWD.**

#### 5.3.4 MODELOS BASEADOS EM TWD+RDVS

No caso dos modelos baseados em Representação de Dados com Vetores Suporte e TWD, além da frequência de amostragem, as janelas de análise (subsequentes) e o deslocamento entre as janelas, é necessário definir os parâmetros associados ao próprio modelo de representação<sup>4</sup> ( $\sigma$  e  $C$ ), o número de níveis da TWD que será aplicado e a *wavelet*-mãe a ser utilizada na decomposição do sinal. Maiores detalhes desse método podem ser encontrados nas subseções 4.8.5, 4.8.6 e 4.8.7.

A *wavelet*-mãe selecionada foi a Daubechies-20, pois conforme demonstra Delmont (2007), as características dessa função são as mais adequadas para os problemas de detecção. Já o número de níveis selecionados na decomposição está intimamente ligado à janela de análise. Caso a janela selecionada para a análise fosse de um ciclo para os modelos baseados em Representação de Dados com Vetores Suporte, não seria possível verificar diferentes transições em um intervalo menor que dois ciclos, já que para realizar essa representação são necessárias duas janelas subsequentes e qualquer transição existente nesse intervalo seria considerada como um único evento, pois a distância entre as hiperesferas no espaço de alta dimensionalidade seria proporcional aos transientes ora contidos no primeiro ciclo, ora contidos no segundo ciclo da janela. Por outro lado, janelas muito pequenas permitem utilizar apenas os primeiros ciclos

<sup>4</sup>Serão apresentados também os resultados para a segmentação, realizando uma escolha automática desses parâmetros, conforme discutido na introdução teórica deste trabalho.

da decomposição dentro desse critério, o que pode comprometer a detecção dos eventos de diferentes conteúdos de frequências transitórias.

Utilizando esse critério, procurou-se selecionar a menor janela de representação que pudesse conter informações suficientes para a detecção dos eventos, de forma que pudessem ser verificados diferentes eventos em um mesmo ciclo de 60 Hz. Desse modo, obteve-se que janelas de 1/8 de ciclo, para as duas janelas de observação, permitiriam a detecção de diferentes eventos com resolução de 1/4 de ciclo, resultando em uma decomposição em 10 níveis para sinais com frequência de amostragem de 983.040 Hz, no caso dos dados simulados. Desse modo, o vetor a ser representado nesse método possui 10 atributos e comprimento de 2048 amostras.

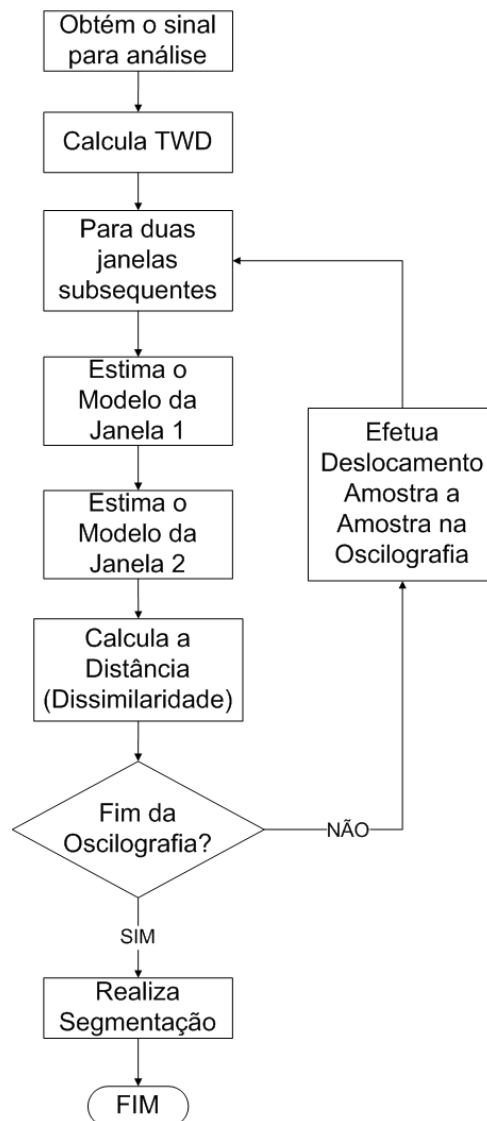
Observou-se que nessa representação para os dados reais, o tempo computacional gasto para realizar toda a segmentação da oscilografia era excessivo (essa questão será retomada no próximo capítulo), dada a grande quantidade de amostras a serem mapeadas e utilizadas na definição da hiperesfera, além de apresentar resultados inferiores para frequências de amostragem mais baixas. Sendo assim, optou-se por realizar testes reduzindo a frequência de amostragem para 7.680 Hz, conforme será exposto no Capítulo 6.

A escolha dos parâmetros ( $\sigma$  e  $C$ ) também será discutida no próximo capítulo. Depois de definidos os parâmetros, são calculadas as distâncias para realização da segmentação, conforme mostra a Figura 63. Vale ressaltar que o processo de limiarização automática é feito de forma similar ao OET e a suavização para o modelo baseado na RDVS não é necessário, uma vez que a própria saída do modelo já é suavizada.

### 5.3.5 ANÁLISE ESTATÍSTICA PARA SEGMENTAÇÃO DE EVENTOS

Uma vez definidos os parâmetros e o processo de segmentação para cada modelo, é feita a análise ROC dos resultados, tendo por base o padrão-ouro estabelecido para cada evento. Para tanto, efetua-se inicialmente a análise do mapeamento resultante do modelo (resíduos ou distâncias), de forma a estabelecer o comportamento da estimação, independentemente do processo de limiarização e segmentação automática. Nesse sentido, é possível observar o comportamento do mapeamento calculado para diferentes valores de limiares, avaliando a qualidade da resposta da própria estimação de cada modelo, gerando diferentes pontos na curva ROC, de acordo com cada limiar. A Figura 64 mostra um resumo desse processo.

Cada ponto nessa figura está associado a um dado limiar para a segmentação. Com isso, é possível verificar o comportamento do mapeamento resultante de cada método para diferentes limiares de segmentação. Nesses diferentes limiares, poderão ser observados pontos

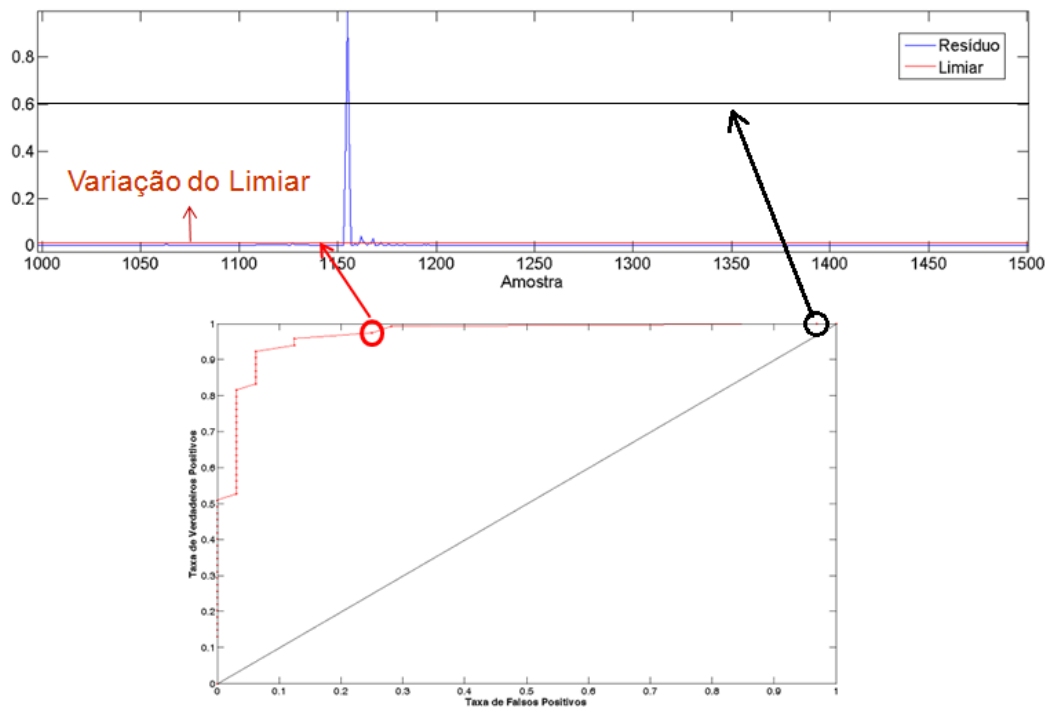


**Figura 63: Metodologia de segmentação com base em RDVS.**

na curva ROC com alta taxa de verdadeiros positivos e baixa taxa de falsos positivos.

Isso indica que podem existir limiares que fornecem uma segmentação mais próxima do padrão-ouro estabelecido. Essa análise é importante para um processo completamente automático, pois se espera que o limiar determinado automaticamente se aproxime do desempenho obtido para os limiares do lado superior esquerdo da curva ROC em questão, uma vez que a resposta gerada pela segmentação com base no limiar automaticamente calculado representa apenas um ponto na curva ROC representada.

Uma possível forma de avaliar o desempenho de uma segmentação baseada no cálculo automático do limiar, é utilizar o acerto de classificação discutido na Subseção 4.9.2. O problema de segmentação pode ser visto como um problema de classificação binária, na qual cada instante de tempo pode ser classificado como pertencente ao estado transitório ou estado quase-



**Figura 64: Detalhe da análise ROC para diferentes limiares.**

estacionário do sinal. Utilizando essa ideia, é possível construir a matriz de confusão do processo de segmentação, calculando o acerto por classe e o acerto médio global, dado pela média de desempenho de cada classe. Com isso, é possível saber o acerto médio do processo de segmentação.

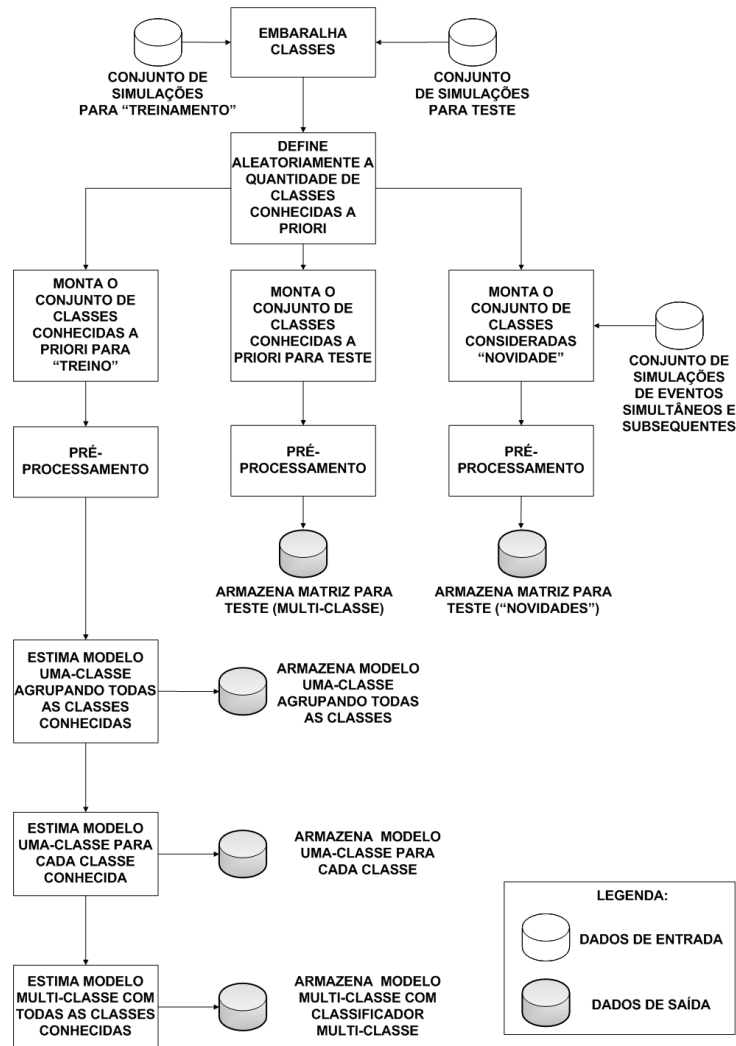
Uma terceira possibilidade é apresentar o erro médio dos instantes de tempo classificados como transitórios. Ou seja, é possível identificar qual o erro de detecção do instante inicial e final de um dado transitório, assinalando a diferença temporal existente entre a segmentação proposta e os instantes transitórios do padrão-ouro (considerados verdadeiros). Vale ressaltar que esse erro pode ser em função de uma detecção antecipada ou posterior aos instantes de referência do padrão-ouro.

#### 5.4 CLASSIFICAÇÃO DE EVENTOS E DETECÇÃO DE NOVIDADES

Para classificação de eventos e detecção de novidades foram avaliadas as abordagens apresentadas anteriormente, utilizando classificadores uma-classe e multiclasse, tanto para os dados reais quanto para os dados simulados. Foram adotados os dois cenários definidos no início do trabalho: DNG-CMC e DNI-CBS. A seguir, serão detalhadas as metodologias aplicadas para o estágio de treinamento e de teste (avaliação) de cada modelo para as bases de dados reais e simulados.

### 5.4.1 DADOS SIMULADOS

A metodologia aplicada para o treinamento dos modelos de classificação e detecção de novidades, no caso dos dados simulados, está resumida na Figura 65.



**Figura 65: Metodologia de treinamento dos modelos aplicada aos dados simulados.**

Inicialmente, os dados das 29 classes de eventos gerados por simulação são embaralhados de forma que possam ser definidas as classes conhecidas *a priori* e as classes que serão caracterizadas como novidades. Nesse processo, são escolhidas de forma aleatória a quantidade e quais classes formarão a base do aprendizado supervisionado da metodologia, sendo que o número mínimo de classes conhecidas para este trabalho foi definido como três<sup>5</sup>.

Na realidade, do grupo de classes conhecidas *a priori* são extraídos dois conjuntos, sendo que o primeiro deles, denominado conjunto de treinamento, é o conjunto efetivamente

<sup>5</sup>Com esse número mínimo de classes, é possível obter uma boa avaliação do processo de classificação multi-classe realizado, sendo esse o principal motivo dessa escolha.



utilizado para estimação dos modelos, utilizando como base os 414 eventos gerados para essa finalidade via simulação. O outro conjunto, denominado conjunto de teste, é utilizado no estágio de avaliação desses modelos, tanto para detecção de novidades quanto para a classificação multiclasse, utilizando os demais 207 eventos restantes. É importante ressaltar que as classes definidas para o conjunto de treinamento devem ser as mesmas para o conjunto de teste, durante o processo de embaralhamento de classes.

O terceiro grupo de classes consiste nas classes não conhecidas *a priori*, ou seja, as novidades. Esse conjunto é formado a partir das classes não selecionadas para o aprendizado supervisionado, além do próprio conjunto de eventos subsequentes e simultâneos gerados conforme descrito anteriormente.

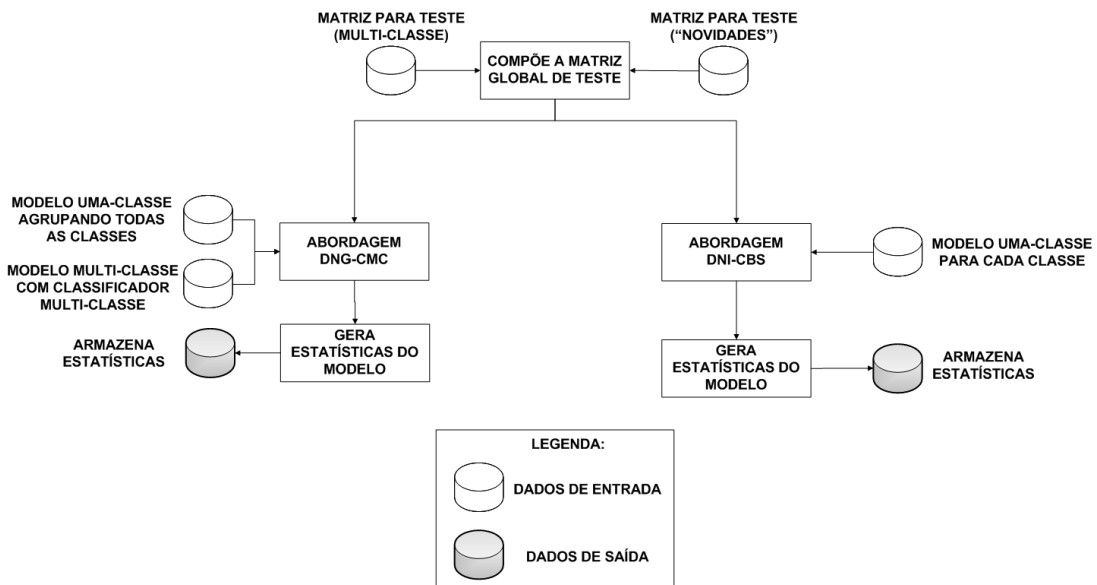
Depois de definidos os conjuntos, todos os eventos selecionados passam por um estágio de pré-processamento. Nesse estágio, cada forma de onda de cada fase é decomposta utilizando a TWD ou a TWPD e, em seguida, são extraídas as treze características do ciclo em que ocorre o evento no caso dos eventos isolados e simultâneos, ou do primeiro evento no caso dos eventos subsequentes. Uma vez realizados esses cálculos, as matrizes correspondentes ao conjunto de teste e novidade são armazenados para serem utilizados no estágio de avaliação dos métodos.

Já o conjunto de dados de treinamento é utilizado para construção de todos os modelos a serem avaliados, tais como: um modelo uma-classe agrupando todas as classes conhecidas, um modelo uma-classe para cada uma dessas classes e o próprio classificador multiclasse que irá operar sobre os eventos conhecidos, visando à definição específica das classes. Depois de estimados todos os modelos, estes são armazenados para o estágio de avaliação da metodologia, a qual contempla as abordagens DNG-CMC e DNI-CBS.

Convém ressaltar que foram avaliados os seguintes classificadores uma-classe: *K*-Vizinhos-Mais-Próximos, *X*-Médias e a Representação de Dados Utilizando Vetores Suporte. No caso do classificador multiclasse convencional foi utilizado o método baseado em Máquinas de Vetor Suporte, anteriormente descrito. É importante ressaltar que o método MVS é normalmente aplicado para problemas de duas classes. Para aplicação em um cenário multiclasse foi utilizada uma codificação *um-contra-um* no treinamento dos modelos (SCHÖLKOPF; SMOLA, 2001). Nessa codificação, é levantado um modelo para cada par das  $N$  classes do problema. Esse processo é repetido para todas as combinações entre pares de classes, resultando em  $N(N - 1)/2$  modelos, responsáveis por identificar cada classe. Na fase de avaliação, o padrão é avaliado em todos os  $N(N - 1)/2$  modelos, sendo definido pela classe com maior número de atribuições (votos) no fim do processo.

Depois de estimados e armazenados os modelos e as matrizes de teste, pode-se efetuar

a avaliação global do método proposto. Essa avaliação considera as duas diferentes abordagens de classificação utilizadas: DNG-CMC e DNI-CBS, conforme resumido na Figura 66. Para cada uma dessas abordagens, utiliza-se a composição da matriz de teste (multiclasse) e a matriz com os eventos caracterizados como novidades. Essa matriz final é denominada de matriz global de teste. Nela, estão todas as características extraídas no pré-processamento, além da própria classe  $K$  associada com cada um dos padrões (rótulo). A classe *novidade* é rotulada com o índice  $K + 1$ .



**Figura 66: Metodologia de teste dos modelos aplicada aos dados simulados.**

Com relação à abordagem DNG-CMC, o processo de detecção de novidades e classificação multiclasse são feitos de forma independente. Inicialmente, é verificado se um dado padrão pertence ou não ao conjunto de classes previamente conhecidas através da detecção de novidades realizada pelo classificador uma-classe (parametrizado através do agrupamento de todas as classes). Em seguida, caso o padrão seja definido como conhecido, aplica-se uma classificação multiclasse para verificar a qual classe esse padrão está associado. Nessa abordagem, a classificação é segmentada em dois estágios: detecção de novidades e classificação multiclasse.

A segunda abordagem adotada para classificação é DNI-CBS. Nesse cenário, aplica-se um classificador uma-classe para cada uma das classes conhecidas e, com isso, é possível realizar toda a classificação de forma acoplada, ou seja, detecção de novidades e classificação multiclasse de forma conjunta. Inicialmente, é verificado se um dado padrão pertence a um dos modelos uma-classe. Caso esse padrão pertença a apenas um dos modelos definidos, é atribuído ao padrão, a classe correspondente ao modelo definido; caso contrário, o padrão é definido como novidade. Uma terceira possibilidade contempla o caso do padrão estar associado a mais de um

modelo. Nesse cenário, a classificação multiclasse é feita com base na medida de similaridade entre os modelos candidatos e o padrão sob análise, atribuindo ao padrão, o modelo com maior similaridade.

Com relação às estatísticas calculadas, foram selecionados três índices de acerto de classificação para avaliar o desempenho de cada um dos estágios do processo (detecção de novidades e classificação multiclasse) e o desempenho global. O primeiro deles diz respeito ao desempenho médio de cada uma das classes no classificador multiclasse. Para tanto, foi utilizado apenas o conjunto de dados de teste (multiclasse). Desse modo, foi possível mensurar o desempenho médio para cada classe sem considerar a influência das novidades. Para realizar esse cálculo, foi calculada a média entre o acerto por classe dado pela Equação (104) na página 117, para todas as  $K$  classes definidas como conhecidas *a priori*, considerando que a classe  $K + 1$  não está definida nessa modelagem.

O segundo índice já considera a influência dos padrões caracterizados como novidades. Essa estatística também considera o acerto médio entre classes, porém são consideradas apenas as classes normal e novidade. Ou seja, não é considerada a classe associada a um dado padrão, no contexto de classificação multiclasse, e sim, o desempenho médio entre as classes normal e novidade, dadas pelas suas respectivas Equações (105) e (106):

$$Acerto_{normal} = \frac{N_{normal}}{N_{total}}, \quad (105)$$

$$Acerto_{novidade} = \frac{N_{novidade}}{N_{total}}, \quad (106)$$

sendo  $N_{normal}$  o número de exemplos classificados como uma das classes conhecidas e  $N_{novidade}$  o número de exemplos classificados como novidade. O último índice considera o desempenho global da abordagem apresentada. Para tanto, pode-se construir uma matriz de confusão com as  $K$  classes conhecidas *a priori*, incluindo a classe novidade como a classe  $K + 1$  dessa matriz. Com isso, pode-se calcular o desempenho médio para todas as classes, incluindo as novidades, através do desempenho individual dado pela Equação (104). Esse desempenho reflete a capacidade do método em identificar novidades, bem como em realizar a classificação multiclasse de forma conjunta e, por isso, é denominado desempenho global.

#### 5.4.2 PARÂMETROS AJUSTÁVEIS DOS DADOS SIMULADOS

Com o intuito de mapear todos os parâmetros ajustáveis desde o pré-processamento até a construção dos modelos e verificar sua influência no desempenho global dos métodos, optou-se por descrever aqui, os parâmetros envolvidos em todos os estágios do processo indivi-

dualmente, como demonstrado a seguir. A influência será detalhada nos resultados apresentados no Capítulo 6.

- Pré-Processamento:

Frequência de amostragem: foram utilizadas duas frequências de amostragem para os dados simulados: 7.680 Hz e 15.360 Hz. Com esses valores é possível cobrir toda a faixa de frequência associada à análise automática dos eventos simulados (LAZZARETTI et al., 2009).

Transformada *wavelet*: foram comparados os desempenhos da TWD e da TWPD para o pré-processamento como um todo. Para os cálculos da TWD e da TWPD é necessário definir o número de níveis da decomposição e a *wavelet*-mãe utilizada. No caso da *wavelet*-mãe optou-se por testar apenas dois bancos de filtros da família Daubechies (DAUBECHIES, 1992): *db4* e *db20*. Esses filtros apresentaram os melhores desempenhos na classificação de eventos apresentada em Lazzaretti et al. (2009). Com relação ao número de níveis, optou-se por manter a frequência fundamental (60 Hz) no sinal aproximação para a TWD. Com isso, são necessários seis níveis de decomposição para uma frequência de amostragem de 7.680 Hz e sete níveis para 15.360 Hz. Já para a TWPD optou-se por uma separação de 960 Hz em cada banda da decomposição – foram utilizados apenas os nós-folha da árvore de decomposição da TWPD (MALLAT, 1999). Para tanto, são necessários três níveis de decomposição para uma frequência de amostragem de 7.680 Hz e quatro níveis para 15.360 Hz.

Extração de características: é realizada utilizando a relação de energia para os diferentes níveis da TWD ou da TWPD, conforme apresenta a Seção 4.5.

- Classificadores uma-classe:

*K*-Vizinhos Mais Próximos: para esse modelo é necessário definir o número de Vizinhos-Mais-Próximos adotado e a fração de eventos a serem utilizados no cálculo do limiar. No caso de um classificador construído agrupando todas as classes, foram utilizados os valores [1;2;10;50] para o número de vizinhos mais próximos e para o cálculo do limiar [0,01;0,1;0,2]. Já para um classificador uma-classe construído para cada classe, foram utilizados os valores [1;2;3;5] para o número de vizinhos mais próximos e para o cálculo do limiar [0,01;0,1;0,2]. Esses parâmetros foram determinados utilizando uma validação cruzada repetida cinco vezes, com 20% dos padrões deixados de fora para teste, escolhidos aleatoriamente em cada rodada.

*X*-Médias: para esse modelo é necessário definir a fração de eventos para o cálculo do

limiar. No caso de um classificador construído agrupando todas as classes e para o classificador uma-classe construído para cada classe, foram utilizados os seguintes valores  $[0,01;0,1;0,2]$ . Esses parâmetros foram determinados utilizando uma validação cruzada repetida cinco vezes, com 20% dos padrões deixados de fora para teste, escolhidos aleatoriamente em cada rodada.

Representação de Dados Utilizando Vetores Suporte: para esse método é necessário definir o parâmetro do *kernel*  $\sigma$  adotado e a fração de eventos a serem utilizados no cálculo do limiar. No caso de um classificador construído agrupando todas as classes, foram utilizados os valores  $[1;2;10;50]$  para  $\sigma$  e para o cálculo do limiar  $[0,01;0,1;0,2]$ . Já para um classificador uma-classe construído para cada classe, foram utilizados os valores  $[1;2;3;5]$  para  $\sigma$  e para o cálculo do limiar  $[0,01;0,1;0,2]$ . Esses parâmetros foram determinados utilizando uma validação cruzada repetida cinco vezes, com 20% dos padrões deixados de fora para teste, escolhidos aleatoriamente em cada rodada. Vale ressaltar que, nesse caso, também foi avaliada a metodologia de seleção automática dos parâmetros.

- Classificador Multiclasse:

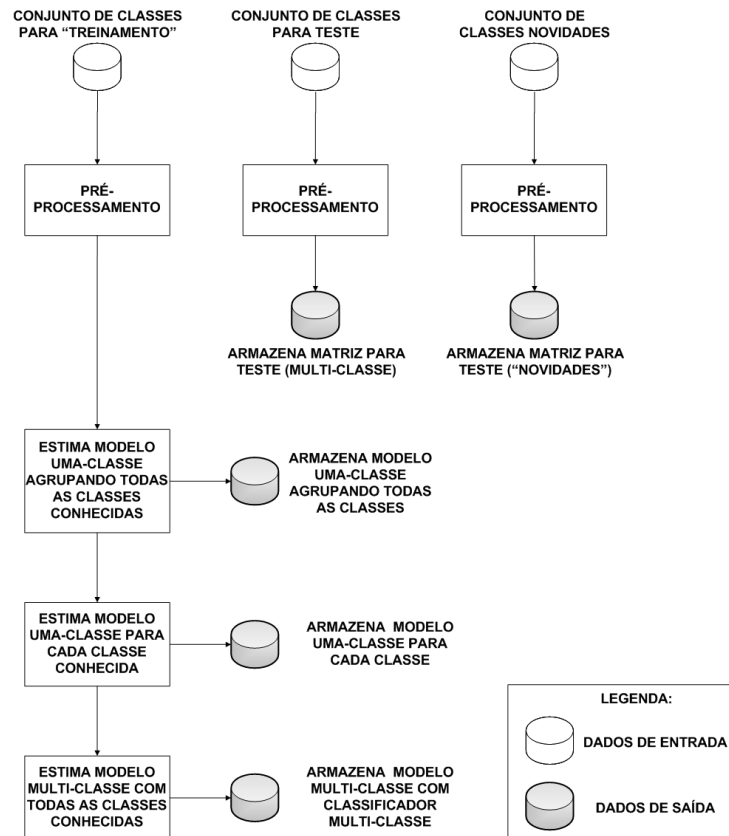
Máquinas de Vetor Suporte: no método escolhido não há necessidade de intervenção no ajuste de parâmetros, uma vez que a minimização do funcional relacionado ao limite superior do erro de generalização fornece estimativas para os parâmetros envolvidos na construção do hiperplano ótimo de separação.

- Número de classes conhecidas *a priori*: a quantidade e quais classes conhecidas previamente foram variadas aleatoriamente no intervalo  $[3,29]$ , sendo esse processo repetido 100 vezes para se obter uma média (e desvio padrão) dos desempenhos calculados, visando estabelecer a comparação de forma independente das classes previamente selecionadas.

### 5.4.3 DADOS REAIS

Para os dados reais, seguiu-se a mesma metodologia aplicada para os dados simulados. A principal diferença é que no caso dos dados reais, não foram escolhidas aleatoriamente as classes conhecidas *a priori*, uma vez que essas já estavam definidas, juntamente com os eventos caracterizados como novidades. Com isso, não é necessária a etapa de escolha de classes, fazendo com que a metodologia aplicada aos dados reais seja mais simples, como demonstra a Figura 67.

Na fase de teste não são verificadas alterações em relação aos dados simulados. Nesse



**Figura 67: Metodologia de treinamento dos modelos aplicada aos dados reais.**

caso, também são aplicadas as duas abordagens apresentadas e as mesmas estatísticas foram obtidas para cada abordagem.

Uma observação importante é que, no caso dos dados reais, foi realizada uma inclusão de dados para permitir um maior equilíbrio no número de eventos por classes. Utilizando apenas os dados da Tabela 7 (página 135), observa-se que existe uma grande diferença entre o número de exemplos para cada classe, conforme demonstra o resumo a seguir:

- Classe 1: Curto-Circuito Monofásico na Fase A (7 exemplos), Curto-Circuito Monofásico na Fase B (2 exemplos) e Curto-Circuito Monofásico na Fase C (17 exemplos), totalizando 26 exemplos;
- Classe 2: Descarga Atmosférica Induzida com 85 exemplos. Do total, 41 eventos foram completamente descartados pois existe uma incerteza significativa na caracterização desses eventos, por parte dos especialistas;
- Classe 3: Religamento Trifásico com 12 exemplos;
- Classe 4 (novidades): Demais classes, em um total de 26 exemplos. Eventos simultâneos e subsequentes menores que um ciclo foram inicialmente retirados da análise, pois existe

uma incerteza muito grande na classificação realizada por especialistas.

Nesse ponto pode-se observar que, caso fosse construído um classificador multiclasse supervisionado utilizando a base de dados disponível até o momento, não existiria um número suficiente de exemplos para caracterizar cada classe.

Devido à grande diferença no número de exemplos por classe, optou-se por incluir dados de curto-circuito e religamento trifásico simulados, de forma que essa diferença fosse reduzida. A utilização desses dois eventos simulados em conjunto com a base de dados real baseia-se na validação realizada para esses eventos no modelo utilizado em Lazzaretti et al. (2009).

Com isso, a seguinte distribuição de dados foi criada:

- Classe 1: Curto-Circuito Monofásico na Fase A (23 exemplos simulados e 7 exemplos reais);
- Classe 2: Curto-Circuito Monofásico na Fase B (23 exemplos simulados e 2 exemplos reais);
- Classe 3: Curto-Circuito Monofásico na Fase C (23 exemplos simulados e 17 exemplos reais);
- Classe 4: Religamento Trifásico (13 exemplos simulados e 12 exemplos reais);
- Classe 5: Descarga Atmosférica Induzida (85 exemplos reais);
- Classe 6 (novidades): Demais classes (26 exemplos reais - utilizados apenas na fase de teste);

Outro aspecto relevante é que com essa inclusão dos dados simulados, torna-se possível individualizar os eventos de curto-circuito monofásico. No Capítulo 6, os resultados levarão em conta apenas a utilização da base de dados definida aqui, a qual contempla os dados reais e simulados de acordo com a distribuição acima apresentada. Convém ressaltar que os dados simulados aqui incluídos na análise em conjunto com os dados reais foram validados com situações reais, durante o trabalho apresentado por Ravaglio et al. (2010).

Uma última observação é que os eventos caracterizados como novidades podem estar inseridos em um dos três cenários descritos no primeiro capítulo deste trabalho: falta de registro prévio para sua caracterização, dificuldade na interpretação do fenômeno ou desconhecimento

com relação à ocorrência desse evento com a forma de onda registrada. Para todos os casos analisados neste trabalho, as novidades serão consideradas completamente desconhecidas no momento da definição dos detectores de novidades, uma vez que esse é o cenário mais provável para classificação de oscilografias em sistemas de distribuição.

#### 5.4.4 PARÂMETROS AJUSTÁVEIS DOS DADOS REAIS

Os parâmetros variados para os dados reais seguem a mesma abordagem apresentada para os eventos simulados. A seguir, são detalhados os parâmetros selecionados para a análise do desempenho da metodologia proposta, nos dados reais:

- Pré-Processamento:

Frequência de amostragem: foram utilizadas duas frequências de amostragem para os dados reais: 245.760 Hz e 122.880 Hz.

Transformada *wavelet*: foram comparados os desempenhos da TWD e da TWPD para o pré-processamento como um todo. No caso da *wavelet*-mãe optou-se por testar apenas o banco de filtro da família Daubechies *db20*. Com relação ao número de níveis, optou-se por manter a frequências inferiores a 240 Hz no sinal aproximação para a TWD. Com isso, são necessários nove níveis de decomposição para uma frequência de amostragem de 245.760 Hz e oito níveis para 122.880 Hz. Já para a TWPD optou-se por uma separação em quatro níveis. Isso reflete uma separação de 15.360 Hz para cada nível em uma frequência de amostragem de 245.760 Hz e 7.680 Hz para uma frequência de amostragem de 122.880 Hz.

Extração de características: na extração de características foi utilizada basicamente a energia, conforme apresenta a Seção 4.5. Além disso, foram incluídos testes com a magnitude e fase do sinal com o intuito de obter uma melhor caracterização dos curtos-circuitos. Desse modo, são adicionadas seis características ao vetor de características obtido com a TWD ou a TWPD.

- Classificadores uma-classe e classificador multiclasse: foram utilizados os mesmos parâmetros apresentados para os dados simulados. É importante ressaltar que, para o modelo RDVS, foram avaliadas duas formulações adicionais para os dados reais: RDVS com exemplos negativos e RDVS com exemplos negativos e otimização da margem. Nesses modelos, é necessário definir a fração de exemplos negativos que devem ser rejeitados (relacionado ao parâmetro  $C_2$ ). Esse parâmetro foi selecionado dentro os valores:  $[0,01;0,1;0,2]$ . Adicionalmente, para o modelo RDVS com exemplos negativos e



otimização da margem, é necessário selecionar o parâmetro da margem  $M$ . Os valores avaliados para esse parâmetro foram [1;5;10;20;50]. Todos os parâmetros envolvidos nesses modelos foram determinados utilizando uma validação cruzada repetida cinco vezes, com 20% dos padrões deixados de fora para teste, escolhidos aleatoriamente em cada rodada.

## 5.5 AGRUPAMENTO AUTOMÁTICO E IDENTIFICAÇÃO DE NOVAS CLASSES

A metodologia selecionada para avaliar os métodos de agrupamento como parte do processo de identificação de novas classes é bastante parecida com a metodologia apresentada nas seções anteriores, para a etapa de classificação e detecção de novidades. Sendo assim, optou-se por apresentar a mesma sequência demonstrada anteriormente, com relação aos parâmetros ajustáveis, estatísticas calculadas e métodos e classes utilizadas, assinalando apenas os aspectos que diferem no caso do agrupamento automático.

Convém ressaltar que o agrupamento automático proposto opera apenas como referência preliminar para o processo de identificação de classes. Conforme citado anteriormente, uma identificação e caracterização completa de uma nova classe, depende da intervenção do especialista. A proposta do agrupamento automático aqui apresentado, é fornecer uma medida que demonstre para o especialista a similaridade existente entre os padrões, facilitando a interpretação dos registros efetuados, permitindo que a identificação de uma nova classe seja agilizada através da correlação dos eventos agrupados e demais bases de dados utilizadas pelo especialista para definição dos eventos.

Os métodos de agrupamento avaliados como identificador de classes, tanto para os dados simulados, quanto para os dados reais foram:

- *Maximum Weighted Likelihood via Rival Penalized Expectation Maximization* ou RPEM: esse modelo utiliza, em linhas gerais, uma representação através de uma Mistura de Gaussianas, sendo que o número de funções Gaussianas na mistura é determinado automaticamente (CHEUNG, 2005);
- Agrupamento baseado em Vetores Suporte: esse modelo considera o agrupamento realizado através do mapeamento para o espaço de características, sendo possível representar grupos com características geométricas (formas) bastante variadas (LEE; LEE, 2005);
- *Self-Tuning Spectral Clustering*: para esse modelo, o problema de agrupamento pode ser visto como um problema de partição de grafos, que pode ser resolvido através do uso da

teoria espectral de grafos (ZELNIK-MANOR; PERONA, 2004). Convém ressaltar que esse modelo também permite grupos com características geométricas (formas) bastante variadas.

- *X-Médias*: esse modelo incorpora ao *K-Médias*, três novas abordagens que visam, principalmente, acelerar a execução do *K-Médias* dentro de um contexto de determinação automática e dinâmica do número de grupos (PELLEG; MOORE, 2000).

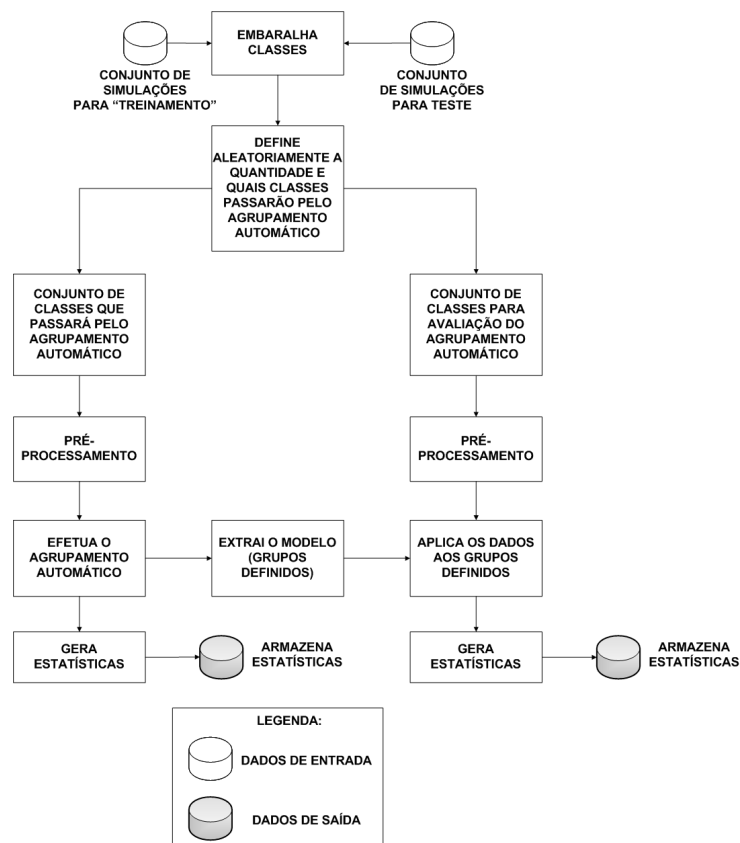
### 5.5.1 DADOS SIMULADOS

Para avaliar o desempenho do método de agrupamento automático nos dados simulados, é necessário definir inicialmente as classes que servirão de base para o método de agrupamento. Nesse sentido, optou-se por utilizar os mesmos cem diferentes arranjos de classes utilizadas para os métodos de classificação e detecção de novidades. A diferença, nesse caso, é a não utilização dos eventos simultâneos e subsequentes. Para avaliar os métodos de agrupamento automático, optou-se por utilizar apenas as 29 classes definidas por simulação, uma vez que facilita a interpretação dos resultados do método de agrupamento como parte do processo de identificação de classes. O resumo da metodologia utilizada para determinação dos conjuntos de avaliação, estimação dos modelos e resultados está apresentado na Figura 68.

A definição das classes que serão analisadas pelo agrupamento automático considera quais classes foram consideradas como conhecidas na etapa de classificação e detecção de novidades da Figura 65. Nessa figura, se, por exemplo, as classes de 1 a 12 são definidas como conhecidas, as classes de 12 a 29 serão utilizadas para avaliar o detector de novidades, em conjunto com os eventos subsequentes e simultâneos. Nesse exemplo, para o método de agrupamento serão utilizadas apenas as classes de 12 a 29. Isso se repete para todos os arranjos determinados.

Depois de definidas as classes, são construídos os grupos de eventos que serão utilizados para definição dos modelos de agrupamento, bem como os eventos que serão utilizados para avaliação desses agrupamentos. Essa divisão é análoga à divisão do conjunto de treinamento e de teste dos modelos. Na sequência, os sinais de cada conjunto são pré-processados da mesma forma que foi apresentado na Seção 5.4.

A definição dos modelos de agrupamento é feita através do conjunto de treinamento. Utilizando esse modelo, aplica-se o conjunto de teste para verificar o desempenho do agrupamento para dados ainda não conhecidos, porém provenientes das mesmas classes definidas inicialmente, tal qual era feito para os modelos de classificação e detecção de novidades.



**Figura 68: Metodologia para agrupamento automático aplicado aos dados simulados.**

Por último, é aplicada a métrica apresentada na Subseção 4.9.3 para os dados de treinamento e de teste, com o intuito de verificar a capacidade do método de agrupamento automático inserir eventos de uma mesma classe em um mesmo grupo. No final do processo, espera-se que o agrupamento indique quantos grupos existem em uma dada distribuição de dados e que esses grupos sirvam como indicativo da distribuição de classes existente.

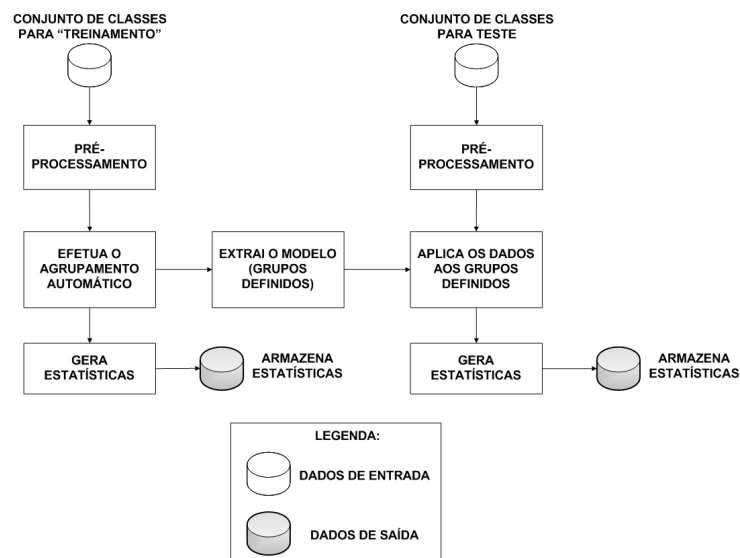
### 5.5.2 PARÂMETROS AJUSTÁVEIS DOS DADOS SIMULADOS

Os parâmetros de pré-processamento seguem a mesma proposta da Seção 4.9.3. A principal diferença é que para os modelos de agrupamento, optou-se por utilizar apenas a frequência de amostragem de 15.360 Hz, pois contém um conteúdo mais amplo de frequência para identificação das classes. Sobre os parâmetros envolvidos nos métodos de agrupamento, foi selecionado automaticamente o número grupos em um intervalo entre  $[0, 1; 10]$  vezes o número de classes utilizadas na modelagem. Demais detalhes de cada modelo de agrupamento serão discutidos no Capítulo 6.

### 5.5.3 DADOS REAIS

A análise dos dados reais segue a mesma metodologia utilizada para os dados simulados, porém o conjunto de classes é fixo e contém a mesma distribuição apresentada na Subseção 5.4.3, considerando os dados reais e os dados simulados de eventos de curto-circuito e religamento trifásico. A principal diferença no caso dos dados reais é que para o método de agrupamento automático foram consideradas as classes conhecidas *a priori*: curto-circuito monofásico na fase A, fase B, fase C, religamento trifásico e descarga atmosférica induzida. Os testes foram focados em verificar a capacidade dos métodos de agrupamento automático em identificar essas cinco classes, uma vez que existiam muitas incertezas por parte do especialista na rotulação dos eventos caracterizados como novidades para os dados reais.

A Figura 69 mostra o resumo dessa metodologia. Vale ressaltar que é muito semelhante à metodologia dos dados simulados, salvo a seleção prévia das classes realizada apenas para os dados simulados. Depois de definidos os conjuntos de treinamento e teste, é feito o pré-processamento e calculada a métrica baseada na matriz de confusão, definida na Subseção 4.9.3, para ambos os conjuntos.



**Figura 69: Metodologia para agrupamento automático aplicada aos dados reais.**

### 5.5.4 PARÂMETROS AJUSTÁVEIS DOS DADOS REAIS

Os parâmetros de pré-processamento seguem a mesma proposta da Subseção 5.4.3. Sobre os parâmetros envolvidos nos métodos de agrupamento, foi selecionado automaticamente o número grupos em um intervalo entre  $[0, 1; 10]$  vezes o número de classes utilizadas na modelagem. Demais detalhes de cada modelo de agrupamento serão discutidos no Capítulo 6.

## 5.6 PACOTES DE FUNÇÕES UTILIZADAS

Convém citar finalmente, que as funções implementadas pelo autor deste trabalho utilizaram como referência as funções já disponíveis nos seguintes pacotes de funções (Matlab):

- O pré-processamento com base na transformada *wavelet* foi realizado utilizando o pacote *Wavelab* (BUCKHEIT et al., 2005);
- Os classificadores uma-classe (detectores de novidade) e a análise estatística foi feita com base no conjunto de ferramentas desenvolvidas por David M. J. Tax, denominada *Data Description Toolbox* (TAX, 2012);
- Para classificação multiclasse com base em MVS, foi utilizado como base principal o desenvolvimento da *LIB-SVM* (CHANG; LIN, 2011), com a definição automática de parâmetros disponibilizadas em Canu et al. (2005);
- Para o agrupamento automático *X-Médias*, foi utilizado como base principal o desenvolvimento apresentado por (PELLEG; MOORE, 2000). Para os demais métodos de agrupamento automático, foram utilizadas as implementações propostas pelos autores dos artigos originais.

Adicionalmente, os próprios trabalhos originais de cada método foram considerados para ajustes e implementações finais.

## 6 RESULTADOS

### 6.1 SEGMENTAÇÃO DE EVENTOS

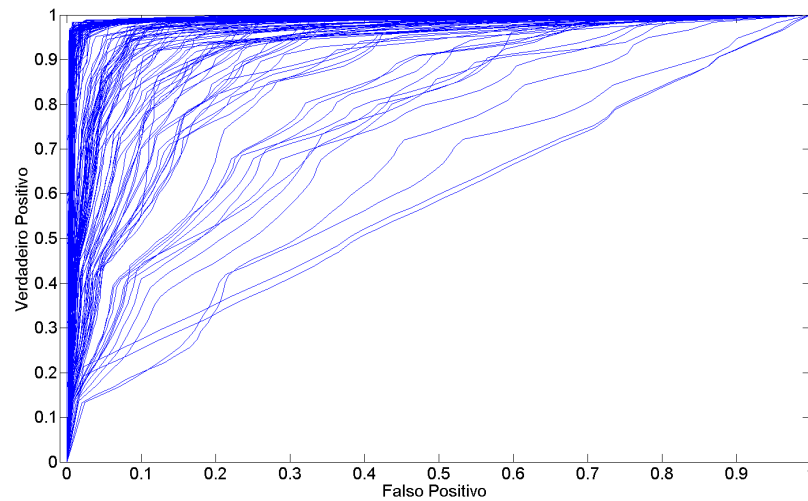
Para avaliar a segmentação dos eventos, optou-se por separar a análise para duas bases de dados distintas, sendo a primeira delas para oscilografias geradas via simulação e a segunda para dados reais obtidos do sistema de monitoramento. Com isso, foi possível testar e validar os métodos para os arquivos simulados para eventos das mais distintas naturezas no *ATPDraw* e, em seguida, aplicá-los nas oscilografias reais. A seguir, serão descritos os principais resultados para cada base de dados, sendo um maior detalhamento dedicado aos testes na base de dados reais.

#### 6.1.1 BASE DE DADOS SIMULADOS

Na Figura 70 está representada a curva ROC para diferentes limiares no caso das 170 oscilografias geradas por simulação, utilizando o método de segmentação baseado nos modelos autorregressivos. Observa-se nessa figura a grande concentração dos pontos na região superior esquerda do gráfico, o que implica em um bom desempenho do sistema de detecção, para determinados limiares na análise ROC.

Conforme apresentado anteriormente, visando resumir a análise de desempenho do método de detecção e permitir uma melhor comparação para os diferentes métodos, optou-se por calcular a AAC da curva ROC para diferentes limiares de comparação para a detecção. Nesse sentido, a Figura 70 fornece uma análise mais qualitativa dos desempenhos de detecção obtidos, bem como do comportamento de cada método para diferentes oscilografias. Para comparação entre métodos, convém analisar a Tabela 8, a qual resume o desempenho para os quatro modelos de detecção testados. Para o modelo baseado na TWD, foi utilizada a proposta baseada na segmentação automática através do limiar universal, apresentada por Ukil e Zivanovic (2005a).

Aplicando o teste de Wilcoxon, pode-se afirmar que os desempenhos dos métodos FK



**Figura 70: Análise ROC das 170 oscilografias simuladas. Cada curva representa uma das 170 oscilografias simuladas para diferentes limiares.**

**Tabela 8: Desempenho do método de segmentação para dados simulados.**

Modelo	AAC para Diferentes Limiares (Média $\pm$ Desvio Padrão)	Tempo Médio de Simulação (s)
AR	0,93 $\pm$ 0,11	650
TWD	0,93 $\pm$ 0,09	665
TWD+RDVS	0,97 $\pm$ 0,07	820
FK	0,92 $\pm$ 0,14	620
OET	0,92 $\pm$ 0,13	560

e OET são equivalentes, bem como os modelos AR e TWD. As demais distribuições podem ser consideradas estatisticamente distintas entre si, sendo que o melhor desempenho foi do método TWD+RDVS, considerando diferentes limiares para os dados simulados.

Com relação ao tempo de simulação, os sinais foram analisados com frequência de amostragem de 983.040 Hz e duração de 18 ciclos. Com essas características, observa-se um tempo relativamente alto para o processo de segmentação, em especial para os modelos baseados em TWD+RDVS, passando de 10 minutos para realizar toda a segmentação. Vale ressaltar que todos os experimentos foram realizados utilizando um microcomputador com processador Intel *Core2-quad* 32 bits, 4 GB de memória RAM e frequência de operação de 3GHz, em um código Matlab não otimizado.

Conforme demonstrado até aqui, a base de dados de simulação fornece uma estimativa inicial importante para os métodos de detecção, uma vez que é possível visualizar a influência de diferentes parâmetros como a natureza do evento e os parâmetros associados (instante de ocorrência, resistência de falta, etc) e eventos subsequentes em diferentes intervalos e de dife-

rentes naturezas. No entanto, a validação desses métodos e sua devida comparação deve ser feita para os dados reais que se procura analisar. Nesse sentido, a subseção seguinte trata dos resultados obtidos para os dados reais coletados pelos sistemas de monitoramento apresentados anteriormente.

### 6.1.2 BASE DE DADOS REAIS

Para avaliar o desempenho dos métodos de segmentação nos sistemas de monitoramento, optou-se por apresentar os resultados para Apucarana-1, São Mateus-1 e São Mateus-2, dada a diversidade de eventos registrados nessas estações. Foram avaliados os quatro métodos de detecção para cada uma dessas estações, admitindo diferentes limiares e limiares automaticamente calculados. Além disso, todo o processo de segmentação foi realizado considerando uma subamostragem dos sinais de entrada, visando reduzir o custo computacional envolvido nessa etapa, bem como seus efeitos no desempenho global no processo de segmentação.

Primeiramente, os sinais foram reamostrados para uma frequência de amostragem de 7.680Hz. Os instantes de tempo obtidos para os transitórios presentes no sinal com a frequência de amostragem de 7.680Hz são aplicados para o sinal com a frequência de amostragem original. Ou seja, a partir dos sinais subamostrados, obtêm-se os instantes de tempo correspondentes à transição entre os estados quase-estacionários e estes são utilizados para a segmentação do sinal na frequência de amostragem original. Com isso, é possível comparar os instantes transitórios obtidos através desse artifício com o padrão-ouro original determinado para todos os sinais.

Adicionalmente, os resultados para o limiar automaticamente determinado estão representados através do acerto médio de classificação. Para determinar esse acerto, considera-se que o problema de segmentação consiste em determinar para cada amostra da oscilografia, qual sua classe correspondente: transitório ou não-transitório. Desse modo, é possível construir uma matriz de confusão utilizando como base o padrão-ouro existente para cada oscilografia avaliada. Uma vez determinada a matriz de confusão, é possível calcular o acerto de cada classe e, conseqüentemente, o acerto médio de classificação. Essa métrica fornece uma possibilidade de comparação mais prática entre os desempenhos dos métodos, quando comparada à AAC calculada para apenas um limiar – métrica normalmente adotada para problemas de detecção de variações abruptas em sinais (DESOBRY et al., 2005). Os resultados desse processo podem ser observados na Tabela 9.

Um detalhe importante sobre os resultados apresentados é que a subamostragem do sinal não impacta em reduções significativas de desempenho, quando utilizada como base para a segmentação do sinal na frequência de amostragem original. Foi verificado no caso particular



**Tabela 9: Desempenho do método de segmentação para Apucarana-1.**

Modelo	AAC para Diferentes Limiares (Média $\pm$ Desvio Padrão)	Acerto Médio para Limiar Calculado Automaticamente (Média $\pm$ Desvio Padrão)	Tempo Médio de Simulação (s)
AR	0,76 $\pm$ 0,06	0,67 $\pm$ 0,08	6,7
TWD+RDVS	0,96 $\pm$ 0,11	0,93 $\pm$ 0,11	23,5
FK	0,77 $\pm$ 0,12	0,77 $\pm$ 0,13	9,2
OET	0,68 $\pm$ 0,07	0,63 $\pm$ 0,10	5,5
TWD	0,90 $\pm$ 0,09	0,86 $\pm$ 0,10	13,6

dos modelos TWD+RDVS que, quando aplicado o método de segmentação na frequência de amostragem original durante todo o processo, obtém-se um valor médio para a AAC de  $0,65 \pm 0,07$ , considerando diferentes limiares. Esse baixo desempenho está associado ao fato dos primeiros níveis da TWD não possuírem características relevantes ao processo de detecção para os sinais reais com frequência de amostragem de 983.040 Hz, ao contrário do observado quando se utiliza a frequência de 7.680 Hz para se obter os instantes dos transitórios e, em seguida, utilizá-los para segmentação do sinal na frequência de amostragem original. Adicionalmente, o tempo computacional gasto para efetuar a segmentação de uma oscilografia pode ultrapassar 40 minutos nesse método.

Já para os métodos de estimação espectral o cenário é bastante distinto, pois diferentemente dos eventos simulados, as transições amostra a amostra são menos abruptas, o que impacta diretamente no cálculo dos resíduos, fazendo com que, em alguns casos, não seja possível identificar os transitórios entre estados quase-estacionários utilizando os resíduos. Sendo assim, a redução da frequência de amostragem se mostra necessária em alguns casos, já que promove indiretamente, uma variação mais significativa entre amostras sem perder a informação temporal associada com os transitórios, conforme demonstra a comparação de desempenho para a frequência de amostragem original e de 7.680 Hz, realizada para o modelo TWD+RVDS. Apenas para ilustrar, quando aplicada a segmentação com a frequência de amostragem original para os modelos de estimação espectral, o desempenho global médio é consideravelmente inferior ao obtido para a frequência de amostragem de 7.680 Hz, sendo esse desempenho em torno de 0,55 para a AAC, ou seja, pouco superior ao desempenho aleatório de segmentação. Sendo assim, a subamostragem proposta se faz necessária para um bom desempenho de detecção, além de otimizar o processamento envolvido, já que podem ser observados tempos de segmentação superiores a 20 minutos para uma oscilografia nesses métodos.

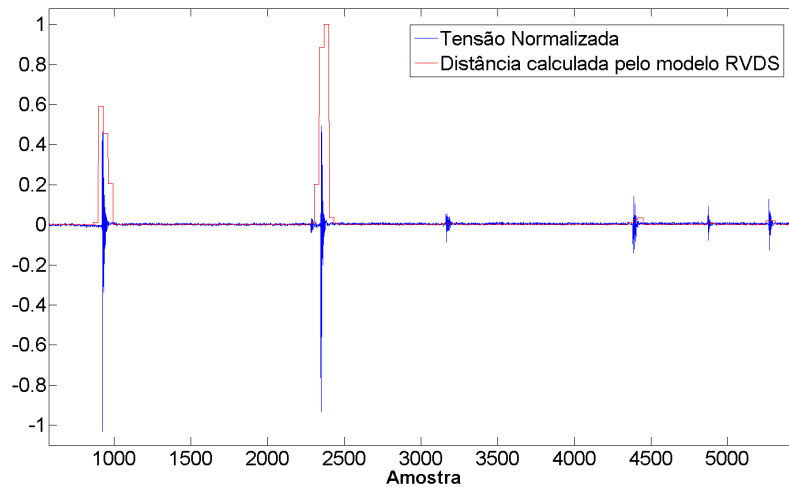
Vale ressaltar que, além da redução no número de amostras em função da subamostragem, também foi feita uma redução do número de níveis na decomposição utilizando TWD

para o método baseado na TWD+RDVS e para o próprio modelo com base na TWD, de modo a manter a mesma janela de observação de 1/8 ciclo. Para tanto, foram utilizados os quatro primeiros níveis da TWD (sinal detalhe). Essa janela de observação também foi mantida para os demais modelos, facilitando o processo de comparação de desempenho.

O teste de Wilcoxon aplicado aos dados da Tabela 9 indicou uma diferença estatisticamente significativa entre todas as distribuições geradas para cada método avaliado, tanto para os diferentes limiares, quanto para o limiar calculado automaticamente. Sendo assim, o modelo baseado em TWD+RDVS possui melhor desempenho para diferentes limiares e para o limiar determinado automaticamente, com um acerto global superior aos demais métodos, o que indica a robustez dessa metodologia, mesmo para oscilografias de diferentes eventos na mesma forma de onda. Uma observação importante é a comparação entre o modelo TWD+RDVS e o modelo baseado somente na TWD. Utilizando a RDVS em conjunto com o pré-processamento baseado na TWD, é possível obter um acréscimo significativo na classificação dos instantes transitórios e não-transitórios. Nesse caso, esse acréscimo foi, em média, de 7%.

Em algumas oscilografias, o desempenho de todos os métodos foi bem inferior ao desempenho médio. Nessas oscilografias são observados diferentes fenômenos. O primeiro deles, em relação às oscilografias de descargas atmosféricas, observam-se transitórios de pequena magnitude e duração, em diferentes pontos da forma de onda, juntamente com um ou mais transitórios principais, de maior magnitude e duração. Os transitórios de menor intensidade podem estar associados a descargas de menor intensidade, subsequentes a uma descarga de grande magnitude responsável pelo transitório principal da forma de onda (Figura 71). Sendo assim, a segmentação é dificultada, pois um dado limiar pode definir a segmentação de algumas descargas atmosféricas subsequentes com uma dada intensidade, mas, ao mesmo tempo, deixar de considerar outras com diferentes intensidades. Esse cenário pode ser repetido para qualquer oscilografia que contenha diferentes transitórios na forma de onda, com diferentes intensidade e duração, uma vez que a intensidade e a duração estão diretamente relacionadas com o cálculo dos resíduos nos modelos de estimação espectral, distâncias no modelo TWD+RDVS ou a saída do OET, os quais apresentam características mais proeminentes para transitórios de maior intensidade e duração.

Adicionalmente, foram verificados eventos de abertura seguidos de curto-circuito em um intervalo inferior a um ciclo (Figura 72). Nesses casos, a segmentação de cada um desses eventos fica comprometida, pois os transitórios decorrentes do curto-circuito são estendidos (mesmo em menor intensidade) até o início da abertura dos mecanismos de proteção. Com isso, o processo de segmentação acaba definindo esses dois processos transitórios como um tran-

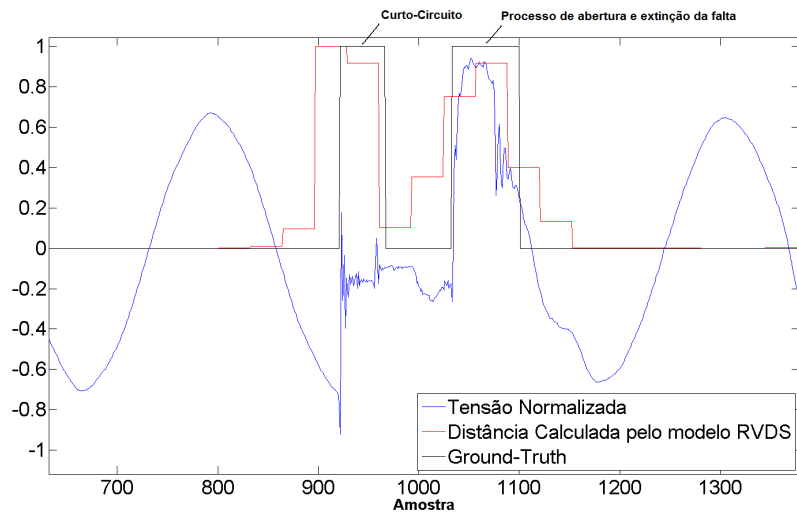


**Figura 71: Detalhe do método TWD+RVDS para descargas atmosféricas subsequentes. Nota-se que as maiores distâncias calculadas pelo modelo estão associadas com os transitórios de maior intensidade. Neste caso, esperava-se obter um cálculo de distância que destacasse de maneira uniforme todos os transitórios presentes na forma de onda. Vale ressaltar que a rede de distribuição encontrava-se desenergizada no instante do registro.**

sitório único, o que não corresponde ao previsto pela segmentação visual prévia, realizada para definição do padrão-ouro. Nesse sentido, tanto os diferentes limiares quanto os limiares automaticamente calculados não promovem a segmentação adequada dos dois eventos. Assinala-se que essa foi a principal dificuldade encontrada em todos os métodos de detecção propostos, uma vez que os transitórios que definem a mudança de estados podem estar sobrepostos, dificultando a segmentação dos eventos. Uma alternativa para a interpretação desses fenômenos seria trabalhar com a classificação de eventos simultâneos ou subsequentes em intervalos muito próximos de tempo, descartando a segmentação refinada desses eventos.

Uma observação importante é que os pontos resultantes da segmentação automática localizavam-se em sua maioria na parte esquerda do gráfico ROC, ou seja, com taxa de falsos positivos mais próxima de zero, garantindo que o início do transitório fosse detectado, mesmo que de forma antecipada, ocasionando uma menor taxa de verdadeiros positivos. Nesse sentido, é garantido que o instante inicial dos eventos foi detectado de forma satisfatória, mesmo para registros contendo diferentes eventos na mesma oscilografia, sendo que apenas a duração de cada transitório é feita com menor acurácia.

Do ponto de vista de classificação, em muitos casos apenas o instante inicial dos eventos é necessário para aplicação dos algoritmos de classificação de eventos. No entanto, o que está proposto aqui é a extração de todo o transiente entre as mudanças de estados quase-estacionários presentes na oscilografia. Isso implica em poder maximizar a probabilidade de



**Figura 72: Detalhe do método de detecção TWD+RVDS para abertura subsequente a um curto-circuito. Nesse caso, limiares inferiores a 0,1 poderiam agrupar os dois eventos, curto-circuito e abertura, em um mesmo transitório, ao passo que limiares superiores poderiam segmentar a oscilografia de forma mais adequada para posterior análise e correta classificação.**

detecção do início do evento, mesmo que de forma antecipada, como mostra a Figura 72, sem no entanto, comprometer a classificação dos eventos. Além disso, pode-se calcular com boa precisão a duração do evento transitório, o que corresponde a uma informação bastante relevante, principalmente para eventos relacionados à descargas atmosféricas.

Os resultados obtidos para as estações São Mateus-1 e São Mateus-2 podem ser observados nas tabelas 10 e 11.

**Tabela 10: Desempenho do método de segmentação para São Mateus-1.**

Modelo	AAC para Diferentes Limiares (Média $\pm$ Desvio Padrão)	Acerto Médio para Limiar Calculado Automaticamente (Média $\pm$ Desvio Padrão)	Tempo Médio de Simulação (s)
AR	0,78 $\pm$ 0,07	0,72 $\pm$ 0,11	12,2
TWD+RDVS	0,96 $\pm$ 0,08	0,92 $\pm$ 0,09	20,2
FK	0,80 $\pm$ 0,09	0,70 $\pm$ 0,09	9,1
OET	0,79 $\pm$ 0,10	0,69 $\pm$ 0,11	5,8
TWD	0,89 $\pm$ 0,12	0,83 $\pm$ 0,13	15,8

Do ponto de vista global, o método baseado em TWD+RDVS novamente apresentou o melhor desempenho para diferentes limiares e para uma segmentação completamente automática, já que o teste de Wilcoxon indicou diferença estatisticamente significativa entre os resultados obtidos para o modelo TWD+RDVS e os demais nesses dois casos. A mesma observação feita com relação ao acréscimo de desempenho utilizando o método RDVS, em relação à utilização da TWD, vale para as estações de São Mateus. Já o método baseado no

**Tabela 11: Desempenho do método de segmentação para São Mateus-2.**

Modelo	AAC para Diferentes Limiares (Média $\pm$ Desvio Padrão)	Acerto Médio para Limiar Calculado Automaticamente (Média $\pm$ Desvio Padrão)	Tempo Médio de Simulação (s)
AR	0,70 $\pm$ 0,07	0,67 $\pm$ 0,15	11,1
TWD+RDVS	0,95 $\pm$ 0,08	0,89 $\pm$ 0,11	23,1
FK	0,80 $\pm$ 0,07	0,68 $\pm$ 0,11	12,1
OET	0,64 $\pm$ 0,08	0,62 $\pm$ 0,10	6,8
TWD	0,86 $\pm$ 0,13	0,83 $\pm$ 0,13	15,4

OET apresentou o menor desempenho para as três estações de monitoramento avaliadas.

É importante observar que, caso seja analisada apenas a eficiência do mapeamento gerado pelos resíduos, distâncias ou a saída do OET, observa-se um desempenho superior do método TWD+RDVS para todas as estações analisadas, com valores de AAC superiores a 0,95, representando uma alta capacidade de segmentação dos transientes existentes no sinal, em relação aos demais métodos. Nesse sentido, observa-se que o processo de segmentação automática proposto para o modelo RDVS é capaz de extrair de forma satisfatória as informações contidas no mapeamento realizado nesse método.

Ainda com relação ao desempenho de detecção de eventos, é possível analisar os erros envolvidos nos tempos de detecção do instante inicial dos estados transitórios para todas as oscilografias reais registradas. Para tanto, optou-se por apresentar para cada estação de monitoramento, o erro médio de detecção do instante inicial para transitórios que são identificados antes do instante verdadeiro do seu início, bem como transitórios identificados depois de seu início verdadeiro. Os resultados para o modelo TWD+RDVS estão resumidos a seguir, para as três unidades de monitoramento:

- São Mateus-1:

Erro médio de detecção para instantes iniciais determinados antes dos instantes verdadeiros, verificado em 99% dos eventos: 1,5 ms;

Erro médio de detecção para instantes iniciais determinados depois dos instantes verdadeiros, verificado em apenas 1% dos eventos: 4,9 ms;

- São Mateus-2:

Erro médio de detecção para instantes iniciais determinados antes dos instantes verdadeiros, verificado em 98% dos eventos: 2 ms;

Erro médio de detecção para instantes iniciais determinados depois dos instantes verdadeiros, verificado em 2% dos eventos: 7 ms;

- Apucarana-1:

Erro médio de detecção para instantes iniciais determinados antes dos instantes verdadeiros, verificado em 97% dos eventos: 2,2 ms;

O método de detecção identificou o instante inicial de forma exata em 3% dos eventos para essa estação.

Em termos de custo computacional, o método TWD+RDVS apresenta um tempo para realizar a segmentação superior aos demais métodos. No entanto, ressalva-se que uma série de otimizações podem ser implementadas nos códigos, a fim de reduzir o custo computacional envolvido. Essas otimizações podem, inclusive, permitir que alguns desses métodos possam ser implementados em tempo real. Entretanto, essas questões fogem ao escopo deste trabalho, sendo caracterizadas como sugestões para trabalhos futuros.

Um último aspecto a ser abordado aqui, diz respeito à escolha dos parâmetros e sua influência em cada método de detecção. Detalhes desse tópico serão discutidos na subseção a seguir.

### 6.1.3 DEFINIÇÃO DOS PARÂMETROS

Para seleção dos parâmetros utilizados em cada modelo foram realizados diversos experimentos, tendo por base de comparação a AAC calculada para diferentes limiares na curva ROC.

Para os modelos de estimação espectral, o parâmetro mais impactante na avaliação do desempenho de detecção é a ordem selecionada para o modelo. Já no caso do método baseado em TWD+RDVS (e indiretamente o modelo baseado somente na TWD), os parâmetros relevantes são  $\sigma$  e  $\nu$  (relacionado ao parâmetro  $C$  do modelo RDVS), bem como a função *wavelet* utilizada. Para realizar essa comparação optou-se por apresentar os resultados para a estação de monitoramento Apucarana-1, dada sua maior variedade de eventos registrados.

Na Tabela 12 está representado o resultado para o método baseado no FK. Observa-se que, quanto maior a ordem do modelo, melhor o desempenho de detecção. No entanto, a partir da vigésima ordem, o desempenho de detecção não apresenta variações significativas, inserindo apenas uma maior complexidade computacional na detecção. Sendo assim, optou-se por utilizar a representação espectral até a vigésima ordem para modelos baseados no FK. A comparação através do teste de Wilcoxon indica uma equivalência apenas entre os desempenhos obtidos para  $K = 5$  e  $K = 10$ , sendo as demais comparações consideradas estatisticamente distintas.

**Tabela 12: Variação da ordem do Filtro de Kalman.**

Modelo	AAC para Diferentes Limiares (Média $\pm$ Desvio Padrão)
$K = 1$	$0,66 \pm 0,06$
$K = 5$	$0,72 \pm 0,11$
$K = 10$	$0,73 \pm 0,12$
$K = 20$	$0,77 \pm 0,12$

Já na Tabela 13, apresenta-se essa mesma análise para os modelos AR. Para esses modelos vale a mesma observação apresentada para os baseados no FK, sendo que, nesse caso, os três desempenhos são estatisticamente equivalentes do ponto de vista do teste de Wilcoxon. Nesse caso, optou-se por utilizar modelos AR de ordem cinco.

**Tabela 13: Variação da ordem do modelo autorregressivo.**

Modelo	AAC para Diferentes Limiares (Média $\pm$ Desvio Padrão)
$K = 1$	$0,75 \pm 0,06$
$K = 5$	$0,76 \pm 0,05$
$K = 10$	$0,76 \pm 0,06$

Com relação ao modelo OET, não existem parâmetros a serem definidos previamente, uma vez que nessa abordagem, utilizam-se apenas as amostras atual, prévia e posterior para o cálculo utilizado na detecção de eventos. Essa característica pode ser apontada como uma vantagem desse método.

Finalmente, os modelos TWD+RDVS possuem uma maior quantidade de parâmetros a serem definidos *a priori*. O primeiro conjunto está relacionado ao pré-processamento utilizando a TWD. O desempenho para diferentes funções *wavelet* está representado na Tabela 14. O teste de Wilcoxon indica que existe equivalência estatística apenas entre os desempenhos das *wavelets db8* e *db20*. Nesse caso, optou-se por uma *wavelet* com ordem elevada, em especial a *wavelet* Daubechies-20, já que o fator  $Q$  dos filtros das funções *wavelet* de maior ordem é mais elevado, limitando a banda passante nos diferentes níveis da TWD. No entanto, poder-se-ia utilizar uma função *wavelet* de menor ordem, como a *wavelet db8*, reduzindo o tempo de cálculo da transformação e, conseqüentemente, o tempo total para realização da detecção.

**Tabela 14: Variação das funções *wavelet* para o modelo RDVS.**

Modelo	AAC para Diferentes Limiares (Média $\pm$ Desvio Padrão)
<i>db1</i>	$0,78 \pm 0,12$
<i>db8</i>	$0,93 \pm 0,11$
<i>db20</i>	$0,93 \pm 0,11$

O segundo conjunto de parâmetros do modelo TWD+RDVS está associado à própria representação dos dados utilizando os vetores suporte (ver subseção 4.8.5 e 4.8.6). Nesse

sentido, duas abordagens foram utilizadas, sendo a primeira delas uma variação manual dos parâmetros, como demonstra a Tabela 15. Nessa variação manual, o teste de Wilcoxon indica equivalência entre todos os desempenhos apresentados.

**Tabela 15: Variação dos parâmetros para o modelo RDVS.**

Modelo	AAC para Diferentes Limiares (Média $\pm$ Desvio Padrão)
$\sigma = 0,1$ e $\nu = 0,05$	$0,90 \pm 0,11$
$\sigma = 0,1$ e $\nu = 0,1$	$0,90 \pm 0,11$
$\sigma = 0,1$ e $\nu = 0,2$	$0,90 \pm 0,11$
$\sigma = 1$ e $\nu = 0,05$	$0,92 \pm 0,12$
$\sigma = 1$ e $\nu = 0,1$	$0,91 \pm 0,12$
$\sigma = 1$ e $\nu = 0,2$	$0,91 \pm 0,12$
$\sigma = 5$ e $\nu = 0,05$	$0,93 \pm 0,12$
$\sigma = 5$ e $\nu = 0,1$	$0,92 \pm 0,12$
$\sigma = 5$ e $\nu = 0,2$	$0,93 \pm 0,12$
$\sigma = 20$ e $\nu = 0,2$	$0,93 \pm 0,11$

Uma segunda simulação realizada envolveu a determinação dos parâmetros  $\sigma$  e  $\nu$  de forma automática, utilizando apenas 20 oscilografias da estação Apucarana-1. Para tanto, os parâmetros  $\sigma$  e  $\nu$  foram determinados automaticamente de acordo com a metodologia exposta na subseção 4.8.6, em um intervalo de  $[0,1;20]$  e  $[0,05;0,2]$  para  $\sigma$  e  $\nu$ , respectivamente. O custo computacional associado a essa determinação automática impede que testes sejam realizados com maiores detalhes para demais oscilografias. Apenas para ilustrar, o tempo médio para analisar uma forma de onda com 24 ciclos, amostrados a 7.680 Hz, é de aproximadamente 1600 segundos. A Tabela 16 resume o desempenho para essa metodologia, em comparação à mesma análise feita para  $\sigma = 20$  e  $\nu = 0,2$ . Essa comparação demonstra que o desempenho obtido para os valores escolhidos de  $\sigma$  e  $\nu$  estão muito próximos dos valores otimizados desses parâmetros, além do teste de Wilcoxon indicar equivalência estatística entre os desempenhos obtidos. Desse modo, optou-se por manter os valores  $\sigma = 20$  e  $\nu = 0,2$  para a segmentação através do modelo RDVS.

**Tabela 16: Modelo RDVS com parâmetros ajustados automaticamente.**

Modelo	AAC para Diferentes Limiares (Média $\pm$ Desvio Padrão)
Automático (DUIN, 2001)	$0,97 \pm 0,13$
$\sigma = 20$ e $\nu = 0,2$	$0,96 \pm 0,12$

## 6.2 CLASSIFICAÇÃO E DETECÇÃO DE NOVIDADES EM DADOS SIMULADOS

Para apresentar o desempenho da classificação de eventos e detecção de novidades, optou-se por separar os resultados apresentados para as bases de dados reais e simuladas, tal



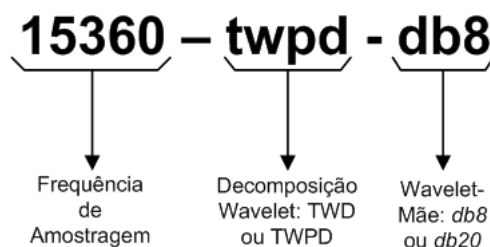
qual foi feito para os métodos de segmentação. Essa seção será dedicada aos dados simulados e inicialmente será discutida a influência dos parâmetros ajustáveis no desempenho global de classificação, para todos os métodos propostos. Em seguida, serão detalhados os desempenhos das abordagens DNG-CMC e DNI-CBS, utilizando diferentes classes conhecidas *a priori*.

### 6.2.1 ANÁLISE INICIAL DOS PARÂMETROS PARA OS DADOS SIMULADOS

Para a avaliação da influência dos parâmetros, foi definido um conjunto para as classes normais e novidades e esse conjunto foi mantido fixo durante toda a avaliação. Esses conjuntos foram definidos conforme segue:

- Classes de eventos normais (conhecidas): curto-circuito monofásico nas fases A, B e C; curto-circuito bifásico para terra nas fases AB, BC e CA; curto-circuito bifásico nas fases AB, BC e CA; curto-circuito trifásico para terra; curto-circuito trifásico.
- Classes de eventos caracterizados como novidades (não conhecidas): abertura monofásica na fase A, B e C; abertura bifásica na fase AB, BC e CA; abertura trifásica; fechamento monofásico na fase A, B e C; fechamento bifásico na fase AB, BC e CA; fechamento trifásico; fechamento de banco de capacitor; curto-circuito monofásico à montante da subestação; curto-circuito bifásico para o terra à montante da subestação; partida de um grande motor.

Na Tabela 17 estão representados os desempenhos globais (desempenho global do detector de novidades e classificação multiclasse), para a abordagem DNG-CMC, considerando diferentes formas de pré-processar o sinal. Essas diferentes formas são representadas através de uma codificação, detalhada na Figura 73.



**Figura 73: Detalhamento da codificação apresentada para o pré-processamento.**

Pela análise da Tabela 17, observa-se que o melhor desempenho obtido foi para o método *X-Médias*, para o pré-processamento utilizando 15.360 Hz de frequência de amostragem, a TWPD para a decomposição e a *wavelet*-mãe *db20*, com 89,25% de acerto. Com relação

**Tabela 17: Desempenho global (%) para diferentes pré-processamentos (abordagem DNG-CMC).**

Pré-Processamento	X-Médias	KVMP	RDVS	Média
15360-twpd-db8	58,21	80,72	48,47	62,47
15360-twd-db8	73,34	74,77	66,67	71,59
7680-twpd-db8	28,72	28,56	28,55	28,61
7680-twd-db8	84,18	77,44	73,23	78,28
15360-twpd-db20	<b>89,25</b>	80,15	47,47	72,29
15360-twd-db20	84,90	85,71	73,99	<b>81,53</b>
7680-twpd-db20	39,30	38,34	28,81	35,48
7680-twd-db20	82,15	80,12	74,89	79,05
Média	<b>67,15</b>	<b>68,23</b>	55,26	–

aos métodos de classificação, o melhor desempenho médio para as diferentes formas de pré-processamento foi do método KVMP, seguido do método de agrupamento X-Médias. Se for comparado o desempenho médio, independentemente do método de classificação, observa-se que o melhor desempenho médio fica a cargo do pré-processamento *15360-twd-db20*.

Na Tabela 18 são apresentados os resultados para a abordagem DNI-CBS. Nesse caso, os modelos uma-classe operam como detector de novidades para cada uma das classes conhecidas, além de realizar a classificação multiclasse de maneira acoplada. Pela análise dessa tabela, fica claro que o desempenho médio global é inferior ao anteriormente apresentado. Nessa tabela, também observa-se que o método com os melhores desempenhos foi o baseado em RDVS. Para o treinamento do modelo RDVS, foi utilizada a formulação RDVS sem exemplos negativos. Com relação ao pré-processamento, os melhores desempenhos foram obtidos por *15360-twpd-db20*.

**Tabela 18: Desempenho global (%) para diferentes pré-processamentos (abordagem DNI-CBS).**

Pré-Processamento	X-Médias	KVMP	RDVS	Média
15360-twpd-db8	45,80	19,91	68,77	44,83
15360-twd-db8	28,56	41,89	22,56	31,00
7680-twpd-db8	18,48	31,40	14,02	21,30
7680-twd-db8	27,54	45,99	30,25	34,59
15360-twpd-db20	54,63	22,69	<b>71,30</b>	<b>49,54</b>
15360-twd-db20	30,28	42,77	21,74	31,60
7680-twpd-db20	17,21	31,76	15,97	21,65
7680-twd-db20	35,04	44,20	25,43	34,89
Média	32,19	<b>35,08</b>	33,76	–

A conclusão obtida pela análise de todos os métodos mencionados, juntamente com os pré-processamentos aplicados, é que o melhor desempenho médio observado foi para o pré-processamento com as características extraídas apenas utilizando a energia de cada nível da

transformação, com a *wavelet*-mãe *db20*. Para a frequência de amostragem optou-se por manter 15.360 Hz, incluindo um conteúdo de frequência mais amplo dos eventos, juntamente com uma decomposição via TWPD, muito embora os resultados não apontem para uma escolha única em relação a esses dois fatores. Desse modo, os demais resultados para os dados simulados serão apresentados utilizando apenas o pré-processamento *15360-twpd-db20*.

### 6.2.2 ABORDAGEM DNG-CMC

Fixada a forma de pré-processar o sinal, é necessário avaliar a influência de outro parâmetro: número de classes conhecidas *a priori*. Vale lembrar que os demais parâmetros associados aos classificadores uma-classe são estimados via validação cruzada.

Para cada uma das abordagens de classificação propostas, DNG-CMC e DNI-CBS, será detalhado o desempenho individual do detector de novidades e o desempenho global do processo de classificação. Além disso, será apresentado o desempenho do classificador multi-classe MVS isolado, quando aplicado para as diferentes classes conhecidas do problema.

### 6.2.3 DESEMPENHO INDIVIDUAL DO DETECTOR DE NOVIDADES

A Tabela 19 apresenta a média percentual e o desvio padrão do desempenho individual do detector de novidades, para 100 diferentes arranjos de classes conhecidas *a priori*. Vale lembrar que nesse arranjo a quantidade e as classes são escolhidas aleatoriamente. Para os quatro classificadores uma-classe avaliados, o desempenho do detector de novidades é, em média, inferior a 75%.

Uma justificativa para esse desempenho relativamente baixo será apresentada posteriormente. O teste de Wilcoxon aplicado às distribuições da Tabela 19 indicam diferença estatisticamente significativa entre todos os resultados obtidos.

**Tabela 19: Desempenho individual (%) do detector de novidades para a abordagem DNG-CMC (dados simulados).**

Método	Desempenho Médio
X-Médias	71,15 ± 6,85
KVMP	74,48 ± 7,41
RDVS	66,25 ± 9,88

#### 6.2.4 DESEMPENHO INDIVIDUAL DO CLASSIFICADOR MULTICLASSE

Quando considerado apenas o desempenho do classificador multiclasse para o conjunto de teste, sem incorporar as classes novidades, observa-se um desempenho médio muito próximo ao apresentado em (LAZZARETTI et al., 2009), o qual contempla apenas classes relacionadas a curto-circuito, abertura e fechamento trifásicos. Na Tabela 20 está representado o desempenho para o classificador MVS.

**Tabela 20: Desempenho individual (%) do classificador multiclasse com base em MVS (dados simulados).**

Método	Desempenho Médio
MVS	88,78 ± 5,70

#### 6.2.5 DESEMPENHO GLOBAL

Ao se estabelecer uma análise do desempenho global da abordagem DNG-CMC, é possível observar que o desempenho apresentado pela detecção de novidades em conjunto com a classificação multiclasse é bastante reduzido em relação ao desempenho da classificação multiclasse isolada, i.e. sem incluir a possibilidade de existirem novidades, como mostra a Tabela 21.

**Tabela 21: Desempenho global (%) para a abordagem DNG-CMC (dados simulados).**

Método	Desempenho Médio
X-Médias e MVS	61,31 ± 14,28
KVMP e MVS	60,27 ± 13,53
RDVS e MVS	36,72 ± 17,82

Como mostra a Tabela 21, o desempenho médio para todos os métodos é inferior a 62%. Além disso, convém ressaltar que o teste de Wilcoxon indica equivalência estatística entre os desempenhos dos modelos X-Médias e MVS e KVMP e MVS.

A inclusão do detector de novidades afeta a classificação global em dois aspectos. O primeiro deles, está relacionado à classificação de padrões normais como novidades. Para esse erro de classificação, todos os índices calculados – apresentados na Seção 5.4 – são afetados. O desempenho individual de cada classe é modificado, alterando significativamente o desempenho global e o desempenho do detector de novidades. O segundo fator de impacto é decorrente desse mesmo processo, porém está associado à classificação de novidades como padrões normais. Esse erro de classificação é tão impactante quanto o erro citado anteriormente, já que os mesmos índices de desempenho são afetados.

A maior fonte de erro para ambos os casos está associada a determinados arranjos de classes que incluem um curto-circuito trifásico como classe normal e um curto-circuito trifásico para terra como uma novidade. Esses eventos, por apresentarem grande similaridade entre si, podem ser confundidos pelo detector de novidades. O mesmo ocorre com eventos de manobra, como as aberturas e os fechamentos no alimentador. Apenas para ilustrar, um modelo MVS aplicado ao problema de separação binária entre as classes mencionadas leva a desempenhos de aproximadamente 75% de acerto. Quando se utiliza um detector de novidades, considerando uma das classes como normal e a outra como novidade, os desempenhos de classificação são da ordem de 55% de acerto – pouco superior ao desempenho aleatório de classificação binária.

Outra observação relevante está associada ao uso de eventos simultâneos e subsequentes como novidades. Para exemplificar esse caso, será apresentada aqui a matriz de confusão para um arranjo de classes considerando inicialmente apenas uma classificação multiclasse entre as classes de curto-circuito monofásico e bifásico, utilizando a abordagem DNG-CMC com o detector de novidades sendo o *X-Médias*, sem a possibilidade de novidades na classificação. Essa matriz de confusão está representada na Tabela 22. Para esse caso, o desempenho médio global entre as seis classes foi de 96,3% de acerto.

**Tabela 22: Resultado da classificação multiclasse para um arranjo contendo classes de curto-circuito monofásico e bifásico (classe atribuída representada nas colunas).**

Classes	Falta Monof. Fase A	Falta Monof. Fase B	Falta Monof. Fase C	Falta Bif. Fases AB	Falta Bif. Fases BC	Falta Bif. Fases CA
Falta Monof. Fase A	15	3	0	0	0	0
Falta Monof. Fase B	0	18	0	0	0	0
Falta Monof. Fase C	0	0	18	0	0	0
Falta Bif. Fases AB	0	0	0	17	0	1
Falta Bif. Fases BC	0	0	0	0	18	0
Falta Bif. Fases CA	0	0	0	0	0	18

Quando são inseridos eventos simultâneos e subsequentes para esse mesmo problema

– incluindo a classe de eventos caracterizados como novidades – a matriz de confusão resultante passa ser a matriz representada na Tabela 23. Nesse caso, o desempenho médio global foi de 92,5%.

**Tabela 23: Resultado da classificação multiclasse para um arranjo contendo classes de curto-circuito monofásico e bifásico incluindo detecção de novidades (classe atribuída representada nas colunas).**

Classes	Falta Monof. Fase A	Falta Monof. Fase B	Falta Monof. Fase C	Falta Bif. Fases AB	Falta Bif. Fases BC	Falta Bif. Fases CA	Novidade
Falta Monof. Fase A	15	2	1	0	0	0	0
Falta Monof. Fase B	0	17	0	0	0	0	1
Falta Monof. Fase C	0	0	18	0	0	0	0
Falta Bif. Fases AB	0	0	0	16	0	0	2
Falta Bif. Fases BC	0	0	0	0	18	0	0
Falta Bif. Fases CA	0	0	0	0	0	17	1
Novidade	6	2	24	21	0	1	322

Comparando as Tabelas 22 e 23, observa-se que o acerto por classe para os eventos de curto-circuito monofásico e bifásico reduz de maneira pouco significativa quando são incluídos os eventos simultâneos e subsequentes. Uma pequena parcela dos eventos que eram corretamente classificados (quatro no total), passa a ser considerada como *novidades*, em função da presença desses novos eventos no processo de classificação. Por outro lado, uma parcela mais significativa de *novidades* é erroneamente classificada como uma das classes conhecidas previamente (54 no total). O desempenho global para o processo de detecção de novidades na Tabela 23 é de aproximadamente 91% e pode ser determinado através da média entre os seguintes acertos individuais:

$$\text{Acerto}_{\text{normal}} = \frac{(15 + 2 + 1) + (17) + (18) + (16) + (18) + (17)}{108} = 96,3\%$$

$$\text{Acerto}_{\text{novidade}} = \frac{322}{376} = 85,6\%,$$

Com base no desempenho individual do detector de novidades e na comparação entre os desempenhos globais das Tabelas 22 e 23, pode-se afirmar que, para o arranjo de classes em questão, é possível utilizar o classificador proposto para identificar eventos simultâneos e subsequentes, já que o impacto final na classificação é pouco significativo em termos globais.

No entanto, existe um último fator relevante na análise aqui apresentada, relacionado ao próprio arranjo de classes conhecidas previamente. Determinados arranjos de classes promovem um desempenho global muito superior em relação a outros arranjos. Por exemplo, o arranjo de classes considerando curtos-circuitos como classe normal e os demais eventos considerados novidades (incluindo eventos simultâneos e subsequentes), apresenta um desempenho global de mais de 89% de acerto. No entanto, um arranjo que considera apenas as aberturas monofásicas como conhecidas e todos os demais eventos como novidades, apresenta um desempenho global inferior à 40%. Um desempenho similar é obtido se eventos de manobra forem considerados como conhecidos e eventos simultâneos e subsequentes como novidades. Mesmo não refletindo um cenário realizável na prática, essa observação auxilia na interpretação dos resultados apresentados.

#### 6.2.6 ABORDAGEM DNI-CBS

Para avaliação da DNI-CBS foram utilizados os mesmos 100 diferentes arranjos de classes conhecidas *a priori* adotados para a abordagem DNG-CMC. Desse modo, é possível realizar a comparação de desempenho entre essas duas abordagens.

No caso da abordagem DNI-CBS serão apresentados os resultados para o detector de novidades individualmente e os desempenhos globais considerando o classificador multiclasse acoplado.

#### 6.2.7 DESEMPENHO INDIVIDUAL DO DETECTOR DE NOVIDADES

Considerando que os modelos uma-classe são utilizados apenas para separar padrões normais de novidades, a Tabela 24 resume o desempenho médio e o desvio padrão associado para os 100 diferentes arranjos de classes selecionados. Em comparação com a Tabela 19 através do teste de Wilcoxon, pode-se afirmar que somente o método RDVS obteve um incremento no seu desempenho. Isso implica que esse método, quando parametrizado de acordo com uma distribuição individualizada por classe, aumenta a possibilidade de identificar corretamente

padrões normais e novidades. Os modelos X-Médias e KVMP apresentaram decréscimo de desempenho no caso DNI-CBS. Na comparação global da Tabela 24, o modelo RDVS apresentou o melhor desempenho, quando comparado com os demais métodos, nas duas abordagens, através do teste de Wilcoxon.

**Tabela 24: Desempenho individual do detector de novidades para a abordagem DNI-CBS (dados simulados).**

Método	Desempenho Médio (%)
X-Médias	63,79 ± 5,86
KVMP	69,76 ± 4,27
RDVS	76,21 ± 7,22

### 6.2.8 DESEMPENHO GLOBAL

Considerando agora o desempenho global do método com o classificador multiclasse acoplado (Tabela 25), observa-se uma melhoria para todos os métodos (estatisticamente significativa pelo teste de Wilcoxon), ao se comparar os resultados com os desempenhos apresentados na Tabela 21.

**Tabela 25: Desempenho global da abordagem DNI-CBS (dados simulados).**

Método	Desempenho Médio (%)
X-Médias	71,09 ± 7,71
KVMP	68,64 ± 7,99
RDVS	51,12 ± 12,28

As mesmas observações feitas para o método DNG-CMC, com relação aos erros de classificação incluindo a detecção de novidades, valem para a abordagem DNI-CBS. No entanto, quando comparados os desempenhos globais entre as duas abordagens, o modelo X-Médias com MVS apresenta o melhor desempenho na abordagem DNI-CBS, considerando a comparação estatística através do teste de Wilcoxon.

### 6.2.9 CONCLUSÕES INICIAIS PARA OS DADOS SIMULADOS

A análise preliminar dos resultados dos dados simulados fornece algumas conclusões iniciais em relação à classificação multiclasse e detecção de novidades para formas de onda de tensão em redes de distribuição.

A primeira delas mostra que o pré-processamento baseado em TWD (ou TWPD) também pode ser utilizado quando classes desconhecidas no estágio de treinamento surgem no



estágio de teste. É importante destacar que essa forma de pré-processamento, combinada com o cálculo de energia para cada nível de decomposição, é bastante usual para classificação multiclasse supervisionada em oscilografias de tensão (LAZZARETTI, 2009). No entanto, convém ressaltar que essa não é uma análise definitiva, sendo que mais detalhes referentes a essa etapa devem ser verificados, de acordo com a finalidade de classificação e os dados disponíveis. Posteriormente, na análise de desempenho para dados reais (Seção 6.3), essa questão será retomada e melhor detalhada.

Adicionalmente, a inclusão da etapa de detecção de novidades impacta na classificação como um todo, reduzindo o desempenho global de classificação, principalmente por impor superfícies fechadas para as classes conhecidas, ao contrário do que se observa em classificadores multiclasse supervisionados, os quais visam apenas a discriminação entre múltiplas classes conhecidas. Para realizar essa tarefa, todos os detectores de novidade utilizam como parâmetro de entrada a fração de padrões normais que podem ser rejeitados durante o treinamento. Esse parâmetro influencia diretamente o resultado final de classificação, indicando que alguns padrões normais do conjunto de treinamento, e conseqüentemente do conjunto de teste, serão classificados como novidades.

O arranjo de classes conhecidas *a priori* está diretamente associado ao desempenho final de classificação. Determinados arranjos podem fornecer um desempenho final superior. Quando se utiliza a separação entre classes de forma supervisionada, através de um classificador baseado em MVS, podem-se obter superfícies de decisão relativamente complexas para separação de classes localizadas muito próximas no espaço de características, mantendo uma boa capacidade de generalização para o modelo. No entanto, do ponto de vista de classificadores uma-classe, essa tarefa é dificultada, pois não se pode considerar a informação de uma das classes, resultando em superfícies de decisão com menor capacidade de generalização. Esse é o caso observado para as classes de manobra (abertura e fechamento) e curto-circuito trifásico, as quais possuem grande similaridade, mesmo quando comparadas visualmente por especialistas.

Uma forma de contornar tal limitação é utilizar os modelos que consideram exemplos de outras classes previamente conhecidas para determinar as fronteiras finais de decisão, a saber: RDVS com exemplos negativos e RDVS com exemplos negativos e otimização da margem. Maiores detalhes sobre a utilização desses modelos serão apresentados nas próximas seções, dedicadas à análise de dados reais (Seção 6.3).

A detecção de novidades aqui proposta pode auxiliar na identificação de eventos simultâneos e subsequentes. No entanto, isso depende fortemente do arranjo de classes disponíveis inicialmente e da característica de simultaneidade. De um modo geral, o processo de

classificação por si só não é a forma mais adequada de identificar eventos com essas características. Tal diferenciação deve ser mais bem analisada durante a extração de características dos sinais, uma vez que se deve procurar por características que ressaltem as dissimilaridades, principalmente para eventos simultâneos. Ao se aplicar a TWD e o cálculo de energia aqui proposto, a informação de ambas as classes acaba sendo combinada, dificultando a caracterização completa do fenômeno. Estudos recentes demonstram que uma forma possível de estabelecer tal diferenciação, seria aplicar a Análise das Componentes Independentes como parte do processo de extração de características (LIMA et al., 2014). Com isso, garante-se ao longo do processo restrições para facilitar a distinção entre eventos isolados, subsequentes e simultâneos, facilitando a classificação final e levando a resultados globais superiores.

Ambas as abordagens propostas, DNG-CMC e DNI-CBS, podem ser utilizadas para a classificação em questão. No entanto, a abordagem DNI-CBS se mostrou mais adequada para a classificação de um modo geral. Isso pode ser função da características dos dados no espaço definido pelas características. Como a quantidade de classes é relativamente alta, é de se esperar que a extensão ocupada pelas classes seja proporcionalmente alta no espaço de características. Desse modo, construir um detector de novidades agrupando todas as classes conhecidas (abordagem DNG-CMC) é comparativamente mais difícil do que separar o problema para cada classe (abordagem DNI-CBS), principalmente devido ao fato das novas classes poderem ocupar regiões intermediárias entre as classes no espaço de características (ver Figura 2).

### 6.3 CLASSIFICAÇÃO E DETECÇÃO DE NOVIDADES EM DADOS REAIS

Os resultados apresentados para os dados simulados fornecem apenas uma ideia inicial para o processo de classificação aqui proposto. Para avaliar e comparar o desempenho dos classificadores visando à aplicação final da metodologia em um sistema real, a presente seção discute os resultados incluindo dados reais do sistema de monitoramento apresentado.

Sendo assim, a mesma metodologia adotada para a análise de desempenho para eventos simulados foi empregada para os dados reais. A principal diferença é que, no caso dos dados reais, as classes conhecidas já estão definidas, de forma que o parâmetro de número de classes conhecidas *a priori* não foi variado nesse caso.

A seguir, é apresentada uma análise inicial das diferentes abordagens de classificação aplicadas aos dados reais, focando nos modelos apresentados para os dados simulados (DNG-CMC e DNI-CBS), bem como diferentes formas de pré-processar o sinal. Na sequência serão discutidas diferentes formas de melhorar o desempenho para o conjunto de dados apresentado.

### 6.3.1 ANÁLISE INICIAL DOS PARÂMETROS PARA OS DADOS REAIS

Inicialmente, os conjuntos de treinamento e de teste foram fixados de modo a utilizar aproximadamente 80% dos padrões disponíveis para treinamento e o restante para teste, de acordo com a distribuição de dados apresentada na Subseção 5.4.3, incluindo padrões simulados e reais. Foi selecionado o subconjunto mais representativo para treinamento, incluindo diferentes situações de curto-circuito, manobra e descarga atmosférica, tanto para os dados reais, quanto para os dados simulados. Os conjuntos de treinamento e teste utilizados nessa análise inicial estão detalhados a seguir:

- Curto-Circuito Monofásico na Fase A: 18 exemplos simulados e 5 exemplos reais para treinamento; 5 exemplos simulados e 2 exemplos reais para teste;
- Curto-Circuito Monofásico na Fase B: 18 exemplos simulados e 1 exemplo real para treinamento; 5 exemplos simulados e 1 exemplo real para teste;
- Curto-Circuito Monofásico na Fase C: 18 exemplos simulados e 13 exemplos reais para treinamento; 5 exemplos simulados e 4 exemplos reais para teste;
- Religamento Trifásico: 9 exemplos simulados e 8 exemplos reais para treinamento; 4 exemplos simulados e 4 exemplos reais para teste;
- Descarga Atmosférica Induzida: 68 exemplos reais para treinamento; 17 exemplos reais para teste;
- Novidades: demais classes, com 26 exemplos reais, utilizados apenas na fase de teste.

### 6.3.2 ABORDAGEM DNG-CMC

Na Tabela 26 estão representados os resultados obtidos para o estágio de detecção de novidades da abordagem DNG-CMC, em diferentes métodos de classificação e diferentes formas de pré-processar o sinal. A codificação adotada para representar o pré-processamento segue a mesma metodologia apresentada na Figura 73. Nesta tabela, destaca-se o desempenho máximo obtido para detecção de novidades utilizando o método KVMP com 78,26% de acerto, para o pré-processamento *245.760-twd-db20*. Essa forma de pré-processamento merece destaque, pois os desempenhos obtidos foram acima de 70% de acerto para todos os métodos avaliados.

**Tabela 26: Desempenho individual (%) do detector de novidades para a abordagem DNG-CMC utilizando dados reais e simulados.**

Pré-Processamento	X-Médias	KVMP	RDVS
245.760-twd-db20	74,84	<b>78,26</b>	71,47
245.760-twpd-db20	70,64	63,99	57,62
122.880-twd-db20	76,97	77,27	69,47
122.880-twpd-db20	47,53	59,19	58,41

Na Tabela 27 são demonstrados os resultados para o classificador multiclasse MVS aplicado aos dados reais (conjunto de teste), sem considerar as classes rotuladas como novidades. Observa-se que para o pré-processamento *245.760-twd-db20*, o desempenho foi de aproximadamente 76% de acerto.

**Tabela 27: Desempenho individual do classificador MVS utilizando dados reais e simulados.**

Pré-Processamento	MVS (%)
<b>245.760-twd-db20</b>	<b>76,37</b>
245.760-twpd-db20	71,40
122.880-twd-db20	68,99
122.880-twpd-db20	63,55

Analisando o desempenho global da abordagem DNG-CMC (Tabela 28), destaca-se que o desempenho de 64,63% de acerto para o método X-Médias, com o pré-processamento *245.760-twd-db20*.

**Tabela 28: Desempenho global (%) para a abordagem DNG-CMC utilizando dados reais e simulados.**

Pré-Processamento	X-Médias e MVS	KVMP e MVS	RDVS e MVS
245.760-twd-db20	<b>64,63</b>	59,77	54,97
245.760-twpd-db20	63,64	54,02	48,43
122.880-twd-db20	54,07	55,92	53,08
122.880-twpd-db20	43,73	33,69	41,17

### 6.3.3 ABORDAGEM DNI-CBS

Considerando agora a abordagem DNI-CBS, o desempenho máximo é de 71,96% de acerto para o pré-processamento *245.760-twd-db20* com o método KVMP, de acordo com a Tabela 29, que é ligeiramente inferior ao desempenho obtido na abordagem DNG-CMC (KVMP com 78,26% de acerto).

Se o modelo uma-classe construído para cada classe for utilizado também como classificador multiclasse, obtêm-se os resultados demonstrados na Tabela 30.

**Tabela 29: Desempenho individual (%) do detector de novidades para a abordagem DNI-CBS utilizando dados reais e simulados.**

Pré-Processamento	X-Médias	KVMP	RDVS
245.760-twd-db20	63,66	<b>71,96</b>	61,39
245.760-twpd-db20	42,16	47,31	66,74
122.880-twd-db20	64,54	65,17	69,26
122.880-twpd-db20	53,50	46,71	62,90

**Tabela 30: Desempenho global (%) da abordagem DNI-CBS utilizando dados reais e simulados.**

Pré-Processamento	X-Médias	KVMP	RDVS
245.760-twd-db20	44,86	33,87	35,54
245.760-twpd-db20	47,16	54,80	45,33
122.880-twd-db20	47,76	50,62	48,01
122.880-twpd-db20	53,51	46,45	40,37

Para a abordagem DNI-CBS, não se verificam desempenhos globais superiores a 55% em todos os casos avaliados. O desempenho da abordagem DNG-CMC foi superior tanto para o detector de novidades, quanto para o processo de classificação como um todo. No caso da abordagem DNG-CMC, o melhor desempenho para a detecção de novidades foi com o método KVMP, com acerto de 78,26%. Já para o processo de classificação multiclasse e detecção de novidades em conjunto, o melhor desempenho foi de 64,63% de acerto para o método X-Médias. De modo geral, o pré-processamento mais indicado para os dados reais está fundamentado em uma frequência de amostragem de 245.760 Hz, com o uso da TWD e cálculo de relação de energia entre os diferentes níveis.

#### 6.3.4 ANÁLISE DE MELHORIA DE DESEMPENHO PARA OS DADOS REAIS

Realizada a análise preliminar e definida a forma de pré-processar o sinal (245.760-twd-db20), optou-se por identificar possíveis formas de incrementar os desempenho globais e do detector de novidades para os dados reais. Além disso, foi feita uma análise das possíveis causas dos baixos desempenhos obtidos em determinados casos.

A primeira modificação realizada foi incluir o módulo e a fase do sinal no primeiro ciclo após a ocorrência do evento, em relação ao módulo e fase do sinal no ciclo imediatamente anterior à ocorrência do evento. Dessa modo, é possível obter as diferenças do fasor de tensão nos instantes de interesse. Essa característica permite identificar de forma mais apropriada os curtos-circuitos (principalmente), de acordo com a classificação ABC (BOLLEN, 1999). O cálculo do módulo e fase são baseados na estimação através do Filtro de Kalman formulado na Seção 4.2 e discutido na Seção 4.6. Com isso, o vetor de características passa a conter as

seguintes informações:

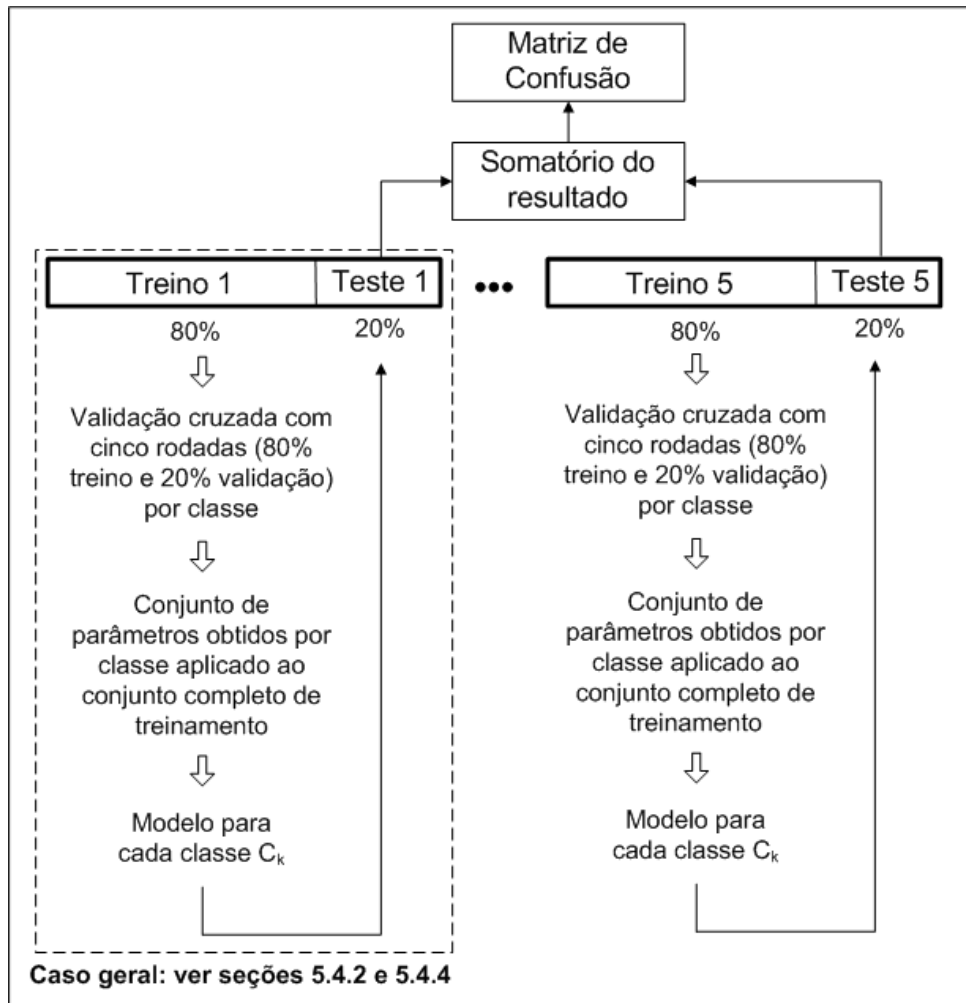
- Relação de energia de cada nível da TWD em nove níveis, conforme apresentado nas seções 4.5 e 5.4.3, totalizando dez características (nove características para sinal detalhe e um para o sinal aproximação) por fase;
- Relação do módulo e fase do sinal de tensão trifásica, calculados através do filtro de Kalman descrito na seção 4.2 – duas novas características por fase;
- Número total de características do vetor: 36.

Além disso, os resultados aqui apresentados foram repetidos para diferentes conjuntos de treinamento e teste, escolhidos aleatoriamente com base na distribuição de dados apresentada no início da Subseção 6.3.1, mantendo a proporção de 80% dos dados para treinamento e 20% para teste, de modo a facilitar a comparação final, visando escolher o melhor método de classificação. Um resumo desse processo está apresentado na Figura 74.

Com a inclusão do módulo e fase, o desempenho para o problema de classificação multiclasse, somente com classes conhecidas *a priori*) e com o modelo MVS, passa a ser de aproximadamente 96% de acerto. As matrizes de confusão para ambos os casos estão representadas nas Tabela 31 e 32.

Ao se comparar os desempenhos acima apresentados, observa-se um incremento de aproximadamente 20% de acerto no desempenho total ao se inserir informações de módulo e fase do sinal. Isso é função da melhoria de desempenho individual das classes de curto-circuito, que passam a ser corretamente classificadas ao se inserir a informação de módulo e fase, ao invés de serem incorretamente classificadas como relacionadas a descargas atmosféricas (maior fonte de erro de classificação para os resultados apresentados na Tabela 31).

Uma possível justificativa para a sobreposição existente entre as classes de curto-circuito e descarga atmosférica com o pré-processamento *245.760-twd-db20* está relacionada ao conteúdo de frequência observado nos sinais registrados para descargas atmosféricas. De um modo geral, curtos-circuitos possuem um estado transitório inicial de baixa frequência (se comparado à descarga atmosférica), seguido de um estado quase-estacionário caracterizado por uma variação de tensão de curta duração. Já a descarga atmosférica é completamente caracterizada pela estado transitório de relativa alta frequência. Em um cenário ideal a própria energia nos diferentes níveis da TWD seria suficiente para realizar a separação entre essas classes. Porém, na maioria dos registros reais disponíveis para os experimentos aqui apresentados, o conteúdo de frequência dos sinais de descarga atmosférica estão abaixo dos valores sugeridos



**Figura 74: Validação cruzada para comparação final.** Inicialmente, o conjunto completo de dados é dividido em cinco diferentes conjuntos de treino e teste definidos aleatoriamente. Para cada conjunto de treinamento é feita a validação cruzada (cinco rodadas) para definir os parâmetros e, com base no modelo selecionado, efetua-se a avaliação no conjunto de teste. A matriz de confusão final é obtida através da soma de cada um dos cinco conjuntos de teste que, no total, serão todos os exemplos disponíveis inicialmente.

na literatura (BOLLEN, 1999). Esse conteúdo de frequência relativamente baixo pode estar associado à distância das descargas atmosféricas induzidas registradas – superiores a 200 metros na maioria dos casos, conforme mostra Lazzaretti et al. (2011).

Essa mesma alteração (introdução do módulo e fase) foi aplicada para a classificação como um todo, considerando a detecção de novidades, através das abordagens DNG-CMC e DNI-CBS. Para ambas as abordagens, o procedimento apresentado na Figura 74 é repetido dez vezes, de modo a obter o resultado médio de cada classificador. Também é apresentado o desempenho para o conjunto de treinamento, de modo a verificar a capacidade de generalização de cada modelo.

A Tabela 33 mostra os resultados para a abordagem DNG-CMC, enquanto a Tabela 34

**Tabela 31: Matriz de confusão para a classificação multiclasse isolada com o modelo MVS e o pré-processamento 245.760-twd-db20 sem módulo e fase (classe atribuída representada nas colunas). Nesse caso, o desempenho médio global é de aproximadamente 76%.**

Classes	Falta Monof. Fase A	Falta Monof. Fase B	Falta Monof. Fase C	Religamento Trifásico	Descarga Atmosférica Induzida
Falta Monof. Fase A	25	0	0	0	5
Falta Monof. Fase B	0	17	0	0	8
Falta Monof. Fase C	0	0	31	2	7
Religamento Trifásico	0	0	0	25	0
Descarga Atmosférica Induzida	9	13	17	0	45

**Tabela 32: Matriz de confusão para a classificação multiclasse isolada com o modelo MVS e o pré-processamento 245.760-twd-db20 com módulo e fase (classe atribuída representada nas colunas). Nesse caso, o desempenho médio global é de aproximadamente 96%.**

Classes	Falta Monof. Fase A	Falta Monof. Fase B	Falta Monof. Fase C	Religamento Trifásico	Descarga Atmosférica Induzida
Falta Monof. Fase A	30	0	0	0	0
Falta Monof. Fase B	0	25	0	0	0
Falta Monof. Fase C	0	0	35	0	5
Religamento Trifásico	0	0	0	25	0
Descarga Atmosférica Induzida	0	0	5	0	80

apresenta os desempenhos para o modelo DNI-CBS. Em ambos os casos, observa-se um desempenho médio superior aos desempenhos obtidos utilizando apenas o pré-processamento



245.760-*twd-db20*. Em particular, o incremento de desempenho está associado ao desempenho global, uma vez que eventos de curto-circuito passam a ser mais bem caracterizados com a inserção de módulo e fase, incrementando o acerto por classe para as classes conhecidas previamente.

**Tabela 33: Detecção de novidades e classificação multiclasse com a abordagem DNG-CMC.**

Modelo	Desempenho Global (%)		Desempenho Individual do Detector de Novidades (%)	
	Treino	Teste	Treino	Teste
X-Médias	78,15 ( $\pm 1,89$ )	69,11 ( $\pm 1,78$ )	83,56 ( $\pm 1,43$ )	73,08 ( $\pm 3,34$ )
KVMP	93,44 ( $\pm 3,66$ )	78,15 ( $\pm 3,80$ )	91,23 ( $\pm 2,34$ )	78,34 ( $\pm 4,12$ )
RDVS	80,70 ( $\pm 4,54$ )	71,26 ( $\pm 5,60$ )	79,98 ( $\pm 2,97$ )	70,74 ( $\pm 5,56$ )

**Tabela 34: Detecção de novidades e classificação multiclasse com a abordagem DNI-CBS.**

Modelo	Desempenho Global (%)		Desempenho Individual do Detector de Novidades (%)	
	Treino	Teste	Treino	Teste
X-Médias	68,11 ( $\pm 2,27$ )	59,03 ( $\pm 7,41$ )	65,12 ( $\pm 3,36$ )	62,20 ( $\pm 3,40$ )
KVMP	95,42 ( $\pm 1,12$ )	79,04 ( $\pm 6,37$ )	86,37 ( $\pm 3,44$ )	76,81 ( $\pm 3,54$ )
RDVS	79,21 ( $\pm 2,55$ )	75,57 ( $\pm 5,69$ )	77,05 ( $\pm 2,42$ )	74,11 ( $\pm 3,70$ )

Até o momento, para o caso particular da abordagem DNI-CBS, todos os modelos uma-classe avaliados não consideraram exemplos das demais classes conhecidas para definição das superfícies de separação. Sendo assim, com o intuito de explorar em detalhes as diferentes formulações do modelo RDVS que incluem tal possibilidade, assinalando as vantagens obtidas por tais formulações, foram incluídas as duas versões apresentadas no capítulo de fundamentação teórica (Subseções 4.8.5, 4.8.8 e 4.8.9):

- RDVS com exemplos negativos (RDVS-N);
- RDVS com exemplos negativos e otimização da margem (RDVS-M).

Os resultados para esses modelos estão apresentados na Tabela 35. Comparando os resultados dessa tabela, com as Tabelas 33 e 34, é possível afirmar, com base no teste de Wilcoxon, que o modelo RDVS-M fornece os melhores resultados para ambas as métricas avaliadas, ultrapassando 80% nos dois casos.

É possível notar, pela análise da Tabela 35, um incremento significativo no resultado médio ao se inserir, primeiramente, os exemplos negativos na formulação (RDVS-N) e, posteriormente, a separação desses exemplos com margem máxima (RDVS-M).

**Tabela 35: Detecção de novidades e classificação multiclasse com a abordagem DNI-CBS com modelos baseados na RDVS.**

Modelo	Desempenho Global (%)		Desempenho Individual do Detector de Novidades (%)	
	Treino	Teste	Treino	Teste
RDVS	79,21 ( $\pm 2,55$ )	75,57 ( $\pm 5,69$ )	77,05 ( $\pm 2,42$ )	74,11 ( $\pm 3,70$ )
RDVS-N	86,93 ( $\pm 2,11$ )	80,04 ( $\pm 5,91$ )	84,62 ( $\pm 4,22$ )	80,65 ( $\pm 5,03$ )
RDVS-M	86,33 ( $\pm 3,91$ )	<b>83,19</b> ( $\pm 6,42$ )	86,72 ( $\pm 4,81$ )	<b>85,01</b> ( $\pm 5,34$ )

Ao se comparar os desempenhos dos conjuntos de treinamento e de teste, pode-se observar que o modelo RDVS-M apresenta a melhor generalização (i.e. a menor diferença de acerto entre treinamento e teste) para ambos os desempenhos avaliados. Isso demonstra que, através da minimização do raio da hipersfera contendo os exemplos de uma dada classe conhecida e, ao mesmo tempo, maximizando a margem de separação entre essa classe e as demais classes conhecidas, obtém-se um incremento tanto no desempenho global, quanto na detecção de novidades, principalmente através da melhoria na capacidade de generalização do modelo. Esta pode ser considerada a maior vantagem do termo de separação com margem máxima, que é a base da formulação do modelo RDVS-M e sua principal diferença em relação ao modelo RDVS-N.

Com o intuito de analisar os detalhes do resultado de classificação do modelo RDVS-M, a matriz de confusão média obtida para os experimentos é apresentada na Tabela 36. Essa matriz considera, para cada experimento, a soma dos resultados do conjunto de teste para cada rodada do processo de validação cruzada (com um total de cinco rodadas). Isso justifica a quantidade de elementos observados para cada classe, em relação à distribuição inicialmente apresentada. Essa representação facilita a visualização dos erros de classificação mais proeminentes. Nesse caso, o desempenho global é de aproximadamente 83% de acerto, enquanto o desempenho do detector de novidades é de 85% de acerto (valores correspondentes à média demonstrada na Tabela 35).

**Tabela 36: Matriz de confusão para o modelo RDVS-M.**

Classe Verdadeira	Classe Atribuída					
	Falta Monof. (A)	Falta Monof. (B)	Falta Monof. (C)	Relig. Trif.	Desc. Atm.	Novidade
Falta Monof. (A)	26	0	0	0	0	4
Falta Monof. (B)	0	22	0	0	0	3
Falta Monof. (C)	0	0	32	1	3	4
Relig. Trif.	0	0	0	21	0	4
Desc. Atm.	0	0	8	3	66	8
Novidade	4	0	5	5	9	107

Assim como observado para os dados simulados, uma parcela dos eventos relacionados a classes conhecidas previamente é erroneamente classificada como novidade. Essa é a parcela mais significativa dos erros verificados. Já para as novidades, o mesmo conjunto (com 26 exemplos) é aplicado para cada uma das cinco rodadas da validação cruzada. Com base neste fato, uma observação importante pode ser extraída a partir da matriz de confusão: a maior parte dos erros de classificação dessa classe são devidos a exemplos desconhecidos previamente (novidades) classificados como relacionados a descargas atmosféricas.

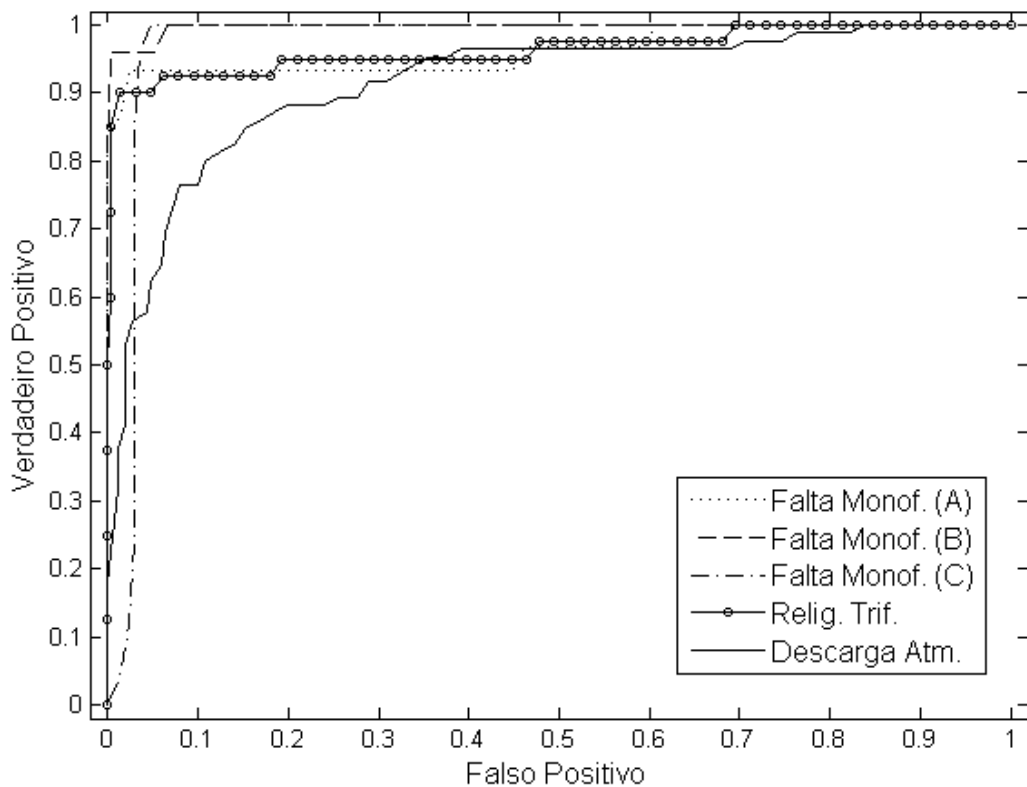
A Tabela 37 mostra os desempenhos por classe para essa matriz de confusão. Destaca-se que a classe com pior desempenho é a classe de eventos associados com Descarga Atmosférica.

**Tabela 37: Desempenho por classe para o modelo RDVS-M.**

Classe	Desempenho (%)
Falta Monof. (A)	86,67
Falta Monof. (B)	88,00
Falta Monof. (C)	80,00
Relig. Trif.	84,00
Desc. Atm.	77,65
Novidade	82,31

Os resultados apresentados na Tabela 37 fornecem apenas uma ideia inicial para o resultado de classificação, utilizando um dado limiar de exemplos normais que devem ser rejeitados (parâmetro  $C_1$  na formulação RDVS) e novidades que podem ser aceitas pelo modelo (parâmetro  $C_2$  na formulação RDVS), os quais foram selecionados via validação cruzada. No entanto, para verificar a saída do classificador definido para cada classe com diferentes limiares, é necessário combinar os resultados da Tabela 37 com a análise ROC por classe. Nessa análise, diferentes erros de classificação podem ser analisados de maneira conjunta, conforme demonstra Tax (2001). A Figura 75 mostra a curva ROC média para diferentes limiares para cada classe, utilizando os mesmos conjuntos de teste utilizados para definir os resultados mostrados na Tabela 36.

Na Figura 75 é possível observar diferentes características de acordo com cada classe. A AAC para a classe relacionada a descargas atmosféricas (pior desempenho relativo) é 0,91. Para as demais classes, a AAC é superior a 0,96. Para a classe curto-circuito monofásico na fase C, podem haver algumas novidades ou exemplos de outras classes que se sobrepõem a essa classe, uma vez que a taxa de falsos positivos é relativamente elevada, mesmo para baixos valores de limiar. Esse resultado também pode ser observado na matriz de confusão da Tabela 36. Para altas taxas de falsos positivos na curva ROC, existem alguns erros significativos,



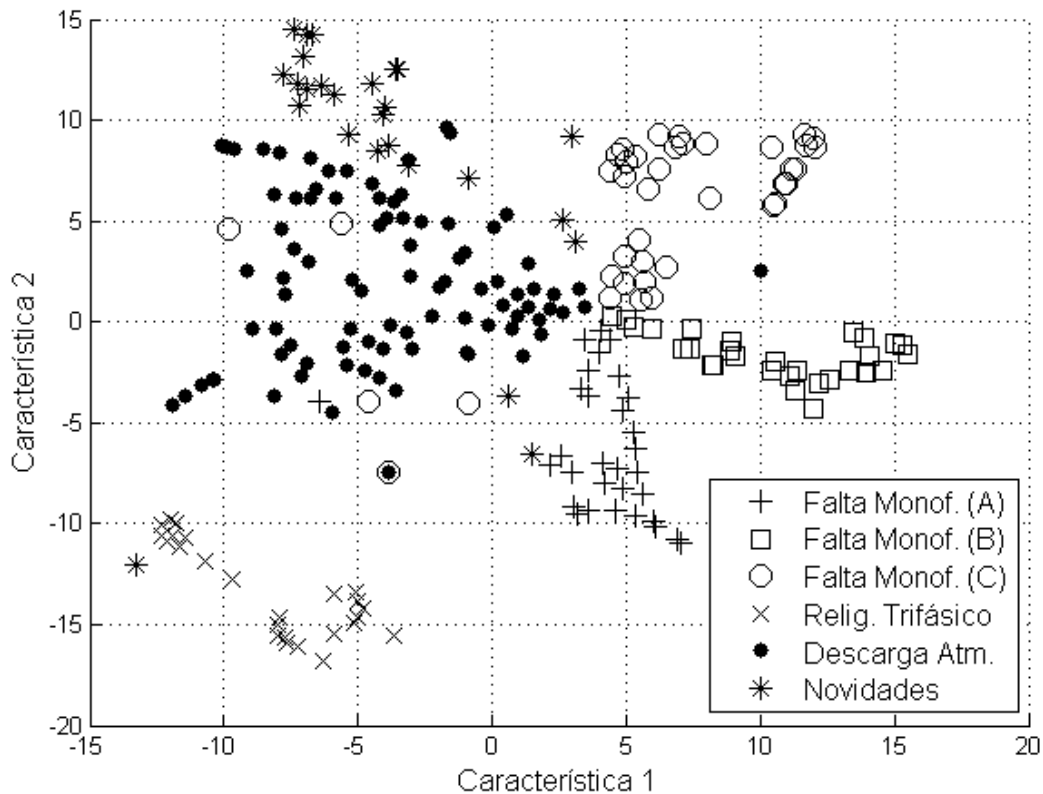
**Figura 75: Análise ROC para as cinco classes conhecidas. A análise da curva ROC da classe relacionada com Descarga Atmosférica confirma seu desempenho relativamente baixo, como apresentado na Tabela 37.**

principalmente para a classe relacionada a descargas atmosféricas. Isso significa que alguns exemplos dessa classe estão localizados relativamente distantes no espaço de características. Como conclusão geral, podemos afirmar que a classe relacionada a descargas atmosféricas é a classe mais impactante nos erros globais de classificação.

Algumas das conclusões apresentadas nas Tabelas 36 e 37, bem como na Figura 75, podem ser melhor observadas através de um método de redução de dimensionalidade. A Figura 76 mostra o resultado do método conhecido como *t-Distributed Stochastic Neighbor Embedding* (MAATEN; HINTON, 2008), aplicado ao conjunto de dados aqui analisado. A vantagem desse método é que ele utiliza um processo de otimização que visa manter a estrutura dos dados no novo espaço projetado, de forma que as principais características observadas no espaço original sejam mantidas no novo espaço de representação.

O algoritmo compreende duas etapas principais. Em um primeiro momento, o método estima uma distribuição de probabilidade sobre pares de objetos no espaço original (de alta dimensionalidade), de tal maneira que os objetos similares têm uma elevada probabilidade de serem escolhidos, enquanto pontos dissimilares têm uma probabilidade infinitesimal de serem

escolhidos. Em um segundo momento, o método define uma probabilidade de distribuição semelhante ao longo dos pontos no espaço de baixa dimensionalidade e minimiza a divergência de KullBack-Leibler<sup>1</sup> entre as duas distribuições, em relação às localizações dos pontos em ambas as representações.



**Figura 76: Visualização em duas dimensões das cinco classes conhecidas previamente e as novidades utilizando o método *t-Distributed Stochastic Neighbor Embedding*. Como pode ser observado, é possível identificar a estrutura de cada classe e também as sobreposições existentes, principalmente para as novidades e a classe relacionada a descargas atmosféricas. Além disso, os exemplos classificados como novidade estão concentrados em regiões específicas do espaço de características. Isso está relacionado à fração de novidades que estavam disponíveis no conjunto de dados e não representa uma generalização para o caso estudado.**

A primeira observação sobre o resultado apresentado na Figura 76 está relacionada à proximidade entre as novidades e a classe relacionada a descargas atmosféricas. Convém destacar que os 26 exemplos de novidades são definidos como tal por apresentarem uma das seguintes características: falta de registros prévio para sua caracterização, dificuldade na interpretação do fenômeno ou desconhecimento com relação à ocorrência desse evento com a forma de onda registrada. Em algumas das formas de onda registradas, os especialistas indicaram possível

<sup>1</sup>A divergência de KullBack-Leibler é uma medida não-simétrica da diferença entre duas distribuições de probabilidade  $P$  e  $Q$ , sendo que o seu resultado informa a quantidade de informação que é perdida quando  $Q$  é utilizada para aproximar  $P$ .

correlação com descarga atmosférica, principalmente curtos-circuitos que ocorrem de maneira simultânea a descargas atmosféricas. Porém, dada a incerteza na classificação, esses eventos foram caracterizados como desconhecidos. Isso pode justificar a proximidade entre alguns desses eventos e as classes de curto-circuito monofásico (fase C) e descargas atmosféricas – principal fonte de erro nesse caso. De modo geral, tal qual foi observado para os dados simulados, a caracterização de eventos simultâneos (e subsequentes) deve ser melhor abordada na etapa de pré-processamento, para que a classificação final possa ser melhor definida. Assinala-se que tal abordagem foge ao escopo do presente trabalho.

Além disso, pode-se observar que a estrutura de cada classe sugere uma distribuição complexa dos dados para o uso de métodos de classificação de uma-classe, uma vez que os dados de uma mesma classe podem ocupar diferentes regiões do espaço de características, com diferentes distribuições – particularmente a classe relacionada a descargas atmosféricas. Para modelos com base na RDVS, os quais contêm um termo que penaliza padrões mais distantes, essa característica pode dificultar a classificação final.

Adicionalmente, diferentes registros de descarga atmosférica foram obtidos em campo. Alguns relacionados com tensões induzidas por descargas atmosféricas localizadas relativamente distantes dos pontos de medição (distâncias elétricas superiores a cinco quilômetros). Sendo assim, o conteúdo de frequência observado difere bastante (isto é, consideravelmente inferior) dos casos em que as tensões são induzidas por descargas próximas (COORAY, 2003), conforme demonstrado em Lazzaretti et al. (2011). Isso pode justificar a presença de exemplos em regiões distintas no espaço de características – ou mesmo no espaço projetado pelo algoritmo *t-Distributed Stochastic Neighbor Embedding*.

Com relação aos eventos de curto-circuito, a presença de eventos com diferentes resistências de falta, incluindo resistências relativamente altas (SARLAK; SHAHRTASH, 2011), pode justificar a presença de eventos em diferentes regiões do espaço, eventualmente formando mais de um grupo no espaço de características, como pode ser observado na Figura 76 para a classe curto-circuito monofásico na fase C. Assinala-se que os dados simulados ocupam as mesmas regiões que os dados reais avaliados, não impactando no resultado aqui apresentado como previamente estabelecido, dada a validação realizada para esses dados.

### 6.3.5 CONCLUSÕES PARA OS DADOS REAIS

Ao se analisar os resultados obtidos para os dados reais, duas conclusões fundamentais podem ser obtidas:

- A inclusão do módulo e fase no vetor de características impacta na classificação de classes conhecidas, e conseqüentemente no desempenho global dos métodos DNG-CMC e DNI-CBS, uma vez que auxilia na separação de classes de curto-circuito e eventos relacionados a descargas atmosféricas;
- A utilização de modelos que consideram exemplos de outras classes, como exemplos negativos na formulação (RDVS-N) e, principalmente, a separação desses exemplos com margem máxima (RDVS-M), pode contribuir nos desempenhos globais e do detector de novidades dos métodos apresentados.

De maneira geral, a escolha final para o pré-processamento, extração de características, abordagem e modelo de classificação pode ser resumida em:

- Pré-processamento e extração de características: Baseado na TWD e na relação de energia para cada nível, utilizando a *wavelet*-mãe *db20*, na frequência de amostragem de 245.760Hz e a inclusão de módulo e fase dos sinais de cada fase;
- Abordagem de classificação: Feita de forma individual por classe, seguindo a metodologia aqui denominada de DNI-CBS;
- Modelo de classificação: Representação de Dados Utilizando Vetores Suporte com Exemplos Negativos e Otimização da Margem (RDVS-M).

Com base na escolha apresentada, é possível afirmar que a classificação multiclasse em conjunto com a detecção de novidades é viável para classificação de oscilografias de tensão em redes de distribuição, fornecendo resultados relevantes do ponto de vista prático (por exemplo, acima de 80% de acerto), tanto para a classificação multiclasse, quanto para a detecção de novidades, sendo que a análise aqui proposta pode ser estendida para outras finalidades de classificação de formas de onda em SEP.

#### 6.4 IDENTIFICAÇÃO DE NOVAS CLASSES

Até o momento, foi dada ênfase apenas na classificação dos eventos, seja ela relacionada à identificação da classe associada ao padrão, ou com a possibilidade desse padrão ser uma novidade. O objetivo dessa seção é demonstrar os resultados do processo de identificação de classes, com base no agrupamento automático aplicado aos padrões classificados como novidades, de forma que os agrupamentos encontrados possam ser um indicativo da distribuição de classes.

O primeiro passo da apresentação desses resultados consiste em determinar o método de agrupamento automático. Em seguida, serão discutidos os resultados obtidos para o método de agrupamento escolhido no processo de identificação de classes, tanto para os dados simulados, quanto para os dados reais.

#### 6.4.1 SELEÇÃO DO MÉTODO DE AGRUPAMENTO

Para seleção do método de agrupamento utilizado como identificador de classes, foram escolhidos previamente os quatro diferentes modelos citados anteriormente na Seção 5.5:

- *Maximum Weighted Likelihood via Rival Penalized Expectation Maximization* ou RPEM;
- Agrupamento baseado em Vetores Suporte ou *Support Vector Clustering* (SVC);
- *Self-Tuning Spectral Clustering* (STSC);
- *X-Médias*;

Para realizar a comparação de desempenho desses métodos de agrupamento, optou-se por utilizar a métrica de validação externa baseada na matriz de confusão dos grupos. Vale ressaltar que a principal justificativa para utilização de uma métrica dessa natureza é avaliar a capacidade do método de colocar eventos de uma mesma classe, em um mesmo grupo.

A base de dados selecionada para avaliar essa capacidade consiste de três classes de curto-circuito monofásico, geradas através de simulação no ATP, com as seguintes características:

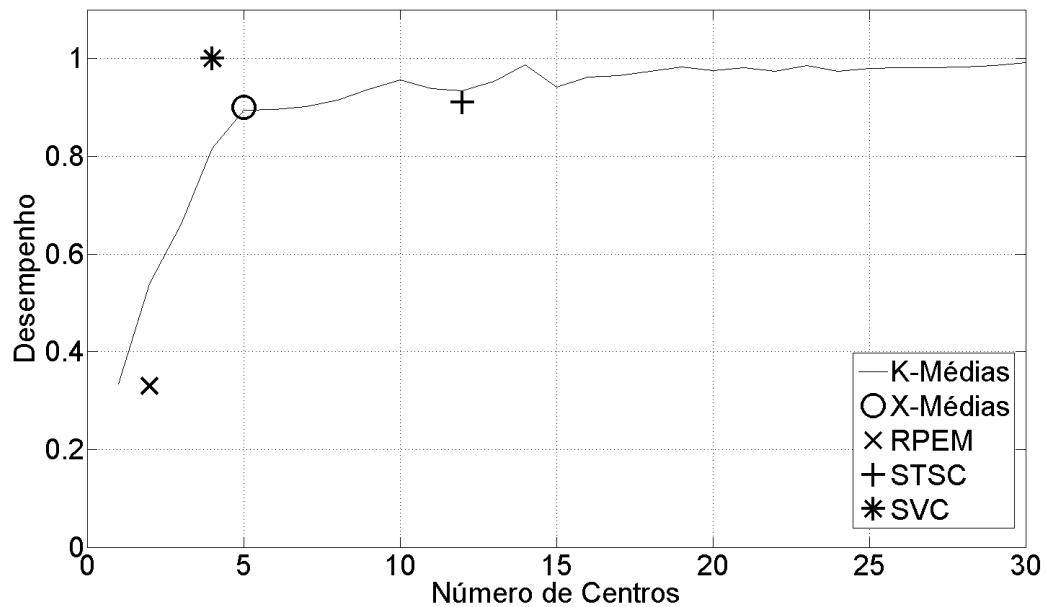
- Classe 1: Curto-Circuito Monofásico na Fase A com 54 padrões por classe;
- Classe 2: Curto-Circuito Monofásico na Fase B com 54 padrões por classe;
- Classe 3: Curto-Circuito Monofásico na Fase C com 54 padrões por classe.

O objetivo de trabalhar com uma base de dados com poucas classes é facilitar o processo de interpretação dos resultados em um primeiro momento, apontando as características relevantes e limitações de cada método para um problema cuja separabilidade e possibilidade de agrupamento são maiores, quando comparados a um conjunto com um número elevado de classes.



Com relação ao pré-processamento, foi utilizada a TWPD em quatro níveis e uma frequência de amostragem de 15.360 Hz. A extração de características foi feita através da relação de energia, descrita na Seção 4.5.

Utilizando as características acima descritas, foram aplicados os quatro métodos de agrupamento automático na base de dados com as três classes de curto-circuito monofásico. A Figura 77 mostra o resultado dessa simulação.



**Figura 77: Comparação de desempenho para diferentes métodos de agrupamento para as três classes de curto-circuito monofásico. Nesse gráfico, fica claro o comportamento assintótico da acurácia na classificação do modelo K-Médias, no limite em que o número de agrupamentos tende a ser igual ao número de padrões por classe. Além disso, observa-se que os modelos X-Médias, SVC e STSC possuem as melhores acurácias na classificação**

Na Figura 77, está representado o desempenho dos quatro métodos de agrupamento automático listados anteriormente, para as três classes de curto-circuito monofásico. Além disso, está representado o desempenho para o modelo K-Médias para diferentes centros, com o intuito de fornecer uma referência para comparação entre os modelos.

Para compreender como essa curva pode fornecer uma referência de comparação para os métodos de agrupamento, é necessário primeiramente analisar o seu comportamento, no limite em que o número de centros definidos tende a ser igual ao número de padrões disponíveis para o agrupamento. Conforme descrito anteriormente, na Seção 4.9, a métrica de validação externa baseada na matriz de confusão dos grupos não considera o número de grupos que um dado método produz. Ou seja, no limite em que o número de centros é igual ao número de padrões, o acerto do agrupamento é 100% para essa métrica, conforme é possível observar na

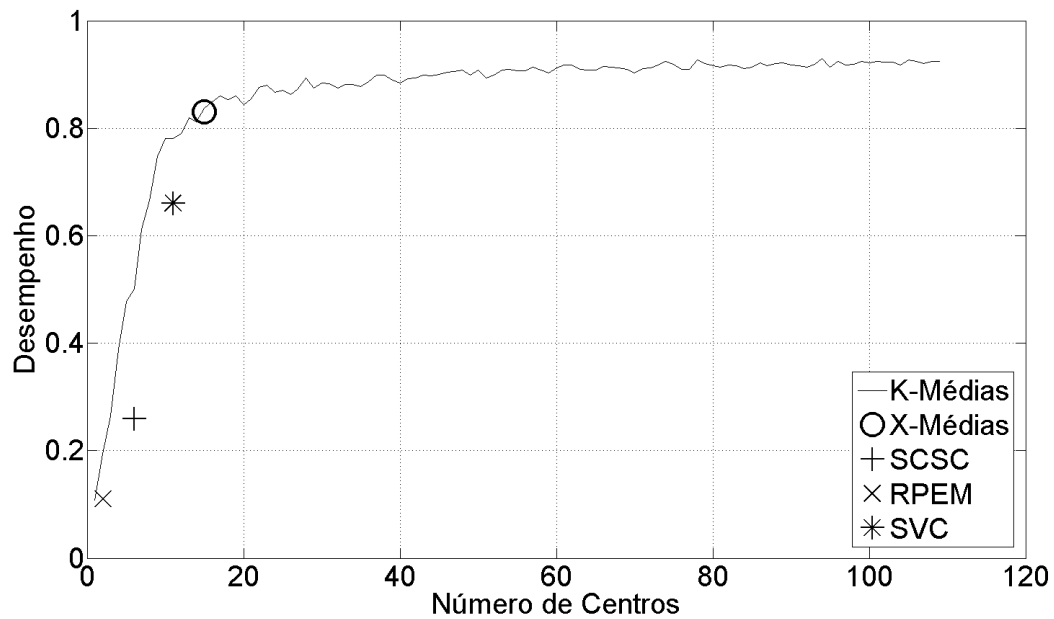
Figura 77.

Analisando a resposta obtida para o modelo  $K$ -Médias para diferentes centros, observa-se que existe certa saturação no desempenho obtido a partir de, aproximadamente, cinco centros. Isso indica que a inserção de novos centros não acrescenta informações relevantes para identificação de novas classes através dos grupos formados, podendo ser, inclusive, considerada desnecessária para a representação em questão. Desse modo, é de se esperar que qualquer modelo de agrupamento automático que apresente um bom desempenho de agrupamento tenha o seu resultado localizado na região do início da saturação da curva (ou acima) do modelo  $K$ -Médias, desde que seja aplicado aos mesmos dados modelados pelo  $K$ -Médias. Essa região do início de saturação indica o compromisso entre um bom desempenho e um número de centros adequado ao número de classes do problema no caso do modelo  $K$ -Médias. Nesse sentido, pode-se afirmar que a curva do modelo  $K$ -Médias serve de referência para a comparação proposta.

Com isso, é possível observar que, em princípio, tanto o modelo  $X$ -Médias, quanto o modelo de agrupamento baseado em Vetores Suporte são candidatos mais adequados ao problema de agrupamento em questão. No entanto, para obter uma ideia mais clara da capacidade de agrupamento de cada método, optou-se por avaliar o desempenho para uma base de dados contendo todas as onze classes de curto-circuito, geradas via simulação no ATP (todas com 54 padrões por classe). Os resultados dessa análise estão representados na Figura 78.

Para detalhar as limitações e as vantagens dos resultados apresentados nas Figuras 77 e 78 é necessário analisar cada um dos quatro métodos de forma individual.

Com relação ao método de agrupamento baseado em Vetores Suporte, é possível observar que o aumento do número de classes leva a uma queda considerável de desempenho. No caso das três classes de curto-circuito monofásico, é possível obter um desempenho de 100% utilizando apenas quatro grupos. No entanto, quando se consideram todas as onze classes de curto-circuito, o desempenho máximo obtido foi de aproximadamente 67%. Observou-se que, para esse modelo, a sensibilidade do ajuste dos parâmetros do *kernel* e de controle de complexidade impactam de forma significativa no formato e na quantidade de grupos gerados. Desse modo, o resultado final do agrupamento fica comprometido, do ponto de vista de identificação de possíveis classes através dos grupos. Muito embora o modelo determine de forma automática o formato e o número de grupos em uma dada distribuição de dados, a forte dependência da seleção do parâmetro do *kernel* e de controle de complexidade faz com que sua utilização no contexto de agrupamento automático seja dificultada e, para determinados arranjos de classes, inviável. A seleção automática desses parâmetros ainda é um tema em aberto na literatura (TAX,



**Figura 78:** Comparação de desempenho para diferentes métodos de agrupamento para todas as classes de curto-circuito. Nesse gráfico, novamente observa-se comportamento assintótico da acurácia do modelo *K-Médias*, no limite em que o número de agrupamentos tende a ser igual ao número de padrões por classe. Nesse caso, porém, os modelos *X-Médias* e *SVC* apresentaram as melhores acurácias na classificação do agrupamento.

2001).

As mesmas observações apontadas para o método de agrupamento baseado em Vetores Suporte com relação à seleção dos parâmetros, valem para o método *Self-Tuning Spectral Clustering*. Uma diferença fundamental, é que esse último permite a determinação automática dos parâmetros envolvidos do processo de agrupamento espectral. No entanto, os experimentos realizados demonstraram que nem sempre a seleção automática proposta pelos autores (ZELNIK-MANOR; PERONA, 2004) leva a um desempenho superior a uma seleção manual (via validação cruzada, por exemplo). Além disso, a sensibilidade na escolha dos parâmetros impacta diretamente no desempenho final. Para ambas as bases de dados avaliadas, os melhores desempenhos obtidos para diferentes parâmetros avaliados estão representados nas Figuras 77 e 78.

Já para o modelo RPEM, existem dois fatores que interferem no desempenho: o número de padrões por classe e a dimensionalidade do problema. Para analisar essa questão, optou-se por criar uma base de dados sintética (CHEUNG, 2005), incluindo três funções Gaussianas bem separadas (sem sobreposição) para diferentes dimensionalidades. Sobre essa base de dados é aplicado um agrupamento automático baseado no modelo RPEM. Para todos os casos, é de se esperar que o modelo forneça a caracterização das três funções Gaussianas, utili-

zando apenas um centro para cada função, sendo que os demais centros podem ser desprezados na modelagem. A Tabela 38 mostra o resultado do agrupamento RPEM, indicando o número de padrões definido por grupo, para 200 padrões divididos de forma equilibrada entre as três funções Gaussianas (aproximadamente 66 padrões por classe). Esse processo foi repetido para diferentes dimensionalidades dos dados de entrada.

**Tabela 38: Resultado para o método RPEM com 200 padrões divididos em três grupos.**

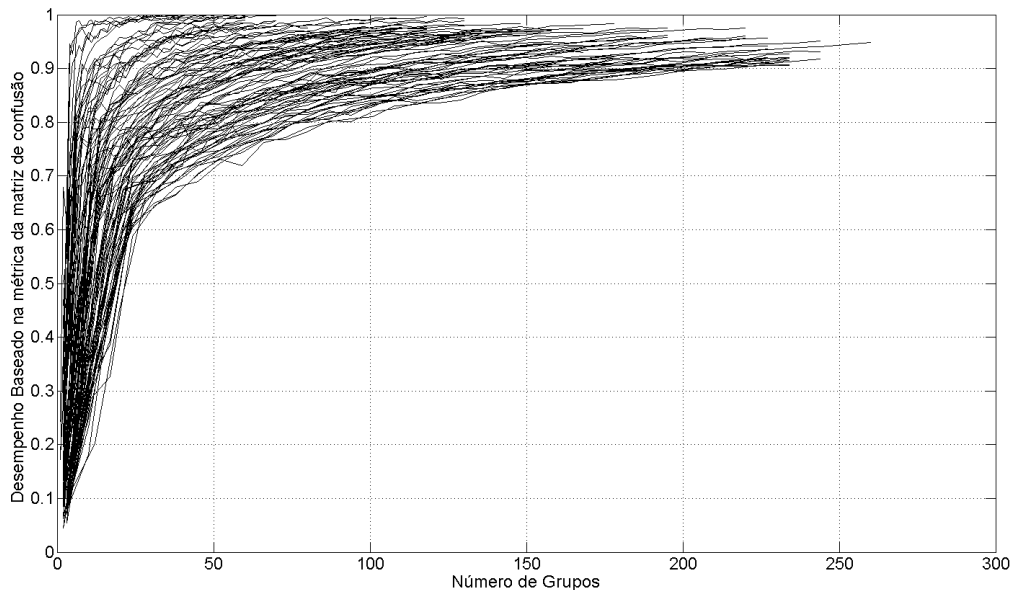
Número de Dimensões	Número de Padrões do Grupo 1	Número de Padrões do Grupo 2	Número de Padrões do Grupo 3
Duas Dimensões	61	66	73
Três Dimensões	61	73	66
Quatro Dimensões	61	73	66
Cinco Dimensões	61	66	73
Dez Dimensões	61	66	73
Vinte Dimensões	6	0	194
Cinquenta Dimensões	200	0	0

Observa-se que dados de entrada a partir de vinte dimensões, passam a ser modelados por praticamente uma função Gaussiana. Isso é função da esparsidade existente nos dados com alta dimensionalidade, uma vez que o número de padrões por distribuição Gaussiana permanece constante ao longo de todas as simulações. Com um número maior de padrões por classe (aproximadamente 300 padrões por classe), esse problema passa a surgir apenas com um número elevado de dimensões (aproximadamente 50), o que reforça o fato desse resultado estar associado à esparsidade dos dados em cenários de alta dimensionalidade e à dificuldade em modelar essas distribuições utilizando funções Gaussianas. Essa característica pode justificar a limitação do modelo RPEM em no máximo dois grupos, conforme mostram as Figuras 77 e 78.

Por último, o modelo baseado no agrupamento *X-Médias* apresentou os desempenhos mais próximos do início da saturação da curva de desempenho do *K-Médias*. De certo modo, esse resultado era esperado, uma vez que o *X-Médias* utiliza como base o agrupamento *K-Médias*. No entanto, uma característica importante é que o modelo *X-Médias* realiza de forma adequada o controle de complexidade, mantendo o número de grupos dentro da região limite esperada para as duas bases de dados apresentadas.

Outra observação que merece destaque nesse contexto, é que esse comportamento de saturação é observado para os mesmos 100 diferentes arranjos de classes avaliados na para classificação e detecção de novidades, utilizando o modelo *K-Médias*, conforme mostra a Figura 79. Para cada arranjo de classe, foi feita uma análise do desempenho do modelo *X-Médias* e foi confirmado que esse desempenho se encontra na região de início de saturação para todas

os 100 arranjos avaliados.



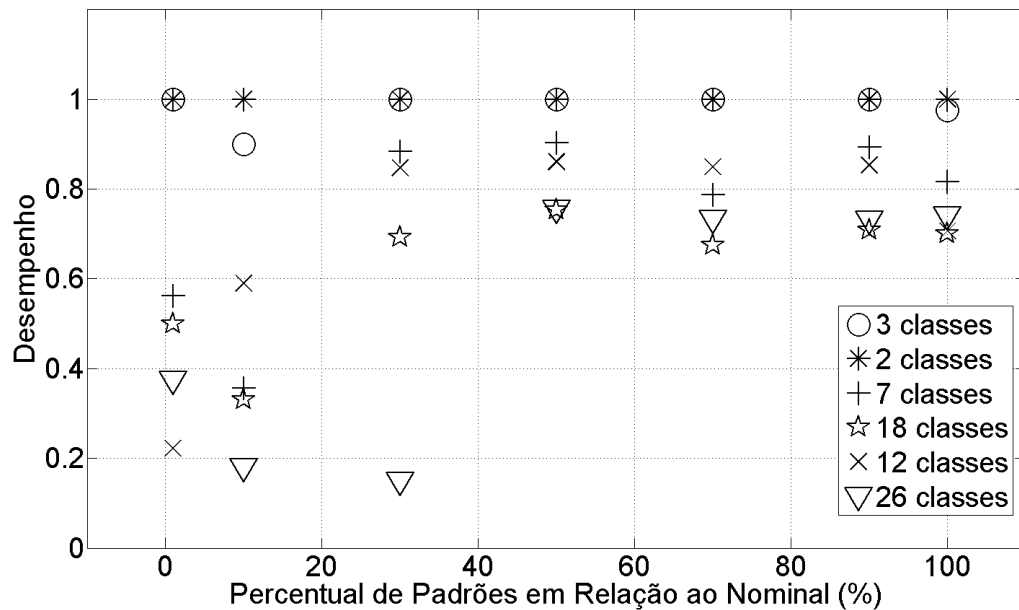
**Figura 79: Comparação de desempenho para diferentes arranjos de classes utilizando o método K-Médias.**

Analisando os critérios anteriormente discutidos, optou-se por utilizar o modelo X-Médias para o agrupamento automático associado ao processo de identificação de classes. Acredita-se que esse modelo seja capaz de atender às restrições impostas para o problema de classificação e agrupamento em questão, principalmente em termos de desempenho, além de não necessitar uma intervenção na seleção de parâmetros para os diferentes arranjos de classes avaliados.

Por fim, existem dois fatores que merecem destaque dentro do escopo proposto, relacionados à identificação de classes através de métodos de agrupamento: número de padrões por classe e número de classes existentes no momento do agrupamento automático. Conforme discutido anteriormente, espera-se que, para o problema de classificação de oscilografias, o agrupamento automático a ser aplicado nas classes consideradas como novidades seja feito periodicamente e, no instante em esse agrupamento for aplicado, espera-se que não existam muitas classes a serem identificadas.

Desse modo, uma análise bastante relevante, é a capacidade do método de agrupamento em identificar os grupos, mesmo em casos em que o número de padrões por classe seja reduzido. Nesse sentido, a Figura 80 mostra o resultado do desempenho obtido para diferentes números de padrões por classe (em termos percentuais do nominal – número de exemplos inicialmente disponíveis) e diferentes arranjos de classes. Observa-se que, para os casos em que existem

poucas classes, que é o caso esperado, o número de padrões por classe não interfere no desempenho do agrupamento. Porém, quando o número de classes aumenta, o desempenho passa a ser fortemente afetado, para os casos em que estão disponíveis poucos padrões por classe. Nesse sentido, verifica-se que é fundamental que o agrupamento automático seja aplicado com a maior frequência possível aos dados assinalados como novidades, mesmo que existam poucos padrões por classe, pois isso maximiza a possibilidade de identificar novas classes a partir do agrupamento automático.



**Figura 80:** Comparação de desempenho para diferentes números de padrões por classe e diferentes classes utilizando o método X-Médias. Para essa comparação, foram utilizadas bases de dados reduzidas em relação à base de dados original, utilizando os valores de 2%, 10%, 30%, 50%, 70%, 90% e 100% dos padrões, em relação ao número total de padrões disponíveis (base de dados original). Para cada caso, foi avaliada a acurácia para diferentes arranjos e números de classes, dentre as 29 classes de eventos disponíveis por simulação. Com isso, é possível avaliar o impacto do número de padrões por classe, o arranjo e o número de classes na acurácia de classificação do agrupamento.

#### 6.4.2 RESULTADOS PARA OS DADOS SIMULADOS

Aplicando o método X-Médias aos 100 diferentes arranjos de classes, de acordo com a metodologia exposta na Figura 68, obtém-se os desempenhos apresentados na Tabela 39. Essa tabela considera a métrica proposta na Seção 4.9.3, tanto para o conjunto de treinamento, quanto para o conjunto de teste.

Observa-se que o desempenho obtido para o conjunto de treinamento é próximo ao obtido para o conjunto de teste. O teste de Wilcoxon indica que as duas distribuições são equi-

**Tabela 39: Desempenho (%) do método de agrupamento automático X-Médias para diferentes arranjos de classes (dados simulados).**

Método	Desempenho Médio (%)
Conjunto de Treinamento	73,4 ± 11,6
Conjunto de Teste	73,4 ± 12,5

valentes e que, portanto, os desempenhos podem ser considerados equivalentes. Esse resultado é justificado em função dos dados de treinamento e teste terem sido gerados de forma bastante similar, conforme demonstrado na Subseção 5.1.9.

A mesma conclusão obtida para os resultados de classificação e detecção de novidades, com relação ao desempenho de determinados arranjos de classes, pode ser estendida para os resultados dos métodos de agrupamento automático. Se considerado o resultado do agrupamento automático aplicado apenas nas classes de curto-circuito, o desempenho para o conjunto de teste é próximo a 83% de acerto. Ou seja, para esse arranjo de classes é possível obter um desempenho superior ao desempenho médio apontado na Tabela 39.

Os resultados apresentados para os diferentes arranjos de classes indicam a possibilidade de identificar os grupos existentes nas classes com desempenho comparável aos resultados apresentados para classificação e detecção de novidades. Isso indica que para os diversos arranjos de classes analisados, o agrupamento automático pode ser utilizado como parte do processo de identificação de novas classes, auxiliando o especialista na identificação de similaridades existentes nos dados sob análise. No entanto, a generalização dessa afirmação necessita da avaliação dos resultados incluindo dados reais, os quais são apresentados na Subseção 6.4.3.

#### 6.4.3 RESULTADOS PARA OS DADOS REAIS

A Tabela 40 apresenta os resultados obtidos para os dados reais, utilizando o agrupamento automático X-Médias, considerando diferentes formas de pré-processar os sinais, tanto para o conjunto de treinamento, quanto para o conjunto de teste. Os conjuntos de treinamento e teste utilizados nessa análise inicial estão detalhados a seguir:

- Curto-Circuito Monofásico na Fase A: 18 exemplos simulados e 5 exemplos reais para treinamento; 5 exemplos simulados e 2 exemplos reais para teste;
- Curto-Circuito Monofásico na Fase B: 18 exemplos simulados e 1 exemplo real para treinamento; 5 exemplos simulados e 1 exemplo real para teste;
- Curto-Circuito Monofásico na Fase C: 18 exemplos simulados e 13 exemplos reais para

treinamento; 5 exemplos simulados e 4 exemplos reais para teste;

- Religamento Trifásico: 9 exemplos simulados e 8 exemplos reais para treinamento; 4 exemplos simulados e 4 exemplos reais para teste;
- Descarga Atmosférica Induzida: 68 exemplos reais para treinamento; 17 exemplos reais para teste;
- Novidades: demais classes, com 26 exemplos reais, utilizados apenas na fase de teste.

**Tabela 40: Desempenho (%) do método de agrupamento automático X-Médias para dados reais.**

Pré-Processamento	Conjunto de Treino	Conjunto de Teste
245.760-twd-db20	68,60	54,33
245.760-twpd-db20	73,26	59,06
122.880-twd-db20	79,07	65,35
122.880-twpd-db20	60,47	47,24

Observa-se que os desempenhos acima apresentados são inferiores a 65% de acerto. Ao se aplicar o mesmo pré-processamento anteriormente apresentado, incluindo o módulo e fase de cada sinal de tensão trifásica no vetor de características composto por 245.760-twd-db20, o desempenho aumenta de forma significativa. A matriz de confusão do conjunto de teste para esse pré-processamento está representada na Tabela 41.

**Tabela 41: Matriz de confusão para o pré-processamento 245.760-twd-db20 (com módulo e fase) para o agrupamento automático através do X-Médias (classe atribuída representada nas colunas).**

Classes	Falta Monof. Fase A	Falta Monof. Fase B	Falta Monof. Fase C	Religamento Trifásico	Descarga Atmosférica Induzida
Falta Monof. Fase A	6	0	0	0	1
Falta Monof. Fase B	0	5	0	0	1
Falta Monof. Fase C	0	0	5	0	4
Religamento Trifásico	0	0	0	8	0
Descarga Atmosférica Induzida	0	0	2	0	15



O desempenho total do modelo é de aproximadamente 82% de acerto. Para esse caso, foram definidos automaticamente 14 grupos para representar as cinco classes analisadas, sendo a distribuição de grupos por classe dada pelos seguintes valores:

- Falta Monofásica (Fase A): dois grupos;
- Falta Monofásica (Fase B): três grupos;
- Falta Monofásica (Fase C): três grupos;
- Religamento Trifásico: dois grupos;
- Descarga Atmosférica: quatro grupos.

Ao se comparar o desempenho do modelo *X*-Médias com o desempenho do modelo MVS para a mesma base de dados (desempenho de aproximadamente 96% de acerto), observa-se que a ausência de rótulos fornece uma classificação não muito distante da classificação com rótulos. A diferença é de quase 15% de acerto mais para o treinamento supervisionado.

Para os dados reais, assim como apresentado para os dados simulados, é possível afirmar com base nos desempenhos apresentados, que o agrupamento automático pode operar como parte do processo de identificação de novas classes, considerando que para o caso da Tabela 40, as cinco classes são desconhecidas *a priori*.

No entanto, para que o processo de identificação de novas classes seja feito de forma completa, é necessária a intervenção final do especialista, de modo a definir se determinados grupos fazem parte de uma mesma classe. Além disso, a atuação do especialista pode indicar eventuais erros de classificação no agrupamento automático, refinando a análise como um todo. Ressalta-se que o agrupamento automático é apenas o estágio inicial do processo de identificação de novas classes, dentro do contexto proposto no presente trabalho.

## 7 CONCLUSÕES

Este trabalho apresentou novas perspectivas para a análise automática de oscilografias em redes de distribuição, iniciando pela detecção (segmentação) da oscilografia e finalizando com o processo de classificação, permitindo a identificação de novas classes. Dentre as principais contribuições apresentadas ao longo desse processo, pode-se destacar:

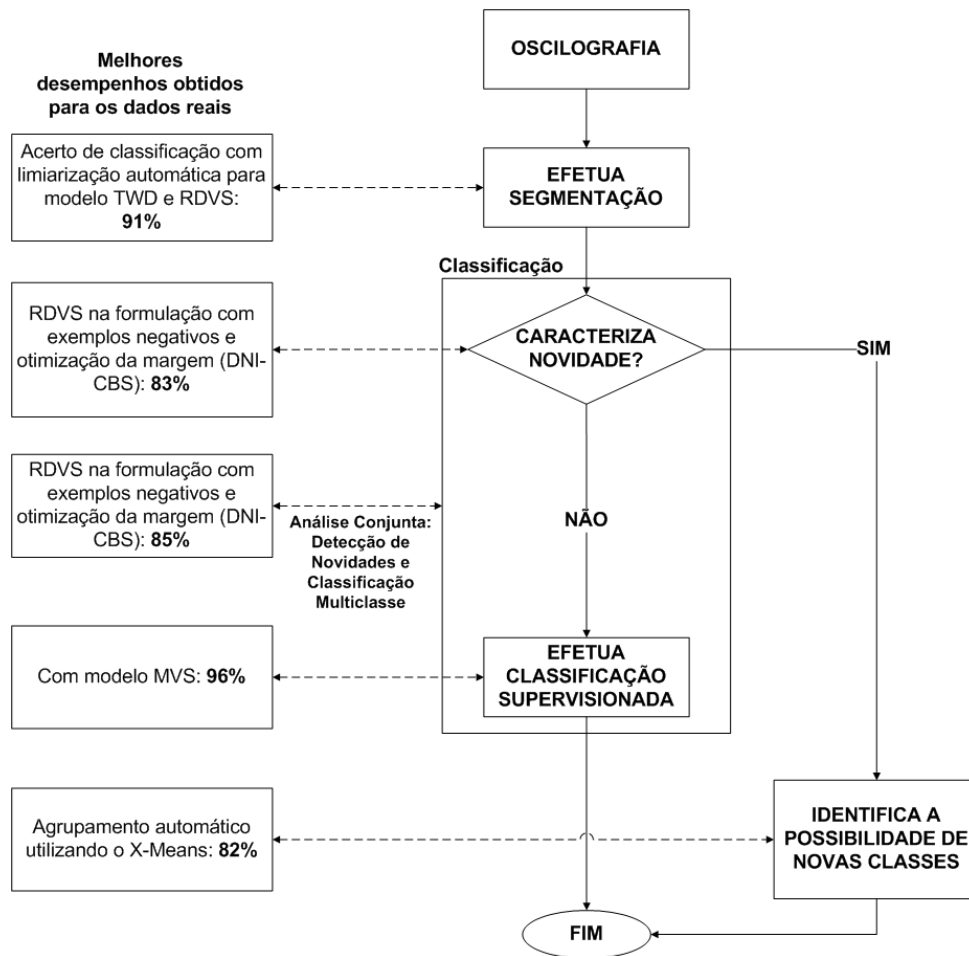
- Detecção (segmentação) de eventos transitórios e estados quase-estacionários: Nesta etapa, destaca-se a proposição de um novo método, baseado na TWD e na RDVS, que leva a resultados significativamente superiores em comparação com os métodos normalmente utilizados para essa finalidade em oscilografias de SEP. O método apresentado pode ser aplicado para qualquer problema de detecção de variação abrupta em sinais, uma vez que não existem restrições para o uso somente em oscilografias. Além disso, permite a composição de diferentes níveis da TWD, maximizando a possibilidade de identificar diferentes transitórios nas formas de onda.
- Classificação multiclasse incluindo a detecção de novidades ou classes previamente desconhecidas: O aspecto dinâmico de aprendizado é comum em muitas aplicações de reconhecimento de padrões, sendo tema de pesquisa frequente (e recente) nessa área. No SEP, em particular nos sistemas de distribuição, essa característica também é presente e vem se tornando mais impactante à medida que novas mudanças são aplicadas no setor, em especial com o advento de redes elétricas inteligentes. Sendo assim, modelos de classificação automática devem considerar essa característica, principalmente modelos que utilizam informações em tempo real do sistema de distribuição, como oscilografias de unidades de monitoramento. No presente trabalho, foram comparados diferentes modelos para essa tarefa, considerando o reconhecimento automático de novas classes (novidades) e a classificação dentre múltiplas classes conhecidas *a priori*. O destaque é uma formulação que deriva do modelo RDVS, o qual considera exemplos negativos e otimização da margem durante o estágio de treinamento. Os resultados demonstram a viabilidade e as limitações desse tipo de classificação no problema de classificação automática de oscilografias, principalmente utilizando dados reais.

- Agrupamento automático como parte do processo de identificação de novas classes: Depois de classificados como novidades, os exemplos desconhecidos previamente podem estar associados a uma nova classe ou simplesmente um registro espúrio. Com o intuito de auxiliar o especialista nessa tarefa, foram comparados e avaliados diferentes modelos de agrupamento automático, os quais visam identificar similaridades entre os padrões classificados como novidades. No presente trabalho, os exemplos das classes previamente disponíveis para os dados reais e simulados foram utilizados para comparação entre os modelos, sob uma abordagem não-supervisionada. O modelo X-Médias obteve os melhores acertos de classificação, agrupando eventos de uma mesma classe em um, ou eventualmente mais de um grupo para cada classe. O resultado final do agrupamento automático fornece indicativos para o especialista decidir, comparando com informações de outras bases de dados, se um determinado conjunto de eventos está associado ao surgimento de uma nova classe.

Com o intuito de apresentar um resumo global dos melhores resultados obtidos para a metodologia proposta neste trabalho (Figura 3), optou-se por inserir os resultados de cada estágio do processo, incluindo a segmentação, classificação, detecção de novidades e a possível identificação de classes através do agrupamento automático. A Figura 81 mostra esquematicamente esse resumo.

Comparando inicialmente o desempenho de segmentação, foi demonstrado que o modelo RDVS em conjunto com a TWD apresenta os melhores desempenhos, tanto para os dados simulados, quanto para os dados reais. No caso dos dados reais foi feita uma análise para diferentes limiares, bem como para um processo de limiarização automática, resultando em uma segmentação completamente automática das oscilografias analisadas. Para o modelo de segmentação proposto neste trabalho, em conjunto com o processo de limiarização automática, verificou-se um acerto médio de classificação 91% dos instantes da oscilografia, que podem ser relacionados a estados quase-estacionários ou transitórios. Vale ressaltar que os dados reais analisados envolvem diferentes eventos registrados em diferentes localidades.

Uma vez estabelecido que um determinado intervalo de tempo corresponde a um transitório (realizada a segmentação), é importante realizar sua classificação. Para esse processo, optou-se por dividir a identificação das classes considerando a possibilidade de um dado transitório (evento) estar relacionado a uma classe ainda não conhecida pelo especialista (e, conseqüentemente, pelo classificador automático), além da identificação da classe do evento, caso esse seja considerado como conhecido previamente. Esse processo incorpora à classificação multiclasse convencional, a detecção de novidades.



**Figura 81: Resultados de classificação, detecção de novidades e identificação de classes.**

Na Figura 81 estão assinalados os desempenhos obtidos apenas para o processo de detecção de novidades, o processo de classificação multiclasse convencional e a classificação como um todo, considerando a detecção de novidades e a classificação multiclasse de forma conjunta. O primeiro estágio do processo, a detecção de novidades, pode ser considerado como um problema de classificação binária, no qual procura-se atribuir ao evento sob análise o rótulo de conhecido ou desconhecido previamente (novidade). Para a base de dados incluindo exemplos reais, o desempenho foi de 83% de acerto utilizando o método RDVS com Exemplos Negativos e Otimização da Margem, na abordagem DNI-CBS. Ou seja, para esse método nessa abordagem, é possível identificar corretamente se um padrão é conhecido previamente (padrão normal) ou desconhecido (novidade) em 83% dos casos. Nesse ponto vale ressaltar que os padrões caracterizados como novidades são utilizados apenas na fase de teste, não sendo incluídos no estágio de treinamento de nenhum método.

Com relação à classificação multiclasse convencional, sem considerar a possibilidade de existirem eventos desconhecidos *a priori*, o desempenho médio foi de 96% para os dados reais. Se for feita a análise conjunta da detecção de novidades e classificação multiclasse super-

visionada, o acerto de classificação para a análise conjunta das cinco classes conhecidas mais os padrões caracterizados como novidades, é de 85% utilizando o modelo RDVS na formulação com exemplos negativos e otimização da margem. Vale lembrar que para realizar essa avaliação de forma conjunta, considera-se que as novidades formam a classe  $K + 1$  de um problema de classificação contendo  $K$  classes na matriz de confusão. Outro detalhe importante, é que um erro de classificação na primeira etapa – detecção de novidade – se propaga para o restante da análise. Esse é a principal fator de influência no desempenho obtido para a análise conjunta aqui proposta.

A última etapa do processo considera a possibilidade de identificação de novas classes, a partir do agrupamento automático aplicado nos dados caracterizados como novidade. A ideia do agrupamento automático é fornecer para o especialista alguma informação relevante que possibilite a identificação de novas classes, a partir de um conjunto de dados sobre os quais não se tem conhecimento *a priori*. Nesse sentido, o agrupamento automático identifica as similaridades entre os padrões de entrada, sendo que os grupos formados são um indício de uma provável distribuição de classes que pode existir nesses dados, mesmo que determinadas classes possam ser formadas por mais de um grupo. Com o método *X-Médias*, obteve-se um desempenho médio de 82% de acerto para as cinco classes disponíveis para esse teste incluindo dados reais. Esse desempenho indica a capacidade do método de agrupamento automático em colocar eventos de uma mesma classe em um mesmo grupo, sendo que o controle de complexidade – aqui representado pelo número de centros – é realizado pelo próprio método de agrupamento.

Com relação aos objetivos traçados inicialmente, pode-se afirmar que todos foram atingidos integralmente. No caso particular do objetivo geral, como foi demonstrado pelos resultados expostos no Capítulo 6, esta tese apresentou um método (ou um conjunto de possíveis métodos) que permitem a detecção e a classificação de instantes transitórios em oscilografias de tensão em redes de distribuição. No caso particular dos objetivos específicos, é possível destacar que:

- Foi modelado e simulado, utilizando o software *Alternative Transient Program*, um sistema de distribuição de energia reduzido, para validação inicial da metodologia proposta. O modelo foi apresentado na Seção 5.1, enquanto os resultados foram discutidos nas Seções 6.1 e 6.2.
- Foram selecionados os principais eventos a serem simulados, de acordo com os eventos mais frequentes em redes de distribuição. As escolhas e os resultados são tratados nos Capítulos 2 e 6, respectivamente.

- Foi desenvolvido um método de detecção de eventos (ou segmentação da oscilografia), baseado na TWD e na RDVS. Sua fundamentação teórica foi apresentada nas Subseções 4.4.3 e 4.8.7 e os resultados na Seção 6.1.
- Foram analisadas diferentes formas de pré-processar os sinais, visando à escolha da melhor e mais compacta forma de extrair as características dos sinais sob análise. O pré-processamento é baseado na TWD e seus resultados são discutidos ao longo do Capítulo 6.
- Foram avaliadas e propostas diversas técnicas de detecção de novidades, permitindo a detecção de novidades e classificação multiclasse conjunta. O destaque é a utilização da formulação RDVS-M, conforme demonstra a Seção 6.3.
- Foram avaliadas diversas técnicas de agrupamento automático, resultando na escolha de um método (*X-Médias*) que pode auxiliar o especialista na identificação de novas classes. Os resultados estão descritos na Seção 6.4.
- Todos os métodos propostos foram discutidos e comparados com base em eventos simulados e reais, conforme demonstram os Capítulos 5 e 6.

De modo geral, os resultados aqui resumidos indicam a viabilidade da proposta completa deste trabalho, iniciando pelo processo de detecção de instantes transitórios, passando pela classificação dentre classes conhecidas e finalizando com a identificação de novas classes. Tal processo de classificação é considerado a maior contribuição do trabalho, sendo caracterizado por um tema pouco explorado nas mais diversas áreas que envolvem reconhecimento de padrões. Sendo assim, é possível apontar alguns trabalhos futuros relevantes nesse tema, como resume a Seção 7.1.

## 7.1 TRABALHOS FUTUROS

De modo geral, os principais pontos para pesquisa futura são:

- Pré-processamento de formas de onda admitindo a detecção de novidades. Quando se considera uma classificação com possibilidade de novas classes na etapa de teste, deve-se procurar realizar um pré-processamento e extração de características que maximize essa possibilidade do ponto de vista de classificação automática. Alguns trabalhos recentes consideram isso, principalmente no tratamento de eventos simultâneos, como mostram Lima et al. (2014);

- Aprendizado de métrica para classificadores uma-classe. Como demonstrado anteriormente, o resultado de extração de características pode resultar em um grande volume na representação dos dados no espaço de características. Aprendizado de métrica pode auxiliar nesse processo (WEINBERGER; SAUL, 2009), adequando medidas de distância para o espaço em questão, facilitando a representação de modelos uma-classe. No entanto sua aplicação em um cenário de detecção de novidades ainda é um tema em aberto na literatura;
- Aprendizado de métrica para agrupamento automático;
- Aquisição de novos dados em campo. Os sistemas de aquisição estão sendo atualmente reinstalados em campo e novos dados estão sendo adquiridos. Com isso, espera-se poder aplicar os métodos propostos neste trabalho aos novos dados, auxiliando diretamente no processo de mitigação de eventos em redes de distribuição. Em paralelo, está sendo desenvolvido um aplicativo para ser utilizado por engenheiros de rede de distribuição, incluindo as metodologias aqui propostas;
- Investigação de formas alternativas de determinação automática de limiar para o modelo TWD+RDVS não dependentes de um conjunto prévio de oscilografias normais;
- Avaliação das técnicas propostas em sistemas embarcados. O desenvolvimento de sistemas de monitoramento poderá incluir em um futuro próximo, sistemas inteligentes na sua solução final. Um tema de pesquisa relevante é a inclusão dos métodos aqui apresentados em dispositivos eletrônicos embarcados, com novas arquiteturas que permitam tal inclusão.

## REFERÊNCIAS

- ABDEL-GALIL, T. Power quality event detection using Adaline. **Electric Power Systems Research**, v. 64, p. 137–144, 2003.
- ADU, T. An accurate fault classification technique for power system monitoring devices. **IEEE Transactions on Power Delivery**, v. 17, p. 684–690, 2002.
- 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.
- AGUIRRE, L. A. **Introdução à Identificação de Sistemas**. [S.l.]: Editora UFMG, 2007.
- AL-KHATEEB, T.; MASUD, M. M.; KHAN, L.; AGGARWAL, C.; HAN, J.; THURAI-SINGHAM, B. Stream classification with recurring and novel class detection using class-based ensemble. In: **12th International Conference on Data Mining**. [S.l.: s.n.], 2012. p. 31–40. ISBN 978-1-4673-4649-8.
- ANEEL. **Procedimentos de Distribuição, Módulo 8**. [S.l.], 2008.
- AYDIN, T. Analysis of asymmetrical faults in power systems using dynamic phasors. **Power**, v. 15, p. 1062–1068, 2000.
- BASSEVILLE, M.; NIKIFOROV, I. V. **Detection of Abrupt Changes: Theory and Application**. [S.l.]: Prentice-Hall, Inc., 2006.
- BISHOP, C. M. **Neural Networks for Pattern Recognition**. Oxford: Oxford University Press, 1995.
- BODESHEIM, P.; FREYTAG, A.; RODNER, E.; KEMMLER, M.; DENZLER, J. Kernel null space methods for novelty detection. In: **IEEE Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2013. p. 3374–3381.
- BOLLEN, M. **Understanding Power Quality Problems: Voltage Sags and Interruptions**. [S.l.]: Wiley-IEEE Press, 1999.
- BOLLEN, M.; GU, I. **Signal Processing of Power Quality Disturbances**. [S.l.]: John Wiley & Sons, 2006.
- BOLLEN, M. H. J.; GU, I. Y. H.; AXELBERG, P. G. V.; STYVAKTAKIS, E. Classification of underlying causes of power quality disturbances: Deterministic versus statistical methods. **EURASIP Journal on Advances in Signal Processing**, v. 1, p. 1–18, 2007.
- BORGHETTI, A.; MORCHED, A.; NAPOLITANO, F.; NUCCI, C. A.; PAOLONE, M. Lightning-induced overvoltages transferred through distribution power transformers. **IEEE Transactions on Power Delivery**, v. 24, p. 360–372, 2009.



BUCKHEIT, J.; CHEN, S.; DONOHO, D. L.; JOHNSTONE, I. M.; SCARGLE, J. D. **WaveLab Reference Manual**. 2005.

CAD, M. M. **Estratégias de Modelagem Dinâmica e Simulação Computacional do Motor de Indução Trifásico**. Dissertação (Mestrado) — Universidade de São Paulo, 2000.

CANU, S.; GRANDVALET, Y.; GUIGUE, V.; RAKOTOMAMONJY, A. **SVM and Kernel Methods Matlab Toolbox**. 2005.

CARVALHO, J. R.; COURRY, D. V.; CALHAU, D.; DELBEM, J. A. C. B.; OES, E. V. S. Algoritmos Genéticos Aplicados à Medição Fasorial e Projeto de PMUs. In: **Congresso Brasileiro de Qualidade de Energia Elétrica**. [S.l.: s.n.], 2009.

CASELLA, G.; BERGER, R. L. **Inferência Estatística**. [S.l.]: Cengage Learning, 2010.

CASTRO, V. E.; CHALUP, S. Support vector clustering through proximity graph modelling. In: **9th International Conference on Neural Information Processing**. [S.l.: s.n.], 2002. v. 2, p. 898–903.

CHANG, C.-C.; LIN, C.-J. LIBSVM: A library for support vector machines. **ACM Transactions on Intelligent Systems and Technology**, v. 2, p. 27:1–27:27, 2011.

CHAPELLE, O.; VAPNIK, V.; BOUSQUET, O.; MUKHERJEE, S. Choosing multiple parameters for support vector machines. **Machine Learning**, v. 46, n. 1-3, p. 131–159, January 2002.

CHEUNG, Y.-M. K-Means: A new generalized k-means clustering algorithm. **Pattern Recognition Letters**, v. 24, p. 2883–2893, 2003.

CHEUNG, Y.-M. On rival penalization controlled competitive learning for clustering with automatic cluster number selection. **IEEE Transactions on Knowledge and Data Engineering**, v. 17, p. 1583–1588, 2005.

CHIEN, C.; CHEN, S.; LIN, Y. Using Bayesian network for fault location on distribution feeder. **IEEE Transactions on Power Delivery**, v. 17, p. 785–793, 2002.

COORAY, V. **The Lightning Flash**. [S.l.]: The Institution of Electrical Engineers, 2003.

COSTA, F. B.; SOUZA, B. A.; BRITO, N. S. D. Detecção de faltas evolutivas e múltiplos distúrbios em registros oscilográficos baseada na transformada wavelet discreta. **Revista Controle e Automação**, v. 21, p. 173–184, 2010.

COSTA, F. B.; SOUZA, B. A.; DANTAS, K. M. C.; BRITO, N. S. D. A method for fault classification in transmission lines based on ann and wavelet coefficients energy. In: **International Joint Conference on Neural Networks**. [S.l.: s.n.], 2006. p. 3700–3705.

COSTA, F. O.; SILVA, E.; ECKMANN, M.; SCHEIRER, W. J.; ROCHA, A. Open set source camera attribution and device linking. **Pattern Recognition Letters**, v. 36, p. 1–15, 2014.

DAS, S.; ABRAHAM, A.; KONAR, A. Automatic kernel clustering with a multi-elitist particle swarm optimization algorithm. **Pattern Recognition Letters**, v. 29, p. 688–699, 2008.

DASH, P.; CHILUKURI, M. Hybrid S-Transform and Kalman filtering approach for detection and measurement of short duration disturbances in power networks. **IEEE Transactions on Instrumentation and Measurement**, v. 53, n. 2, p. 588–596, 2004.

DASH, P.; PRADHAN, A.; PANDA, G. Frequency estimation of distorted power system signals using extended complex Kalman filter. **IEEE Transactions on Power Delivery**, v. 14, p. 761–766, 1999.

DASH, P.; SAMANTARAY, S.; PANDA, G. Fault classification and section identification of a series-compensated transmission line using support vector machine. **IEEE Transactions on Power Delivery**, v. 22, n. 1, p. 67–73, 2007.

DAUBECHIES, I. **Ten Lectures on Wavelets**. [S.l.]: SIAM, Philadelphia, 1992.

DAVIDSON, E.; MCARTHUR, S.; MCDONALD, J.; CUMMING, T.; WATT, I. Applying multi-agent system technology in practice: Automated management and analysis of SCADA and digital fault recorder data. **IEEE Transactions on Power Systems**, v. 21, p. 559–567, 2006.

DAVY, M.; DESOBRY, F.; GRETTON, A.; DONCARLI, C. An online support vector machine for abnormal events detection. **IEEE Transactions on Signal Processing**, v. 86, p. 2009–2025, 2006.

DAVY, M.; GODSILL, S. Detection of abrupt spectral changes using support vector machines. In: **IEEE International Conference on Acoustics, Speech, and Signal Processing**. [S.l.: s.n.], 2002. p. 1313–1316.

DELMONT, 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.

DEMIR, Y. A new algorithm for automatic classification of power quality events based on wavelet transform and SVM. **Expert Systems with Applications**, v. 37, p. 4094–4102, 2010.

DEMIR, Y. Wavelet-based feature extraction and selection for classification of power system disturbances using support vector machines. **Electric Power Systems Research**, v. 80, p. 743–752, 2010.

DESOBRY, F.; DAVY, M. Support vector-based online detection of abrupt changes. In: **IEEE International Conference on Acoustics, Speech, and Signal Processing**. [S.l.: s.n.], 2003.

DESOBRY, F.; DAVY, M.; DONCARLI, C. An online kernel change detection algorithm. **IEEE Transactions on Signal Processing**, v. 53, p. 2961–2974, 2005.

DONG, X.; MEMBER, S.; KONG, W.; CUI, T. Fault classification and faulted-phase selection based on the initial current traveling wave. **IEEE Transactions on Power Delivery**, v. 24, p. 552–559, 2009.

DONOHU, D. L.; JOHNSTONE, I. M. Ideal spatial adaptation by wavelet shrinkage. **Biometrika**, v. 81, p. 425, 1994.

DUGAN, R. C.; MCGRANAGHAN, M. F.; SANTOSO, S.; BEATY, H. W. **Electrical Power System Quality**. 2nd. ed. New York: McGraw-Hill, 2002.

DUIN, R. P. W. Uniform object generation for optimizing one-class classifiers. **Journal of Machine Learning Research**, v. 2, p. 155–173, 2001.

DUQUE, C. A.; RIBEIRO, M. V.; CARVALHO, J. R.; BALDWIN, T. L.; RIBEIRO, P. F. A comparative performance of signal processing methods applied to power system phasor estimations under time-varying high distortion conditions. In: **Congresso Brasileiro de Qualidade de Energia**. [S.l.: s.n.], 2007.

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

ESCALERA, S.; MASIP, D.; PUERTAS, E.; RADEVA, P.; PUJOL, O. Online error correcting output codes. **Pattern Recognition Letters**, Elsevier B.V., v. 32, p. 458–467, 2011.

FAIZ, J.; LOTFI-FARD, S.; SHAHRI, S. H. Prony-Based optimal Bayes fault classification of overcurrent protection. **IEEE Transactions on Power Delivery**, v. 22, p. 1326–1334, 2007.

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.

FARID, D. M.; ZHANG, L.; HOSSAIN, A.; RAHMAN, C. M.; STRACHAN, R.; SEXTON, G.; DAHAL, K. An adaptive ensemble classifier for mining concept drifting data streams. **Expert Systems with Applications**, Elsevier Ltd, v. 40, p. 5895–5906, 2013. ISSN 09574174.

FAWCETT, T. An introduction to ROC analysis. **Pattern Recognition Letters**, v. 27, p. 861–874, 2006.

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

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

FERREIRA, W. A. **Modelagens e Simulações de Motores de Indução Trifásicos no Programa ATP**. [S.l.], 2007.

FILIPPONE, M.; CAMASTRA, F.; MASULLI, F.; ROVETTA, S. A survey of kernel and spectral methods for clustering. **Pattern Recognition**, v. 41, p. 176–190, 2008.

GAO, Y.; LI, Y. Improving Gaussian process classification with outlier detection, with applications in image classification. In: **10th Asian conference on Computer vision**. [S.l.: s.n.], 2011.

GENCER, O.; OZTURK, S.; ERFIDAN, T. A new approach to voltage sag detection based on wavelet transform. **International Journal of Electrical Power & Energy Systems**, v. 32, p. 133–140, 2010.

GEREK, O.; ECE, D. Power quality event analysis using higher order cumulants and quadratic classifiers. **IEEE Transactions on Power Delivery**, v. 21, p. 883–889, 2006.

GEREK, O.; ECE, D.; BARKANA, A. Covariance analysis of voltage waveform signature for power quality event classification. **IEEE Transactions on Power Delivery**, v. 21, p. 2022–2031, 2006.

- GOLDBERG, D. E. **Genetic Algorithms in Search, Optimization and Machine Learning**. [S.l.]: Addison-Wesley Professional, 1989.
- GOLUB, T. R.; SLONIM, D. K.; TAMAYO, P.; HUARD, C.; GAASENBEEK, M.; MESIROV, J. P.; COLLIER, H.; LOH, M. L.; DOWNING, J. R.; CALIGIURI, M. A.; BLOOMFIELD, C. D.; LANDER, E. S. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. **Science (New York, N.Y.)**, v. 286, p. 531–7, 1999.
- GREENWOOD, A. **Electrical Transients in Power Systems**. [S.l.]: John Wiley and Sons, 1991.
- GUO, H.; JACK, L. B.; NANDI, A. K. Feature generation using genetic programming with application to fault classification. **IEEE Transactions on Systems, Man and Cybernetics-PART B: Cybernetics**, v. 35, p. 89–99, 2005.
- GUSTAFSSON, F. **Adaptive Filtering and Change Detection**. [S.l.]: John Wiley & Sons, 2000.
- HANLEY, A.; MCNEIL, J.; D, P. The meaning and use of the area under a receiver operating characteristic curve. **Radiology**, v. 143, p. 29–36, 1982.
- HAO, P. Y.; CHIANG, J. H.; LIN, Y. H. A new maximal-margin spherical-structured multi-class support vector machine. **Applied Intelligence**, v. 30, p. 98–111, 2007.
- HAYKIN, S. **Redes Neurais: Princípios e Prática**. [S.l.]: Bookman, 2001.
- HEMPSTALK, K.; FRANK, E. Discriminating against new classes : One-class versus multi-class classification. In: **21st Australasian Joint Conference on Artificial Intelligence: Advances in Artificial Intelligence**. [S.l.: s.n.], 2008.
- HERBEI, R.; WEGKAMP, M. H. Classification with reject option. **Canadian Journal of Statistics**, v. 34, p. 1–28, 2005.
- HODGE, V. J.; AUSTIN, J. I. M. A survey of outlier detection methodologies. **Artificial Intelligence Review**, v. 22, p. 85–126, 2004.
- ILONEN, J.; PAALANEN, P.; KAMARAINEN, J.-K.; KALVIAINEN, H. Gaussian mixture PDF in one-class classification: Computing and utilizing confidence values. In: **18th International Conference on Pattern Recognition**. [S.l.: s.n.], 2006.
- IOANNIS, P. **Implementing mRVMs and Investigating the Sparsity vs Accuracy Trade-off**. Dissertação (Mestrado) — University of Glasgow, 2009.
- JAIN, A. K. Data clustering: 50 years beyond K-means. **Pattern Recognition Letters**, Elsevier B.V., v. 31, p. 651–666, 2010.
- JANIK, P.; LOBOS, T. Automated classification of power quality disturbances using SVM and RBF networks. **IEEE Transactions on Power Delivery**, v. 21, n. 3, p. 1663–1669, jul. 2006.
- JINGLONG, F.; WANLIANG, W.; XINGQI, W.; ZHE, L. A SVDD method based on maximum distance between two centers of spheres. **Chinese Journal of Electronics**, v. 21, p. 107–111, 2012.

- KAISER, J. F. Some useful properties of Teager's energy operators. In: **International Conference on Acoustics, Speech, and Signal Processing**. [S.l.: s.n.], 1993.
- KALMAN, E. R. A new approach to linear filtering and prediction problems. **Transactions of the ASME – Journal of Basic Engineering**, v. 82, p. 35–45, 1960.
- KEERTHIPALA, W. Fuzzy-neuro approach to fault classification for transmission line protection. **IEEE Transactions on Power Delivery**, v. 13, n. 4, p. 1093–1104, 1998.
- KENNEDY, K.; LIGHTBODY, G.; YACAMINI, R. Power system harmonic analysis using the Kalman filter. In: **IEEE Power Engineering Society General Meeting**. [S.l.: s.n.], 2003.
- KINDERMANN, G. **Curto-Circuito**. [S.l.]: Universidade Federal de Santa Catarina, 2007.
- KRISTAN, M.; SKOCAJ, D.; COMPUTER, F.; SCIENCE, I. Incremental learning with Gaussian mixture models. **Pattern Recognition**, v. 8, p. 25–32, 2008.
- KULKARNI, S.; LEE, D.; ALLEN, A. J.; SANTOSO, S.; SHORT, T. A. Waveform characterization of animal contact, tree contact, and lightning induced faults. In: **IEEE Power and Energy Society General Meeting**. [S.l.: s.n.], 2010.
- KULKARNI, S.; SANTOSO, S. SVM-based classification methodology for overhead distribution fault events. In: **14th International Conference on Harmonics and Quality of Power**. [S.l.: s.n.], 2010.
- LAZZARETTI, A. E. **Classificação de Eventos em Redes de Distribuição de Energia Elétrica utilizando Modelos Neurais Autônomos**. Dissertação (Mestrado) — Universidade Tecnológica Federal do Paraná, 2009.
- LAZZARETTI, A. E.; FERREIRA, V. H.; NETO, H. V.; RIELLA, R. J.; OMORI, J. S. Autonomous neural models for the classification of events in power distribution networks. **Journal of Control, Automation and Electrical Systems**, v. 24, n. 5, p. 612–622, 2013.
- LAZZARETTI, A. E.; FERREIRA, V. H.; Vieira-Neto, H.; RIELLA, R. J.; OMORI, J. Classification of events in distribution networks using autonomous neural models. In: **15th International Conference on Intelligent System Applications to Power Systems**. [S.l.: s.n.], 2009.
- LAZZARETTI, A. E.; RAVAGLIO, M. A.; TOLEDO, L. F. R. B.; TEIXEIRA-JÚNIOR, J. A.; ROJAS, P. M.; PINTO, C. L. S. Measurements of lightning discharges in overhead distribution feeders. In: **XI International Symposium on Lightning Protection**. [S.l.: s.n.], 2011.
- LE, C. D.; GU, I. Y. H.; BOLLEN, M. H. J. Joint causal and anti-causal segmentation and location of transitions in power disturbances. In: **Power and Energy Society General Meeting**. [S.l.: s.n.], 2010.
- LEE, J.; LEE, D. An improved cluster labeling method for support vector clustering. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 27, p. 461–4, 2005.
- LEE, J.; LEE, D. Dynamic characterization of cluster structures for robust and inductive support vector clustering. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 28, p. 1869–74, 2006.

- LIMA, M. A. A.; COURY, D. V.; CERQUEIRA, A. S.; NASCIMENTO, V. H. A method based on independent component analysis for adaptive decomposition of multiple power quality disturbances. **Journal of Control, Automation and Electrical Systems**, v. 25, p. 80–92, 2014.
- LOBOS, T.; LEONOWICZ, Z.; REZMER, J.; SCHEGNER, P. High-resolution spectrum-estimation methods for signal analysis in power systems. **IEEE Transactions on Instrumentation and Measurement**, v. 55, p. 1–8, 2006.
- MA, J.; PERKINS, S. Time-series novelty detection using one-class support vector machines. In: **Proceedings of the International Joint Conference on Neural Networks**. [S.l.]: IEEE, 2003. p. 1741–1745.
- MAATEN, L. van der; HINTON, G. Visualizing high-dimensional data using t-SNE. **Journal of Machine Learning Research**, v. 9, p. 2579–2605, 2008.
- MACGORMAN, D. R.; RUST, W. D. **The Electrical Nature of Storms**. [S.l.]: Oxford University Press, 1998.
- MAHANTY, R. N.; GUPTA, P. B. D. Application of RBF neural network to fault classification and location in transmission lines. **IEEE Generation, Transmission and Distribution**, v. 151, p. 201–212, 2004.
- MAHANTY, R. N.; GUPTA, P. B. D. A fuzzy logic based fault classification approach using current samples only. **Electric Power Systems Research**, v. 77, p. 501–507, 2007.
- MALATHI, V.; MARIMUTHU, N. S.; BASKAR, S. Intelligent approaches using support vector machine and extreme learning machine for transmission line protection. **Neurocomputing**, Elsevier, v. 73, p. 2160–2167, 2010.
- MALLAT, S. **A Wavelet Tour of Signal Processing**. 2nd. ed. [S.l.]: Academic Press, 1999.
- MARKOU, M. Novelty detection: a review-part 1: Statistical approaches. **Signal Processing**, v. 83, p. 2481–2497, 2003.
- MARKOU, M. Novelty detection: a review-part 2: Neural network based approaches. **Signal Processing**, v. 83, p. 2499–2521, 2003.
- MASON, R. **The Art and Science of Protective Relaying**. [S.l.]: General Electric, 2000.
- MASUD, M. M.; CHEN, Q.; GAO, J. Classification and Novel Class Detection of Data Streams in a Dynamic Feature Space. In: **European Conference on Machine learning and Knowledge Discovery in Databases: Part II**. [S.l.: s.n.], 2010. p. 337–352.
- MASUD, M. M.; CHEN, Q.; KHAN, L.; AGGARWAL, C.; GAO, J.; HAN, J.; THURAI-SINGHAM, B. Addressing concept-evolution in concept-drifting data streams. In: **IEEE International Conference on Data Mining**. [S.l.: s.n.], 2010. p. 929–934.
- MASUD, M. M.; GAO, J. Integrating novel class detection with classification for concept-drifting data streams. In: **European Conference on Machine Learning and Knowledge Discovery in Databases: Part II**. [S.l.: s.n.], 2009. p. 79–94.
- MATHWORKS. **Wavelet toolbox users guide - The Mathworks Inc**. [S.l.], 2009.

- MELO, K.; COSTA, F. B.; SAYONARA, S.; BRITO, N. S. D.; SOUZA, B. A.; DANTAS, K. M. C. Detecção e classificação de faltas a partir da análise de registros oscilográficos via redes neurais artificiais e transformada wavelet. **Revista Controle e Automação**, v. 18, p. 163–172, 2007.
- MICHALIK, M.; REBIZANT, W.; LUKOWICZ, M.; LEE, S.-J. High-impedance fault detection in distribution networks with use of wavelet-based algorithm. **IEEE Transactions on Power Delivery**, v. 21, p. 1793–1802, 2006.
- MILLER, D.; BROWNING, J. A mixture model and EM-based algorithm for class discovery, robust classification, and outlier rejection in mixed labeled/unlabeled data sets. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 25, p. 1468–1483, 2003.
- MORAIS, J.; PIRES, Y.; CARDOSO, C.; KLAUTAU, A. Data mining applied to the electric power industry: classification of short-circuit faults in transmission lines. In: **Proceedings of the 7th International Conference on Intelligent Systems Design and Application**. [S.l.: s.n.], 2007. p. 943–948.
- MORETO, M.; ROLIM, J. G. Segmentação automática de registros oscilográficos: Um método voltado para geradores de energia. In: **XVIII Congresso Brasileiro de Automática**. [S.l.: s.n.], 2010. p. 1272–1279.
- MORETO, M. M.; ROLIM, J. G. Análise automática de oscilografias em sistemas elétricos de potência. **Revista Controle e Automação**, v. 21, p. 347–362, 2010.
- NEHMZOW, U. **Scientific Methods in Mobile Robotics**. [S.l.]: Springer, 2006.
- NISHIYAMA, K. A nonlinear filter for estimating a sinusoidal signal and its parameters in white noise: On the case of a single sinusoid. **IEEE Transactions on Signal Processing**, v. 45, p. 970–981, 1997.
- NISHIYAMA, K. Nonlinear estimation of harmonic signals. **IEEE Transactions Fundamentals**, E90–A, p. 1021–1027, 2007.
- NUNEZ, V. B.; KULKARNI, S.; SANTOSO, S.; MELENDEZ, J. Feature analysis and classification methodology for overhead distribution fault events. In: **IEEE/PES IEEE Transmission and Distribution Committee**. [S.l.: s.n.], 2010.
- OLIVEIRA, A. L.; COSTA, F. R.; FILHO, C. O. Novelty detection with constructive probabilistic neural networks. **Neurocomputing**, v. 71, p. 1046–1053, 2008.
- OLIVEIRA, C. C. B.; SCHMIDT, H. P.; KAGAN, N.; ROBBA, E. J. **Introdução a Sistemas Elétricos de Potência**. [S.l.]: Bulcher, 2000.
- PAN, Z.-W.; XIANG, D.-H.; XIAO, Q.-W.; ZHOU, D.-X. Parzen windows for multi-class classification. **Journal of Complexity**, v. 24, p. 606–618, 2008.
- PAOLONE, M.; RACHIDI, F.; MEMBER, S.; BORGHETTI, A.; NUCCI, C. A.; RUBINS-TEIN, M.; RAKOV, V. A.; UMAN, M. A. Lightning electromagnetic field coupling to overhead lines: Theory, numerical simulations, and experimental validation. **IEEE Transactions on Electromagnetic Compatibility**, v. 51, p. 532–547, 2009.

- PELLEG, D.; MOORE, A. Accelerating exact  $k$ -means algorithms with geometric reasoning. In: **5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. [S.l.: s.n.], 1999.
- PELLEG, D.; MOORE, A. X-means: Extending  $k$ -means with efficient estimation of the number of clusters. In: **17th International Conference on Machine Learning**. [S.l.: s.n.], 2000.
- PRATI, R. C.; MONARD, G. E. A. P. A.; BATISTA, M. C. Curvas ROC para avaliação de classificadores. **IEEE Latin America Transactions**, v. 6, p. 1–8, 2008.
- RAVAGLIO, M.; TOLEDO, L.; BACCA, I. A.; CABUSSU, M. S.; SALLES, R. N.; CERQUEIRA, D. P. Metodologia para gestão e avaliação de danos elétricos. In: **Anais do XIX Seminário Nacional de Distribuição de Energia Elétrica**. [S.l.: s.n.], 2010.
- RAVIKUMAR, B.; THUKARAM, D.; KHINCHA, H. P. Application of support vector machines for fault diagnosis in power transmission system. **IET Generation, Transmission and Distribution**, v. 2, p. 119–130, 2008.
- REDDY, M. J.; MOHANTA, D. K. A wavelet-fuzzy combined approach for classification and location of transmission line faults. **Electrical Power and Energy Systems**, v. 29, p. 669–678, 2007.
- SAMANTARAY, S. R.; DASH, P. K. Transmission line distance relaying using machine intelligence technique. **IET Generation, Transmission and Distribution**, v. 2, p. 53–61, 2008.
- SANTAMAR, I.; PANTALEON-PRIETO, C.; FIGUEIRAS-VIDAL, A. R. Joint segmentation and AR modeling of quasistationary signals using the EM algorithm. In: **IEEE International Conference on Acoustics, Speech, and Signal Processing**. [S.l.: s.n.], 1999.
- SANTOSO, S.; MCGRANAGHAN, M. F.; CROSSLEY, P. A.; RIBEIRO, M. V.; RIBEIRO, P. F. Assessing power system quality using signal processing techniques. **IEEE Signal Processing Magazine**, v. 12, p. 12–31, 2009.
- SARLAK, M.; SHAHRTASH, S. High impedance fault detection using combination of Multi-Layer Perceptron neural networks based on multi-resolution morphological gradient features of current waveform. **IET Generation, Transmission & Distribution**, v. 5, p. 588, 2011.
- SCHEIRER, W. J.; JAIN, L. P.; BOULT, T. E. Probability models for open set recognition. **IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)**, v. 36, p. 1–8, 2014.
- SCHEIRER, W. J.; ROCHA, A. R.; SAPKOTA, A.; BOULT, T. E. Toward open set recognition. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 35, p. 1757–72, 2013.
- SCHÖLKOPF, B.; SMOLA, A. J. **Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond**. [S.l.]: MIT Press, 2001.
- SILVA, K.; SOUZA, B.; BRITO, N. Fault detection and classification in transmission lines based on wavelet transform and ANN. **IEEE Transactions on Power Delivery**, v. 21, p. 2058–2063, 2006.



SIMARD, G.; CHARTRAND, D.; CHRISTOPHE, P. Distribution automation: Applications to move from today's distribution system to tomorrow's smartgrid. In: **2009 IEEE Power & Energy Society General Meeting**. [S.l.: s.n.], 2009.

SPINOSA, E. J. **Detecção de Novidade com Aplicação a Fluxos Contínuos de Dados**. Tese (Doutorado) — Universidade de São Paulo, 2008.

SPINOSA, E. J.; CARVALHO, C. P. L. F. Combining one-class classifiers for robust novelty detection in gene expression data. **Advances in Bioinformatics and Computational Biology**, v. 3594, p. 54–64, 2005.

STYVAKTAKIS, E. **Automating Power Quality Analysis**. Tese (Doutorado) — Chalmers University of Technology, 2002.

STYVAKTAKIS, E.; BOLLEN, M.; GU, I. Expert system for voltage dip classification and analysis. In: **IEEE Power Engineering Society Summer Meeting**. [S.l.]: Ieee, 2001. p. 671–676.

STYVAKTAKIS, E.; BOLLEN, M. H. J.; MEMBER, S.; GU, I. Y. H. Power system events. **IEEE Transactions on Power Delivery**, v. 17, p. 423–428, 2002.

SU, S.; DUAN, X.; ZENG, X. ATP-based automated fault simulation. **IEEE Transactions on Power Delivery**, v. 23, p. 1687–1689, 2008.

TAN, S.; LIM, C.; RAO, M. A hybrid neural network model for rule generation and its application to process fault detection and diagnosis. **Engineering Applications of Artificial Intelligence**, v. 20, n. 2, p. 203–213, mar. 2007.

TAVARES, J.; PERETTA, I. S.; LIMA, G. F. M.; YAMANAKA, K.; PAIS, S. Algoritmo baseado no operador de energia de teager (TEO) para segmentação de linhas e palavras em textos manuscritos e impressos. In: **VII Workshop de Visão Computacional**. [S.l.: s.n.], 2011.

TAX, D. **DDtools, the Data Description Toolbox for Matlab**. Maio 2012. Versão 1.9.1.

TAX, D.; DUIN, R. Growing a multi-class classifier with a reject option. **Pattern Recognition Letters**, v. 29, p. 1565–1570, 2008.

TAX, D. M. J. **One-class Classification**. Tese (Doutorado) — Technische Universiteit Delft, 2001.

TAX, D. M. J.; DUIN, R. P. W. Combining one-class classifiers. In: **Second International Workshop on Multiple Classifier Systems**. [S.l.: s.n.], 2001. p. 299–308.

THEODORIDIS, S.; KONSTANTINOS, K. **Pattern Recognition**. [S.l.]: Elsevier, 2009.

TORO, V. D. **Fundamentos de máquinas elétricas**. [S.l.]: Editora LTC, 1999.

UKIL, A.; IVANOVI, R. Application of abrupt change detection in power systems disturbance analysis and relay performance monitoring. **IEEE Transactions on Power Delivery**, v. 22, p. 59–66, 2007.

UKIL, A.; ZIVANOVIC, R. Abrupt change detection in power system fault analysis using wavelet transform. In: **International Conference on Power Systems Transients**. [S.l.: s.n.], 2005.

- UKIL, A.; ZIVANOVIC, R. Detection of abrupt changes in power system fault analysis: A comparative study. In: **Southern African Universities Power Engineering Conference**. [S.l.: s.n.], 2005.
- UKIL, A.; ZIVANOVIC, R. Abrupt change detection in power system fault analysis using adaptive whitening filter and wavelet transform. **Electric Power Systems Research**, v. 76, p. 815–823, 2006.
- UKIL, A.; ZIVANOVIC, R. Automatic signal segmentation based on abrupt change detection for power systems applications. In: **IEEE Power India Conference**. [S.l.: s.n.], 2006. p. 33–40.
- UKIL, A.; ZIVANOVIC, R. The detection of abrupt changes using recursive identification for power system fault analysis. **Electric Power Systems Research**, v. 77, p. 259–265, 2007.
- UKIL, A.; ZIVANOVIC, R. Adjusted Haar wavelet for application in the power systems disturbance analysis. **Digital Signal Processing**, v. 18, p. 103–115, 2008.
- UMAN, M. A. **The Lightning Discharge**. [S.l.]: Dover Publications, INC, 2001.
- VAPNIK, V. N. **Statistical Learning Theory**. New York: Wiley-Interscience, 1998.
- VASILIC, S.; KEZUNOVIC, M. Fuzzy ART neural network algorithm for classifying the power system faults. **IEEE Transactions on Power Delivery**, v. 20, n. 2, p. 1306–1314, abr. 2005.
- VIEIRA-NETO, H. **Visual Novelty Detection for Autonomous Inspection Robots**. Tese (Doutorado) — University of Essex, England, 2006.
- VISACRO-FILHO, S. **Descargas Atmosféricas: Uma Abordagem de Engenharia**. [S.l.]: Artiliber Editora, 2005.
- WANG, C.-D.; LAI, J.-H.; HUANG, D. Kernel-based clustering with automatic cluster number selection. In: **11th International Conference on Data Mining Workshops**. [S.l.: s.n.], 2011.
- WEINBERGER, K. Q.; SAUL, L. K. Distance metric learning for large margin nearest neighbor classification. **Journal of Machine Learning Research**, v. 10, p. 207–244, 2009.
- WIOT, D. A new adaptive transient monitoring scheme for detection of power system events. **IEEE Transactions on Power Delivery**, v. 19, p. 42–48, 2004.
- WU, M.; YE, J. A small sphere and large margin approach for novelty detection using training data with outliers. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 31, p. 2088–92, 2009. ISSN 1939-3539.
- XU, L.; CHOW, M.-Y. A classification approach for power distribution systems fault cause identification. **IEEE Transactions on Power Systems**, v. 21, p. 53–60, 2006.
- XU, L.; CHOW, M.-Y.; TAYLOR, L. S. Power distribution fault cause identification with imbalanced data using the data mining-based fuzzy classification E-algorithm. **IEEE Transactions on Power Systems**, v. 22, p. 164–171, 2007.
- XU, L.; CHOW, M.-Y.; TIMMIS, J.; TAYLOR, L. S. Power distribution outage cause identification with imbalanced data using artificial immune recognition system (AIRS) algorithm. **IEEE Transactions on Power Systems**, v. 22, n. 1, p. 198–204, 2007.

XU, Z. E.; WEINBERGER, K. Q.; CHAPELLE, O. **Distance Metric Learning for Kernel Machines**. [S.l.], 2011.

ZAMORA, I.; MAZON, A. J.; SAGASTABEITIA, K. J.; ZAMORA, J. J. New method for detecting low current faults in electrical distribution systems. **IEEE Transactions on Power Delivery**, v. 22, p. 2072–2079, 2007.

ZELNIK-MANOR, L.; PERONA, P. Self-tuning spectral clustering. In: **Advances in Neural Information Processing Systems**. [S.l.: s.n.], 2004. v. 17.

ZENG, H.; CHEUNG, Y.-M. A new feature selection method for Gaussian mixture clustering. **Pattern Recognition**, v. 42, p. 243–250, 2009.

ZHANG, N.; KEZUNOVIC, M. A real time fault analysis tool for monitoring operation of transmission line protective relay. **Electric Power Systems Research**, v. 77, p. 361–370, 2007.

ZHAO, X.-M.; CHEUNG, Y.-M.; HUANG, D.-S. Analysis of gene expression data using RPEM algorithm in normal mixture model with dynamic adjustment of learning rate. **International Journal of Pattern Recognition and Artificial Intelligence**, v. 24, p. 651–666, 2010.

ZWEIG, A.; ESHAR, D.; WEINSHALL, D. **Detection and Identification of Rare Audiovisual Cues**. [S.l.]: Springer, 2012.

## APÊNDICE A – PUBLICAÇÕES RELACIONADAS

A seguir serão detalhadas as publicações e submissões relacionadas aos temas apresentados, no período da realização do presente trabalho de doutorado, o qual corresponde aos anos de 2011 a 2015.

Artigos em periódicos internacionais:

- André Eugênio Lazzaretti, Vitor Hugo Ferreira, Hugo Vieira Neto, Rodrigo Jardim Riella and Julio Shigeaki Omori, Autonomous Neural Models for the Classification of Events in Power Distribution Networks. *Journal of Control, Automation and Electrical Systems*, 24(5):612-622, 2013;

Artigos em conferências internacionais:

- André Eugênio Lazzaretti, Luiz Felipe Ribeiro Barrozo Toledo, Marcelo Antônio Ravaglio, José Arinos Teixeira Júnior; Patrício Enrique Muñoz Rojas e Cleverson Luiz da Silva Pinto. Measurements of Lightning Discharges in Overhead Distribution Feeders. *XI International Symposium on Lightning Protection*, 2011, Fortaleza, Ceará, Brazil;
- André Eugênio Lazzaretti, Vitor Hugo Ferreira, Hugo Vieira Neto, Luiz Felipe Ribeiro Barrozo Toledo and Cleverson Luiz da Silva Pinto, A New Approach for Event Classification and Novelty Detection in Power Distribution Networks. In Proceedings of the *IEEE PES General Meeting*, 2013, IEEE PES Press, Vancouver, Canada, 2013. Selected for the Best Papers on System Operations and Market Economics paper session;
- André Eugênio Lazzaretti, Vitor Hugo Ferreira, Hugo Vieira Neto, Luiz Felipe Ribeiro Barrozo Toledo and Cleverson Luiz da Silva Pinto, A New Approach for Event Detection in Smart Distribution Oscillograph Recorders. In Proceedings of the *II IEEE PES Conference on Innovative Smart Grid Technologies - Latin America*, IEEE PES Press, São Paulo, Brazil, 2013.

Artigos em conferências nacionais:

- André Eugênio Lazzaretti, Vitor Hugo Ferreira, Hugo Vieira Neto and Cleverson Luiz da Silva Pinto, Agrupamento Automático de Oscilografias em Redes de Distribuição. In *Anais do XX Congresso Brasileiro de Automática*, Belo Horizonte, Brazil, 2014, pp. 715-722.

Artigos submetidos para periódicos internacionais:

- *Expert Systems with Applications* (Elsevier): André Eugênio Lazzaretti, David M. J. Tax, Hugo Vieira Neto e Vitor Hugo Ferreira. Novelty Detection and Multi-Class Classification in Power Distribution Voltage Waveforms (submetido em Novembro de 2014).

Artigos aceitos em periódicos internacionais:

- *IEEE Transactions on Power Delivery*: André Eugênio Lazzaretti, Hugo Vieira Neto e Vitor Hugo Ferreira. An Accurate Approach for Automatic Segmentation in Power Distribution Voltage Waveforms (aceito em Fevereiro de 2015).