

CURVAS PRINCIPAIS APLICADAS NA IDENTIFICAÇÃO DE DESCARGAS  
PARCIAIS EM EQUIPAMENTOS DE POTÊNCIA

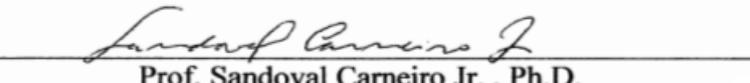
José Márcio Faier

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS  
PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE  
FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS  
NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM  
ENGENHARIA ELÉTRICA.

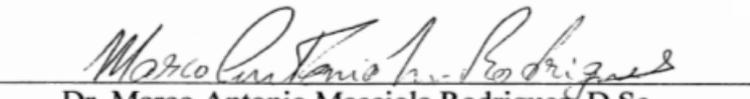
Aprovada por:



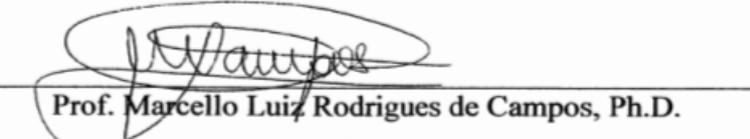
Prof. José Manoel de Seixas, D.Sc.



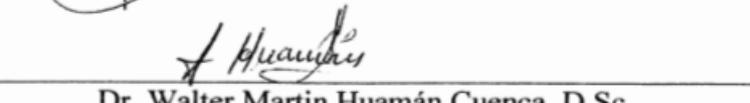
Prof. Sandoval Carneiro Jr., Ph.D.



Dr. Marco Antonio Macciola Rodrigues, D.Sc.



Prof. Marcello Luiz Rodrigues de Campos, Ph.D.



Dr. Walter Martin Huamán Cuenca, D.Sc.

RIO DE JANEIRO, RJ - BRASIL

MARÇO DE 2006

FAIER, JOSÉ MÁRCIO

Curvas Principais Aplicadas na  
Identificação de Descargas Parciais em  
Equipamentos de Potência [Rio de Janeiro] 2006

IX, 102 p. 29,7 cm (COPPE/UFRJ, M.Sc.,  
Engenharia Elétrica, 2006)

Dissertação - Universidade Federal do Rio  
de Janeiro, COPPE

1. Curvas Principais 2. Descargas Parciais 3.  
Reconhecimento de Padrões 4. Classificação 5.  
Equipamentos de Potência

I. COPPE/UFRJ II. Título ( série )

# Agradecimentos

Agradeço aos meus pais, que me deram todas as condições para que eu pudesse ir a busca dos meus objetivos.

Agradeço aos pais de minha namorada e, especialmente, agradeço a minha namorada, Mônica, por todo amor que tem por mim e pela força que me ajudaram a terminar este trabalho.

Agradeço ao meu orientador, professor Seixas, por toda a ajuda, paciência e sabedoria. Um amigo.

Agradeço a Alain Levy, que colaborou com o seu conhecimento no desenvolvimento deste trabalho.

Agradeço a todos que, direta ou indiretamente me ajudaram a terminar mais uma etapa de minha vida. Em particular, agradeço a Helena Fernandez e Walter Martin que contribuíram com informações importantes para a realização desta dissertação.

Agradeço também a COPPE por me aceitar como estudante e ao CEPTEL pela aceitação como bolsista e pelo apoio financeiro durante este tempo.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

## CURVAS PRINCIPAIS APLICADAS NA IDENTIFICAÇÃO DE DESCARGAS PARCIAIS EM EQUIPAMENTOS DE POTÊNCIA

José Márcio Faier

Março/2006

Orientador: José Manuel de Seixas

Programa: Engenharia Elétrica

Neste trabalho, aplica-se o método das curvas principais (CPs) para identificar e classificar padrões em descargas parciais (DPs), a partir de seus modelos experimentais. Os dados são representados através de histogramas bidimensionais preenchidos de acordo com a amplitude das descargas e a fase da tensão de teste na qual elas ocorrem. A informação é analisada através da concatenação ou da acumulação dos sinais de descarga obtidos.

A distância dos padrões de DPs às curvas principais (extraídas para cada um dos três modelos experimentais disponíveis) é utilizada como informação por diferentes métodos de classificação, a saber: distância Euclidiana mínima, redes neurais e pontuação percentual. Devido às restrições estatísticas, os modelos baseados em CPs são avaliados através de dois enfoques: priorizando a informação especialista em DPs ou a estrutura estatística dos dados disponíveis.

A extração das curvas principais, através da concatenação dos canais, mostrou-se simples, robusta e eficaz. A eficiência do classificador neural foi superior a 98%. Já os classificadores com informação acumulativa não se mostraram robustos, mas se mostraram atraentes na visualização dos modelos extraídos em três dimensões.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

## PRINCIPAL CURVES FOR PARTIAL DISCHARGE IDENTIFICATION IN POWER EQUIPAMENTS

José Márcio Faier

March/2006

Advisor: José Manuel de Seixas

Department: Electrical Engineering

This work applies Principal Curve Method (PC) to identify and classify partial discharges (PD) patterns, based on PD experimental models. Data are represented by bi-dimensional histograms, filled in by the discharge amplitudes and test voltage phase in which they occur. The information is analyzed by concatenation or accumulation of the DP signals.

The PD pattern to Principal Curve (extracted for each one of the three available experimental models) distance is used as input by different classification methods: minimum Euclidean distance, neural networks and percent score. Due to data statistical restrictions, the PC models are evaluated by two different approaches: focusing on expert PD knowledge or statistical representation of the available data.

The Principal curves extraction, by means of concatenation, proved to be simple, robust and efficient. The neural classifier efficiency was found to be above 98%. The method based on data accumulation did not show robustness, but allowed model visualization.

# Sumário

<b>1 Introdução.....</b>	<b>1</b>
1.1 Contexto.....	2
1.2 Motivação.....	2
1.3 Escopo do Trabalho.....	3
<b>2 Descargas Parciais.....</b>	<b>6</b>
2.1 Ionização.....	6
2.2 Conceito.....	8
2.3 Detecção e Medição.....	9
2.3.1 Métodos Elétricos.....	10
2.3.2 Aspectos Relevantes da Detecção e Medição de DPs.....	12
2.4 Dados de Descargas Parciais.....	12
2.4.1 Dados Associados à Fase.....	13
2.4.2 Dados Associados ao Tempo.....	14
2.4.3 Dados não Associados nem à Fase nem ao Tempo.....	14
2.5 Conclusão.....	14
<b>3 Reconhecimento de Padrões em DPs.....</b>	<b>16</b>
3.1 Extração de Características.....	16
3.1.1 Métodos Estatísticos.....	17
3.1.2 Ferramentas de Caracterização do Pulso Temporal.....	19
3.1.3 Ferramentas de Processamento de Sinais.....	21
3.1.3.1 Transformada de Fourier.....	21
3.1.3.2 Transformada Haar e Walsh.....	22
3.1.3.3 Transformada Wavelet.....	22
3.1.4 Ferramentas de Processamento de Imagens.....	23
3.1.4.1 Algoritmos de Análise de Textura.....	23
3.1.4.2 Características Fractais.....	24
3.1.4.3 Decomposição de Imagens Baseada em Wavelets.....	24
3.1.5 Aproximação de Séries Temporais.....	25
3.2 Classificação.....	25
3.2.1 Classificadores Baseados na Distância.....	25

3.2.1.1 Distância Mínima.....	26
3.2.1.2 Votação ou Pontuação Percentual.....	26
3.2.2 Classificadores Estatísticos.....	27
3.2.2.1 Classificador de Bayes.....	27
3.2.2.2 Classificador Baseado na Taxa de Reconhecimento.....	27
3.2.3 Classificadores Baseados em Redes Neurais Artificiais.....	28
3.2.2.1 Redes Neurais Backpropagation.....	29
3.2.2.2 Mapas Auto-Organizáveis de Kohonen.....	30
3.2.2.3 Rede de Quantização do Vetor Aprendizagem.....	30
3.2.2.3 Redes Neurais Counterpropagation.....	31
3.2.2.3 Redes Neurais Modulares.....	32
3.2.2.3 Redes Neurais Cascadeadas.....	32
3.2.4 Classificadores baseados em lógica nebulosa.....	33
3.3 Análise do Desempenho.....	33
3.3.1 Classificação usando dados com informação de fase.....	33
3.3.2 Classificação usando dados com informação temporal.....	35
3.3.3 Classificação usando dados sem informação de fase/tempo.....	36
3.3 Conclusão.....	36
<b>4 Curvas Principais.....</b>	<b>37</b>
4.1 Componentes Principais.....	37
4.2 Definição de Curvas Principais.....	40
4.2.1 Definição Clássica.....	40
4.2.2 Definição de Pontos Orientados.....	42
4.3 Algoritmos de Extração de Curvas Principais.....	43
4.3.1 O algoritmo K-segmentos Não-Suave.....	44
4.4 Aplicações.....	46
4.4 Conclusão.....	48
<b>5 Base de Dados.....</b>	<b>49</b>
5.1 Coleta dos Dados.....	50
5.1.1 Modelo experimental.....	50
5.1.1 Equipamento de Medição.....	52
5.2 Tratamento das informações.....	53
5.3 Separação da Base de Dados.....	54

5.3.1 Enfoque Especialista.....	54
5.3.2 Enfoque estatístico.....	55
5.4 Conclusão.....	57
<b>6 Curvas Principais Aplicadas no Reconhecimento de Padrões de DPs.....</b>	<b>58</b>
6.1 Extração de Características.....	58
6.1.1 Mapas Concatenados (1024 Dimensões).....	59
6.1.2 Mapas Acumulativos (3D).....	60
6.2 Classificação.....	62
6.2.1 Classificação em 1024 dimensões.....	63
6.2.1.1 Classificação pela Distância Mínima.....	63
6.2.1.2 Classificação utilizando Redes Neurais.....	63
6.2.2 Classificação 3D.....	64
6.2.2.1 Classificação pela Distância Mínima.....	64
6.2.2.2 Classificação utilizando Redes Neurais.....	65
6.2.2.3 Classificação por Votação.....	65
6.3 Conclusão.....	66
<b>7 Resultados.....</b>	<b>67</b>
7.1 Extração de Características.....	67
7.1.1 Extração em 1024 Dimensões.....	68
7.1.1.1 Visão Especialista.....	68
7.1.1.2 Visão Estatística.....	70
7.1.2 Extração 3D.....	73
7.1.2.1 Visão Especialista.....	73
7.1.2.2 Visão Estatística.....	76
7.2 Classificação.....	78
7.2.1 Classificação em 1024 Dimensões.....	78
7.2.1.1 Distância Euclidiana Mínima.....	78
7.2.1.2 Rede Neural.....	80
7.2.2 Classificação 3D.....	83
7.2.2.1 Distância Euclidiana Mínima.....	84
7.2.2.2 Rede Neural.....	87
7.2.2.3 Votação.....	89
7.3 Quadro Comparativo.....	92

7.4 Conclusão.....	93
<b>8 Análises e Conclusões.....</b>	<b>94</b>
8.1 Perspectivas Futuras.....	96
<b>Referências .....</b>	<b>97</b>
<b>Apêndice – Produção Científica.....</b>	<b>102</b>

# Capítulo 1

## Introdução

O reconhecimento de padrões em descargas parciais (DPs) [1] tem despertado um grande interesse na comunidade científica. Procuram-se métodos simples, robustos e eficazes capazes de diagnosticar o estado de funcionamento de equipamentos elétricos sem a necessidade da paralisação de suas operações. Dessa forma, diversas técnicas de análise de dados têm sido aplicadas no estudo do comportamento das DPs a fim de encontrar padrões relacionados ao estado de funcionamento de um equipamento [2].

A complexidade desse estudo está no fato das informações contidas nos dados estarem dispersas em dimensões elevadas. Além dessa característica, geralmente, não se dispõe de uma grande quantidade de eventos - devido à dificuldade na aquisição destes dados. Considerando-se as restrições estatísticas práticas, o desenvolvimento de modelos que tenham poucos parâmetros é bastante atraente, pois os graus de liberdade são reduzidos, o que limita a complexidade do modelo [3]. Dessa forma, modelos compactos podem ser de considerável importância para o processo de extração de características. A compactação deve ser eficiente a ponto de se obter somente as informações que sejam relevantes ao objetivo de reconhecimento de padrões de DPs. O que resultará em um menor tempo de processamento, dando possibilidades para um processamento em tempo real desta tarefa.

Avanços tecnológicos recentes têm facilitado o trabalho de reconhecimento de padrões em DPs. O uso de equipamentos digitais de detecção e medição possibilitou que informações relativas às descargas pudessem ser extraídas e armazenadas em bancos de dados. A crescente capacidade de processamento dos computadores e as novas técnicas de processamento estão possibilitando, cada vez mais, o avanço na análise de dados em geral.

## 1.1 Contexto

Atualmente, a concorrência entre as empresas está cada vez mais acirrada. Para sobreviverem a um mercado tão competitivo, as empresas precisam ser o mais eficiente possível. No setor de geração, transmissão e distribuição de energia elétrica esta regra não é diferente. Há necessidade da oferta de energia acompanhar as variações da curva de demanda ao longo do tempo, não podendo haver risco de “apagão”. Se por um lado a demanda exige um serviço de qualidade, por outro lado os custos envolvidos para se atender esta expectativa são elevados. Sendo assim, as companhias do setor elétrico estão sempre procurando uma maior eficiência na operação de seus equipamentos.

A manutenção preditiva é o conceito que vai ao encontro da necessidade deste aumento de eficiência. Neste tipo de manutenção, diagnosticam-se possíveis falhas com base em indicadores de defeitos que o próprio equipamento fornece durante o seu funcionamento. A pergunta que se faz é: quais são os melhores indicadores? Para ajudar a responder esta pergunta, diversos métodos, técnicas e praticas têm sido utilizados no intuito de determinar os indicadores de defeitos mais precisos. Para o caso específico de equipamentos de potência, por exemplo, transformadores de alta-tensão, a avaliação das descargas parciais tem se mostrado um excelente indicador de falhas futuras.

As descargas parciais são descargas elétricas que ocorrem de maneira intermitente e que podem ser acumuladas em histogramas. Estas informações são avaliadas por sofisticadas ferramentas de análise de sinais, para que sejam extraídos padrões representativos do comportamento de um equipamento. Este trabalho contribui com o estudo de DPs, aplicando-se a técnica de curvas principais (CPs) [4], produzindo classificadores compactos e que permitam a visualização dos modelos extraídos.

## 1.2 Motivação

Estudos anteriores mostraram um avanço no desempenho de classificação de mapas de DPs ao se utilizar métodos não lineares (NLPCA) [5] – quando comparado com estudos através de métodos lineares (PCA) [6] – indicando uma possível correlação de ordem superior nos dados, ou seja, não linearidades. Uma vez que, nestes trabalhos, são utilizados enfoques diferentes na seleção de sinais de DPs - que comporão o conjunto de eventos para o desenvolvimento do classificador e o conjunto para o teste da capacidade de generalização - a comparação dos resultados entre estes trabalhos não

deve ser realizada de forma direta. Assim, a técnica das curvas principais, que pode ser interpretada com uma generalização não linear das componentes principais, é utilizada na investigação de mapas de DPs. Além disso, a técnica tem se mostrado eficaz no reconhecimento e classificação de padrões em diferentes tipos de dados [7]. E, ainda, de acordo com a configuração dos mapas de DPs, as CPs oferecem a possibilidade de visualização dos modelos extraídos, o que ajuda na prospecção dos dados.

### **1.3 Escopo do trabalho**

Este trabalho aplica um método de obtenção das curvas principais através do *algoritmo K-segmentos não suave* [8]. As curvas são extraídas de um conjunto de eventos selecionados criteriosamente para o desenvolvimento do classificador e para o teste da capacidade de generalização. Em seguida, os eventos são classificados, utilizando o método da distância mínima, processamento neural e votação [1]. Os conjuntos de desenvolvimento e de teste são escolhidos a partir de dois enfoques: o especialista [6] e o estatístico [5].

O primeiro enfoque seleciona os eventos para o conjunto de desenvolvimento (do qual serão extraídas as características) a partir de critérios estabelecidos por especialistas no assunto. O objetivo é separar para este conjunto os eventos menos ruidosos e também aqueles mais representativos do modelo - aproveitando-se todo o conhecimento acumulado em anos de experiência das pessoas que atuam na área. Este enfoque procura levar em consideração que, em campo, os sinais estarão envolvidos por ruídos. Assim, os eventos restantes (com mais ruídos e menos representativos) comporão o conjunto de teste, para que a capacidade de generalização do classificador seja avaliada.

No segundo enfoque, busca-se extrair o melhor classificador baseado no conhecimento da estatística dos dados disponíveis. Nesta visão, admite-se que o ruído faz parte dos modelos. Nesse caso, separam-se, para o conjunto de desenvolvimento, aqueles eventos que representam o modelo estatisticamente e, para o conjunto de teste, os eventos que melhor generalizam o modelo.

Neste contexto, analisam os mapas de DPs, que são na realidade histogramas bidimensionais preenchidos de acordo com a amplitude da descarga e a fase da tensão de teste em que elas ocorrem, de duas formas distintas. Ou concatenam-se os canais dos

histogramas na direção de uma serpentina, ou os canais com descargas acumuladas são indexados para formar amostras de três dimensões. Neste último caso, permite-se a visualização dos modelos no espaço 3D. No caso de histogramas com canais concatenados, os resultados aqui obtidos foram utilizados para a comparação com o desempenho de classificadores lineares (PCA) e não lineares (NLPCA) - uma vez que trabalhos que utilizaram estas técnicas também utilizaram a metodologia de canais concatenados.

Toda teoria, metodologia, resultados, análises e conclusões são apresentados neste documento, organizado da forma seguinte, sendo que cada capítulo é terminado através de uma conclusão que lhe conecta ao capítulo seguinte. Inicia-se com uma breve introdução, apresentando o problema de reconhecimento de padrões em DPs. No capítulo 2, apresenta-se o fenômeno de descargas parciais. Estuda-se o processo de ionização de moléculas que irá culminar em uma descarga parcial e apresenta-se o conceito de DP. Em seguida, são mostrados métodos de detecção e medição das descargas e alguns aspectos relevantes nesta medição. Apresentam-se, ainda, os parâmetros que definem os tipos de dados de DPs e como eles são classificados.

No capítulo 3, apresenta-se o estado da arte do reconhecimento de padrões em DPs. São discutidos os métodos utilizados por diversos autores na extração de características de descargas parciais. Em seguida, as técnicas de classificação aplicadas nas DPs são mostradas e faz-se uma revisão das redes neurais aplicadas no reconhecimento de padrões. Por fim, discutem-se os resultados obtidos pelos métodos pelos diversos métodos já aplicados em DPs.

No capítulo 4, apresentam-se as curvas principais. As curvas são definidas segundo a literatura original de CPs e de acordo com um conceito baseado em pontos orientados. Esta última definição é base para o algoritmo de extração de curvas principais utilizado nas análises deste trabalho – o K-segmentos não-suave. Por fim, algumas aplicações de curvas principais são mostradas.

No capítulo 5, mostra-se a base de dados utilizada nas análises deste trabalho. São apresentados os procedimentos de coleta dos dados, o modelo experimental e o equipamento de medição utilizado. O tratamento dado às informações, para tornar possível a aplicação do algoritmo de curvas principais, é então discutido. São ainda apresentados os métodos de separação da base de dados nos conjuntos de teste e de desenvolvimento.

No capítulo 6, apresenta-se o que foi realizado neste trabalho para se aplicar a técnica de curvas principais em dados de descargas parciais. Discute-se a extração das curvas principais a partir do conjunto de desenvolvimento. São mostrados os procedimentos tanto para dados concatenados quanto para dados acumulados.

No capítulo 7, apresentam-se os resultados obtidos. São mostrados os resultados da extração de características para mapas de DPs. Os procedimentos utilizados para se definir a quantidade de segmentos que compõem as curvas são mostrados. Os modelos passíveis de visualização são exibidos. Ainda, os resultados da classificação nas diferentes configurações são apresentados. Por fim, exibe-se um quadro comparativo dos resultados obtidos em todas as análises.

No capítulo 8, são feitas as análises e conclusões finais e apresentadas sugestões para a realização de trabalhos futuros.

# Capítulo 2

## Descargas Parciais

As descargas parciais (DPs) são sucessões de descargas elétricas que não chegam a completar um caminho fechado no interior de um meio isolante [1]. Estas descargas podem levar o sistema de isolamento de um equipamento a uma deterioração física e química. As DPs podem apresentar padrões que se relacionam às causas de sua origem. Isto é de grande interesse prático, uma vez que as descargas podem ser detectadas e medidas sem a necessidade de acessar os componentes internos do equipamento. Na medição devem ser levados em consideração os atributos que irão compor os padrões de dados de DPs, a saber: amplitude, duração e taxa de ocorrência dos pulsos, entre outros. Baseados nestes parâmetros existem, basicamente, três categorias em que os dados de DPs podem pertencer: descargas relacionadas à fase da tensão aplicada, DPs relacionadas ao tempo de duração e descargas parciais não relacionadas nem a fase nem ao tempo [2].

Além de classificar os dados de DPs, este capítulo apresentará o conceito clássico de descargas parciais presente em uma vasta quantidade de trabalhos sobre o assunto [1, 5, 6, 9]. Iniciaremos o entendimento do processo de formação das DPs discutindo a ionização de moléculas no interior de um meio isolante. Em seguida, serão analisados os aspectos relacionados aos processos práticos de detecção e medição de descargas parciais.

### 2.1 Ionização

O fenômeno de descargas parciais decorre de uma avalanche de elétrons provocada pelo processo de ionização dos átomos que compõem um material isolante [6, 9]. Para compreendermos melhor este processo, imaginemos um campo elétrico uniforme no interior de um material isolante. O campo elétrico será ocasionado por uma diferença de potencial ( $\vec{E} = \Delta V$ ).

O campo elétrico aplicado ao material dielétrico exerce uma força nos elétrons da camada mais externa dos átomos do material isolante. O que ocorre é a tentativa de “arrancá-los” da camada de valência. No entanto, como os elétrons são fortemente unidos a seus átomos, o que ocorre é uma polarização do átomo. Ou seja, há um deslocamento de cargas positivas e negativas de suas posições de equilíbrio original para posições mais periféricas no átomo. Os elétrons ficam sujeitos a forças contrárias ao sentido do campo e o núcleo fica sujeito a forças de mesmo sentido do campo. São estas forças que distorcem a estrutura atômica. Neste momento, os átomos comportam-se como dipolos, ou seja, os átomos comportam-se como duas cargas pontuais  $+q$  e  $-q$ . Aumentando-se o campo elétrico (através do aumento da diferença de potencial, por exemplo), as forças internas não serão mais capazes de manter os elétrons da última camada presos aos átomos. Haverá um desprendimento dos elétrons desta camada. Neste momento, o átomo encontra-se com mais cargas positivas (ionizado positivamente) e o elétron liberado encontra-se livre pelo dielétrico.

Este processo libera energia e pode culminar em uma descarga. No entanto, este não é o processo principal que ocasiona as descargas elétricas. Existe um processo subsequente que é a ionização ocasionada pelo elétron que foi liberado. Este elétron é acelerado pelo campo elétrico e colide com os átomos de moléculas neutras. Se a intensidade do campo for superior a um certo valor crítico, a colisão com as moléculas neutras será capaz de liberar elétrons destas. Os elétrons liberados agem da mesma forma que o primeiro elétron liberado, ou seja, as colisões se multiplicam cada vez mais no interior do dielétrico. Esse processo, ocorrendo sucessivamente, é a avalanche de elétrons (veja figura 2.1). Imaginando-se que o campo elétrico é produzido entre um eletrodo positivo e outro negativo, os íons positivos surgidos pelos desprendimentos dos elétrons movem-se vagorosamente na direção do eletrodo negativo. Assim, existe uma grande possibilidade de atração dos elétrons que estão vagando nas proximidades, voltando a se ter uma molécula neutra. Sabe-se que o nível de energia de uma molécula neutra é menor que o nível de energia de um íon positivo, logo, um quantum de energia deve ser liberado e irradia-se uma onda eletromagnética. A onda eletromagnética, conseqüência de absorção de potência da fonte, faz surgir uma centelha. Se esta centelha não atravessar completamente o material dielétrico, o que normalmente acontece, ocorre uma descarga parcial.

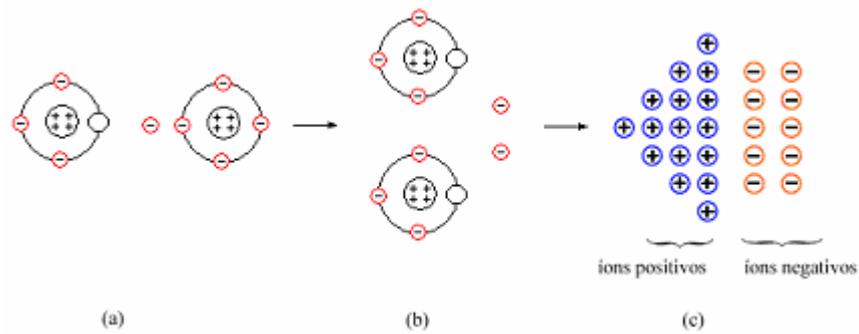


Figura 2.1 – Processo de avalanche de elétrons. (a) Desprendimento e colisão do elétron, (b) multiplicação dos elétrons livres e (c) avalanche de elétrons. Adaptada de [9].

## 2.2 Conceito

Descarga parcial é o nome dado à descarga elétrica envolvendo somente uma porção de um dielétrico, envolvido entre dois eletrodos, e que não chegam a fechar um circuito [1]. Esta é a definição clássica citada em vários textos. Descargas parciais também podem ocorrer em cavidades gasosas, em isolantes gasosos, líquidos e sólidos, em superfícies e em pontas de materiais sólidos. Esta última sentença é um ponto de discussão entre os especialistas do assunto. Descargas em volta de pontas devem ser definidas com descarga corona? Deveria uma descarga na superfície de um material isolante ser chamada de descarga parcial? Uma grande quantidade de trabalhos utiliza vários termos para descrever o mesmo fenômeno. Em Danikas [1], são analisadas e discutidas as várias definições dadas às DPs ao longo dos anos. Não entrando no mérito destas questões, as DPs serão aqui definidas de acordo com a definição clássica abordada em vários textos relacionados ao tema.

Descargas parciais são sucessões de descargas elétricas incompletas, rápidas e intermitentes, que ocorrem em um meio gasoso em série com isolantes sólidos ou líquidos [6, 9]. A conceituação de descargas incompletas se dá pela restrição quanto à zona do dielétrico. As DPs ocorrem em regiões definidas e não chegam a completar um circuito fechado (curto-circuito). As DPs são também consideradas rápidas pelo fato de ocorrerem em períodos curtíssimos de tempo, ao comparar-se com o período da tensão de teste (1/60 segundos). O tempo de duração de uma descarga parcial é da ordem de nano-segundos (*ns*), o que caracteriza as DPs como um fenômeno de alta-frequência. Ademais, as DPs são ditas intermitentes, o que significa dizer que, após o seu

surgimento, elas se extinguem rapidamente. Este fenômeno determina impulsos nos circuitos elétricos.

Quanto à ocorrência, as DPs podem surgir pela proximidade entre duas partes condutoras em um meio isolante. Tais descargas também podem ocorrer na zona de contato de isolantes sólidos e superfícies metálicas submetidas a uma diferença de potencial. Além disso, as ocorrências são favorecidas pela presença de cavidades gasosas no interior destes materiais.

Na figura 2.2 observa-se o formato de uma descarga parcial típica, medida através do pulso de corrente ao longo do tempo. Observa-se que, neste exemplo, o intervalo de duração da descarga é aproximadamente 10 ns e o pico de corrente é aproximadamente 120 mA.

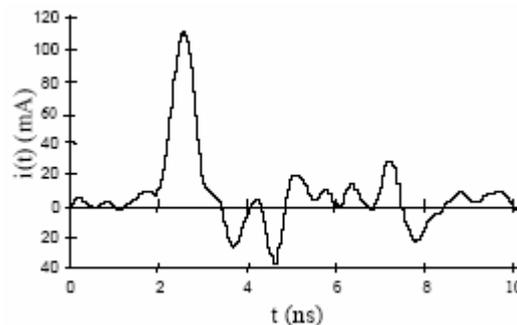


Figura 2.2 – Pulso de uma descarga parcial medida através da variação da corrente no tempo com pico em 2,5 nano-segundos. Extraído de [9].

### 2.3 Detecção e Medição das DPs

As descargas parciais apresentam diversas características que permitem sua detecção e medição. As DPs podem produzir pulsos de corrente, luminescências, ondas eletromagnéticas, ondas acústicas, consumo de energia, variações térmicas, variações químicas, vibrações mecânicas etc [6, 9]. As técnicas de detecção e medição partem da identificação destes fenômenos para reconhecer a respectiva descarga. Em sentido amplo, as técnicas de detecção e medição podem ser divididas em dois grupos: elétricos e não elétricos. O método elétrico insere o circuito de medição e detecção no circuito onde as descargas estão ocorrendo. Já os métodos não elétricos, geralmente, servem apenas de suporte aos métodos elétricos. Existem diversos métodos de detecção e medição não elétricos, dos quais pode se destacar:

- Métodos acústicos: capta ondas ultra-sônicas, não audíveis, através de sensores apropriados. Esta técnica tem merecido uma grande atenção no auxílio da localização de DPs no equipamento.
- Método óptico: Detecta a ocorrência de descargas através das emissões de luz provocadas no momento de ocorrência.
- Método químico: Detecta a presença de produtos dissolvidos no dielétrico. Por exemplo, uma análise do óleo isolante é capaz de determinar se houve ocorrência de descargas parciais no transformador.
- Método de rádio frequência: um receptor capta interferências provocadas pelas ondas eletromagnéticas. Assim, é possível identificar o lugar de ocorrência de DPs.

### 2.3.1 Métodos Elétricos

No sistema de medição elétrico, o circuito de detecção e medição, aqui denominado somente circuito de medição, torna-se parte integrante do meio onde estão ocorrendo as descargas parciais. Este meio comporta-se da mesma forma que indutâncias com capacitâncias em paralelo ou capacitâncias aterradas. O circuito de medição, que representa o equipamento de medição e algum circuito auxiliar, pode ser modelado por uma impedância RLC (resistiva, indutiva, capacitiva) ou uma impedância RC (resistiva, capacitiva) [6, 9].

As normas IEC 60270 [10] e IEEE P1434 [11] regem os procedimentos para a detecção e medição das descargas parciais. Nesse sentido, estas normas fornecem, dentre outras coisas, alternativas de circuitos de ensaio. Basicamente, têm-se três circuitos de medição, os quais detectam uma queda de tensão em uma impedância de medição, provocada por impulsos de corrente em um circuito exterior. A impedância de medição ( $Z_m$ ) pode ser dada por um resistor de medição, um indutor da ordem de  $mH$  e uma capacitância parasita inerente ao circuito. Na figura 2.3 é mostrado como a impedância de medição é configurada em conjunto com o circuito de ensaio ( $C_t$ ). Além desses componentes, filtros ( $Z$ ) e capacitores de acoplamento ( $C$ ) podem ser inseridos.

Os circuitos mostrados nas figuras 2.3-a e 2.3-b são de implementação mais simples e rápidas, porém são mais sensíveis a interferências. O circuito da figura 2.3-c é menos sensível a interferências, mas a calibração e sincronização do equipamento de

medição são dificultadas. O circuito da figura 2.3-a é utilizado em casos em que o objeto de ensaio tem uma extremidade aterrada. Essa configuração protege o equipamento de medição de falhas advindas do objeto de ensaio. O circuito da figura 2.3-b é mais indicado para o caso em que o lado de baixa tensão do objeto sob ensaio é isolado do referencial da terra. O circuito da figura 2.3-c é utilizado em casos que exigem menos interferências externas.

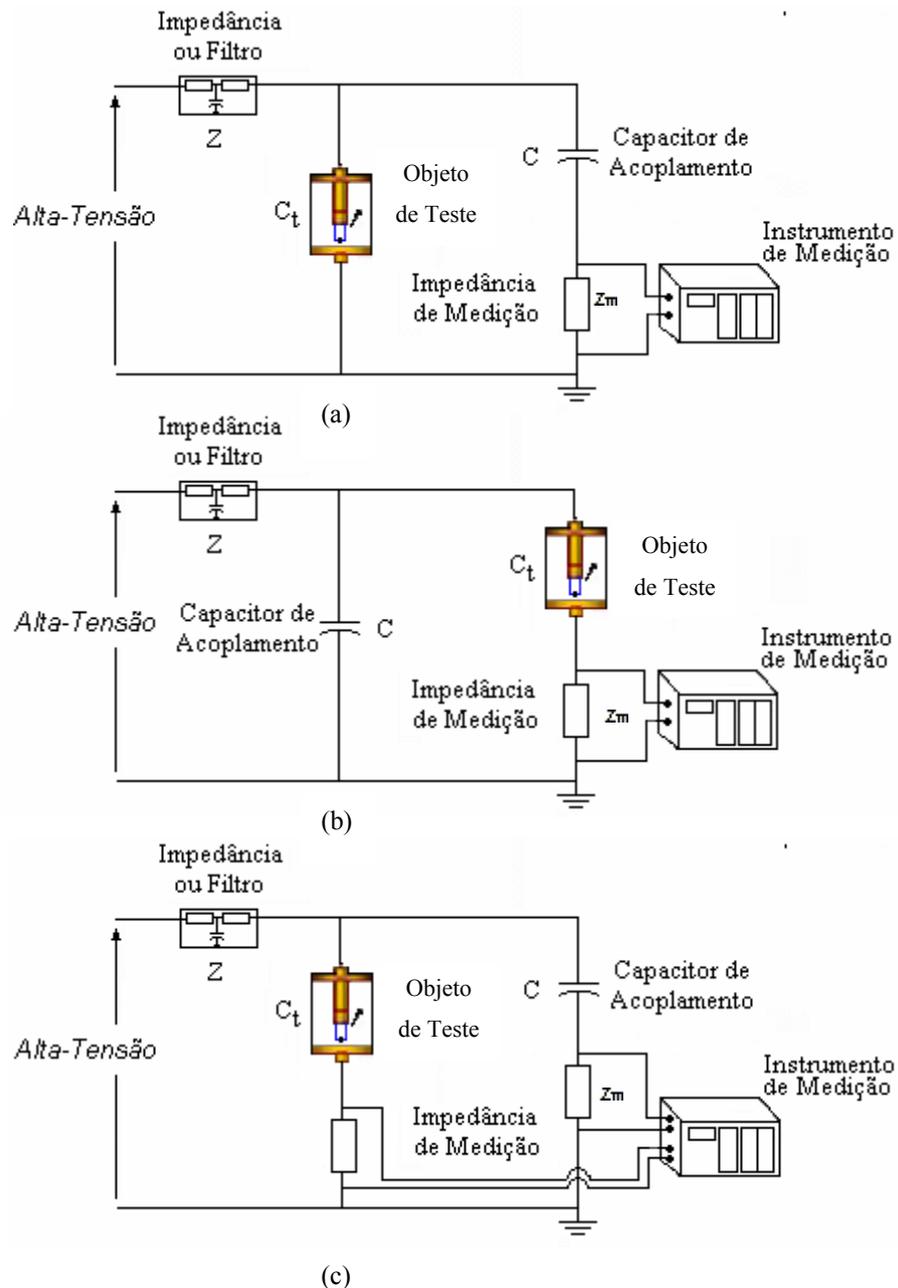


Figura 2.3 – Configurações sugeridas pelas normas para a detecção e medição de descargas parciais. Objeto de teste com (a) extremidade aterrada, (b) isolado da terra e (c) isolado da terra e configurado para haver menos interferência externa. Adaptada de [9].

### **2.3.2 Aspectos Relevantes da Detecção e Medição das DPs**

Ao se medir um sinal de descarga parcial, este pode estar atenuado ou deformado [12, 13]. A própria complexidade do circuito de medição é um fator que altera o sinal original. Além disso, existe a dificuldade de saber se a descarga medida advém da parte em teste do equipamento ou de uma parte distinta. E, ainda, existem inúmeras fontes de ruído que podem adicionar ruído ao sinal de DP.

Entre toda essa complexidade, há de se considerar a alteração que a própria descarga causa no seu meio, influenciando a próxima DP. Este fenômeno é conhecido como efeito de propagação ou efeito memória [13]. Podem ocorrer, ainda, sinais sobrepostos uns aos outros durante o processo de detecção, ou as descargas serem tão rápidas que o equipamento não consiga detectá-las.

Do ambiente exterior podem vir interferências propagadas através de condução ou irradiação de ondas eletromagnéticas, assim como: ruídos senoidais (a partir de serviços de comunicações, senóide da rede), ruídos pulsantes (a partir de comutadores de máquinas elétricas) e ruídos aleatórios (descargas corona etc). Todas estas interferências causam deformações no formato original das descargas parciais e são intrínsecos ao processo de medição. Assim, tanto em campo quanto em laboratório, sempre haverá a presença de algum tipo ruído, mais ou menos atenuado.

Apesar destes problemas, consegue-se, em laboratório principalmente, controlar o nível de interferência até valores toleráveis; quer seja por ajuste nos equipamentos de medição, quer seja por medições em ambientes mais apropriados. Além disso, técnicas de tratamento de ruído tornam os sinais de DPs mais “limpos” e possibilitam análises bastante satisfatórias [5].

## **2.4 Dados de descargas parciais**

Os processos de medição devem selecionar os parâmetros relativos às descargas parciais para armazená-los em um banco de dados. Basicamente, os dados de descargas parciais são classificados em três categorias distintas, de acordo com os parâmetros medidos, a saber: descargas associadas à fase da tensão aplicada, descargas associadas ao tempo de duração da descarga e DPs não associadas nem à fase nem ao tempo.

### 2.4.1 Dados associados à fase

Inicialmente, supõe-se que a tensão de teste é constante e os ângulos de fase desta onda de tensão são discretizados em faixas (“bins”). O detector digital mede a amplitude das descargas ( $q$ ), relacionadas às fases da tensão de teste ( $\varphi$ ), durante um intervalo de tempo. Durante este intervalo de tempo, os valores das amplitudes são acumulados e podem ser expressos pela razão ou taxa de ocorrência ( $n$ ). Estes dados de DP são classificados como dados associados à fase. Os *softwares* de análise exibem esses dados como distribuições univariadas em função da fase  $\varphi$  ou distribuições bivariadas - distribuições do tipo  $(\varphi, q, n)$ . As distribuições univariadas mais comuns [14] são:

- $(q_a, \varphi)$ : carga média em cada janela de fase vs.  $\varphi$
- $(q_p, \varphi)$ : pico de carga em cada janela de fase vs.  $\varphi$
- $(n, q)$ : quantidade de pulsos de DP detectadas em cada janela de fase vs.  $\varphi$
- $(I, \varphi)$ : corrente média da descarga em cada janela de fase vs.  $\varphi$

Os dados com distribuição bivariada são geralmente exibidos no espaço tridimensional pelos parâmetros  $(\varphi, q, n)$ . Esta exibição em 3 dimensões é a forma mais completa de representação gráfica do pulso de descarga parcial (veja figura 2.4). Histogramas das amplitudes dos pulsos ( $n$  vs.  $q$ ) podem ser obtidas diretamente ou indiretamente através de  $(\varphi, q, n)$  [15].

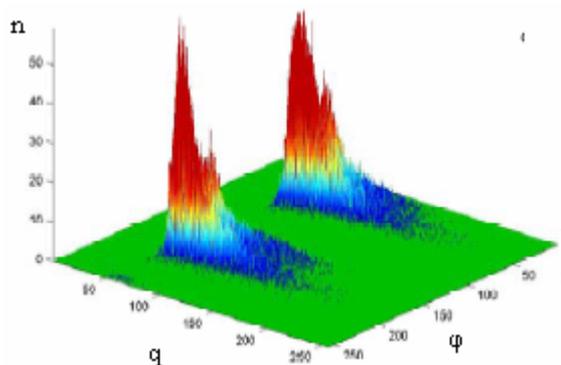


Figura 2.4 – Mapa típico de descarga parcial com informações associadas à fase. Extraído de [9].

### 2.4.2 Dados associados ao tempo

Nos dados associados à fase não é considerada a informação relacionada à forma do pulso de descarga parcial. Ao contrário, nos padrões de dados associados ao tempo ( $q, t$ ), exibem formato de um pulso individual de DP [16]. Novamente, a tensão de teste é tratada como se fosse constante. Estes padrões de dados são atraentes desde que se tenha um relacionamento entre a física do defeito e a forma do sinal. Ainda, dados relacionados ao envelhecimento de materiais isolantes podem ser observados com mais precisão. Na figura 2.5 são mostradas algumas descargas parciais medidas pelo pulso de corrente em função do tempo.

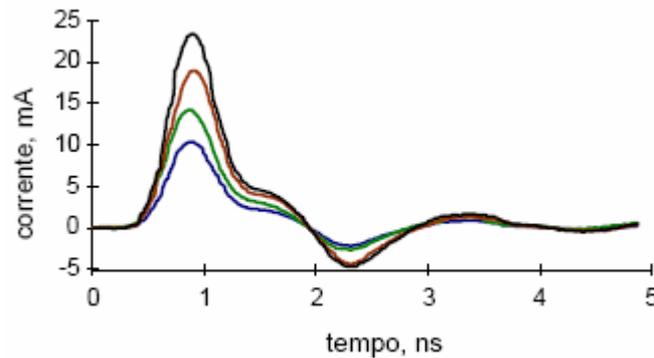


Figura 2.5 – Descargas com informações do tempo – Extraído de [9].

### 2.3.2 Dados não associados nem ao tempo nem à fase

Um exemplo típico de pulso de descarga parcial, sem informação de tempo nem fase, pode ser obtido pela medida da amplitude da DP variando-se a tensão de teste, de maneira a demonstrar o efeito de histerese da DP no ciclo de teste [17]. O efeito observado de histerese é diretamente relacionado ao fato de que a aplicação de pouca tensão é suficiente para sustentar o processo de descarga uma vez que tenha sido iniciada.

## 2.5 Conclusão

As descargas parciais são ocasionadas por um campo elétrico elevado e seguido por um processo de ionização e polarização (provocado por elétrons livres) que

produzem uma avalanche dos elétrons em um meio isolante. Problemas neste isolante podem ser o motivo do surgimento das DPs. Assim, diversos pesquisadores têm realizado estudos para identificar padrões em dados de descargas parciais, relacionando-os com as suas causas. Normas específicas estabelecem parâmetros para que os processos de detecção e medição das DPs sejam realizados. E, dessa forma, pode-se criar uma base de dados para o desenvolvimento de estudos. Diversos métodos de análise têm sido aplicados pelos pesquisadores ao longo dos anos. Estas técnicas e ferramentas serão abordadas no capítulo seguinte.

# Capítulo 3

## Reconhecimento de Padrões em DPs

Ao longo dos anos, diversos pesquisadores têm estudado técnicas e desenvolvido ferramentas de análise para identificar padrões em dados de DPs. Estes dados podem ser sintetizados em três categoriais: dados com informação temporal, dados com informação de fase e dados sem informação de tempo ou de fase, conforme explicado no capítulo anterior. Além da divisão em categorias, o problema de reconhecimento de padrões em DPs pode ser sub dividido em duas etapas: a **extração de características** e a **classificação** dos padrões.

A extração de características é a etapa na qual os parâmetros de um conjunto-padrão são coletados. Assim, forma-se um modelo de DP que comporá uma base de dados. Posteriormente, coletam-se os mesmos parâmetros de um vetor de dados qualquer (vetor-teste) e compara-se com a base-padrão. Esta é a etapa de classificação. Este capítulo apresenta uma abordagem geral sobre o reconhecimento de padrões em DPs. Apresentam-se métodos de extração de características e métodos de classificação aplicados em descargas parciais.

### 3.1 Extração de Características

Quaisquer padrões que possam ser reconhecidos e classificados têm um número de atributos ou características discriminatórias. Assim, o primeiro passo no processo de reconhecimento é buscar a identificação de quais características discriminatórias devem ser selecionadas e como extrair características relevantes. As ferramentas usadas para seleção de vetores característicos (um conjunto peculiar de características) geralmente dependem da aplicação. Nas duas décadas passadas, diferentes técnicas foram desenvolvidas para a escolha das características que vão fornecer a informação relevante para o problema de reconhecimento de padrões em DPs. Essas técnicas podem ser agrupadas em 5 categorias diferentes, a saber: métodos estatísticos, análise do

formato dos pulsos, processamento de sinais, processamento de imagens e aproximações de série temporal [2].

### 3.1.1 Métodos Estatísticos

Os métodos estatísticos têm sido amplamente utilizados na extração de características das DPs, dentre elas, destacam-se: análise de componentes principais (PCA) [5, 6, 18], curvas principais (CP), análises de componentes independentes (ICA) [9], filtros casados [6] e momentos estatísticos[14].

A caracterização por PCA extrai componentes que procuram compactar a informação ordenando as direções de máxima energia (máxima variância). Isso ajuda a revelar a informação que está aparentemente oculta. A característica valiosa da PCA é o seu poder de compactação. Em Cuenca [18] é realizado um processo de extração por PCA linear e, em Diniz [5], utiliza-se PCA não linear. A seção 4.1 traz mais detalhes sobre PCA aplicada às DPs.

As curvas principais podem ser entendidas como uma generalização de análise de componentes principais [4]. A curva é composta por segmentos que são extraídos de regiões definidas no espaço de dados. Estes segmentos correspondem a primeira componente principal daquela região. As CPs foram aplicadas neste trabalho no reconhecimento de padrões de descargas parciais e os respectivos resultados serão discutidos nos capítulos subseqüentes.

A análise de componentes independentes busca mapear os dados de forma a encontrar fontes de sinais independentes. A técnica tem a capacidade de extrair os sinais independentes sem o conhecimento a priori das fontes. Existem diversos algoritmos que implementam ICA [19]. A técnica foi sugerida em Cuenca [9] para a extração de padrões de DPs livres de ruídos.

Os filtros casados são baseados na teoria de sistemas lineares de filtragem ótima [20]. Os filtros casados são bastante utilizados em sistemas de decisão. Nestes sistemas supõe-se que o sinal a se detectar não é aleatório e encontra-se mascarado por um ruído do tipo aditivo. O ruído pode ser branco ou colorido. No caso de ruído colorido é necessária uma etapa de branqueamento. Após o branqueamento, detecta-se o sinal a partir de um banco de filtros. Em Cuenca [6] esta técnica foi utilizada para a classificação de DPs em conjunto com uma rede neural.

Momentos estatísticos são aplicados em distribuições univariadas de DPs , do tipo  $(q_a, \varphi)$ ,  $(q_p, \varphi)$ ,  $(n, \varphi)$  [14]. Esses momentos constituem um vetor característico. O perfil dessas distribuições pode ser obtido segundo uma função geral,  $y_i=f(x_i)$ . Assim, os momentos podem ser computados como seguem:

$$\text{Valor Médio} \quad \mu = \frac{\sum_{i=1}^N x_i f(x_i)}{\sum_{i=1}^N f(x_i)} \quad (3.1)$$

$$\text{Variância} \quad \sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2 f(x_i)}{\sum_{i=1}^N f(x_i)} \quad (3.2)$$

$$\text{Assimetria} \quad S_k = \frac{\sum_{i=1}^N (x_i - \mu)^3 f(x_i)}{\sigma^3 \sum_{i=1}^N f(x_i)} \quad (3.3)$$

$$\text{Curtose} \quad K_u = \frac{\sum_{i=1}^N (x_i - \mu)^4 f(x_i)}{\sigma^4 \sum_{i=1}^N f(x_i)} \quad (3.4)$$

Onde N é o número de janelas de fase no meio-ciclo da tensão aplicada (ciclo positivo ou negativo).

Desde que a atividade dos pulsos de DP ocorra nas duas metades da tensão AC de teste (tensão aplicada), esses momentos estatísticos devem ser computados separadamente para ambas metades. Apenas com a média e o desvio padrão pode se determinar um padrão. No entanto, dada a complexidade dos dados de DPs, a classificação pode ficar prejudicada. Estatísticas de ordem superior como a curtose e a assimetria podem se fazer necessárias. A assimetria é a medida do grau de inclinação dos lados da distribuição normal. Se a distribuição é simétrica  $S_k=0$ ; senão  $S_k \neq 0$ ; A Curtose é um indicador do quanto a distribuição é pontiaguda em relação a uma normal. Se a distribuição tem a mesma distribuição da normal,  $K_u=0$ ; senão  $K_u \neq 0$ . A assimetria de ocorrência das descargas também foi avaliada por outras descrições [16].

$$\text{Assimetria da descarga} \quad Q = \frac{Q_s^- / N^-}{Q_s^+ / N^+} \quad (3.5)$$

$$\text{Assimetria de fase} \quad \varphi = \frac{\varphi_{in}^-}{\varphi_{in}^+} \quad (3.6)$$

onde  $Q_s^+$  e  $Q_s^-$  são a soma das descargas das distribuições de alturas do pulso médio e  $N^+$  e  $N^-$  são o número de descargas da distribuição de alturas do pulso médio - na metade positiva e negativa da tensão aplicada, respectivamente. Assim,  $Q=1$  significa nível de descarga igual, e  $Q=0$  significa uma grande diferença. Analogamente,  $\varphi$  refere-se a ocorrência das descargas.

### 3.1.2 Ferramentas de caracterização do pulso temporal

Os parâmetros do vetor característico de DPs são definidos de maneira a que venham revelar vários atributos significativos de sua forma no domínio do tempo [16]. As definições desses parâmetros são:

- Tempo de subida ( $t_r$ ) do pulso: período correspondente entre 10% e 90% do valor de pico.
- Tempo de decaimento ( $t_d$ ): período correspondente entre 90% e 10% do valor de pico.
- Largura do pulso ( $t_w$ ): intervalo de tempo entre os níveis de 50% de ambos os lados do pico.
- Área sobre o pulso: área correspondente a 10% do nível de surgimento e extinção do pulso.

A figura 3.1 mostra variação da amplitude de pulso em função do tempo e indica os seus parâmetros.

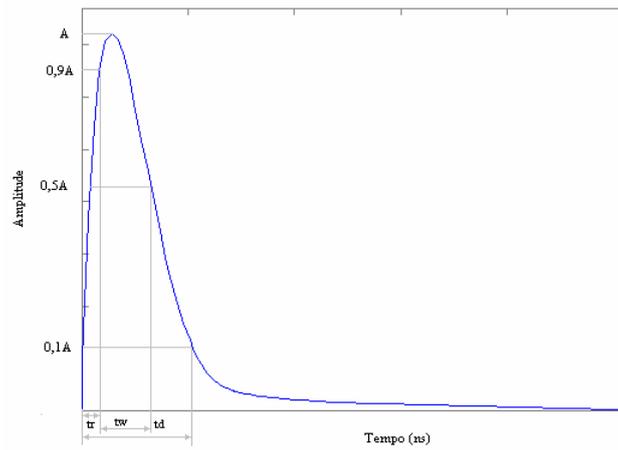


Figura 3.1 – Modelo de pulso de DP

Além destes parâmetros anteriores, outros quantificadores dos pulsos DPs são avaliados em [2]. Esses parâmetros correspondem a valores da carga aparente, e da tensão de teste  $V_i$ .

$$\text{Descarga de corrente média} \quad I = \frac{1}{\Delta t} \sum_{i=1}^N |q_i| \quad (3.7)$$

$$\text{Razão quadrática} \quad D = \frac{1}{\Delta t} \sum_{i=1}^N q_i^2 \quad (3.8)$$

$$\text{Potencia de descarga} \quad P = \frac{1}{\Delta t} \sum_{i=1}^N q_i V_i \quad (3.9)$$

$$\text{Razão de repetição} \quad n = N / \Delta t \quad (3.10)$$

$$\text{Magnitude do pico de descarga} \quad q_{\max} = \max(q_1, \dots, q_n) \quad (3.11)$$

$$\text{Magnitude da descarga média} \quad q_{\text{médio}} = \sum_{i=1}^N |q_i| / N \quad (3.12)$$

onde  $q_i$  é a carga aparente e  $N$  é o número total de descargas dentro de um intervalo de observação  $\Delta t$ , geralmente 1 segundo.

Outra categoria de vetores característicos é constituída pela observação da magnitude da tensão AC de teste para cada pulso de DP nas partes positiva e negativa do ciclo de tensão. O vetor característico consiste de somente dois parâmetros calculados pela diferença de voltagem entre dois pulsos consecutivos [21],  $\Delta V_k (V_{k+1} - V_k)$  e  $\Delta V_{k-1} (V_k - V_{k+1})$ , onde,  $k=1,2,3,\dots,n$  é o índice que identifica o pulso.

### 3.1.3 Ferramentas de Processamento de Sinais

Muitas técnicas de processamento de sinais podem ser aplicadas no reconhecimento de padrões em DPs. Dentre elas, as Transformadas de Fourier [22], *Wavelet* [23], Haar e Walsh [22] são as mais conhecidas e serão discutidas brevemente nesta subseção.

#### 3.1.3.1 Transformada de Fourier

A análise de Fourier em sinais de DPs é baseada no fato de haver uma forte correlação entre a tensão AC de teste e os sinais de descarga parcial. Desde que a tensão seja periódica, os sinais de DP, particularmente o perfil de distribuição de carga  $q(\varphi)$ , deve exibir um comportamento aproximadamente periódico. O vetor característico consiste das componentes espectrais da distribuição de carga normalizada no domínio da frequência [22]:

$$\bar{a}_k = \frac{1}{P_0} \sum_{p=1}^{P_0} \frac{\int_{\varphi=0}^{2\pi} q_p(\varphi) \cos(k\varphi) d\varphi}{\sqrt{\int_{\varphi=0}^{2\pi} q_p^2(\varphi) d\varphi}} \quad (3.13)$$

$$\bar{b}_k = \frac{1}{P_0} \sum_{p=1}^{P_0} \frac{\int_{\varphi=0}^{2\pi} q_p(\varphi) \sin(k\varphi) d\varphi}{\sqrt{\int_{\varphi=0}^{2\pi} q_p^2(\varphi) d\varphi}} \quad (3.14)$$

onde  $q_p(\varphi)$  é a carga aparente na posição de fase  $\varphi$  no  $p$ -ésimo ciclo,  $q_p(3\pi) = q_{p+1}(\pi)$ , e  $P_0$  denota o número de ciclos considerados.

A carga média normalizada  $\bar{Q}(\varphi)$  é, assim, dada por:

$$\bar{Q}(\varphi) = \frac{\bar{a}_0}{2} + \sum_{k=1}^{\infty} \left( \bar{a}_k \cos(k\varphi) + \bar{b}_k \text{sen}(k\varphi) \right) \quad (3.15)$$

onde  $\bar{a}_0$  é o valor médio da serie de Fourier correspondente.

### 3.1.3.2 Transformadas de Haar e Walsh

O processo básico por trás da computação dos coeficientes  $a_k$  e  $b_k$  nas aproximações por série de Fourier é a estimação da correlação dos coeficientes do sinal de DP  $q(\varphi)$  para  $\text{sen } \varphi$  e  $\text{cos } \varphi$ , e seus harmônicos de ordem mais elevada. Um exame mais atento dos pulsos irá revelar que um típico pulso de DP não se encontra significativamente correlacionado com funções senoidais. Preferivelmente, as bordas rápidas e a duração pontual dos pulsos de DP relacionam-se muito melhor com outras funções, assim como Haar e Walsh [22]. Usando essas funções, o vetor característico pode ser construído a partir dos coeficientes correspondentes e suas diferenças absolutas de uma maneira similar àquela realizada para a aproximação da transformada de Fourier.

### 3.1.3.3 Transformada *Wavelet*

O ponto fraco da transformada de Fourier está no fato de que as características extraídas são expressas apenas no domínio da frequência. Para superar esta limitação, transformadas *wavelets* (TW) têm sido aplicadas para realizar análises no domínio do tempo-frequência em vários problemas [23].

Uma transformada *wavelet* contínua (TWC) é geralmente utilizada para obter a informação variante no tempo em detalhes. É a medida de similaridade entre um sinal e um conjunto de funções oscilantes de decaimento rápido, conhecidas como *wavelets*. Cada membro (*wavelet*) deste conjunto é construído pela compressão ou a dilatação de uma função original chamada *wavelet mãe*. A TWC de um sinal  $x(t)$  com relação a *wavelet mãe*  $\varphi(t)$  é definida como:

$$TWC(a,b) = \frac{1}{\sqrt{a}} \int_{\varphi=0}^{2\pi} x(t) \psi\left(\frac{t-b}{a}\right) dt, \quad a > 0 \quad (3.16)$$

onde **a** é a dilatação ou fator de escala e **b** é o fator de dilatação, e ambas as variáveis são contínuas.  $TWC(a,b)$  são os coeficientes *wavelets*.

Em Jang et. al. [24], o vetor característico é construído usando-se o coeficiente de aproximação e o coeficiente de detalhes em uma TWD de seis níveis, baseado na *wavelet* Daubechies'12 (db12). Em Carminati et. al. [25], as *wavelets* Malvar foram aplicadas na classificação do fenômeno de envelhecimento de através da análise da degradação dos pulsos de descarga parcial.

### 3.1.4 Ferramentas de Processamento de Imagens

As ferramentas de processamento de imagens consideram as informações providas pelos parâmetros ( $\varphi$ ,  $q$ ,  $n$ ) como uma imagem. Assim, algoritmos de processamento de imagem são usados para distinguir as características das imagens. Especificamente, três aproximações diferentes têm sido utilizadas para isto: os algoritmos de análise de textura [26], a teoria de fractais [27] e a decomposição de imagem baseada em *wavelets* [28]. Uma breve explanação sobre estas técnicas será dada nesta seção.

#### 3.1.4.1 Algoritmos de Análise de Textura

Os algoritmos de análise de textura avaliam a variação das imagens em escala de cinza. Os algoritmos operam na menor porção da imagem – o *pixel*. Assim, uma imagem é decomposta em *pixels* nas direções horizontal e vertical. Na análise da imagem de DP, a amplitude do pulso é substituída pelos valores da escala de cinza. Os algoritmos utilizados em sinais de DP são: método de dependência do nível de cinza espacial (SGLDM) [29], método do histograma da diferença de nível de cinza (GLDHM) [30], método da largura corrida do nível de cinza (GLRLM) [31] e o método do espectro de potência (PSM) [32]. Todos os quatro algoritmos investigam a amplitude

relativa e a distribuição espacial dos pulsos de DP (desse modo, leva-se em consideração o efeito da propagação de memória [13]) no mesmo ciclo (direção horizontal) e entre os pulsos de DP no mesmo ângulo de fase em diferentes ciclos (direção vertical). Os três primeiros algoritmos operam no domínio espacial, enquanto o último usa o domínio da frequência. As texturas características são obtidas para os ciclos positivos e negativos individualmente.

#### 3.1.4.2 Características Fractais

A natureza complexa das imagens de padrões de DP e a habilidade dos fractais na a modelagem de imagens complexas é o fator que motiva uso de fractais na extração de características [2].

Fractais são gerados por um processo iterativo iniciado por uma construção de blocos, assumindo uma forma em que as partes, quando ampliadas, revelam muito mais detalhes do que o inteiro. Esta propriedade é chamada de auto-similaridade [27]. O objeto fractal pode ser imaginado como a composição de muitas cópias de si mesmo. Fractais são associados a um número (não limitado a um inteiro) denominado dimensão fractal. Em [27], a dimensão fractal é utilizada para constituir o vetor característico. Em um outro trabalho [33], a técnica da compressão de imagem fractal (CIF) foi utilizada para combinar as tarefas de compressão de padrões de DP e com a extração de características. As duas tarefas são, assim, implementadas em um simples passo. A tarefa de compressão consiste em um algoritmo de codificação. Ele é baseado na partição da imagem de DP e a geração subsequente de transformação relevante.

#### 3.1.4.3 Decomposição de Imagens Baseada em *Wavelets*

Transformadas *wavelets* multi-resolução têm sido usadas com sucesso na decomposição de imagens [34]. As seqüências horizontal e vertical são filtradas e sub amostradas separadamente. Quatro sub-imagens são criadas, a saber, um coeficiente de aproximação (A) e três conjuntos de coeficientes detalhados (variações da imagem vertical (V), horizontal (H) e diagonal (D)). Esta aproximação foi utilizada em [28] para classificação de fontes múltiplas de DPs. Neste sentido, os coeficientes *wavelets* mostram as variações vertical, horizontal e diagonal da imagem de DP. Assim, neste

caso, constituem o vetor característico, a saber, cada sub-imagem é um componente do vetor característico.

### **3.1.5 Aproximação por Séries Temporais**

A aproximação por séries temporais foi realizada em [17] para determinar o número de características de entrada a ser usado por uma rede neural para a classificação. Um modelo de série temporal auto-regressivo foi utilizado para modelar o efeito de histerese da característica  $(q, V)$  das DPs. Esta aproximação foi utilizada para gerar leituras  $q_{t-1}, q_{t-2}, q_{t-3}$  e  $V_{t-1}, V_{t-2}, V_{t-3}$ , etc., as quais constituem os componentes do vetor característico (com possíveis inclusões de outros parâmetros como a área aproximada pela curva de histerese, por exemplo).

## **3.2 Classificação**

Existe uma vasta quantidade de classificadores descritos na literatura para o reconhecimento de padrões [35]. As várias aproximações são baseadas em funções de decisão, distância ou verossimilhança, entre outras. Assim, os classificadores propostos para DPs podem ser agrupados como classificadores de distância, classificadores estatísticos, classificadores baseados em redes neurais artificiais e classificadores baseados em lógica nebulosa.

### **3.2.1 Classificadores baseados na distância**

Os algoritmos de classificação baseados na função distância assumem que existe uma biblioteca de referência para cada classe padrão  $\omega_i$  (onde  $i = 1, 2, \dots, M$  e  $M$  é o número de classes). Essa biblioteca de referência pode ser constituída a partir de um conjunto inicial de medidas de vetores característicos pertencentes a classes diferentes. Os principais algoritmos utilizados para a classificação na respectiva classe são: classificadores de distância mínima, classificador do vizinho mais próximo e classificador de pontuação percentual.

### 3.2.1.1 Classificadores de Distância Mínima

O conceito de classificação de padrões usando funções distância é baseado no fato de que a simples média das amostras de um vetor característico, representadas por pontos no espaço euclidiano, determina a sua proximidade aos *clusters* padrões. O classificador determina ao vetor desconhecido a classe que está mais próxima (distância mínima). Valores situados nas fronteiras de dois *clusters* são resolvidos arbitrariamente. Existem várias maneiras de calcular a distância  $d_i$  entre o vetor característico desconhecido ( $\mathbf{x}$ ) e as classes de referência. As classes de referência são representadas por um vetor característico médio  $m_i$  ( $i=1, 2, \dots, M$ ). No contexto das DPs,  $M$  denota o numero de possíveis padrões de descarga parcial. As várias aproximações para computação da distância podem ser derivadas do classificador de distância generalizada:

$$d_i^z(x) = z \sqrt{\frac{1}{N} \sum_{j=1}^N \omega_{ij} |x_j - m_{ij}|^z}, i = 1, 2, \dots, M \quad (3.17)$$

onde  $\omega_{ij}$  é um fator de peso.

Fazendo-se  $z=1$  e  $\omega_{ij}=1$  obtém-se o classificador de Manhattan [36]. A distância euclidiana normalizada é outro caso especial com  $z=2$  e  $\omega_{ij}=1$  (o fator  $1/N$  pode ser visto como um fator de escala) [2]. O fator  $\omega_{ij}$  pode ainda ser substituído por  $1/\sigma_{ij}^2$  onde  $\sigma_{ij}$  representa o desvio padrão do  $j$ -ésimo componente do vetor característico da  $i$ -ésima classe.

### 3.2.1.2 Classificadores de Votação ou Pontuação percentual

O classificador baseado na pontuação percentual também leva em consideração a função distância. A distância euclidiana normalizada  $\|(r_{ik} - m_i)/\sigma_i\|$  é calculada para o  $k$ -ésimo vetor de referencia  $r_{ik}$  pertencente a  $i$ -ésima classe e, a seguir, para o vetor medido  $x$ . O classificador calcula a quantidade (ou percentagem) de referências à  $i$ -ésima classe. A probabilidade é da medida pertencer à classe com o maior pontuação percentual. A classificação é rejeitada se a diferença entre a maior e a menor pontuação é inferior a um limite estabelecido.

### 3.2.2 Classificadores Estatísticos

Esses algoritmos são baseados em alguma consideração estatística para a classificação de padrões. Em descargas parciais, existem dois algoritmos que são bastante utilizados: classificador de Bayes e classificador baseado na taxa de reconhecimento.

#### 3.2.2.1 Classificador de Bayes

O classificador de Bayes [35] baseia-se na função de verossimilhança para minimizar o risco através da menor probabilidade de erro na classificação. Ele usa propriedades estatísticas de dados previamente observados e cuja densidade de probabilidade é resumida pelo valor médio e a pela matriz de covariância. A função normal multivariada pode ser usada como a função densidade de probabilidade [2].

Para M classes de padrões, a função densidade normal multivariada  $p(x/\omega_i)$  do vetor medido  $x$  quando o vetor pertence a classe  $\omega_i$  é dada por:

$$p(x/\omega_i) = \frac{\exp\left[-0,5(x - m_i)' C_i^{-1} (x - m_i)\right]}{(2\pi)^{N/2} |C_i|^{0,5}} \quad i=1,2,\dots,M \quad (3.18)$$

onde a densidade de cada classe é descrita pelo seu vetor médio  $m_i$  e a matriz covariância  $C_i$ . A função decisão para a classe  $\omega_i$  é dado por:

$$d_i(x) = p(x/\omega_i) \cdot p(\omega_i) \quad (3.19)$$

onde  $p(\omega_i)$  é a probabilidade a priori da classe  $\omega_i$ . Assim, um padrão  $x$  é determinado para a classe  $\omega_i$  se  $d_i(x) > d_j(x)$  para  $j \neq i$ , sendo  $i = 1, 2, \dots, M$ .

#### 3.2.2.2 Classificador Baseado na Taxa de Reconhecimento

No diagnóstico de fontes de DPs, o classificador baseado na taxa de reconhecimento tem sido usado com vetores característicos constituídos de operadores estatísticos (OE). O classificador baseado na taxa de reconhecimento funciona da

seguinte forma: para cada sinal conhecido de DP várias amostras são extraídas. A média aritmética e o desvio padrão de seus parâmetros característicos são calculados. A seguir, obtém-se o intervalo de confiança (IC) - geralmente 95% da distribuição normal - sobre esta média, como segue:

$$IC_1 = M_{oe} - \frac{\lambda\sigma}{\sqrt{N}} \quad IC_2 = M_{oe} + \frac{\lambda\sigma}{\sqrt{N}} \quad (3.20)$$

onde  $M_{oe}$  é a média aritmética do operador estatístico obtida da série de N observações de um mesmo tipo de defeito,  $\sigma$  é o desvio padrão da série e  $\lambda$  é um parâmetro de teste estatístico dependente de N.

Para um padrão desconhecido, todos os seus parâmetros característicos são examinados segundo um intervalo de confiança de 95%. Se um parâmetro característico encontra-se dentro do intervalo de confiança, este é computado àquela característica. Este procedimento é repetido para todas as características constituindo um vetor característico. A taxa de reconhecimento pode variar entre zero e o número máximo de características. Esse procedimento de pontuação é realizado para todas as classes de referência. A classe padrão que obtiver a maior quantidade de parâmetros é a vencedora.

### 3.2.3 Classificadores baseados em Redes Neurais Artificiais

Redes neurais artificiais (RNA) têm sido aplicadas com bastante sucesso em problemas de classificação de padrões de DPs. Pode se dizer que uma vantagem das RNA sobre outros classificadores é a sua habilidade de aprender por exemplos. O conhecimento adquirido do conjunto de treinamento é extraído e armazenado nos conectores dos neurônios (“sinapses” e “bias”) durante a fase de aprendizado. Diversos tipos de estruturas de RNA têm sido utilizadas no reconhecimento de padrões de descargas parciais: rede neural *backpropagation*, mapas auto-organizáveis de Kohonen, rede neural *counterpropagation*, rede de quantização do vetor aprendido, redes neurais modulares e cascadeadas.

### 3.2.3.1 Redes Neurais *Backpropagation*

A rede neural *backpropagation* é a mais utilizada em diversos tipos de aplicações. Ela recebe este nome pela forma como manipula a diferença entre o valor desejado na saída e o valor de saída da rede. Esta diferença (erro) é propagada na rede em sentido inverso e, nesse percurso de volta, os coeficientes da rede são atualizados [37]. A rede é composta de uma camada de entrada, pelo menos uma camada escondida e uma camada de saída (veja a figura 3.4). Cada camada é conectada a camada seguinte. O número de nós na camada de entrada é igual ao número de parâmetros selecionados para a classificação. O número de neurônios na camada de saída é igual ao número de fontes de descargas parciais a serem classificadas. A determinação do número de neurônios da camada escondida depende de diversos fatores, incluindo a estatística disponível para o treinamento. Detalhes podem ser encontrados em vários artigos sobre o assunto [9, 38, 39]. No trabalho de reconhecimento de DPs, o vetor característico, pertencente a um conjunto DPs utilizadas para o treinamento da rede, é apresentado na entrada da rede de uma maneira aleatória. Assim, a rede vai armazenando essas informações aprendidas em coeficientes  $w_{ij}$  (onde  $i$  é o neurônio anterior e  $j$  o neurônio posterior). O treinamento é realizado até alcançar-se um índice de desempenho satisfatório. Dessa forma, a rede está apta a ser utilizada para classificar vetores de classes desconhecidas apresentados em sua entrada. A figura 3.2 mostra a topologia de uma rede neural *backpropagation* típica.

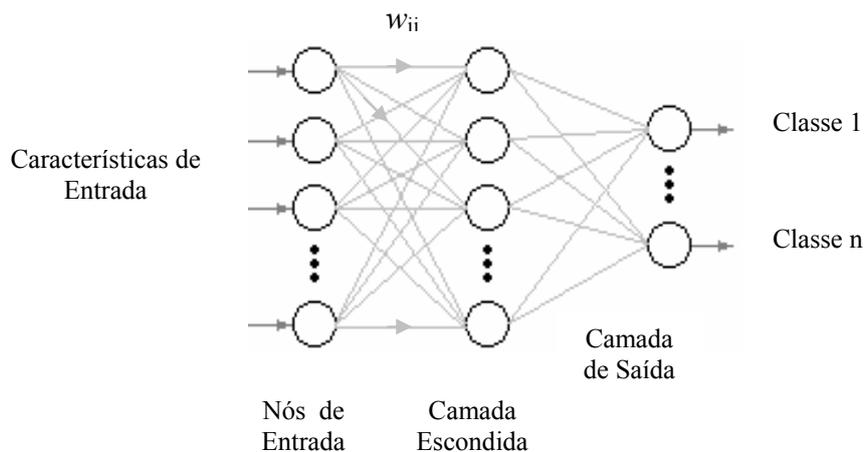


Figura 3.2 – Rede neural *backpropagation*

### 3.2.3.2 Mapas Auto-Organizáveis de Kohonen

Os mapas auto-organizáveis de Kohonen [40] traduzem um espaço de entrada multidimensional em uma rede neural de duas dimensões de uma maneira não linear [41] e preservando a ordem topológica dos dados de entrada simultaneamente. Estes procedimentos são realizados da seguinte forma: dois vetores característicos próximos um do outro são mapeados para neurônios também próximos um do outro, numa camada bidimensional da rede (camada de Kohonen). Uma arquitetura típica deste tipo de rede é mostrada na figura 3.3. Cada neurônio na camada de entrada e saída são completamente conectados a todos os neurônios da camada de Kohonen. Cada neurônio da camada de Kohonen mede a distância euclidiana entre seus pesos e os vetores característicos de entrada. Durante o aprendizado, o neurônio de Kohonen com a menor distância ajusta seus pesos próximo ao vetor de entrada. Os neurônios vizinhos também ajustam seus pesos próximos do mesmo vetor de entrada. Este aprendizado é realizado de forma não supervisionada. Depois de completar a fase de aprendizado, os pesos da camada de saída são aprendidos de um modo supervisionado. Os detalhes sobre o algoritmo de aprendizagem podem ser encontrados em Kohonen [42].

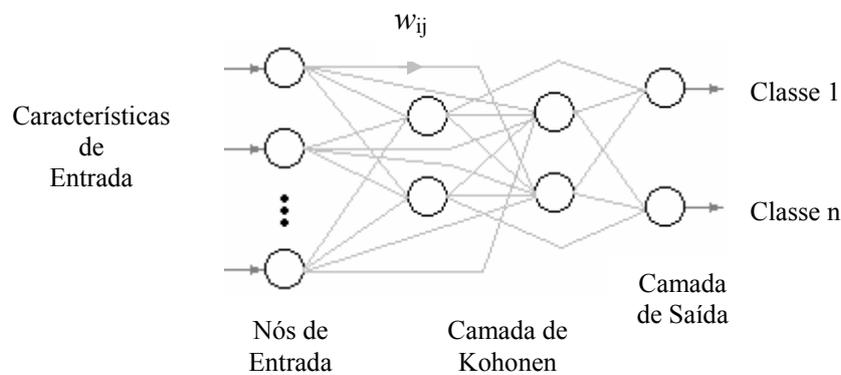


Figura 3.3 – Mapas auto-organizáveis de Kohonen

### 3.2.3.3 Rede de Quantização do Vetor Aprendizagem

A rede de quantização do vetor aprendizagem (LQV) foi utilizada na classificação de DPs em Mazroua et. al. [16] e Gulski e Krivda [40]. A estrutura básica é mostrada na figura 3.4. Ela consiste de uma camada de entrada, uma camada de Kohonen com o mesmo número de neurônios das classes disponíveis e uma camada de

saída. Durante a fase de aprendizado, a distância euclidiana entre o vetor de aprendizagem e os pesos de cada neurônio da camada de Kohonen é computado. O neurônio mais próximo é declarado o vencedor. Se o neurônio vencedor pertence à classe do vetor aprendizagem, então sua distância a classe particular é diminuída pelo movimento dos seus pesos em direção ao vetor aprendizagem, desse modo, aumentando sua probabilidade de vencer da próxima vez. Se o neurônio vencedor não pertence à classe do vetor aprendizagem, então sua distância à classe particular é aumentada pelo movimento de seus pesos para longe do vetor de entrada, desse modo, diminuindo sua probabilidade de vencer da próxima vez. Na etapa de classificação, a distância entre o vetor desconhecido e cada neurônio é computada. Aquele mais próximo é declarado o vencedor.

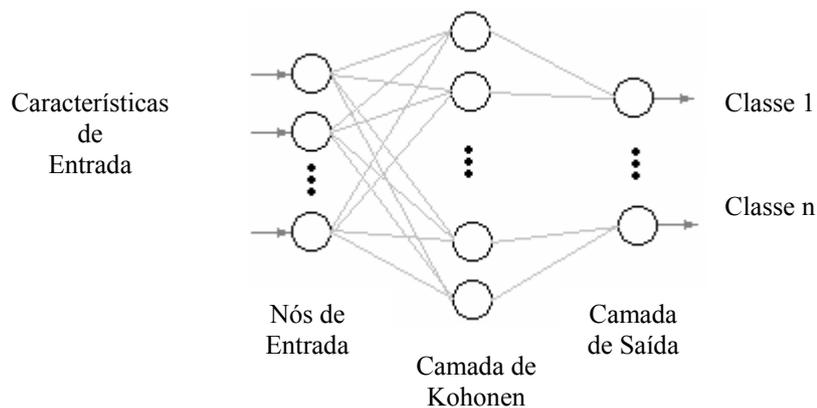


Figura 3.4 – Rede de quantização do vetor aprendizagem

### 3.2.3.4 Redes Neurais *Counterpropagation*

Redes *Counterpropagation* (RCP) são mais usadas para aproximar uma função do tipo  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ . Assim, mapeia um espaço de entrada n-dimensional em um espaço m-dimensional. Um esquema é mostrado na figura 3.5. Todas as três camadas da rede são completamente interconectadas. As funções comparam um determinado padrão com os protótipos armazenados e codificados nos pesos entre as camadas de entrada e escondida. Os pesos de maior similaridade com a entrada são selecionados. O aprendizado é realizado de forma não supervisionada, parecida com as redes de *Kohonen*. Já os pesos entre as camadas escondida e de saída armazenam os resultados

de saída. Estes são aprendidos de uma forma supervisionada. Os detalhes podem ser encontrados em Hecht-Nielsen [41].

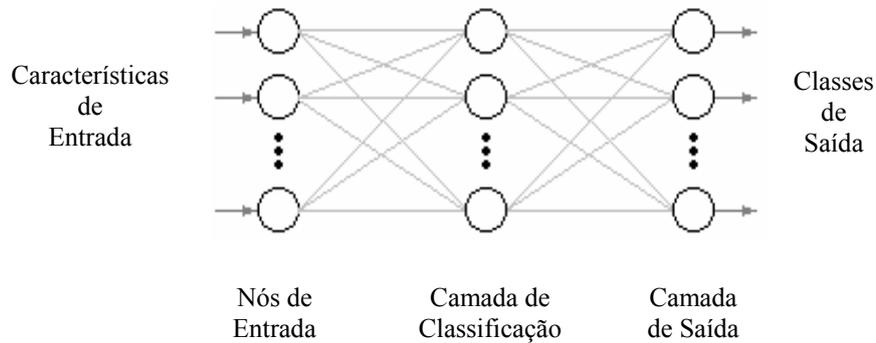


Figura 3.5 – Rede neural counterpropagation

### 3.2.3.5 Redes Neurais Modulares

O conceito por trás da construção de uma rede neural modular está na decomposição de um problema multitarefa em suas sub-tarefas constituintes. A maior vantagem está no fato de que o treinamento das sub-redes para fazerem tarefas individuais é mais simples, quando comparado com o requerido por uma rede única a realizar uma tarefa complexa. Cada sub-rede é responsável pela identificação de uma fonte de descarga parcial e a estrutura utilizada pode ser uma rede *backpropagation*. Este tipo de rede foi realizado por Hong et. al. [43].

### 3.2.3.6 Redes Neurais Cascadeadas

Redes neurais *backpropagation* simples também podem ser cascadeadas no intuito de chegar-se a uma melhor classificação. Além de cada estágio ter entradas próprias, os estágios intermediários têm entradas correspondentes às saídas do estágio anterior. A rede opera utilizando uma entrada correspondente a saída do estágio inicial. Esta é a entrada mais importante do estágio seguinte. Em Salama e Bartinikas [17] foi utilizada esta rede com dois estágios, conforme o esquema mostrado na figura 3.9.

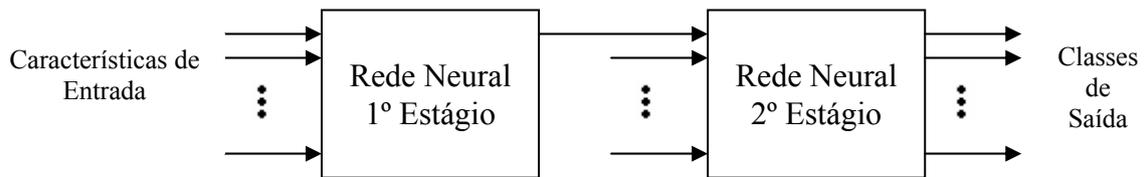


Figura 3.6 – Redes neurais cascadeadas. Extraída de [17].

### 3.3.4 Classificadores baseados em lógica nebulosa

A classificação de descargas parciais usando lógica nebulosa é baseada na observação básica de que os vários parâmetros característicos são valores não confusos, preferencialmente eles devem ser tratados como variáveis nebulosas [44]. Assim os parâmetros característicos são mapeados nos valores associados a diferentes conjuntos nebulosos definidos sobre seus respectivos domínios.

Em uma aproximação para classificar DPs, a lógica nebulosa é usada para uma partição nebulosa do espaço característico em diferentes classes padrão [44]. Assim, um dado ponto padrão no espaço característico pertence a todos os cluster/classes padrão, mas com graus de associação diferentes. O desenho desse classificador é uma extensão dos classificadores convencionais de máxima verossimilhança com algum grau de nebulosidade incorporada.

## 3.3 Análise do Desempenho

Nesta seção, será discutido o desempenho dos extratores e classificadores em descargas parciais. Tipicamente, a análise é realizada considerando-se a etapa de extração e classificação simultaneamente.

### 3.3.1 Classificação usando dados com informação de fase

Em Kreuger et. al. [45], a extração das características das DPs foi realizada usando-se operadores estatísticos e a etapa de classificação baseou-se na taxa de reconhecimento e na pontuação percentual. Fontes de dados de DPs de diversas naturezas foram consideradas neste trabalho, a saber: descargas no ar, no SF<sub>6</sub>, no óleo

do transformador, nas cavidades do dielétrico, em cabos, etc. Os resultados são indicadores de um bom potencial de classificação (superior a 90%). Foi mostrado experimentalmente que os dois esquemas de classificação tiveram uma capacidade de distinção parecida. Ainda utilizando a extração com operadores estatísticos, em Kranz [46], utilizou-se o a distância mínima e também redes neurais artificiais na etapa classificação. Neste trabalho, identificaram-se as faltas em pontas de alta tensão e pontas aterradas, partículas livres, *gaps* internos, partículas fixas no dielétrico e ruído. O classificador obteve um bom nível de acertos na classificação (superior a 90%). Existem outros trabalhos [40, 47] que usam operadores estatísticos como parâmetros característicos. Em Gulski e Dreuger[47], diferentes defeitos artificiais são testados: cavidade no dielétrico, múltiplas cavidades, etc. No entanto, o uso de muitos operadores estatísticos demandou uma grande quantidade de memória e um grande tempo computacional para criação e operação do banco de dados. Em Gulski e Krivda [40], um pequeno estudo comparativo da eficiência de classificação com vários classificadores de RNA, entre eles, *backpropagation*, mapas auto-organizáveis de Kohonen e rede LQV, foi realizado para a classificação de DPs em cavidades do isolamento, descargas superficiais no ar e efeito corona. Em Hong et. al.[43] um classificador baseado em uma rede neural modular mostrou resultados promissores em relação a uma simples estrutura de rede neural . A taxa de reconhecimento, para diferentes defeitos, foi de 94% contra 88% com uma rede neural simples. A rede modular exibiu uma maior velocidade de convergência, alta precisão e melhor taxa de reconhecimento. Classificadores de rede *couterpropagation* foram utilizados em Hoof et. al. [21]. Neste trabalho, a diferença da tensão externa entre dois pulsos consecutivos é tratada como parâmetro característico adicional. Existe uma considerável melhora no desempenho (uma taxa de acerto próximo de 99%), comparando-se com outros algoritmos de RNA para a classificação de tipos diferentes de defeitos em DPs. O classificador estatístico de *Bayes* obteve bons resultados com vetores característicos estatísticos [48]. Neste trabalho, o classificador é usado em conjunto com um algoritmo genético (AG) para a estimação da matriz covariância, especialmente para uma classe de defeitos com poucas medidas no banco de dados. A taxa média de acerto na classificação é bem elevada (maior que 96%). Em Cuenca [6, 9] utilizou-se PCA, filtros casados e ICA. Utilizou-se três classes distintas de DPs em óleo isolante, relacionadas a transformadores de alta tensão . O índice geral de acertos obtido com PCA, conjuntamente com a rede neural, chegou a 93%. Os filtros casados utilizaram também uma rede neural. Utilizaram-se três filtros casados, relativos

aos três padrões analisados, funcionando como uma etapa de pré-processamento. A seguir redes *backpropagation* realizaram o processo de classificação. O classificador atingiu um desempenho geral de 97%. ICA foi utilizada [9] para recuperar sinais ou fontes não observadas, explorando hipóteses como independência de sinais. Com isso, ICA foi útil para retirar o ruído de natureza gaussiana dos sinais de DPs.

Em Jang et. al. [24] e Kreuger et. al. [45], os vetores característicos foram obtidos usando técnicas de processamento de sinais, a saber, transformadas de Fourier, transformada Haar, transformada Walsh. Um estudo comparativo de classificação de características é feito com varias técnicas [22], a saber: classificador de mínima distância, classificador neural, classificador *fuzzy* etc. Todos os classificadores obtiveram bons resultados, mas o classificador de distância mínima atingiu os melhores índices de desempenho. Classificadores *Fuzzy* foram utilizados com sucesso em Cotin et. al. [44]. Este classificador resulta em um cluster de sinais homogêneos em termos de características estocásticas dos pulsos de DP.

Ferramentas de processamento de imagens, a saber, algoritmos de análise de textura [49], fractais [27], e *wavelets* [28], foram aplicadas com sucesso para a extração de vetores característicos de mapas de 3 dimensões. Em Rahman et. al. [49], além da aplicação de vários algoritmos de análise de textura, a transformação de componentes principais foi também utilizada como uma técnica de redução de características - reduzindo o tempo de identificação. O algoritmo de classificação usou o conceito da distância mínima. A precisão da classificação para uma variedade de fontes de DPs foi bem satisfatória. Os fractais foram aplicados com sucesso em classificadores convencionais para diagnóstico de DPs em Satish e Zaengl [27]. O ponto atrativo da aproximação usando fractais foi o eficiente uso da memória do computador e do tempo, dado que o vetor característico ficou composto apenas de dois componentes. Um estudo de *wavelets* em 3D foi realizado em Lalitha e Satish [28] para classificação de múltiplas fontes de DP por inspeção visual de varias sub-imagens das imagens originais. A inovação do algoritmo é que a classificação pode ainda ser conseguida com razoável precisão mesmo se a fonte individual de DP produza padrões sobrepostos.

### **3.3.2 Classificação usando dados com informação temporal**

Em James e Phung [14], os parâmetros característicos determinísticos dos dados usados com uma RNA obtiveram uma classificação acima de 90% com as varias classes

de DPs estudadas. Desempenhos similares foram mencionadas em vários outros trabalhos usando parâmetros determinísticos [21, 36, 46] com classificadores baseados em estruturas de redes neurais, lógica nebulosa e classificadores de distância (distância mínima e vizinho mais próximo). Os resultados são encorajadores para diferentes tipos de fontes de DPs. Em Mazroua et. al. [36], estruturas diferentes de RNA foram utilizadas no reconhecimento de cavidades artificiais de diversas profundidades (entre 1 e 2 mm) e realizaram as classificações corretamente. No entanto, observa-se também que, quando é necessária uma discriminação entre cavidades menores, uma rede LQV mostra resultados superiores. A transformada *wavelet* foi também utilizada para a extração de vetores característicos, consistindo de diferentes coeficientes *wavelets*, no trabalho reportado em Jang et. al. [24].

### **3.3.1 Classificação usando dados sem informação de fase/tempo**

Esse esquema foi adotado em Salama e Bartinikas [17] usando aproximações de series temporais para extração de características. Um modelo auto-regressivo (AR) foi usado para modelar o efeito de histerese das características ( $q$ ,  $V$ ) das DPs. Esta aproximação foi utilizada para obter as leituras de  $q_{t-1}$ ,  $q_{t-2}$ ,  $q_{t-3}$  e  $V_{t-1}$ ,  $V_{t-2}$ ,  $V_{t-3}$ , etc., constituintes dos componentes do vetor característico. Redes neurais de estágio único, dois estágios, cascadeada, etc foram utilizadas. Os estudos comparativos indicam que rede cascadeada apresentou o melhor desempenho de classificação.

## **3.4 Conclusão**

Uma explanação sobre a extração e classificação de características de sinais de DPs foi apresentada neste capítulo. Os três diferentes tipos de dados (fase, tempo, e sem fase/tempo) de DP que são utilizados em trabalhos práticos foram brevemente discutidos. Vários algoritmos e ferramentas propostos, tanto para extração quanto para a classificação, foram analisados. Sem a pretensão de exaurir o assunto, este capítulo procurou apresentar o estado da arte em reconhecimento de padrões de descargas parciais.

# Capítulo 4

## Curvas Principais

As curvas principais foram originalmente definidas em Hastie e Stuetzle [4] como curvas unidimensionais que atravessam o “meio” de um conjunto de dados em um espaço multi-dimensional. Após esta conceituação, surgiram definições de outros autores que procuravam abordar outros aspectos não abordados no trabalho original [50-53]. Em Delicado [54], as curvas principais foram definidas com base em pontos orientados na direção da primeira componente principal dos dados.

As curvas principais são utilizadas na classificação de padrões e têm sido aplicadas em diversos tipos de dados com bastante sucesso. Uma característica das CPs é a capacidade de extrair modelos compactos dos dados. Esta característica é de grande interesse na análise de mapas de DPs, uma vez que, nestes dados, as informações estão dispersas em dimensões elevadas e, devido às restrições estatísticas, os graus de liberdade são reduzidos [3].

Os métodos de classificação apresentados no capítulo anterior podem ser utilizados conjuntamente com as curvas principais. Além da distância euclidiana, outros métodos, como redes neurais, têm sido utilizados de forma híbrida com o extrator de curvas principais [7].

Este capítulo define as curvas principais e apresenta alguns algoritmos utilizados na extração. De forma mais completa, é apresentado o algoritmo *K-segmentos não-suave* proposto por Verbeek et. al. [8] que se baseia na definição de Delicado [54]. Serão mostrados, ainda, alguns resultados obtidos com alguns classificadores de curvas principais.

### 4.1 Componentes Principais

As componentes principais (PCA – *principal components analysis*) representam as direções que maximizam a variância de um conjunto de amostras [55]. A PCA tem

vido extensamente utilizada na compactação de sinais. Esta compactação é obtida utilizando somente alguns eixos do espaço amostral para se definir as amostras – aqueles que representem mais informações. Os outros eixos são desprezados, o que, obviamente, acarreta em uma perda de informação. No entanto a perda é minimizada ao se extrair somente as componentes menos relevantes.

As componentes principais podem ser obtidas através de uma transformação ortonormal do conjunto de dados. A transformação ortonormal é uma transformação linear que permite derivar características estatisticamente decorrelacionadas de um conjunto de características correlacionadas. Para entender como esta transformação é realizada, considere um *cluster* de amostras correlacionadas ( $\mathbf{y}$ ) e assumamos que ele possa ser obtido a partir da transformação linear  $\mathbf{y}=\mathbf{A}\mathbf{x}$ . Esta transformação produz o conjunto  $\mathbf{y}$  (com características correlacionadas) a partir do conjunto de características decorrelacionadas  $\mathbf{x}$  (cluster circular caracterizado por uma matriz covariância identidade). A figura 4.1 mostra como ocorrem estes procedimentos.

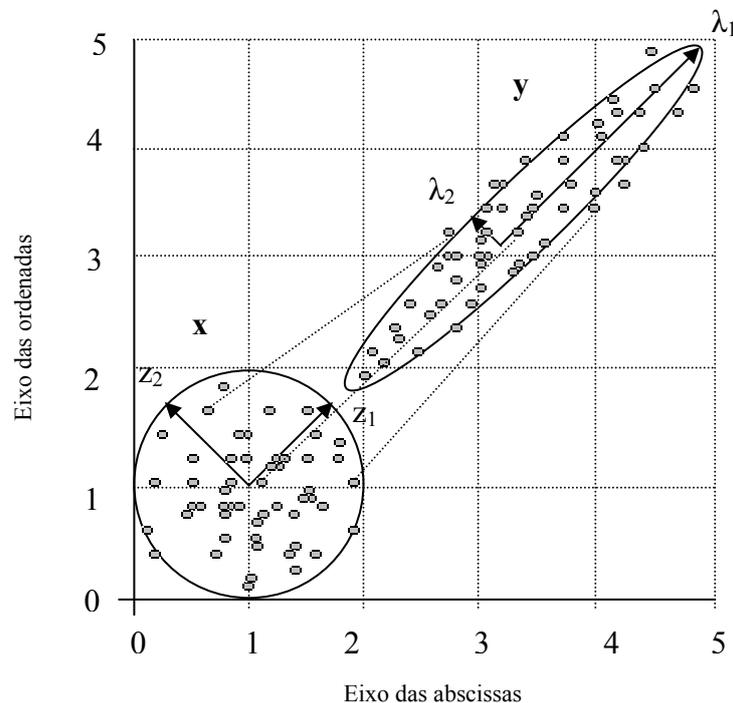


Figura 4.1 – Autovetores de uma transformação linear, mantendo a mesma direção antes e depois da transformação. Os desvios padrão dos autovetores são dados pelos autovalores  $\lambda_1$  e  $\lambda_2$ . Extraído de [8].

Suponha agora que se queira a direção desta transformação, a partir de vetores  $\mathbf{z}$  ortogonais. Assim, tem-se:

$$\mathbf{Az} = \lambda \mathbf{z} \quad (4.1)$$

onde  $\lambda$  é um escalar.

A determinação dos escalares  $\lambda$  e do vetor  $\mathbf{z}$  corresponde à resolução da equação

$$(\mathbf{I}\lambda - \mathbf{A})\mathbf{z} = 0 \quad (4.2)$$

onde,  $\mathbf{I}$  é a matriz identidade  $d \times d$  e  $d$  corresponde à dimensão dos dados (em nosso exemplo  $d=2$ )

Assim, resolvendo o determinante  $|\lambda \mathbf{I} - \mathbf{A}| = 0$ , obtêm-se as soluções não triviais. Existem  $d$  soluções escalares para  $\lambda$  - conhecidas como auto valores de  $\mathbf{A}$ . Com estes autovalores, obtém-se uma família de auto vetores  $\mathbf{z}$ .

Observe que os autovalores e os auto vetores são relacionados à matriz original de transformação linear ( $\mathbf{A}$ ). Ocorre que, em situações práticas, geralmente, não se conhece esta matriz. No lugar, devem-se obter os autovalores e os autovetores da matriz covariância dos dados ( $\mathbf{C}$ ), o que é uma aproximação válida [55]. A forma compacta de representar a matriz  $\mathbf{C}$  é dada na equação 4.1.

$$\mathbf{C} = \frac{1}{n-1} \sum_{k=1}^n (\mathbf{y}_k - \mathbf{m})(\mathbf{y}_k - \mathbf{m})^T \quad (4.3)$$

Onde  $\mathbf{m}$  é o vetor com as coordenadas da média do  $k$ -ésimo ponto de  $\mathbf{y}$ .

Assim, a matriz covariância  $\mathbf{C}$  fica:

$$\mathbf{C} = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1d} \\ c_{21} & c_{22} & \dots & c_{2d} \\ \dots & \dots & \dots & \dots \\ c_{d1} & c_{d2} & \dots & c_{dd} \end{bmatrix}$$

Concluindo, as componentes principais são representadas pelos autovetores (correspondentes às direções ortogonais) associados aos autovalores (representam o módulo de cada direção) da matriz covariância de um conjunto de dados. O método

permite a reconstrução dos dados em eixos cujas direções correspondam a uma ordenação maximizada da energia dos dados em ordem decrescente. Dessa forma, é possível reconstruir os dados apenas com as coordenadas das componentes mais significantes, de modo a obter representações compactadas, sendo a primeira componente aquela que tem maior quantidade de energia. Estas características são fundamentais para o entendimento da definição de curvas principais, que será abordada no item seguinte.

## **4.2 Definição de Curvas Principais**

As curvas principais (CPs) foram apresentadas pela primeira vez por Hastie e Stuetz [4] como uma ferramenta capaz de fornecer uma descrição não-linear de um conjunto de dados. Após esta definição, surgiram outros trabalhos que estendiam e aprimoravam este conceito clássico. Em Fernandez [7] é apresentado um apanhado destes trabalhos. Em Delicado [54], as curvas principais foram definidas com base no conceito de componentes principais. Fundamentados nesta definição, diversos algoritmos foram propostos [56, 57]. Entre eles encontra-se o algoritmo K-segmentos não-suave sugerido em Verbeek et. al. [8]. O algoritmo destaca-se por sua robustez na estimação das curvas, menor susceptibilidade a mínimos locais e convergência garantida.

### **4.2.1 Definição Clássica**

As curvas principais são curvas unidimensionais que passam no “meio” de um conjunto de dados no espaço multi-dimensional, fornecendo uma descrição não linear dos dados; são curvas não paramétricas e a sua forma é sugerida pelo conjunto de dados [4]. A curva é construída por um algoritmo que depende do ponto de partida. Os autores sugeriram a primeira componente principal do conjunto de dados como um ponto de partida. Em seguida, a curva é alterada sucessivamente até que o algoritmo convirja. Assim, a curva principal é extraída.

Matematicamente, as curvas são definidas a partir do conceito de autoconsistência. Para entender este conceito é necessário definir o índice de projeção de um ponto  $x_i$  numa curva  $\mathbf{f}$ .

Uma curva unidimensional no espaço d-dimensional é um vetor  $\mathbf{f}(\lambda)$  de  $d$  funções contínuas e uma única variável  $\lambda$ , ou seja:

$$\mathbf{f}(\lambda) = \begin{bmatrix} f_1(\lambda) \\ f_2(\lambda) \\ \dots \\ f_d(\lambda) \end{bmatrix} \quad (4.4)$$

Estas funções são denominadas funções de coordenadas e o parâmetro  $\lambda$  é relacionado ao ordenamento ao longo da curva.

Seja  $\mathbf{X}$  um vetor aleatório em  $\mathfrak{R}^d$ , com densidade de probabilidade  $h$ , e momentos de segunda ordem finitos. Seja  $\mathbf{f}$  uma curva suave no intervalo fechado  $I \subseteq \mathfrak{R}^1$  que não intercepta a si própria, ou seja,  $\lambda_1 \neq \lambda_2 \rightarrow \mathbf{f}(\lambda_1) \neq \mathbf{f}(\lambda_2)$ . E, ainda, o seu comprimento é finito dentro de uma esfera de dimensões finitas em  $\mathfrak{R}^d$ .

O índice de projeção  $\lambda_f: \mathfrak{R}^d \rightarrow \mathfrak{R}^1$  é definido como:

$$\lambda_f(\mathbf{x}) = \sup_{\lambda} \{ \lambda : \|\mathbf{x} - \mathbf{f}(\lambda)\| = \inf_{\mu} \|\mathbf{x} - \mathbf{f}(\mu)\| \} \quad (4.5)$$

onde  $\mathbf{x}$  é o evento arbitrário pertencente a  $\mathbf{X}$  e  $\mu$  é uma variável auxiliar definida em  $\mathfrak{R}^1$ . O índice de projeção  $\lambda_f(\mathbf{x})$  é o valor de  $\lambda$  para o qual a curva principal  $\mathbf{f}(t)$  está mais próxima de  $\mathbf{x}$ . Se houver mais de um valor possível, o maior deles é selecionado. Na figura 4.2 é mostrado o índice de projeção relativo a uma curva principal.

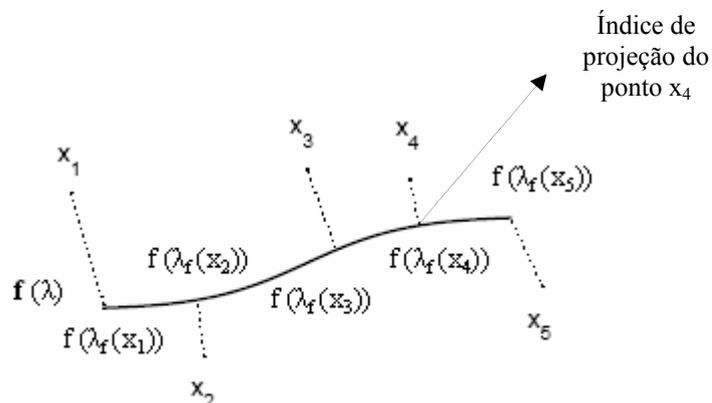


Figura 4.2 – Projeção dos dados na curva principal

Assim, de acordo com a definição de Hastie e Stuetzle [4], a curva principal é autoconsistente. Ou seja, os pontos que compõem a curva principal constituem a média dos dados que nela projetam, conforme equação:

$$f(\lambda) = E \left[ \mathbf{X} / \lambda_f(\mathbf{X}) = \lambda \right] \quad (4.6)$$

A figura 4.3 mostra um ponto auto-consistente ( $\lambda_i$ ) de uma curva para uma distribuição de dados em duas dimensões, representada parcialmente.

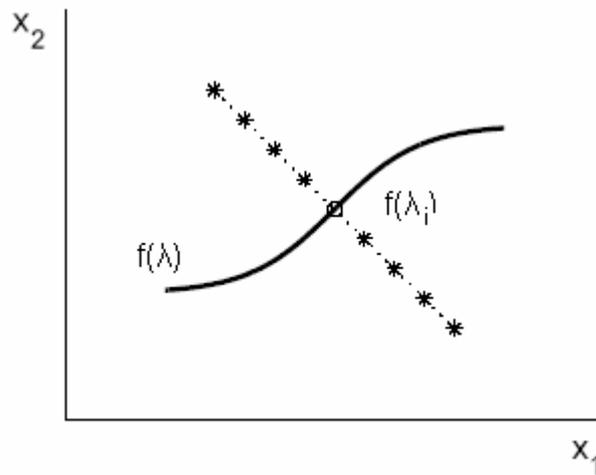


Figura 4.3 – Ponto autoconsistente de uma curva obtida de uma distribuição de dados

#### 4.2.2 Definição de Pontos Orientados

No trabalho de Delicado e Huerta [57] propôs-se uma definição de curvas principais a partir de pontos orientados autoconsistentes. Sabe-se que a projeção de uma variável aleatória normal no hiperplano ortogonal à primeira componente principal tem a menor variância total dentre todas as projeções da variável em qualquer outro hiperplano [55]. Seja  $\mathbf{X}$  uma variável aleatória de distribuição  $h$  em  $\mathfrak{R}^d$ , tal que  $\mathbf{X}=(X_1, X_2, \dots, X_d)$  e seja  $\mathbf{x}$  um evento arbitrário gerado conforme essa mesma distribuição. Considere  $\mathbf{b}$  um vetor unitário em  $\mathfrak{R}^d$  e  $\mathbf{H}(\mathbf{x}, \mathbf{b})$  o hiperplano ortogonal a  $\mathbf{b}$  passando por  $\mathbf{x}$ , ou seja:

$$\mathbf{H}(\mathbf{x}, \mathbf{b}) = \{ \mathbf{y} \in \mathfrak{R}^d : (\mathbf{y} - \mathbf{x})^T \mathbf{b} = 0 \} \quad (4.7)$$

Considere, também,  $\mu(\mathbf{x}, \mathbf{b})$  a média condicional de  $\mathbf{X}$  no hiperplano  $\mathbf{H}$ , ou seja:

$$\mu(\mathbf{x}, \mathbf{b}) = E[\mathbf{X} / \mathbf{X} \in H(\mathbf{x}, \mathbf{b})] \quad (4.8)$$

Define-se  $\mathbf{b}^* : \mathfrak{R}^d \rightarrow \mathfrak{R}^1$  como:

$$\mathbf{b}^*(\mathbf{x}) = \arg \min \varphi(\mathbf{x}, \mathbf{b}) \quad (4.9)$$

onde cada elemento de  $\mathbf{b}^*$  é uma direção principal de  $\mathbf{x}$  e  $\varphi(\mathbf{x}, \mathbf{b})$  é o ângulo entre  $\mathbf{x}$  e  $\mathbf{b}$ .

E também define:

$$\mu^* = \mu(\mathbf{x}, \mathbf{b}^*(\mathbf{x})) \quad (4.10)$$

O conjunto  $\Gamma(\mathbf{X})$  de pontos principais orientados de  $\mathbf{X}$  é um conjunto de pontos fixos de  $\mu^*$ :  $\Gamma(\mathbf{X}) = \{\mathbf{x} \in \mathfrak{R}^d : \mathbf{x} \in \mu^*(\mathbf{x})\}$ . Uma curva  $f: \mathfrak{R}^d \rightarrow I = [a, b]$  é uma curva principal de pontos orientados de  $\mathbf{X}$  se, para todo  $\lambda \in [a, b]$ , os pontos da curva  $f(\lambda)$  são pontos principais orientados. Esta definição pode ser interpretada como uma generalização não linear de componentes principais, visto que uma curva principal de pontos orientados maximiza a variância total dos dados nela projetados, assim como a primeira componente principal é a direção que maximiza a variância da projeção dos dados linearmente (seção 4.1).

### 4.3 Algoritmos de Extração de Curvas Principais

No mesmo trabalho no qual foram inicialmente definidas as curvas principais [4], também se apresentou um algoritmo de extração. O algoritmo apresenta alguns problemas potenciais apontados pelos próprios autores, como: tendências em trechos de curvaturas, convergência do algoritmo e a existência de uma curva principal não ocorria para qualquer distribuição. Estes fatores deram margem para novos trabalhos. Em Duchamp e Stuetzle [50], as curvas foram obtidas através de um sistema de equações diferenciais. Em Branfield e Raftery [58] foi proposta uma correção para a tendência de estimação. Em Tibishrani [60] foi proposto um modelo paramétrico e Eibek [56] propôs um algoritmo baseado na definição de Delicado e Huerta [57]. Diversos outros algoritmos são discutidos e analisados em Fernandez [7]. Um algoritmo importante foi

proposto no trabalho de Verbeek et. al. [8], o algoritmo de K-segmentos não-suave que será analisado a seguir.

#### 4.3.1 O algoritmo K-segmentos Não-Suave

Este algoritmo fundamenta-se na definição de Delicado e Huerta [57]. Este algoritmo é bastante robusto, tem convergência garantida e é menos susceptível a mínimos locais [8].

O algoritmo K-segmentos não-suave constrói uma curva adicionando segmentos através da adição de um segmento a cada passo de iteração, a partir de um segmento inicial. Este algoritmo se desenvolve da seguinte forma:

Passo 0:

- Consideram-se todos os eventos do conjunto de dados de desenvolvimento como um único agrupamento.
- Define-se o centro do agrupamento pelo valor médio dos dados.
- O primeiro segmento corresponde à direção da primeira componente principal do conjunto de dados e o comprimento equivalente a  $3/2\sigma$  associado a esta componente.
- Assim, considerando-se o centro do agrupamento igual à média dos dados, traça-se o primeiro segmento a partir deste ponto.

Passo 1:

- Na inserção do segundo segmento é realizado um teste envolvendo todas as amostras do conjunto de dados. Verificam-se quais amostras estão mais próximos entre si do que do primeiro segmento. Aquelas mais próximas formam novos agrupamentos. O segundo segmento será inserido no maior agrupamento formado. Os procedimentos de inserção deste são os mesmos do primeiro segmento. Todo este processo é realizado pelo algoritmo K-means [55].
- Durante a inserção do segundo segmento algumas amostras do primeiro agrupamento passam a fazer parte do novo agrupamento. Dessa forma, uma vez que a região inicial alterou-se, realiza-se um novo cálculo para determinar o primeiro segmento. De posse destes dois segmentos, une-se um

com o outro através de uma linha reta (daí a qualificação *não-suave* do algoritmo).

- Na inserção de outros segmentos, os mesmos procedimentos anteriores são realizados.

Passo 2:

- É verificado se o algoritmo atingiu o número máximo de segmentos definido pelo usuário ou se houve a convergência. O algoritmo converge quando o maior agrupamento possível tem menos de três amostras.
- Caso não ocorra nenhuma das duas hipóteses anteriores, o algoritmo volta ao passo 1.

Ao final da extração o algoritmo sugere uma determinada quantidade de segmentos para a curva principal. Esta sugestão baseia-se na variação do comprimento total da curva após a inserção de um segmento, da seguinte forma. O comprimento total é obtido através da medição do comprimento de todos os segmentos inseridos mais o comprimento da ligação entre eles. Observando-se que o algoritmo seleciona as regiões de Voronoi em ordem decrescente, em regra, a cada iteração o segmento inserido é menor que o anterior. Assim, haverá um momento em que o segmento inserido não representa uma grande modificação na curva como um todo. Dessa forma, o método de extração sugere uma quantidade de segmentos que resulte em uma curva não super treinada. O fluxo do algoritmo é mostrado na figura 4.4. A ordem de complexidade computacional do método é proporcional à multiplicação do número de amostras pela quantidade de segmentos, elevado ao quadrado.

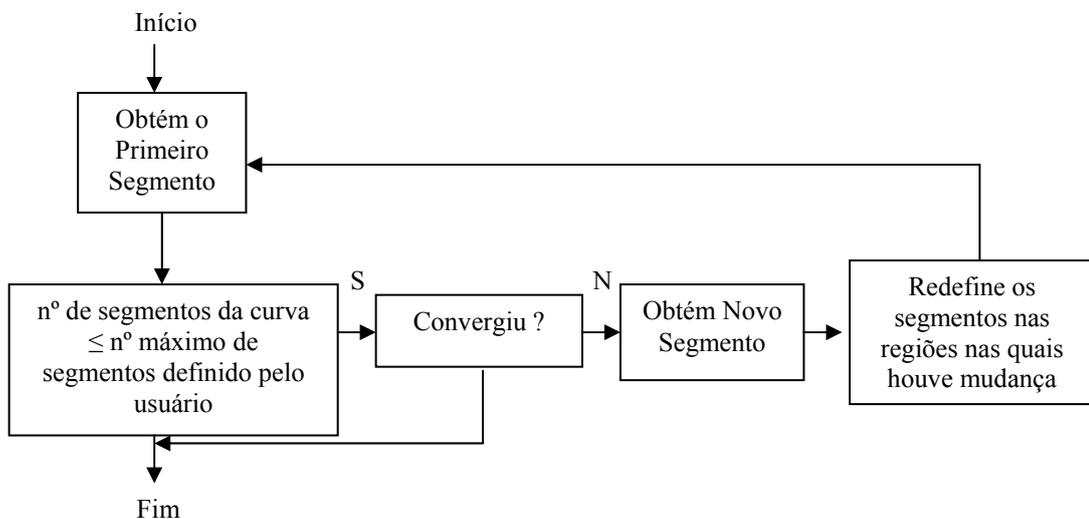


Figura 4.4 – Fluxograma do algoritmo *K-segmentos não-suave*

O algoritmo de Verbeek et. al. [8] não busca diretamente uma curva autoconsistente, conforme proposto por Hastie e Stuetzle [4]. O algoritmo de K-segmentos não-suave busca inserir segmentos em regiões de Voronoi [55]. Estas regiões são compostas por amostras que estão mais próximas de um ponto (centro da região) do que de um dos segmentos que compõe a curva principal e são obtidas com o algoritmo k-means [55]. A curva obtida é uma curva principal devido a dois fundamentos:

- Os segmentos com direção da primeira componente principal são autoconsistentes [7].
- Se a região é uma região de Voronoi, então os k pontos pertencentes a ela são autoconsistentes [60].

Na figura 4.5 é mostrada uma curva principal obtida de um conjunto de dados de descargas parciais utilizando o algoritmo K-segmentos não-suave. O algoritmo inicia traçando uma linha reta correspondente à direção da primeira componente principal. Em seguida, o algoritmo vai adicionando segmentos à curva e interligando-os. Neste exemplo definiu-se uma curva com 3 segmentos.

#### **4.4 Aplicações**

As curvas principais têm sido utilizadas em diversos tipos de aplicações [4, 7, 34, 52, 58, 61]. Em Hastie e Stuetzle [4], relatou-se aplicações no auxílio do alinhamento de equipamentos de colisão de partículas. Em Banfield e Raftery [58], as curvas principais foram empregadas para extrair contornos de blocos de gelo em imagens de satélite. Em Young [34], utilizou-se as curvas para digitalizar caracteres de textos escritos a mão. Em Dewang [61] utilizaram-se as curvas para definir fluxos mais intensos de transito de veículos. Em Chang e Ghosh [52], as curvas principais são aplicadas na classificação de padrões. Neste trabalho, são utilizados vários conjuntos de dados, como o conhecido conjunto Íris [62]. As curvas obtiveram uma grande taxa de acertos na classificação destes sinais comparado com outros métodos. Em Fernandez [7], as curvas também foram utilizadas em classificadores para identificar padrões de navios e bons resultados foram obtidos. Neste trabalho foram empregados os métodos da distância euclidiana mínima e redes neurais para classificar padrões. Os resultados

são bastante motivadores atingindo elevados índices de acertos, tanto com o classificador de distância euclidiana mínima quanto com o classificador híbrido com redes neurais.

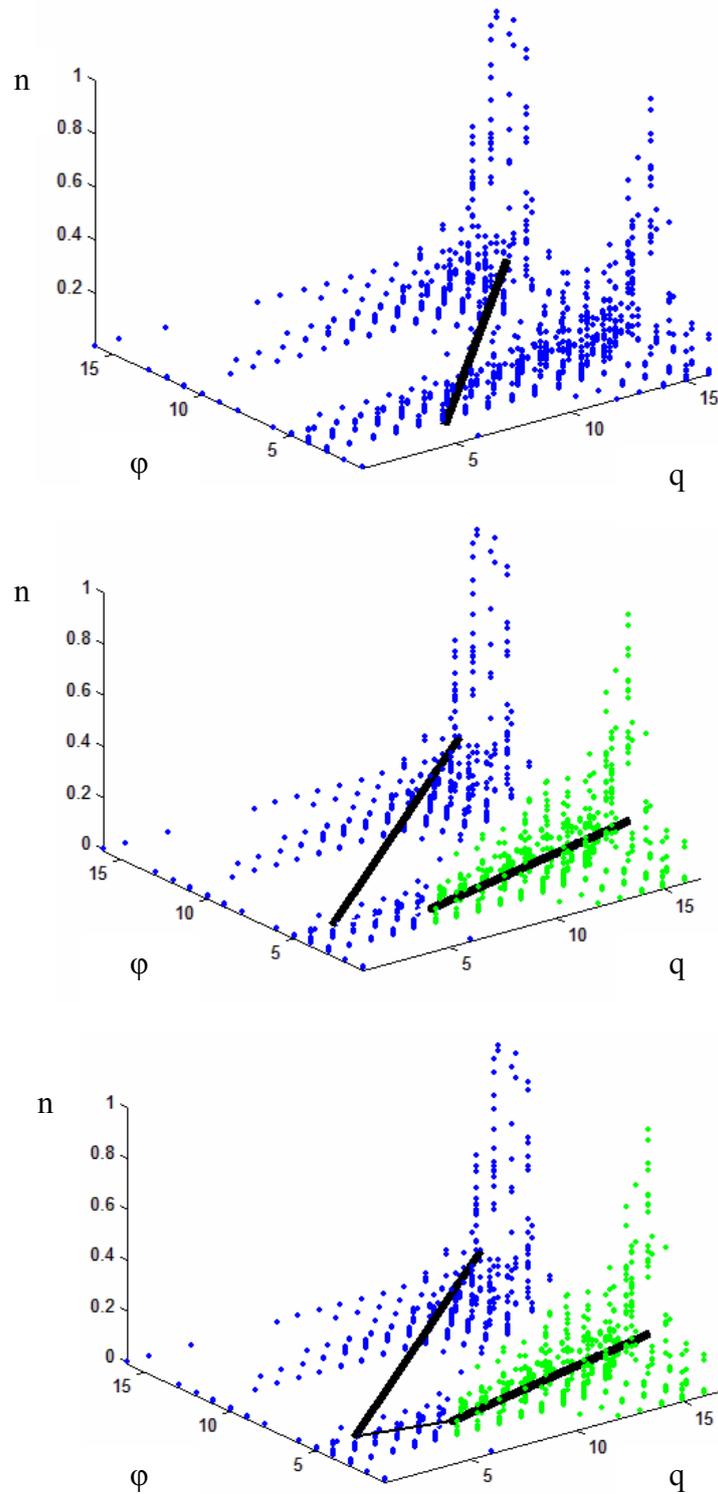


Figura 4.5 – Obtenção da curva principal de um mapa de descargas parciais padrão Ponta Flutuante (FL). Os segmentos mais grossos correspondem à direção da primeira componente principal em uma região de Voronoi (região destacada) e o segmento mais tênue é a ligação entre eles.

## 4.5 Conclusão

As curvas principais podem ser interpretadas como uma generalização não-linear de componentes principais. E, assim como as componentes principais, as curvas são capazes de extrair características de um conjunto de dados e representá-la em modelos compactos. As curvas principais podem alimentar sistemas classificadores, assim como os relatados no capítulo 3. E, com o aperfeiçoamento dos algoritmos de extração de curvas principais, esta técnica tem se mostrado simples, robusta e eficaz ao lidar com dados reais.

# Capítulo 5

## Base de Dados

Em situações práticas, a aquisição de dados consistentes não é uma tarefa fácil. Em campo é complicado devido à presença de ruído. Já em laboratório, além da sofisticação das instalações e dos equipamentos envolvidos, são exigidos grandes esforços na montagem de modelos apropriados para a coleta de DPs. Sendo assim, não se têm grandes quantidades de dados disponíveis para estudos. Como reflexo das dificuldades na medição de eventos reais de DPs, a base de dados deste trabalho contou com apenas 115 eventos. Além disso, devido a esta estatística reduzida, foi necessário considerar o problema de definição dos conjuntos de desenvolvimento e de teste, visando à generalização do projeto do classificador de padrões de DPs. No caso do desenvolvimento, inclui-se também a extração de características. A questão é: quantos e quais eventos devem compor o conjunto de extração de características (conjunto de desenvolvimento) e quantos e quais devem compor o conjunto de teste de capacidade de generalização do projeto classificador (conjunto de teste)? Neste contexto, os enfoques de separação de conjuntos, indicados em Cuenca [6] e Diniz [5], foram adotados no presente trabalho.

Os conjuntos de desenvolvimento e de teste em Cuenca [6] foram definidos com base no enfoque especialista. O conjunto composto pelos mapas de DPs mais “limpos” corresponde ao conjunto desenvolvimento. Os eventos que melhor simulam uma situação real compõem o conjunto de teste. A ênfase dada nesta separação é a capacidade de generalização do classificador, ou seja, a capacidade de classificar corretamente sinais ruidosos e menos representativos dos modelos.

Já em Diniz [5], a definição dos dois conjuntos baseou-se em um critério estatístico: o critério de validação cruzada [63]. A figura de mérito deste enfoque estatístico foi a quantidade de energia acumulada em cada conjunto. Buscou-se uma divisão da base de dados em conjuntos de desenvolvimento e de teste que apresentassem uma extensão de energia, quando se usa uma análise de componentes principais, aproximadamente iguais para os dois conjuntos. A ênfase desta separação é

a obtenção de um conjunto que represente estatisticamente os dados (conjunto de desenvolvimento) e outro conjunto que melhor generalize este conjunto de desenvolvimento (conjunto de teste).

## **5.1 Coleta dos dados**

Os dados utilizados neste trabalho foram coletados nos laboratórios do Centro de Pesquisas Elétricas da Eletrobrás (CEPEL) e foram utilizados em [5, 6, 9]. Um modelo experimental composto de três células-teste, baseado em norma específica, foi montado. Os sinais de DP foram artificialmente provocados e, em seguida, coletados pelo equipamento digital de descargas parciais - o ICM *system* [64].

### **5.1.1 Modelo experimental**

O circuito de ensaio e medição foi montado conforme os procedimentos da norma IEC 60270 [10]. Foram utilizadas três células como objeto de ensaio que simula regiões internas de um transformador de potência, as quais são críticas e sujeitas a possível formação de DPs. Cada uma delas é constituída de material acrílico e translúcido, o que permite a visualização da formação de descargas. Na figura 5.1, um esquema simplificado mostra as três configurações implementadas para a aquisição dos dados disponibilizados para o presente trabalho. Os pulsos são gerados por eletrodos de ponta de aço de 1mm e 2mm de comprimento e raio de curvatura de 50  $\mu\text{m}$ . A distância entre os eletrodos varia entre 1mm e 4 mm e o dielétrico utilizado é óleo isolante próprio para transformador. O objetivo é reproduzir DPs que representem defeitos reais em campo, conforme descrição:

- Descargas no eletrodo de alta tensão (ponta-plano - PP): representam descargas parciais envolvendo condutores de alta tensão, tais como: descargas nas saídas de alta tensão, descargas em objetos condutores flutuantes, descargas na bucha do transformador ou em grampos, descargas provocadas por mau contato nos conectores e descargas nos enrolamentos de alta tensão.
- Descargas no eletrodo intermediário (eletrodo flutuante - FL): representam descargas entre camadas de bobinas e descargas induzidas envolvendo o enrolamento do lado de alta e baixa tensão.

- Descargas no eletrodo aterrado (ponta-terra - PT): representam descargas entre a bucha e o tanque aterrado do transformador, DPs nas bobinas e nos núcleos de alta e baixa tensão e em grampos de fixação, conectores, tirantes e jugos.

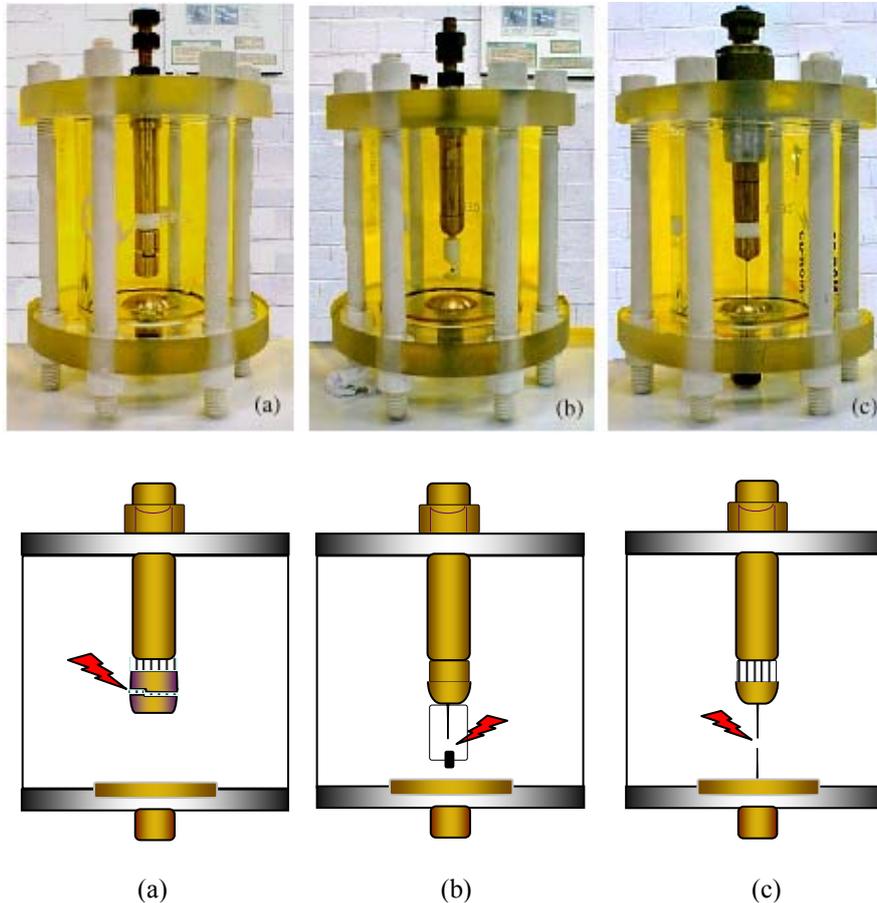


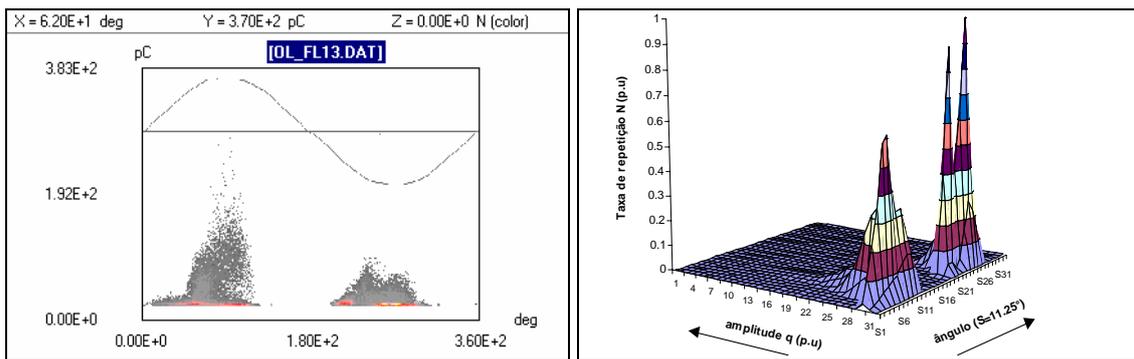
Figura 5.1 – Modelos experimentais montados para a aquisição de descarga parciais nas configurações (a) Ponta Flutuante, (b) Ponta-Plano e (c) Ponta-Terra. Na figura superior é mostrado o modelo real e na inferior o esquema simplificado. Extraído de [9]

Para cada uma das três células foi realizada uma seqüência de medições independentes, de tal forma que se pudesse ter uma assinatura representativa de cada uma das configurações. Os objetivos da simulação em laboratório são:

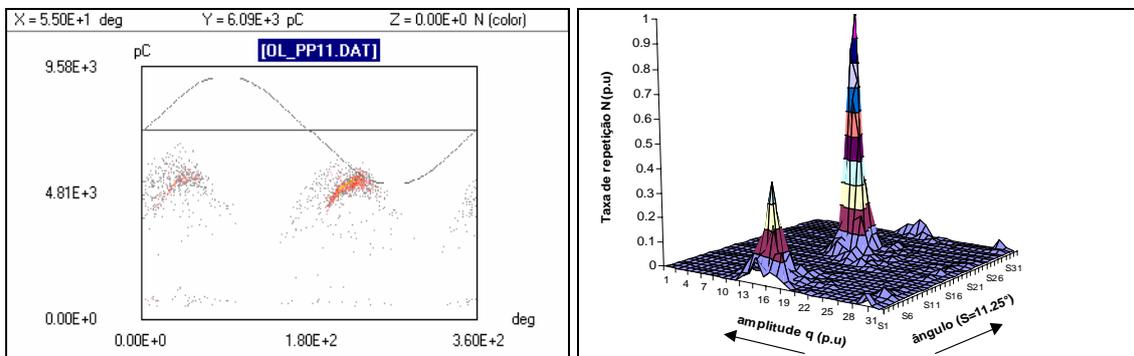
- Simplificar a complexidade dos equipamentos;
- Conhecer previamente os parâmetros envolvidos;
- Conhecer a fonte de descargas parciais;
- Sintonizar melhor os sinais;
- Discriminar e atenuar os ruídos;
- Formar padrões consistentes e livres de perturbações.

### 5.1.2 Equipamento de Medição

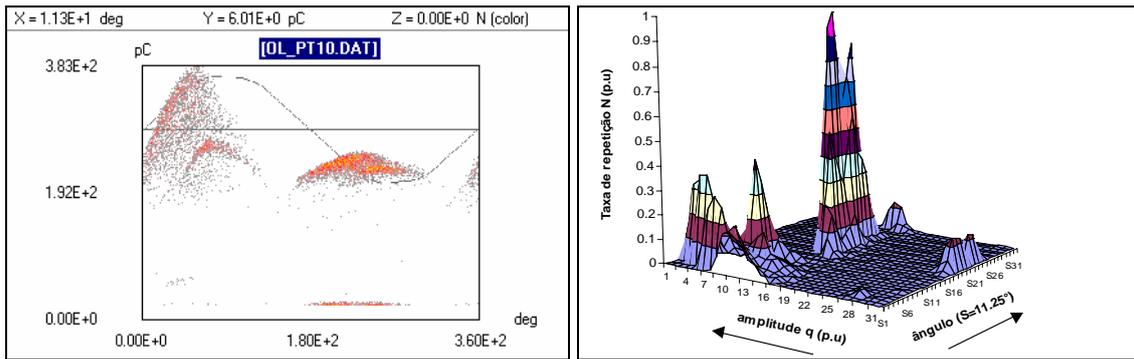
O sistema digital de medição - ICM [64] - é o equipamento responsável por detectar e medir as descargas parciais. Este equipamento coleta e armazena os sinais de DPs produzidos nas células de teste. Além disso, é possível conectá-lo a um computador portátil para que os sinais coletados sejam visualizados em função de seus parâmetros. O sistema de visualização do ICM constrói mapas com os três parâmetros relacionados às DPs: a fase ( $\phi$ ) da tensão aplicada (senóide), a amplitude ( $q$ ) e taxa de ocorrência ( $n$ ). Estes mapas podem ser visualizados em duas ou três dimensões. Em duas dimensões a quantidade de descargas é diferenciada por uma grade de cores. Já em 3 dimensões, a taxa de ocorrência, a fase e a amplitude são dadas como coordenadas dos 3 eixos do gráfico ( $n, \phi, q$ ) (veja figura 5.2).



(a)



(b)



(c)

Figura 5.2 – Sistema de visualização do ICM dos padrões (a) Ponta Flutuante, (b) Ponta-Plano e (c) Ponta-Terra, em duas dimensões do lado esquerdo e em três dimensões do lado direito. Extraído de [6].

## 5.2 Tratamento das informações

As informações fornecidas pelo instrumento de medição são armazenadas em arquivo binário de uso exclusivo do programa de gerenciamento do instrumento. Para que se possa trabalhar com estas informações, foi feita a tradução do código binário para código decimal. Neste formato, os eventos são disponibilizados pelo instrumento de medição como histogramas bidimensionais 256x256. A resolução dos canais corresponde à amplitude das descargas e a fase da tensão de teste na qual elas ocorrem. Dessa forma, a amplitude da DP é discretizada em 256 níveis, de zero até um valor máximo (definido a partir de uma configuração do *trigger* do instrumento). A fase da senóide é discretizada de zero até 360 graus.

A compactação em histogramas reduzidos é realizada acumulando-se descargas em canais de menor resolução [6, 9] (veja figura 5.3). A compactação é realizada somando-se as descargas acumuladas em agrupamentos de canais vizinhos no histograma. A seguir, os valores são normalizadas pela quantidade de DPs no canal de maior acúmulo. A compressão diminui a resolução original, sendo necessária devido a grande quantidade de amostras coletadas pelo equipamento de medição. No entanto, estudos anteriores [5, 6, 9] mostram que é possível obter-se bons resultados de classificação em mapas reduzidos.

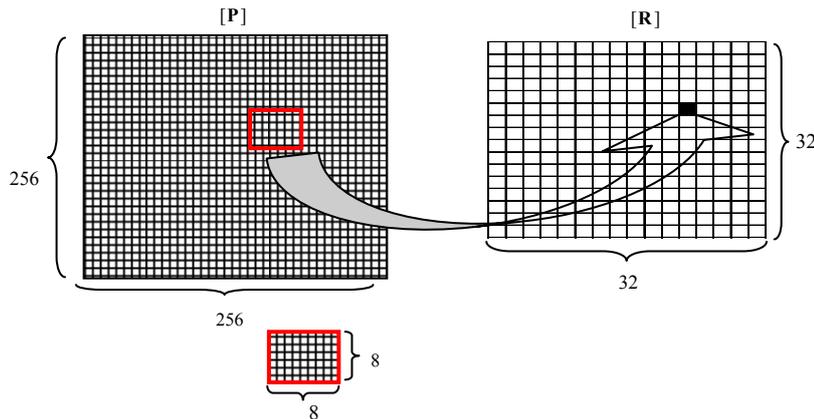


Figura 5.3 – Ilustração da redução de dimensionalidade do histograma primitivo [P] 256x256 em um histograma reduzido [R] 32x32, acumulando-se as DPs em canais maiores. Extraído de [6].

### 5.3 – Separação da Base de Dados

A separação da base de dados em conjuntos de desenvolvimento e teste foi realizada de duas formas distintas. Em Cuenca [6] foi utilizado o conhecimento *a priori* de um especialista para fazer esta separação. Já em Diniz [5] utilizou-se informações estatísticas extraídas dos dados disponíveis. Cada um dos trabalhos deu ênfase a aspectos diferentes do problema. O primeiro enfatizou a capacidade de generalização do classificador, sob condições práticas de ruído, e o segundo enfatizou o desempenho do classificador segundo sua eficiência de classificação.

#### 5.3.1 Enfoque Especialista

A separação realizada em Cuenca [6] baseou-se no conhecimento de especialistas no assunto de descargas parciais para separar os sinais em conjuntos para o desenvolvimento e o teste do classificador. Neste enfoque, eventos com características topológicas obtidas por pré-visualização dos mapas de DPs, por modelos típicos e por informações acumuladas dos especialistas, além de algumas amostras menos contaminada por ruído, foram utilizadas para o desenvolvimento do projeto do classificador. Conseqüentemente, os eventos mais contaminados por ruídos e aqueles menos representativos do respectivo padrão pertenceriam ao conjunto de teste.

A justificativa para este critério de separação é de que os sinais mais “limpos” representam mais fielmente um mapa ideal de descargas parciais e o conjunto de teste poderia avaliar a capacidade de generalização do projeto em condições reais de operação. Os conjuntos de desenvolvimento e teste ficaram compostos de 61 e 54 eventos, respectivamente. Os conjuntos correspondentes aos três tipos de padrões (FL, PP e PT) ficaram configurados da seguinte forma:

- Padrão Ponta Flutuante (FL): 20 eventos para o conjunto de desenvolvimento e 15 para o conjunto de teste.
- Padrão Ponta Plano (PP): 26 eventos para o conjunto de desenvolvimento e 29 para o conjunto de teste.
- Padrão Ponta Terra (PT): 15 eventos para o conjunto de desenvolvimento e 10 para o conjunto de teste.

### **5.3.2 Enfoque estatístico**

O mesmo banco de dados foi dividido com base no método da validação cruzada [63]. Este critério foi utilizado após a constatação de que os dois conjuntos na separação especialista apresentavam uma grande discrepância de energia, quando se retinham as dez primeiras componentes principais de cada conjunto. Na separação especialista, observou-se que o conjunto de desenvolvimento correspondia a somente 8% da energia total dos eventos, enquanto o conjunto de teste correspondia a 79%. Como o conjunto de desenvolvimento é composto por eventos com menos ruídos, a variância (e conseqüentemente a energia) deste conjunto é menor. Já o conjunto de teste, por ser composto de eventos com presença significativa de ruído, tem uma variância bem maior. Dessa forma, era de se esperar que a energia retida pelo conjunto de teste correspondesse a maior parte da energia.

Para contornar o problema da discrepância de energia no enfoque especialista, em Diniz [5] propôs-se a escolha dos conjuntos de desenvolvimento e de teste empregando-se o método da validação cruzada. Neste caso, assume-se que o ruído faz parte dos dados. Assim, selecionam-se, para o conjunto de desenvolvimento, os eventos que representam os dados estatisticamente. A seguir, selecionam-se, para o conjunto de

teste, os eventos que melhor generalizam o conjunto de desenvolvimento. A obtenção dos conjuntos de desenvolvimento e teste através da validação cruzada seguiu os seguintes procedimentos:

- Distribuição de forma aleatória de todos os eventos disponíveis em dois conjuntos com aproximadamente a mesma quantidade de eventos.
- A seguir, varia-se a constituição de cada conjunto, transferindo eventos de um para o outro. Cada modificação do conjunto original forma um grupo.
- Formaram-se 5 grupos de dois conjuntos.
- Para cada um dos dois conjuntos do bloco, mede-se o acúmulo de energia retida nas dez primeiras componentes principais.

O resultado desta separação é apresentado na tabela V.1 . Observa-se que a melhor separação de conjuntos se dá para a terceira combinação; uma vez que este grupo tem a menor diferença entre os percentuais de energia dos conjuntos de desenvolvimento e de teste.

*Tabela V.1 – Resultado da separação por validação cruzada. Extraído de [11]*

<b>Combinação</b>	<b>Percentual de Energia do Conjunto de Desenvolvimento</b>	<b>Percentual de Energia do Conjunto de Teste</b>
1 <sup>a</sup>	74%	49%
2 <sup>a</sup>	74%	54%
<b>3<sup>a</sup></b>	<b>72%</b>	<b>63%</b>
4 <sup>a</sup>	73%	53%
5 <sup>a</sup>	76%	42%

Os novos conjuntos de desenvolvimento e teste passaram a ser compostos por 70 e 45 eventos, respectivamente. Os conjuntos correspondentes aos três tipos de padrões ficaram configurados da seguinte forma:

- Padrão Ponta Flutuante (FL): 21 eventos no conjunto de desenvolvimento e 14 no conjunto de teste.
- Padrão Ponta Plano (PP): 32 eventos no conjunto de desenvolvimento e 23 no conjunto de teste.
- Padrão Ponta Terra (PT): 17 eventos no conjunto de desenvolvimento e 8 no conjunto de teste

## **5.4 Conclusão**

Os dados de descargas parciais coletados dos modelos experimentais foram divididos em conjuntos de desenvolvimento e conjunto de teste, segundo dois enfoques diferentes de projeto. No presente trabalho, avaliam-se estas duas visões simultaneamente, aplicando-se a técnica de curvas principais.

## Capítulo 6

# Curvas Principais Aplicadas no Reconhecimento de Padrões de DPs

As curvas principais têm sido utilizadas para a classificação de padrões em diversos tipos de dados, conforme foi visto no capítulo 4. Este trabalho propõe a aplicação das CPs na análise de dados de descargas parciais. O que se procura é um método compacto e que também permita a visualização dos modelos extraídos.

O classificador baseado em curvas principais foi realizado a partir de uma configuração dos histogramas de DPs que podia operar no espaço de 1024 ou de 3 dimensões. No primeiro caso, concatenaram-se os canais do histograma na direção de uma serpentina. No segundo caso, indexaram-se os canais com DPs acumuladas como amostras em 3 dimensões. Nestas configurações, as curvas principais dos respectivos modelos (FL, PP ou PT) foram extraídas.

Realizada a extração, alimentaram-se os classificadores com as curvas principais. Em seguida, utilizou-se a distância entre os eventos e os segmentos de cada modelo como informação para se treinar uma rede neural, aplicar o método da distância Euclidiana mínima ou o método de pontuação (votação).

### 6.1 Extração de Características

A extração das características dos dados de DPs, através das curvas principais, foi realizada utilizando-se os mapas de descargas parciais no formato 32x32 [6, 9] ou 16x16. A primeira configuração foi utilizada para formar padrões em 1024 dimensões, já a segunda foi utilizada para formar padrões que podiam ser visualizados no espaço tri-dimensional.

### 6.1.1 Mapas Concatenados (1024 Dimensões)

Os dados utilizados neste trabalho foram disponibilizados em matrizes 32x32, já compactadas a partir dos dados originais 256x256 (seção 5.2). Os dados compactados compõem-se por 1024 elementos em cada matriz. O conjunto de dados é composto de um total de 115 eventos, ou seja, 115 matrizes 32x32. Estes eventos foram separados em dois conjuntos: conjunto de desenvolvimento e conjunto de teste.

O conjunto de desenvolvimento foi utilizado na extração das características das descargas parciais. Cada evento deste conjunto foi concatenado na direção de uma serpentina, formando eventos em 1024 dimensões (veja figura 6.1).

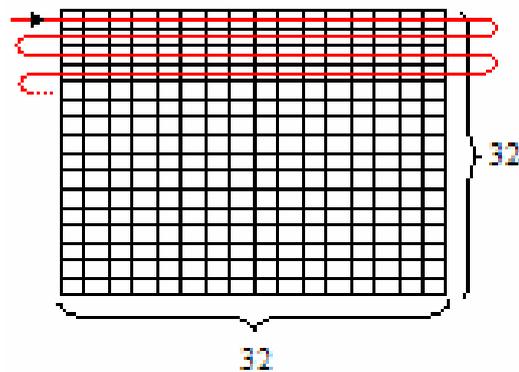


Figura 6.1 – Concatenação do histograma de 1024 canais, formando eventos em 1024 dimensões.

Os eventos pertencentes ao conjunto de desenvolvimento, relacionados a um determinado padrão (FL, PP ou PT), representam um modelo. As curvas principais são extraídas pelo algoritmo K-segmentos (seção 4.3.1) a partir do modelo de 1024 dimensões do respectivo padrão. A figura 6.2 esquematiza como são realizados estes procedimentos.

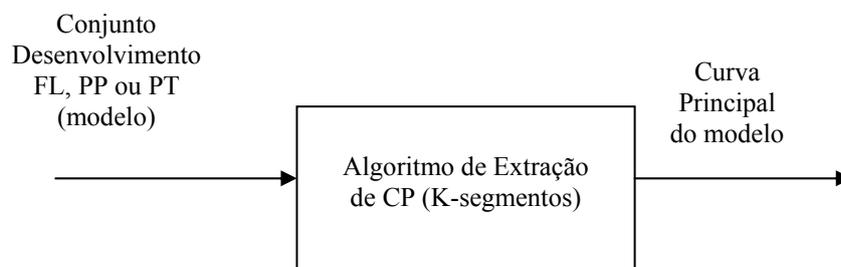


Figura 6.2 – Extração das CPs dos modelos de DPs.



explicado anteriormente. O modelo de mapa tridimensional FL, obtido com estes procedimentos, é mostrado na figura 6.4.

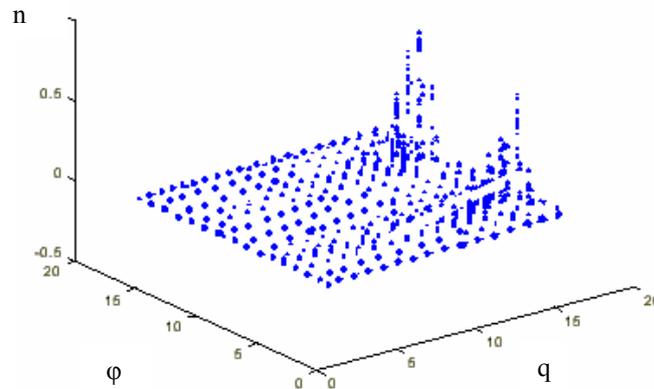


Figura 6.4 – Padrão FL típico, em 3 dimensões.

A classificação nesta configuração apresentou um desempenho demasiadamente baixo. Um fator relevante que explica este resultado é indexação de canais do histograma que não tiveram sido efetivamente preenchidos por sinais de DPs. Os sinais se discriminam melhor através dos canais preenchidos. Isso pôde ser constatado ao se obter resultados de classificação com erros sistemáticos. A solução encontrada para resolver este problema foi retirar os canais que não tivessem sido preenchidos por DPs (veja figura 6.5).

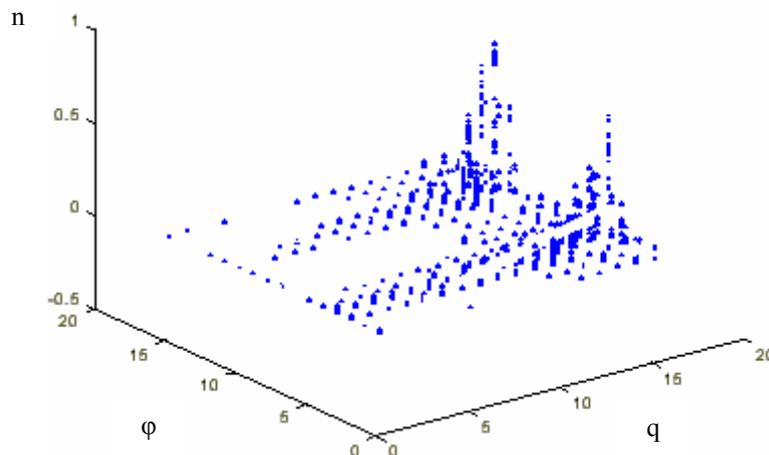


Figura 6.5 – Um mapa de DPs em 3 dimensões sem os elementos nulos.

Assim, um evento é composto de várias amostras de 3 dimensões, podendo chegar até 256 amostras por evento. Com esta configuração, o método de extração obtém as curvas principais para o modelo de cada padrão (FL, PP ou PT). Na figura 6.6 estes procedimentos são esquematizados.

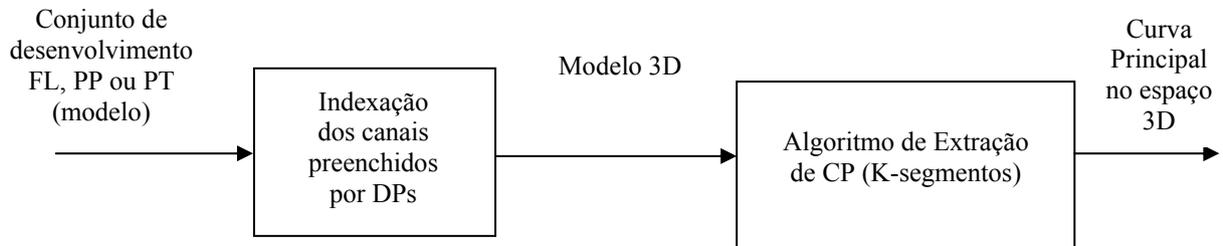


Figura 6.6 – Extração das curvas principais no espaço de 3 dimensões.

## 6.2 Classificação

O classificador trabalha alimentado pelas curvas principais de cada um dos padrões (FL, PP ou PT). Em uma primeira etapa de classificação, calculam-se as distâncias euclidianas entre o evento de entrada e os segmentos das curvas principais extraídas para cada modelo. Em uma segunda etapa, estas distâncias são avaliadas segundo alguns critérios de classificação abordados no capítulo 4, a saber: distância mínima, redes neurais e votação.

Na figura 6.7 são mostrados os procedimentos de obtenção das distâncias euclidianas, que representam os vetores característicos de cada evento de teste. O evento de teste é apresentado na entrada do classificador de curvas principais. As distâncias do evento aos segmentos de cada curva (FL, PP e PT) são medidas e apresentadas na saída na forma de vetores distância ( $\mathbf{d}_{fl}$ ,  $\mathbf{d}_{pp}$  e  $\mathbf{d}_{pt}$ ).

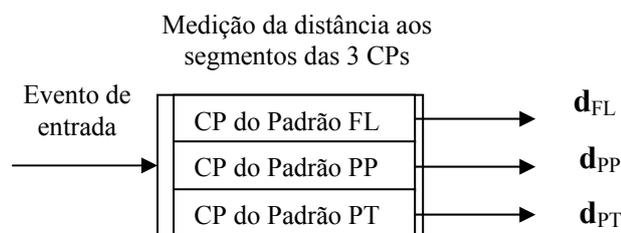


Figura 6.7 – Etapa de Classificação. A distância euclidiana é medida entre o evento de teste e os segmentos das curvas principais.

## 6.2.1 Classificação em 1024 dimensões

Em eventos no espaço de 1024 dimensões foram utilizados dois métodos de classificação: o método da distância euclidiana mínima e o método baseado em redes neurais.

### 6.2.1.1 Classificação pela Distância Mínima

A forma mais simples e direta de classificação foi obtida pela medida mínima da distância euclidiana entre o evento de entrada e a curva principal extraída para cada modelo de DP. O evento é associado àquela classe cuja distância é a menor distância encontrada, ou seja, o evento é associado à classe mais próxima. Apesar de simples, este método é bastante eficaz, como veremos no capítulo 7, sendo capaz de classificar padrões eficientemente sem muito esforço computacional.

### 6.2.1.2 Classificação utilizando Redes Neurais

Os vetores distância, correspondentes aos dados no espaço de 1024 dimensões, também foram analisados através de uma rede neural (veja a figura 6.8) treinada através da retro propagação do erro - algoritmo *backpropagation*. O vetor de distâncias, referentes ao evento que se deseja classificar, alimentou a entrada da rede. O número de nós da camada de entrada foi feito igual ao somatório do número de segmentos extraídos para os modelos, (inclusive as ligações entre os segmentos que compõem a curva). O número de neurônios da camada intermediária variou entre a quantidade de neurônios da camada de entrada e a da camada de saída. Esta última, composta de 3 neurônios, com cada um associado a uma classe de padrão (FL, PP e PT) dos eventos.

A rede foi treinada com os vetores de distância euclidiana dos eventos que compõem o conjunto de desenvolvimento. Denominaremos o grupo destes vetores de conjunto de treinamento. O conjunto de treinamento constitui-se de dados do tipo *número de eventos x distância aos segmentos das curvas FL, PP e PT*. Por exemplo, dado o conjunto de desenvolvimento especialista, com 61 eventos, foram extraídas curvas do padrão FL, PP e PT com 1, 1 e 2 segmentos, assim, o conjunto de treino da rede será uma matriz do tipo 61x5. O número 5 refere-se aos quatro segmentos dos padrões mais a ligação entre os segmentos do padrão PT. A rede neural foi treinada com

os pares entrada-saída do conjunto de treinamento. As saídas da rede correspondem a apenas um neurônio ativado, aquele referente à classe de entrada, enquanto os outros permanecem desativados. Estes pares foram misturados aleatoriamente. Os vetores formam normalizados entre os valores 0 e 1 para não saturar a rede. No teste o neurônio que apresenta a maior saída é considerado vencedor.

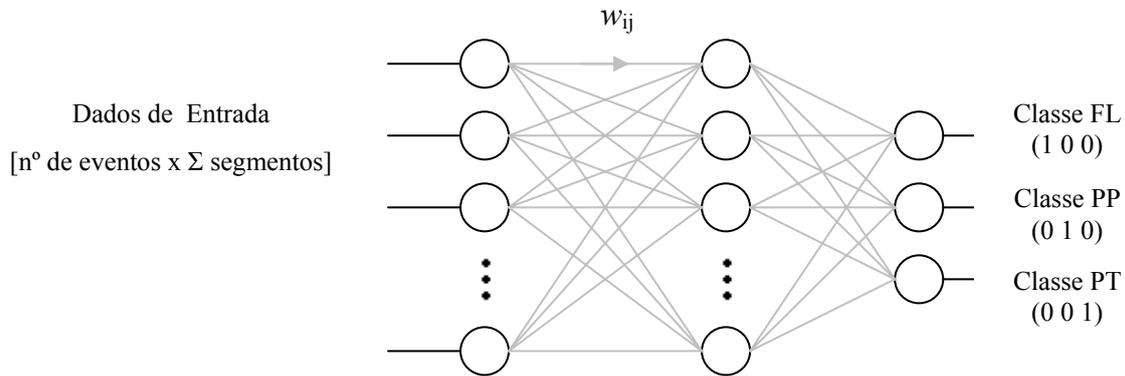


Figura 6.8 – Classificador baseado em Redes neurais. O neurônio ativo (saída=1) indica a qual classe o evento pertence.

## 6.2.2 Classificação 3D

A classificação em mapas acumulativos (seção 6.1.2) utilizou o método da distância euclidiana mínima, as redes neurais e também o método da votação. Nesta configuração, consegue-se visualizar os modelos extraídos para cada padrão.

### 6.2.2.1 Classificação pela Distância Mínima

Uma forma para se avaliar a distância, no método 3D, é através da média e do valor do erro médio quadrático (RMS) do vetor distância obtido com o conjunto de teste referente a cada classe (FL, PP e PT). Por exemplo, o vetor distância entre um evento e uma curva corresponde às médias das 256 distâncias das amostras aos segmentos do modelo (este valor refere-se a mapas 16x16 caso todos os canais tivessem sido preenchidos). Assim, se a curva for composta por três segmentos, haverá, para cada evento, 256x3 distâncias à curva principal, o que proporciona um vetor distância 1x3; ou seja, este vetor é composto por três distâncias médias. Na figura 6.9 é mostrada a

distância da  $i$ -ésima amostra de um evento FL a um segmento da curva principal, que será utilizado para encontrar a média e o valor RMS em relação àquele segmento.

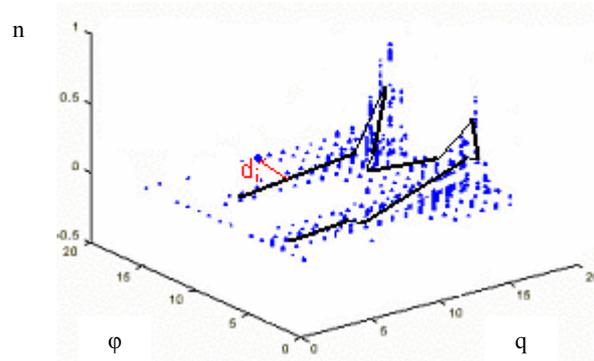


Figura 6.9 – Ilustração da medida da distância entre um segmento da CP e uma amostra do evento.

#### 6.2.2.2 Classificação utilizando Redes Neurais

A classificação por redes neurais procedeu-se de forma semelhante ao caso de mapas concatenados, com a diferença de utilizar um valor médio no lugar de um valor simples para as distâncias. Os procedimentos se realizaram da seguinte forma. Os elementos nulos do conjunto de desenvolvimento foram extraídos. Computaram-se as distâncias médias dos elementos não-nulos aos segmentos das três curvas. Formaram-se, assim, vetores com as médias das distâncias entre as amostras dos eventos do conjunto de desenvolvimento a cada segmento das curvas -  $\mathbf{d}_{fl}$ ,  $\mathbf{d}_{pp}$  e  $\mathbf{d}_{pt}$ . Estes dados são utilizados para treinar a rede e têm a mesma configuração dos dados utilizados nas redes neurais aplicadas em mapas concatenados; ou seja, *quantidade de eventos  $\times$  somatório da quantidade de segmentos das curvas*. Com os eventos do conjunto de teste realizaram-se os mesmos procedimentos de extração da distância média aos segmentos das curvas e, dessa forma, foram apresentados na entrada da rede para serem classificados.

#### 6.2.2.3 Classificação por Votação

O procedimento de classificação utilizando a votação (seção 3.2.1.2) utiliza a informação da distância média. No entanto, diferentemente do método da distância

mínima, agora a figura de mérito é a quantidade de vezes que um padrão é referenciado. Assim, para cada amostra do evento, é calculada a distância a cada uma das curvas do respectivo padrão (FL, PP e PT). Aquele mais próximo ao evento recebe um voto. Assim, o evento pertencerá àquele padrão com o maior número de votos. Quantidades iguais de votos para dois padrões são resolvidas arbitrariamente.

### **6.3 – Conclusão**

Este capítulo mostrou os procedimentos utilizados para a aplicação de curvas principais no reconhecimento de padrões de descargas parciais. Foram mostradas as configurações utilizadas na etapa de extração das curvas principais e na etapa de classificação dos eventos de teste, tanto para o método de mapas concatenados quanto para mapas acumulativos. As curvas principais extraídas pelo algoritmo K-segmentos foram submetidas aos classificadores de distância euclidiana mínima, de votação e de redes neurais. Os resultados obtidos são apresentados no capítulo seguinte.

# Capítulo 7

## Resultados

Os resultados obtidos neste trabalho originaram-se de duas análises dos mapas de descargas parciais: a extração de características e a classificação de eventos. A extração de características é a fase na qual as curvas principais são obtidas. A etapa de classificação é a etapa na qual os eventos são comparados à curva principal de cada modelo. Neste contexto, também são obtidos os resultados através de dois tipos enfoques: especialista e estatístico.

A similaridade entre um evento e os padrões possíveis é medida de acordo com critérios apropriados (seção 3.2). Os resultados deste trabalho foram obtidos com os métodos da distância euclidiana mínima, redes neurais e votação. Esses métodos foram utilizados de acordo com a configuração dos dados experimentais. Em histogramas de canais concatenados (1024 dimensões), foram utilizadas a distância euclidiana mínima e as redes neurais. Já para o caso de histogramas com descargas acumuladas (3D), utilizou-se, além dos dois métodos anteriores, o critério de votação. Este capítulo mostra os resultados obtidos nas análises da extração de características e da classificação.

### 7.1 Extração de Características

A etapa de extração de características dos mapas de DPs é realizada utilizando-se o algoritmo K-segmentos não-suave (seção 4.3.1). Para se obter uma curva principal, os modelos do conjunto de desenvolvimento devem ser fornecidos ao algoritmo (seção 5.3). Este conjunto de dados é disposto em duas configurações mostradas no capítulo anterior (1024D ou 3D). A saída do algoritmo corresponde à curva principal do modelo-padrão de entrada (FL, PP ou PT) e é composta de vários segmentos dispostos no espaço de 1024 ou de 3 dimensões.

### 7.1.1 Extração em 1024 Dimensões

A curva principal extraída para cada modelo compõe-se de vários segmentos que são obtidos de forma incremental (um a um), a cada iteração do algoritmo. O processo de incremento de segmentos na curva é paralisado quando o algoritmo chega a uma região de Voronoi com menos de 3 amostras. A determinação da quantidade de segmentos de uma curva principal é um ponto crítico da extração de características. Quando há muitos segmentos pode haver um super aprendizado, afetando-se a capacidade de generalização do método. Por outro lado, uma quantidade reduzida de segmentos poderia afetar a eficiência do classificador.

Para determinar a quantidade de segmentos necessários para cada curva principal, foram analisados dois índices: o desempenho de classificação obtido dos conjuntos de desenvolvimento e o desempenho obtido do conjunto de teste. Estes procedimentos foram realizados tanto para classificadores estatísticos quanto para classificadores especialistas. As análises teóricas mostrarão que se devem buscar classificadores que sejam compostos por curvas mais segmentadas, em casos de extração no espaço de 1024 dimensões e, no espaço 3D, devem-se observar o melhor desempenho obtido do conjunto de desenvolvimento. No caso de classificadores estatísticos, a sugestão oferecida pelo método de extração, através da análise da modificação no comprimento total da curva após a inserção de um novo segmento, mostrou-se um bom método de escolha dos segmentos.

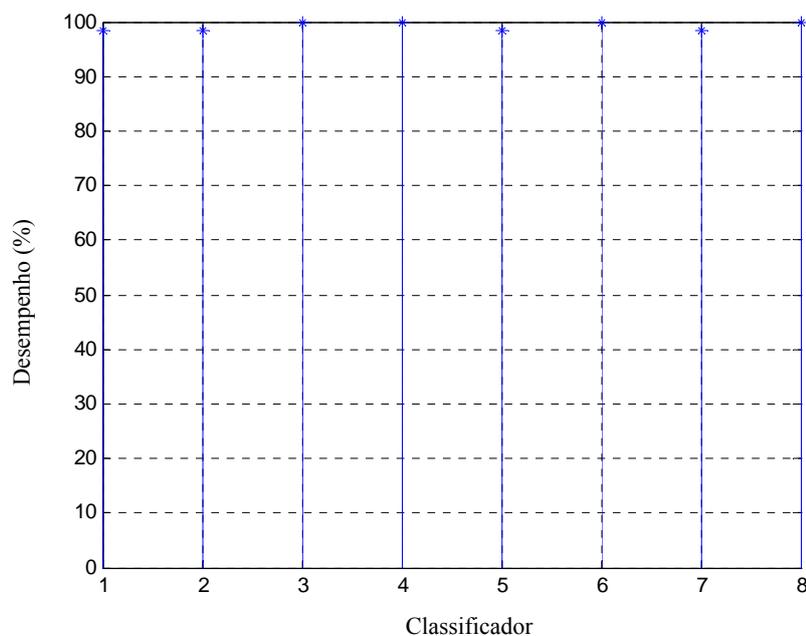
#### 7.1.1.1 Visão Especialista

No conjunto de desenvolvimento com mapas em 1024 dimensões - obtido da separação de conjuntos baseada em conhecimento de especialista -, o número máximo de segmentos alcançado pelo algoritmo foi 1, 4 e 2, para os padrões FL, PP e PT, respectivamente. Dessa forma, é possível se ter oito classificadores diferentes baseados nas combinações possíveis entre as curvas. A tabela VII.1 mostra as diversas combinações de curvas que podem compor o classificador. Para cada classificador, os desempenhos podem ser visualizados na figura 7.1 e 7.2. A figura 7.1 indica que o desempenho obtido do conjunto de desenvolvimento está acima de 98%. Isto mostra que, neste caso, a quantidade de segmentos contida em cada curva não têm alterado o desempenho de classificação de maneira significativa. Dessa forma, com base apenas

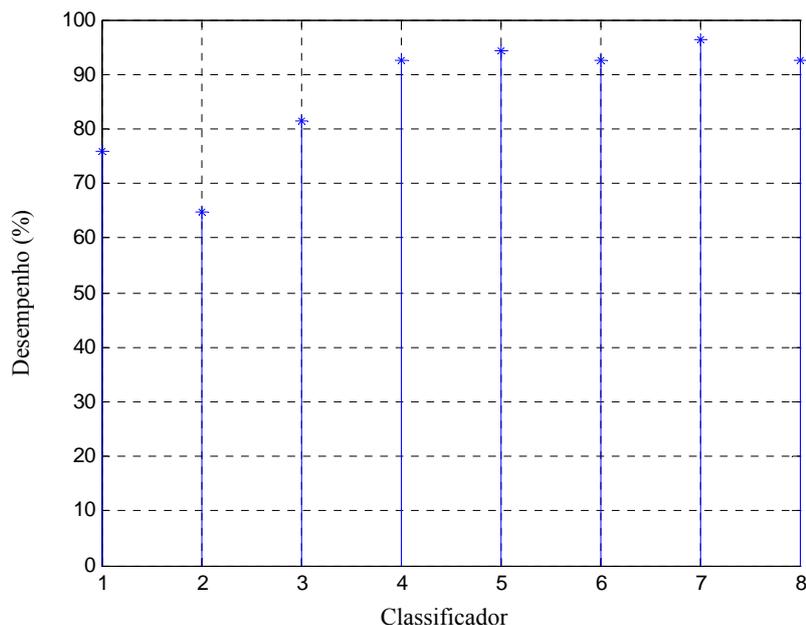
neste índice, é indiferente escolher a combinação de segmentos para a curva principal. No entanto, observando o desempenho obtido do conjunto de teste na figura 7.2, percebe-se que as curvas com maior quantidade de segmentos proporcionam desempenhos elevados. Assim, uma boa sugestão é utilizar curvas mais segmentadas. Nota-se, também, que a sugestão oferecida pelo método de extração (classificador 2) atingiu o pior desempenho de classificação dos eventos do conjunto de teste.

*Tabela VII.1 – Combinação de segmentos no enfoque especialista*

Classificador	Quantidade de Segmentos por Modelo		
	FL	PP	PT
<b>1</b>	1	1	1
<b>2</b>	1	1	2
<b>3</b>	1	2	1
<b>4</b>	1	2	2
<b>5</b>	1	3	1
<b>6</b>	1	3	2
<b>7</b>	1	4	1
<b>8</b>	1	4	2



*Figura 7.1–Desempenho geral obtido do conjunto de desenvolvimento especialista com vários classificadores no espaço de 1024 dimensões, conforme a tabela VII.1.*



*Figura 7.2–Desempenho geral obtido do conjunto de teste especialista com vários classificadores no espaço de 1024 dimensões, conforme a tabela VII.1.*

#### 7.1.1.2 Visão Estatística

Da mesma forma como foi realizado o procedimento na visão especialista, as análises também foram feitas para os conjuntos de desenvolvimento e de teste obtidos com critérios estatísticos. Na extração das curvas do conjunto de desenvolvimento, o número máximo de segmentos em cada curva é 1, 7 e 2 para os padrões FL, PP e PT, respectivamente. O desempenho de classificação dos diversos classificadores, obtidos através da combinação dos segmentos das curvas, é mostrado na figura 7.3 (referente ao conjunto de desenvolvimento) e na figura 7.4 (referente ao conjunto de teste). A tabela VII.2 indica quais as combinações que compõem cada classificador.

Observa-se, na figura 7.3, que o desempenho de classificação obtida com o conjunto de desenvolvimento varia em uma estreita faixa, entre 97% e 98%. Há de se atentar também para a forma de variação do índice de acertos. A variação sistemática entre o máximo e o mínimo se dá pela variação da quantidade de segmentos do padrão PT (1 ou 2 segmentos). Ora tem-se uma curva PT com um segmento, ora tem-se uma curva PT com dois segmentos. Ou seja, apesar da quantidade de combinações possíveis, o índice de acerto não se altera de maneira significativa e varia apenas quando o número

de segmentos de PT é modificado. Observando o desempenho do classificador obtida com o conjunto de teste, percebe-se que o índice varia entre 98% e 100%. Atenta-se ainda que, apesar das várias possibilidades de combinação possíveis, o índice atinge os valores mais elevados para 2 segmentos na curva do padrão PT. Assim, qualquer classificador com esta característica é uma boa sugestão. Mais especificamente, desejando-se classificadores mais compactos, devem-se utilizar aqueles classificadores com menores quantidades de segmentos por curva.

*Tabela VII.2 – Combinação de segmentos no enfoque estatístico*

<b>Classificador</b>	<b>Quantidade de Segmentos por Modelo</b>		
	<b>FL</b>	<b>PP</b>	<b>PT</b>
<b>1</b>	1	1	1
<b>2</b>	1	1	2
<b>3</b>	1	2	1
<b>4</b>	1	2	2
<b>5</b>	1	3	1
<b>6</b>	1	3	2
<b>7</b>	1	4	1
<b>8</b>	1	4	2
<b>9</b>	1	5	1
<b>10</b>	1	5	2
<b>11</b>	1	6	1
<b>12</b>	1	6	2
<b>13</b>	1	7	1
<b>14</b>	1	7	2

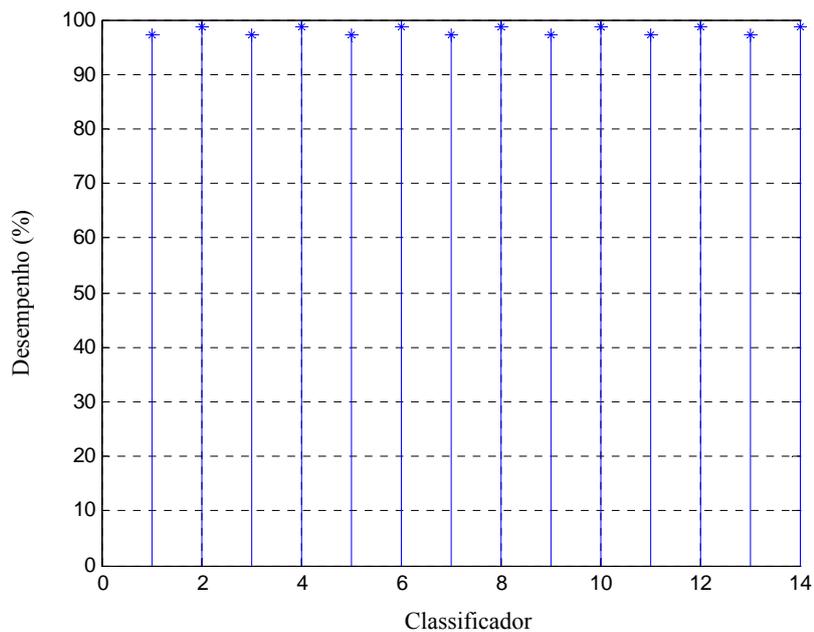


Figura 7.3- Desempenho geral obtido do conjunto de desenvolvimento estatístico com vários classificadores no espaço de 1024 dimensões, conforme a tabela VII.2.

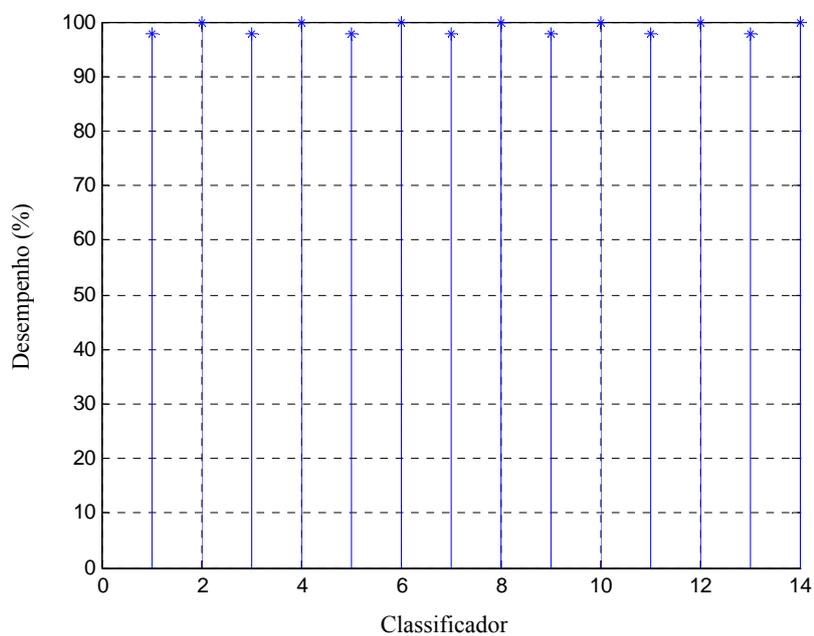


Figura 7.4 –Desempenho geral obtido do conjunto de teste estatístico com vários classificadores no espaço de 1024 dimensões, conforme a tabela VII.2.

### 7.1.2 Extração 3D

Os conjuntos de dados utilizados para a análise em mapas concatenados (1024 dimensões) foram também utilizados nas análises em mapas acumulativos (3D). A grande vantagem de se trabalhar nesta configuração é a possibilidade de se visualizar as curvas principais obtidas dos modelos no espaço tri-dimensional.

Apesar das vantagens de visualização, a informação em 3 dimensões apresenta a desvantagem de aumentar a complexidade de execução do algoritmo, proporcionado pela indexação de cada canal. Para resolver esta dificuldade, compactou-se a matriz de dados 32x32 no formato 16x16 (seção 6.1.2). Isso permitiu que o algoritmo fosse implementado sem ultrapassar os limites práticos de memória e de processamento das plataformas computacionais tipicamente disponíveis.

Para determinar o número de segmentos em cada curva, utilizou-se um critério parecido com aquele utilizado em 1024 dimensões; ou seja, analisa-se o desempenho de classificação obtido dos conjuntos de desenvolvimento e de teste para várias possibilidades de classificadores. No entanto, agora, o desempenho do classificador é avaliado através da informação da média das distâncias (seção 6.2.2). Uma outra diferença para o caso anterior é o fato de que, neste classificador, considera-se a mesma quantidade de segmentos para os três padrões. Isto é devido à grande quantidade de segmentos possíveis em cada curva principal, dado que o algoritmo chegou a atingir mais de cem segmentos por curva. Dessa forma, analisaram-se os dois enfoques utilizados na separação de conjunto de desenvolvimento e de teste (especialista e estatístico).

#### 7.1.2.1 Visão Especialista

A quantidade de classificadores, obtidos a partir da combinação dos segmentos das curvas principais, é relativamente elevada, uma vez que as possibilidades de segmentos por curva aumentaram significativamente para modelos em 3 dimensões. Diferente do caso das curvas em 1024 dimensões, no qual havia poucos eventos em dimensões elevadas; agora, há muitas amostras em dimensões reduzidas. Isso, do ponto de vista do algoritmo, significa uma maior quantidade de informação. Assim, mais segmentos são necessários para compor a curva que represente o respectivo modelo.

O desempenho de classificação, obtida com dados em 3 dimensões, é mostrada nas figuras 7.5 e 7.6, para os conjuntos de desenvolvimento e de teste, respectivamente. Avaliou-se o desempenho até cinqüenta segmentos, a partir do qual o desempenho não se altera mais de maneira significativa e a complexidade computacional começa a tornar-se um problema. Na figura 7.5, tem-se o desempenho obtido do conjunto de desenvolvimento. Nota-se que os melhores resultados ocorrem para classificadores alimentados por curvas compostas de cinco a nove segmentos por padrão. A partir de dez segmentos o desempenho tende a se estabilizar. Nota-se no desempenho de classificação do conjunto de teste que os melhores resultados também ocorrem entre cinco e nove segmentos por curva. Assim, sugere-se que, para o caso de classificação em 3 dimensões, o desempenho do conjunto de desenvolvimento possa ser uma boa sugestão para o número de segmentos que comporá as curvas principais.

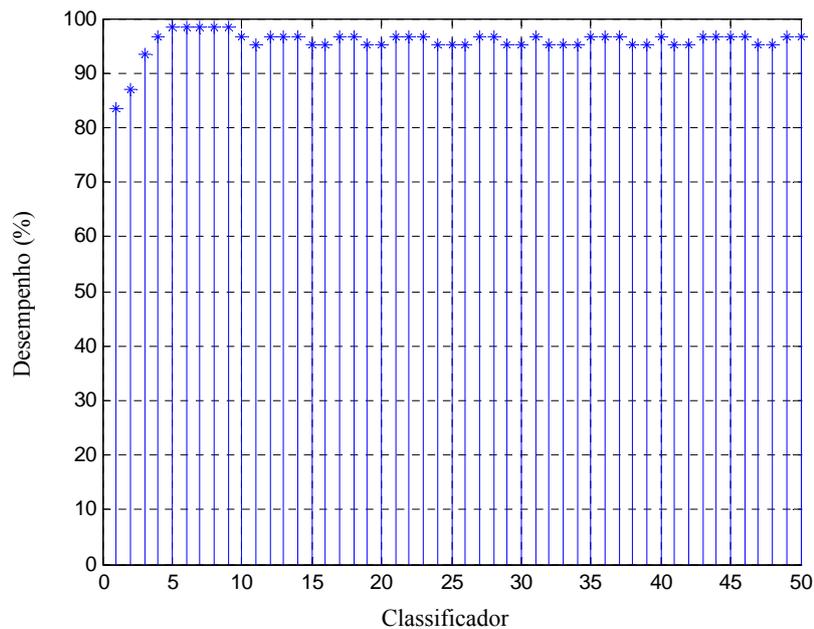


Figura 7.5 - Desempenho geral obtido do conjunto de desenvolvimento especialista com classificadores no espaço 3D, alimentados por curvas com quantidades iguais de segmentos e variando de 1 a 50.

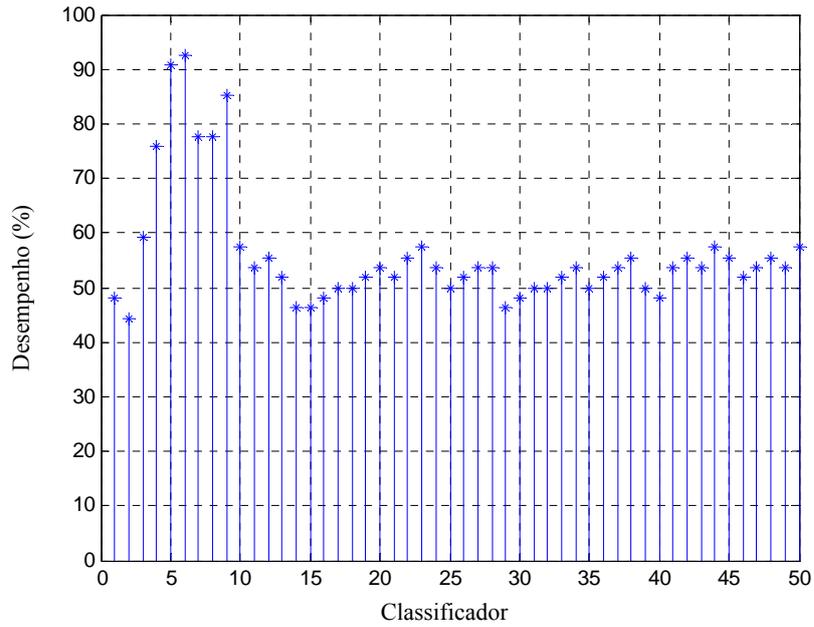
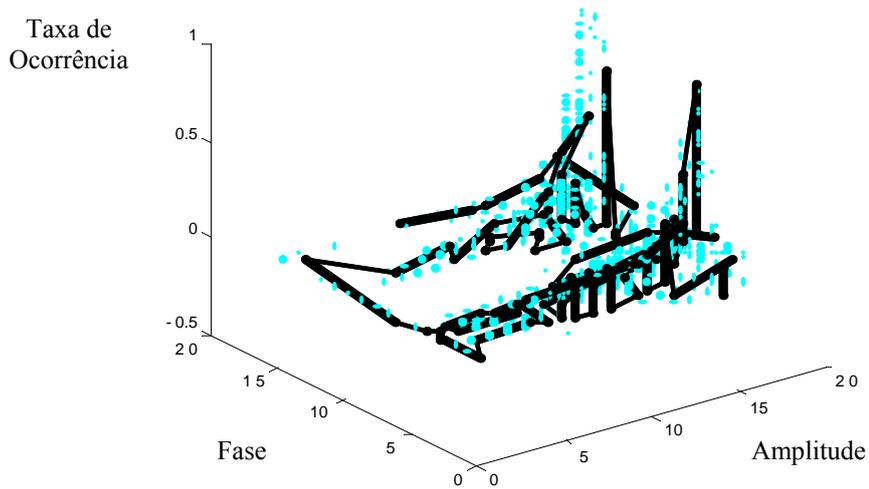
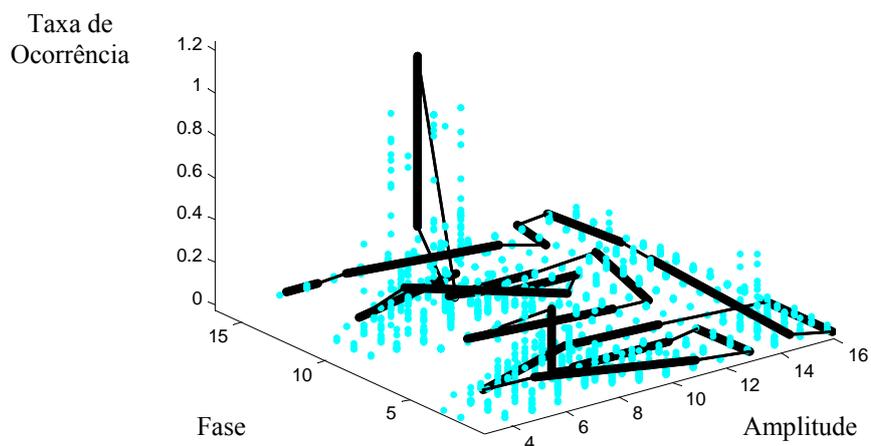


Figura 7.6 – Desempenho geral obtido do conjunto de teste especialista com classificadores no espaço 3D, alimentados por curvas com quantidades iguais de segmentos e variando de 1 a 50.

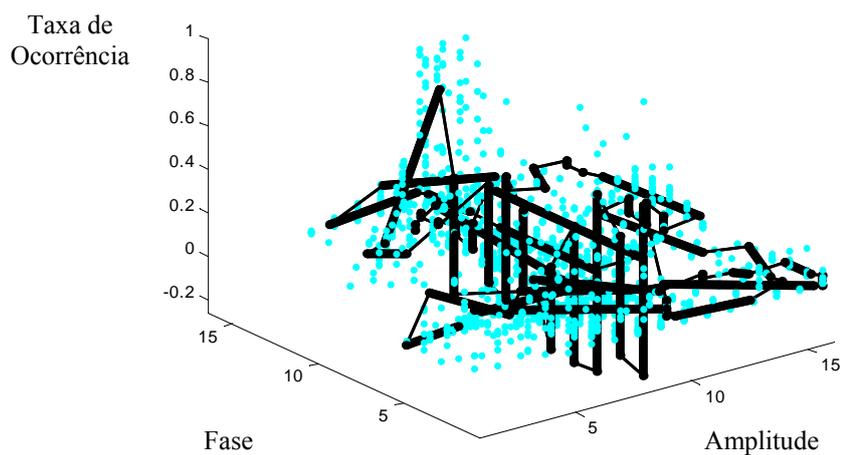
Alguns exemplos de modelos extraídos em 3 dimensões do conjunto de desenvolvimento podem ser visualizados na figura 7.7.



(a)



(b)



(c)

Figura 7.7 – Curva principal correspondente ao padrão (a) Ponta Flutuante (b) Ponta-Plano e (c) Ponta-Terra. Extraídas com o algoritmo K-segmentos.

#### 7.1.2.2 Visão Estatística

As análises também foram realizadas para os classificadores na visão estatística. O desempenho do classificador, obtido do conjunto de desenvolvimento, é mostrado na figura 7.8. Observa-se que o índice atinge os melhores resultados com classificadores alimentados a partir de 6 segmentos por curva principal. Na figura 7.9, observa-se que o

melhor desempenho encontra-se na faixa entre 6 e 12 segmentos. Assim, pela visão estatística em 3 dimensões, um número sugerido de segmentos para compor as curvas principais seria em torno de 6 segmentos, uma vez que, neste valor, atinge-se um bom desempenho de classificação do conjunto de desenvolvimento e um desempenho máximo para o conjunto de teste.

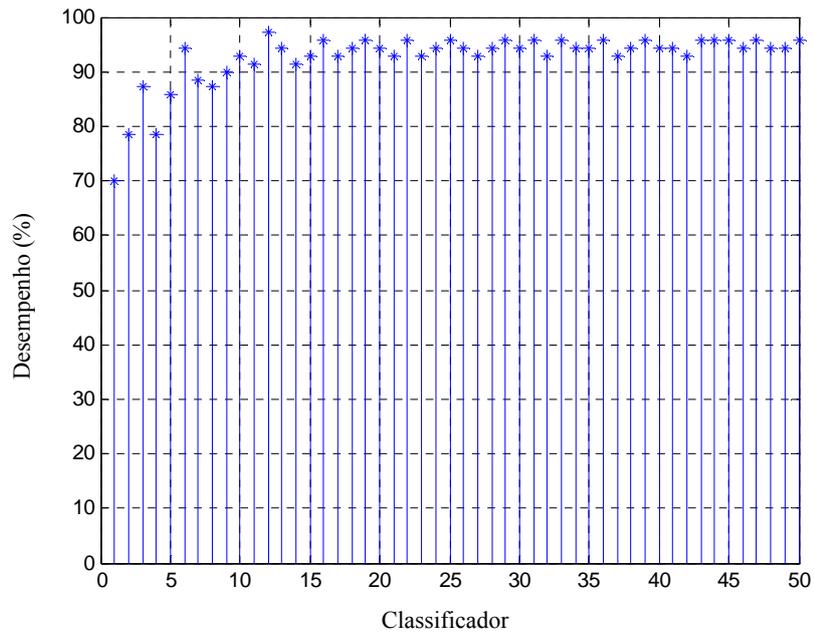


Figura 7.8 – Desempenho geral obtido do conjunto de desenvolvimento estatístico com classificadores no espaço 3D, alimentados por curvas com quantidades iguais de segmentos e variando de 1 a 50.

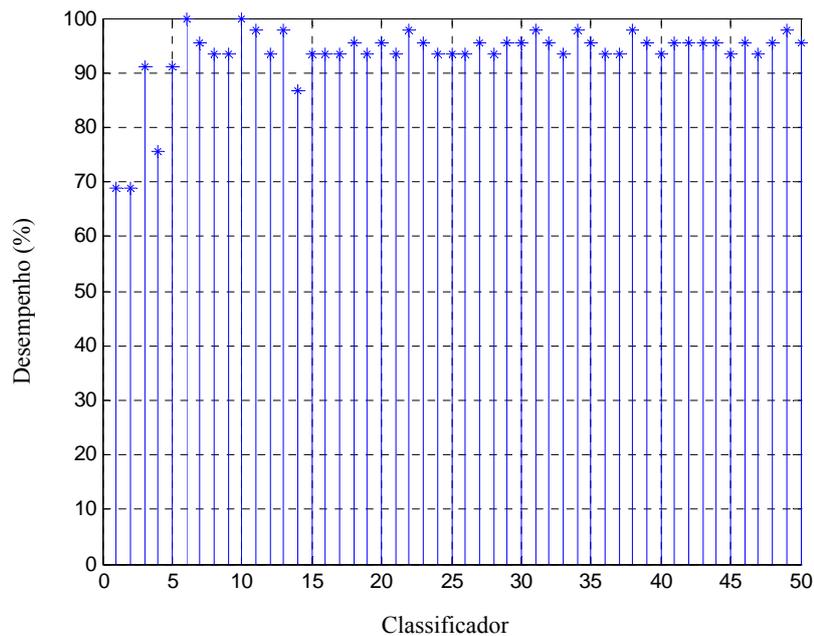


Figura 7.9 – Desempenho geral obtido do conjunto de teste estatístico com classificadores no espaço 3D, alimentados por curvas com quantidades iguais de segmentos e variando de 1 a 50.

## 7.2 Classificação

Ao fim da etapa de extração de características, temos cada padrão (FL, PP ou PT) expressado por uma dada curva principal. A etapa de classificação vai, então, avaliar qual das curvas que melhor representa um dado evento de descarga parcial.

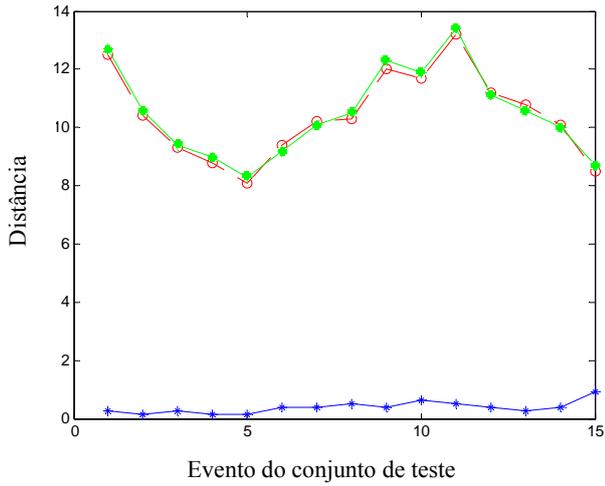
### 7.2.1 Classificação em 1024 Dimensões

Utilizou-se a informação da distância (distâncias euclidianas do evento de teste aos segmentos de uma curva principal) – seção 6.2 - para serem aplicados os métodos da distância mínima (seção 3.2.1.1) e o processamento neural (seção 3.2.3.1). A distância dos eventos do conjunto de teste ao segmento mais próximo foi utilizada pelo método da distância mínima para a classificação. A rede neural utilizou a informação completa das distâncias (distância do evento a todos os segmentos das três curvas), para treinar a rede e classificar os eventos.

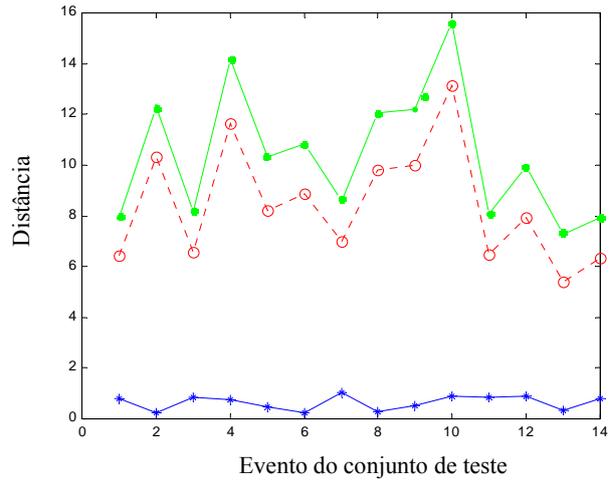
Para o caso dos modelos de DPs em 1024 dimensões, na visão especialista, utilizou-se o classificador relativo à oitava combinação de segmentos (ver tabela VII.1); ou seja, 1, 4 e 2 segmentos para as curvas extraídas para os padrões FL, PP e PT, respectivamente. Para o caso da visão estatística, utilizou-se o classificador relativo à segunda combinação de segmentos (ver tabela VII.2); ou seja, 1, 2 e 1 segmentos para as curvas extraídas para os padrões FL, PP e PT.

#### 7.2.1.1 Distância Euclidiana Mínima

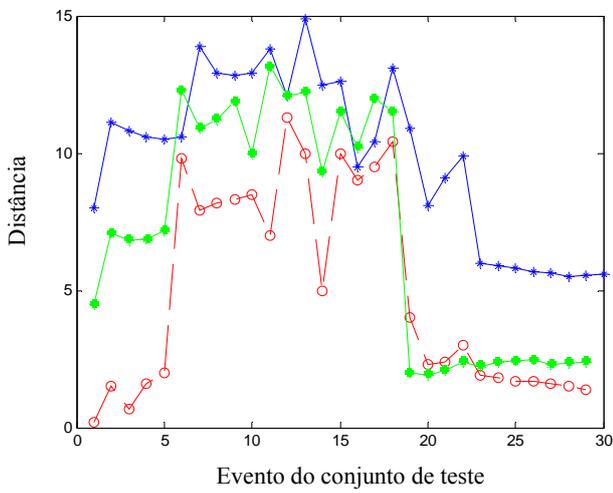
Utilizando-se a distância euclidiana mínima (seção 3.2.1.1) para classificar os eventos de teste, a eficiência do classificador foi bastante elevada. O classificador classificou corretamente 50 dos 54 eventos pertencentes ao conjunto de teste especialista. Já os 45 eventos do conjunto estatístico de teste, foram todos corretamente classificados. Os resultados são mostrados na figura 7.10. Os gráficos do lado esquerdo representam as distâncias aos padrões FL, PP e PT, na visão especialista. Os gráficos do lado direito representam as distâncias aos padrões na visão estatística. Observa-se que o padrão FL diferencia-se bem dos outros dois padrões. Já os padrões PP e PT estão mais próximos um do outro.



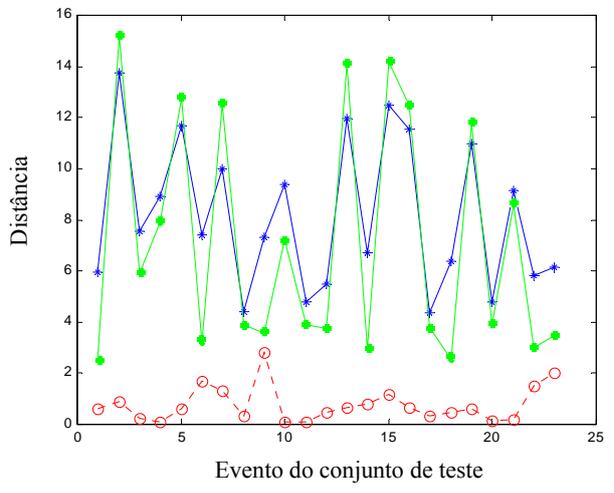
(a)



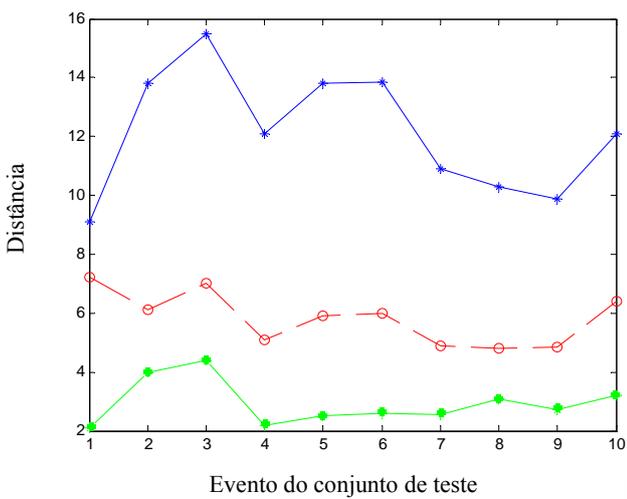
(d)



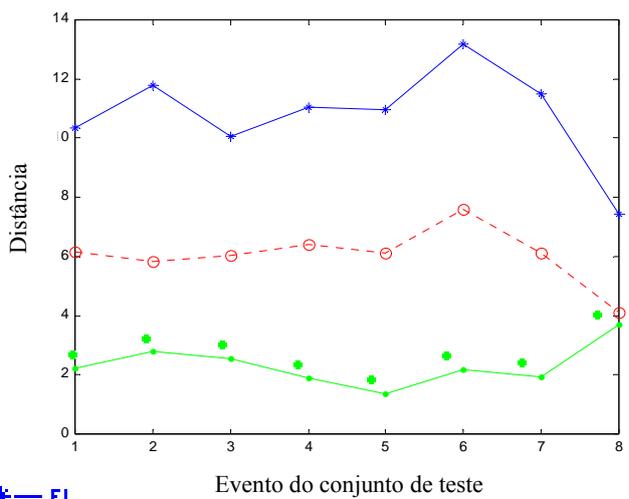
(b)



(e)



(c)



(f)



Figura 7.10 – Distâncias mínimas dos eventos do conjunto de teste às curvas principais do padrão (a) e (d) FL, (b) e (e) PP e (c) e (f) PT, para as separações especialista e estatística, respectivamente.

As tabelas VII.3 e VII.4 apresentam os desempenhos de classificação dos padrões FL, PP e PT, relativos às visões especialista e estatística. Nota-se que, do ponto de vista especialista, 13% dos eventos de teste do padrão PP confundiram-se com o padrão PT. O desempenho geral do classificador especialista foi de 92%.

Já o classificador estatístico não confundiu nenhum padrão, atingindo um desempenho de 100%. Na realidade, devido à incerteza referente à quantidade de eventos, pode se afirmar que, para o classificador estatístico, o desempenho geral é superior a 98%.

*Tabela VII.3 – Desempenho no enfoque especialista utilizando o classificador de distância mínima no espaço de 1024 dimensões*

<b>Enfoque Especialista</b>			
<b>Desempenho</b>	<b>FL</b>	<b>PP</b>	<b>PT</b>
<b>FL</b>	100%	0%	0%
<b>PP</b>	0%	87%	0%
<b>PT</b>	0%	13%	100%
<b>Geral</b>	<b>92%</b>		

*Tabela VII.4 – Desempenho no enfoque estatístico utilizando o classificador de distância mínima no espaço de 1024 dimensões*

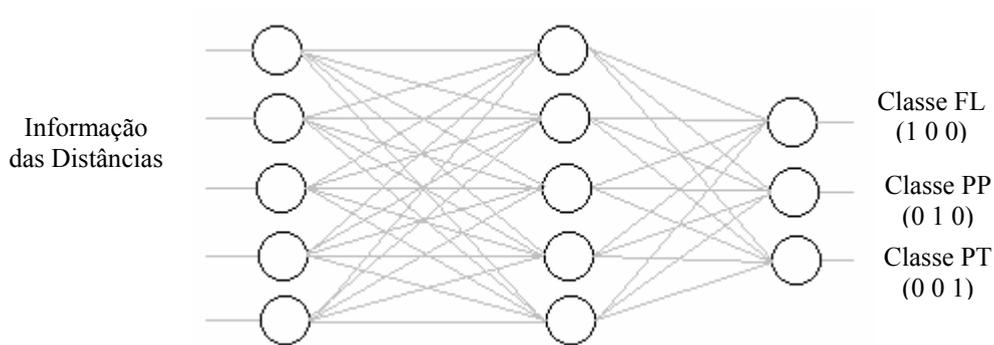
<b>Enfoque Estatístico</b>			
<b>Desempenho</b>	<b>FL</b>	<b>PP</b>	<b>PT</b>
<b>FL</b>	100%	0%	0%
<b>PP</b>	0%	100%	0%
<b>PT</b>	0%	0%	100%
<b>Geral</b>	<b>100%</b>		

#### 7.2.1.2 Rede Neural

No caso da aplicação com redes neurais, o treinamento foi realizado com as informações das distâncias do evento a todos os segmentos das curvas (ver explicações na seção 6.2.1.1). Estas informações compõem-se de 54 matrizes no formato 1024x11 e 45 matrizes no formato 1024x5, para o treinamento dos classificadores neurais especialista e estatístico. Assim, configuraram-se, respectivamente, redes neurais com

11 e 5 nós na entrada, 11 e 5 neurônios na camada intermediária e 3 neurônios na camada de saída.

Com esta configuração, classificaram-se corretamente 53 dos 54 eventos do conjunto de teste especialista. Já os 45 eventos do conjunto de teste estatístico foram todos corretamente classificados. A figura 7.11 mostra a configuração da rede neural para o classificador estatístico.



*Figura 7.11 – Configuração da rede neural para a classificação dos mapas em 1024 dimensões, na visão estatística.*

Os resultados obtidos com a rede neural são mostrados na tabela VII.5. Os valores mostrados nesta tabela representam os valores de saída dos neurônios com precisão de duas casas decimais. O neurônio vencedor (relativo a FL, PP ou PT) é aquele que possui o maior valor de saída.

Tabela VII.5 – Saídas da rede para classificadores neurais em 1024 dimensões

Evento	Conjunto Teste Especialista			Conjunto Teste Estatístico		
	FL	PP	PT	FL	PP	PT
1	<b>0,97</b>	0,05	-0,01	<b>1,00</b>	0,07	0,01
2	<b>1,00</b>	-0,05	0,03	<b>1,00</b>	0,03	0,03
3	<b>0,97</b>	0,04	0,04	<b>1,00</b>	0,05	0,04
4	<b>1,00</b>	0,00	-0,02	<b>1,00</b>	0,02	0,02
5	<b>1,00</b>	0,02	0,02	<b>1,00</b>	0,02	0,02
6	<b>1,00</b>	0,00	0,00	<b>1,00</b>	0,03	0,00
7	<b>0,97</b>	0,05	0,05	<b>1,00</b>	0,05	0,05
8	<b>1,00</b>	0,01	0,01	<b>1,00</b>	0,01	0,01
9	<b>1,00</b>	-0,14	0,24	<b>1,00</b>	0,24	0,24
10	<b>1,00</b>	0,04	0,02	<b>1,00</b>	0,02	0,02
11	<b>0,96</b>	0,06	0,06	<b>1,00</b>	0,01	0,06
12	<b>1,00</b>	-0,02	-0,01	<b>0,98</b>	-0,03	-0,01
13	<b>0,98</b>	0,05	0,04	<b>0,99</b>	0,04	0,04
14	<b>0,97</b>	0,04	0,05	<b>0,97</b>	0,05	0,05
15	<b>1,00</b>	0,02	0,01	0,02	<b>0,96</b>	0,01
16	0,03	<b>0,97</b>	0,06	0,01	<b>1,00</b>	0,01
17	0,04	<b>0,99</b>	-0,01	0,09	<b>0,97</b>	0,03
18	0,02	<b>1,00</b>	0,04	0,05	<b>0,99</b>	0,04
19	0,02	<b>0,97</b>	0,05	0,02	<b>1,00</b>	0,02
20	0,00	<b>0,99</b>	0,01	0,07	<b>0,97</b>	0,02
21	0,05	<b>0,98</b>	0,04	0,03	<b>0,99</b>	0,00
22	0,01	<b>0,93</b>	0,05	0,04	<b>0,98</b>	0,05
23	0,24	<b>1,00</b>	0,24	0,06	<b>0,93</b>	0,01
24	0,02	<b>1,00</b>	0,02	0,12	<b>0,81</b>	0,65
25	0,06	<b>0,99</b>	0,01	0,13	<b>0,74</b>	0,62
26	-0,01	<b>1,00</b>	0,05	0,08	<b>0,71</b>	0,13
27	0,04	<b>0,97</b>	0,01	0,05	<b>0,73</b>	0,21
28	0,05	<b>1,00</b>	0,24	-0,05	<b>0,97</b>	0,19
29	0,01	<b>0,95</b>	0,02	0,02	<b>1,00</b>	0,04
30	0,01	<b>0,91</b>	0,06	0,07	<b>0,95</b>	0,01
31	0,03	<b>1,00</b>	0,05	0,03	<b>0,91</b>	-0,01
32	0,04	<b>0,97</b>	0,01	0,08	<b>1,00</b>	0,04
33	0,02	<b>0,99</b>	0,24	0,13	<b>0,97</b>	0,05
34	0,02	0,22	<b>0,76</b>	0,07	<b>0,99</b>	0,01
35	0,00	<b>0,92</b>	0,06	0,03	<b>0,96</b>	0,01
36	0,05	<b>0,51</b>	0,42	0,07	<b>0,92</b>	0,03
37	0,01	<b>0,65</b>	0,52	0,13	<b>1,00</b>	0,04
38	0,24	<b>0,53</b>	0,44	0,08	0,02	<b>0,88</b>
39	0,02	<b>0,72</b>	0,39	0,13	0,02	<b>0,84</b>
40	0,06	<b>0,89</b>	0,05	0,08	0,00	<b>0,81</b>
41	-0,01	<b>0,99</b>	0,01	-0,02	0,05	<b>0,92</b>
42	0,04	<b>0,91</b>	0,24	0,06	0,01	<b>0,95</b>
43	0,05	<b>0,97</b>	0,02	0,09	0,05	<b>0,94</b>
44	0,01	<b>0,98</b>	0,06	0,05	0,01	<b>0,97</b>
45	0,08	-0,01	<b>0,95</b>	0,23	0,46	<b>0,74</b>
46	0,02	0,01	<b>0,94</b>	-	-	-
47	0,06	0,24	<b>1,00</b>	-	-	-
48	0,09	0,02	<b>0,99</b>	-	-	-
49	0,06	0,02	<b>0,88</b>	-	-	-
50	0,09	0,06	<b>0,89</b>	-	-	-
51	0,13	0,05	<b>0,93</b>	-	-	-
52	0,06	0,01	<b>0,96</b>	-	-	-
53	0,12	0,14	<b>0,79</b>	-	-	-
54	0,13	0,56	<b>0,74</b>	-	-	-

As tabelas VII.6 e VII.7 apresentam o desempenho de classificação dos padrões FL, PP e PT, relativos às visões especialista e estatística. Nota-se que, do ponto de vista especialista, para o padrão PP, houve um acréscimo no desempenho, em relação ao método da distância mínima, atingindo-se 96% de acertos deste padrão. Isto levou o desempenho geral deste classificador a atingir 98%.

Já o classificador estatístico, através das redes neurais, permaneceu inalterado em seu desempenho máximo.

*Tabela VII.6 – Desempenho no enfoque especialista utilizando o classificador neural no espaço de 1024 dimensões*

<b>Enfoque Especialista</b>			
<b>Desempenho</b>	<b>FL</b>	<b>PP</b>	<b>PT</b>
<b>FL</b>	100%	0%	0%
<b>PP</b>	0%	96%	0%
<b>PT</b>	0%	7%	100%
<b>Geral</b>	<b>98%</b>		

*Tabela VII.7 – Desempenho no enfoque estatístico utilizando o classificador neural para o espaço de 1024 dimensões*

<b>Enfoque Estatístico</b>			
<b>Desempenho</b>	<b>FL</b>	<b>PP</b>	<b>PT</b>
<b>FL</b>	100%	0%	0%
<b>PP</b>	0%	100%	0%
<b>PT</b>	0%	0%	100%
<b>Geral</b>	<b>100%</b>		

### **7.2.2 Classificação 3D**

No caso da classificação em mapas acumulativos, utilizou-se o método de classificação baseado na votação, além dos dois métodos utilizados em mapas de 1024 dimensões (distância mínima e processamento neural). Neste contexto, avaliaram-se os classificadores especialista e estatístico. A quantidade de segmentos escolhida para compor cada curva principal dos classificadores baseou-se nas análises realizadas em

7.1. Utilizaram-se ambos os classificadores alimentados pelas curvas dos padrões FL, PP e PT, com seis segmentos cada curva.

#### 7.2.2.1 Distância Euclidiana Mínima

A figura 7.12 exibe a média das distâncias das amostras dos eventos de teste às respectivas curvas principais e suas barras de erro (ver explicações na seção 6.2.2.1). Os gráficos do lado esquerdo mostram os resultados referentes ao classificador especialista. Nota-se que o padrão FL distinguiu-se bem dos outros dois padrões. No entanto, os padrões PP e PT estão bem próximos entre si. Observando-se as flutuações em relação à média, percebe-se a pouca precisão do método 3D, quando se utiliza um classificador baseado na distância mínima. Um comportamento parecido também é observado nos resultados do classificador estatístico (lado direito da figura 7.12). No entanto, os valores das distâncias distinguem-se um pouco melhor, comparado com o caso especialista.

O classificador especialista classificou corretamente 50 dos 54 eventos de teste. Já o classificador estatístico classificou corretamente todos os 45 eventos do seu respectivo conjunto de teste. Apesar do desempenho geral dos classificadores ser igual o desempenho obtido no espaço de 1024 dimensões para a distância mínima, os desvios em relação à média indicam grandes flutuações das distâncias entre as amostras do evento e os segmentos da curva.

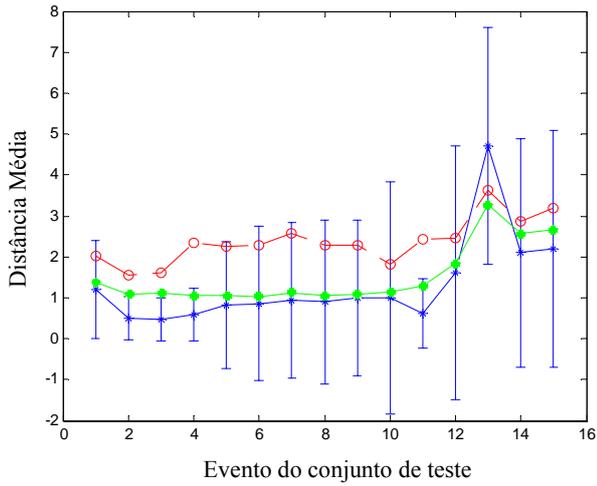
As tabelas VII.8 e VII.9 apresentam o desempenho de classificação dos padrões FL, PP e PT relativos às visões especialista e estatística, bem como o desempenho geral.

*Tabela VII.8 – Desempenho no enfoque especialista utilizando classificador de distância mínima no espaço 3D*

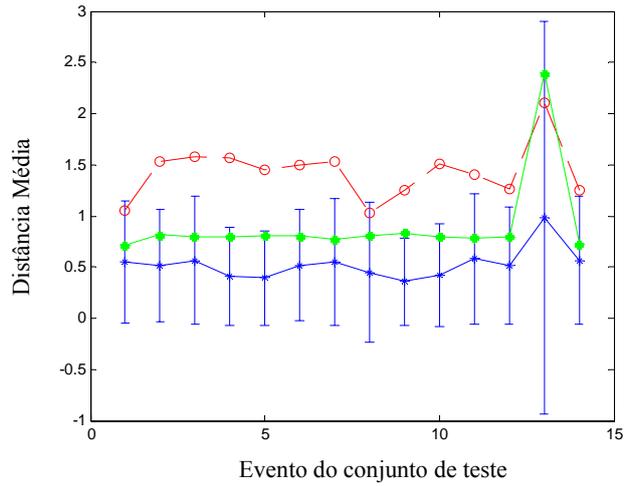
<b>Enfoque Especialista</b>			
<b>Desempenho</b>	<b>FL</b>	<b>PP</b>	<b>PT</b>
<b>FL</b>	93%	0%	0%
<b>PP</b>	0%	93%	10%
<b>PT</b>	7%	7%	90%
<b>Geral</b>	<b>92%</b>		

*Tabela VII.9 – Desempenho no enfoque estatístico utilizando o classificador de distância mínima no espaço 3D*

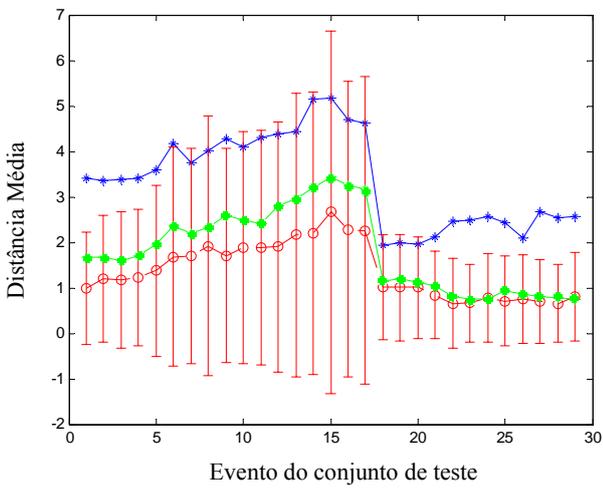
<b>Enfoque Estatístico</b>			
<b>Desempenho</b>	<b>FL</b>	<b>PP</b>	<b>PT</b>
<b>FL</b>	100%	0%	0%
<b>PP</b>	0%	100%	0%
<b>PT</b>	0%	0%	100%
<b>Geral</b>	<b>100%</b>		



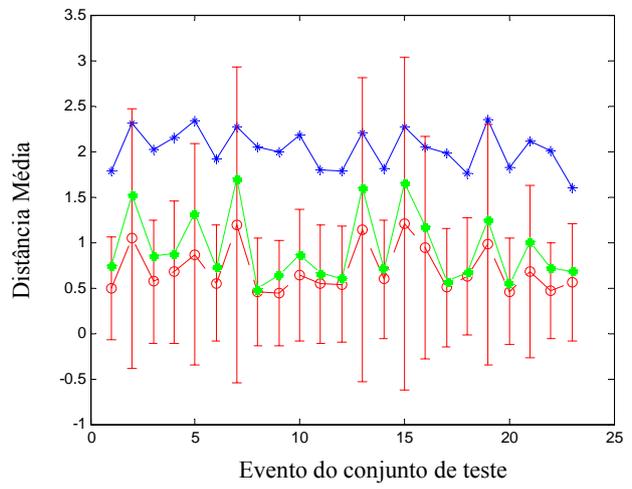
(a)



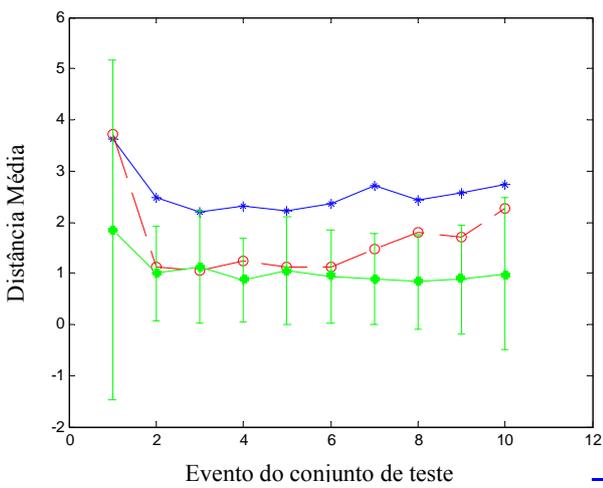
(d)



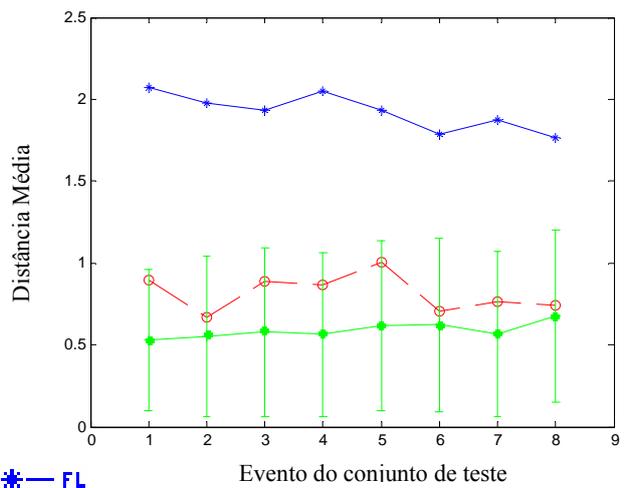
(b)



(e)



(c)



(f)

\* FL  
○ PP  
■ PT

Figura 7.12 – Média das distâncias das amostras dos eventos de teste aos padrões (a) e (d) FL, (b) e (e) PP e (c) e (f) PT, nos enfoque especialista e estatístico, respectivamente.

### 7.2.2.2 Rede Neural

Uma rede neural, treinada por *backpropagation*, também foi utilizada em mapas indexados (veja seção 6.2.2.2). O conjunto de treinamento é apresentado na entrada da rede no formato 54x33 e 45x33, para o treinamento dos classificadores especialista e estatístico, respectivamente.

Assim, para ambos os classificadores, a rede ficou configurada com 33 nós de entrada, 33 neurônios na camada intermediária e 3 neurônios na camada de saída.

Com estas configurações, classificaram-se, corretamente, 52 dos 54 eventos do conjunto de teste especialista. Já os 45 eventos do conjunto de teste estatístico foram classificados corretamente. O modelo da rede é mostrado na figura 7.13. Os resultados obtidos com esta rede neural são mostrados na tabela VII.10. O neurônio vencedor (relativo a FL, PP ou PT) é aquele que possui o maior valor de saída.

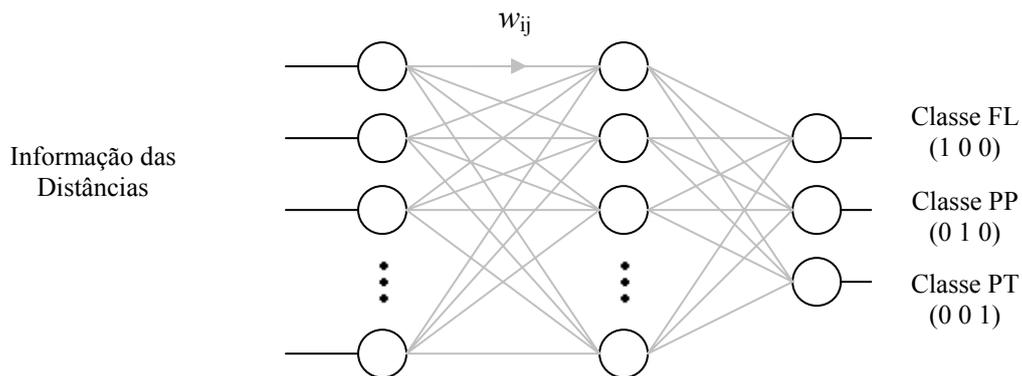


Figura 7.13 – Configuração da rede neural para a classificação dos mapas em 3 Dimensões

Tabela VII.10 – Saídas da rede para classificadores neurais em 3 dimensões

Evento	Conjunto Teste Especialista			Conjunto Teste Estatístico		
	FL	PP	PT	FL	PP	PT
1	<b>0.84</b>	0.16	0.21	<b>0.94</b>	-0.06	0.11
2	<b>0.93</b>	0.10	0.01	<b>0.98</b>	0.03	-0.03
3	<b>0.89</b>	0.23	0.32	<b>0.96</b>	-0.02	0.02
4	<b>0.93</b>	-0.25	0.26	<b>0.97</b>	-0.01	-0.03
5	<b>0.89</b>	-0.07	0.23	<b>0.97</b>	0.07	-0.03
6	<b>0.80</b>	-0.09	0.22	<b>0.96</b>	-0.01	0.00
7	<b>0.75</b>	0.03	0.25	<b>0.95</b>	-0.01	0.06
8	<b>0.43</b>	0.53	0.12	<b>0.99</b>	-0.04	0.08
9	<b>0.88</b>	-0.02	0.13	<b>0.93</b>	0.04	-0.01
10	<b>0.70</b>	0.33	0.02	<b>0.99</b>	-0.03	0.00
11	<b>0.69</b>	0.38	0.41	<b>0.99</b>	0.00	-0.04
12	<b>0.99</b>	0.22	0.12	<b>0.98</b>	0.07	0.01
13	<b>0.76</b>	0.37	0.51	<b>0.96</b>	0.05	-0.05
14	<b>0.98</b>	0.17	0.35	<b>0.98</b>	-0.00	-0.00
15	<b>0.98</b>	0.46	0.52	-0.00	<b>0.96</b>	-0.22
16	-0.01	<b>0.92</b>	-0.04	-0.08	<b>0.73</b>	0.61
17	-0.02	<b>0.88</b>	0.21	-0.07	<b>0.93</b>	0.15
18	-0.01	<b>0.93</b>	0.02	-0.00	<b>0.97</b>	-0.03
19	-0.05	<b>0.89</b>	0.27	0.10	<b>0.96</b>	-0.06
20	-0.03	<b>0.80</b>	0.51	0.03	<b>0.65</b>	0.17
21	-0.08	<b>0.73</b>	0.65	0.11	<b>0.86</b>	0.42
22	0.06	<b>0.77</b>	0.51	0.01	<b>0.92</b>	-0.04
23	-0.05	<b>0.83</b>	0.42	0.05	<b>0.90</b>	-0.06
24	-0.11	<b>0.86</b>	0.50	0.10	<b>0.88</b>	0.21
25	-0.05	<b>0.80</b>	0.60	0.04	<b>0.90</b>	-0.02
26	-0.07	<b>0.89</b>	0.40	-0.01	<b>0.87</b>	0.05
27	-0.05	<b>0.82</b>	0.51	0.02	<b>0.86</b>	0.64
28	-0.05	<b>0.83</b>	0.48	-0.04	<b>0.90</b>	-0.09
29	-0.09	<b>0.83</b>	0.59	-0.00	<b>0.84</b>	0.55
30	-0.04	<b>0.79</b>	0.64	-0.06	<b>0.95</b>	-0.07
31	-0.10	<b>0.74</b>	0.61	0.06	<b>0.82</b>	0.11
32	-0.09	<b>0.81</b>	0.60	-0.13	<b>0.93</b>	0.05
33	0.12	0.47	<b>0.60</b>	0.03	<b>0.67</b>	0.47
34	0.26	<b>0.49</b>	0.21	-0.01	<b>0.87</b>	0.24
35	0.00	<b>0.60</b>	0.48	0.02	<b>0.82</b>	0.06
36	0.01	<b>0.46</b>	0.43	0.03	<b>0.88</b>	0.03
37	-0.03	<b>0.62</b>	0.32	0.00	<b>0.88</b>	0.15
38	-0.02	<b>0.61</b>	0.44	-0.02	0.11	<b>0.84</b>
39	-0.03	0.54	<b>0.70</b>	0.01	0.63	<b>0.75</b>
40	-0.00	<b>0.49</b>	0.41	0.00	0.12	<b>0.79</b>
41	0.53	<b>0.56</b>	0.33	0.02	0.27	<b>0.89</b>
42	-0.10	<b>0.62</b>	0.46	0.00	0.31	<b>0.87</b>
43	-0.03	<b>0.75</b>	0.32	-0.05	0.17	<b>0.86</b>
44	-0.05	<b>0.71</b>	0.57	-0.00	0.29	<b>0.67</b>
45	0.00	0.34	<b>0.76</b>	0.04	0.45	<b>0.91</b>
46	-0.11	0.56	<b>0.71</b>			
47	-0.11	0.43	<b>0.54</b>			
48	0.02	0.48	<b>0.83</b>			
49	0.06	0.41	<b>0.58</b>			
50	0.04	0.19	<b>0.86</b>			
51	-0.10	0.65	<b>0.74</b>			
52	0.08	0.10	<b>0.67</b>			
53	-0.02	0.18	<b>0.77</b>			
54	0.16	0.11	<b>0.84</b>			

As tabelas VII.11 e VII.12 apresentam o desempenho de classificação dos padrões FL, PP e PT relativos às visões especialista e estatística. Percebe-se uma ligeira melhora do desempenho geral do classificador no enfoque especialista, em comparação ao método da distância mínima no espaço tri-dimensional. Já o desempenho do classificador no enfoque estatístico permaneceu inalterado em relação à distância mínima.

*Tabela VII.11 – Desempenho no enfoque especialista utilizando classificador neural no espaço 3D*

<b>Enfoque Especialista</b>			
<b>Desempenho</b>	<b>FL</b>	<b>PP</b>	<b>PT</b>
<b>FL</b>	100%	0%	0%
<b>PP</b>	0%	93%	0%
<b>PT</b>	0%	7%	100%
<b>Geral</b>	<b>96%</b>		

*Tabela VII.12 – Desempenho no enfoque estatístico utilizando o classificador neural no espaço 3D*

<b>Enfoque Estatístico</b>			
<b>Desempenho</b>	<b>FL</b>	<b>PP</b>	<b>PT</b>
<b>FL</b>	100%	0%	0%
<b>PP</b>	0%	100%	0%
<b>PT</b>	0%	0%	100%
<b>Geral</b>	<b>100%</b>		

### 7.2.2.3 Votação

Em dados no espaço 3D, utilizou-se também o critério de votação (seção 3.2.1.3). A tabela VII.13 mostra os resultados obtidos com o método. O padrão vencedor é aquele que recebe o maior número de votos. Cada voto está relacionado a uma amostra de um evento no espaço tri-dimensional. Por exemplo, o evento 1 do conjunto especialista tem um total de 76 amostras. Este valor corresponde às amostras não nulas do evento 1, que por sua vez tem 256 elementos. Assim, o padrão FL recebeu 29 votos e os padrões PP e PT receberam 24 e 23 votos, respectivamente. Portanto, o evento 1 do padrão FL foi corretamente escolhido, segundo a maioria dos votos. O

mesmo procedimento é repetido para todos os eventos pertencentes aos conjuntos de teste (ver mais detalhes na seção 6.2.2.3).

Da tabela VII.13, nota-se que o desempenho do classificador atingiu também bons níveis de classificação. O padrão atribuído (FL, PP ou PT) ao modelo é aquele com maior quantidade de votos. Observa-se que classificador especialista classificou corretamente 51 dos 54 eventos de teste e o classificador estatístico acertou todos os eventos.

Tabela VII.13 – Resultado do método de votação no espaço 3D

Evento	Conjunto Teste Especialista			Conjunto Teste Estatístico		
	FL	PP	PT	FL	PP	PT
1	<b>29</b>	23	25	<b>26</b>	21	22
2	<b>38</b>	24	28	<b>31</b>	15	21
3	<b>39</b>	25	26	<b>36</b>	19	22
4	<b>40</b>	27	34	<b>41</b>	21	32
5	<b>42</b>	32	36	<b>48</b>	20	22
6	<b>43</b>	30	37	<b>46</b>	25	26
7	<b>40</b>	34	35	<b>44</b>	24	35
8	<b>41</b>	33	37	<b>39</b>	30	37
9	<b>40</b>	30	36	<b>37</b>	23	25
10	<b>35</b>	27	31	<b>36</b>	25	27
11	<b>39</b>	26	31	<b>45</b>	23	31
12	<b>29</b>	24	25	<b>39</b>	21	29
13	27	28	<b>34</b>	<b>37</b>	25	30
14	<b>26</b>	20	24	<b>36</b>	24	24
15	<b>31</b>	25	26	23	<b>38</b>	25
16	26	<b>53</b>	35	26	<b>40</b>	35
17	27	<b>57</b>	38	27	<b>43</b>	38
18	24	<b>54</b>	30	24	<b>40</b>	30
19	30	<b>59</b>	34	30	<b>49</b>	34
20	26	<b>52</b>	42	26	<b>43</b>	32
21	28	<b>51</b>	37	28	<b>43</b>	39
22	26	<b>58</b>	35	26	<b>39</b>	32
23	27	<b>49</b>	37	27	<b>37</b>	36
24	24	<b>55</b>	30	24	<b>42</b>	30
25	30	<b>53</b>	29	30	<b>46</b>	29
26	26	<b>69</b>	43	26	<b>45</b>	39
27	28	<b>55</b>	30	28	<b>47</b>	34
28	28	<b>53</b>	33	28	<b>50</b>	33
29	23	<b>41</b>	23	23	<b>44</b>	33
30	25	<b>60</b>	32	25	<b>39</b>	44
31	25	<b>54</b>	31	25	<b>41</b>	45
32	23	<b>49</b>	43	18	<b>43</b>	48
33	20	<b>58</b>	51	20	<b>46</b>	39
34	24	<b>56</b>	44	24	<b>41</b>	44
35	32	<b>58</b>	52	32	<b>47</b>	31
36	30	<b>54</b>	40	30	<b>52</b>	47
37	35	<b>57</b>	35	35	<b>49</b>	37
38	37	<b>46</b>	36	27	37	<b>41</b>
39	29	30	<b>34</b>	20	33	<b>37</b>
40	19	<b>41</b>	30	19	38	<b>44</b>
41	21	<b>46</b>	33	23	36	<b>41</b>
42	20	<b>42</b>	35	20	26	<b>39</b>
43	33	<b>38</b>	33	28	31	<b>35</b>
44	23	36	<b>39</b>	23	27	<b>34</b>
45	18	30	<b>46</b>	18	34	<b>39</b>
46	23	42	<b>48</b>			
47	27	39	<b>42</b>			
48	29	34	<b>35</b>			
49	30	32	<b>40</b>			
50	32	35	<b>39</b>			
51	26	31	<b>47</b>			
52	28	23	<b>38</b>			
53	21	33	<b>51</b>			
54	23	31	<b>52</b>			

As tabelas VII.14 e VII.15 apresentam o desempenho de classificação dos padrões FL, PP e PT relativos às visões especialista e estatística.

*Tabela VII.14 – Desempenho no enfoque especialista utilizando o método da votação no espaço 3D*

<b>Enfoque Especialista</b>			
<b>Desempenho</b>	<b>FL</b>	<b>PP</b>	<b>PT</b>
<b>FL</b>	93%	0%	0%
<b>PP</b>	0%	93%	0%
<b>PT</b>	7%	7%	100%
<b>Geral</b>	<b>94%</b>		

*Tabela VII.15 – Desempenho no enfoque estatístico utilizando o método da votação no espaço 3D*

<b>Enfoque Estatístico</b>			
<b>Desempenho</b>	<b>FL</b>	<b>PP</b>	<b>PT</b>
<b>FL</b>	100%	0%	0%
<b>PP</b>	0%	100%	0%
<b>PT</b>	0%	0%	100%
<b>Geral</b>	<b>100%</b>		

### 7.3 Quadro Comparativo

As tabelas VII.16 e VII.17 mostram um quadro comparativo entre os enfoques especialista e estatístico nos espaços de 1024 e de 3 dimensões. Os desempenhos de classificação são apresentados lado a lado. Para o caso de classificadores de 1024 dimensões, no enfoque estatístico, o que pode se afirmar é que o desempenho máximo foi superior a 98%, dado que se analisaram apenas 45 eventos. Isto implica em uma resolução estatística que não permite afirmar desempenho superior a este. Já no enfoque especialista, com 54 eventos para teste, os melhores resultados corresponderam ao classificador neural, atingindo um desempenho geral de 98%. Os resultados para o classificador no espaço 3D também atingiram desempenhos elevados. No entanto, esta configuração introduz um desvio (barra de erro) elevado em relação à média, mostrando que estes classificadores não são robustos.

Tabela VII.16 – Quadro comparativo entre os classificadores que operam no espaço de 1024 dimensões

<b>Formato</b>	<b>Método</b>	<b>Enfoque</b>	<b>Desempenho Geral</b>
Mapas 1024	Distância Mínima	Especialista	92%
		Estatístico	>98%
	Redes Neurais BP	Especialista	98%
		Estatístico	>98%

Tabela VII.17 – Quadro comparativo entre classificadores que operam no espaço tri-dimensional

<b>Formato</b>	<b>Método</b>	<b>Enfoque</b>	<b>Desempenho Geral</b>
Mapas 3 dimensões	Distância Mínima	Especialista	92%
		Estatístico	>98%
	Redes Neurais BP	Especialista	96%
		Estatístico	>98%
	Votação	Especialista	94%
		Estatístico	>98%

## 7.5 Conclusão

As curvas principais foram utilizadas no reconhecimento de padrões em descargas parciais. No enfoque especialista, modelos mais segmentados se mostraram mais eficazes na classificação. No enfoque estatístico, modelos com menos segmentos, sugeridos pelo método de extração, obtêm bons resultados. A eficiência total da classificação atingiu índices elevados, principalmente em mapas de 1024 dimensões, produzindo classificadores eficientes, robustos e simplificados. Já os classificadores configurados para atuar no espaço 3D não se mostraram robustos. A utilidade deste tipo de configuração é a possibilidade de visualização dos modelos extraídos.

# Capítulo 8

## Análises e Conclusões

Este trabalho buscou dar uma contribuição na análise de sinais de descargas parciais. O método das curvas principais (CPs) foi aplicado na extração de características de mapas de DPs e diferentes procedimentos de classificação foram propostos. Avaliaram-se também critérios de separação da base de dados experimental, em termos dos conjuntos de desenvolvimento e de teste, por meio especialista e estatístico. Além disso, apresentou-se um método para a prospecção de dados de descargas parciais, permitindo a visualização dos modelos de DPs.

O método das curvas principais, aplicado na configuração de mapas concatenados (1024 dimensões), apresentou como características principais: simplicidade durante a extração das curvas, robustez do classificador e eficiência do desempenho de classificação. Nesta configuração, os melhores desempenhos obtidos com o conjunto de teste foram: 98% no enfoque especialista e superior a 98% no enfoque estatístico.

O método das CPs aplicado em mapas acumulativos (3D) também apresentou bons resultados, chegando a classificar corretamente mais de 98% dos eventos do conjunto de teste. Apesar deste desempenho, o grau de incerteza produzido pela média é elevado, mostrando que este tipo de classificador não é robusto. Além disso, o método 3D eleva a complexidade computacional no processo de extração de características. A grande vantagem deste tipo de configuração é a visualização dos modelos extraídos no espaço tri-dimensional.

Durante a implementação dos classificadores anteriores, avaliou-se a forma de seleção de eventos para os conjuntos de desenvolvimento e de teste. Diante das duas visões analisadas (especialista e estatística), os desempenhos de classificação foram obtidos para os respectivos conjuntos. Notou-se que, no enfoque estatístico, todos os eventos do conjunto de teste foram classificados corretamente, tanto para o método da distância mínima quanto para os outros métodos utilizados (RN e votação). O que corrobora os resultados obtidos em Diniz [5] através do método não linear (NLPCA). Já

no enfoque especialista, os melhores resultados foram atingidos através do classificador neural (classificando corretamente 98% dos eventos do conjunto de teste). Além disso, utilizando o classificador de distância Euclidiana mínima, o desempenho do método chegou a 92%, corroborando os resultados obtidos em Cuenca [6].

Os resultados obtidos no presente trabalho dão um indicativo de que a diferença nos desempenhos de classificação, utilizando métodos lineares e não lineares, não é devido à presença de estatísticas de ordem superior nos dados de DPs. O principal motivo para esta discrepância está relacionado aos enfoques de separação de conjuntos utilizados em cada um dos trabalhos. O fato de obterem-se desempenhos superiores no enfoque estatístico já era esperado, uma vez que se buscam eventos, para compor o conjunto de teste, que melhor generalizem o modelo. Já os eventos do conjunto de teste no enfoque especialista são mais ruidosos e menos representativos dos padrões, característica típica de uma situação real em campo. O que levou a uma perda natural na eficiência de teste do método linear (PCA), operando sob este último enfoque.

Analisando-se a configuração de dados em 1024 dimensões, cujos resultados são mais consistentes, a extração das curvas - no enfoque especialista - mostrou que os modelos mais segmentados apresentam melhores desempenhos de classificação. Uma explicação para este fato é a composição do conjunto de desenvolvimento, o qual tem eventos mais característicos dos sinais de DPs. Assim, quanto mais segmentos, mais a curva principal se aproxima do modelo real sugerido por especialistas. Já no enfoque estatístico, os resultados mostraram que a quantidade de segmentos não afeta o desempenho significativamente, ou seja, curvas mais segmentadas não representam, necessariamente, o melhor o modelo. Uma vez que se procuram por classificadores eficientes e, ao mesmo tempo, simplificados (para processamento em tempo real), modelos menos segmentados são preferíveis. Uma boa escolha da quantidade de segmentos, no enfoque estatístico, é a sugestão dada pelo método de extração de curvas principais (seção 4.3.1).

Concluindo, o classificador baseado em CPs mostrou-se simples, robusto e eficaz na classificação de sinais de descargas parciais, tanto no enfoque especialista quanto no enfoque estatístico. Adicionalmente, o classificador oferece a possibilidade de visualização dos modelos extraídos, o que é útil na prospecção dos dados.

## 8.1 Perspectivas Futuras

O reconhecimento de padrões em descargas parciais é um assunto que está longe de ser exaurido. Diversos outros métodos podem ainda ser aplicados em classificadores de DPs. Da mesma forma, as curvas principais também são uma fonte de estudo na qual ainda há muito por explorar, principalmente em relação às descargas parciais. Algoritmos alternativos de extração de curvas podem ser utilizados na modelagem de padrões de DPs. A versão suave do algoritmo K-segmentos [8] é um exemplo. O método busca curvas que são compostas por segmentos interconectados por *splines*.

Outro trabalho que pode ser desenvolvido é a aplicação de outros métodos de classificação, abordados no capítulo 3. A técnica de ICA é interessante quanto à remoção de ruídos, podendo ser explorada em conjunto com as curvas principais. Além disso, ao invés de caracterizar a distância através da média e do valor RMS, métodos utilizando estatística de ordem superior podem ser desenvolvidos no intuito de encontrar classificadores mais robustos para a configuração 3D.

O projeto de filtros analógicos de descargas parciais, para atuar em condições ruidosas de campo, está em fase de implementação. Para operarem em série com tais filtros, em operações de tempo real, classificadores compactos, baseados em curvas principais, podem ser implementados em processadores de sinais digitais (eventualmente DSPs).

# Referências

- [1] DANIKAS, M. G., “The Definitions Used for Partial Discharge Phenomena”, IEEE Transactions on Electrical Insulation, v. 28, n. 6, pp. 1075-1081, 1993.
- [2] SAHOO, N. C., SALAMA, M. A., BARTINIKAS, R., “Trends in Partial Discharge Pattern Classification: A Survey”, *IEEE Transactions on Dielectrics and Electrical Insulation*, v.12, n. 2, pp. 248-264, 2005.
- [3] MEDEIROS, M., C., TIMO, T., REACH, C.,”Building Neural Network Model for Time Series: a Statistical Approach”, SSE/EFI Working Paper in Economics and Finance, n. 508, 2002.
- [4] HASTIE, T. J., STUETZLE, W., “Principal Curves”, Journal of the American Statistical Association, v. 84, n. 406, pp. 502-516, 1989.
- [5] DINIZ, F. C. C. B., *Supressão de Ruído, Detecção e Classificação de Sinais de Descargas Parciais em Transformadores de Potência*. Tese de Mestrado, Programa de Engenharia Elétrica, COPPE/UFRJ, Rio de Janeiro, RJ, 2005.
- [6] HUÁMAN CUENCA, W., *Aplicação de Sistemas Inteligentes no Reconhecimento de Padrões de DPs em Transformadores de Potência*. Tese de Mestrado, Programa de Engenharia Elétrica, COPPE/UFRJ, Rio de Janeiro, RJ, 1998.
- [7] FERNANDES, H. L., *Classificação de Navios Baseado em Curvas Principais*, Tese de Mestrado, Programa de Engenharia Elétrica, COPPE/UFRJ, Rio de Janeiro, RJ, 2005.
- [8] VERBEEK, J. J., VLASSIS, N. , KRÖSE, B., A K-segments Algorithm for Finding Principal Curves, Report, Intelligent Autonomous System Technical Report Series, nr. IAS-UVA-00-11, 2000.
- [9] HUÁMAN CUENCA, W., *Caracterização dos Sinais de Descargas Parciais em Equipamentos de Alta Tensão a Partir dos Modelos Experimentais*. Tese de Doutorado, Programa de Engenharia Elétrica, COPPE/UFRJ, Rio de Janeiro, RJ, 2005.
- [10] IEC Publication-60270, *High Voltage Test Technique-Partial Discharge Measurement*, Third Edition, 2000.
- [11] IEEE P1434, *Guide to the Measurement of Partial Discharges in Rotating Machinery*, Electrical Machinery Committee of the IEEE Power Society, 1997.
- [12] LATINI, M., et. al., “Medição de Descargas Parciais em Hidrogerador da Usina de Itaipu – Uma Avaliação Experimental das Limitações, Dificuldades de Medição e Interpretação de resultados”. In: XV SNPTEE, Foz do Iguaçu – Paraná, 1999.

- [13] LEVY, A., et. al., “Avaliação do Efeito de Propagação de Memória no Desempenho de Materiais Elétricos – Medição de DP em Ultra-Alta Freqüência ao Ar Ambiente e Óleo Isolante”, Relatório Técnico ACET 095/97, Fev. 1997.
- [14] JAMES, R. E., PHUNG, B. T., “Development of Computer-based Measurements and their Application to PD Pattern Analysis”, IEEE Transactions on Electrical Insulation, v. 2, pp. 838-856, 1995
- [15] CACCIARI, M., CONTIN, A., RABACH, G., MONANARI, C. G., “An Approach to Partial Discharge Investigation by Height Distribution Analysis”, IEEE Proc.: Sci., Meas. Techol., v.42, pp. 102-108, 1995.
- [16] MAZROUA, A. A., BARTINIKAS, R., SALAMA, M. M., “Discrimination Between PD Pulse Shapes Using Different Neural Network Paradigms”, IEEE Transactions on Electrical Insulation, v. 1, n. 6, pp. 1119-1131, 1994.
- [17] SALAMA, M. M., BARTINIKAS, R., “Determination of Neural Network Topology for Partial Discharge Pulse Pattern Recognition”, IEEE Transactions Neural Networks, v. 13, pp. 446-456, 2002.
- [18] CUENCA, W. M., SEIXAS, J. M., LEVY, A. F., “Análise de Componentes Principais para Identificar Descargas Parciais em Transformadores de Potência”, In VI Brazilian Symposium on Neural Networks, Aplicações Cuenca46, PUC, Rio de Janeiro Nov. 2001.
- [19] HIVÄRINEN, A., KARHUNEN, J. & OJA, E., *Independent Component Analysis*, John Wiley & Sons. Inc. 2001.
- [20] SHANMUGAN, K. S., AM Breipoh, *Random Signal Detection, Estimate and Data Analysis*, John Wiley Sons. 1988.
- [21] HOOFF, M., FREISLEBEN, B. PATSH, R., “PD Source Identification with Novel Discharge Parameters Using Counterpropagation Neural Networks”, IEEE Transactions Neural Networks, v. 4, pp. 17-32, 1997.
- [22] HUCKER, T., KRANZ, G. H., “Requirements of Automated PD Diagnosis System for Fault Identification in Noisy Conditions”, IEEE Transactions on Electrical Insulation, v. 2, pp. 544-556, 1995.
- [23] CHUI, C. K., *An Introduction to Wavelets*, Academic Press, Inc., 1992.
- [24] JANG, J. K., KIM, S. H. LEE, Y. S., “Classification of Partial Discharge Electrical Signals Using Wavelet Transforms”, IEEE 13<sup>th</sup> Intern. Conf. Dielectr. Liquids, Nara, Japan, pp. 552-555, 1999.
- [25] CARMINATI, E., CRISTALDI, L. LAZZARONI, M., MONTI, A., “A Neuro-fuzzy Approach for the Detection of Partial Discharge”, IEEE Trans. Instrum. Meas, v. 50, pp. 1413-1417, 2001.

- [26] RAHMAN, M. K. A., ARORA, R. SRIVASTAVA, S. C., “Partial Discharge Classification Using Principal Component Transformation”, IEEE Proc., Part A, v. 147, pp. 7-13, 2000.
- [27] SATISH, L., ZAENGL, W. S., “Can Fractal Features be Used for Recognizing 3-D Partial Discharge Patterns?”, IEEE Transactions on Electrical Insulation, v. 2, pp. 352-359, 1995.
- [28] LALITHA, E. M., SATISH, L., “Wavelet Analysis for Classification of Multi-source PD Patterns”, IEEE Transactions on Electrical Insulation, v. 7, pp. 40-47, 2000.
- [29] HARALIK, R. M., SHAMUGAN, K., DINSTEN, I., “Textural Features for Image Classification”, IEEE Trans. Syst., Man and Cybernetics, SMC-3, pp. 610-621, 1973.
- [30] CONNERS, R. W., HARLOW, C. A., “A Theoretical Comparison of Texture Algorithms”, IEEE Trans. Pattern Analysis and Machine Intelligence, PAMI-2, pp. 204-222, 1980.
- [31] GALLOWAY, M. M., “Texture Analysis Using Gray Level Run Lengths”, Computer Vision, Graphics and Image Processing, v. 4, pp. 172-179, 1975.
- [32] PRATT, W. K., *Digital Signal Processing*, Wiley, 1978
- [33] LALITHA, E. M., SATISH, L., “Fractal Image Compression for Classification of PD Sources”, IEEE Transactions on Electrical Insulation, v. 5, pp. 550-557, 1998.
- [34] YOUNG, R. K., *Wavelet Theory and Applications*, Kluwer Academic Publisher, 1993.
- [35] TOU, J. T., GONZALES, R. C., *Pattern Recognition Principles*, Addison-Wesley, 1974.
- [36] MAZROUA, A. A., SALAMA, M. M. A., BARTINIKAS, R., “PD Patterns Recognition with Neural Networks Using the Multilayer Perceptron Technique”, IEEE Transactions on Electrical Insulation, v. 28, pp. 1082-1089, 1993.
- [37] HAYKIN, S., *Neural Network: A Comprehensive Foundation*, 2da. Edition, ISBN 0-02-352761-7 Prentice Hall, 1999.
- [38] WANG, M. -H., “Partial Discharge Pattern Recognition of Current Transformers Using an ENN”, Transactions on Power Delivery, v. 20, pp. 1984-1990- 2005.
- [39] CUENCA, W. M., SEIXAS, J. M., LEVY, A. F., “Análise Sistema Caracterizador Híbrido para Identificar Descargas Parciais em Transformadores de Potência”, Proceedings of the V Brazilian Conference on Neural Networks, pp. 169-174, PUC, Rio de Janeiro – RJ, Brazil, April 2-5, 2001.
- [40] GULSKI, E., KRIVDA, A., “Neural Networks as a Tool for Recognition of Partial Discharges”, IEEE Transactions on Electrical Insulation, v. 28, pp. 984-1001, 1993.

- [41] HECHT-NIELSEN, R., "Application of Counterpropagation Networks", *Neural Networks*, v. 1, pp. 131-139, 1988
- [42] Kohonen, T., "The Self-organizing Map", *IEEE Proc.*, v. 78, pp. 1464-1480, 1990.
- [43] HONG, T., FANG, M. T. C., HILDER, D., "PD Classification by a Modular Neural Network Based on Task Decomposition", *IEEE Transactions on Electrical Insulation*, v. 3, pp. 207-212, 1996.
- [44] COTIN, A., CAVALLINI, A., MONTANARI, G. C., PASINI, G., "Digital Detection and Fuzzy Classification of Partial Discharge Signals", *IEEE Transactions on Electrical Insulation*, v. 9, pp. 335-348, 1993.
- [45] KREUGER, F. H., GUSKI, E., KRIVDA, A., "Classification of Partial Discharges", *IEEE Transactions on Electrical Insulation*, v. 28, pp. 917-931, 1993.
- [46] KRANZ, H. –G., "Diagnosis of Partial Discharge Signals Using Neural Networks and Minimum Distance Classification", *IEEE Transactions on Electrical Insulation*, v. 28, pp. 1016-1024, 1993.
- [47] GULSKI, E. DREUGER, F. H., "Computer-aided Recognition of Discharge Sources", *IEEE Transactions on Electrical Insulation*, v. 27, pp. 82-92, 1992.
- [48] ZIOMEK, W. REFORMAT, M., KUFFEL, E., "Application of Genetic Algorithms to Pattern Recognition of Defects in GIS", *IEEE Transactions on Electrical Insulation*, v. 7, pp. 161-168, 2000.
- [49] RAHMAN, M. K. A., ARORA, R., SRIVASTAVA, S. C., "Partial Discharge Classification Using Principal Component Transformation", *IEEE Proc. Part A*, V. 147, pp. 7-13, 2000.
- [50] DUCHAMP, T., STUETZLE, W., "Geometric Properties of Principal Curves in the plane", *Robust Statistics, Data Analysis and Computer Intensive Methods: in Honor of Petr Huber's 60th Birthday*, v. 109, pp. 135-152, 1996.
- [51] CHANG, K., GHOSH, J., "Probabilistic Principal Surfaces", *Proceeding of the International Joint Conference on Neural Networks*, p. 605, 1999.
- [52] CHANG, K., GHOSH, J., "A Unified Model for Probabilistic Principal Surfaces", *Transactions on the Pattern Analysis and Machine Intelligence*, v. 23, pp. 22-41, 2001.
- [53] KÉGL, B., KRZYŻAK, A., LINDER, T., et al., "Learning and Design of Principal Curves", *Transactions on the Pattern Analysis and Machine Intelligence*, v. 22, pp. 281-297, 2000.
- [54] DELICADO, P., "Another Look at Principal Curves and Surfaces", *Journal of Multivariate Analysis*, v. 77, pp. 84-116, 2001.
- [55] DUDA, R., HART, P. and STORK, D., *Pattern Classification*, 2nd. ed. 1997.

- [56] EINBECK, J., TUTZ, G., EVERS, L., Local Principal Curves, Report, SFB Discussion Report Number 320, 2003.
- [57] DELICADO, P., HUERTA, M., “Principal Curves of Oriented Points: Theoretical and Computational Improvements”, *Computational Statistics*, v. 18, n.2, pp. 293-315, 2003.
- [58] BANFIELD, J., RAFTERY, A., “Ice Flow Identification in Satellite Images Using Mathematical Morphology and Clustering About Principal Curves”, *Journal of the American Statistical Association*, v. 87, n. 417, pp. 7-16, 1992.
- [59] TIBSHIRANI, R., “Principal Curves Revisited”, *Statistics and Computation*, v. 2, pp. 183-190, 1992.
- [60] TARPEY, T., LI, L., FLURY, B. D., “Principal Points and Self-consistent Points of Elliptical Distributions”, *Annals of Statistics*, v. 23, n. 6, pp. 103-112, 1995.
- [61] DEWANG, C., ZANG, J., TANG, SHUMING, T. WANG, J., “Freeway Traffic Stream Modeling Based on Principal Curves and Its Analysis”, *IEEE Transactions on Intelligent Transportation Systems*, v. 5. n.4, 2004.
- [62] MURPHY, P. M., “The u.c. Irvine repository of Machine Learning Databases and Domain Theories”, <http://www.ics.uci.edu/mllearn/MLRepository.html>.
- [63] STONE, M., “Asymptotic for and Against Cross-Validation”, *Biometrika*, v. 64, pp. 29-35, 1977.
- [64] ICM: *Digital Partial Discharge Measurement System Operation Manual vs. 2.20*.

# Apêndice

## Produção Científica

Neste apêndice, apresenta-se o artigo produzido como resultado do presente trabalho de pesquisa, submetido ao Simpósio Brasileiro de Sistemas Elétricos, Campina Grande – PB, julho de 2006.

Título: Curvas Principais Aplicadas na Identificação de Descargas Parciais em Equipamentos de Potência.

Autores: José Márcio Faier, José Manuel de Seixas, Alain Levy e Helena Lopez Fernandez.

Resumo: O artigo discute a aplicação do método das curvas principais (CPs) na identificação de padrões em equipamentos de potência. A técnica de processamento de sinais apresentada consegue uma eficiente extração de características de dados de descargas parciais (DPs), permitindo classificadores simples e compactos. Baseado na assinatura das curvas principais, a eficiência de classificação foi superior a 98% para as três classes de DPs disponíveis. Como resultado adicional, o método proporcionou a visualização dos modelos de CPs no espaço tri-dimensional.