

UNIVERSIDADE FEDERAL DO MARANHÃO  
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA  
CURSO DE CIÊNCIA DA COMPUTAÇÃO

**CLASSIFICAÇÃO DE FALHAS GEOLÓGICAS EM DADOS SÍSMICOS  
USANDO CORRELOGRAMA E APRENDIZADO DE MÁQUINA**

Antônio Gomes da Silva Júnior

São Luís  
2013

Antônio Gomes da Silva Júnior

**CLASSIFICAÇÃO DE FALHAS GEOLÓGICAS EM DADOS SÍSMICOS  
USANDO CORRELOGRAMA E APRENDIZADO DE MÁQUINA**

Monografia apresentada ao Curso de Ciência da Computação da Universidade Federal do Maranhão, como parte dos requisitos necessários para a obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Anselmo Cardoso Paiva

São Luís  
2013

Silva Júnior, Antônio Gomes

Classificação de falhas geológicas em dados sísmicos usando correlograma e aprendizado de máquina / Antônio Gomes da Silva Júnior. – 2013.

50f.

Orientador: Anselmo Cardoso Paiva.

Monografia (Graduação em Ciência da Computação) – Universidade Federal do Maranhão, 2013.

1.Aprendizado de máquina 2.SVM 3.Falhas sísmicas 4.Atributos sísmicos 5.Geoestatística 6.Correlograma I. Título.

CDU 550.832.4

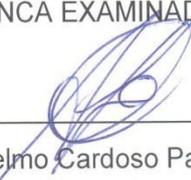
Antônio Gomes da Silva Júnior

**CLASSIFICAÇÃO DE FALHAS GEOLÓGICAS EM DADOS SÍSMICOS  
USANDO CORRELOGRAMA E APRENDIZADO DE MÁQUINA**

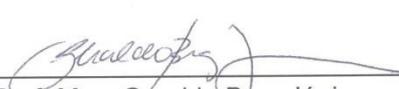
Monografia apresentada ao Curso de Ciência da Computação da Universidade Federal do Maranhão, como parte dos requisitos necessários para a obtenção do grau de Bacharel em Ciência da Computação.

Aprovado em 18 de fevereiro de 2013

BANCA EXAMINADORA

  
\_\_\_\_\_  
Prof. Dr. Anselmo Cardoso Paiva (Orientador)  
Universidade Federal do Maranhão

  
\_\_\_\_\_  
Prof. Dr. Aristófanés Córrea Silva  
Universidade Federal do Maranhão

  
\_\_\_\_\_  
Prof. Msc. Geraldo Braz Júnior  
Universidade Federal do Maranhão

## **AGRADECIMENTOS**

A Deus, o que seria de mim sem a fé que eu tenho nele.

À minha família, por sua capacidade de acreditar em mim e investir em mim. Mãe, seu cuidado e dedicação foram o que deram, em alguns momentos, a esperança para seguir. Pai, sua presença significou segurança e certeza de que não estou sozinho nessa caminhada.

À Nicélia, pessoa com quem amo partilhar a vida. Com você tenho me sentido mais vivo de verdade. Obrigado pelo carinho, a paciência e por sua capacidade de me trazer paz nos momentos difíceis.

Ao professor Dr. Anselmo Cardoso Paiva pela paciência na orientação e incentivo que tornaram possível a conclusão desta monografia.

A todos os professores do curso, que foram tão importantes na minha vida acadêmica, em especial, ao professor Aristóфанes Corrêa, Geraldo Braz, Simara Rocha e ao colega João Dallyson.

Ao PRH-39, o Programa de Recursos Humanos da Agência Nacional do Petróleo, Gás Natural e Biocombustíveis – ANP pelo incentivo.

Aos meus amigos, pelas alegrias, tristezas e dores compartilhadas. Com vocês, as pausas entre um parágrafo e outro de produção melhora tudo o que tenho produzido na vida, em especial, a Márcio Frazão, Afonso Lopes, Jefferson Sá, Nadson Timbó, Pablo Fernandes, Wagner Júnior e Marcos Corrêa.

A todos aqueles que de alguma forma estiveram e estão próximos de mim, fazendo esta vida valer cada vez mais a pena.

“O homem não é nada além daquilo que a educação faz dele.”  
(Immanuel Kant)

## RESUMO

A sísmica de reflexão, em relação a outras técnicas de mapeamento do subsolo, tem como principais vantagens a não necessidade de perfuração de poços e o levantamento de dados de áreas de grande extensão. A busca por uma representação adequada dos reservatórios de petróleo através de dados sísmicos e o fato desse processo gerar uma grande quantidade de informações sobre as condições geológicas da área antes de se iniciar a perfuração de um poço pode evitar prejuízos e maximizar o lucro, além de reduzir o impacto ambiental com perfurações desnecessárias. O presente trabalho tem como objetivo desenvolver um método que auxilie o intérprete na classificação de possíveis falhas em dados sísmicos utilizando a função geoestatística Correlograma como descritor de características e técnicas de aprendizado de máquina. A interpretação sísmica visa identificar estruturas onde o óleo possa estar aprisionado em um reservatório. O método apresentado obteve resultados de sensibilidade igual a 96,8%, especificidade a 93,63%, e acurácia igual a 96,6%.

Palavras-chave: Aprendizado de máquina, SVM, falhas sísmicas, atributos sísmicos, geoestatística, correlograma.

## **ABSTRACT**

The seismic reflection, relative to other subsurface mapping techniques, has as its main advantages is no need for drilling and data collection areas to great extent. The search for an adequate representation of petroleum reservoirs using seismic data and the fact that process generate a lot of information on geological conditions of the area before they start drilling a well can avoid losses and maximize profit and reduce environmental impact with perforations unnecessary. This study aims to develop a method to assist the interpreter in the classification of possible faults in seismic data using geostatistics correlogram function as a descriptor of features and machine learning techniques. The seismic interpretation is to identify structures where oil may be trapped in a reservoir. The method presented results obtained sensitivity equal to 96.8%, specificity equal to 93.63% and accuracy equal to 96.6%.

Keywords: Machine learning, classification, SVM, seismic faults, seismic attributes, geostatistics, correlogram.

## LISTA DE FIGURAS

Figura 2.1: Trapa Estrutural, anticlinal. (TEIXEIRA et al., 2000) .....	15
Figura 2.2: Trapa Estrutural, falhas. (TEIXEIRA et al., 2000) .....	15
Figura 2.3: Trapa Estratigráfica, discordâncias. (TEIXEIRA et al., 2000) .....	15
Figura 2.4: Geometria de aquisição (TECNO PETRO BR, 2008). .....	16
Figura 2.5: Exploração (óleo e gás) baseada em sísmica (ROBINSON, TREITEL, 1980)....	18
Figura 2.6: Caminhão vibrador (PGS, 2010).....	19
Figura 2.7: Cabos sísmicos flutuantes e canhões de ar (PGS, 2010). .....	19
Figura 2.8: Aquisição Sísmica (GERHARDT, 1998).....	20
Figura 2.9: Dados sísmicos. Traço sísmico (a). Linha sísmica (b). Volume sísmico (c) (MACHADO, 2008). .....	21
Figura 2.10: Exemplos de falhas. Bom selante (a). Mal selante (b) (MACHADO, 2008). .....	22
Figura 2.11: Atributos e as configurações dos parâmetros da Rede Neural (TINGDAHL, ROOIJ, 2005). .....	24
Figura 2.12: Dados sintéticos com falhas e os atributos calculados ao longo da linha branca (TINGDAHL, ROOIJ, 2005). .....	24
Figura 2.13: Amostras correlacionadas (a) e amostras não-correlacionadas (b) (ALMEIDA, 2010). .....	27
Figura 2.14: Hiperplano ótimo, margem de separação e vetores-suporte (ALMEIDA, 2007).30	
Figura 3.1: Volume sísmico. ....	35
Figura 3.2: Fatia do volume sísmico. ....	36
Figura 3.3: Cálculo da Energia Vertical (ALMEIDA, 2007). .....	37
Figura 3.4: Energia Vertical. ....	37
Figura 3.5: Cálculo da Similaridade Vertical (ALMEIDA, 2007).....	38
Figura 3.6: Similaridade Vertical. ....	38
Figura 3.7: Distâncias e direções dos lags (CAMARGO, 1998). .....	39

## LISTA DE TABELAS

Tabela 2.1: Parâmetros de aquisição. ....	17
Tabela 2.2: Paradigmas de Aprendizado de Máquina. ....	29
Tabela 2.3: Principais Kernels utilizados nas SVMs. Adaptado de (HAYKIN, 1999). ....	32
Tabela 2.4: Equações dos métodos estatísticos utilizados pra validação. ....	34
Tabela 4.1: Resultados da acurácia para cada atributo.....	41
Tabela 4.2: Resultados da acurácia para associação dos atributos.....	42
Tabela 4.3: Média e desvio padrão para os melhores valores de acurácia.....	43
Tabela 4.4: Sensibilidade, especificidade e acurácia dos melhores resultados. ....	43

# SUMÁRIO

1	INTRODUÇÃO .....	12
2	FUNDAMENTAÇÃO TEÓRICA.....	14
2.1	Dados Sísmicos .....	14
2.1.1	Sísmica de Reflexão .....	16
2.1.2	Aquisição .....	18
2.1.3	Processamento.....	20
2.1.4	Interpretação.....	21
2.2	Atributos Sísmicos .....	23
2.2.1	Energia.....	25
2.2.2	Energia Vertical .....	25
2.2.3	Similaridade.....	25
2.2.4	Correlograma.....	26
2.3	Máquina de Vetores de Suporte .....	28
2.4	Métrica de Desempenho.....	33
3	METODOLOGIA PROPOSTA .....	35
3.1	Aquisição, preparação e tratamento das imagens sísmicas .....	35
3.2	Cálculo dos atributos.....	36
3.2.1	Cálculo da Energia .....	36
3.2.2	Cálculo da Similaridade .....	37
3.2.3	Cálculo do Correlograma.....	38
3.3	Classificação usando SVM .....	39
4	RESULTADOS .....	41
5	CONCLUSÃO .....	45
6	REFERÊNCIAS.....	47

# 1 INTRODUÇÃO

A perfuração de poços de petróleo pode consumir até 85% do custo total da exploração, em uma estimativa grosseira (RUSSO et al., 2004) e a descoberta de petróleo em uma determinada área não é uma tarefa fácil. Para que ocorra uma possível perfuração em regiões que apresentam a suspeita da existência de hidrocarbonetos, realizam-se estudos sobre as condições geológicas da área e a viabilidade econômica, dentre esses estudos, o método sísmico é o mais decisivo.

Obviamente, a perfuração é necessária para a extração de óleo e gás, porém como método de identificação de reservatórios de hidrocarbonetos em subsuperfície apresenta desvantagens, fornecendo apenas informações pontuais e tendo um custo elevado. A sísmica de reflexão por sua vez, como método indireto de investigação é mais barata, gera informações sobre grandes áreas, possibilita a investigação de alvos com profundidade de centenas a milhares de metros e com uma alta resolução.

Para que o intérprete localize seus alvos, as imagens sísmicas da subsuperfície devem ser de boa qualidade. O método sísmico e a pesquisa por novas técnicas para classificação de artefatos sísmicos tornaram-se objetos de grandes investimentos.

Na aquisição dos dados o método sísmico utiliza uma fonte sísmica artificial e receptores dispostos na superfície, explosões de dinamite e geofones, no caso terrestre, ou canhões pneumáticos e hidrofones, no caso marítimo, as explosões geram ondas mecânicas que se propagam através das camadas rochosas, já os receptores registram os sinais que são refletidos pelas interfaces entre camadas geológicas diferentes.

Com o objetivo de retirar ou minimizar eventos indesejáveis utiliza-se um conjunto de técnicas que atuam sobre os dados adquiridos, este é o chamado processamento sísmico que visa facilitar a interpretação desses dados (MACHADO et al., 2008), de forma que melhor represente a realidade da geologia e dos reservatórios de petróleo.

A localização de petróleo ou gás em bacias sedimentares se baseia no fato do petróleo se alojar em uma estrutura chamada de armadilha ou trapa, resultantes de modificações sofridas pelas rochas ao longo do tempo geológico que ficam

localizadas na parte mais alta de um compartimento de rocha porosa, isolada por camadas impermeáveis. É necessário o conhecimento da geometria destas estruturas para que seja feito um mapeamento correto do campo petrolífero, pois, em determinadas situações os fatores estruturais, principalmente a presença de falhas determinam o acúmulo de petróleo.

O processo de utilizar técnicas de reconhecimento de padrões para representar adequadamente os reservatórios de petróleo através de dados sísmicos, tem sido alvo de inúmeras publicações científicas (JOHANN, RIBET, 1999). Nesse contexto, utilizaremos o atributo geoestatístico correlograma na extração de características e técnicas de aprendizado de máquina, em particular a máquina de vetores de suporte (SVM).

Este trabalho visa estudar o problema de classificação de falhas geológicas em volume de dados sísmico e aplicar técnicas de aprendizado de máquina para este fim. Especificamente pretende-se pesquisar aplicabilidade do atributo correlograma como descritor de característica.

O presente trabalho será composto de mais quatro capítulos, conforme descrição sumária a seguir:

Capítulo 2 – apresenta a fundamentação teórica com os conceitos básicos sobre dados sísmicos, sísmica de reflexão, as três etapas de levantamento sísmico: aquisição, processamento e interpretação; os atributos sísmicos: energia, similaridade e correlograma; e a máquina de vetores de suporte.

Capítulo 3 – nele mostramos a metodologia proposta para alcançar a solução.

Capítulo 4 – apresenta os resultados obtidos.

Capítulo 5 – traz as conclusões e as considerações finais sobre o trabalho realizado.

## 2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo serão abordados temas importantes para compreensão deste trabalho, onde partiremos dos conceitos sobre dados sísmicos, a aquisição, o processamento e a interpretação, passando pelos atributos sísmicos de energia e similaridade até chegarmos às funções geoestatísticas e, enfim, o atributo correlograma que é o objeto de estudo dessa monografia. Serão apresentados ainda conceitos sobre máquina de vetores de suporte e as métricas de desempenho.

### 2.1 Dados Sísmicos

Encontra-se em rochas sedimentares a grande maioria dos acúmulos de hidrocarbonetos como óleo e gás (GERHARDT, 1998). Os geofísicos utilizam a sísmica de reflexão na descoberta desses reservatórios, buscando identificar estruturas onde o óleo possa estar aprisionado.

Através de processos geológicos de baixa temperatura, como desidratação, reações minerais e outros que sucedem à deposição de sedimentos, são formadas as rochas sedimentares, esses processos são conhecidos como diagênese. Com o acúmulo dos sedimentos a pressão na bacia sedimentar aumenta e acionam os processos de compactação, cimentação, recristalização, hidratação, lixiviação, ação de bactérias, etc., que compõem a diagênese. O início do processo de geração de hidrocarbonetos é dado pelos sedimentos orgânicos, também conhecidos como rochas geradoras.

A formação de uma acumulação de petróleo em uma bacia sedimentar requer ainda a existência de rochas com porosidade e permeabilidade necessárias à acumulação e produção dos hidrocarbonetos, denominada de rochas reservatório. A maior parte das reservas conhecidas encontra-se em arenitos e rochas carbonáticas<sup>1</sup>, embora acumulações de petróleo também ocorram em folhelhos, conglomerados ou mesmo em rochas ígneas e metamórficas.

O óleo é mais leve do que a água, portanto um reservatório não é possível a menos que o óleo esteja impedido de escapar. Assim, é necessária a ocorrência de

---

<sup>1</sup> Rocha carbonática é um tipo de rocha sedimentar cuja composição primária são os carbonatos. Os principais tipos de rochas carbonáticas são o calcário, o cré e o tufo.

uma camada de rochas impermeáveis que retenha o óleo, normalmente folhelho ou sal, formando uma trapa (armadilha), denominada de rocha selante ou capeadora, onde os hidrocarbonetos podem ser trapeados e acumulados. Trapas são situações geológicas em que o arranjo espacial de rochas reservatório e selante possibilitam a acumulação de petróleo, formando assim um reservatório.

As Figuras 2.1, 2.2 e 2.3 ilustram exemplos de trapas e tipos de reservatórios, a primeira mostra uma trapa estrutural anticlinal, formando uma estrutura capeadora convexa; a segunda é uma trapa estrutural composta por deslocamentos lineares (falhas geológicas), as trapas estruturais são formadas por alguma deformação local, como resultado de falhamentos e de dobramentos; a terceira ilustra as trapas estratigráficas associadas a discordâncias.

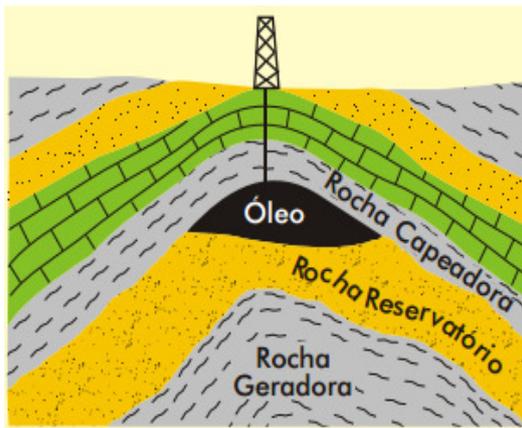


Figura 2.1: Trapa Estrutural, anticlinal. (TEIXEIRA et al., 2000)

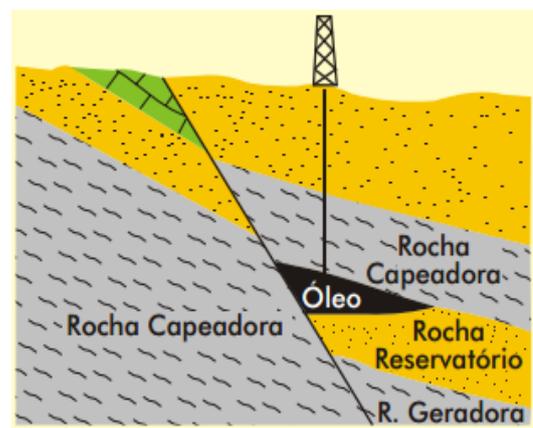


Figura 2.2: Trapa Estrutural, falhas. (TEIXEIRA et al., 2000)

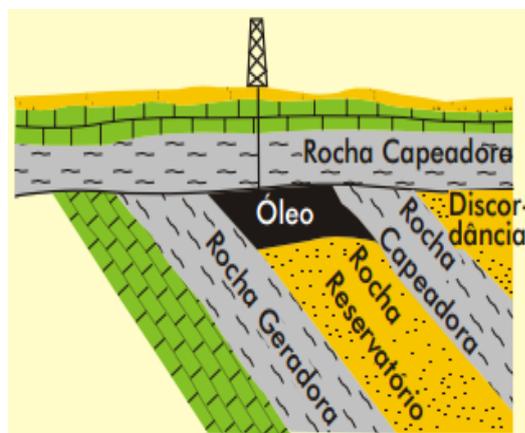


Figura 2.3: Trapa Estratigráfica, discordâncias. (TEIXEIRA et al., 2000)

### 2.1.1 Sísmica de Reflexão

A sísmica de reflexão é, sem dúvida, a técnica geofísica mais utilizada, graças à indústria do petróleo que a emprega em larga escala, sendo também empregada em pesquisas sobre a origem das pedras que compõem a crosta terrestre, em inspeção ambiental e aplicações rasas.

Graças ao seu poder de resolução, a sísmica de reflexão vem sendo usada intensivamente pela indústria do petróleo, constituindo-se por isso em um dos métodos geofísicos mais utilizados (BUSTAMANTE, 2008).

A sísmica de reflexão se divide em dois tipos, a convencional e a de alta resolução. A convencional é a utilizada para petróleo, onde a profundidade de investigação é de centenas a milhares de metros e o tamanho dos alvos é de dezenas a centenas de metros. O termo sísmica de reflexão de alta resolução é utilizado para alvos mais rasos e menores, que por isso mesmo precisa ter uma maior resolução, onde a altura de investigação é de dezenas de metros e o tamanho dos alvos é de metros.

O princípio físico utilizado na metodologia consiste na reflexão parcial da energia das ondas elásticas na sua passagem através de um plano de descontinuidade de velocidade sísmica. A energia refletida será maior para contrastes mais destacantes de velocidade.

Como ilustrado na Figura 2.4, a onda sísmica provocada pela fonte de energia, por exemplo, um fuzil sísmico ou por um explosivo, se propaga no subsolo e é parcialmente refletida contra a superfície pelas descontinuidades geológicas. Uma sequência de geofones alinhados na superfície permite ressaltar simultaneamente as formas de onda sísmica refletidas.

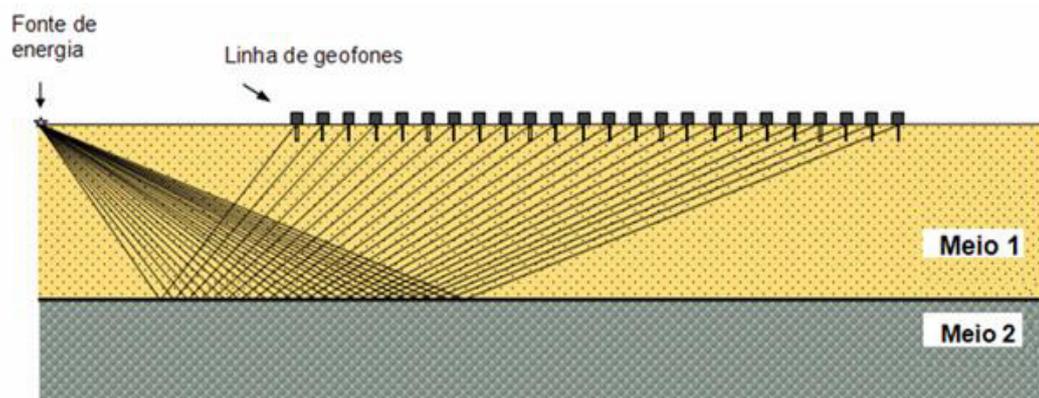


Figura 2.4: Geometria de aquisição (TECNO PETRO BR, 2008).

Essa geometria de base é deslocada ao longo do alinhamento dos geofones a um passo geofônico, a cada energização. Esse procedimento permite, portanto, registrar mais vezes as informações provenientes do mesmo refletor em profundidade, proporcionando assim uma "cobertura múltipla", além da vantagem das ondas sísmicas provenientes do mesmo ponto em profundidade têm seguido diversos percursos caracterizados por diversos ângulos de incidência.

A qualidade dos terrenos e os parâmetros de aquisição escolhidos são cruciais para garantir a qualidade dos dados obtidos na fase de aquisição. A definição dos parâmetros, segundo representação da Tabela 2.1, depende da experiência do operador ou dos resultados de um perfil de teste denominado *walk-away*, realizado com o escopo de testar as características dos terrenos e assim permitir escolher as melhores geometrias e os melhores parâmetros de aquisição.

Tabela 2.1: Parâmetros de aquisição.

<b>Parâmetros</b>	<b>Definição</b>
Tempo de medição	O tempo que precisa para o equipamento receber toda informação de interesse;
Procedimento de amostragem	Tempo de aquisição de cada valor, cuja sequência constitui a forma da onda;
Distância entre a origem e o último geofone <sup>2</sup>	Determina a profundidade explorada para a qual é ainda possível uma boa análise das velocidades;
Distância entre a origem e o primeiro geofone <sup>2</sup>	Não deve ser muito grande, de modo a permitir uma boa avaliação das espessuras e das velocidades da primeira camada areada;
Distância entre os geofones <sup>2</sup>	Depende dos parâmetros precedentes e do número de canais disponíveis no sismógrafo. Quanto mais curta for, maior será a resolução superficial;
Filtros analógicos	Uma boa escolha dos filtros permite a eliminação dos ruídos e o melhor aproveitamento da dinâmica do sismógrafo.

<sup>2</sup> Instrumentos que servem para registrar ondas sísmicas.

Além da definição dos parâmetros, outros conceitos são muito importantes para garantir a qualidade dos dados obtidos, como a fonte de energia, o sistema de recepção e o sistema de aquisição.

A exploração (óleo e gás) baseada em sísmica pode ser dividida em três etapas: aquisição dos dados; processamento, correção de erros ocorridos durante a etapa de levantamento sísmico; interpretação, caracterização e monitoramento dos reservatórios de petróleo (ROBINSON, TREITEL, 1980), conforme ilustrado na Figura 2.5.

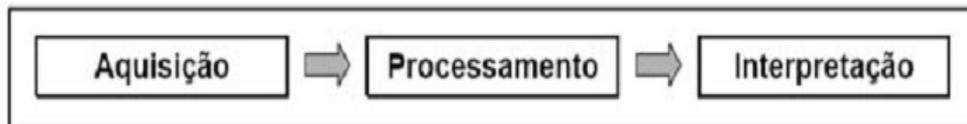


Figura 2.5: Exploração (óleo e gás) baseada em sísmica (ROBINSON, TREITEL, 1980).

### 2.1.2 Aquisição

Utiliza-se na aquisição uma fonte de energia para gerar ondas sísmicas que se propagam abaixo da superfície da terra, habitualmente se usa:

- um peso ou marreta, caso a máxima profundidade a explorar não seja superior a 30-40 metros;
- um fuzil sísmico, caso a máxima profundidade a explorar seja entorno de 100 metros;
- um explosivo (explosões de dinamite), que permite regular a energia imersa no terreno para uma pesquisa a maiores profundidades;
- em aquisições marinhas são usados normalmente dispositivos pneumáticos como canhões de ar (*Airguns*).

A Figura 2.6 ilustra como Caminhões Vibradores são utilizados para gerar ondas acústicas em aquisições terrestres. A Figura 2.7 demonstra a utilização de *Streamers*, cabos que contém receptores que flutuam e podem ser rebocados no mar, e os *Airguns*, canhões de ar comprimido que produzem ondas acústicas e servem como fonte de energia sísmica no mar.



Figura 2.6: Caminhão vibrador (PGS, 2010).

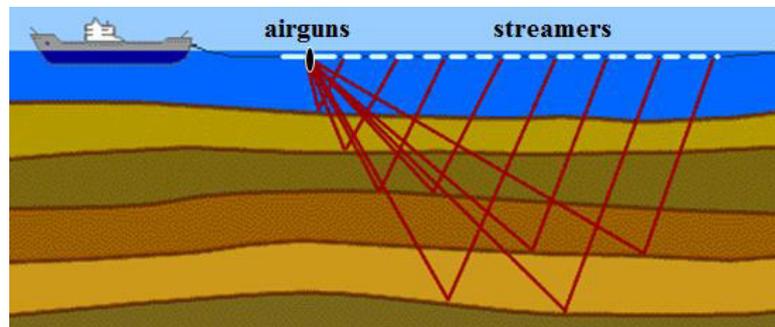


Figura 2.7: Cabos sísmicos flutuantes e canhões de ar (PGS, 2010).

A onda sísmica ao atingir uma interface entre duas camadas de rocha com valores de impedância acústica (o produto da velocidade sísmica de um intervalo de rochas pela sua densidade) diferentes, parte da onda é refratada e continua viajando para baixo; outra parte é refletida e retorna a superfície (SILVA, 2004).

Ao regressar à superfície a parte da onda refletida é captada pelo sistema de recepção, em caso de aquisições terrestres, podem ser utilizados de 12 a 48 geofones ou mais, que medem a velocidade do pulso sísmico que retorna à superfície, em aquisições marinhas utiliza-se hidrofones, que medem a variação de pressão gerada na água. Após a captura, as formas de onda simultaneamente presentes nos geofones/hidrofones são adquiridas pelo sistema de aquisição com a utilização de um sismógrafo multicanal, como por exemplo, *EEG BR-24* ou *BISON 9048*, que armazena tanto o tempo de chegada da onda quanto à intensidade medida neste momento. A Figura 2.8 demonstra o esquema de aquisição sísmica terrestre e marítima, à esquerda e à direita, respectivamente (GERHARDT, 1998).

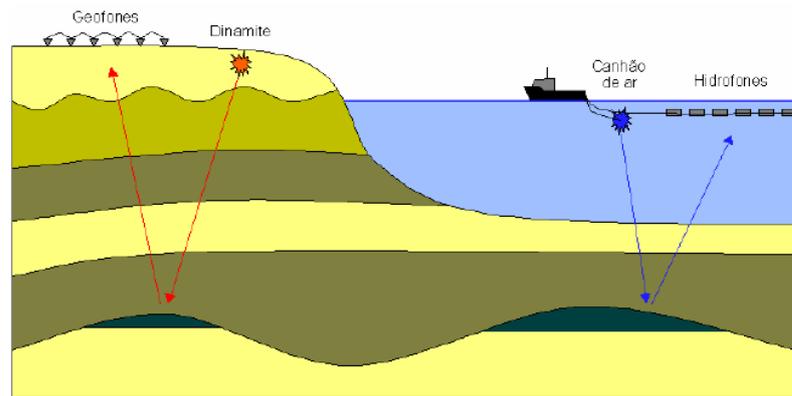


Figura 2.8: Aquisição Sísmica (GERHARDT, 1998).

Depois de alterar a posição da fonte de energia, do sistema de recepção e realizar várias detonações, todos os dados armazenados são enviados para a segunda etapa, a de processamento, para que as estruturas geológicas com condições de possuir acumulações de petróleo ou gás possam ser identificadas através da interpretação da energia captada e convertida em sinais elétricos.

### 2.1.3 Processamento

Nesta etapa, os dados adquiridos são submetidos ao chamado processamento sísmico, a fim de aumentar a relação sinal-ruído, aplica-se sobre os dados um conjunto de técnicas que atuam de forma a retirar ou minimizar eventos indesejáveis. O objetivo do processamento sísmico é facilitar a interpretação dos dados sísmicos.

Além da correção de erros, os dados são reorganizados para formarem em caso de imagens 2D, fatias (*slice*) onde as amostras de amplitude sísmica são representadas pelos pontos do plano, ou no caso de imagens 3D (volume sísmico), uma grade tridimensional onde as amostras de amplitude sísmica são representadas pelos vértices da grade (*voxel*).

Os dados resultantes correspondem a uma função escalar tridimensional das variáveis discretas espaciais  $x$ ,  $y$  e do tempo  $t$ , associado ao tempo decorrido entre a geração do pulso em superfície, sua reflexão por um refletor em subsuperfície e, finalmente, seu registro nos receptores, ou seja, o tempo de cada amostra corresponde ao tempo que a onda leva para viajar até uma interface mais o tempo da volta à superfície. Cada valor da função, chamada de amplitude sísmica, é

proporcional à energia refletida pela interface. Um conjunto de amostras de amplitudes como uma função do tempo  $t$  para  $x, y$  constantes é denominado *traço sísmico* (SILVA, 2004).

A Figura 2.9 mostra a função de amplitudes sísmicas do traço sísmico (a), uma *linha sísmica* (b), extraída de um volume sísmico (c), com duas direções espaciais, que são chamadas de *inline* (direção das linhas sísmicas) e *crossline* (direção perpendicular às linhas sísmicas), além de uma direção temporal.

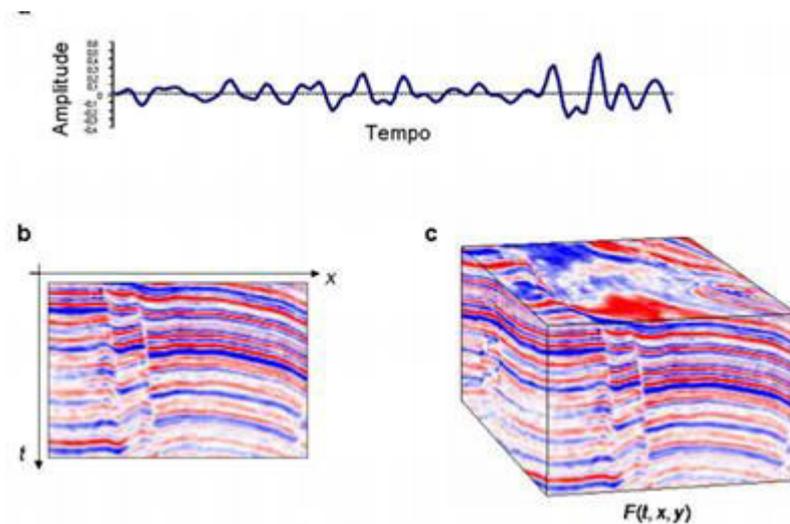


Figura 2.9: Dados sísmicos. Traço sísmico (a). Linha sísmica (b). Volume sísmico (c) (MACHADO, 2008).

#### 2.1.4 Interpretação

Na etapa de interpretação sísmica o intérprete, em geral um geólogo ou geofísico, analisa os dados sísmicos processados a fim de desenvolver um modelo geológico que represente a área onde foi realizado o levantamento. Para a indústria de petróleo, a interpretação sísmica visa identificar estruturas onde o óleo possa estar aprisionado em um reservatório. A existência de um reservatório só é possível se o óleo, uma vez mais leve do que a água, esteja impedido de escapar. Para que isso ocorra é necessária à existência de uma camada de rochas impermeáveis, folhelho (rocha de granulação muito fina, formada pela compactação de argila ou silte) ou sal, formando assim uma trapa (armadilha).

A interpretação sísmica pode ser classificada em dois tipos: estrutural e estratigráfica. A interpretação estrutural consiste principalmente em criar horizontes e determinar superfícies de falha. Um horizonte sísmico ou refletor é uma superfície

que separa duas camadas diferentes de rocha, onde tal superfície (mesmo sem ter sido identificada) está associada com uma reflexão que se estende por uma grande área (SHERIFF, 1991). A ocorrência de falhas, fraturas de rochas em subsuperfície causada por forças tectônicas, geram descontinuidades na estrutura das camadas tornando assim a criação de horizontes mais difícil.

Quando ocorrem deslocamentos de camadas impermeáveis às falhas podem criar trapas estruturais onde o óleo é aprisionado, ou gerar vazamentos, ao romper a capa selante de uma trapa, por isso falhas são extremamente importantes na identificação de trapas de origem estrutural e possíveis reservatórios.

As falhas também podem atuar no sentido de gerar um selante, bloqueando o fluxo entre diferentes compartimentos de um reservatório ou podem agir como condutores, permitindo a comunicação entre diferentes compartimentos (PEDERSEN, 2005).

A Figura 2.10 mostra onde uma falha gera um bom selante (a) devido à justaposição de camadas impermeáveis contra o reservatório e um mal selante (b) onde os fluidos podem migrar para fora do reservatório.

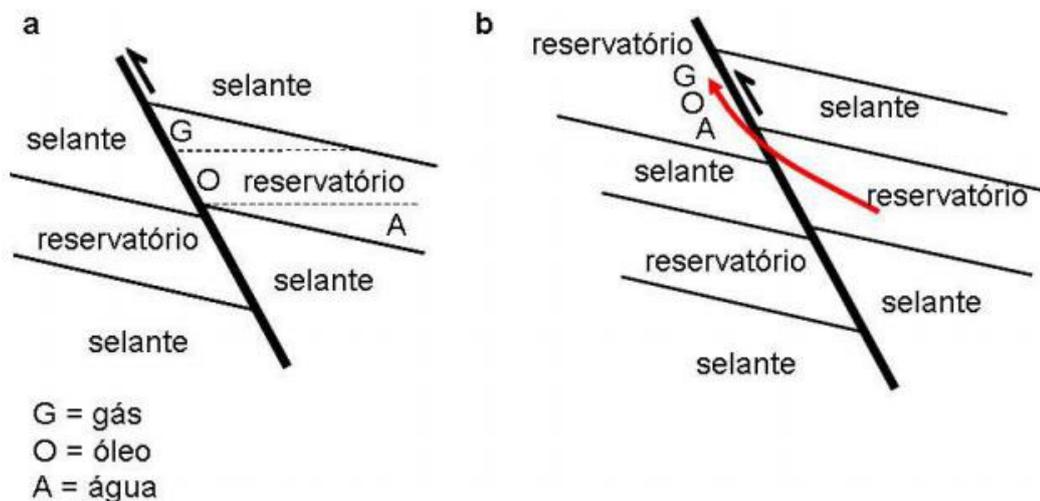


Figura 2.10: Exemplos de falhas. Bom selante (a). Mal selante (b) (MACHADO, 2008).

O outro tipo de interpretação sísmica, a estratigráfica visa entender a maneira como as camadas foram se formando ao longo do tempo.

O resultado da interpretação sísmica depende de uma série de fatores, como por exemplo, a experiência do intérprete, a qualidade dos dados sísmicos, a complexidade do sistema geológico e etc., embora o rastreamento de horizontes já seja feito de forma automática, a interpretação tridimensional de falhas permanece como uma tarefa manual.

## **2.2 Atributos Sísmicos**

Atributos sísmicos são todas as informações obtidas de um dado sísmico capazes de revelar de forma mais clara as informações geológicas contida nos sinais sísmicos, seja através de medidas diretas, por lógica (operações matemáticas) ou experimentalmente.

Existem vários tipos de atributos: pré-empilhados, inversão, velocidade, horizonte, multi-componente e outros, cada um com sua funcionalidade. Por exemplo, para um determinado material geofísico existe um atributo que melhor o identifica, outro para identificar falhas geológicas ou indicadores de hidrocarbonetos e assim por diante, ou seja, dependendo do tipo de análise em questão, deve-se escolher o tipo adequado de atributo (MATOS, 2004). Atributos também são usados como filtros para revelar tendências, padrões ou combinado para prever as características sísmicas.

Os atributos utilizados como indicadores de falha partem da suposição de que os horizontes sísmicos podem ser modelados como localmente planares e que as falhas geram perturbações nesse modelo (MACHADO, 2008). Utilizam-se técnicas de reconhecimento de padrões e combinações de atributos sísmicos para realizar a análise das características sísmicas e buscar a identificação de características geológicas dos reservatórios.

O principal objetivo dos atributos sísmicos é ajudar o intérprete a gerar um modelo geológico de uma determinada área sísmica. Quanto mais detalhado for o modelo geológico feito pelo intérprete maior será o índice de sucesso na exploração da área em questão.

A Figura 2.11 mostra os atributos sísmicos e suas respectivas configurações de parâmetros para a Rede Neural utilizados nos trabalhos sobre detecção semiautomática de falha em dados sísmicos 3D (TINGDAHL, ROOIJ, 2005). A

Figura 2.12 (a) e (b) demonstram dados sintéticos com falhas e os atributos da Figura 2.11 calculados ao longo da linha branca.

Attribute	Time gate (ms)	Lateral position	Other settings
Energy	[-32,32]	-	-
Similarity	[-24,24]	(0,1) and (0,-1) 90° rotated	Minimum similarity is output
Steered similarity	[-24,24]	Positions (0,1) and (0,-1) 90° rotated	Minimum similarity is output
Cube similarity	[-24,24]	Full block	Full steering, minimum similarity is output
Cube similarity variance	[-24,24]	Full block	Full steering
Frequency-AFS	[-36,36]	-	-
Frequency-FSF	[-36,36]	-	-
Dip angle	-	-	-
Dip variance	[-32,32]	Stepout 1	-
Most positive curvature	-	Stepout 2	-
Most negative curvature	-	Stepout 2	-
Disc variance	-	Stepout 3	Full steering

Figura 2.11: Atributos e as configurações dos parâmetros da Rede Neural (TINGDAHL, ROOIJ, 2005).

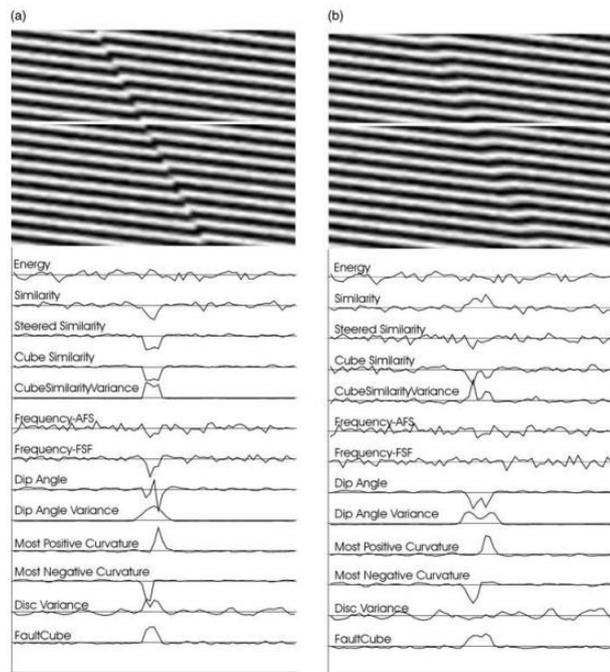


Figura 2.12: Dados sintéticos com falhas e os atributos calculados ao longo da linha branca (TINGDAHL, ROOIJ, 2005).

A seguir é feita uma breve apresentação de alguns atributos sísmicos citados na literatura como indicadores de falha.

### 2.2.1 Energia

O atributo de energia pode ser enunciado pela Equação 2.1 (TINGDAHL, ROOIJ, 2005):

$$E = \frac{\sum_{i=0}^{N-1} f(t_0+i.dt)^2}{N} \quad (2.1)$$

onde,  $f$  representa a amplitude do traço,  $t_0$  o limite superior da janela,  $N$  o número de amostras na janela e  $d_t$  o intervalo amostral.

Desta forma, o atributo de energia pode ser definido como o quadrado da soma da amplitude dentro de uma janela (*gate*) ao redor da amostra de investigação.

### 2.2.2 Energia Vertical

O atributo de energia vertical combina o cálculo da média quadrática em uma janela deslizante sobre cada traço, com um processo de afinamento da imagem em cada fatia de tempo constante (MACHADO et al., 2006a; 2006b).

A média quadrática ignora as oscilações do sinal da amplitude sísmica além de reforçar o efeito de diminuição de valores de amplitude, enquanto o afinamento preserva os pontos com valores mínimos de média quadrática ao longo de estruturas em forma de linha, funcionando assim como uma restrição geométrica sobre os dados sísmicos, uma vez que, falhas usualmente geram superfícies verticais; tais superfícies, quando vistas em fatias de tempo constante, geram linhas com baixos valores de média quadrática de amplitude sísmica.

### 2.2.3 Similaridade

O atributo de similaridade pode ser enunciado pela expressão mostrada na Equação 2.2 e os respectivos valores para os vetores  $v$  e  $u$  (TINGDAHL, ROOIJ, 2005).

$$s = 1 - \frac{|v-u|}{|v|+|u|} \quad (2.2)$$

$$v = \begin{bmatrix} f(t_1, x_v, y_v) \\ f(t_1 + dt, x_v, y_v) \\ \vdots \\ f(t_2 - dt, x_v, y_v) \\ f(t_2, x_v, y_v) \end{bmatrix}, u = \begin{bmatrix} f(t_1, x_u, y_u) \\ f(t_1 + dt, x_u, y_u) \\ \vdots \\ f(t_2 - dt, x_u, y_u) \\ f(t_2, x_u, y_u) \end{bmatrix} \quad (2.3)$$

onde  $t$  representa o intervalo de tempo considerado,  $dt$  o intervalo amostral,  $t_1$  e  $t_2$  os limites do tempo da janela,  $(x_v, y_v)$  e  $(x_u, y_u)$  as duas posições de traço que serão comparadas, e  $f$  é o valor de amplitude no cubo.

Sendo assim, o atributo de similaridade pode ser definido como a distância no hiperespaço entre os vetores dos segmentos, normalizados à soma dos comprimentos dos vetores, sendo calculado entre dois traços sísmicos.

## 2.2.4 Correlograma

O método proposto neste trabalho é buscar o reconhecimento de falhas geológicas através da análise de sua textura, logo, é interessante investigar métodos de representação que possam analisar e extrair informações das mesmas. Estas propriedades, chamadas características, são elementos que fazem a distinção entre classes de objetos (NADLER, SMITH, 1993).

No contexto da estatística, a textura pode ser descrita em termos dos dois componentes principais associados a *pixels* (ou outra unidade): variabilidade e autocorrelação espacial.

Sendo assim, utilizaremos a medida geoestatística correlograma, e uma combinação dela com outros atributos sísmicos já existentes na literatura a fim de extrair características para identificar e sugerir possíveis regiões de falhas no dado sísmico.

O conceito de geoestatística surgiu em 1951 e pode ser definida como sendo a estatística sobre uma população com endereço conhecido (KRIGE, 1951). O correlograma, assim como outras funções geoestatísticas tem como principal característica descrever a continuidade espacial entre um conjunto de pares de pontos de uma amostra em função de uma distância e uma direção. Esta continuidade é importante, pois determina o relacionamento entre os pontos da amostra, quanto maior a proximidade menor será sua similaridade (ISAACS, SRIVASTAVA, 1990).

As funções geoestatísticas são aplicadas em diversas áreas, tais como: Ciências da Terra, Sensoriamento Remoto, Processamento de Imagens, etc.

A função correlograma ou função de correlação é a relação do coeficiente de correlação de um diagrama de dispersão e uma distância  $h$ . Podendo ser definida

como sendo a versão normalizada da função de covariância, onde os coeficientes de correlação estão na faixa de  $-1$  a  $1$ .

Um diagrama de dispersão representa todas as possíveis combinações dos pares de valores cuja localização é separada por uma distância e certa direção.

O correlograma apresenta valores altos para distâncias pequenas, sendo que os mesmos tendem a decrescer à medida que a distância aumenta, ou seja, espera-se que a correlação seja alta para unidades que estão próximas umas das outras (correlação igual a  $1$  para distância zero) e que tenda a zero quando a distância entre as unidade aumenta. A Figura 2.13 (a) ilustra um exemplo de amostras correlacionadas, enquanto a Figura 2.13 (b) mostra um exemplo de amostras não-correlacionadas (ALMEIDA, 2010).

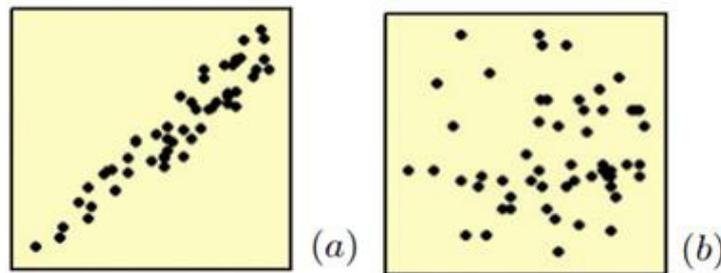


Figura 2.13: Amostras correlacionadas (a) e amostras não-correlacionadas (b) (ALMEIDA, 2010).

O correlograma é definido pela Equação 2.4:

$$\rho(h) = \frac{C(h)}{\sigma_{-h}\sigma_{+h}} \quad (2.4)$$

onde  $C(h)$  é a correlação, definida por:

$$C(h) = \left[ \frac{1}{N(h)} \sum_{i=1}^{N(h)} x_i y_i - m_{-h} m_{+h} \right] \quad (2.5)$$

$\sigma_{-h}$  é o desvio padrão dos valores das origens dos vetores e o seu valor é mostrado na Equação 2.6:

$$\sigma_{-h} = \left[ \frac{1}{N(h)} \sum_{i=1}^{N(h)} x_i^2 - m_{-h}^2 \right]^{\frac{1}{2}} \quad (2.6)$$

e  $\sigma_{+h}$  é o desvio padrão dos valores das extremidades dos vetores, cujo seu valor é mostrado na Equação 2.7:

$$\sigma_{+h} = \left[ \frac{1}{N(h)} \sum_{i=1}^{N(h)} x_i^2 - m_{+h}^2 \right]^{\frac{1}{2}} \quad (2.7)$$

## 2.3 Máquina de Vetores de Suporte

Máquina de Vetores de Suporte (*Support Vector Machines* - SVM) são técnicas de Aprendizado de Máquina (AM) embasadas na Teoria de Aprendizado Estatístico (VAPNIK, 1995), que vem recebendo grande atenção nos últimos anos. As SVMs vêm sendo utilizadas em diversas tarefas de reconhecimento de padrões, obtendo resultados superiores aos alcançados por técnicas similares em várias aplicações.

Aprendizado de Máquina (AM) é um campo de pesquisa da Inteligência Computacional que estuda o desenvolvimento de métodos capazes de extrair conceitos (conhecimento) a partir de amostras de dados (MITCHELL, 1997).

Em geral, os diversos algoritmos de AM são utilizados de forma a gerar classificadores para um conjunto de exemplos. Por classificação entende-se o processo de atribuir, a uma determinada informação, o rótulo da classe a qual ela pertence (RUSSEL, NORVIG, 1995). Portanto, as técnicas de AM são empregadas na indução (a partir de um conjunto de treinamento) de um classificador, que deve ser capaz (idealmente) de prever a classe de instâncias quaisquer do domínio em que ele foi treinado.

Três paradigmas podem ser utilizados na geração de um preditor por meio de técnicas de AM: aprendizado supervisionado, aprendizado não-supervisionado e por reforço (HAYKIN, 1999). A escolha de um paradigma de aprendizado determina a maneira como o algoritmo de AM se relaciona com seu meio ambiente, ou seja, o modo como ocorrerá o seu aprendizado por meio de um conjunto de dados.

A Tabela 2.2 apresenta os três paradigmas de AM, seus conceitos e exemplos de algoritmos que as utilizam.

Tabela 2.2: Paradigmas de Aprendizado de Máquina.

Paradigma	Conceito	Exemplos
Supervisionado	Tem-se a figura de um “professor externo”, o qual apresenta um conhecimento do ambiente representado por conjuntos de exemplos na forma entrada-saída. Neste caso, o algoritmo de AM é treinado a partir de conjuntos de exemplos rotulados com o objetivo de aprender uma função desejada.	Máquina de vetores de suporte; Redes neurais artificiais do tipo <i>multilayer perceptron</i> ; Árvores de decisão; Algoritmos genéticos.
Não-supervisionado	Não há a presença de um professor, ou seja, não existem instâncias rotuladas da função a ser aprendida. O algoritmo de AM aprende a representar (ou agrupar) as entradas submetidas segundo uma medida de qualidade.	Redes neurais do tipo mapa auto-organizáveis; Agrupamento hierárquico; K-médias.
Por reforço	O aprendizado se dá por meio de recompensas ou não ao indutor, dependendo de seu desempenho em aproximar a função desejada.	<i>Q-learning</i> ; <i>Sarsa</i> ; $Q(\lambda)$ .

Este trabalho foca a utilização das SVMs em tarefas de classificação por meio de treinamento supervisionado, algumas das principais características das SVMs que tornam seu uso atrativo são (SMOLA et al., 1999b):

- **Boa capacidade de generalização:** os classificadores gerados por uma SVM em geral alcançam bons resultados de generalização. A capacidade de generalização de um classificador é medida por sua eficiência na classificação de dados que não pertençam ao conjunto utilizado em seu treinamento.
- **Robustez em grandes dimensões:** as SVMs são robustas diante de objetos de grandes dimensões, como, por exemplo, dados sísmicos.
- **Convexidade da função objetivo:** a aplicação das SVMs implica na otimização de uma função quadrática, que possui apenas um mínimo global.

- **Teoria bem definida:** as SVMs possuem uma base teórica bem estabelecida dentro da Matemática e Estatística.

Com a utilização do algoritmo supervisionado, assume-se a existência de um conjunto de amostras de padrões cujas classes são conhecidas, conforme:

$$\mathfrak{S} = (x_1, k_1), (x_2, k_2), \dots, (x_m, k_m) \quad (2.8)$$

onde,  $X_i$  representa a  $i$ -ésima observação ou padrão,  $K_i$  representa a classe, ou rótulo, do padrão acima, e  $m$  representa o número de amostras disponíveis. O algoritmo gera um preditor ou classificador, capaz de prever precisamente o rótulo de novos dados, esse processo de indução é denominado treinamento (LORENA, CARVALHO, 2002).

O objetivo do SVM é produzir um classificador que consiga separar os dados de treinamento de forma a efetivamente maximizar a generalização do classificador obtido, ou seja, apresente um bom desempenho junto a amostras não observadas durante o treinamento.

A Figura 2.14 mostra o classificador linear denominado hiperplano ótimo (em vermelho), dentre os vários classificadores lineares possíveis que podem separar as amostras disponíveis sem nenhum erro, somente o hiperplano maximiza a distância entre o classificador e a amostra mais próxima de cada classe e espera-se que ele generalize melhor que os demais classificadores a classificação das amostras não utilizadas durante a fase de treinamento.

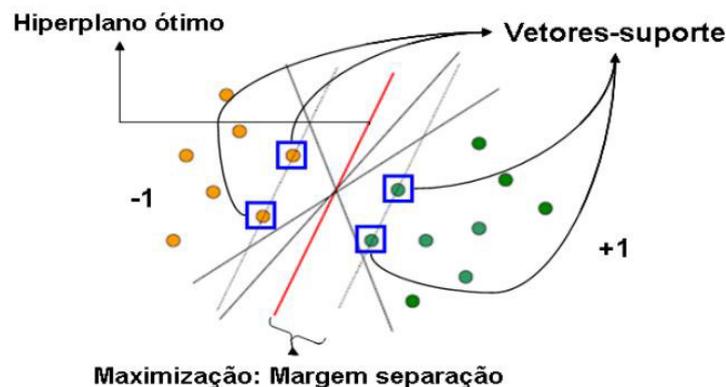


Figura 2.14: Hiperplano ótimo, margem de separação e vetores-suporte (ALMEIDA, 2007).

O SVM busca um compromisso entre a minimização do erro de treinamento e a complexidade da função escolhida para separação dos dados em classes. Essas

condições são alcançadas através do controle da dimensão de *Vapnik-Chervonenkis* (dimensão VC) do conjunto de funções do qual o classificador deve ser extraído. A dimensão VC mede a complexidade das hipóteses examinadas pelo algoritmo na busca por uma solução.

Considerando um conjunto de treinamento de pares de rótulo (valor objetivo):

$$(x_i, y_i), i = 1 \dots l \quad (2.9)$$

onde  $x_i \in R^n$  e  $y_i \in \{1, -1\}$ . A maximização da margem de separação entre os dados de treinamento e a fronteira de decisão induzida para separação das classes, tem como objetivo encontrar um hiperplano tal que a distância dos vetores-suporte seja maximizada. Na Equação 2.9 para cada ponto  $x$  no espaço de entrada é mapeado para o espaço de característica, onde os dados são linearmente separados por um hiperplano, definido pelo ponto  $z = \Phi(x)$  de alta dimensionalidade.

Uma forma mais simples que o do mapeamento  $\Phi$ , de se escrever as função de decisão é utilizando as funções *Kernel*, isso ocorre de maneira geral, pois em grande parte dos casos,  $\Phi$  assume formas bastante complexas. Um *Kernel*  $K$  é uma função que recebe dois pontos  $x_i$  e  $x_j$  do espaço de entradas e computa o produto escalar  $\Phi(x) \cdot \Phi(x_i)$  no espaço de características, como descrito em:

$$k(x, x_i) = \Phi(x) \cdot \Phi(x_i) \quad (2.10)$$

Sendo assim, a função de decisão é definida pela Equação 2.11:

$$f(x) = \sum_{i=1}^l \alpha_i y_i K(x, x_i) + b \quad (2.11)$$

onde  $k(x, x_i)$  é definido pela Equação 2.10, e coeficientes  $\alpha_i$  e  $b$ , são soluções das equações:

$$\min_{\omega, b, \xi} \frac{1}{2} w^T \cdot w + C \sum_{i=1}^l \xi_i \quad (2.12)$$

sujeito a:

$$y_i [w^T \cdot \phi(x_i) + b] \geq 1 - \xi_i \quad (2.13)$$

onde  $w$  representa o vetor de pontos perpendicular no hiperplano de separação,  $C > 0$  é um parâmetro escolhido pelo usuário que corresponde a penalidade do erro e os  $\xi_i$ 's são variáveis de folga que penalizam os erros de treinamento.

A utilidade dos *Kernels* está, portanto, na simplicidade de cálculo e na capacidade de representar espaços muito abstratos. Alguns dos *Kernels* mais utilizados são os polinomiais, os Gaussianos ou RBF (*Radial-Basis Function*) e os *Sigmoidais*, apresentados na Tabela 2.3.

Tabela 2.3: Principais *Kernels* utilizados nas SVMs. Adaptado de (HAYKIN, 1999).

Tipo de Kernel	Função $K(x_i, x_j)$ correspondente	Comentários
Linear	$x^T y$	
Polinomial	$(x_i^T \cdot x_j + 1)^p$	A potência $p$ deve ser especificada pelo usuário.
Gaussiano	$\exp\left(-\frac{1}{2\sigma^2} \ x_i - x_j\ ^2\right)$	A amplitude $\sigma^2$ é especificada pelo usuário.
Sigmoidal	$\tanh(\beta_0 x_i \cdot x_j + \beta_1)$	Utilizado somente para alguns valores de $\beta_0$ e $\beta_1$ .
Função básica radial (RBF)	$e^{-\gamma \ x-y\ ^2}$	O parâmetro $\gamma$ é definido pelo usuário.

A obtenção de um classificador por meio do uso de SVMs envolve a escolha de uma função *Kernel*, além de parâmetros desta função. A escolha do *Kernel* e dos parâmetros considerados tem efeito no desempenho do classificador obtido (MULLER et al., 2001), pois eles definem a fronteira de decisão induzida.

Há uma grande variedade de implementações de SVMs disponíveis para uso não comercial, como por exemplo: SMO, *SVMlight*, *mySVM*, LIBSVM, *SVMtoolbox* para *Matlab* e *SVMtorch*. No presente trabalho foi utilizado o LIBSVM, uma biblioteca de funções para SVMs. Possui fontes em C++ e Java e interface para *Matlab*. As implementações são baseadas nas ferramentas SMO e *SVMlight* e é utilizado principalmente em problemas de classificação multiclases e regressão. Com o LIBSVM foi utilizado o *Kernel* RBF com o parâmetro  $\gamma > 0$  definido pelo usuário, sem a necessidade então de se considerar o *Kernel* Linear (KEERTHI, LIN, 2003).

## 2.4 Métrica de Desempenho

A validação dos resultados encontrados por meio das técnicas de Aprendizado de Máquina supervisionado utilizadas nas SVMs geralmente se dá através da análise do desempenho do classificador gerado pelo mesmo na classificação das classes de teste.

O presente trabalho utiliza os métodos estatísticos: Sensibilidade ( $S$ ), Especificidade ( $E$ ) e a Acurácia ( $A$ ), para validar os resultados encontrados, para os valores de Acurácia são apresentados ainda a Média ( $M$ ) e a medida de dispersão: Desvio Padrão ( $Dp$ ).

Na classificação de falhas geológicas em dados sísmicos, os possíveis resultados são positivo e negativo para a existência ou não de falhas nas imagens, podem-se obter erros de precisão ou mesmo de interpretação errada. Quando o resultado for positivo, os pontos definidos na imagem como Falha foram classificados como tal (verdadeiro positivo - VP) ou os pontos definidos como Falha foram classificados como de Não-Falha (falso positivo - FP). Da mesma forma, quando o resultado for negativo, os pontos definidos como Não-Falha foram classificados como tal (verdadeiro negativo - VN) ou os pontos definidos como Não-Falha foram classificados como Falha (falso negativo - FN).

A sensibilidade mede a capacidade do teste em predizer a condição Falha para aqueles casos que realmente a apresentam, ou seja, o quão sensível é o teste. A especificidade mede a capacidade do método de decisão de apontar ausência da condição para aqueles casos que realmente não a têm, ou seja, o quão específico o teste é. Já a acurácia é a probabilidade de predições corretas, as Equações para os métodos estatísticos utilizados são mostradas na Tabela 2.4.

Tabela 2.4: Equações dos métodos estatísticos utilizados pra validação.

<b>Método</b>	<b>Equação</b>
Sensibilidade (S)	$S = \frac{VP}{VP + FN}$
Especificidade (E)	$E = \frac{VN}{VN + FP}$
Acurácia (A)	$A = \frac{VP + VN}{VP + VN + FP + FN}$
Média (M)	$M = \frac{1}{n} \sum_{i=1}^n a_i$
Desvio Padrão (Dp)	$Dp = \sqrt{V}$

### 3 METODOLOGIA PROPOSTA

O objetivo do trabalho é desenvolver um método que auxilie o intérprete na classificação de possíveis falhas geológicas em dados sísmicos utilizando técnicas de aprendizado de máquina. A interpretação sísmica visa identificar estruturas onde o óleo possa estar aprisionado em um reservatório. Especificamente pretende-se pesquisar a aplicabilidade do atributo correlograma como descritor de característica. Para atingir o objetivo seguimos os seguintes passos:

1. Aquisição, preparação e tratamento das imagens sísmicas;
2. Cálculo dos atributos: Energia, Similaridade e Correlograma;
3. Classificação usando SVM.

#### 3.1 Aquisição, preparação e tratamento das imagens sísmicas

Para o desenvolvimento da metodologia proposta utilizou-se um volume sísmico sintético com as direções espaciais, *inline*<sup>3</sup> e *crossline*<sup>4</sup> definidas de 1 a 256 e a direção temporal de 1 a 255, ou seja, dados de amplitude de 256 x 256 x 255. A Figura 3.1 mostra o volume sísmico utilizado.



Figura 3.1: Volume sísmico.

Na etapa de preparação das imagens sísmicas os dados utilizados armazenados no formato *begin-endian*<sup>5</sup> foram convertidos para *little-endian*<sup>6</sup>, por fim o volume sísmico foi dividido em *slices* (fatias). Cada fatia foi digitalizada na

<sup>3</sup> Direção das linhas sísmicas.

<sup>4</sup> Direção perpendicular às linhas sísmicas.

<sup>5</sup> Byte menos significativo com o maior endereço.

<sup>6</sup> Byte menos significativo no menor endereço.

resolução 256 x 256 *pixels* e os dados de amplitude, com valores entre -1699.143604 e +1769.342529, quantizados para de 8 bits, gerando imagens conforme a Figura 3.2.

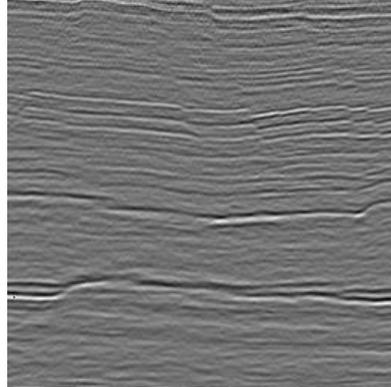


Figura 3.2: Fatia do volume sísmico.

## 3.2 Cálculo dos atributos

Dentre a centena de atributos sísmicos existentes na literatura, foram escolhidos dois: o de energia e o de similaridade, calculados ao longo do traço sísmico, na vertical da imagem.

### 3.2.1 Cálculo da Energia

A Figura 3.3 demonstra o cálculo da energia vertical, onde para o elemento (*pixel*) circulado em vermelho, com valor 103, é calculado o quadrado da soma da amplitude,  $(8^2 + 11^2 + 103^2 + 5^2 + 4^2)$ , dentro da janela ao redor da amostra de investigação e dividido por 5, quantidade de elementos da janela, que sempre percorre a imagem na vertical, o tamanho dessa janela é variável e configurado pelo usuário (ALMEIDA, 2007), porém no presente trabalho utilizou-se a janela de tamanho 15.

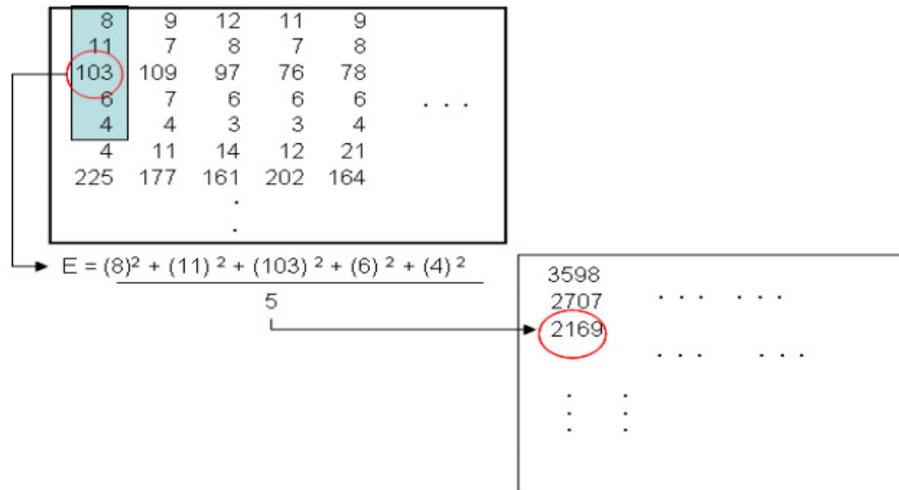


Figura 3.3: Cálculo da Energia Vertical (ALMEIDA, 2007).

As equações para o cálculo deste atributo foi descrita na Seção 2.2.1 e 2.2.2. A Figura 3.4 mostra um exemplo de uma imagem com o atributo de energia vertical calculado.

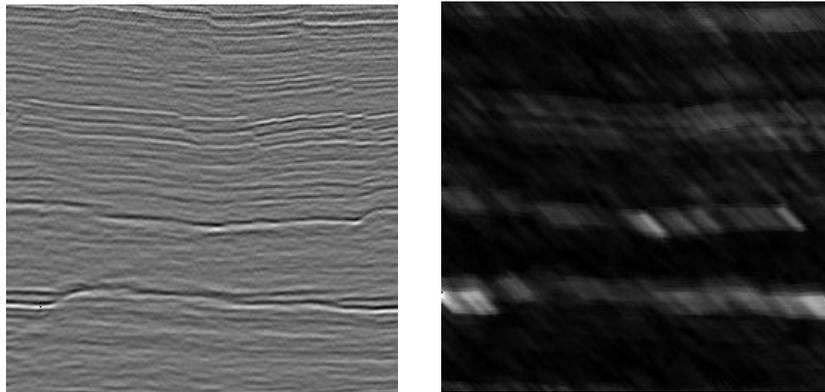


Figura 3.4: Energia Vertical.

### 3.2.2 Cálculo da Similaridade

A Figura 3.5 mostra o cálculo da similaridade vertical, que é realizado entre dois segmentos de traço, sempre definindo onde está contido o *pixel* de referência (onde é armazenado o resultado dos cálculos), do vetor  $v$  e o vetor adjacente  $u$ , ambos com o mesmo *timegate*.

Na Figura 3.5 para o *pixel* circulado em vermelho, com valor 103, foi calculado  $|v - u|$ , fazendo  $|(8 - 9), (1 - 7), (103 - 109), (6 - 7), (4 - 4)|$ , dividindo pelo  $|v| + |u|$  e subtraindo de 1 (ALMEIDA, 2007), o resultado é armazenado no *pixel* de referência e a janela vai percorrendo a imagem na vertical realizando o cálculo para os demais *pixels*.

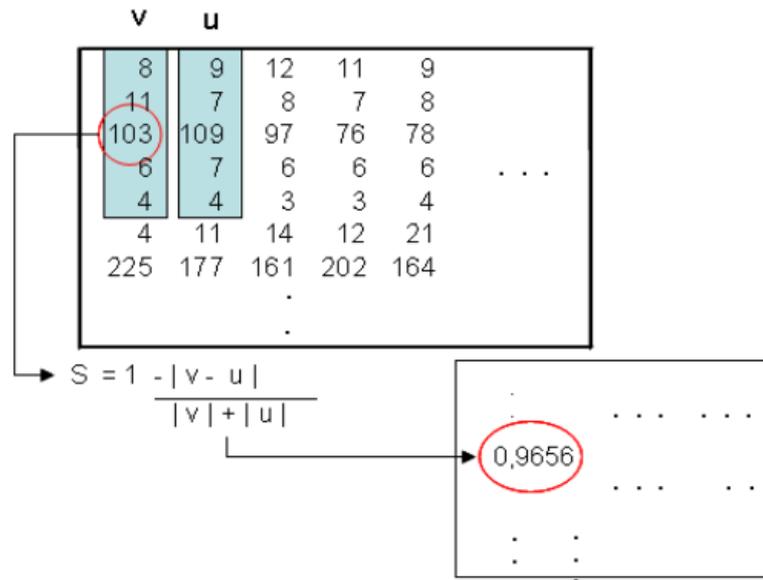


Figura 3.5: Cálculo da Similaridade Vertical (ALMEIDA, 2007).

As equações para o cálculo deste atributo foi descrita na Seção 2.2.3. A Figura 3.6 mostra um exemplo de uma imagem com o atributo de similaridade vertical calculado.

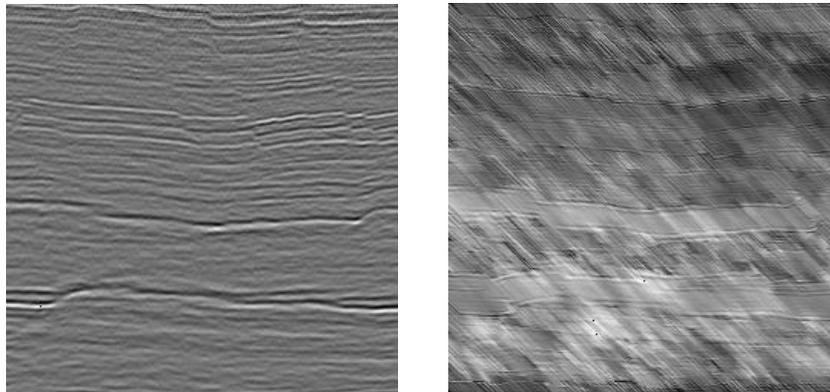


Figura 3.6: Similaridade Vertical.

### 3.2.3 Cálculo do Correlograma

O atributo geoestatístico correlograma é utilizada neste trabalho para descrever a textura de objetos representantes de falhas nas imagens sísmicas. As equações para o cálculo desta função foi descrito na Seção 2.3.1.

Os parâmetros usados pelo correlograma para extração das características em cada amostra foram:

- As direções:  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  e  $135^\circ$  com tolerância angular de  $22.5^\circ$ .
- Incremento de *lag* (distância): 1, 2 e 3 correspondendo a 15, 8 e 5 *lags*.
- Tolerância de cada distância de *lag*: 0,5, 1,0 e 1,5, a metade do incremento de *lag*, escolha mais comum (ISAACS, SRIVASTAVA, 1990).

Considerando uma imagem de  $15 \times 15$  *pixels*, que é o tamanho da janela utilizada neste trabalho, o número máximo de *lags* que teremos em qualquer uma das direções será de 15 *lags* para  $d = 1$ , 8 *lags* para  $d = 2$ , e 5 *lags* para  $d = 3$ . A Figura 3.7 mostra como essas distâncias (*lags*) e suas direções são utilizadas.

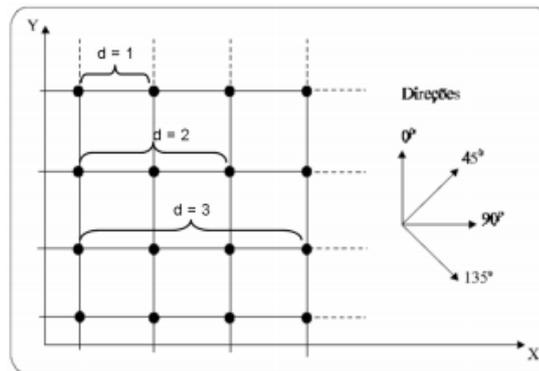


Figura 3.7: Distâncias e direções dos lags (CAMARGO, 1998).

A assinatura da amostra é representada pelo Vetor de Características e para a sua criação foram extraídas 112 características por amostra correspondentes as 4 direções de 28 *lags* ( $15 + 8 + 5$ ).

### 3.3 Classificação usando SVM

Nesta etapa fizemos a classificação de cada objeto em Falha e Não-Falha, utilizando o reconhecimento de padrões de acordo com a análise de textura obtida pelo correlograma e os atributos sísmicos Energia e Similaridade. Utilizamos também a biblioteca LIBSVM, descrito na Seção 2.4, que facilita a aplicação do SVM como uma ferramenta.

Para classificar o dado sísmico usando SVM, seguimos o seguinte fluxo de atividades:

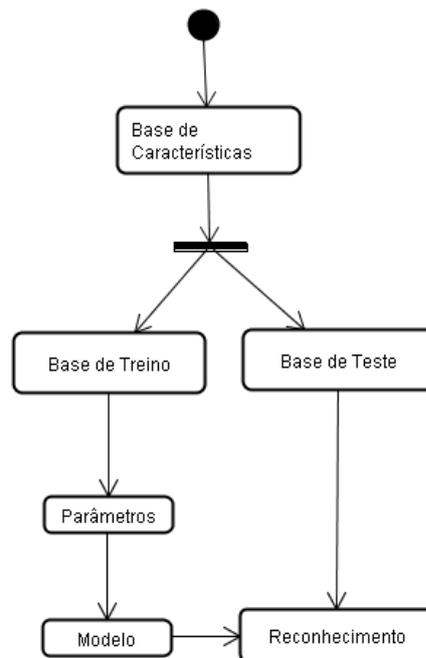


Figura 3.8: Fluxo de atividades realizadas durante a etapa de classificação (BRAZ JUNIOR, 2008).

- Geração do arquivo (base de características) no formato LIBSVM;
- Utilização do *script subset.py* que divide a base de características de forma aleatória em duas outras bases, a base de treino e base de teste;
- Utilização do *svm-scale* para normalizar a base de treino e salvar os parâmetros dessa normalização para ser utilizados na base de teste;
- Construção do modelo;
- Classificação do arquivo de teste, utilizando o *svm-predict*;

Definimos duas classes, Falha (+1) e Não-Falha (-1) e geramos arquivos de características com 1, 2 e 3 atributos, com todas as combinações possíveis entre o atributo de energia, similaridade e correlograma.

Todo o fluxo descrito anteriormente foi aplicado cinco vezes de forma aleatória sobre a base de características para garantir que a SVM não se tornasse viciada, ou seja, foram realizadas cinco análises. O próximo Capítulo apresenta os resultados alcançados com o uso da metodologia proposta.

## 4 RESULTADOS

Aplicamos os atributos energia, similaridade e correlograma para 50 fatias, dentro de um volume com 256 fatias e numerosas falhas, para cada fatia foi extraído 25 pontos de Falha e 25 pontos de Não-Falha totalizando 2500 amostras. Com o objetivo de verificar a eficiência do atributo correlograma em comparação com os outros dois atributos sísmicos calculados, indicando a eficiência do método proposto.

Os atributos de energia, similaridade e correlograma foram calculados utilizando uma janela de tamanho 15. Foram realizados testes para cada atributo e para todas as combinações possíveis entre eles, com diferentes quantidades de amostra.

Na primeira bateria de testes, separamos 50% (1250) das amostras para treinamento e 50% (1250) para teste. Na segunda bateria de testes, separamos 80/20% (2000/500) das amostras para treinamento e teste. Cada bateria de testes foi executada cinco vezes de forma aleatória sobre a base de características, ou seja, foram realizadas cinco análises.

A Tabela 4.1 mostra os resultados da acurácia para as amostras calculadas da base de teste para cada atributo separadamente, Correlograma (CR), Energia Vertical (EV) e Similaridade Vertical (SV).

Tabela 4.1: Resultados da acurácia para cada atributo.

		Análise				
Atributo	%	1ª	2ª	3ª	4ª	5ª
CR	50/50	87,2	88,56	86,96	85,68	89,68
	80/20	92,8	89,2	90,6	90,0	91,6
EV	50/50	88,0	89,12	88,96	89,12	88,64
	80/20	90,4	89,8	90,0	89,2	88,0
SV	50/50	53,44	55,2	53,2	51,44	55,12
	80/20	51,4	57,4	56,0	53,2	57,0

Analisando os resultados apresentados na Tabela 4.1, verificamos que os melhores resultados foram alcançados para análises onde 80% das amostras foram utilizadas no treinamento e 20% nos testes.

A Tabela 4.2 apresenta os resultados obtidos para as combinações entre os atributos, CR com EV, CR com SV, EV com SV e os três juntos.

Tabela 4.2: Resultados da acurácia para associação dos atributos.

		Análise				
Atributo	%	1ª	2ª	3ª	4ª	5ª
CR / EV	50/50	92,56	93,52	92,88	93,36	93,52
	80/20	93,4	94,6	94,8	96,0	93,6
CR / SV	50/50	87,12	89,28	87,76	87,04	88,08
	80/20	89,8	91,4	90,6	89,8	90,6
EV / SV	50/50	89,2	89,28	88,88	89,52	88,4
	80/20	88,4	90,2	87,2	90,2	90,0
CR / EV / SV	50/50	94,0	95,04	94,08	93,28	93,04
	80/20	95,0	94,4	<b>96,6</b>	95,6	95,2

Avaliando os resultados apresentados na Tabela 4.2, observamos que os melhores resultados foram os encontrados com a associação dos 3 atributos, sendo que o melhor resultado encontrado atingiu 96.6%, como destacado na Tabela 4.2.

A Tabela 4.3 mostra o cálculo da média (M) e a medida de dispersão: desvio padrão (Dp) para os valores de acurácia dos atributos. É mostrado ainda o melhor valor encontrado.

Tabela 4.3: Média e desvio padrão para os melhores valores de acurácia.

<b>Atributo</b>	<b>%</b>	<b>Melhor acurácia</b>	<b>Média</b>	<b>Desvio Padrão</b>
CR	50/50	89,68%	87,616%	1,37
	80/20	92,8%	90,84%	1,25
EV	50/50	89,12%	88,768%	0,42
	80/20	90,4%	89,48%	0,83
SV	50/50	55,2%	53,68%	1,39
	80/20	57,4%	55,0%	2,32
CR / EV	50/50	93,52%	93,168%	0,38
	80/20	96,0%	94,48%	0,93
CR / SV	50/50	89,28%	87,856%	0,81
	80/20	91,4%	90,44%	0,59
EV / SV	50/50	89,52%	89,056%	0,38
	80/20	90,2%	89,2%	1,20
CR / EV / SV	50/50	95,04%	93,888%	0,70
	80/20	96,6%	95,36%	0,73

A Tabela 4.4 mostra os valores de sensibilidade e especificidade para os melhores resultados mostrados na Tabela 4.3.

Tabela 4.4: Sensibilidade, especificidade e acurácia dos melhores resultados.

<b>Atributo</b>	<b>%</b>	<b>Acurácia (%)</b>	<b>Sensibilidade (%)</b>	<b>Especificidade (%)</b>
CR	80/20	92,8%	91,66%	90,8%
EV	80/20	90,4%	92,33%	85,81%
SV	80/20	57,4%	55,15%	63,12%
CR / EV	80/20	96,0%	97,13%	94,92%
CR / SV	80/20	91,4%	89,96%	92,94%
EV / SV	80/20	90,2%	97,18%	85,01%
CR / EV / SV	80/20	96,6%	96,8%	93,63%

Analisando os dados da Tabela 4.4, podemos observar que o melhor resultado encontrado nos testes foi alcançado associando os atributos energia, similaridade e correlograma, onde a sensibilidade é igual a 96,8% , especificidade é igual a 93,63% , e acurácia 96,6%, indicando a eficiência da utilização do atributo geoestatístico correlograma.

## 5 CONCLUSÃO

Para a indústria de petróleo, a interpretação sísmica visa identificar estruturas onde o óleo possa estar aprisionado em um reservatório. Com a identificação dessas estruturas pode-se evitar prejuízos e aumentar o lucro, além de reduzir o impacto ambiental com perfurações desnecessárias.

Portanto, antes de iniciar a perfuração de um poço é necessário coletar o máximo de informações possíveis sobre as condições geológicas da área em questão. A complexidade do sistema geológico, a qualidade dos dados sísmicos, o bom processamento, e a experiência do intérprete são fatores que influenciam diretamente no resultado da interpretação.

Foi apresentado, neste trabalho, uma investigação sobre a aplicabilidade do atributo geoestatístico correlograma como descritor de característica para a classificação de falhas geológicas em dados sísmicos através do aprendizado de máquina. Na busca pela melhor solução, aplicamos ainda os atributos sísmicos de energia e similaridade a fim de comparar com os resultados obtidos pelo correlograma e constatou-se que os melhores resultados foram obtidos com a associação destes dois atributos com o atributo geoestatístico.

O método apresentado aqui mostra que falhas geológicas podem ser classificadas utilizando os atributos sísmicos existentes na literatura juntamente com o atributo correlograma, como foi fundamentado pelos resultados encontrados, no melhor caso onde a sensibilidade foi igual a 96,8%, a especificidade a 93,63%, e a acurácia a 96,6%, indicando a eficiência da utilização do atributo geoestatístico correlograma.

No entanto, vê-se necessário a realização de testes mais profundos, com uma quantidade maior de volumes sísmicos, de dados reais, e com diferentes características de falhas, ou seja, para que seja possível comprovar a eficiência da técnica é necessário aplicá-la à imagens mais complexas.

Observa-se que a indústria petrolífera necessita de um conjunto de ferramentas que dê suporte com precisão a análise de dados sísmicos. Com isso a proposta para trabalhos futuros é associação do atributo correlograma com outras

funções geoestatísticas (semivariograma, semimadograma e covariograma) e com outros atributos sísmicos, como por exemplo, o semblance, o cubo de variância, o cubo de coerência por autoestrutura e outros. Outra proposta é realizar a análise para volumes de dados sísmicos 3D.

## 6 REFERÊNCIAS

ALMEIDA, J. D. S de. Detecção de Falhas em Dados Sísmicos usando Máquinas de Suporte Vetorial, Monografia, UFMA, São Luís, MA, 2007.

ALMEIDA, J. D. S de. Metodologia Computacional para Detecção Automática de Estrabismo em Imagens Digitais através do Teste de Hirschberg, Dissertação de Mestrado, UFMA, São Luís, MA, 2010.

BUSTAMANTE, S. G. H. Inversão de Parâmetros Geofísicos em Três Dimensões a partir de Dados de Reflexão Sísmica por Algoritmos Genéticos Híbridos. Dissertação de Mestrado. PUC – Rio, Rio de Janeiro, 2008.

BRAZ JUNIOR, G. Classificação de Regiões de Mamografias em Massa e Não Massa Usando Estatística Espacial e Máquina de Vetores de Suporte. Dissertação de Mestrado, UFMA, São Luís, MA, 2008.

CAMARGO, E. C. G. Curso de Geoprocessamento para Projetos Ambientais do Instituto Nacional de Pesquisas Espaciais – INPE. Cap. 5. São José dos Campos, SP, 1998

GERHARDT, A.. Aspectos da Visualização Volumétrica de Dados Sísmicos. Dissertação de Mestrado, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, RJ, 1998.

HAYKIN, S. Neural Networks: A Comprehensive Foundation, 1999.

ISAACS, E. H.; SRIVASTAVA, R. M. An Introduction to Applied Geostatistics, 1990.

JOHANN, P.; RIBET, B. D. Multivariate Seismic Pattern Recognition and Kohonen Maps Applied on a Deep-water Turbidite Reservoir in Campos Basin, Brazil. In: Internacional Congress of the Brazilian Geophysical Society, Rio de Janeiro, 1999.

KRIGE, D. G. A Statistical Approach to Some Mine Valuations Problems At The Witwatersrand. Journal of the Chemical, Metallurgical and Mining Society of South Africa, 1951.

KEERTHI, S. S.; LIN C. J. Asymptotic Behaviors of Support Vector Machines with Gaussian Kernel. Taiwan, 2003.

LORENA, A. C.; CARVALHO, A. As máquinas de Vetores Suporte, relatório interno. Laboratório de Inteligência Computacional, ICMC/USP, São Carlos - SP, 2002.

MATOS, M. C. de. Reconhecimento de Padrões Sísmicos Utilizando Análises Tempo-frequência. Tese doutorado, Pontifícia Universidade Católica do Rio de Janeiro, RJ, 2004.

MITCHELL, T. Machine Learning. McGraw Hill, New York, 1997.

MACHADO, M. de C. Determinação de Malhas de Falhas em Dados Sísmicos por Aprendizado Competitivo. Tese de Doutorado. PUC-Rio. Rio de Janeiro, 2008.

MACHADO, M.; VELLASCO, M.; SILVA, P.M.; GATTASS, M. Using neural networks to evaluate the effectiveness of a new seismic fault attribute. In: Biannual Brazilian Symposium On Artificial Neural Networks, SBRN'2006, 2006. Conference Proceedings.

MACHADO, M.; VELLASCO, M.; SILVA, P.M.; GATTASS, M. Providing Spatial Coherence Information to Fault Meta-Attribute. In: 76TH Annual International Meeting, SEG, 2006. Expanded Abstracts.

MULLER, K. R.; MIKA, S.; RATSCH, G.; TSUDA, K.; SCHOLKOPF, B. An Introduction to Kernel-based Learning Algorithms. IEEE Transactions on Neural Networks, p. 181–201. 2001.

NADLER, M.; SMITH, E. P. Pattern Recognition Engineering. New York: Wiley, 1993.

PGS, Geofísica Sísmica 3D. 2010. Disponível em: <<http://www.pgs.com/>>. Acesso em: fevereiro de 2013.

ROBINSON, E. A.; TREITEL, S. Geophysical Signal Analysis. Prentice-Hall, 1980.

RUSSELL, S.; NORVIG, P. Artificial Intelligence: A Modern Approach, 1995.

RUSSO, E. E. R.; RAPOSO, A. B.; TERRENCE, F.; GATTASS, M. A Realidade Virtual na Indústria de Exploração e Produção de Petróleo. In: Realidade Virtual: Conceitos e Tendências - Livro do Pré-Simpósio SVR 2004. São Paulo: Mania de Livro. 2004. Cap. 21, p. 283-288.

SILVA, P. M., Visualização Volumétrica de Horizontes em Dados Sísmicos 3D. Tese de Doutorado, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2004.

SILVA, A. C.; GATTASS, M.; CARVALHO, P. C. P. Analysis of Spatial Variability Using Geostatistical Functions for Diagnosis of Lung Nodule in Computerized Tomography Images. Pattern Analysis and Applications, 2004.

SILVA, P. M. C. e. Visualização Volumétrica de Horizontes em Dados Sísmicos 3D. Tese de Doutorado, Pontifícia Universidade Católica do Rio de Janeiro, 2004.

SOUSA JUNIOR, O. S. de. Método de Reconhecimento Pessoal Através da Íris Usando Funções Geoestatísticas. Dissertação de Mestrado, UFMA, São Luís, MA, 2006.

SMOLA, A. J.; BARLETT, P.; SCHOLKOPF, B.; SCHUURMANS, D. Introduction to Large Margin Classifiers, Cap. 1, p. 1–28. 1999.

TECNO PETRO BR. Sísmica de Reflexão. 2008. Disponível em: <<http://tecnopetrobr.blogspot.com.br/2008/11/ssmica-de-reflexo.html>>. Acesso em: fevereiro de 2013.

TEIXEIRA, W.; TOLEDO, M. C. M. de; FAIRCHILD, T. R.; TAIOLI, F. (Orgs.)  
Decifrando a Terra. São Paulo: Oficina de Textos, 568 p, 2000.

VAPNIK, V. N. The Nature of Statistical Learning Theory. Springer Verlag, New York,  
1995.