

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

AFONSO PEREIRA LOPES

**DETECÇÃO DE FALHAS GEOLÓGICA EM DADOS SÍSMICO
USANDO SEMIVARIOGRAMA E APRENDIZADO DE MÁQUINA**

São Luis

2013

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

AFONSO PEREIRA LOPES

**DETECÇÃO DE FALHAS GEOLÓGICA EM DADOS SÍSMICO
USANDO SEMIVARIOGRAMA E APRENDIZADO DE MÁQUINA**

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

Orientador: Prof. Dr. Anselmo Cardoso Paiva

São Luis
2013

Lopes, Afonso Pereira

Detecção de falhas geológicas em dados sísmicos usando semivariograma e aprendizado de máquina / Afonso Pereira Lopes. – 2013.

53f.

Orientador: Anselmo Cardoso Paiva.

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

1.Dados sísmicos 2.Falhas geológicas 3.Aprendizado de máquina 4.SVM 5.Função geoestatística 6.Semivariograma I. Título.

CDU 004.9:550.832.4

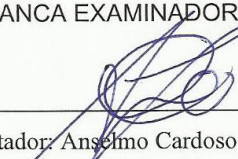
Afonso Pereira Lopes


**Detecção de Falhas Geológicas em Dados Sísmico usando
Semivariograma e Aprendizado de Máquina**

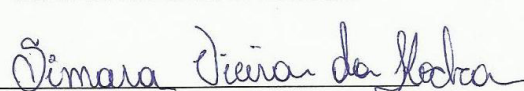
Monografia apresentada ao curso de Ciência da
Computação da Universidade Federal do Maranhão,
como parte dos requisitos necessários para obtenção
do grau de Bacharel em Ciência da Computação.

Aprovada em: 13/03/2013

BANCA EXAMINADORA


Orientador: Anselmo Cardoso Paiva
Prof. Dr. em Informática – UFMA
Universidade Federal do Maranhão


Geraldo Braz Junior
Prof. em Informática – UFMA
Universidade Federal do Maranhão


Simara Vieira da Rocha
Prof. em Informática – UFMA
Universidade Federal do Maranhão

AGRADECIMENTOS

Agradeço primeiramente a Deus, pois sem ele não teria mais forças para essa longa jornada;

Ao meu pai, amigo de sempre, homem pelo qual tenho maior orgulho, que me deu todo apoio e me fez acreditar que nada é impossível, pessoa que sigo como exemplo, pai dedicado, batalhador, pessoa pelo qual mostrou total esforço para formação de seus filhos, que espero hoje, ser motivo de mais orgulho para ele;

A minha mãe, por ser a pessoa que também me apóia e acredita na minha capacidade, meu agradecimento pelas horas em que ficou ao meu lado não me deixando desistir e me mostrando que sou capaz de chegar onde desejo, sem dúvida foi quem me deu o maior incentivo para conseguir concluir esse trabalho;

Aos meus irmãos Adriano Lopes e Andrea Lopes, pelos conselhos e apoio;

Aos professores, orientador Dr. Anselmo Cardoso Paiva, Dr. Aristófanês Corrêa Silva, Geraldo Braz Junior e Simara Vieira da Rocha pela orientação, dedicação, amizade, incentivo, conselhos, apoio e principalmente a paciência durante estes anos de convivência e trabalho;

Aos amigos que fiz durante o curso, pela verdadeira amizade que construímos em particular aqueles que estavam sempre ao meu lado (Antonio Gomes, Pablo Fernandes, Jefferson Sá, Marcio Frazão) por todos os momentos que passamos durante esses cinco anos meu especial agradecimento. Sem vocês essa trajetória não seria tão prazerosa;

Aos que considero meus amigos irmãos, Guilherme Lima, Carlos Junior, Paulo Victor, Ricardo Campelo e Flávio Rodrigues pela força, pela amizade;

A minha namorada, Cleane Dantas, que apareceu em minha vida no meio da minha jornada, no momento mais delicado, e que me deu forças para continuar, mostrando-me que era capaz;

A todos os professores do curso Ciência da Computação, pelos ensinamentos disponibilizados nas aulas, cada um de forma especial contribuiu para a conclusão desse trabalho e conseqüentemente para minha formação profissional;

Ao Programa de Recursos Humanos da ANP – PRH 39 pela bolsa de iniciação científica concedida.

Por fim, aos meus familiares que até aqui não citei, pelo carinho e pela compreensão nos momentos em que a dedicação aos estudos foi exclusiva, a todos que contribuíram direta ou indiretamente para que esse trabalho fosse realizado, meu eterno AGRADECIMENTO.

RESUMO

O método direto de identificação de reservatórios de petróleo envolve altos custos para a indústria petrolífera, esse método consiste na perfuração de poços, isso faz com que tenham uma crescente busca por métodos indiretos que possam identificar uma região onde esse hidrocarboneto possa está armazenado. O método indireto mais utilizado é a sísmica de reflexão.

As falhas geológicas são um dos fatores extremamente importantes para a identificação de estruturas de aprisionamento assim como na compressão do fluxo de fluídos em um reservatório. Com isso, objetivo deste trabalho é desenvolver um método usando a função Semivariograma como descritores de característica e aprendizado de máquina, que auxilie o intérprete na geração de um modelo que melhor representa a estrutura geológica para detecção de falhas em dados sísmicos. O método obteve resultados de acurácia igual a 95,2%, sensibilidade igual a 95,22% e especificidade igual a 96,49%.

Palavras-chave: Dados sísmicos. Falhas geológicas. Aprendizado de Máquina. MVS. Função Geoestatística. Semivariograma.

ABSTRACT

The direct method of identifying oil reservoirs involves high costs for the oil industry, this method consists in drilling wells, it makes you have a growing search by indirect methods that can identify a region where this is stored hydrocarbon can. The indirect method is the most widely used seismic reflection.

The faults are an extremely important factor for the identification of trapping structures as well as compression of fluid flow in a reservoir. With this objective is to develop a method using the function as Semivariogram feature descriptors and machine learning, the interpreter to assist in the generation of a model that best represents the geological structure for fault detection in seismic data. The method was accuracy equal to 95,2%, the sensitivity was 95,22% and specificity of 96,49%.

Keywords: Seismic data. Geological faults. Machine Learning. MVS. Function geostatistics. Semivariogram.

Sumário

Lista de Figuras	10
Lista de Tabelas	11
1. Introdução	12
2. Fundamentação Teórica	14
2.1. Dados Sísmicos.....	14
2.1.1. Método Sísmico de Reflexão	15
2.1.2. Aquisição Sísmica.	17
2.1.3. Processamento Sísmico.	19
2.1.4. Interpretação Sísmica.....	22
2.2. Atributos Sísmicos.....	24
2.2.1. Importância dos atributos sísmicos na geofísica.....	25
2.2.2. Atributo de Energia.....	26
2.2.3. Atributo de Similaridade.....	26
2.3. Funções Geoestatísticas	27
2.3.1. Semivariograma	28
2.4. Máquina de Vetores Suporte	31
2.4.1. Avaliação dos Resultados.....	35
3. Materiais e Métodos	36
3.1. Aquisição e tratamento das imagens sísmicas	36
3.2. Marcações de pontos de Falhas e Não – Falhas	37
3.3. Cálculo de atributos sísmicos de Falha.....	38
3.3.1. Calculo da Energia Vertical.....	38
3.3.2. Calculo da Similaridade Vertical	39
3.3.3. Calculo do Atributo de Semivariograma.....	41
3.4. Classificação usando MVS.	42
4. Resultados Obtidos	44
5. Conclusão	48
Referências Bibliográficas	50

Lista de Figuras

Figura 2.1 – Estrutura capeadora convexa.	17
Figura 2.2 – Falha Sísmica.	17
Figura 2.3 – Incidência de onda de um meio para outro.	19
Figura 2.4 – Exploração Sísmica [Robinson & Treitel, 1980].....	19
Figura 2.5 – Esquema de aquisição de dado sísmico marítimo [Schlumberger Oilfield Glossary, 2003].....	20
Figura 2.6 – Aquisição Sísmica Terrestre.	20
Figura 2.7 – Esquema de aquisição 3D marítima [Sercel, 2003].	21
Figura 2.8 – Traço sísmico (esquerda), linha sísmica (direita) (Adptada de [Kristofer M. Tingdahl & Matthijs de Rooij]).	23
Figura 2.9 – Volume sísmico (Adptada de [Kristofer M. Tingdahl & Matthijs de Rooij]).	23
Figura 2.10 – Horizonte Sísmico [Robson & Treitel, 1980].....	23
Figura 2.11 – Falha Sísmica [Robson & Treitel, 1980].	23
Figura 2.12 – Atributos utilizados por Tingdahl [Tingdahl & de Rooij, 2005].	25
Figura 2.13 – Definição de vetor distância.	29
Figura 2.14 – Características do semivariograma.	29
Figura 2.15 – Parâmetros utilizados para cálculo da função geoestatística [Deutsch & Journel, 1992].....	30
Figura 2.16 – Hiperplano ótimo.....	33
Figura 3.1 – Fatia gerada a partir dos dados de amplitude.	397
Figura 3.2 – Arquivo contendo pontos de Falhas e Não-Falhas.....	397
Figura 3.3 – Cálculo da Energia Vertical.....	38
Figura 3.4 – Energia Vertical.....	39
Figura 3.5 – Cálculo de Similaridade Vertical.....	40
Figura 3.6 – Similaridade Vertical.	40
Figura 3.7 – Distância e direções dos lags. Adaptada : [CAMARGO,1988].....	41
Figura 4.1 – Resultados obtidos nos testes.	47

Lista de Tabelas

Tabela 4.1 – Resultados da acurácia para 2500 amostras, calculada para cada atributo e suas associações.....	44
Tabela 4.2 – Desvio padrão dos atributos e suas associações.....	45
Tabela 4.3 – Resultados de sensibilidade e especificidade dos atributos e suas associações.....	46

1. Introdução

O Petróleo é a principal fonte de energia do mundo, identificar uma região onde possa estar armazenado não é uma tarefa simples, ela envolve um estudo do comportamento das diversas camadas do subsolo através de métodos geológicos e geofísicos.

Para o acúmulo do petróleo ou gás numa bacia sedimentar, é necessário que o petróleo se aloje em uma estrutura localizada na parte mais alta de um compartimento de rocha porosa, sendo isolada por camadas impermeáveis. Essa estrutura é denominada de armadilha ou trapa, que são resultantes de modificações sofridas pelas rochas ao longo do tempo geológico, especialmente a sua deformação, através do desenvolvimento de dobras ou falhas na crosta terrestre.

As falhas podem ser extremamente importantes na identificação de trapas de origem estrutural. Deslocando uma camada impermeável, uma falha pode criar uma trapa estrutural onde o óleo é aprisionado, ou pode romper a capa selante de uma trapa, gerando um vazamento [Lines et al., 2004]. De forma semelhante, as falhas também são importantes na compreensão do fluxo de fluidos em um reservatório. Elas podem agir 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, 2007].

Existem vários métodos de identificação destes reservatórios. O método direto consiste na perfuração de poços, que fornecem apenas dados pontuais e envolve altos custos, o que se torna uma desvantagem para a indústria petrolífera. No entanto, os métodos indiretos de investigação possuem custos consideravelmente menores e abrange alvos de investigações bem maiores, com grandes profundidades e maior resolução dos dados. Entre os mais utilizados encontra-se a sísmica de reflexão.

Obter o máximo de informações possíveis das condições geológica da área antes de iniciar a perfuração de um poço é de fundamental importância,

logo que, cada perfuração de poço de petróleo pode atingir, em uma estimativa, até 85% do custo total da exploração. É relevante saber como se apresenta toda a estrutura do reservatório, numa região onde há suspeita da existência desse hidrocarboneto, utilizam-se técnicas para detectá-los e faz-se, então, um estudo de viabilidade para evitar prejuízos e garantir máxima lucratividade. A partir daí, os geofísicos desenvolvem o processo de busca inicial por hidrocarbonetos através da exploração sísmica.

A confiança no resultado da interpretação depende da complexidade do sistema geológico, da qualidade dos dados sísmicos, do bom processamento, e da experiência do intérprete [Rankey & Mitchell, 2003]. Logo, qualquer erro de interpretação pode levar a resultados incoerentes.

A busca por uma representação adequada dos reservatórios de petróleo através de dados sísmicos, utilizando técnicas de reconhecimento de padrões, tem sido alvo de inúmeras publicações científicas [Johann & Ribet, 1999], o que nos motivou a utilizar as técnicas de aprendizado de máquina.

Contudo, este trabalho visa o estudo da função geoestatística de semivariograma como descritor de características que possam levar a identificação em volumes de dados sísmico e aplicar técnicas de reconhecimento de padrões e aprendizado de máquina para o reconhecimento de falhas geológicas. Para isto, este trabalho está estruturado em 5 capítulos. O Capítulo 2 apresenta a Fundamentação Teórica. O Capítulo 3, a metodologia utilizada para chegar à solução. O Capítulo 4 os resultados apresentados. E por fim, no Capítulo 5 fazemos as considerações finais do trabalho realizado.

2. Fundamentação Teórica

Neste capítulo serão dados alguns conceitos que serão importantes na compreensão dos capítulos seguintes. Aborda-se o conceito de dados sísmico, sísmica de reflexão, aquisição, processamento de imagem, atributos sísmico, método de análise de textura baseado em medidas geoestatísticas: semivariograma e método de classificação e reconhecimento de padrão utilizando Máquinas de Vetores de Suporte.

2.1. Dados Sísmicos

A grande vantagem do método sísmico de reflexão está na alta taxa de aquisição de dados em um curto período de tempo. Os dados, após serem devidamente processados, fornecem informações valiosas sobre a área investigada, permitindo assim que o número e localização de amostras sejam determinados do modo mais eficiente, reduzindo os custos finais do projeto.

Para acúmulo do petróleo é necessário que exista rocha denominada de geradora, assim chamada por se tratar de um mineral formado principalmente pelo acúmulo de fragmentos de outros minerais e detritos orgânicos. Segundo Gehardt [Gehardt, 1998], a grande maioria dos acúmulos de petróleo (hidrocarbonetos como óleo e gás) se encontra em rochas sedimentares.

Com a sua formação, para que se acumule futuramente, é necessária que ocorra a migração do petróleo, devida a sua densidade inferior que a da água, da rocha geradora para outra rocha, porosa e permeável, denominada de rocha reservatório, e que continue seu fluxo no interior da mesma até que encontre uma armadilha, isto é, uma estrutura geológica compreendida dentro de uma rocha selante (impermeável), que interrompa o seu fluxo, formando um reservatório.

Não havendo a presença de uma rocha selante e de uma armadilha (trapa), o petróleo não se acumularia e escaparia pela superfície da terra. As Figuras 2.1 e 2.2 ilustram dois tipos de reservatórios, o primeiro associado a uma estrutura capeadora do tipo convexo (anticlinal) e a seguinte com um deslocamento linear (falha geológica).

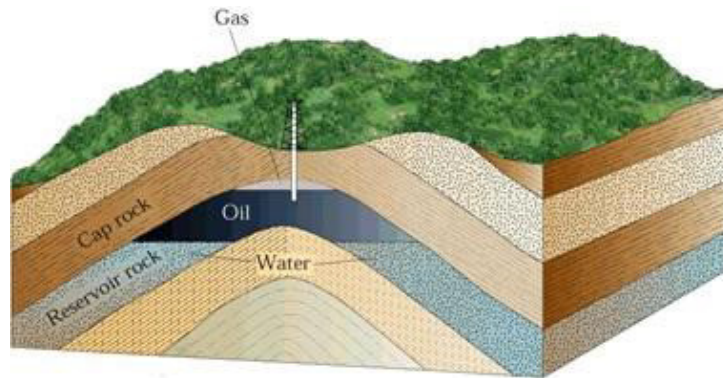


Figura 2.1 – Estrutura capeadora convexa.

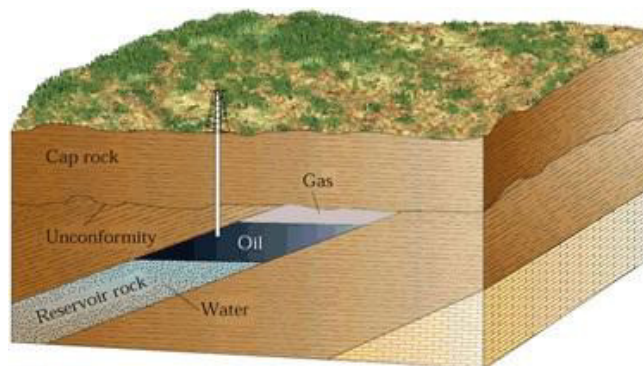


Figura 2.2 – Falha Sísmica.

2.1.1. Método Sísmico de Reflexão

O método sísmico tem como objetivo, a criação de modelos de dados que contenham informações fundamentais a respeito da geologia terrestre onde foram aplicadas. O método sísmico de reflexão vem sendo usada

intensivamente pela indústria de petróleo, por possuir várias vantagens tais como altas definições das configurações geológicas e a ampla abrangência de grandes áreas de aquisição [Dallyson, 2007].

Este conceito é baseado na propagação de ondas sísmicas, que se propagam através da terra de forma artificial causadas por explosões. Estas são estudadas pelos sismólogos, e medidas por sismógrafos, sismômetros ou geofones. Também podendo ser utilizadas os hidrofones.

No estudo da sísmica dois tipos de ondas são de fundamental importância, as ondas primárias (ondas P) e as secundárias (ondas S). As ondas P, conhecidas como “*compressional waves*” têm maior velocidade de propagação, que varia de acordo com o meio em que se propagam, sendo característicos os valores de 330 m/s no ar, 1450 m/s e 5000 m/s no granito. São ondas longitudinais que fazem a rocha vibrar paralelamente em direção a onda e não destrutivas como as ondas S ou as ondas de superfície que se seguem.

As ondas S (shear or rotational waves) têm velocidade de propagação ligeiramente superior a metade das ondas P. São ondas transversais ou de cisalhamento, deslocando o solo de forma perpendicularmente à direção de propagação como um chicote. Elas provocam alterações morfológicas sem alterações de volume.

Como ilustrada na Figura 2.3, quando uma frente de onda sísmica atinge a interface entre dois materiais distintos, parte da energia é refletida para o meio no qual a onda estava se propagando, e parte é transmitida para o outro meio. Ilustrada na Figura 2.3.

A mudança de velocidade de propagação das ondas causadas pela mudança de um meio para outro, é uma informação utilizada na sísmica. A complexidade da descrição destas ondas em meios heterogêneos é extremamente alta, por isso, para se obterem equações que descrevam essa propagação adequadamente, é necessário admitir condições simplificadoras.

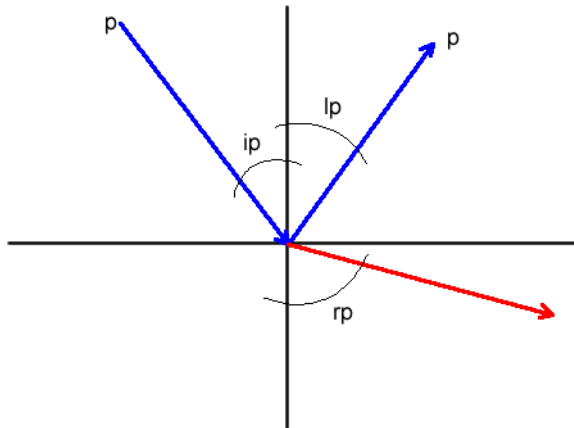


Figura 2.3 – Incidência de onda de um meio para outro.

Segundo Robinson e Treitel [Robinson & Treitel, 1980], o método sísmico utilizado na exploração de hidrocarboneto, óleo e gás, baseado em sísmica pode ser dividido em três etapas: aquisição, processamento e interpretação. Conforme Figura 2.4.

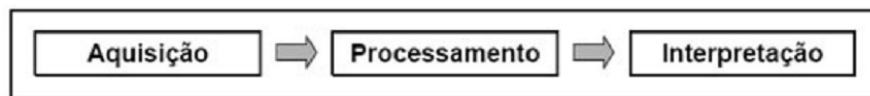


Figura 2.4 – Exploração Sísmica [Robinson & Treitel, 1980].

2.1.2. Aquisição Sísmica.

Na aquisição sísmica se faz necessária a existência de dois componentes essenciais, que são as fontes geradoras de ondas e os receptores, que captam as ondas na superfície. Fontes geradoras artificiais são utilizadas para geração de ondas mecânicas, as mais utilizadas são explosões de dinamites, no caso terrestre, ou canhões pneumáticos, no caso marítimo.

As vibrações causadas pelas fontes geradoras viajam ao longo das subsuperfícies até que alcancem uma interface entre duas camadas no qual parte é refletida, sendo captadas pelos receptores tanto o tempo de chegada

quanto a intensidade medida neste momento registrada no sismógrafo, e parte é refratada continua a viajar no meio da nova camada.

Os receptores podem ser geofones, que medem a velocidade do pulso sísmico que retorna a superfície, na aquisição de dados terrestres, ou hidrofones que medem a variação de pressão gerada na água [Machado, 2000]. A Figura 2.5 mostra o esquema da aquisição sísmica marítima, seguido da Figura 2.6 que mostra o esquema em aquisição terrestre.

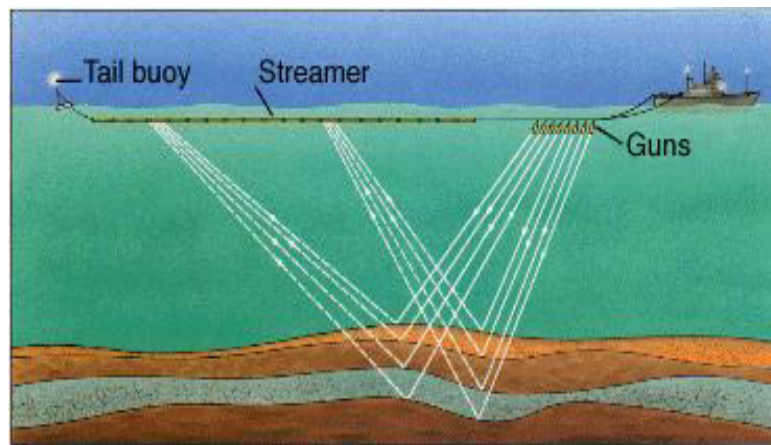


Figura 2.5 – Esquema de aquisição de dado sísmico marítimo [Schlumberger Oilfield Glossary, 2003].

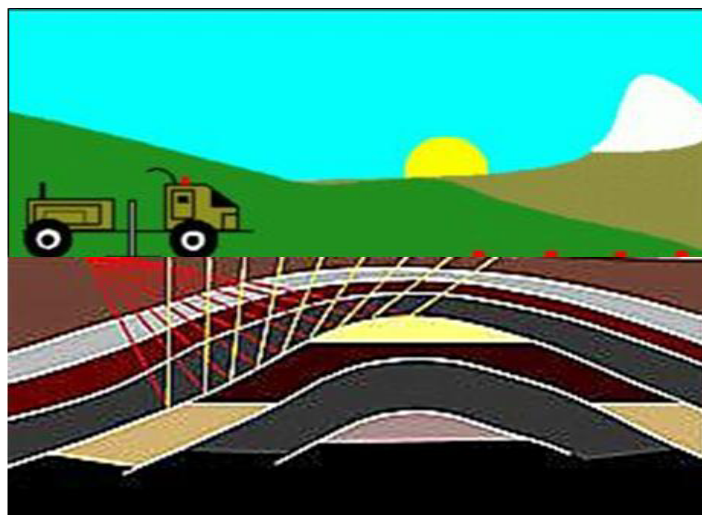


Figura 2.6 – Aquisição Sísmica Terrestre.

A aquisição sísmica marítima é feita por navios equipados com canhões pneumáticos de ar comprimido, que disparam. A energia criada por esses canhões se propaga até o fundo do mar, sob a forma de ondas acústicas. Ao atingirem o fundo do mar, essas ondas sofrem um fenômeno físico denominado “partição de energia”, onde parte da energia é refletida e parte é refratada. A parte refratada continua a se propagar na camada subsequente até sofrerem o mesmo fenômeno, sendo que a energia refletida por estas são captadas pelos hidrofones dispostos em intervalos regulares dos cabos sismógrafos, que são cabos preenchidos com fluídos de baixa densidade para facilitar a flutuação, com comprimentos que podem variar entre 4 e 16 km como podemos ver na Figura 2.7.

Ao serem captadas pelos receptores, no caso os hidrofones, a energia, é posteriormente convertida em sinais elétricos, enviados para o sistema de registro e processamento, que serão interpretados, permitindo assim, que as estruturas geológicas com condições de possuir acumulações de petróleo ou gás sejam identificadas.

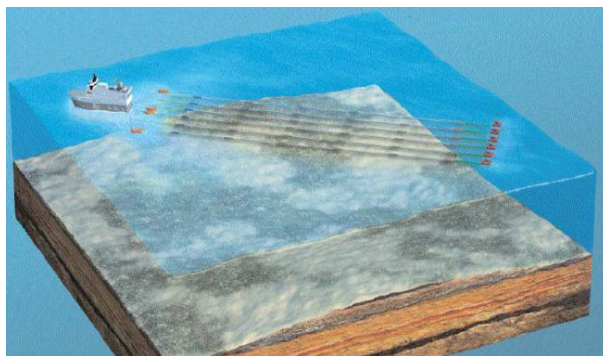


Figura 2.7 – Esquema de aquisição 3D marítima [Sercel, 2003].

2.1.3. Processamento Sísmico.

Depois de feita a coleta dos dados sísmicos, os mesmos são levados para o chamado processamento, que consiste numa aplicação de conjunto de técnicas aos dados adquiridos com intuito de se obter uma imagem representativa do subsolo. Essas técnicas têm por objetivo melhorar a

qualidade do sinal e reduzir o ruído, corrigindo os erros inerentes pelo processo de aquisição.

Nessa etapa os dados são reorganizados para formarem uma fatia (slice), no caso 2D, ou em grades tridimensionais. Nas fatias, as amostras de amplitudes sísmicas são representadas pelos pontos do plano. Nas imagens 3D uma grade tridimensional com uma amostra de amplitude sísmica de cada vértice da grade (voxel) em que as posições das fontes e receptores estão relacionadas a duas dimensões do conjunto de dados e a terceira dimensão do conjunto de dados é a temporal e a propagação da onda é feita apenas na direção vertical. Assim como a fonte e o receptor estão na mesma posição na superfície, o tempo de cada amostra corresponde ao tempo que a onda leva para viajar até uma interface mais o tempo de volta à superfície.

A imagem sísmica obtida em cada ponto receptor da superfície é composta por um respectivo conjunto de valores verticais de amplitudes, formando uma coluna. Essa coluna de amostras com as mesmas coordenadas espaciais, variando apenas o tempo, é denominada de traço sísmico. Na Figura 2.8 mostra o exemplo de traço sísmico, a esquerda, como função de amplitudes sísmicas, no caso o traço de amplitudes é representado como um sinal ondulante. Do lado direito da mesma figura, temos o que chamamos de fatia (*slice*), formado por um conjunto de traços sísmicos, que é chamada de linha sísmica, com uma dimensão espacial e outra temporal.

Já os volumes sísmicos, são formadas por varias linhas sísmicas. Neste caso, temos duas direções espaciais, que são as chamadas de *inline* (direção das linhas sísmicas) e *crossline* (direção perpendicular a linha sísmica), além de uma direção temporal, Figura 2.9.

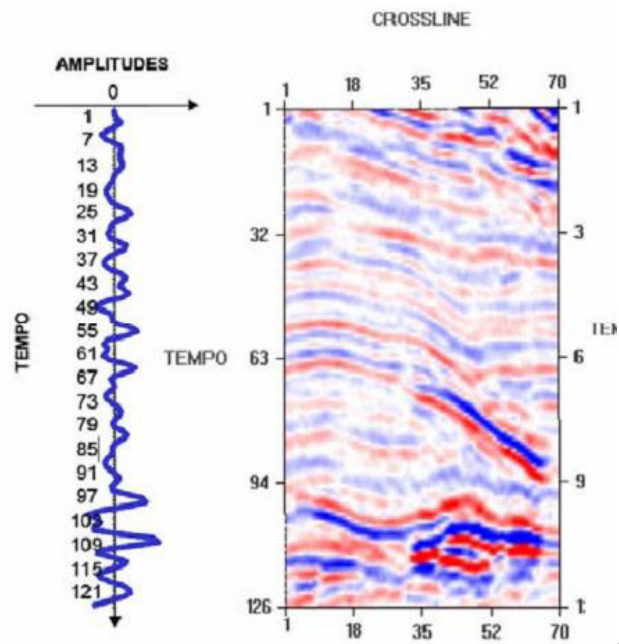


Figura 2.8 – Traço sísmico (esquerda), linha sísmica (direita) (Adptada de [Kristofer M. Tingdahl & Matthijs de Rooij]).

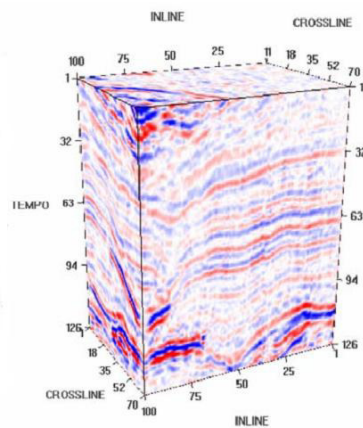


Figura 2.9 – Volume sísmico (Adptada de [Kristofer M. Tingdahl & Matthijs de Rooij]).

As imagens sísmica 2D e 3D são formadas a partir de uma escala de cores, onde cada cor representa a amplitude da onda em cada ponto, deixando de ser feita a partir de um sinal ondulante como no traço sísmico.

2.1.4. Interpretação Sísmica.

Nesta etapa, o intérprete, em geral geólogo ou geofísico, analisa os dados sísmicos e tenta criar um modelo que represente a geologia da área levantada.

No final da década de 1970, a interpretação era feita com o auxílio de mapas desenhados em papel, em cima dos quais os especialistas apontavam as áreas de interesse para a perfuração. Os especialistas precisavam ser capazes de criar um modelo tridimensional mental sobre a espessura, constituição e profundidade das camadas de rocha. O trabalho se tornou muito mais simples, uma vez que os especialistas não precisam mais usar seus conhecimentos e imaginações para desenharem modelos mentais. Um mapeamento que antes demorava meses para ser realizado, agora pode ser concluído em poucas horas [Petrobrás, 1999].

A interpretação pode ser classificada, de acordo com o foco, em dois tipos: estrutural e estratigráfica. Em que a estrutural basicamente tenta identificar as camadas geológicas ou, de forma equivalente, as interfaces entre as camadas, bem como as falhas geológicas que recortam as camadas. Já a estratigráfica, o objetivo está em entender a maneira como as camadas estão dispostas ao longo do tempo. As Figuras 2.10 e 2.11 mostram um modelo geológico que poderia ser resultante de uma interpretação de uma linha sísmica.

A etapa baseia-se nas operações de *picking* e *tracking*. *Picking* é o processo de determinar as coordenadas de um determinado evento sísmico dentro de um volume de dados. Utiliza-se do critério de máximo local de amplitude [Machado, 2000]. Já a operação de *tracking*, consiste em continuar o *picking* para todo o volume sísmico. Têm sido utilizados modernos softwares para oferecer assistência ao intérprete na interpretação de horizonte e falhas sísmicas.

Um horizonte sísmico é definido como sendo a superfície que separa duas camadas diferentes de rocha, onde tal superfície está associada com uma reflexão que se estende por uma grande área, à esquerda Figura 2.10.

As falhas sísmicas discretas dos horizontes, através dos quais são medidos os deslocamentos da rocha, podendo ser percebidas através das descontinuidades dos horizontes sísmicos, à direita da Figura 2.11.

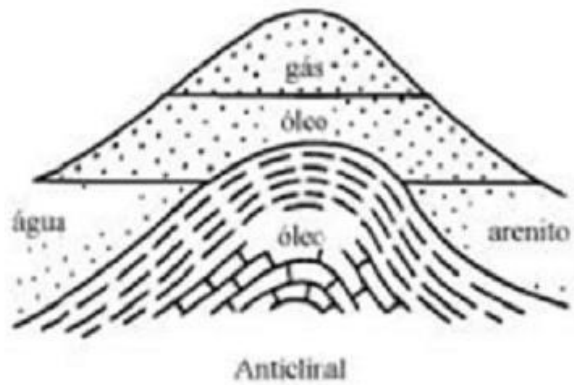


Figura 2.10 – Horizonte Sísmico [Robson & Treitel, 1980].

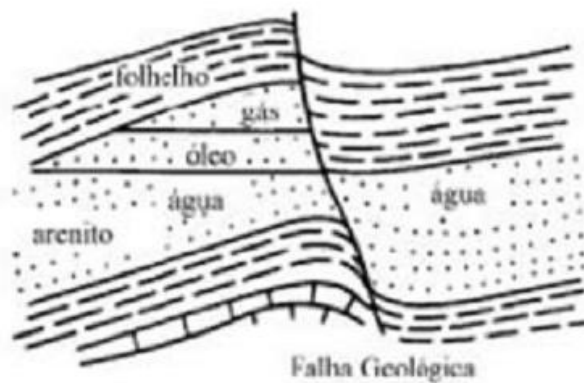


Figura 2.11 – Falha Sísmica [Robson & Treitel, 1980].

2.2. Atributos Sísmicos.

Os atributos sísmicos são subconjuntos de toda a informação geológica contidas nos sinais sísmicos, pois quantificam características específicas dos dados. Para Taner define-se atributo sísmico como sendo “quaisquer informações obtidas através de dados sísmicos, seja através de medições diretas, lógicas ou baseadas na experiência e conhecimento anteriores” [Taner, 2000]. Seu principal objetivo é ajudar o interprete a gerar um modelo geológico de determinada área sísmica. Quanto mais detalhado for o modelo geológico feito pelo intérprete, maior a chance de sucesso na área em questão.

Se os sinais sísmicos contêm informações geológicas, podem existir operações matemática capazes de revelar de forma mais clara essas informações. Logo, os cálculos de atributos decompõem dados sísmicos em atributos consistentes. Esta decomposição é informal na qual não existem regras que comandem como são calculados esses atributos ou o que eles podem ser. Conseqüentemente, qualquer quantidade calculada de dados sísmicos pode ser considerada um atributo [Barnes, 2001].

Os atributos sísmicos permitem a obtenção de informações detalhadas precisas sobre os parâmetros estruturais, estratigráficas e litológicas da perspectiva sísmica [Taner, 2001] como uma solução integrada [Cooper, 2008] para modelagem de reservatório de hidrocarbonetos, identificação e caracterização de incerteza reduzida [Sheline, 2005].

A análise de características sísmica é realizada através do emprego de técnicas de reconhecimento de padrões, onde através da combinação adequada de atributos sísmicos, busca-se a identificação de características geológicas dos reservatórios. Objetivando detectar as falhas sísmicas no dado, neste trabalho foi usada a técnica de máquina de vetorial de suporte (MVS), para classificar as imagens sísmicas.

2.2.1. Importância dos atributos sísmicos na geofísica.

Os atributos sísmicos são gerados para se enfatizar a característica desejada dos dados, que não podem ser diretamente identificada. Através de um registro sísmico é possível avaliar vários parâmetros, tais como: espessura de camada, mergulho, presença de falhamentos, ocorrência de acumulações rasas de gás biogênico e deslizamentos submarinos.

O estudo dos atributos sísmicos pode ser utilizado para fazer inúmeras interpretações geológicas e geofísicas tornando de extrema importância para o geofísico na etapa de interpretação. Dentre vários atributos existentes, alguns podem ser usados como indicadores de hidrocarboneto, outros como indicadores de falhas geológicas, ou seja, dependendo do tipo de análise em questão, deve-se escolher o tipo adequado de atributo [Matos, 2004].

Na Figura 2.12, lista alguns atributos utilizados por Tingdahl & Rooij [Tingdahl & de Rooij, 2005] em seus trabalhos sobre detecção semi-automática de falhas em dados 3D.

Neste trabalho serão aplicados os atributos de similaridade e energia descritos por Tingdahl [Tingdahl & de Rooij, 2005] na tentativa de detectar a falha para comparação e associação com o atributo geoestatístico Semivariograma.

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 2.12 – Atributos utilizados por Tingdahl [Tingdahl & de Rooij, 2005].

2.2.2. Atributo de Energia

A energia E é definida, segundo Tingdahl e Rooij [Tingdahl & de Rooij, 2005], como sendo a soma da amplitude dentro de uma janela (gate) ao redor da amostra de investigação:

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

onde, f é a amplitude do traço, t_o é o limite superior da janela, N é o número de janelas e dt é o intervalo amostral.

2.2.3. Atributo de Similaridade

A similaridade S é calculada entre dois traços, sendo definida como a distância no hiperespaço entre os vetores dos segmentos, normalizados à soma dos comprimentos dos vetores. Descrito em [Tingdahl & Rooij, 2005], sendo expresso por:

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

onde,

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

e t é o intervalo do tempo (profundidade) considerado, dt é o intervalo amostral, t_1 e t_2 são os limites do tempo da janela (x_v, y_v) e (x_u, y_u) são as duas posições de traço que serão comparadas, f é o valor de amplitude no cubo. Comparado com a coerência tradicional [Bahorich & Farmer, 1995], a similaridade também leva em conta as diferenças de amplitude entre os dois segmentos de traço. Falhas são descontinuidades no dado que dá uma baixa resposta á similaridade.

2.3. Funções Geoestatísticas

A geoestatística é um método topo-probabilístico, isto é, que utiliza um modelo probabilístico e a posição espacial das observações para estudar a variabilidade dos valores observados.

Krige (1951), trabalhando com dados de concentração de ouro, concluiu que somente a informação dada pela variância não era suficiente para explicar o fenômeno em estudo. Sendo necessário levar em consideração a distancia entre as observações. Surgindo então o conceito da geoestatística, que leva em consideração a localização geográfica e a dependência espacial.

Baseado nas observações de Krige, Matheron(1963, 1971), desenvolveu a teoria das variáveis regionalizadas. Em que uma variável regionalizada é uma função numérica com distribuição espacial, que varia de um ponto para outro com continuidade aparente, mas cujas variações não podem ser representadas por uma função matemática simples.

Neste trabalho será utilizada a medida geoestatística, semivariograma, na extração de características para identificar e sugerir regiões de falhas geológicas em dados sísmicos. Assim como os métodos estimadores geoestatísticos da autocorrelação espacial, como o variograma, covariograma e o correlograma, o semivariograma é usado como ferramenta de continuidade espacial, usada para investigar a magnitude da correlação entre as amostras e sua similaridade ou não, com a distância.

A vantagem do uso de técnicas estatísticas espacial é que tanto a variabilidade e a autocorrelação espacial podem ser medidos em conjunto. E ambos os componentes, tanto a variabilidade quanto a autocorrelação espacial, estão associados a pixels (ou outra unidade).

2.3.1. Semivariograma

Duarte [Duarte, 2000] define o semivariograma como um método geoestatístico para diagnosticar a presença da correlação entre as unidades amostradas. Preferencialmente para caracterizar a estrutura de continuidade espacial da característica avaliada. Ela representa uma função de semivariâncias em relação às respectivas distâncias. A semivariância é definida como a metade da variância de diferença entre observações de uma variável aleatória Z , separadas por uma distância h .

A função variograma é denominada de $2\gamma(h)$, é a esperança matemática do quadrado da diferença entre pares de pontos separados por um vetor distância (h) [Journel & Huijbregts, 1978].

$$2\gamma(h) = E\{[Z(x) - Z(x + h)]^2\} \quad (2.4)$$

A função semivariograma é denominada como sendo a metade da função variograma. Em que o estimador da semivariância $\gamma(h)$ é igual a média aritmética das diferenças do quadrado entre pares de valores experimentais, em todos os pontos separados por um vetor distância h .

$$\gamma(h) = \frac{1}{2N(h)} \sum_i^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \quad (2.5)$$

onde $\gamma(h)$ é a semivariância estimada para cada vetor distância h , $N(h)$ é o número de pares de pontos separados por um vetor distância h , $Z(x_i)$ é o valor da variável regionalizada no ponto x_i e $Z(x_i + h)$ é o valor no ponto $x_i + h$. O vetor distância é exemplificado na Figura 2.13.

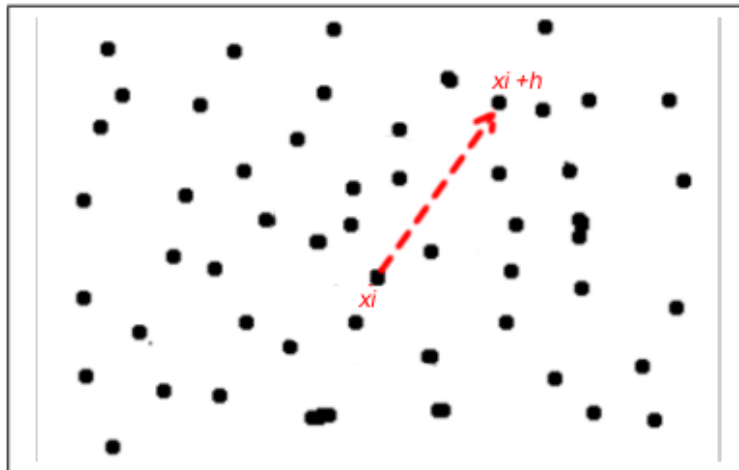


Figura 2.13 – Definição do vetor distância.

A função apresentada neste trabalho apresenta 3 características principais: alcance, patamar e efeito pepita.

A Figura 2.14 mostra uma função semivariograma próximas do ideal. Em que na origem ocorre uma descontinuidade denominada de efeito pepita que pode ser explicado pela variabilidade de pequena escala não detectada na amostragem ou simplesmente pela aleatoriedade espacial do conjunto de dados. A curva então cresce suavemente e a partir de um dado valor de h torna-se constante.

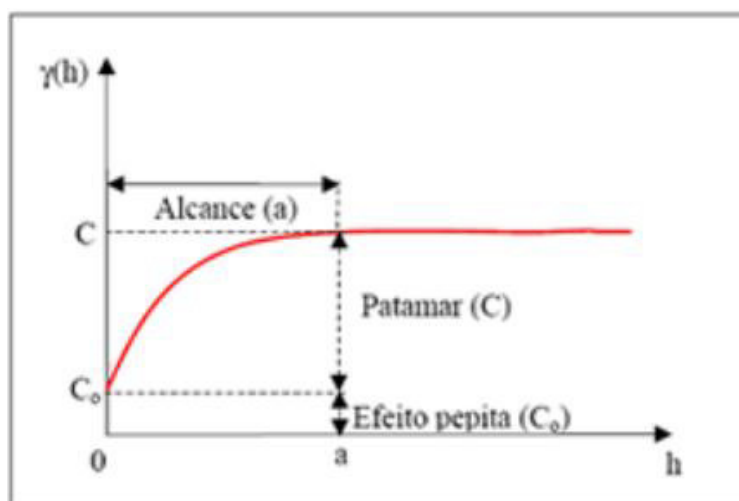


Figura 2.14 – Características do semivariograma.

O valor da variável h onde a curva se torna constante é denominada alcance, e o valor correspondente da função $y(h)$ é chamado de patamar. Considera-se que a partir do alcance não há mais dependência espacial entre os pixels da amostragem.

- **Alcance:** É utilizado para verificar o grau de continuidade espacial dos dados, representando a distância máxima que uma característica está correlacionada espacialmente.

- **Patamar:** Pode ser usado para medir a variabilidade do semivariograma. É aproximadamente igual à variância dos dados. À medida que a distancia h entre os pontos amostrados aumentam, a semivariância $y(h)$ também aumenta até se estabilizar.

- **Efeito Pepita:** Mostra a descontinuidade do semivariograma para distâncias menores que a menor distância entre as amostras. Na prática isso acontece quando h tende para 0 (zero), fazendo conseqüentemente, com que $y(h)$ se aproxime de um valor positivo chamado de Efeito Pepita (C_0).

O incremento do *lag* (*lag spacing*), tolerância do *lag* (*lag tolerance*), direção (*direction*), tolerância angular (*angular tolerance*) e a largura máxima de banda (*maximum bandwidth*), como ilustrados na Figura 2.15, são outros parâmetros usados para calcular o semivariograma.

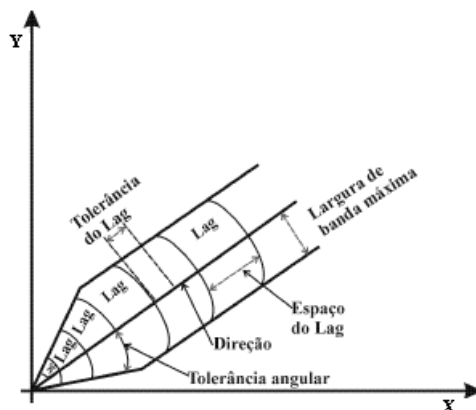


Figura 2.15 – Parâmetros utilizados para cálculo da função geoestatística [Deutsch e Journel, 1992].

Difícilmente um conjunto de pares de pontos da amostra vai estar separado precisamente por uma distância h , logo, são utilizados erros para mais e para menos em função da distância (incremento do *lag*) e em função da direção (tolerância angular) com o objetivo de alcançar um conjunto mais adequado de pares de pontos que estão relacionados com a distância em questão. O que torna esses parâmetros necessários.

2.4. Máquina de Vetores Suporte

Desenvolvida por Vapnik e Colaboradores [Gunn, 1988], as máquinas de vetores de suporte (MVS) são algoritmos de aprendizado bastante utilizados na área de aprendizado de máquina (AM), que fazem uso do princípio da inferência chamado indução, a partir do qual é possível obter conclusões genéricas a partir de entradas fornecidas. Podendo este aprendizado indutivo ser de dois tipos: supervisionado e não supervisionado [Lorena, 2007]

No aprendizado supervisionado há a presença externa do que chamamos de professor, ou seja, é dado um conjunto de exemplos para que a máquina possa aprender sobre o domínio do problema. Esses exemplos fornecem um conjunto de entrada e suas respectivas saídas, com isso ele extrai o conhecimento necessário para generalizar as saídas esperadas de futuras entradas.

Ao contrário, no aprendizado não supervisionado a presença externa do professor é dispensada, ou seja, não é dado um conjunto de exemplo para que a máquina possa aprender. Aprendendo, assim, com a ajuda de uma medida de qualidade. Usada geralmente quando o objetivo do algoritmo é encontrar padrões e/ou tendências em um conjunto de dados complexos.

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

$$S = (x_1, k_1), (x_2, k_2), \dots, (x_m, k_m) \quad (2.5)$$

onde, x_m representa a m-ésima observação ou padrão. k_m representa a classe, ou rótulo, do padrão acima, e m representa o número de amostras disponíveis. Deve-se produzir um classificador, também denominado modelo, preditor ou hipótese, capaz de prever precisamente o rótulo de novos dados. Esse processo de indução de um classificador a partir de uma amostra é denominado treinamento.

A tarefa de classificação envolve normalmente treinamento e teste de dados, consiste em alguns exemplos de dados. Cada conjunto de treinamento contém um “valor objetivo” (rotulo de classes) e vários atributos (características), e freqüentemente, a variável de interesse é definida da seguinte forma:

$$y_i = \begin{cases} 1, & \text{se o indivíduo pertence ao grupo 1} \\ 0, & \text{se o indivíduo pertence ao grupo 2} \end{cases} \quad (2.6)$$

onde o principal interesse é obter $Prob(y = 1/x) = f(x, \beta)$, e na prática, $f(x, \beta)$ é sempre uma função desconhecida.

O objetivo do MVS é produzir um classificador desempenho satisfatório junto a amostras não-observadas durante a fase de treinamento, ou seja, consiga generalizar. Classificador é uma máquina que fornece classificação. No exemplo da Figura 2.16, nota-se a existência de vários classificadores lineares possíveis que podem separar em duas classes as amostras disponíveis sem nenhum erro, mas há somente um (em verde) que maximiza a distância entre o classificador e a amostra mais próxima de cada classe. O classificador linear que maximiza a distância é denominado hiperplano ótimo. Da mesma forma podemos observar na Figura 2.16 que o hiperplano ótimo criado tem outros hiperplanos como suporte (tracejadas em vermelho) denominado vetores-suporte e que passam por pontos das classes.

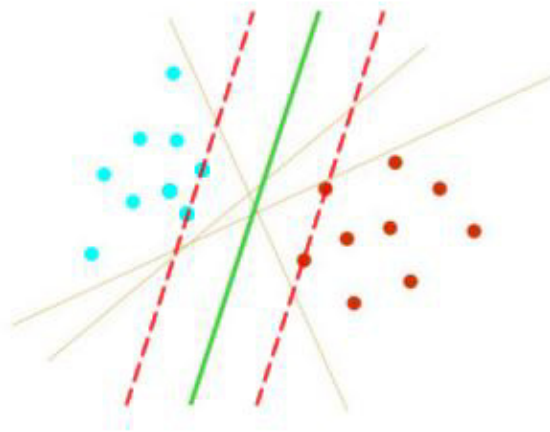


Figura 2.16 – Hiperplano ótimo.

A solução ótima a ser encontrada pelo MVS utiliza o princípio de minimização do risco estrutural baseando-se no fato de que a taxa de erro nos dados de teste (taxa de erro de generalização) é limitada pela soma da taxa de erro de treinamento e por um termo que depende da dimensão de Vapnik-Chervonenkis (dimensão VC)³

Sendo dado um conjunto de treinamento de pares de rótulo (x_i, y_i) , $i = 1 \dots l$ onde $x_i \in R^n$ e $y_i \in \{1, -1\}$. A classificação da margem de separação ao máximo por finalidade separar as duas classes por um hiperplano tal que a distância dos vetores suporte seja maximizada. Cada ponto x no espaço de entrada é mapeado para um ponto $z = \Phi(x)$ de alta dimensionalidade, denominado espaço de característica, onde os dados são linearmente separados por um hiperplano.

Dividido em três categorias, os dados podem ser chamados linearmente separáveis, não linearmente separáveis e não separáveis. Assim podemos escrever a nossa função de decisão usando uma função de *kernel* $K(x, y)$, por ser mais simples que o do mapeamento Φ , e que é definida pela função:

$$f(x) = \sum_{i=1}^l \alpha_i y_i K(x, x_i) + b \quad (2.7)$$

onde $K(x, x_i) = \Phi(x) \cdot \Phi(x_i)$, e os coeficientes α_i e b , são soluções das equações:

$$\min_{w,b,\xi} \frac{1}{2} w^T \cdot w + C \sum_{i=1}^l \xi_i \quad (2.8)$$

sujeito a:

$$y_i [w^T \cdot \phi(x_i) + b] \geq 1 - \xi_i \quad (2.9)$$

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

As MVS possuem diferentes *kernels* que são utilizados para resolver problemas de espaços não lineares, sendo os mais utilizados os tipos Linear, Polinomial (que manipula uma função polinomial cujo grau pode ser definido durante os treinamentos), Sigmoidal (permite que a MVS tenha o comportamento de uma rede MLP4) e Gaussiano (corresponde a um espaço de características de dimensão infinita; a utilização desse tipo de função permite que a MVS se comporte com características de uma rede RBF5). Temos as quatro funções básicas de *kernel*:

- Linear: $K(x, y) = x^T y$;
- Polinomial: $K(x, y) = (\gamma x^T y + r)^d, \gamma > 0$;
- Sigmoidal : $K(x, y) = \tanh(\gamma x^T y + r)$;
- Função básica radial (RBF): $k(x, y) = e^{-\gamma \|x-y\|^2}$.

As diferentes definições do *kernel* e seus respectivos parâmetros provocam alterações nos resultados fornecidos por uma MVS. Utilizamos nesse trabalho o *kernel RBF* com o parâmetro $\gamma > 0$ definido pelo usuário.

2.4.1. Avaliação dos Resultados

Como métricas de avaliação serão utilizadas neste trabalho a Sensibilidade (s), Especificidade (e) e a Acurácia (a), por serem métodos estatísticos comumente utilizados, para validar resultados.

A sensibilidade (s) é calculada como a probabilidade de verdadeiros positivos, ou seja, a medida da capacidade do método de decisão de predizer a condição de Falha para aqueles casos que realmente são, segundo a equação:

$$s = \frac{VP}{VP+FN} \quad (2.9)$$

em que VP é o número de verdadeiros positivos, e FN é o número de falso negativo, no qual pontos foram definidos como Falha são classificados como Não-Falhas erroneamente.

A especificidade (e) é calculada como a probabilidade de verdadeiros negativos, a medida da capacidade do método de decisão apontar ausência de condição de Falha, para aqueles casos que realmente não tem, como apresentado:

$$e = \frac{VN}{VN+FP} \quad (2.10)$$

onde VN é o número de verdadeiros negativos, Não-Falha classificado como Não-Falha e FP é o número de falsos positivos, Não-Falhas classificados Falhas.

A acurácia (a) é a probabilidade de predições corretas, e é dada pela Equação abaixo:

$$a = \frac{VP+VN}{VP+VN+FP+FN} \quad (2.11)$$

3. Materiais e Métodos

Neste trabalho, o objetivo é o estudo da função geoestatística semivariograma para a detecção de falhas geológicas em dados sísmicos, para isso foi realizado usando os seguintes passos:

1. Aquisição e tratamento das imagens sísmicas;
2. Marcação dos pontos de Falhas e Não-Falhas das imagens sísmicas;
3. Cálculo dos atributos sísmico de falha: Energia Vertical e Similaridade Vertical;
4. Cálculo do atributo Semivariograma como descritores de característica.
5. Classificação usando MVS.

3.1. Aquisição e Tratamento das Imagens Sísmicas

Neste trabalho foi utilizado um volume sísmico sintético de 256 x 256 x 255 dados de amplitude pós-empilhados. No tratamento, os dados foram convertidos para o formato *little-endian*, em que byte menos significativo é armazenado na memória nos menores endereços, já que os estavam sendo armazenados estavam no formato *begin-endian*, byte menos significativo nos maiores endereços.

Além disso, a partir do volume, foram obtidas as imagens como fatias (slices) digitalizadas em tons de cinza de 8 bits com resoluções de 256 x 256. Os dados de amplitude, que são valores entre -1699.143604 e +1769.342529, são quantizadas para 8 bits gerando imagens como mostrada na Figura 3.1.

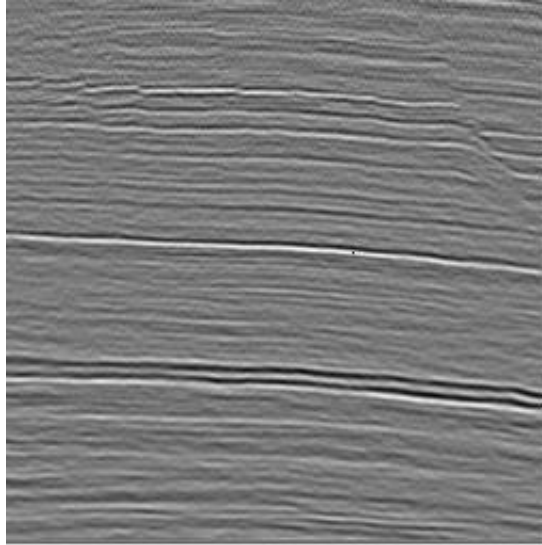


Figura 3.1 – Fatia gerada a partir dos dados de amplitude.

3.2. Marcações de pontos de Falhas e Não – Falhas

Para facilitar o processo de marcação manual de Falhas e Não-Falhas, que servem de base para o treinamento e teste dos atributos aqui estudados, foi desenvolvida uma ferramenta intuitiva.

Os pontos de Falhas foram salvos em um arquivo, nomeado com o número que identifica a fatia, contendo o valor +1 e sua respectiva posição na imagem, continuamente, com os postos de Não-Falhas com valor -1 e sua respectiva posição. Ver exemplo na Figura 3.2.

36.bt			
1	+1	228	77
2	+1	229	78
3	+1	229	79
4	-1	188	160
5	-1	188	151
6	-1	198	145

Figura 3.2 – Arquivo contendo pontos de Falhas e Não-Falhas.

3.3. Cálculo de Atributos Sísmicos de Falha.

Características específicas das imagens sísmicas são identificadas através dos atributos sísmicos. Para efeito de comparação e combinação com o atributo de Semivariograma, foram escolhidos os atributos de Energia Vertical (EV) e Similaridade Vertical (SV), ambos são calculados ao longo do traço sísmico, vertical da imagem.

3.3.1. Calculo da Energia Vertical

A Energia Vertical é calculada ao longo do traço sísmico, segundo Equação 2.1, o calculo para o pixel em questão é o dado como a soma do quadrado dos valores de amplitude dentro de uma janela definida, dividido pelo tamanho da janela. No trabalho, o tamanho da janela utilizada foi o tamanho 15.

A Figura 3.3, mostra um exemplo de como é calculada o atributo. Onde é selecionado o pixel de valor 8 e definida o tamanho da janela, 5. É feita a soma do quadrado dos valores dentro da janela ($5^2 + 2^2 + 8^2 + 2^2 + 7^2$) e dividido 5, referente ao tamanho da janela.

A janela vai sempre percorrer a imagem na direção vertical. O tamanho da janela é definida pelo usuário.

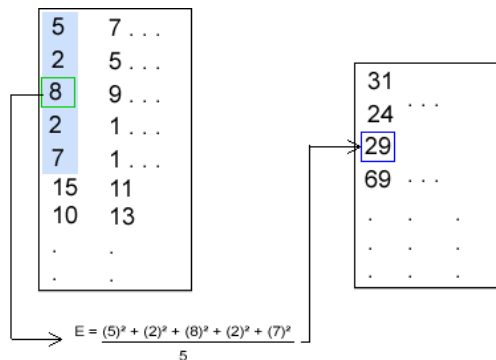


Figura 3.3 – Calculo da Energia Vertical.

Na Figura 3.4 temos um exemplo de uma imagem com o atributo de energia vertical calculado.

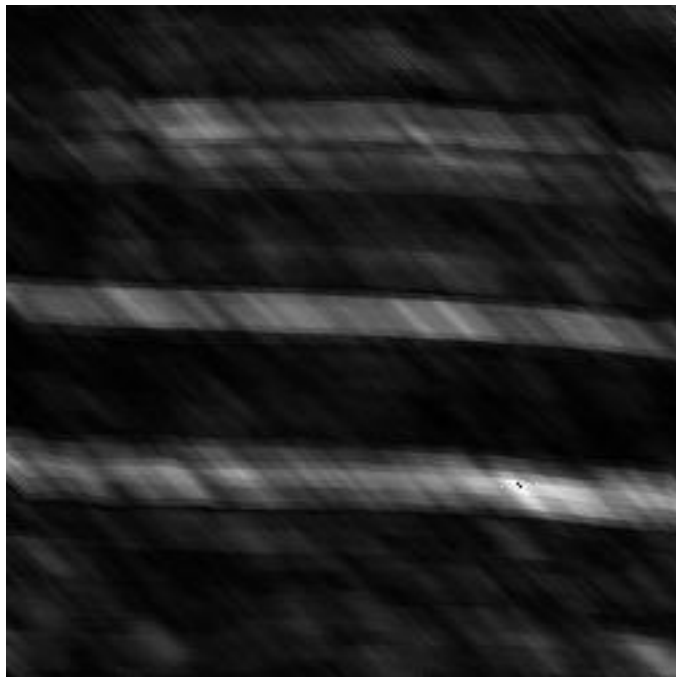


Figura 3.4 – Energia Vertical.

3.3.2. Calculo da Similaridade Vertical

A similaridade vertical (SV) é realizada entre dois segmentos de traço, calculadas, segundo a Equação 2.2, fatia por fatia sempre definido o vetor onde está contido o pixel de referência, para a qual será calculado o atributo, de vetor v e o vetor adjacente u , ambos com o mesmo tamanho de janela. Depois de calculado, o resultado é guardado em uma posição equivalente do pixel de referência. A Figura 3.5, exemplifica o calculo.

Como podemos ver na Figura 3.5, o cálculo para o elemento 103, circulado, foi calculado $|v - u|$ fazendo $|(8 - 9), (11 - 7), (103 - 109), (6 -$

7), $(4 - 4)$], dividido pelo $|v| + |u|$ e subtraindo de 1. A janela vai percorrendo a imagem na vertical assim como no cálculo da EV.

Na Figura 3.6 um exemplo de imagem gerada após o cálculo de do atributo de SV.

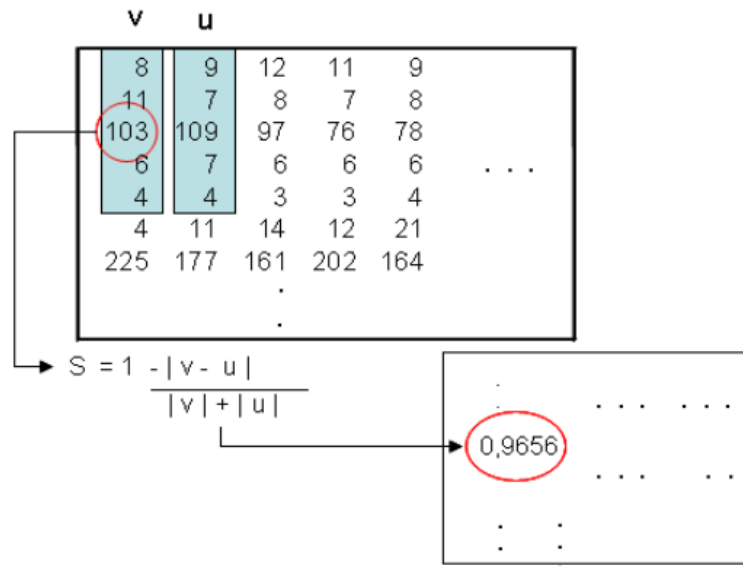


Figura 3.5 – Cálculo da Similaridade Vertical.

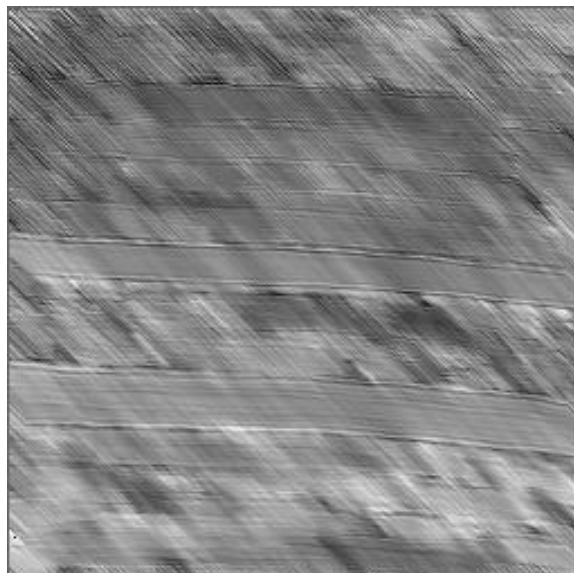


Figura 3.6 – Similaridade Vertical.

3.3.3. Calculo do Atributo de Semivariograma

No calculo do atributo de Semivariograma, um vetor de características (VC) é gerado com o objetivo de analisar a eficiência da área de interesse do dado sísmico.

Para a criação desse VC, foram extraídas 72 características por amostra correspondentes as 4 direções de 26 lags (15 + 8 + 3) para função geoestatística semivariograma.

Os parâmetros usados para a extração de características pela função semivariograma, em cada amostra, foram as direções 0° , 45° , 90° e 135° com uma tolerância de $22,5^\circ$ e incremento de *lag* (distância) igual a 1, 2 e 3 correspondendo a 29, 14 e 9 *lags* e a tolerância de cada distancia de *lag* igual a respectivamente 0,5, 1,0 e 1,5. As direções adotadas são as mais utilizadas na literatura para análise de imagens, já para escolher a tolerância de *lag* segundo [ISAAKS e SRIVASTAVA, 1989] a escolha mais comum é adotar a metade do incremento de *lag*.

Na Figura 3.7, temos a ilustração da maneira como as distâncias e direções do *lag* são utilizadas. Supondo que essa imagem seja de 30 x 30 pixels, que é o tamanho da janela utilizada, o máximo de *lags* que teremos em qualquer das quatro direções para $d = 1$ será 29, $d = 2$ será 14 e $d = 3$ será 9 justificado assim, a presença desses valores nos parâmetros.

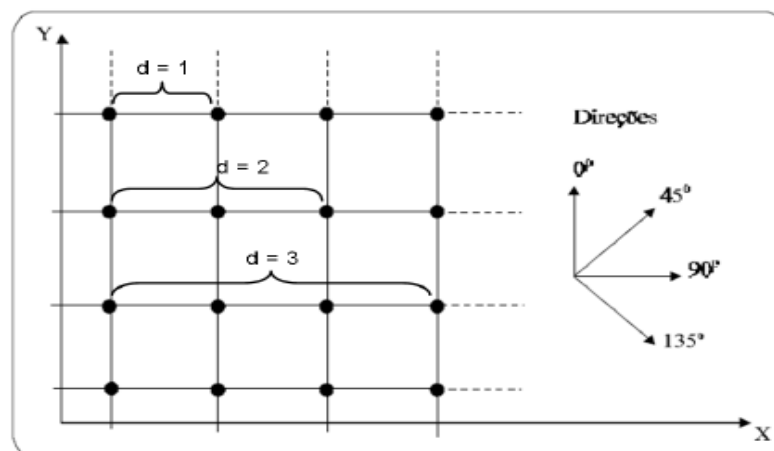


Figura 3.7 – Distância e direções dos *lags* [CAMARGO, 1998].

3.4. Classificação usando MVS.

Foi utilizada nesta etapa, a biblioteca LIBSVM [Chang & Lin, 2001], como uma ferramenta que facilita os usuários a usar MVS. Um aspecto importante desta ferramenta é a existência, nos diretórios da ferramenta, de dois scripts *python - grid.py* e *easy.py*. Estes têm por finalidade ajudar pessoas que nunca treinaram uma MVS antes.

O objetivo principal desta etapa consiste em classificar objeto na imagem sísmica como Falha e Não Falhas. Para isso, tivemos os seguintes passos a seguir em suas respectivas ordens:

- Geração do arquivo de entrada no formato LIBSVM;
- Utilização do *Kernel* $k(x, y) = e^{-\gamma \|x-y\|^2}$;
- Utilizamos a base modelo e transformamos para o intervalo de 0 a 1, usando *svm-scale.exe*;
- Dividimos o esse arquivo em arquivo de teste e arquivo de treino na proporção de 50%, usamos para essa finalidade o script *subset.py*;
- Utilizamos o script *grid.py* que implementa a busca automatizada (*search grid*) para encontrar o melhor parâmetro C e (γ);
- Utilização do melhor parâmetro C e γ em todo o conjunto de treinamento;
- Construção do modelo;
- Classificação do arquivo de teste;
- Calculo das métricas;

Os dados de EV submetidos à classificação foram normalizados para o intervalo [0,1] como em:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (2.13)$$

onde, x_{norm} é o valor normalizado correspondente ao valor original, e x_{min} e x_{max} são os valores mínimo e máximo entre todos os valores. Para a SV não foi necessário a normalização, pois o calculo da SV já esta entre valores no intervalo [0,1].

Foram definidas duas classes para essa finalidade, Falha (+1) e Não-Falha (-1). Geramos arquivos com 1 atributo de Semivariograma, com 1 atributo de SV e com 1 atributo de EV, em seguida foi gerado com associação de 2 em 2 e com associação dos três .

Existem dois parâmetros a serem configurados no *kernel*, RBF, escolhido neste trabalho: C e γ . É desconhecida anteriormente qual os melhores valores de C e γ para o problema, deve ser feito algum tipo de modelo de seleção (parâmetros de busca), com a finalidade de identificar qual o melhor C e γ de forma que o classificador possa prever dados desconhecidos com precisão.

Os dados de treinamento foram separados em duas partes, em que, uma é considerada desconhecida no treinamento do classificador. Então a precisão da predição, neste conjunto, pode refletir mais precisamente no desempenho da classificação de dados desconhecidos.

A etapa é finalizada com os treinamentos e testes do MVS, configurada com os parâmetros C e γ para cada conjunto de amostras submetidas ao processo.

Na Seção seguinte trataremos dos resultados atingidos com o uso desta metodologia.

4. Resultados Obtidos

Foram aplicados os atributos propostos para 50 fatias, dentro de um volume sísmico com numerosas falhas, onde de cada fatia foi extraído 25 pontos de falhas e também 25 pontos de não-falhas, no total de 2500 amostras. Com o objetivo da detecção de falhas geológicas, teve muito eficiência o método de classificação com uso do MVS. No trabalho foram rodados cinco vezes cada atributo e suas associações, os resultados mostrados aqui são uma média aritmética das cinco, sendo a acurácia realizada através da validação cruzada.

De acordo com a metodologia proposta na Seção 3.4, foi realizada a etapa de classificação e validação de resultados, obtidos pelo classificador MVS, e relacionadas à Tabela 4.1, cuja melhor resultado de forma individual pode-se ver no atributo de Semivariograma, e com resultados bem melhores quando associamos os três.

Tabela 4.1 – Resultados da acurácia para 2500 amostras, calculada para cada atributo e suas associações.

Atributo	Amostras	Acurácia	
		Treino	Teste
EV	50/50	81,32%	81,23%
SV	50/50	63,84%	62,72%
Semivariograma	50/50	91,68%	91,13%
EV / SV	50/50	84,38%	83,58%
Semivariograma / SV	50/50	91,82%	92,25%
Semivariograma / EV	50/50	94,88%	94,76%
Semivariograma / EV/SV	50/50	95,36%	95,2%

A Tabela 4.2 mostra do desvio padrão da acurácia dos atributos, em que o que possui maior estabilidade nos resultados de acurácia de teste é o atributo de Semivariograma associado ao atributo de Energia Vertical.

Tabela 4.2 – Desvio padrão dos atributos e suas associações.

Atributo	Desvio Padrão de Acurácia	
	Treino	Teste
EV	0,61	0,64
SV	0,78	1,02
Semivariograma	0,59	1,02
EV / SV	0,47	0,71
Semivariograma / SV	0,49	0,90
Semivariograma / EV	0,51	0,23
Semivariograma / EV/SV	0,44	0,75

Na tabela 4.3 são mostrados os resultados de sensibilidade (s) e especificidade (e) para os atributos individualmente e suas associações.

Na Tabela 4.1 e 4.3 podemos observar o melhor desempenho dos resultados alcançados nos testes onde a sensibilidade (s) é igual a 95,22%, especificidade igual a 96,49% e acurácia atingindo 95,2%. O gráfico mostrada na Figura 4.1 mostra os resultados dos testes apresentados, comprovando o melhor resultado como sendo a associação dos 3 atributos.

Tabela 4.3 – Resultados de sensibilidade e especificidade dos atributos e suas associações.

Atributo	Amostras	Teste	
		Sensibilidade	Especificidade
EV	50/50	73,63%	89,16%
SV	50/50	53,88%	47,12%
Semivariograma	50/50	90,61%	92,68%
EV / SV	50/50	81,63%	85,51%
Semivariograma / SV	50/50	91,45%	93,07%
Semivariograma / EV	50/50	94,22%	95,33%
Semivariograma/EV /SV	50/50	95,22%	96,49%

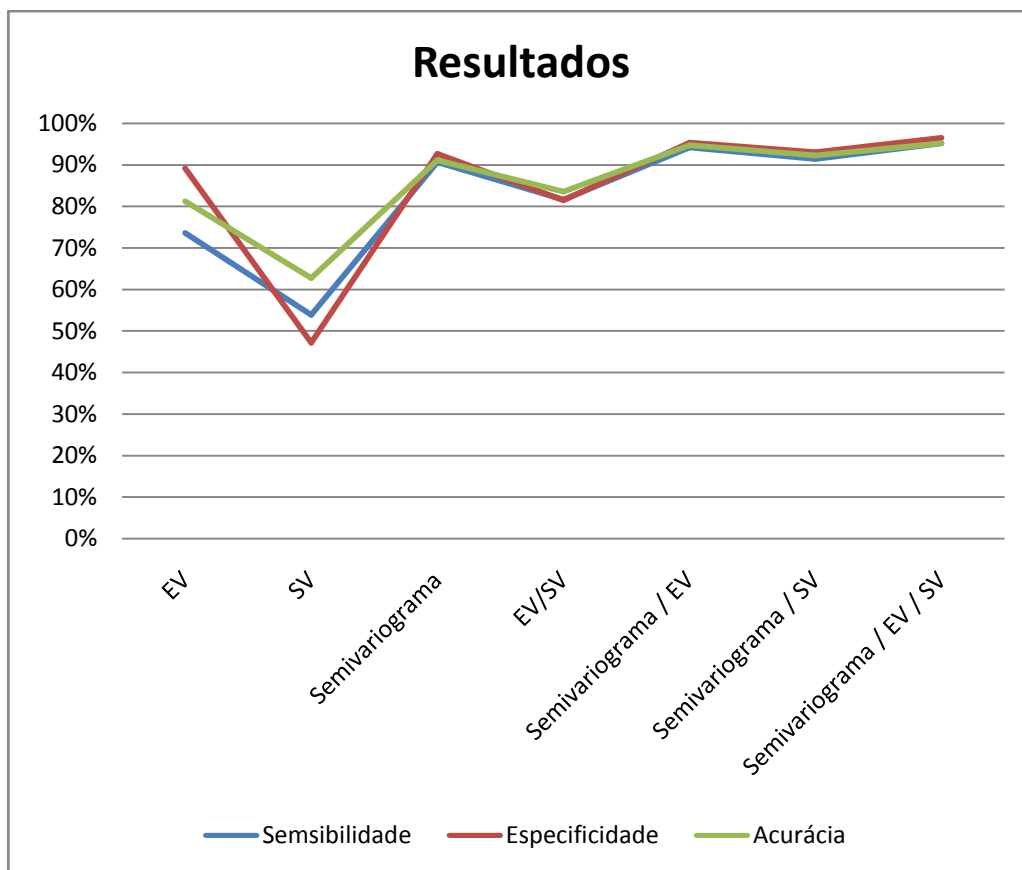


Figura 4.1 – Resultados obtidos nos testes.

5. Conclusão

Neste trabalho foi apresentada a viabilidade do uso da função geoestatística Semivariograma e a Máquina de Vetores de Suporte para a detecção de falhas geológicas em dados sísmicos.

O fato de que os altos custos que envolvem os métodos diretos de identificação de reservatórios de petróleo, faz com que se busque alternativas mais baratas que possam trazer o maior número de informações possíveis antes de se iniciar a perfuração de um poço. Dando importância de como se apresenta toda a estrutura do reservatório.

A confiança no resultado da interpretação depende da complexidade do sistema geológico, da qualidade dos dados sísmicos, do bom processamento, e da experiência do interprete [Rankey & Mitchell, 2003]. Portanto, qualquer erro de interpretação pode levar a resultados incoerentes.

Como visto, a escolha correta dos atributos, para o objeto sísmico que se deseja obter, é de extrema importância. Em busca dos melhores resultados, fizemos várias combinações com os atributos (um a um, dois a dois e todos). Com base nos resultados, verificou-se que o melhor resultado foi aquele aplicado os três atributos juntos.

Fundamentados nos resultados de especificidade igual a 96,49%, sensibilidade igual a 95,22% e acurácia igual a 95,2%, os métodos apresentados mostram que falhas sísmicas podem ser identificadas de forma semi-automática, utilizando o Semivariograma aplicados a uma máquina de suporte vetorial. Contudo, é necessário realizar testes com uma quantidade maior de volumes sísmicos, de dados reais, e com diferentes características de falhas, ou seja, imagens mais complexas para que seja possível comprovar a eficiência da técnica proposta. E como trabalhos futuros propomos:

- Realizar a combinação da função semivariograma com outros atributos existentes, como por exemplo, o de Energia Horizontal.
- A utilização de outros *kernels* que a MVS possuem, como o Linear, Polinomial, Sigmoidal ou Gaussiano.

- O classificador SVM que desempenhou um papel fundamental na classificação, podendo ser trocado por algum outro classificador de maneira a avaliar o desempenho desse novo classificador.

REFERÊNCIAS BIBLIOGRÁFICAS

BAHORICH, M ., & FARMER, S. 1995. 3-D seismic discontinuity for faults and stratigraphic features; the coherence cube. The Leading Edge, 14(10),1053–1058.

CAMARGO, E. C. G. Desenvolvimento, implementação e teste de procedimentos geoestatístico (KRIGEAGEM) no sistema de processamento de informações georeferenciadas (APRING). Dissertação (Mestrado), Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 1997.

CARVALHO, M. M. Determinação de Malhas de Falhas em Dados Sísmicos por Aprendizado Competitivo. Tese (Doutorado), Programa de Pós-Graduação em Informática, PUC-RJ, 2008.

CHEVRON, CORPORATION. Disponível em: <http://ep.chevrontexaco.com.br>. Acesso em 15 jan de 2012.

DALLYSON, S. A. J. Detecção de Falhas de Dados Sísmico usando Máquinas de Vetor de Suporte. Trabalho de Conclusão de Curso (Graduação em Bacharelado Em Ciência da Computação) - Universidade Federal do Maranhão, 2007.

DEUTSCH, C.V.; JOURNEL, A. G. GSLIB: Geostatistical Software Library and user's guide. New York, Oxford University Press, 1992. 339p.

DORN, G. A. Modern 3d seismic interpretation. The Leading Edge, v.17, p. 1262-1273, 1998.

DUARTE, J.B. Sobre o emprego e a análise estatística do delineamento em blocos aumentados no melhoramento genético vegetal. Piracicaba, 2000. 292p.

MELLO, JOSÉ MARCIO DE Geoestatística aplicada ao inventário florestal. Tese (Doutorado) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, 2004.

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

GUNNARSON, F.; HOLM, P.H.; THURESONN, T. On the potential of krigagem for forest management planning. Scandinavian Journal of Forest Research, v.13, n.2, p.237 - 245, 1998.

ISAAK S, E.H.; SRIVASTAVA, R.M. An introduction to applied geostatistics. New York: Oxford University Press, 1989. 560p

JOURNEL, A.G.; HUIJBREGTS, C.J. Mining geostatistics. London: Academic Press, 1978. 600p.

JOHANN, P., & RIBEIRO, B.D. 1999. Multivariate seismic pattern recognition and Kohonen maps applied on a deep-water turbidite reservoir in Campos Basin, Brazil. International Congress of the Brazilian Geophysical Society.

KOURGLI, ASSIA; AICHOUCHE, BELHADJ-AISSA. Characterising textural primitives using variography. IMVIP - Irish Machine Vision and Image Processing Conference. Belfast, Ireland, p. 165–175, 2000.

LINES, L. R.; NEWRICK, R. T. Fundamentals of geophysical interpretation. SEG, Society of Exploration Geophysicists. Department of Geoscience, University of Calgary, Calgary, AB T2N 1N4, Canada, 2004.

LIMA. Uma Abordagem Unificada Empregando Máquinas de Vetores-Suporte. Tese doutorado, Universidade Estadual de Campinas, 2004.

MACHADO, MARCOS. Segmentação de Dados Sísmica Via Hyper-stack para Visualização. Dissertação de Mestrado, Pontifícia Universidade Católica do Rio de Janeiro – RJ, 2000.

MATOS, MARCÍLIO CASTRO. Reconhecimento de padrões sísmicos utilizando análises tempo-frequência. Tese doutorado, Pontifícia Universidade Católica do Rio de Janeiro, RJ, 2004.

MATOS, MARCILIO CASTRO. Reconhecimento de padrões sísmicos utilizando análises tempo-frequencia. Tese doutorado, Pontifícia Universidade Católica do Rio de Janeiro, RJ, 2004.

MELLO, JOSÉ MARCIO DE . Geostatística aplicada ao inventário florestal. Tese doutorado, Escola Superior de Agricultura Luiz de Queiroz, 2004.

MORAES. Mapeamento automático de horizontes e falhas em dados sísmicos 3D baseado no algoritmo de gás neural evolutivo. Tese (Doutorado), PUC – RJ, 2008.

PETROBRÁS. 1999. . Petrobrás Magazine , 7(26), 20–23.

OILFIELD, SCHLUMBERGER GLOSSARY. Disponível em: <http://www.glossary.oilfield.slb.com>. Acesso em 22 de jun 2012.

PEDERSEN, S.; SCHLUMBERGER TECHNOLOGY CORPORATION. Image feature extraction. United States Patent, 7203342, April, 10, 2007.

RANNEY, E. C.; MITCHELL, J. C. Interpreter's Corner– That's why it's called interpretation: Impact of horizon uncertainty on seismic attribute analysis. The Leading Edge, Vol. 22, p. 820, 2003.

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

SERCEL 2003 Marine Acquisition Catalog. Disponível em <<http://www.sercel.com>>. Acesso em 21 nov 2003.

SILVA, PEDRO MARIO CRUS. Visualização Volumétrica de horizontes em Dados Sísmico 3D. Tese doutorado, Pontifícia Universidade Católica do Rio de Janeiro, 2004.

SRIVASTAVA, M. R.; PARKER, H. M. Robust measures of spatial continuity. Geostatistics Proceedings of the Third Int. Geostatistics, Congress, Avignon, France 1, 295–308, 1988.

TANER, M. T. Seismic Attributes. Canadian Society of Exploration Geophysicists. Sept. 2000.

TINGDAHL, K. M.; DE ROOIJ, M. Semi-automatic detection of faults in 3D seismic data. Geophysical Prospecting, vol. 53, p. 533–542, 2005.

THOMAS, J. E. Fundamentos de Engenharia de Petróleo. Ed. Sinopse, 2ª Edição, 2001.