

**CINTIA SCHOENINGER**

Tratamento de Informações Imperfeitas na Análise de Risco de Prospectos em  
Exploração Petrolífera

Florianópolis - SC

2003

CINTIA SCHOENINGER

Tratamento de Informações Imperfeitas na Análise de Risco de Prospectos em  
Exploração Petrolífera

Trabalho Individual submetido à avaliação como requisito parcial para a obtenção do grau de Mestre em Ciências da Computação pela Universidade Federal de Santa Catarina (UFSC) no curso de Mestrado em Ciências da Computação, com área de concentração em Inteligência Artificial.

PROF. DR. EEL. MAURO ROISENBERG

(Orientador)

Florianópolis - SC

2003

## CIP - CATALOGAÇÃO NA PUBLICAÇÃO

Schoeninger, Cíntia

Tratamento de Informações Imperfeitas na Análise de Risco de Prospectos em Exploração Petrolífera / por Cíntia Schoeninger – Florianópolis: CPGCC da UFSC, 2003. 97f.:il.

TI(mestrado) - Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Ciências da Computação, Florianópolis, SC-BR, 2003. Orientador: Roisenberg, Mauro.

UNIVERSIDADE FEDERAL DE SANTA CATARINA

Reitor: Prof. Esp. Dir. Rodolfo Joaquim Pinto da Luz

Pró-Reitor de Pós-Graduação: Prof. PhD Álvaro Toubes Prata

Secretaria Extraordinária de Informática: Prof. Dr. EngProd. Rogério Cide Bastos

Coordenador do CPGCC: Prof. Dr. Eng.Prod. Fernando Álvaro Ostuni Gauthier

Bibliotecária Chefe do Instituto de Informática: Prof.<sup>a</sup> Sigríde Karin Dutra

Tratamento de Informações Imperfeitas na Análise de Risco de Prospectos em  
Exploração Petrolífera

**CINTIA SCHOENINGER**

Esta Dissertação foi julgada adequada para a obtenção do título de Mestre em Ciência da Computação Área de Concentração Inteligência Artificial e aprovada em sua forma final pelo Programa de Pós-Graduação em Ciência da Computação.

Prof. Mauro Roisenberg

**Orientador**

Prof. Fernando Álvaro Ostuni Gauthier

**Coordenador**

Prof. Mauro Roisenberg

**Presidente da Banca**

Prof. Guilherme Bittencourt

**Membro da Banca**

Prof. Pedro Alberto Barbeta

**Membro da Banca**

Geol. Reneu Rodrigues da Silva

**Membro da Banca**

## **Agradecimentos**

Ao Professor Mauro Roisenberg pela orientação, recursos, tempo dedicado e principalmente pela confiança depositada em mim.

À Petrobras, especialmente ao coordenador do projeto RCSUEX, geólogo Reneu Rodrigues da Silva, pela oportunidade de aplicar os conhecimentos teóricos a um problema real e pelas dicas e conhecimentos adquiridos ao longo do projeto.

Ao meu colega Dennis Kerr Coelho, que se empenhou em concretizar as idéias desta Dissertação.

Agradeço à paciência da minha família, sobretudo meu pai, Ernani Schoeninger, e minha mãe, Nadia Schoeninger, que tanto penalizei com minhas ausências, mas que sempre compreenderam a importância deste trabalho.

Ao meu esposo, Rafael Signori, que mais que compreender minhas ausências ajudou-me a concretizar este trabalho.

Aos meus colegas de mestrado, o meu muito obrigado.

E, por fim, gostaria de agradecer a um grande incentivador desta Pós-Graduação, o Prof. Ricardo Vargas Dornelles.

## RESUMO

A exploração petrolífera é uma atividade econômica onde são investidos bilhões de dólares na expectativa de encontrar-se petróleo, que é uma das fontes de energia mais utilizadas no mundo. A exploração petrolífera é uma tarefa difícil porque envolve tomada de decisão sobre informações imperfeitas, isto é, muitos dos dados disponíveis são incertos, imprecisos, vagos, etc. Apesar do alto investimento feito pelas companhias petrolíferas em ferramentas que auxiliem o processo exploratório, este ainda possui muitas variáveis incertas. Além disso, percebe-se a presença de fatores, como a experiência, a capacidade de argumentação e até mesmo certa dose de intuição da equipe que estuda o local do “provável” poço, dificultando a representação do conhecimento na exploração petrolífera.

Este trabalho surgiu como fruto de um projeto denominado “RCSUEX – Representação da Certeza de Sucesso Exploratório”, desenvolvido pelo Laboratório de Conexão e Ciências Cognitivas (L3C) da Universidade Federal de Santa Catarina – UFSC para a Petrobras, a partir da necessidade desta última em estruturar a representação do conhecimento incerto no Processo de Exploração Petrolífera. O objetivo do Projeto RCSUEX é desenvolver um Sistema Especialista que estime os riscos da exploração petrolífera, sobre dados geológicos e dados estatísticos para um novo poço, com o propósito de sistematizar a argumentação e medir a probabilidade de sucesso da descoberta de uma acumulação de petróleo.

Este trabalho estuda os métodos de raciocínio da Inteligência Artificial, buscando um método mais adequado ao tratamento da imperfeição das informações do Processo Exploratório. Analisa-se a aplicabilidade de algumas metodologias e justifica-se a adoção da Lógica Nebulosa como metodologia capaz de representar as incertezas presentes neste Processo. Descreve-se, então, a implementação de um sistema especialista nebuloso-probabilístico capaz de representar a probabilidade de sucesso de exploração petrolífera.

**Palavras-chave:** Sistemas Especialistas, Lógica Nebulosa (“Fuzzy”), Análise de Risco, Exploração Petrolífera.

## ABSTRACT

Petroleum Exploration is an economical activity where billions of dollars are invested with the expectation of finding petroleum, which is one of the most used sources of energy worldwide. The petroleum Exploration is a difficult job because it involves decision taking on imperfect information, that is, many of the available data are uncertain, undetermined, vague, etc. In spite of the high investment made by petroleum companies in tools that help the exploiting process, it still has many uncertain variables. Besides, it is perceived the presence of factors such as experience, capacity of argument and even a certain dose of intuition from the team that studies the site of the “probable” well, thus making knowledge representation difficult in the petroleum Exploration.

The present work emerged as the fruit of a project called “RCSUEX – Representação da Certeza de Sucesso Exploratório (Exploiting Success Certainty Representation)”, developed by the Conexionism and Cognitive Sciences Laboratory (L3C) at UFSC – Universidade Federal de Santa Catarina for Petrobras, from the necessity of this latest to structure the representation of uncertain knowledge in the Petroleum Exploration Process. The Project RCSUEX aim is to develop an Expert System that estimates the risks of petroleum Exploration, on geological and statistic data to a new well, with the purpose of systematizing the arguing and measuring the probability of success of a petroleum accumulation discovery.

This work studies the Artificial Intelligence reasoning methods, looking for a method that is more appropriate to the treatment of the Exploiting Process information imperfections. It is analyzed the application of some methodologies, and the adoption of fuzzy logic is justified as a methodology that is able to represent the uncertainties that are present in this Process. It is described, then, the implementation of a probabilistic-fuzzy expert system that is able to represent the probability of success in petroleum Exploration.

**Keywords:** Expert Systems, Fuzzy Lógica, Petroleum Exploration, Decision Analysis.

## Sumário:

1	Introdução.....	15
1.1	Objetivo Geral .....	17
1.2	Objetivos Específicos .....	17
1.3	Estruturação do texto .....	18
2	Exploração Petrolífera .....	19
2.1	Introdução.....	19
2.2	O processo de Análise do Risco .....	19
2.3	Componentes do Sistema Petrolífero .....	21
2.4	Componentes Geológicos de um Prospecto .....	25
2.4.1	Rocha Geradora .....	25
2.4.2	Migração.....	26
2.4.3	Reservatório.....	27
2.4.4	Geometria (Trapa) .....	28
2.4.5	Selo (Retenção) .....	28
2.4.6	Sincronismo .....	29
2.5	Tecnologia de DHI .....	30
3	Representação de Imperfeições .....	31
3.1	Introdução.....	31
3.2	Mecanismos de Raciocínio.....	33
3.2.1	Abordagem Simbólica .....	33
3.2.1.1	Raciocínio Inferencial: Regras de Produção .....	33
3.2.1.2	Raciocínio Probabilístico (Redes Bayesianas) .....	34
3.2.1.3	Raciocínio Possibilista.....	36
3.2.1.3.1	Conjuntos Nebulosos (“Fuzzy”).....	36
3.2.1.3.2	Distribuição de Possibilidade .....	37
3.2.1.4	Raciocínio por Crença (Teoria de Dempster-Shafer – TDS) .....	39
3.2.1.5	Fatores de Certeza .....	41
3.2.1.6	Raciocínio Baseado em Casos .....	42
3.2.2	Abordagem Conexcionista .....	43
3.2.2.1	Redes Neurais.....	43
4	O Projeto RCSUEX.....	46



4.1	Introdução.....	46
4.2	O que já existe na Indústria do Petróleo .....	48
4.2.1	A Metodologia Chevron .....	49
4.2.2	O Projeto Fuzzy Expert Exploration Tool.....	50
4.3	Estudos das alternativas para tratamento de dados subjetivos .....	51
4.3.1	Redes Neurais .....	51
4.3.2	Redes Bayesianas .....	53
4.3.3	Lógica Nebulosa.....	54
4.4	Compatibilidade dos Dados.....	55
4.5	Metodologias para tratamento de dados Objetivos.....	56
4.5.1	Função de Distribuição Beta.....	56
4.5.2	Simulação de Monte Carlo .....	57
4.5.2.1	Os passos para a realização da Simulação de Monte Carlo.....	58
4.5.3	Incorporação de dados históricos .....	59
5	Conexão entre saída fuzzy e probabilidade .....	61
5.1	Introdução.....	61
5.2	Cálculo do intervalo [a,b] pelo Suporte da Função de Pertinência .....	61
5.3	Cálculo do intervalo [a,b] pelo Ponto de Corte da Função de Pertinência.....	65
5.4	Função de Saída representa a Função de Distribuição de Probabilidade .....	67
5.5	Justificativas da escolha da conexão .....	69
6	O Sistema RCSUEX.....	70
6.1	Introdução.....	70
6.2	O Sistema RCSUEX.....	70
6.3	Representação de dados Subjetivos.....	71
6.4	Os Módulos do Sistema.....	72
7	Conclusão .....	79
7.1	Trabalhos Futuros .....	81
Anexo A .....		82
Lógica Nebulosa.....		82
Introdução.....		82
Conjuntos Nebulosos.....		83
Variáveis Nebulosas .....		84
Operadores dos Conjuntos Nebulosos.....		84

Funções de Pertinência .....	85
Regras Nebulosas .....	87
Sistema da Lógica Nebulosa.....	87
Nebulizar .....	88
Inferência .....	89
Desnebulização .....	90
Aplicações de Lógica Nebulosa: Decisão e Controle.....	91
Referências Bibliográficas.....	93

## Lista de Figuras:

Figura 1: Descobertas de campos de petróleo entre os anos de 1960 a 1999 relacionados ao número de campos descobertos. A escala de novas descobertas decresceu a partir de 1980 (ROSE, 2001). .....	20
Figura 2: Níveis de investigação do Sistema Petrolífero extraído de (ROSE, 2001). .....	22
Figura 3: Exemplo de falha gerada por tremor de terra por onde o petróleo pode vir a migrar. ....	27
Figura 4: Tipos Principais de Trapa: (a) Estrutural, (b) Estratigráfica. ....	29
Figura 5: Ilustra os componentes geológicos de um prospecto [extraído de material publicitário da Petrobras]. ....	30
Figura 6: Ilustração de uma árvore de busca para o raciocínio inferencial. ....	34
Figura 7: Ilustra uma rede Bayesiana. ....	36
Figura 8: Função $\mu(x)$ para graus de pertinência. ....	37
Figura 9: Conjunto nebuloso A = “Alto” extraído de (BITTENCOURT, 1998). ....	38
Figura 10: Conjunto de todas as hipóteses obtidas do conjunto de hipóteses primitivas $\theta = \{ \text{Alergia, Intoxicação, Sarampo, Rubéola} \}$ , extraído de (UCHÔA, 1997). ....	41
Figura 11: Rede de Inferência com fatores de certeza. ....	42
Figura 12: (a) representação do neurônio biológico (KOVÁCS,1996); (b) representação do neurônio artificial (KOVÁCS, 1996) .....	44
Figura 13: Representação da topologia de uma rede neural artificial. ....	45
Figura 14: Tarefas de uma avaliação de prospecto, extraída de Clark (1989). ....	47
Figura 15: Ilustração da representação do processo de definição da probabilidade de um prospecto. ....	48
Figura 16: Mostra a construção de um trapézio da variável de saída nebulosa. ....	62
Figura 17: Representação dos conjuntos nebulosos de saída. ....	62
Figura 18: Caso onde o valor de desnebulizado está entre dois valores lingüísticos. ....	64
Figura 19: Obtenção do intervalo de probabilidade pelos pontos de intersecção entre a área de representação do trapézio e a força de ativação das regras ativadas. ....	66
Figura 20: Representação das forças de ativação das regras de inferência Nebulosa. ....	66
Figura 21: Ilustração da obtenção do intervalo probabilístico quando mais de uma variável lingüística está ativada. ....	67
Figura 22: Ilustra o segundo passo onde se trata da Normalização da distribuição de probabilidade do fator. ....	68

Figura 23: Ilustração do menu principal do Sistema RCSUEX. ....	73
Figura 24: Ilustra a tela de saída da Probabilidade Subjetiva do Sistema RCSEUX .....	75
Figura 25: Representa a saída do Módulo de Indicação de Sucesso a Posteriori: (a) o botão histórico gera um gráfico sobre os valores de poços pioneiros e descobertos; (b) incorpora dados históricos sobre os dados subjetivos, não aplica DHI; (c) incorpora dados históricos à saída do Módulo Subjetivo e DHI. ....	77
Figura 26: Ilustra a saída do módulo de incorporação de DHI: (a) incorpora a DHI sobre os dados subjetivos; (b) incorpora DHI sobre os dados subjetivos operados com dados objetivos. ....	78
Figura 27: (a) Conjuntos Tradicionais (b) Conjuntos Nebulosos (FUZZY,2002).....	83
Figura 28: Conjuntos para a variável Temperatura (FUZZY,2002).....	86
Figura 29: Formas mais comuns da função de pertinência. ....	86
Figura 30: Sistema de Lógica Nebulosa.....	88
Figura 31: Nebulização da variável Temperatura(FUZZY,2002).....	89
Figura 32: Aplicação do método do centróide(FUZZY,2002).....	91

**Lista de Tabelas:**

Tabela 1: a Tabela de Probabilidade Condicional de um nodo. ....	36
Tabela 2: Ilustra os graus de Possibilidade e necessidade para a suposição de $p$ extraído de (BITTENCOURT, 1998).....	39
Tabela 3: Inferência de Regras Nebulosas . ....	89
Tabela 4: Inferência de Regras Nebulosas: selecionar o valor do conjunto Nebuloso de saída. .....	90

**Abreviações:**

SE – Sistema Especialista

UFSC – Universidade Federal de Santa Catarina

RCSUEX – Representação da Certeza de Sucesso Exploratório

IA – Inteligência Artificial

IAS – Inteligência Artificial Simbólica

IAC – Inteligência Artificial Conexionista

IAE – Inteligência Artificial Evolucionária

IAH – Inteligência Artificial Híbrida

RB – Redes Bayesianas

RBC – Raciocínio Baseado em Casos

TDS – Teoria Dempster-Shafer

CF – Fatores de Certeza

DHI – Direct Hydrocarbon Indicators

AVO – Amplitude versus offset

PDF – Função de Densidade Beta

## 1 Introdução

A principal fonte de energia do mundo, hoje, é o petróleo<sup>1</sup>. A exploração petrolífera é uma das grandes atividades econômicas mundiais e envolve investimentos de bilhões de dólares por ano. Apesar do grande e maciço investimento da indústria do petróleo na área de exploração petrolífera, esta é uma atividade complexa que envolve informações imperfeitas e análise de riscos.

Antigamente, quando se tratava de exploração de novos poços, a indústria do petróleo observava apenas variáveis geológicas mais relevantes (volumes, dinâmicas das acumulações, etc.) e algumas tendências gerais do mercado (preço do óleo, demanda de derivados, etc.), e não se preocupava com o uso de metodologias complexas e mais abrangentes, devido à facilidade de descoberta de boas jazidas. Entretanto, esse cenário alterou-se drasticamente em função da diminuição dos indícios de jazidas de petróleo, da crescente globalização dos negócios e do envolvimento de diversos agentes, tornando o processo de tomada de decisão na exploração bastante complexo e nem sempre de fácil solução (SUSLICK, 2003).

Desta forma, durante os anos noventa, muitas companhias de petróleo internacionais incorporaram às suas atividades exploratórias, princípios de análise de risco e administração, em combinação com novas tecnologias. O estudo de análise de decisão considera o risco e incerteza de uma maneira quantitativa e provê meios para incorporar-se a dimensão de risco em uma estratégia de decisão lógica e consistente sob condições de incerteza (NEWENDORP, 2000).

A atividade exploratória trata de informações imperfeitas quando da avaliação dos dados geológicos. Os especialistas do petróleo estudam variáveis críticas da dinâmica da acumulação petrolífera, como geração, reservatórios, migração, sincronismo entre outras, e estas variáveis estão repletas de dados imperfeitos, isto é, o valor de uma variável pode ser um intervalo de diferentes valores possíveis e aceitáveis.

Os especialistas do petróleo buscam indícios de jazidas de petróleo em áreas distantes vários quilômetros da costa terrestre, a uma grande profundidade abaixo do nível do mar. Navios enviam mapas sísmicos de áreas a serem avaliadas como propensas à perfuração de novos poços de petróleo (prospectos). Mas a falta de uma “assinatura” ou de um dado que represente com certeza a existência de petróleo faz que os intérpretes trabalhem com inferências, indícios e analogias. O especialista possui mapas sísmicos onde tenta enxergar

---

<sup>1</sup> Petróleo é uma mistura na forma gasosa, líquida ou sólida de hidrocarboneto que ocorre naturalmente na terra.

semelhanças na estrutura estratigráfica que justifiquem a perfuração de um novo poço petrolífero. Muitas vezes o especialista consegue amostras análogas de outros poços perfurados sobre a mesma área (o que os especialistas do petróleo chamam de “trends”) que ajudam a sustentar ainda mais fortemente sua análise. A indústria do petróleo tem conseguido grande auxílio tecnológico para a área de exploração petrolífera, entre estes, o estudo do AVO (“Amplitude versus offset”) e a aplicação de DHI (“Direct Hydrocarbon Indicators”), que resultaram em um grande avanço na descoberta de novos campos de petróleo, já que quando presentes são indícios fortes de hidrocarboneto<sup>2</sup>.

Desta forma, o especialista trabalha com vários dados subjetivos (dados geológicos), com diferentes níveis de imperfeição (vago, incompleto, impreciso, etc.). Além disso, muitas vezes, nem todos os dados estão disponíveis ao especialista, tornando cada decisão exploratória diferente uma da outra.

A forma de trabalho observada na exploração petrolífera ainda é muito intuitiva, sendo que os especialistas criam uma argumentação desestruturada, e de forma descritiva justificam suas avaliações sobre uma determinada área. Este conhecimento não é representável computacionalmente, e as informações coletadas para o estudo de uma área ficam restritas aos especialistas que estudaram a área. Primeiramente, isto ocasiona perda de informações, porque o conhecimento fica somente para a equipe que estudou a área e não para a empresa. E, por conseguinte, existirá um abismo entre o conhecimento do especialista novato e de um especialista mais experiente, fazendo que o novato para adquirir novos conhecimentos tenha de passar por várias experiências, gerando, assim, um preço bastante alto para a empresa ao formar especialistas.

Ainda que exista uma lista de variáveis (“check-lists”) para a justificativa de uma nova jazida de petróleo, estas são metodologias qualitativas geralmente pouco flexíveis que incorrem no erro de tratar áreas com características diferentes como tendo a mesma avaliação final.

A Inteligência Artificial (IA), por sua vez, disponibiliza um conjunto de metodologias para representação e tratamento das várias formas de imperfeição da informação. O emprego da metodologia adequada deve ser capaz de prover os meios para representar, estruturar e sistematizar o mecanismo de raciocínio envolvido no processo de análise de risco exploratório.

---

<sup>2</sup> O termo hidrocarboneto é comumente usado na indústria do petróleo para indicar óleo cru (bruto) ou gás natural. Em termos químicos, hidrocarboneto é composto de carbono e hidrogênio.



## 1.1 Objetivo Geral

O objetivo geral deste trabalho de dissertação é estudar a exploração petrolífera, a fim de desenvolver um Sistema Especialista denominado Sistema RCSUEX, que auxilie os intérpretes na estimativa do risco exploratório de um prospecto (novo poço). Deve-se contribuir, também, para que a exploração petrolífera trabalhe sobre uma perspectiva mais racional, estruturada, representável computacionalmente e que trate os conhecimentos conforme suas imperfeições.

## 1.2 Objetivos Específicos

Para que os objetivos gerais sejam atingidos, devem-se cumprir alguns objetivos específicos, tais como:

- mapear o processo de exploração petrolífera, entendendo o problema e identificando os conhecimentos que dirigem o processo de análise e decisão de prospecção de poços de petróleo;
- identificar o tipo da imperfeição do conhecimento envolvida neste processo;
- analisar e prototipar os métodos e ferramentas de tratamento de informações imperfeitas adequados para tratamento do conhecimento do processo exploratório;
- analisar os resultados dos protótipos e desenvolvimento da proposta do modelo de conhecimento e raciocínio do problema;
- sugerir uma metodologia que integre argumentação geológica, informação visual e estatística exploratória, contribuindo para a estruturação da estimativa do risco da exploração petrolífera;
- encontrar e justificar uma conexão entre os dados subjetivos e objetivos;
- projetar as interfaces, rotinas e relatórios do Sistema RCSUEX;
- implementar o sistema computacional;
- aferir e calibrar o Sistema RCSUEX de forma que:
  - dados os atributos de um poço que a priori se sabe ser seco, a resposta fornecida pelo Sistema seja um valor de favorabilidade bastante baixo;
  - para atributos de um prospecto que se sabe conter acumulação, o Sistema responderá com um valor de favorabilidade bastante alta.

Esta dissertação é baseada no projeto RCSUEX – Representação da Certeza de Sucesso Exploratório, desenvolvido pela Universidade Federal de Santa Catarina - UFSC, em convênio com a Petrobrás.

### **1.3 Estruturação do texto**

Esta dissertação está dividida em sete capítulos, da seguinte forma:

1. Introdução descreve-se a motivação, justificativa e os objetivos;
2. A Exploração petrolífera – neste capítulo descreve-se a atividade de exploração petrolífera, suas características e ferramentas;
3. Classes de Imperfeição – com o entendimento da atividade petrolífera, buscam-se metodologias que tratem das imperfeições dos conhecimentos envolvidos. Discute-se metodologias, vantagens e desvantagens;
4. O Projeto RCSUEX – neste capítulo descreve-se as alternativas do Projeto para a avaliação de um novo prospecto;
5. Conexão entre Saída Nebulosa e Probabilidade – este capítulo explica a conexão entre os dados subjetivos e objetivos;
6. O Sistema RCSUEX – neste capítulo definem-se as funcionalidades do Sistema, com suas características e módulos.
7. Conclusão e trabalhos futuros – encerra-se esta dissertação com as conclusões e propostas de novos trabalhos que no decorrer do seu desenvolvimento foram surgindo.

## **2 Exploração Petrolífera**

### **2.1 Introdução**

A exploração petrolífera começou em 1859 no estado da Pensilvânia, nos Estados Unidos e de lá para cá muitos modelos geológicos foram propostos com diferentes níveis de investigação.

Para que o entendimento do processo exploratório fosse mais bem compreendido e com maior rapidez por parte da equipe da UFSC, realizou-se uma visita à E&P na EDISE (prédio sede da Petrobras no Rio de Janeiro), entre os dias 18 e 22 de março de 2002. Durante esta semana foram agendadas diversas reuniões, contando com a presença de geólogos, geofísicos, intérpretes, entre outros profissionais da Petrobras envolvidos no processo exploratório e os profissionais da UFSC. Os profissionais da Petrobras apresentaram as etapas do processo exploratório, as ferramentas por eles utilizadas, a forma como manipulavam as informações desde a sua fase de análise até sua decisão final, ou seja, se a locação proposta vai à carteira (isto é, o poço será perfurado) ou é rejeitado (isto é, inviabiliza o poço).

Este capítulo tem por objetivo descrever o processo de análise de um Sistema Petrolífero, principalmente os fatores geológicos, nos quais estão os conhecimentos e informações envolvidas na solução do problema do Projeto RCSUEX.

### **2.2 O processo de Análise do Risco**

A exploração petrolífera é uma atividade que requer muito investimento e está cheia de incertezas. Apesar do alto investimento feito pelas companhias petrolíferas em ferramentas que auxiliem o processo exploratório, estas ainda possuem muitas variáveis imperfeitas<sup>3</sup>.

Exploração Petrolífera, como muitas atividades empresariais, é um processo repetitivo sob condição de incertezas, sendo que cada exploração requer um investimento substancial. Durante os anos noventa, muitas companhias de petróleo internacionais melhoraram o desempenho de suas explorações significativamente usando princípios de análise de risco e administração, em combinação com novas tecnologias. Contudo, se o risco da exploração não pode ser eliminado, pode ser, certamente, reduzido substancialmente (ROSE, 2001).

Em meados dos anos 80, renomadas empresas internacionais de petróleo que estavam engajadas na exploração confessaram que os volumes dos novos poços de petróleo

---

<sup>3</sup> A imperfeição é caracterizada pelo fato de que os dados analisados são indiretos, porque não se pode enxergar se o petróleo está ou não no local estudado. Os métodos de análise, como interpretação visual de mapas sísmicos, geram variáveis indiretas, que são em suma variáveis imperfeitas. No capítulo 3 explica-se a representação de informações imperfeitas e suas metodologias de IA.

descobertos estavam diminuindo (Figura 1). Não coincidentemente, a classe dos prospectos categorizados como "de alto risco / (mas) alto-potencial" estavam mostrando sinais de baixa atuação (ROSE, 2001).

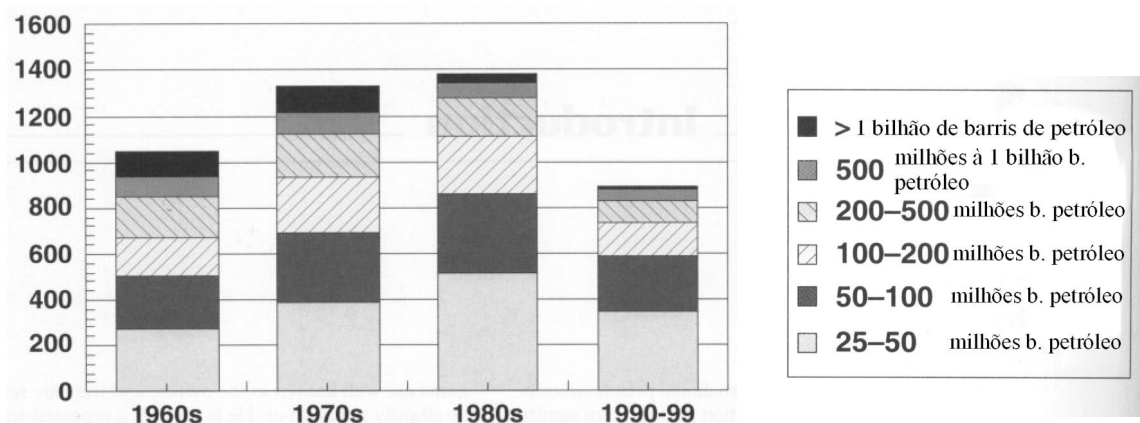


Figura 1: Descobertas de campos de petróleo entre os anos de 1960 a 1999 relacionados ao número de campos descobertos. A escala de novas descobertas decresceu a partir de 1980 (ROSE, 2001).

A exploração petrolífera é uma cadeia produtiva vinculada a modelos geológicos<sup>4</sup> de acumulação, que se esgotam. As grandes companhias buscam novos poços, usando sempre o modelo geológico de acumulação da área estudada, mas em determinados momentos esse modelo geológico se esgota e inicia-se um declínio na descoberta de novos poços e no volumes dos poços descobertos. O modelo geológico mudou bruscamente quando a indústria do petróleo passou de bacias terrestres para bacias marítimas (ambiente de águas profundas) (SUSLICK, 2003).

Mesmo em bacias marítimas existem modelos geológicos que se esgotam e as empresas declinam em suas descobertas, como mostra a Figura 1 nos anos 80. Nesta época as principais empresas, como Shell, Mobil e Amoco, declinaram o número de suas descobertas e o volume das poucas descobertas da época. Ao todo, a exploração de grandes poços foi menor nos anos oitenta e início dos anos noventa (ROSE, 2001).

Por conseguinte, a exploração petrolífera, como uma função incorporada, teve sua credibilidade perdida. A exploração petrolífera precisou ficar mais eficiente, e assim mais lucrativa (ROSE, 2001). Assim, introduziu-se nova tecnologia, e as empresas perceberam que

<sup>4</sup> Modelo Geológico são características que se acredita ser favorável para a descoberta de petróleo. Contudo estas características dependem da área que se está buscando petróleo, pois em águas profundas existem características que não indicam petróleo, por exemplo, em áreas terrestres. Com o passar do tempo este modelo geológico se desgasta, pois todos as jazidas com esta característica já foram descobertas e deve-se procurar por outra característica na busca de novas jazidas.

precisavam adotar procedimentos sistemáticos para melhor administrarem a função de exploração.

No início dos anos 90, muitas empresas começaram a adotar métodos de análise de risco, trazendo objetividade e consistência para avaliarem novos prospectos. A análise de risco utiliza princípios da Estatística, Teoria da Probabilidade e Teoria da Utilidade, as quais começaram a ser reconhecidas como subdisciplinas da matemática e filosofia, através de publicações de Pascal e Fermat, Graunt e Petty (1662), Halley (1763), Jacob Bernoulli (1713), De Moivre (1733), Daniel Bernoulli (1738) e Bayes (1764). O incentivo destes estudos veio de problemas como jogos de azar. Pode-se fazer uma analogia entre a exploração petrolífera e um cassino, visto que ambos são processos repetitivos sob condições de incerteza, que a cada momento requer um investimento e que não se sabe das chances de sucesso e do tamanho do prêmio até que o jogo acabe (KARBAUGH,1995). Por estes motivos, a indústria do petróleo busca ferramentas que melhorem o “palpite” dos seus técnicos ao propor um novo poço. O Sistema Petrolífero é o estudo que descreve a relação de um conjunto de rochas geradoras e o resultado de acumulação de óleo ou gás. O sistema petrolífero pode ser usado como um modelo eficiente de investigação para a descoberta de acumulação de petróleo.

A exploração petrolífera abrange diversas atividades, desde o estudo geológico até cálculos econômicos que vão fornecer subsídios para justificar ou não a perfuração de um poço, porque muitas vezes o poço possui hidrocarboneto, mas a sua quantidade não cobre os custos de extração. Faz-se necessário localizar o espectro de atuação do Projeto RCSUEX, pois não é de intenção desta dissertação abranger toda atividade de exploração.

### **2.3 Componentes do Sistema Petrolífero**

O Processo de exploração petrolífera possui quatro níveis de investigação:

- Bacias sedimentares,
- Sistema petrolífero,
- Play e
- Prospecto.

A bacia sedimentar descreve a seqüência estratigráfica e estilo estrutural das rochas sedimentares. O sistema petrolífero pode ser usado como um modelo de investigação para a descoberta de acumulações de petróleo, pois ele descreve a relação entre um conjunto de rochas geradoras e o resultado de acumulações de óleo ou gás. Os elementos essenciais de um sistema petrolífero são: rocha geradora, rocha reservatório e rocha selante. Já os processos de

um sistema petrolífero são: formação da trapa, geração-migração e sincronismo. O sistema petrolífero tem estratigrafia, geologia e uma evolução temporal.

Play é um ou mais prospectos, e um prospecto é uma situação geológica favorável para conter acumulação de petróleo. Um prospecto deve ser definido como um local específico dentro de uma área, o qual acredita-se ter características geológicas ou econômicas que justifique a perfuração (SILVA, 2000).

Bacia sedimentar e sistema petrolífero não possuem importância econômica, mas são essenciais à avaliação do play ou do prospecto. Já o play e o prospecto possuem importância econômica, pois o sistema petrolífero pode indicar petróleo, isto é, ter chances de haver petróleo, mas este pode ser de pouca qualidade ou um volume pequeno, o que inviabiliza sua extração. A Figura 2 apresenta um resumo dos quatro níveis de investigação do Petróleo.

O Processo de exploração petrolífera envolve várias pessoas, entre elas:

- geólogos,
- geofísicos,
- intérpretes.

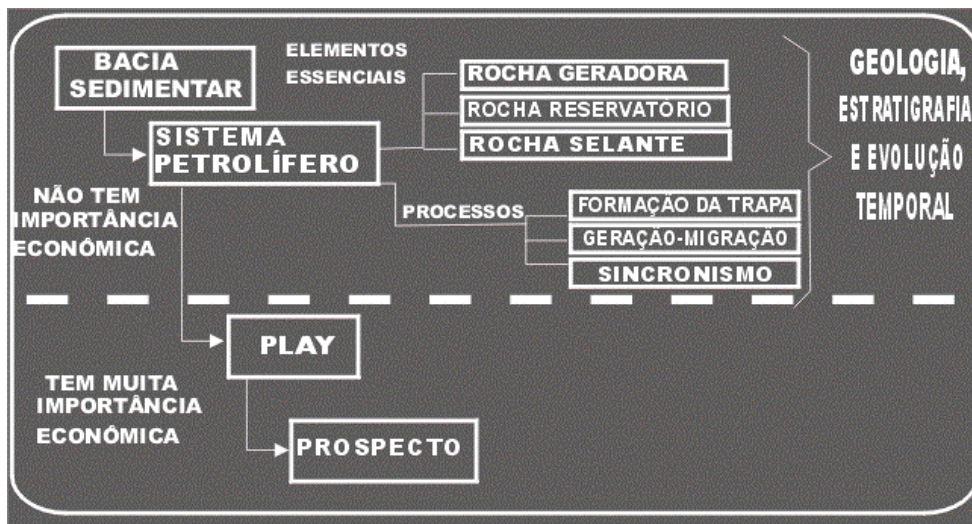


Figura 2: Níveis de investigação do Sistema Petrolífero extraído de (ROSE, 2001).

Estas pessoas formam uma equipe que trabalha um longo tempo sobre uma determinada região ou bacia. O estudo da bacia é feito através de mapas de linhas sísmicas, analogias com outras regiões e, principalmente, do conhecimento de cada integrante da equipe sobre a região (experiências passadas).

A partir do momento em que a equipe recebe uma nova região para analisar, ela passa a procurar características geológicas que indiquem a presença de hidrocarboneto na região.

Algumas dessas características geológicas não estão claras e nem podem ser mensuradas quantitativamente (RENEU, 2000).

Existem três tipos de fontes de informação sobre esta área:

- Informações sísmicas;
- Informações análogas a outros poços da área;
- Teoria Geológica.

Dados sísmicos fornecem, entre outras coisas, informações sobre modelos de camadas de rochas, sendo que estes dados são aproximados e imprecisos. Quando é possível fazer-se analogia com outros poços da região, as informações são fornecidas com maiores detalhes, tais como, profundidade e tipo de rocha, e estas informações são projetadas ao prospecto que está sendo estudado. Junto a estas duas fontes de informações aplica-se a Teoria Geológica, que fornece dados da área em estudo.

Dispondo destes dados, o geólogo busca características que provem a existência de um novo poço na área estudada. Estes dados são agrupados de forma que caracterizem cada elemento (Rocha Geradora, Rocha Reservatório e Rocha Selante), ou processo (formação da trapa, geração-migração e sincronismo), e passam a ser chamados de fatores geológicos. A argumentação dos geólogos, intérpretes e geofísicos é puramente qualitativa, ou seja, eles descrevem características que comprovem a favorabilidade de cada um desses fatores geológicos que formam o Sistema Petrolífero.

Após os estudos feitos, utilizando as fontes de informações disponíveis (dados sísmicos, teoria da geologia e analogia), a equipe deve resumir sua crença na existência de hidrocarboneto sob um número probabilístico que represente cada um dos fatores geológicos. Multiplicando-se a probabilidade dos fatores geológicos, a equipe obtém a probabilidade do prospecto.

A estimativa da reserva do prospecto é importante porque muitas vezes o hidrocarboneto existe, mas seu volume não justifica a sua extração, ou a qualidade do hidrocarboneto é ruim e não retornaria o investimento de perfuração e extração. O interessante é que haja o sucesso geológico porque, afinal de contas, existe o hidrocarboneto onde o geólogo afirmou existir, apenas pode não ser rentável sua extração.

Para a realização destas estimativas econômicas é necessário que a equipe de exploração justifique a favorabilidade de sucesso do novo prospecto sob forma de uma probabilidade. A transição entre a argumentação qualitativa do geólogo para um número probabilístico não é nada fácil, uma vez que para o geólogo este número probabilístico pouco significa. Na verdade o geólogo se interessa pelo sucesso geológico no sentido provar ou não

a existência do hidrocarboneto na área estudada, sem se importar com cálculos econômicos. Existe uma grande dificuldade, por parte do geólogo, em expressar a probabilidade de cada fator geológico, visto que toda a argumentação é qualitativa.

O passo seguinte é levar este prospecto a uma reunião gerencial, onde o geólogo argumenta sobre suas crenças no potencial do prospecto, a fim colocar o prospecto em um “portfólio” ou carteira, onde se aguarda o melhor momento para a sua perfuração. A escolha de qual prospecto perfurar primeiro, neste “portfólio”, requer alguns critérios como:

- a proximidade de sondas e outros equipamentos necessários para perfuração, de modo a serem evitados gastos com transporte de equipamentos;
- uma maior probabilidade de sucesso deste poço ter acumulações de hidrocarboneto;
- quantidade prevista do volume de petróleo.

Uma das maneiras de saber qual prospecto tem maior probabilidade de sucesso de acumulação de hidrocarboneto é através do valor da probabilidade associado ao prospecto. Este número probabilístico indica qual o fator de certeza da equipe em encontrar hidrocarboneto naquele prospecto, levando em consideração o estudo dos fatores geológicos. Quanto maior o valor probabilístico maior a crença da equipe na existência de petróleo ou gás natural. Este valor de probabilidade é calculado pela multiplicação dos seis fatores geológicos que serão explicados detalhadamente na seção 2.3:

- Rocha Geradora,
- Migração,
- Rocha Reservatório,
- Selo (Retenção),
- Geometria (Trapa),
- Sincronismo.

Estes seis fatores geológicos são variáveis dependentes. À medida que um deles for zero, a probabilidade de sucesso geológico será zero. Nem todas as empresas de exploração petrolífera trabalham com estes seis fatores geológicos. Muitas empresas trabalham apenas com quatro ou cinco destes fatores, pelo fato de ter uma maior abrangência de itens dentro dos



fatores escolhidos para trabalhar. Por exemplo, Jim A Mackey<sup>5</sup> sugere que se trabalhe com apenas quatro fatores:

- Carga de hidrocarboneto (engloba qualidade de geradora, geração, migração, sincronismo e preservação);
- Segurança estrutural;
- Selo (Retenção);
- Reservatório.

Pelo fato de este trabalho estar sendo direcionado à Petrobras, vai-se trabalhar com os seis fatores.

## **2.4 Componentes Geológicos de um Prospecto**

As descrições dos componentes geológicos, apresentadas nesta sessão, foram extraídas de leituras em (ROSE, 2001) e (MAGOON,1994), além das conversas durante a semana de palestras realizadas pela Petrobras.

### **2.4.1 Rocha Geradora**

O hidrocarboneto hoje utilizado foi gerado muito tempo atrás. O hidrocarboneto é formado por organismos antigos (fósseis, vegetais) que foram soterrados e convertidos em petróleo pelos efeitos combinados de calor e do tempo. Os materiais orgânicos soterrados são chamados de kerogen, e a fonte de petróleo é qualquer rocha que contenha bastante kerogen para gerar óleo e gás.

Kerogen é uma mistura de produtos de degradação de material orgânico e restos de vários tipos de plantas e de animais, que pode ser distinguida pela sua química e pela sua morfologia que refletem ao uso de um microscópio petrográfico (CLARK, 1989).

Como o próprio nome diz, Rocha Geradora é a rocha onde o óleo é gerado (produzido). Para descrever uma rocha geradora usa-se os termos “marine” ou “terrigenous”, que se referem à proveniência (origem) ou ao ambiente deposicional.

Rochas geradoras devem satisfazer três exigências de geoquímica:

- quantidade de matéria orgânica;
- qualidade ou tipo da matéria orgânica;
- maturidade termal da matéria orgânica.

---

<sup>5</sup> Jim A. Mackay é membro da AAPG – American Association of Petroleum Geologists – e possui diversos trabalhos publicados sob o tema Exploração Petrolífera. Concedeu uma palestra no Auditório CECO na UFRGS – Universidade Federal do Rio Grande do Sul, no dia 07-05-2002, intitulada “Exploration Risk Assessment: What is it? How is it Done? Who is Responsible?”.

Maturidade termal se refere à extensão de temperatura-tempo dirigida a reações que convertam a matéria orgânica sedimentar (rocha geradora) em óleo, gás úmido, e, finalmente, gás seco. Imaturidade termal da rocha geradora ocorre por causa da mudança química, biológica e física na matéria orgânica durante e depois da deposição dos sedimentos, mas a priori isso só ocorre ao se alcançarem temperaturas de soterramento maiores que 60-80°C, aproximadamente. A quantidade e qualidade da matéria orgânica determinam o potencial da rocha geradora. A maioria das rochas geradoras é do tipo xisto, com um conteúdo orgânico total (TOC) de pelo menos 3%.

O geólogo está estudando uma área para determinar a presença de Rocha Geradora, contudo ele não está fisicamente na área estudada e tampouco tem acesso à matéria orgânica do local para fazer testes. As rochas geradoras estão a quilômetros de profundidade no mar, e o geólogo trabalha com dados sísmicos<sup>6</sup> e testes laboratoriais feitos em poços já perfurados próximos à área estudada.

#### **2.4.2 Migração**

A migração é o processo pelo qual o óleo passa de uma camada mais profunda, onde está a rocha geradora, para uma camada mais superficial. Este movimento é feito através das falhas no solo geradas por tremor de terra, como ilustra a Figura 3. A força motriz para migração é a força de fluatibilidade vertical devido à mais baixa densidade de petróleo comparada à água de formação.

Tipos de Migração:

- Migração primária;
- Migração secundária.

Migração primária é o movimento do petróleo ainda na Rocha Geradora.

Migração secundária é o processo pelo qual o petróleo é transportado do conjunto de rocha geradora ativa para a trapa, ou melhor, é o movimento de petróleo além do ponto onde o hidrocarboneto existe - a rocha geradora ativa. O destino deste hidrocarboneto migrando é dependente da eficiência do canal pelo qual ele está se movendo e a natureza e eficiência da trapa a qual ele deve encontrar. Uma trapa inclui uma rocha reservatório e uma rocha selante

---

<sup>6</sup> Quando tremor de terra ou explosão acontece ondas sísmicas se propagam pelo chão do oceano do mesmo modo que ondulam sobre a água. Rochas diferentes refletem freqüentemente de ondas diferente, de modo que um navio ao passar sobre a área a ser examinada pode gerar um plano com um perfil do chão do oceano através das diferentes freqüências de ondas sísmicas refletidas do fundo do mar.

que estão em uma configuração tridimensional capaz de impedir ou armazenar petróleo abaixo da superfície.

Na migração podem ocorrer muitos erros no momento de analisar o seu percentual de sucesso, pois as falhas geológicas por onde o petróleo pode passar, muitas vezes, possuem desvios por fissuras na rocha não perceptíveis à análise do geólogo.

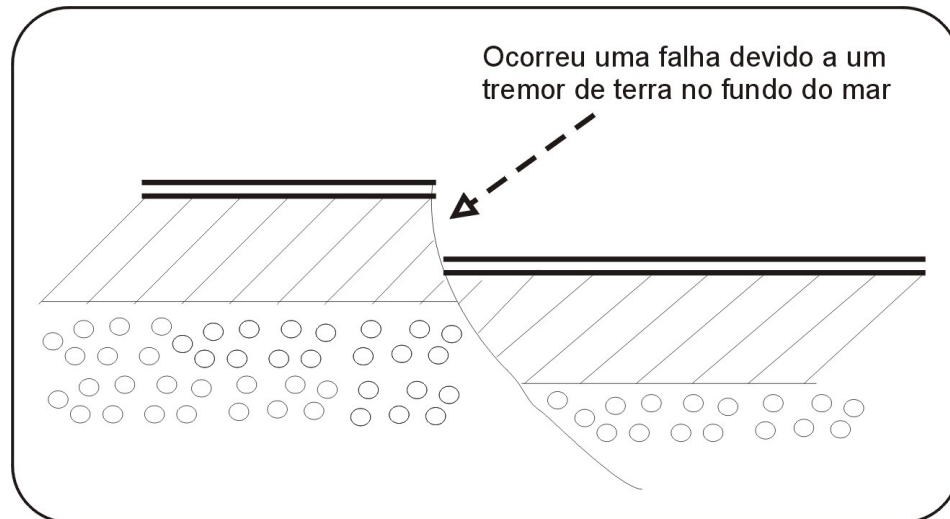


Figura 3: Exemplo de falha gerada por tremor de terra por onde o petróleo pode vir a migrar.

### 2.4.3 Reservatório

Reservatório é uma unidade estratigráfica de rochas capaz de armazenar petróleo comercialmente (SILVA, 2000). Consiste em uma pedra porosa que segura o óleo ou gás, e uma pedra que encapsula ou cobre, que é a rocha selante, prevenindo a saída do óleo ou gás.

Caso uma rocha tenha bastante porosidade e permeabilidade para fluir óleo ou gás, então é um reservatório de potencial. Porosidade é a porcentagem de vácuo em uma pedra comparada à formação sólida e é determinada pelo tamanho dos grãos que formam a rocha. Já permeabilidade é a propriedade de uma formação que quantifica o fluxo de um fluido pelo espaço do poro.

Contudo, muitas vezes a rocha possui porosidade e permeabilidade, mas não contém petróleo e sim água. Se o conteúdo de água no reservatório for maior, então o óleo tende a ficar atrás e o reservatório produz somente água. Se a saturação de água é menor, então o reservatório pode ser produtivo.

#### **2.4.4 Geometria (Trapa)**

A trapa é uma configuração geométrica de rochas que permitem a acumulação de hidrocarboneto, sendo que vários componentes críticos devem fazer parte desta configuração geométrica para se ter uma trapa efetivamente.

O processo de formação do hidrocarboneto e o processo de formação da trapa ocorrem como eventos independentes e comumente em tempos diferentes. O tempo de formação da trapa é importante no estudo do sistema petrolífero, porque caso a trapa se forme antes do processo de formação do hidrocarboneto, a evidência de que existe o sistema petrolífero é preservada. O volume de óleo ou gás depende do tamanho e tipo da trapa, o qual é importante na avaliação do prospecto.

A combinação de componentes críticos de uma trapa pode ter diversas combinações diferentes, fazendo que existam diferentes tipos de classificação de trapas. Diferentes autores focam vários atributos da trapa como elemento-chave de suas classificações. Alguns autores enfatizam a geometria da trapa, enquanto que outros se concentram no mecanismo de formação da trapa. Outros consideram o reservatório ou característica de selo como a principal parte de classificação.

Para ser uma trapa viável, uma superfície deve ser capaz de receber hidrocarboneto e armazená-lo por algum tempo. Isto requer dois componentes fundamentais: rocha reservatório, na qual se armazena o hidrocarboneto, e selo, que guarda o hidrocarboneto vindo da migração para a trapa.

#### **2.4.5 Selo (Retenção)**

O selo é um componente crítico da trapa. Sem um selo efetivo o hidrocarboneto irá migrar para a rocha reservatório, mas não irá acumular.

A extensão geográfica das rochas de selo define os limites efetivos do sistema de petróleo. Onde os limites das rochas selos estão faltando, o hidrocarboneto escapa à superfície, desta forma a rocha selo é um elemento essencial do sistema de petróleo.

Muitos selos de acumulações de hidrocarboneto são formados na parte mais espessa da estrutura, prolongamentos laterais, rochas maleáveis com alta pressão capilar, contudo existem outros tipos de selos que são importantes, como rochas vulcânicas, camadas de terra congeladas, entre outros. Toda trapa requer alguma forma de selo. Quando a base do topo do selo é convexa nas três dimensões, o contorno do desenho representado por esta superfície é completo.

Muitas trapas são mais complicadas e requerem que, além do topo do selo, outro selo deva estar presente. Selos laterais impedem o hidrocarboneto de se movimentar de um lado para outro na trapa e são mais comumente elementos de trapas chamadas estruturais. Já as trapas chamadas de estruturais são formadas por uma geometria (uma estrutura) que envolve o reservatório, usualmente intervalos de selos permitindo a acumulação de hidrocarboneto e, muitas vezes, são denominadas como: dobras (“folds”), falhas (“faults”), furos (“piercements”) ou qualquer combinação destes anteriores.

Muitas pessoas usam a combinação de trapa (Estrutural), que é um termo menos rigoroso, para aplicar a qualquer trapa que tenha os elementos estratigrafia e estrutura, embora estes não sejam requeridos juntos para a trapa ser viável. A Figura 4 ilustra os tipos de trapa estrutural e estratigráfica.

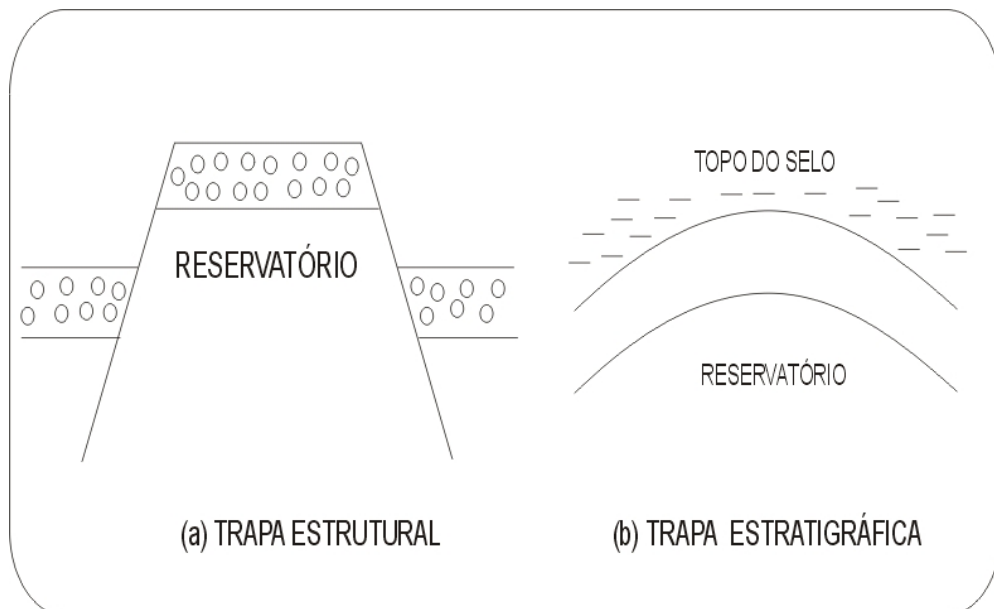


Figura 4: Tipos Principais de Trapa: (a) Estrutural, (b) Estratigráfica.

#### 2.4.6 Sincronismo

O sincronismo é o relacionamento temporal entre rocha geradora madura, reservatório e trapa, especialmente a disponibilidade da trapa ao tempo da carga, isto é, a rocha geradora madura e capacidade de expulsão (migração) para as trapas (ROSE, 2001).

A Figura 5 ilustra todos os componentes geológicos de um prospecto comentado neste capítulo.

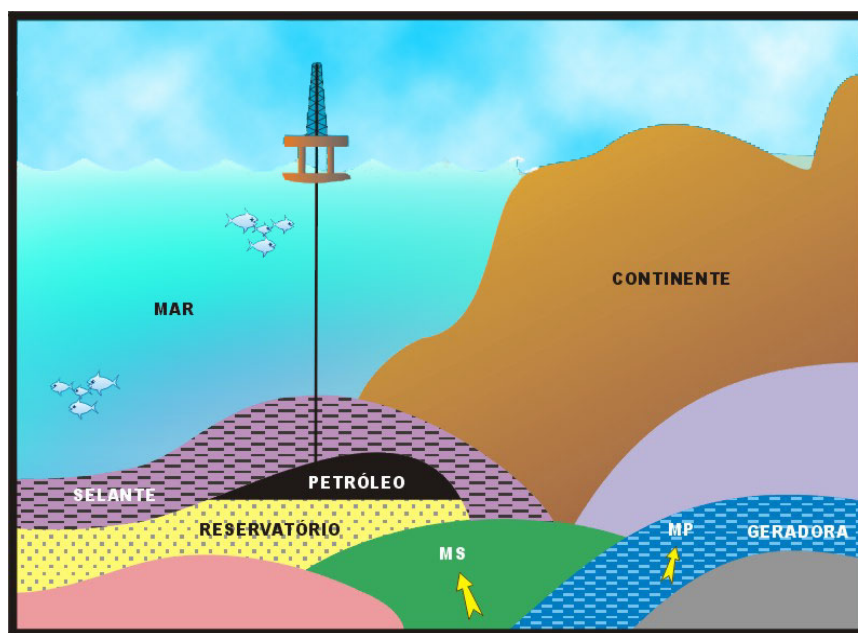


Figura 5: Ilustra os componentes geológicos de um prospecto (extraído de material publicitário da Petrobras).

## 2.5 Tecnologia de DHI

Tecnologias Exploratórias para identificação direta de hidrocarboneto (“Direct Hydrocarbon Indicators”) vêm sendo cada vez mais aplicadas na indústria do petróleo. Visando minimizar riscos geológicos e econômicos na perfuração de poços, elas têm se baseado em princípios geoquímicos e, principalmente, em análises geofísicas como Amplitude Versus Offset, AVO (SILVA, 2000). Esta é uma técnica feita por emissão de ondas a partir de um navio até o fundo do mundo se descrever, para medir o coeficiente de reflexão desta onda ao passar de uma rocha mais mole para outra mais dura. O contraste da variação de uma rocha para outra dá o coeficiente de reflexão que revela a presença de hidrocarboneto.

O DHI é uma evidência de acumulação e, desta forma, ele é aplicado sobre o prospecto como um todo e não em fatores geológicos individuais. A intensidade do DHI pode ser representada através de uma expressão lingüística. Para este trabalho, vai-se tratar dos seguintes termos: “Muito forte”, “Forte”, “Moderado”, “Fraco” e “Muito Fraco”, cada um associado a uma estatística de poços perfurados e ao número de descobertas realizadas.

### 3 Representação de Imperfeições

#### 3.1 Introdução

Muitas vezes ao se deparar com um problema sobre o qual não se possuem todas as informações ou não se sabe se tais informações são completas e precisas, mesmo assim, consegue-se chegar a alguma conclusão sobre o problema. Isto é, tomam-se decisões mesmo com ausência de informações ou ainda com informações conflitantes. Por exemplo, você precisa comprar um presente de aniversário para o seu chefe e você está trabalhando na empresa há apenas um mês, o que você compraria? Primeiramente você observaria o seu chefe para verificar se ele lhe mostra alguma dica. Observaria sua sala e mesa. Pediria a colegas mais antigos informações sobre as preferências do seu chefe. Com estas informações você sairia e compraria algo que acredita que ele iria gostar (com um fator de certeza). Este exemplo ilustra um sistema baseado em conhecimento imperfeito.

Ao utilizar-se do termo sistema baseado em conhecimentos incerto comete-se um erro, pois na verdade existem outras formas de imperfeições da informação, como, por exemplo, imprecisão, conflito, ignorância parcial, etc. (BITTENCOURT, 1998).

Utilizando-se o exemplo da compra de um presente de aniversário ao chefe, apresentam-se alguns tipos de imperfeições da informação:

- 1) Informação perfeita: a premissa é sempre verdadeira: “O chefe toma, somente, vinho tinto Cabernet Sauvignon da cantina XYZ”.
- 2) Informação imprecisa: (vaga, ambígua, indeterminada): “O chefe bebe de vinho.”
- 3) Informação incerta: (hesitante, duvidosa, variável): “Acho que o chefe bebe vinho”.
- 4) Informação vaga: (volúvel, inconstante, indefinida): “Eu vi o chefe comprando vinho no supermercado”.
- 5) Informação probabilística: “É provável que o chefe beba vinho tinto”.
- 6) Informação possibilista: “É possível que o chefe beba vinho tinto”<sup>7</sup>.
- 7) Informação inconsistente: (incoerente) “Um colega de trabalho disse que o chefe bebe uísque, mas outro colega comentou que o chefe bebe vinho”.
- 8) Informação incompleta: “O chefe gosta de um tipo de bebida”.
- 9) Ignorância total: “Não sei do que o chefe gosta”.

---

<sup>7</sup> Note a diferença entre os conceitos de probabilidade e possibilidade: na afirmação probabilista, um evento ocorre ou não, enquanto que em uma afirmação possibilista o evento ocorre ou não é evidente (claro).

Assim como o especialista humano consegue inferir sobre informações imperfeitas (com um grau de certeza), os Sistemas Especialistas baseados em conhecimento devem fazer o mesmo. Para cada uma das imprecisões de informação citada anteriormente se tem uma abordagem da Inteligência Artificial (IA) mais adequada, tanto no modo de manipular o conhecimento quanto em como adquiri-lo, armazená-lo e empregá-lo. Por este motivo a IA subdivide-se em quatro abordagens:

- Inteligência Artificial Simbólica (IAS) – é o resultado da manipulação de símbolos que representam o mundo. Na IAS a inferência se dá pelo uso da Lógica.
- Inteligência Artificial Conexionista (IAC), que se inspira no sistema nervoso responsável pelo comportamento inteligente para solucionar problemas. Esta abordagem trata do estudo dos neurônios e suas conexões, formando as redes neurais.
- Inteligência Artificial Evolucionária (IAE) – utiliza-se de inspiração biológica, assim como a IAC, contudo, trata de seleção natural ou sobrevivência de indivíduos mais aptos.
- Inteligência Artificial Híbrida (IAH) – é uma mescla de todas ou algumas das abordagens anteriores, de forma a resolver um mesmo problema.

A abordagem simbólica é voltada para o tratamento de incerteza em informação incompleta, sendo inadequada para o caso de informação imprecisa, pois não possui mecanismos para quantificar níveis de imperfeição. A abordagem simbólica permite a reconstituição do raciocínio a partir das fontes de informação até as conclusões finais. Essa abordagem utiliza enfoque formal (possui teoria lógica que determina o mecanismo utilizado na avaliação das inferências) ou enfoque heurístico (possui um conjunto de regras dependentes de contexto para manipular estruturas de endosso) (LADEIRA, 1999).

Dentro da abordagem simbólica encontram-se Teoria de probabilidade (estatística bayesiana), fatores de certeza, redes bayesianas e diagramas de influências<sup>8</sup> que são exemplos do enfoque mono valorado; a Teoria da Evidência de Dempster-Shafer exemplifica o enfoque de intervalos; e a Lógica Nebulosa (“Fuzzy”), Teoria das Possibilidades de Zadeh e Normas Triangulares<sup>9</sup> exemplificam sistemas baseados na teoria dos conjuntos Nebulosos (LADEIRA, 1999).

---

<sup>8</sup> Surgiram na década de 80 e podem ser considerados extensões das Redes Bayesianas, porque possuem nodos de decisão e utilidade, permitindo a tomada de decisão. É uma técnica de análise de decisão.

<sup>9</sup> O conjunto de T-normas (intersecção) e T-conormas (União) forma as normas triangulares.



Já a abordagem conexionista geralmente representa incerteza como uma quantidade precisa (único valor ou intervalo) em uma dada escala e requer que o usuário ou especialista faça uma atribuição numérica precisa e consistente para a incerteza dos dados atômicos e das suas relações. Permite definir um cálculo que especifica o mecanismo a ser utilizado para combinar e propagar a incerteza durante o processo de raciocínio. O uso de operadores de agregação de incerteza permite obter uma classificação dos fatos resultantes com certeza resumida, de forma a permitir o seu uso na tomada de decisão. Entretanto, essa abordagem não prevê meios para que se expliquem claramente as razões que conduziram a uma dada conclusão (LADEIRA, 1999).

Portanto, deve-se analisar muito bem o domínio do problema a ser solucionado antes de escolher qual a melhor abordagem ou ferramenta a ser aplicada, porque a escolha da ferramenta interfere na solução do problema. Além disso, tem-se na literatura implementação formal ou “ad-hoc”, para cada um desses modelos (BITTENCOURT, 1998).

## **3.2 Mecanismos de Raciocínio**

### **3.2.1 Abordagem Simbólica**

A IAS obtém a solução do problema manipulando conhecimentos básicos e imitando o modo de raciocínio usado pelos homens. Por isso a IAS é usada quando o problema é bem definido, quando se tem uma boa idéia de como ele deve ser resolvido ou que seja explícito o modo de achar a solução.

#### **3.2.1.1 Raciocínio Inferencial: Regras de Produção**

Raciocínio Inferencial é aquele que se baseia em regras válidas de inferência, ou seja, se inspira no uso dos silogismos enunciados por Aristóteles<sup>10</sup> (BARRETO, 2001). Entretanto, atualmente este raciocínio se apresenta sob formalismos de mais fácil aplicação dentre os quais Regras de Produção.

Em muitos problemas a solução é uma seqüência de passos, ou seja, o conjunto de respostas é um conjunto de seqüências de decisões. Nestes casos é conveniente ver como solucionar o problema sob outro aspecto: o de achar uma seqüência de movimentos em um

---

<sup>10</sup> O silogismo de Aristóteles pode ser definido assim: é um trio de termos, no qual o último, que é a conclusão, contém uma verdade que se chega através das outras duas. A é B, C é A, portanto C é B. O exemplo clássico de silogismo pode ser dado por: (A) Todos os homens são mortais; (B) Sócrates é homem; (C) Logo, Sócrates é mortal.

grafo de decisões (BARRETO, 2001) ou árvore<sup>11</sup>. Pode-se notar que este tipo de problema é de conhecimento total, isto é, tudo que é necessário para solucioná-lo é conhecido, o que facilita sua formalização.

Pode-se exemplificar o raciocínio inferencial através de um jogo do xadrez e, para tanto, especifica-se a posição inicial, as regras que definem movimentos legais (“Se posição atual do tabuleiro, então posição seguinte do tabuleiro”), posições e condições de vitória. A Figura 6 ilustra uma árvore de busca onde a célula em verde representa a posição inicial do tabuleiro e as células em vermelho representam as posições ou condições de vitória.

Este método de raciocínio muitas vezes torna-se inviável, porque necessita da definição completa do problema, sendo que, para o exemplo do tabuleiro de xadrez, tem-se de escrever uma regra distinta para cada uma das aproximadamente  $10^{120}$  posições possíveis do tabuleiro.

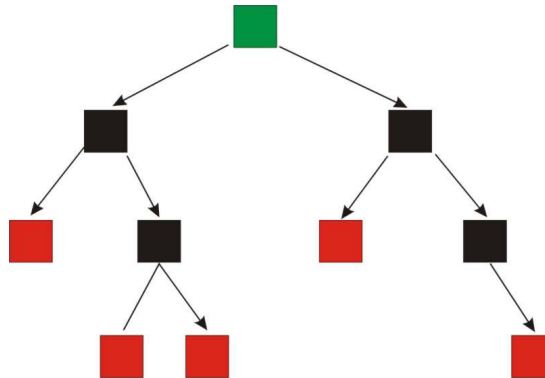


Figura 6: Ilustração de uma árvore de busca para o raciocínio inferencial.

### 3.2.1.2 Raciocínio Probabilístico (Redes Bayesianas)

O raciocínio probabilístico pode também ser chamado de Redes Bayesianas (Bayesian Network). As Redes Bayesianas são compostas por uma parte estrutural que reflete as relações causais entre as variáveis, e pelos valores de probabilidade que refletem a força da relação entre as variáveis.

De uma forma geral, a estrutura de uma rede bayesiana consiste de duas partes:

- *Parte Qualitativa:* Representa o modelo gráfico (grafo acíclico direcionado) onde as variáveis são os nós e as regras, que são as relações de dependência condicional entre essas variáveis. São os arcos direcionados;

<sup>11</sup> Árvores são mais simples para busca que grafos, porque quando um novo nodo é gerado, pode-se estar seguro que ele nunca foi visitado antes nem nunca será gerado depois.

- *Parte Quantitativa:* É o conjunto de probabilidades condicionais associadas aos arcos existentes no modelo gráfico e as probabilidades estimadas a priori das hipóteses diagnósticas.

A Figura 7 ilustra um típico exemplo de distribuição de probabilidade utilizando redes bayesianas. O exemplo trata de um alarme de residência que deverá disparar caso exista um arrombamento ou terremoto, acionando o telefone de João e Maria. Quem causa o disparo do alarme são os eventos de arrombamento ou terremoto; logo se estabeleceu uma relação de dependência para o acionamento do alarme. A definição dos nodos e das relações de causalidade representa a parte qualitativa da rede bayesiana (Figura 7).

Uma vez que a topologia da rede é definida, basta especificar as probabilidades condicionais para os nodos que possuem dependências diretas e usar essas para computar qualquer outro valor de probabilidade (Tabela 1).

Para o domínio de aplicações de Sistemas Especialistas Probabilísticos comumente existem várias hipóteses concorrentes, cada uma com um conjunto distinto de evidências; diz-se então que o sistema tem um conjunto multivalorado de hipóteses. Os Sistemas Especialistas Probabilísticos devem selecionar a hipótese (output) mais provável, a partir de um conjunto de evidências conhecidas (input).

A inferência é feita através da utilização da teoria da probabilidade, mais precisamente do Teorema de Bayes. Este teorema computa a probabilidade de um dado evento, dado um conjunto de observações. Sejam:

$P(H_i | E)$  a probabilidade que a hipótese  $H_i$  seja verdadeira dada a evidência  $E$ .

$P(H_i)$  a probabilidade “a priori” que a hipótese  $H_i$  é verdadeira na ausência de qualquer evidência.

O Teorema de Bayes é formulado como:

$$\mathcal{P}(H_i | E) = \frac{\mathcal{P}(E | H_i) \cdot \mathcal{P}(H_i)}{\sum_{j=1}^k \mathcal{P}(E | H_j) \cdot \mathcal{P}(H_j)}$$

Aplica-se este teorema ao exemplo relacionado à Figura 7 que ilustra o problema de, ao soar um alarme, quais são as chances de ser uma tentativa de roubo. Formula-se a proposição da seguinte maneira:

$$P(\text{alarme} | \text{roubo}) = 0.95$$

Um dos problemas deste método é a dificuldade das pessoas em estimarem probabilidade, sendo que a rede bayesiana necessita ter uma base de conhecimento completa,

isto é, todas as evidências relevantes às hipóteses consideradas devem estar explícitas na base de conhecimento.

		Probabilidade (soar alarme   arrombamento, terremoto)	
Arrombamento	Terremoto	Verdadeiro	Falso
Verdadeiro	Verdadeiro	0.950	0.050
Verdadeiro	Falso	0.940	0.060
Falso	Verdadeiro	0.290	0.710
Falso	Falso	0.001	0.999

Tabela 1: a Tabela de Probabilidade Condicional de um nodo.

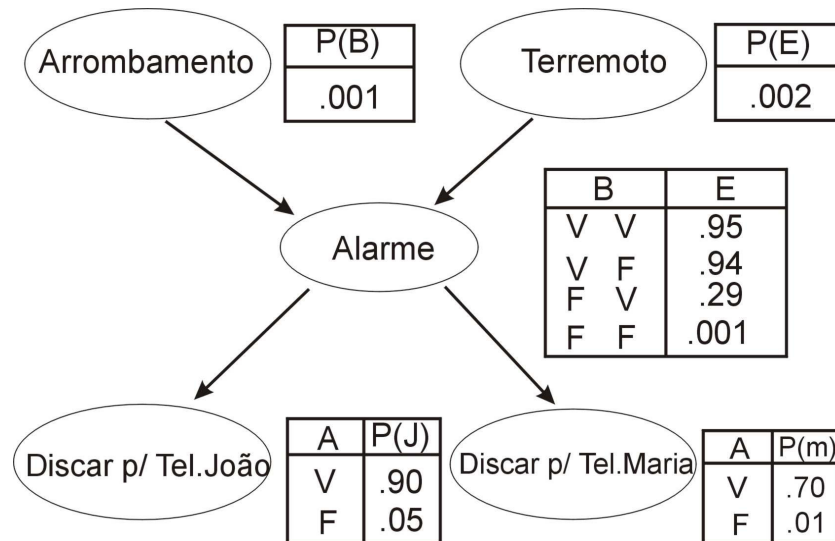


Figura 7: Ilustra uma rede Bayesiana.

### 3.2.1.3 Raciocínio Possibilista

#### 3.2.1.3.1 Conjuntos Nebulosos (“Fuzzy”)

A Teoria dos Conjuntos Nebulosos (“Fuzzy”) é o modelo mais antigo para o tratamento da informação imprecisa e vaga (BITTENCOURT, 1998). Este modelo foi introduzido em 1960 pelo Prof. Lotfi Zadeh, da Universidade da Califórnia, que expandiu a idéia dos conjuntos clássicos, onde um elemento pertencia ou não a um determinado conjunto. Zadeh propôs que em vez da pertinência de um elemento ser completamente falsa ou verdadeira correspondesse a um nível de pertinência ao conjunto. Surgiu assim uma nova ramificação de Lógica Multivalorada chamada de Lógica Nebulosa.

Dado um universo de discurso  $X$ , um subconjunto Nebuloso  $A$  de  $X$  é definido por uma função de pertinência que associa a cada elemento de  $x$  de  $X$  o grau  $\mu_A(x)$ , compreendido entre 0 e 1, com o qual  $x$  pertence a  $A$  (BITTENCOURT, 1998). A função de pertinência indica o grau de compatibilidade entre  $x$  e o conceito expresso por  $\mu_A(x)$ .

Suponha um Universo  $X$  do tipo:  $\{0;0.1;0.2;\dots;0.9;1.0\}$ ,  $x$  representa um elemento genérico de  $X$ . A função de pertinência  $\mu$  irá mostrar o grau com que cada elemento  $x$  pertence ao conjunto Nebuloso. A Figura 8 mostra que o elemento  $x=0.5$  tem grau  $\mu(x)$  de pertinência 1, enquanto que o elemento  $x=0.9$  tem grau 0.2. Veja-se mais sobre Lógica Nebulosa no apêndice A desta dissertação.

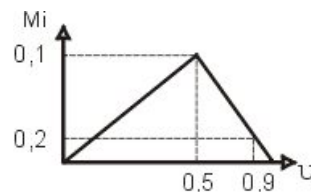


Figura 8: Função  $\mu(x)$  para graus de pertinência.

### 3.2.1.3.2 Distribuição de Possibilidade

Teoria de possibilidade enfoca principalmente informações imprecisas que estão intrínsecas nos idiomas naturais.

Na Teoria da Possibilidade tem-se um conjunto finito de variáveis possíveis do universo denominado  $\Omega$ , sendo que se parte do pressuposto que as variáveis contidas em  $\Omega$  são sempre um valor verdadeiro. Esta é uma razão por que a terminologia e o simbolismo de teoria de possibilidade diferem da Teoria Nebulosa.

Um distribuição de Possibilidade ( $\pi$ ) é mapeada para cada valor de  $x \in \Omega$ . Desta forma uma distribuição de possibilidades  $\pi: \Omega \rightarrow [0,1]$  para qualquer valor de  $x_i \in \Omega$ ,  $\pi(x_i)$  reflete até que ponto é possível que  $x = x_i$  (partindo do pressuposto que se tem somente um valor verdadeiro).

Uma distribuição de possibilidade  $\pi$  relativa à variável  $x$  pode, portanto, ser vista como uma função de pertinência de um conjunto Nebuloso dos valores possíveis de  $X$   $[0,1]$ . Por convenção,  $\pi$  representa algum conhecimento principal sobre o mundo real, isto é,  $\pi(x) = 0$  significa que  $x$  não é possível e  $\pi(x) = 1$  nada impede que  $x$  seja verdadeiro. Quando  $\pi(x) > \pi(x')$ ,  $x$  é um candidato preferencial a  $x'$  para ser um fato real. Em resumo,  $\pi$  seria como um codificador de relação de preferência.

Uma distribuição de possibilidade avalia por dois aspectos:

- Grau de Possibilidade  $\Pi(p)$  – quantifica o quanto a evidência  $p$  não contradiz a hipótese expressa por  $\pi$ . A medida de possibilidade é expressa da seguinte maneira:  $\Pi(p) = \sup_{x \in p} \pi(x)$ .
- Grau de Necessidade  $N(p)$  – quantifica o quanto a evidência disponível suporta a hipótese de que  $p$  contém o verdadeiro valor de  $x$ . A medida de necessidade é expressa da seguinte maneira:  $N(p) = \inf_{x \in A} 1 - \pi(x)$ .

A dualidade entre grau de certeza e possibilidade é expressa por  $N(p) = 1 - \Pi(\neg p)$ . Graus de possibilidade e certeza podem ser usados para induzir relação de ordem no conjunto de sentenças lingüísticas, considerando uma sentença como um conjunto de possíveis palavras verdadeiras.

Para melhor esclarecimento da Teoria da Possibilidade supõe-se que não se saiba a altura de Carlos, mas que se saiba que ele é alto (Figura 9). Então, a distribuição de possibilidade da altura de Carlos pode ser aproximada por aquela do conjunto Nebuloso  $A =$  “alto”, ou seja, faz-se  $\pi(x) = \mu_A(x)$  (BITTENCOURT, 1998).

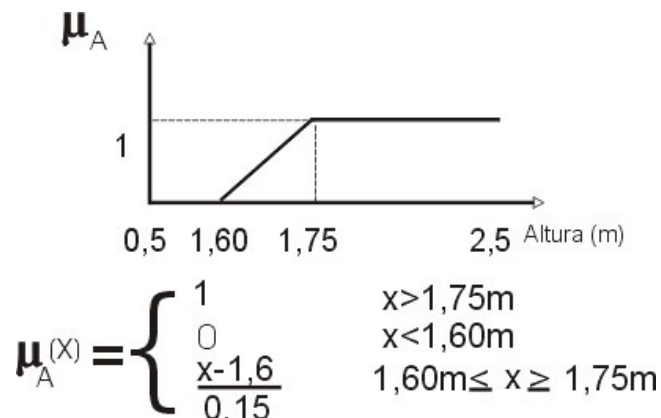


Figura 9: Conjunto nebuloso  $A =$  “Alto” extraído de (BITTENCOURT, 1998).

Intervalos(p)	Grau de Possibilidade $\Pi(p)$	Grau de Necessidade $N(p)$
(0.5-1.6)	$\Pi[(0.5-1.6)] = 0$	$N[(0.5-1.6)] = 0$
(1.60-1.75)	$\Pi[(1.6-1.75)] = 1$	$N[(1.6-1.75)] = 0$
(1.75-2.5)	$\Pi[(1.75-2.5)] = 1$	$N[(1.75-2.5)] = 0.34$
(1.80-2.5)	$\Pi[(1.80-2.5)] = 1$	$N[(1.80-2.5)] = 0$
(1.60-2.5)	$\Pi[(1.60-2.5)] = 1$	$N[(1.60-2.5)] = 1$

Tabela 2: Ilustra os graus de Possibilidade e necessidade para a suposição de  $p$  extraído de (BITTENCOURT, 1998).

Para calcular-se o grau de possibilidade, lembre a formula  $\Pi(p) = \sup x \in p \pi(x)$ , isto é, pega-se o valor superior do intervalo e calcula-se o  $\mu_A(x)$ .

O Grau de Necessidade  $N(p) = \inf x \in A \ 1 - \pi(x)$ , onde  $N(p)$  quantifica o quanto a evidência disponível suporta a hipótese de que  $p$  contém o verdadeiro valor de  $x$ , e  $\Pi(p)$  quantifica o quanto a evidência não contradiz esta hipótese. Para conferir os valores de  $\Pi(p)$  e  $N(p)$  tem-se a tabela 2 extraída de Bittencourt (1998).

#### 3.2.1.4 Raciocínio por Crença (Teoria de Dempster-Shafer – TDS)

O Raciocínio por Crença se baseia essencialmente nos trabalhos feitos originalmente por Dempster, que tentou modelar a incerteza por uma faixa de probabilidades mais do que um simples número probabilístico. Shafer estendeu e refinou o trabalho de Dempster num livro publicado em 1976 chamado “A Mathematical Theory of Evidence” (BUCHANAN, 1984). A teoria de Dempster-Shaper surgiu devido a problemas encontrados na modelagem da incerteza usando métodos puramente probabilísticos, e também pela falta de embasamento matemático do modelo Fator de Certeza do MYCIN. Uma vez que a TDS atribui valores de crença a subconjuntos e a cada elemento do conjunto de hipótese, esta teoria tem condições de refletir mais precisamente o processo de acúmulo de evidências. Além disso, a TDS permite que funções de crença possam ser combinadas, produzindo novas funções de crença num procedimento que independe da ordem na qual as evidências surgem, mas que exige que as hipóteses primitivas sendo consideradas sejam mutuamente exclusivas e exaustivas (UCHÔA, 1997).

A Teoria de Dempster-Shaper supõe que existe um conjunto fixo de elementos (também chamados de hipóteses primitivas) onde é permitido construir subconjuntos mais

elaborados, que são mutuamente exclusivos e exaustivos, chamados "meio" e simbolizados por  $\theta$ .

Suponha-se que um paciente apresente manchas vermelhas pelo corpo, e que isso seja sintoma de qualquer um dos seguintes problemas: alergia, intoxicação, sarampo, ou rubéola. É óbvio que existem muitos outros possíveis problemas, mas para efeito de simplicidade, supõe-se que exista somente estes quatro (UCHÔA, 1997). O conjunto meio -  $\theta$  para este problema ficaria

$$\theta = \{ \text{alergia, intoxicação, sarampo, rubéola} \}$$

A TDS assume para qualquer domínio do problema  $\theta$  que:

- $\theta$  é no sentido de completo, conter todas as possíveis hipóteses primitivas;
- as hipóteses primitivas em  $\theta$  são mutuamente exclusivas.

Suponhe-se agora que uma nova evidência seja levantada pelo médico:

- para um diagnóstico de reação orgânica definida no exemplo pelo conjunto {alergia, intoxicação};
- para um diagnóstico de infecção definida no exemplo como o conjunto {sarampo, rubéola}.

Se o médico, por exemplo, observar uma evidência que confirma com um determinado grau o diagnóstico de reação orgânica, ele irá atribuir uma quantidade de crença ao conjunto {alergia, intoxicação} proporcional ao grau observado de confirmação de evidência (UCHÔA, 1997).

Um subconjunto de hipóteses pode ser visto como uma nova hipótese formada pela disjunção de seus elementos, uma vez que, como se sabe, hipóteses primitivas são mutuamente exclusivas. O conjunto de hipóteses primitivas dá origem a  $2^{|\theta|}$  possíveis hipóteses, como mostra a Figura 10 para o exemplo anterior, onde  $\theta = \{ \text{alergia, intoxicação, sarampo, rubéola} \}$  (UCHÔA, 1997).

Para indicar a crença de uma hipótese, dada uma evidência, calcula-se uma atribuição de probabilidade básica ou, como é mais conhecida, função Massa, que nada mais é do que uma generalização da função de densidade de probabilidade, sendo que esta última associaria um valor numérico do intervalo  $[0,1]$  a todo subconjunto  $\theta$ , de maneira que a soma das evidências seja 1.



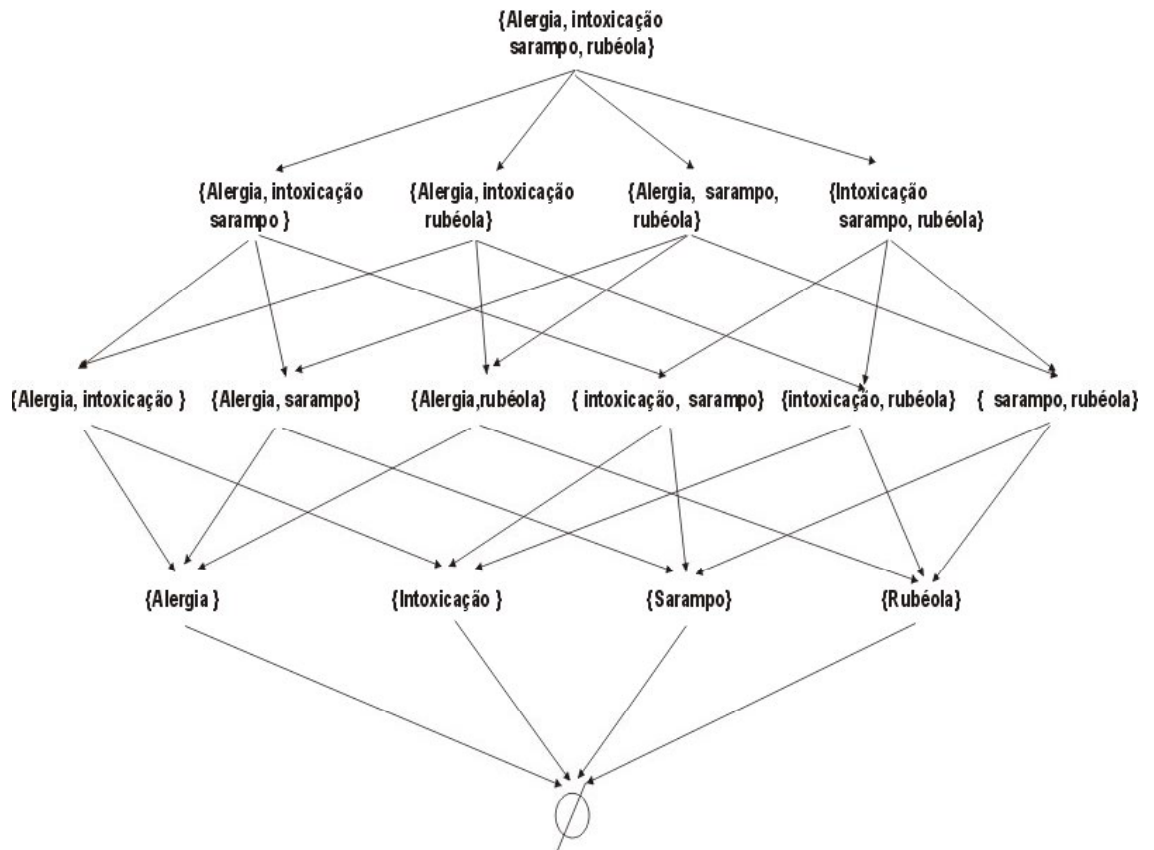


Figura 10: Conjunto de todas as hipóteses obtidas do conjunto de hipóteses primitivas  $\theta = \{ \text{Alergia, Intoxicação, Sarampo, Rubéola} \}$ , extraído de (UCHÔA, 1997).

### 3.2.1.5 Fatores de Certeza

O Fator de Certeza foi desenvolvido por Edward Shortliffe e Bruce Buchanan para ser usado pelo sistema especialista médico MYCIN. Um fator de certeza é um número real no intervalo  $[-1; 1]$ . O usuário do sistema especialista deve fornecer ao sistema fatores de certeza para cada proposição básica. O fator de certeza  $CF(A \rightarrow B)$  da regra de produção  $A \rightarrow B$  (se  $A$ , então  $B$ ) é fornecido pelo especialista. Uma simples pergunta ao especialista de qual certeza ele atribui a regras  $A \rightarrow B$ , ou equivalentemente, qual a certeza atribuída para  $B$  se  $A$  é verdadeiro.

A Figura 11 ilustra uma simplificação da rede de inferência utilizada no MYCIN. Para determinar o fator de certeza  $G$  da Figura 11 necessitam-se propagar os valores pelas regras. Em um modelo de Fator de Certeza a propagação das regras se faz da seguinte maneira:

1. Combinação serial:  $CF[B, \{A\}] = CF[A \rightarrow B] \text{Max}(0, CF[A])$
2. Conjunção:  $CF[A \wedge B] = \text{Min}(CF[A], CF[B])$
3. Disjunção:  $CF[A \vee B] = \text{Max}(CF[A], CF[B])$



idéia é de que se descrevam e se acumulem descrições de ‘casos’<sup>12</sup>, ou instâncias de problemas na área do conhecimento especializado, e se tente descobrir, por analogia, quando um determinado problema é similar a um outro já resolvido. Desta forma, a solução já aplicada ao problema pode ser utilizada novamente ou adaptada para o caso em questão.

A utilização da abordagem de casos para construção de sistemas baseados em conhecimentos tem algumas desvantagens, como:

- dificilmente os casos estão disponíveis de forma confiável, em quantidade suficiente e com boa representatividade sobre o domínio, inviabilizando a aplicação dos algoritmos de recuperação como forma de solução de problemas;
- os algoritmos de recuperação por similaridade não mostram bom desempenho em aplicações reais, onde as comparações feitas por um especialista tendem a ser muito mais sofisticadas e de difícil compreensão do que a implementada em sistemas de casos;
- não existem bons algoritmos de adaptação que permitam ao sistema fornecer soluções adequadas ao problema apresentado pelo usuário, se esse não for muito parecido com algum previamente descrito.

### **3.2.2 Abordagem Conexionista**

A Inteligência Artificial Conexionista - IAC inspira-se no sistema nervoso humano para imitar um comportamento inteligente. Através de modelos baseados na estrutura e funcionamento do cérebro humano, surgiram os neurônios artificiais e as redes neurais.

#### **3.2.2.1 Redes Neurais**

As redes neurais artificiais são inspiradas no Sistema Nervoso Natural. O cérebro é responsável por uma grande quantidade de funções executadas pelo corpo. Inspirando-se no neurônio natural, McCulloch e Pitts propuseram o primeiro modelo de neurônio artificial em 1943.

Os neurônios artificiais são unidades de processamento que recebem estímulos (valores de entrada excitatórios ou inibitórios) e geram saídas. A Figura 12(a) ilustra um neurônio natural com a sua árvore dendrital que recebe informações do corpo, os axônios que

---

<sup>12</sup> Um caso é uma descrição completa de um problema do domínio com a respectiva solução aplicada, mais uma avaliação da eficácia dessa solução. É descrito de forma similar a uma tabela, ou seja, através da enumeração e qualificação de atributos e valores. O conjunto dos casos é indexado por seus atributos de forma a se agilizar a busca e recuperação de casos similares.

se conectam a outro neurônio vizinho, formando a rede neural biológica, e o soma que tem a função de metabolizar os processos da célula nervosa. A Figura 12(b) representa um neurônio artificial com entradas, imitações de axônios naturais, um valor de saída, imitando o funcionamento dos dendritos, formando uma rede neural artificial. Uma rede neural artificial é uma estrutura computacional que consiste de neurônios artificiais conectados entre si através de conexões de pesos. O conhecimento das redes neurais artificiais está representado pelos pesos que ligam os neurônios uns aos outros.

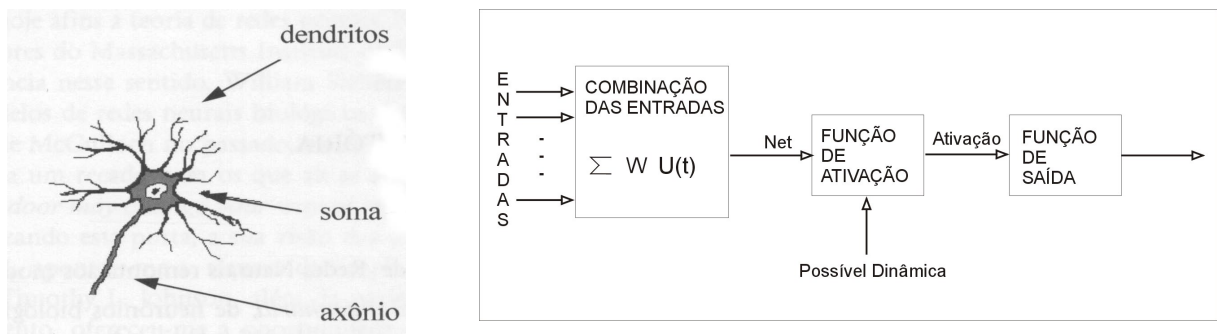


Figura 12: (a) representação do neurônio biológico (KOVÁCS,1996); (b) representação do neurônio artificial (KOVÁCS, 1996)

Um neurônio artificial, segundo apresentado na Figura 12(b), possui alguns aspectos, tais como:

- Combinação das entradas: onde as informações que o neurônio recebe podem ser externas à rede, podem ser saídas de outros neurônios ou ainda polarização (“bias”). A Combinação das entradas (“net”) é a soma da multiplicação das conexões entre os neurônios pelos valores de entrada do neurônio.
- Função de Ativação: produz um novo estado de ativação do neurônio, através de uma função de saída.
- Função de Saída: as mais comuns funções de saída usadas em neurônios artificiais são a função linear, a função logística e a função tangente hiperbólica (AZEVEDO, 2000).

O primeiro passo para se definir uma rede neural artificial é definir sua topologia, isto é, a quantidade de neurônios que a rede vai possuir e suas interligações. Normalmente, as redes neurais são organizadas em camadas da seguinte forma:

- Camada de Entrada, onde as informações entram na rede;
- Camada Intermediária, onde é feito o processamento propriamente dito;

- Camada de Saída, onde o resultado final é apresentado.

A Figura 13 ilustra uma rede neural com uma camada de entrada com dois neurônios, duas camadas intermediárias com quatro neurônios e uma camada de saída contendo um neurônio.

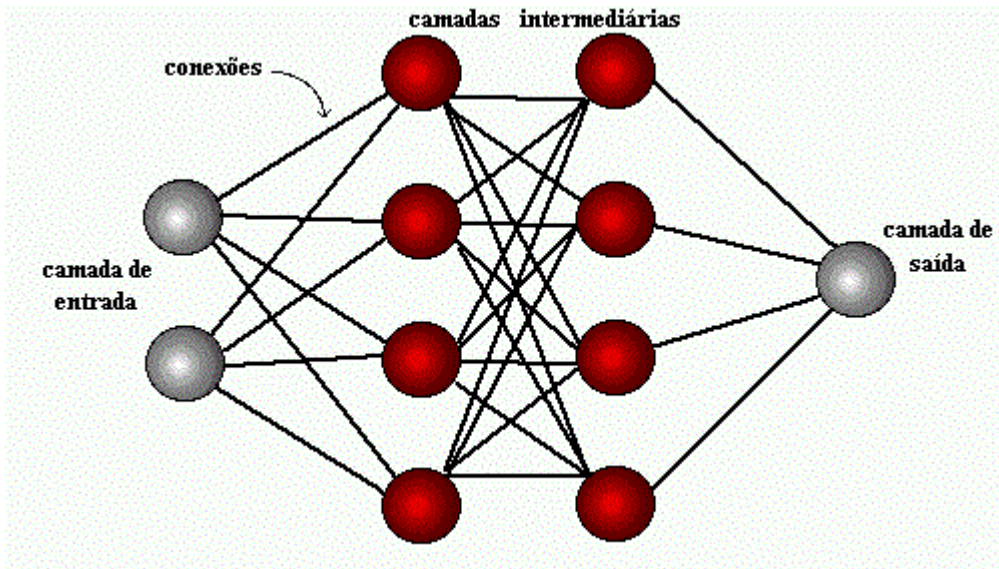


Figura 13: Representação da topologia de uma rede neural artificial.

A capacidade de aprender das redes neurais é obtida através de um processo iterativo de ajustes de pesos chamado treinamento. O treinamento pode ser:

- Supervisionado, onde são apresentados à rede valores de entrada e quais os respectivos valores de saída apropriados àquelas entradas;
- Não supervisionado – consiste somente nos valores de entrada. A rede classifica as entradas conforme padrões e não se pode afirmar como a rede determinou tais padrões.

A grande vantagem das redes neurais é a sua capacidade de aprender através de exemplos e de generalizar este aprendizado de forma que reconheça outras entradas como sendo similar aos exemplos apresentados. Assim, as redes neurais tratam de problemas mal definidos, onde faltam informações sobre o problema.

Em contrapartida, é difícil definir-se uma topologia ideal para uma rede neural, pelo fato de não existir uma formalização para a construção da topologia das redes neurais. É crítica a etapa de escolha dos conjuntos de exemplos, sendo que muitas vezes não se possui a quantidade de exemplos suficientes para a etapa de treinamento. E, além disso, muitas vezes a etapa de treinamento da rede tende a ser lenta.

## **4 O Projeto RCSUEX**

### **4.1 Introdução**

Este trabalho surgiu como fruto de um projeto denominado “RCSUEX – Representação da Certeza de Sucesso Exploratório”, desenvolvido pelo Laboratório de Conexão e Ciências Cognitivas (L3C) da Universidade Federal de Santa Catarina – UFSC para a Petrobras, a partir da necessidade desta última de estruturar a representação do conhecimento incerto no Processo de Exploração Petrolífera. O objetivo do Projeto RCSUEX é desenvolver um Sistema Especialista que estime os riscos da exploração petrolífera, sobre dados geológicos e dados estatísticos, para um novo poço, com o propósito de sistematizar a argumentação e medir a probabilidade de sucesso da descoberta de uma acumulação de petróleo. A partir deste capítulo usa-se a expressão “Projeto RCSUEX” quando se refere ao estudo e planejamento das etapas necessárias à implementação do “Sistema RCSUEX”, o Sistema Especialista (software) que é o produto final desta parceria.

A primeira etapa do projeto RCSUEX foi de entender o processo exploratório. A exploração petrolífera inicia-se pelas avaliações de modelos geológicos, isto é, de variáveis críticas à dinâmica das acumulações petrolíferas. Este modelo geológico é subdividido em fatores geológicos (geração, migração, reservatório, selo, trapa e sincronismo) que foram definidos no capítulo 2 desta dissertação.

As informações sobre as variáveis de cada fator geológico continuam sendo obtidas através de análise sísmica, analogia a outros poços da área e teorias geológicas. Deve-se notar que a abrangência deste projeto não se refere às ferramentas de coleta de dados. O Sistema RCSUEX é utilizado depois de coletados os dados, isto é, em nível de avaliação de dados coletados para o novo prospecto. A Figura 14 representa as atividades, desde a coleta de informações através de dados sísmicos até a obtenção do risco do prospecto, realizadas para se calcular um valor probabilístico, que indica a favorabilidade do prospecto. O projeto RCSUEX atua no nível quatro (4), da Figura 14, avaliando os fatores geológicos.

O intérprete de posse dos valores de cada variável faz analogias com suas experiências anteriores, tentando estimar uma probabilidade para cada fator. Esta etapa de análise das informações refere-se a dados subjetivos, porque o especialista transforma as informações sobre o prospecto, usando sua experiência e sentimento, em uma probabilidade de sucesso. Segundo (BUENO, 1982) a palavra subjetivo é um substantivo masculino que se

refere a “aquilo que é sentido por experiência íntima”. Neste sentido, os dados que o intérprete possui para determinar a existência ou não de um novo prospecto são valores qualitativos, com informações vagas, incertas e imprecisas.

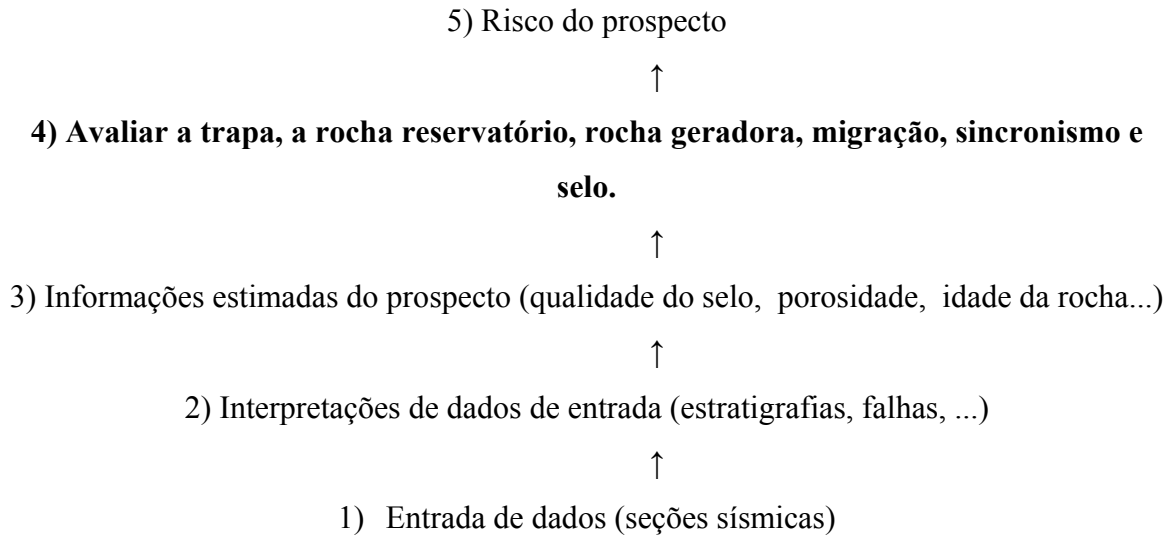


Figura 14: Tarefas de uma avaliação de prospecto, extraída de Clark (1989).

Muitas vezes a estimativa de probabilidade de sucesso, dos dados subjetivos, vêm carregadas de otimismo ou desânimo devido ao tratamento dos dados serem feitos através de sentimentos e experiências do especialista. Desta forma, a indústria do petróleo não se sente confortável apenas com uma avaliação subjetiva.

Tem sido comum na indústria do petróleo utilizarem-se as frequências observadas (taxas ou índices) de sucesso em históricos de poços pioneiros como boas estimativas das probabilidades de sucesso de prospectos exploratórios (SILVA, 2000).

Os dados históricos (taxas ou índices) denominam-se dados objetivos, porque são frequências observadas. Afirma-se que a representação puramente frequencista (objetiva), assim como a puramente subjetiva (argumentativa), não deve ser tratada separadamente, mas propõe-se a união através de um modelo matemático que combina as perspectivas objetiva e subjetiva de forma harmônica (SILVA, 2000).

Os dados históricos usados como parâmetros neste projeto são representações estatística clássicas (frequencistas) e se referem a:

- número total de prospectos perfurados no local;
- número de descobertas com sucesso no local.

Além disso, faz-se importante a incorporação da tecnologia de DHI sobre a favorabilidade do prospecto como um todo, por ser uma evidência muito forte da existência ou não do prospecto. A Figura 15 ilustra o processo que deve ser respeitado para o desenvolvimento do Sistema Especialista que auxiliará o intérprete a estimar o risco exploratório de um prospecto.

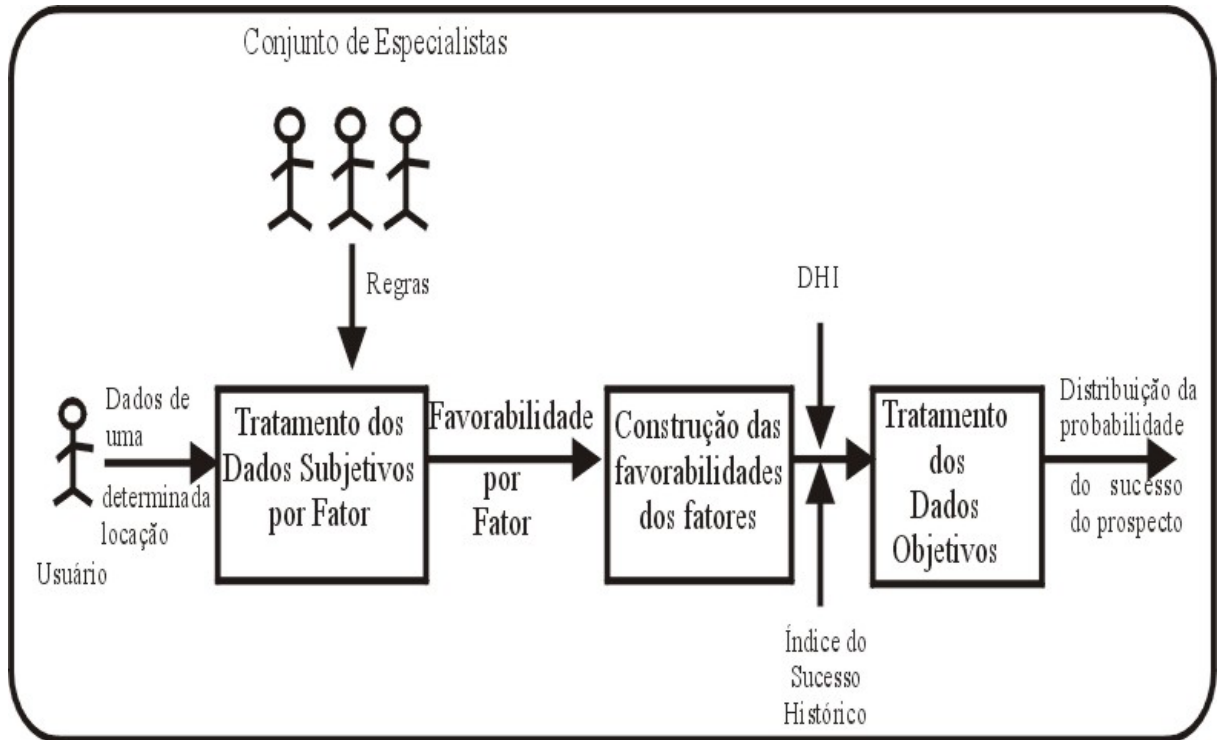


Figura 15: Ilustração da representação do processo de definição da probabilidade de um prospecto.

Desta forma, o projeto RCSUEX possui duas metodologias bem distintas:

- Dados Subjetivos (qualitativos, com informações vagas, incertas e imprecisas) e
- Dados Objetivos (estatísticos).

Neste capítulo define-se cada um dos tipos de dados, subjetivos e objetivos, e a metodologia escolhida para tratar cada um deles. Inicia-se o capítulo apresentando outros trabalhos na área de exploração petrolífera.

## 4.2 O que já existe na Indústria do Petróleo

Inicialmente, pesquisou-se em busca de projetos na área da avaliação de prospecção a fim de comparar com a proposta deste trabalho. Decidiu-se por pesquisar o que a indústria do Petróleo esta fazendo para buscar soluções à área de exploração. Entre outros trabalhos destacam-se a Metodologia Chevron e o Projeto Fuzzy Expert Exploration Tool (FEE TOOL).



#### 4.2.1 A Metodologia Chevron

A partir de 1989, a Chevron Overseas Petroleum desenvolveu um método analítico que permite administrar/gerenciar o processo de exploração petrolífera. O método envolve uma sistematização dos dados de exploração, incorporando tecnologias de gerenciamento (OTIS,1987). Este método é representado por uma planilha eletrônica que, ao ser preenchida, permite obter-se a avaliação do risco do prospecto (risco geral) e por fator geológico. O método Chevron analisa o risco geológico sobre quatro fatores: geração, reservatório, trapa/selo e sincronismo/migração.

O método Chevron é qualitativo, isto é, o geólogo ou intérprete avalia o prospecto de forma puramente lingüística, utilizando termos como “Favorável”, “Encorajador”, “Neutro”, “Questionável” ou ainda “Desfavorável”. Para cada um dos quatro fatores geológicos existem atributos que são avaliados segundo os termos lingüísticos citados acima. Por trás destes termos lingüísticos existem números probabilísticos que geram uma probabilidade de sucesso para cada fator para posterior multiplicação que resulta na probabilidade de sucesso geológico do prospecto. O tratamento puramente qualitativo dado ao método é similar à linguagem usual de avaliação de um prospecto utilizada comumente na indústria petrolífera. Apesar de o método Chevron ter como entradas variáveis lingüísticas, ele se utiliza de combinações aritméticas para gerar uma saída que não permite a investigação de qual raciocínio foi utilizado para chegar a uma conclusão. Logo, este modelo não dispõe de um histórico (explicação) de como chegou à conclusão sobre a avaliação do prospecto.

O fato de as variáveis de entrada serem unicamente qualitativas e bivalentes, no sentido de possuírem apenas um valor “verdadeiro” para apenas uma classe e “falso” para todas as outras classes, provoca uma perda de informação bastante significativa. Tome-se como exemplo o caso em que um dado prospecto, onde o atributo “Permeabilidade da Rocha Reservatório” é “Encorajador” pendendo ao “Neutro”, recebe o mesmo tratamento que outro prospecto onde a “Permeabilidade da Rocha Reservatório” é “Encorajador” pendendo ao “Favorável”. Ao serem tratados os atributos somente como valores lingüísticos perdem-se informações e, com isso, pode-se incorrer no erro de tratar da mesma forma prospectos diferentes.

O método Chevron não incorpora dados históricos, como número de poços pioneiros e números de descobertas, e não diferencia poços com ou sem aplicação de DHI (“Direct Hydrocarbon Indicators”). O peso da aplicação de DHI e sua intensidade são evidências importantes a serem ponderadas.

O método Chevron mostra-se impreciso na avaliação de prospectos porque trabalha de forma qualitativa sobre dados puramente estatísticos. Dois prospectos diferentes podem assumir o mesmo valor de sucesso, mas terem particularidades bem distintas que poderiam ser ressaltadas caso a abordagem não fosse puramente qualitativa.

Uma abordagem interessante dada pelo artigo escrito por Otis e Schneidermann (OTIS, 1987), onde se apresenta o método Chevron, é a maneira como eles calibraram os atributos de cada fator ( i.e., agrupam as variáveis conforme suas características). As variáveis de cada fator geológico foram classificadas segundo três características:

- Dados Análogos: são dados que foram extraídos através de analogias com outros poços da mesma área. Não se possui o dado exato referente ao prospecto que está sendo avaliado.
- Dados Indiretos: não se pode enxergar se o petróleo está ou não no local estudado. Portanto, usa-se de métodos, tais como mapas sísmicos e interpretações visuais para a obtenção de informações que, em suma, são variáveis imperfeitas. Exemplos de variáveis indiretas: evidências de falhas recentes.
- Dados Diretos: são dados que determinam claramente a presença de hidrocarboneto ou que podem claramente excluir a presença de hidrocarboneto do local. Exemplos de dados diretos: proximidade de acumulações, seeps entre outros indícios, que geram um risco baixo de insucesso, ou seja, que a favorabilidade do fator seja alta para a presença de hidrocarboneto caso estas variáveis sejam positivas ao prospecto.

Quando o atributo é uma indicação de analogia com outro poço, este atributo recebe um peso bem pequeno. Quando o atributo se refere a um dado indireto, seu peso aumenta um pouco. Já o atributo que se refere a um dado direto recebe um peso próximo a 1 (um). Isso faz que dados análogos tenham um peso menor do que um dado indireto e, por sua vez, um dado indireto tem peso menor em comparação a um dado direto. Desta forma, o sistema sabe que um dado direto representa mais ao sistema do que um dado análogo.

#### **4.2.2 O Projeto Fuzzy Expert Exploration Tool**

O Projeto Fuzzy Expert Exploration Tool (FEE TOOL) (FUZZY EXPERT, 2002) da Petroleum Recovery Research Center (PRRC), na New Mexico Tech, é um projeto financiado pelo Departamento de Energia dos Estados Unidos, Ministério Nacional de Tecnologia do Petróleo, Estado do Novo México e várias companhias independentes de petróleo. Este projeto se propõe a desenvolver uma abordagem de Inteligência Artificial para incorporar

informações com estimacões de risco modeladas de maneira conveniente e realista. O desenvolvimento do Fuzzy Expert Exploration Tool (FEE Tool) trata as informacões sobre o processo explorat3rio com L3gica Nebulosa, mas disp3e de Redes Neurais para tratamento de dados s3smicos. O Projeto Fuzzy Expert Exploration Tool encontra-se em desenvolvimento.

### **4.3 Estudos das alternativas para tratamento de dados subjetivos**

A partir do estudo do dom3nio do problema, a explorac3o petrol3fera, fez-se uma pesquisa sobre as metodologias que mais se adequavam 3s imperfeic3es dos dados estudados. No cap3tulo 3, foram descritas algumas metodologias que tratam imperfeic3es da informac3o atrav3s de uma abordagem de Intelig3ncia Artificial. Nem todas as metodologias apresentadas no cap3tulo 3 foram testadas, ao n3vel de prot3tipo, porque n3o tratam de imperfeic3es apresentadas no dom3nio do problema de explorac3o petrol3fera, ou a soluic3o apresentada pela metodologia n3o 3 adequada 3 resposta desejada (uma probabilidade).

Durante os estudos do Sistema Petrol3fero, foram analisadas as informac3es, buscando identificar o tipo de imperfeic3o dos dados. Esta an3lise mostrou a incerteza como tipo de imperfeic3o mais presente no Sistema Petrol3fero. As informac3es incertas est3o presentes nas vari3veis que n3o possuem m3tricas, por exemplo, tipo da trapa. Mesmo que o ge3logo acredite ter um tipo de trapa no local do prospecto, ele n3o descarta as demais alternativas. Por esse motivo, pesquisaram-se metodologias espec3ficas de Intelig3ncia Artificial capazes de manipular informac3es imperfeitas do tipo incertas.

Duas metodologias se sobressa3ram 3s demais devido ao fato de tratarem informac3es incertas: L3gica Nebulosa e Redes Bayesianas. Ainda assim foram feitos testes com Redes Neurais para se obter sustentac3o na escolha da metodologia

#### **4.3.1 Redes Neurais**

O primeiro prot3tipo, que testou as Redes Neurais, foi desenvolvido na ferramenta MatLab com o “toolbox” de Redes Neurais. Inicialmente, por quest3o de facilidade, tratou-se em separado cada fator geol3gico, isto 3, pensou-se que existiria uma Rede Neural para cada fator geol3gico (gerac3o, reservat3rio, selo, trapa, migrac3o e sincronismo).

Para a construc3o do prot3tipo de Redes Neurais, precisaram-se definir algumas vari3veis que caracterizassem especificamente cada um dos fatores geol3gicos. Com o estudo do dom3nio do problema petrol3fero apresentado no cap3tulo 3, foi poss3vel distinc3o de algumas vari3veis importantes para cada fator, mas n3o se tinha a experi3ncia suficiente para definir o que resultaria da combinaic3o destas vari3veis. Esta foi sem d3vida a primeira grande

dificuldade encontrada na etapa de escolha da metodologia de tratamento dos dados subjetivos: não havia exemplos de dados para os testes.

Desta forma, optou-se por avaliar as Redes Neurais através do exemplo de benchmark “Visitou a Ásia”, disponível através de redes bayesianas pela ferramenta Netica, que simula um diagnóstico médico para descobrir se o paciente está com bronquite e/ou tuberculose. Construiu-se o conjunto de exemplos para as variáveis, utilizando-se as tabelas de probabilidades do exemplo apresentado pela ferramenta Netica para as variáveis: “Visitou a Ásia”, “fuma”, “Resultado do Raios X”, “Dispneia”, “bronquite”, “tuberculose” ou “câncer”, sendo estas duas últimas as variáveis de saída da rede. Tanto as variáveis de entrada como as de saída tinham valores de zero(0) ou um (1), sendo 0 para representar verdade e 1, falsidade. Criou-se uma rede neural direta, com algoritmo de aprendizado backpropagation, para o protótipo, utilizando o “toolbox” de Redes Neurais do MatLab.

O sentimento inicial de que as redes neurais apresentariam má generalização foi comprovado no protótipo. A rede neural não conseguiu generalizar o seu aprendizado para outros casos não vistos durante o treinamento, fornecendo saídas bem diferentes das esperadas em casos não apresentados na etapa de treinamento. Isso se justifica na medida em que as redes neurais não são uma boa metodologia para tratar dados imprecisos, vagos e incertos.

Outra dificuldade imposta pelas Redes Neurais é a obtenção de exemplos para a fase de treinamento. A indústria do petróleo não dispõe dos valores de cada variável que define um fator geológico. Mesmo que o geólogo possua uma lista de variáveis (“check-list”) para um fator, nem sempre, para todos os prospectos, ele realmente irá avaliar ou estudar todas as variáveis. Muitas vezes, para alguns prospectos o geólogo não possui, por exemplo, informações laboratoriais, mas mesmo assim, com dados incompletos, o geólogo consegue definir uma probabilidade de sucesso para o prospecto. Além disso, a indústria do petróleo não dispõe de um banco de dados com um histórico de prospectos perfurados, o que dificulta, e muito, a construção de exemplos necessários ao treinamento das redes neurais.

Portanto, pela dificuldade de definir exemplos para a fase de treinamento da rede, pelo fato de que as redes neurais não possuem, por si só, uma explicação ou argumentação de como chegaram à solução do problema, descarta-se a metodologia. Sabe-se que existem abordagens híbridas que permitiriam às redes neurais apresentarem argumentações que justifiquem sua saída, mas pela restrição dos exemplos e pela restrição de tempo que o projeto impõe, optou-se pelo estudo de outras metodologias.

### 4.3.2 Redes Bayesianas

A segunda metodologia avaliada foi a Rede Bayesiana, através de um protótipo desenvolvido em Netica. Novamente faz-se importante ressaltar que para cada fator geológico (geração, reservatório, selo, trapa, migração e sincronismo) construiu-se uma rede bayesiana. Para cada fator foram definidas em média quatro variáveis, sendo que a variável de saída teria apenas as hipóteses de “existir” e “não existir” hidrocarboneto.

Sabe-se que as redes bayesianas trabalham com relações causais quantificadas por valores de probabilidade condicional. Desta forma, estudou-se a relação de causalidade entre as variáveis definidas para cada fator geológico e preencheram-se as tabelas de probabilidade de forma subjetiva.

Outra característica importante das Redes Bayesianas é que a sua base de conhecimento deve ser completa, isto é, todas as evidências relevantes às hipóteses devem ser explicitadas na base de conhecimento. Por isso, repetiu-se aqui a necessidade de haver um banco de dados de variáveis para que se pudesse fazer um estudo freqüencista das variáveis para o preenchimento das tabelas condicionais. Contudo, ao apresentar-se este modelo à Petrobras, confirmou-se a primeira grande desvantagem da metodologia: o preenchimento das tabelas condicionais.

A indústria do petróleo, assim como a Petrobras, não dispõe de uma base de dados de atributos do sistema petrolífero. Embora os geólogos e intérpretes possuam uma lista de variáveis (“check-list”) na qual se apóiam para justificar sua defesa a um prospecto, esta é feita de forma descritiva – os geólogos não pensam em números probabilísticos. Desta maneira, as tabelas condicionais das Redes Bayesianas seriam preenchidas de forma subjetiva, isto é, as probabilidades seriam uma mera aproximação da realidade.

O protótipo desenvolvido pelo Projeto RCSUEX possuía os seis fatores geológicos com aproximadamente quatro atributos cada fator, o que resultaria, aproximadamente, em uma tabela de 96 probabilidades condicionais. Note-se que ainda se está em nível de protótipo, pois com o sistema real esta tabela condicional seria bem maior. O crescimento da tabela de probabilidades condicionais é uma função exponencial do número de variáveis do problema, o que dificultaria muito à Petrobras definir probabilidade em nível de variáveis. Outra desvantagem desta metodologia seria o fato de que as tabelas de probabilidade condicionais seriam diferentes para cada bacia ou área estudada, pois se, por exemplo, na bacia hipotética “XYZ” o tipo mais comum de trapa em poços com acumulação de hidrocarboneto for a estratigráfica, o mesmo não necessariamente ocorrerá na bacia hipotética “ABC”, e que significa dizer que seria necessária a criação de uma rede bayesiana para cada

tipo de bacia ou área, fazendo-se necessárias mais tabelas de probabilidades condicionais. Desta forma, inviabilizou-se a metodologia.

### **4.3.3 Lógica Nebulosa**

O terceiro protótipo utilizou a Lógica Nebulosa para solucionar o problema da exploração petrolífera. Este protótipo foi desenvolvido em Matlab e seu “Toolbox” de Lógica Nebulosa.

Para cada fator geológico construiu-se uma máquina de inferência Nebulosa. Desta forma, cada fator possui a sua variável de saída, que representa a probabilidade de sucesso do fator. Este protótipo utilizou as mesmas variáveis do modelo das Redes Bayesianas, em média quatro variáveis por fator. Construíram-se algumas regras para cada fator, não sendo necessária a construção completa da base de conhecimento como nas Redes Bayesianas.

As saídas obtidas no protótipo foram satisfatórias para os atributos ou variáveis apresentadas. A grande vantagem da metodologia é sua linguagem natural para tratar os problemas. Esta metodologia tem um raciocínio (lingüístico) semelhante à linguagem do geólogo ao justificar suas defesas e não exige probabilidades subjetivas. Desta forma, apresenta uma característica importante: ser mais estável se comparada às Redes Bayesianas. A estabilidade está no fato de a máquina de inferência tratar de regras e não probabilidades, que deveriam ser ajustadas de tempos em tempos. A Lógica Nebulosa é tolerante a informações vagas, incertas e incompletas, o que caracteriza muito bem o domínio do problema de exploração petrolífera.

Esta metodologia já havia sido sugerida por (SILVA, 2000) em sua tese de doutorado, que serviu de base ao Projeto RCSUEX, como uma possível solução ao problema da exploração petrolífera. A Lógica Nebulosa possui um histórico de como chegou à solução, o que auxilia o geólogo a fazer a sua justificativa sobre a probabilidade do sucesso do novo prospecto, e facilita a comparação entre prospectos que estão aguardando para serem perfurados.

Outra vantagem da metodologia é a simplicidade dos conceitos matemáticos que envolvem a Lógica Nebulosa, facilitando a implementação da sua máquina de inferência. Desta maneira, o Projeto RCSUEX optou por trabalhar na primeira etapa (dados subjetivos) do Sistema com a metodologia Nebulosa.

#### 4.4 Compatibilidade dos Dados

O fato de a Lógica Nebulosa tratar expressões lingüísticas (qualitativas) para a avaliação de um prospecto, como, por exemplo, “favorável”, “regular” e “desfavorável”, é um fator muito importante para o interprete, por ser mais natural ao seu cotidiano. Contudo, após as avaliações dos intérpretes uma proposta de alocação precisa passar por avaliações econômicas. Assim, o prospecto deve ser expresso por um valor de probabilidade, chamado probabilidade de sucesso a priori (dados geológicos - subjetivos), obtidos pela combinação de probabilidades de cada fator geológico. Ao se incorporarem dados objetivos (dados históricos) à probabilidade de sucesso a priori, passa-se a chamar de probabilidade de sucesso a posteriori, e esta nova probabilidade é que serve de parâmetro para os cálculos econômicos.

Portanto, é necessário que se façam ajustes à saída Nebulosa para que se possa conectar a uma probabilidade. Existem algumas metodologias apresentadas em (KLIR, 1988) que conjugam Conjuntos Nebulosos com Probabilidade. Contudo, tais abordagens não se mostraram válidas para o presente problema. Assim, propõem-se algumas alternativas de conexão entre Conjuntos Nebulosos e Probabilidade, no capítulo 5 desta dissertação.

A metodologia que trata os dados subjetivos, a Lógica Nebulosa, avalia cada fator geológico (geração, migração, reservatório, trapa, selo e sincronismo) quanto à sua “favorabilidade”. A variável de saída de cada fator geológico corresponde a essa “favorabilidade” e pode assumir as seguintes expressões lingüísticas (funções de pertinências): Desfavorável, Questionável, Neutro, Encorajador e Favorável. A variável de saída nebulosa de cada fator geológico é um valor entre 0-100. Contudo, quando se aplica a última etapa do Lógica Nebulosa, a desnebulização, é gerado um número nítido que representa uma probabilidade média (entre 0-100) para o fator geológico.

Para os geólogos, a existência de um consenso sobre um único valor probabilístico para representar a “favorabilidade” do fator não seria fácil. Isto porque se trata de avaliações sobre dados subjetivos e que a transformação para um número probabilístico único seria motivo de discordância, enquanto que um intervalo probabilístico é de mais fácil aceitação.

Assim, tem-se para cada fator geológico (geração, reservatório, migração, trapa, selo e sincronismo) um intervalo de valores probabilísticos [a,b] que representa uma distribuição de probabilidade que descreve os valores prováveis que aquele fator pode ter. Explica-se detalhadamente a obtenção do intervalo [a,b] no capítulo 5 desta dissertação. Por ora basta saber-se que este intervalo [a,b] representa os possíveis valores de favorabilidade para o fator geológico.

#### 4.5 Metodologias para tratamento de dados Objetivos

A partir da obtenção do intervalo [a,b], para cada fator geológico, conjuga-se a Função de Densidade Beta, para estimar uma distribuição a priori. A distribuição Beta a priori representa o conhecimento subjetivo (geológico e nebuloso) sobre a certeza da descoberta, de um poço pioneiro, a ser perfurado em condições geológicas análogas aos poços já perfurados. A função de Densidade Beta transforma as “favorabilidades” dos fatores geológicos em um valor de probabilidade a priori, isto é, sem incorporar dados objetivos. Para incorporar os dados objetivos deve-se conjugar novamente a Função de Densidade Beta (pdf), levando-se em consideração dados históricos da área. A este novo ajuste se chamou de Distribuição Beta a posteriori, que representa o conhecimento subjetivo combinado com o conhecimento objetivo (SILVA, 2000).

Em termos matemáticos, o modelo proposto é válido porque o intervalo de “favorabilidade” [a,b] é uma distribuição uniforme e a distribuição uniforme no intervalo [0,1] é um caso particular de Distribuição Beta (ROHATGI, 1976). Note-se que o intervalo de “favorabilidade” [a,b] para cada fator geológico está entre os valores 0-100.

##### 4.5.1 Função de Distribuição Beta

A função de Distribuição Beta é um tipo de função de densidade que estima distribuições de probabilidade a priori e a posteriori sobre os parâmetros alfa ( $\alpha$ ) e beta ( $\beta$ ). Formalmente defini-se a função Beta por:

Para  $\alpha > 0$ ,  $\beta > 0$ ,  $Beta(\alpha, \beta)$  é definida por:

$$(1) \quad Beta(\alpha, \beta) = \int_{0+}^{1-} x^{\alpha-1} (1-x)^{\beta-1} dx$$

Esta resulta na função de densidade de probabilidade (pdf)

$$(2) \quad f(x) = \frac{x(1-x)}{B(\alpha, \beta)} \quad 0 < x < 1$$

0, caso contrário.

Uma propriedade muito importante para a aplicação da Função de Distribuição Beta no presente trabalho, refere-se à Distribuição Uniforme em [0,1] ser um caso particular da Distribuição Beta. Note-se que no caso de  $\alpha = \beta = 1$ , nas equações (1) e (2) tem-se uma distribuição uniforme em [0,1].



A partir desta propriedade aplica-se o teorema 17 citado em (ROHATGI, 1976 : 216) que diz:

Teorema 17: Seja  $X_1, X_2, \dots, X_k$  são valores de beta independentes com  $X_i \sim B(\alpha_i, \beta_i)$ ,  $i = 1, 2, \dots, k$ . Seja  $\alpha_{i+1} + \beta_{i+1}$ ,  $i = 1, 2, \dots, k-1$ . Então  $\prod_{i=1}^k X_i$  é também uma beta com parâmetros  $(\alpha_k, \beta_1 + \beta_2 + \dots + \beta_k)$ .

Seja  $X_1, X_2, \dots, X_n$  formarem uma distribuição uniforme em  $[0,1]$ .

Assim, a multiplicação de Funções de Distribuição Beta gera novas de Funções de Distribuição Beta e a Distribuição Uniforme é um caso especial de Distribuição Beta.

Contudo, para fazer uso da Função de Distribuição Beta precisam ser primeiramente estimados os parâmetros alfa e beta. Uma maneira de estimar os parâmetros alfa e beta é feita aplicando-se o método de Estimação de Máxima Verossimilhança (Maximum Likelihood Estimation - MLE) a um conjunto de valores entre o intervalo  $[a,b]$ . Precisa-se de uma simulação para se obterem números probabilísticos entre o intervalo  $[a,b]$  de cada fator. Um método bem popular de simulação, usado frequentemente na indústria do petróleo, é o Monte Carlo.

#### 4.5.2 Simulação de Monte Carlo

Simulação é o método mais popular de execução de análise de risco em exploração petrolífera, e é o mais usado pelas principais empresas de petróleo. É fácil entender por que simulação é tão popular dentro da indústria do petróleo. A Simulação é usada por engenheiros para a modelação do desempenho dos reservatórios, projetar plataformas, entre outras aplicações. Já o economista usa a simulação para estudar os possíveis efeitos de mudanças no preço do petróleo, nas taxas de desconto, ou em políticas de investimento (KARBAUGH, 1995).

Porém, a razão mais importante do porquê de a simulação ser tão atraente no presente problema é o fato de permitir que os geólogos obtenham um número probabilístico de forma quase determinística, ou seja, fazendo que o geólogo use uma formalização para obter o número probabilístico ao invés de ser meramente uma estimativa subjetiva. O geólogo possui seis fatores geológicos, cada um composto de algumas variáveis, e, ao final, deve ser estimada uma probabilidade de sucesso para cada fator. O geólogo não conhece o valor certo de cada

variável para calcular a probabilidade do fator; ele apenas possui algumas evidências indiretas que fazem com que ele tenha uma idéia de qual a é variabilidade que cada variável do fator pode assumir, mas nunca a informação específica sobre o fator. Dessa forma, o geólogo deve criar combinações de valores válidos para cada variável (dentro da extensão aceita no modelo), criando, assim, resultados diferentes. De posse de um conjunto grande de resultados forma-se uma distribuição de probabilidade que descreve as chances ou a predisposição que aquelas variáveis têm para o fator geológico. A técnica de Simulação de Monte Carlo é justamente operar repetidamente estas várias combinações para criar uma distribuição de resultados.

A simulação de Monte Carlo foi desenvolvida durante a Segunda Guerra Mundial e tem conexão com as pesquisas da bomba atômica. A Simulação de Monte Carlo é muito aplicada na matemática e na estatística para a solução de problemas numéricos que de outra maneira seriam intratáveis. Em 1960, a técnica de simulação foi aplicada em análise de decisões e logo incorporada pela indústria do petróleo (KARBAUGH, 1995). A Simulação de Monte Carlo permite ao analista a opção de descrever riscos e incertezas introduzidas por variáveis na forma de uma distribuição de probabilidade (NEWENDDORP, 2000).

#### **4.5.2.1 Os passos para a realização da Simulação de Monte Carlo**

Quando o Sistema RCSUEX encerra a avaliação da Lógica Nebulosa, cada fator geológico possui um valor de saída lingüístico e um valor médio (numérico nítido) devido ao processo de desnebulização. A variável de saída Nebulosa de cada fator é uma medida de probabilidade que varia de zero (0) a cem (100) por cento. Contudo, associa-se um intervalo  $[a,b]$  à saída nítida da desnebulização para que se possa criar uma combinação de valores válidos entre este intervalo  $[a,b]$ . Deve-se lembrar que o significado deste intervalo  $[a,b]$  para o geólogo seria o fato de que um único valor probabilístico não seria facilmente aceito como verdadeiro, na medida em que se trata de avaliações sobre dados subjetivos e que a transformação para um número probabilístico único seria motivo de discordância, enquanto que um intervalo probabilístico é de mais fácil aceitação.

Em resumo, tem-se para cada fator geológico (geração, reservatório, migração, trapa, selo e sincronismo) um intervalo de valores probabilísticos que representam uma distribuição de probabilidade que descreve os valores prováveis que aquele fator pode ter. Admite-se que as distribuições de probabilidade são do tipo uniforme. Estas distribuições representam uma codificação das expectativas dos geólogos para um determinado fator, isto é, são probabilidades distintas, mas que são consideradas prováveis para um determinado prospecto.

O resultado da técnica de Simulação Monte Carlo é uma probabilidade conjunta, isto é, o produto das probabilidades individuais dos conjuntos de combinações possíveis sobre o intervalo [a,b]. Assim, a técnica de Monte Carlo envolve multiplicação das probabilidades, adicionando estas novas probabilidades a um conjunto de resultado final. Depois de algumas repetições sucessivas, o resultado é analisado para determinar quais valores estão dentro do intervalo especificado. Quando este conjunto de dados é visualizado na forma de um histograma ou de uma curva acumulativa, o resultado é uma distribuição de probabilidade de sucesso para o prospecto.

#### 4.5.3 Incorporação de dados históricos

Para que seja feita a incorporação dos dados históricos ajusta-se a uma função de Distribuição Beta, de forma que os valores dos parâmetros alfa e beta sejam somados a dados históricos da seguinte forma:

Alfa = alfa (estimado pelo modulo da Probabilidade Subjetiva) + número de sucessos de descobertas.

Beta = beta (estimado pelo modulo da Probabilidade Subjetiva) + (número de poços pioneiros – número de sucessos de descobertas).

Em termos matemáticos, o modelo proposto está suportado pelo seguinte teorema (apud De Groot, 1970, p. 160) (SILVA,2000):

“Suponha que  $W_1, \dots, W_n$  seja uma amostra aleatória de uma distribuição de Bernoulli com o parâmetro  $\Theta$  desconhecido. Suponha também que a distribuição a priori de  $\Theta$  seja a distribuição Beta com parâmetros  $\alpha$  e  $\beta$  sendo  $\alpha > 0$  e  $\beta > 0$ . Então a distribuição a posteriori de  $\Theta$  quando  $W_i = w_i$  ( $i = 1, \dots, n$ ) é uma distribuição Beta com parâmetros  $(\alpha + \delta)$  e  $(\beta + n - \delta)$ , onde  $\delta = \sum_{i=1}^n w_i$ .” (SILVA, 2000).

Em termos mais petrolíferos (SILVA,2000):

- $w_1, \dots, w_n$  corresponde ao histórico de  $n$  poços pioneiros que supostamente teriam sido perfurados com base em concepções geológicas análogas, com  $\delta$  igual ao número de sucessos (descobertas), representando o conhecimento objetivo

(freqüencialista) sobre a certeza de descobertas;  $w_i = 1$ , no caso de descoberta;  $w_i = 0$ , no caso de poço seco (ou descoberta não econômica)<sup>13</sup>;

- o parâmetro  $\Theta$  é a probabilidade de um pioneiro, a ser perfurado em condições geológicas análogas aos poços já perfurados, descobrir um campo de petróleo, representando o conhecimento sobre a certeza da descoberta (SILVA, 2000);
- a distribuição Beta *a priori* de  $\Theta$  representa o conhecimento subjetivo (geológico e difuso) sobre a certeza da descoberta, obtida conforme o item anterior (SILVA, 2000); e
- a distribuição Beta *a posteriori* de  $\Theta$  representa o conhecimento subjetivo combinado com o conhecimento objetivo, de acordo com o teorema (SILVA, 2000).

---

<sup>13</sup> De fato, o histórico também tem uma componente de subjetividade que está na escolha dos poços perfurados em situações geologicamente análogas.

## 5 Conexão entre saída fuzzy e probabilidade

### 5.1 Introdução

Um dos grandes desafios do Projeto foi definir a forma de conexão da metodologia Nebulosa com a metodologia Probabilística. Três metodologias foram propostas e avaliadas como forma de tratar esta conexão:

- Cálculo do intervalo  $[a,b]$  pelo Suporte da Função de Pertinência.
- Cálculo do intervalo  $[a,b]$  pelo Ponto de Corte da Função de Pertinência.
- Função de Saída Representa a Função de Distribuição de Probabilidade.

Neste capítulo apresentam-se formalmente estas metodologias de modo que sejam entendidas as abordagens usadas em cada uma delas. Também, apresentam-se as justificativas que nos levaram a optar por uma metodologia.

### 5.2 Cálculo do intervalo $[a,b]$ pelo Suporte da Função de Pertinência

No Sistema RCSUEX, o conhecimento incerto sobre a presença de petróleo num prospecto exploratório se dá por uma metodologia que combina Lógica Nebulosa com Teoria da Probabilidade, devido ao fato de ser esta última mais adequada às decisões de investimento em exploração de petróleo.

A representação probabilística difusa da certeza do sucesso de um prospecto exploratório segue os seguintes cinco passos:

Primeiro passo: Construção dos conjuntos Nebulosos de saída.

A primeira etapa para a transformação do domínio subjetivo, ou Neubuloso, para o domínio probabilístico, se dá pela interpretação dos valores lingüísticos da variável de saída do sistema Nebuloso como sendo a “favorabilidade” da existência, ou a frequência de sucesso, de determinado fator do sistema petrolífero.

Para o Sistema RCSUEX, os valores lingüísticos para avaliação da favorabilidade foram: “DESAVORÁVEL”, “QUESTIONÁVEL”, “NEUTRO”, “ENCORAJADOR” e “FAVORÁVEL”.

Assim, o administrador do Sistema ao definir um conjunto de saída Nebulosa, intrinsecamente associa valores de probabilidade a esta variável. Supondo-se, por exemplo, que o administrador defina valores **típicos** de “favorabilidade” ENCORAJADOR para um determinado fator como sendo entre 65% e 75% (0,65 e 0,75), mas admita que valores entre 50% e 90% (0,5 e 0,9) ainda possam ser considerados como encorajadores, então a função de

pertinência para este conjunto pode ser definida pela função trapezoidal<sup>14</sup> ( $a=0,5$ ,  $b=0,65$ ,  $c=0,75$ ,  $d=0,9$ ), onde  $a$  e  $d$  representam os valores mínimo e máximo do suporte do trapézio, enquanto os parâmetros  $b$  e  $c$  representam os valores mínimo e máximo da núcleo (base menor do trapézio) como mostra a Figura 16.

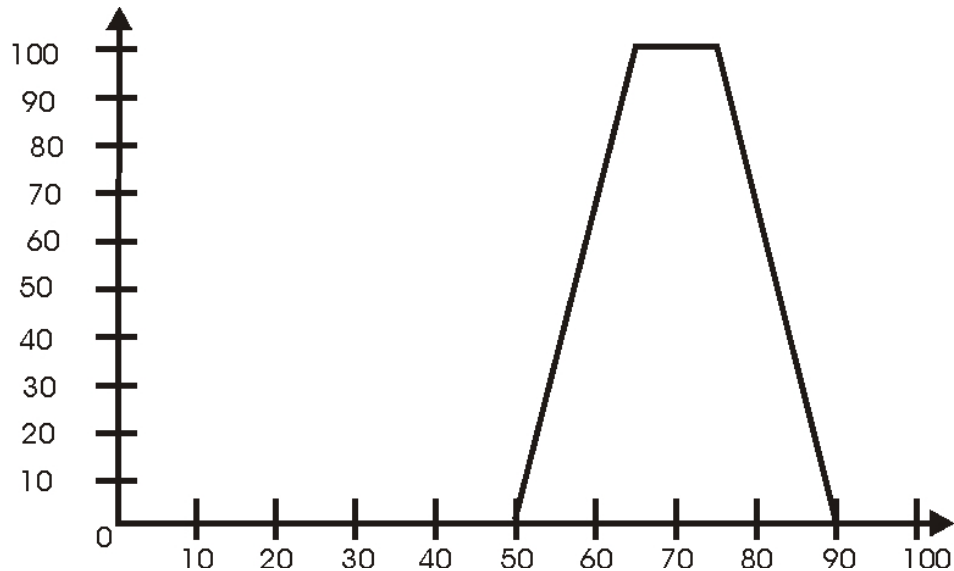


Figura 16: Mostra a construção de um trapézio da variável de saída nebulosa.

A Figura 17 ilustra todos os valores lingüísticos definidos no Sistema RCSUEX.

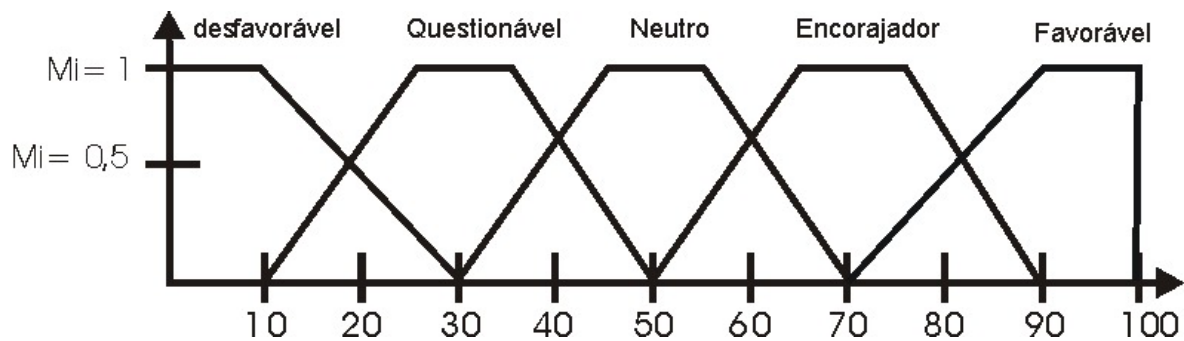


Figura 17: Representação dos conjuntos nebulosos de saída.

### Segundo passo: Desnebulização

<sup>14</sup> Representação trapezoidal ( $a, b, c, d$ ) de um número nebuloso ("fuzzy"): SE  $a \leq b < c \leq d$  e  $M(x)$  é a função pertencimento (*membership*) para o valor  $x$ , ENTÃO  $M(a) = M(d) = 0$ ,  $M(x) = 1$ ,  $x \in [b, c]$ .

A Desnebulização é a última etapa da Lógica Nebulosa que mapeia conjuntos Nebulosos em números nítidos, isto é, para um elemento do conjunto universo. Existem vários métodos que transformam saídas Nebulosas em número nítido, contudo, o método mais utilizado é o Centróide. O Método do Centróide calcula o centro de gravidade da área que leva em consideração a força de ativação das regras de cada conjunto Nebuloso de favorabilidade. Sua fórmula geral é:

$$\frac{\int \mu(x) * x dx}{\int \mu(x) dx}$$

Onde:

$\mu(x)$  : é o valor da força de ativação da função de pertinência.

$x$ : é o valor de favorabilidade da função de pertinência.

Terceiro passo: Associação do valor desnebulização com intervalos de probabilidade.

Segundo esta abordagem, o valor desnebulizado representa a probabilidade média para o fator petrolífero considerado. Este valor pode estar ou não associado a uma única variável lingüística.

Caso o valor desnebulizado seja exatamente igual à base média do trapézio que representa um valor lingüístico de favorabilidade, por exemplo: 0,70 para o caso de favorabilidade ENCORAJADOR, como visto no exemplo acima, então se associa uma distribuição de probabilidade uniforme, que representa probabilidades de sucesso do fator, com o intervalo dado pelos parâmetros **a** e **d** do trapézio.

Caso o valor desnebulizado não corresponda ao valor médio de um valor lingüístico de favorabilidade, por exemplo: 0,55 no exemplo acima, então se observa que este valor está associado a duas soluções lingüísticas (NEUTRO e ENCORAJADOR para o exemplo acima). Neste caso é necessário um processo de “interpolação” para a criação de um conjunto Fuzzy de favorabilidade que poderia ser chamado neste caso de “NEUTRO/ENCORAJADOR” e cujo valor médio do trapézio é, portanto, o valor Desnebulizado

O método para geração deste novo conjunto Nebuloso é o seguinte:

Calcula-se um valor  $\alpha$  que é a constante de proporcionalidade entre o valor desnebulizado e os pontos médios dos trapézios dos conjuntos nebuloso de favorabilidade imediatamente inferior e imediatamente superior. Para o exemplo que se está considerando,

estes conjuntos são o NEUTRO e ENCORAJADOR, que possuem pontos médios 0,50 e 0,70, respectivamente. A fórmula para cálculo de  $\alpha$  é:

$\alpha = (x - m_1) / (m_2 - m_1)$ , onde  $x$  é o valor desnebulizado pelo método do centróide descrito no segundo passo, e  $m_1$  e  $m_2$  são os pontos médios do trapézio inferior e superior, respectivamente.

Para o exemplo acima:  $\alpha = (0,55 - 0,50) / (0,70 - 0,50) = 0,25$

Calculam-se os valores dos parâmetros  $a$ ,  $b$ ,  $c$  e  $d$  do trapézio do novo conjunto Nebuloso pelas seguintes fórmulas:

$$a = a_1 + (a_2 - a_1) * \alpha$$

$$b = b_1 + (b_2 - b_1) * \alpha$$

$$c = c_1 + (c_2 - c_1) * \alpha$$

$$d = d_1 + (d_2 - d_1) * \alpha$$

onde  $a_1$ ,  $b_1$ ,  $c_1$  e  $d_1$  e  $a_2$ ,  $b_2$ ,  $c_2$  e  $d_2$  são os parâmetros  $a$ ,  $b$ ,  $c$  e  $d$  dos trapézios inferior e superior, respectivamente.

Para o exemplo acima:

$$a = 0,30 + (0,50 - 0,30) * 0,25 = 0,35$$

$$b = 0,45 + (0,65 - 0,45) * 0,25 = 0,50$$

$$c = 0,55 + (0,75 - 0,55) * 0,25 = 0,60$$

$$d = 0,70 + (0,90 - 0,70) * 0,25 = 0,75$$

A Figura 18 mostra os conjuntos Nebulosos de favorabilidade NEUTRO e ENCORAJADORA e o conjunto Nebuloso intermediário obtido pela metodologia.

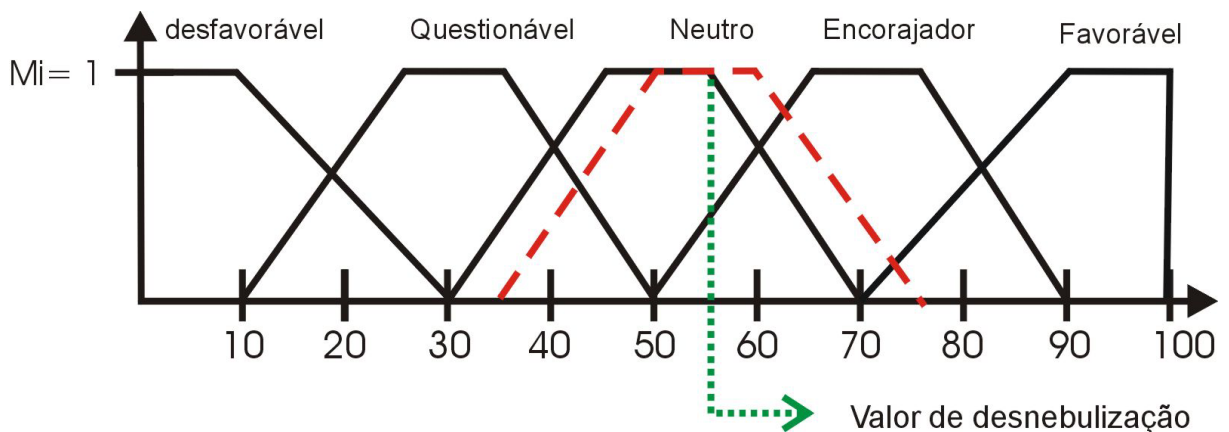


Figura 18: Caso onde o valor de desnebulizado está entre dois valores lingüísticos.



Uma vez obtido este novo conjunto Nebuloso NEUTRO/ENCORAJADOR, associa-se então uma distribuição de probabilidade uniforme, que representa probabilidades de sucesso do fator, com o intervalo dado pelos parâmetros **a** e **d** do novo trapézio.

Quarto passo: Simulação de amostra com possíveis medidas de (certeza) probabilidades de sucesso do prospecto

Admitindo-se que um prospecto será um sucesso se e somente se todos os fatores geológicos indicarem sucesso, combinam-se os fatores da maneira usualmente feita (Rose, 1991a) na indústria do petróleo, simulando pela técnica de Monte Carlo as probabilidades conjuntas para o prospecto. Assim, obtém-se uma distribuição não-paramétrica das possíveis medidas de probabilidade de sucesso do prospecto avaliado.

Quinto passo: Ajuste de uma Distribuição Beta de probabilidades à amostra simulada

Por fim, a distribuição não-paramétrica simulada no quarto passo é ajustada a uma Distribuição Beta de probabilidades.

### **5.3 Cálculo do intervalo [a,b] pelo Ponto de Corte da Função de Pertinência**

Este método é semelhante ao primeiro, contudo a extensão do intervalo da distribuição uniforme não é fixa, isto é, da extensão dada pelos valores dos parâmetros a e d da função de pertinência trapezoidal. Aqui o valor do intervalo é dado pelos pontos de intersecção do limite superior da força de ativação das regras com a função de pertinência.

Para este método, seguem-se os seguintes passos:

Primeiro passo: Construção dos conjuntos Nebulosos de saída

Este passo é igual ao primeiro passo proposto no método anterior.

Segundo passo: Desnebulização

Este passo é igual ao primeiro passo proposto no método anterior.

Terceiro passo: Associação do valor desnebulizado com intervalos de probabilidade.

Segundo esta abordagem, o valor denebulizado representa a probabilidade média para o fator petrolífero considerado. Este valor pode estar ou não associado a uma única variável lingüística.

Pode-se ter dois casos bem distintos:

- Um caso é onde o valor desnebulizado é exatamente igual à base média do trapézio que representa um valor lingüístico de favorabilidade, por exemplo: 0,70 para o caso de favorabilidade ENCORAJADORA da figura 1. Neste caso, o intervalo de probabilidade é obtido pelos pontos de intersecção entre a área de representação do trapézio e a força de ativação das regras ativadas conforme Figura 19.

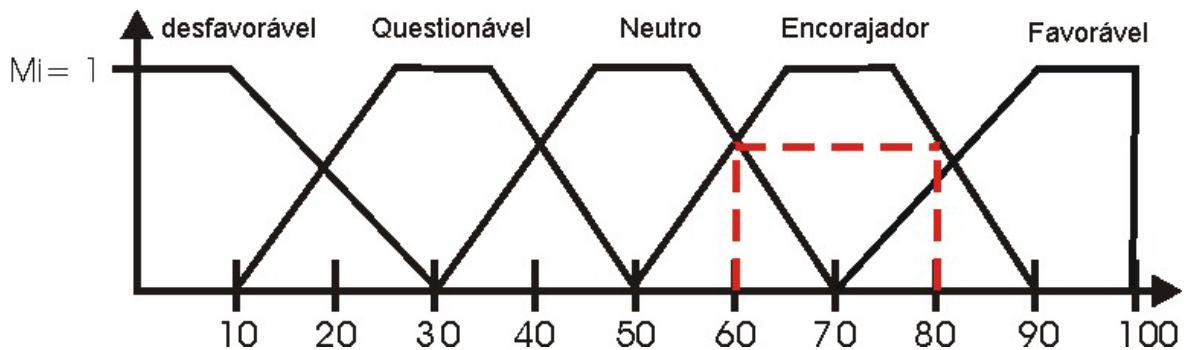


Figura 19: Obtenção do intervalo de probabilidade pelos pontos de intersecção entre a área de representação do trapézio e a força de ativação das regras ativadas.

- Outro caso é onde o valor desnebulizado não corresponda ao valor médio de um valor lingüístico de favorabilidade, por exemplo: 0,55 no primeiro exemplo, então se observa que este valor está associado a duas soluções lingüísticas (NEUTRO e ENCORAJADOR, conforme Figura 20).

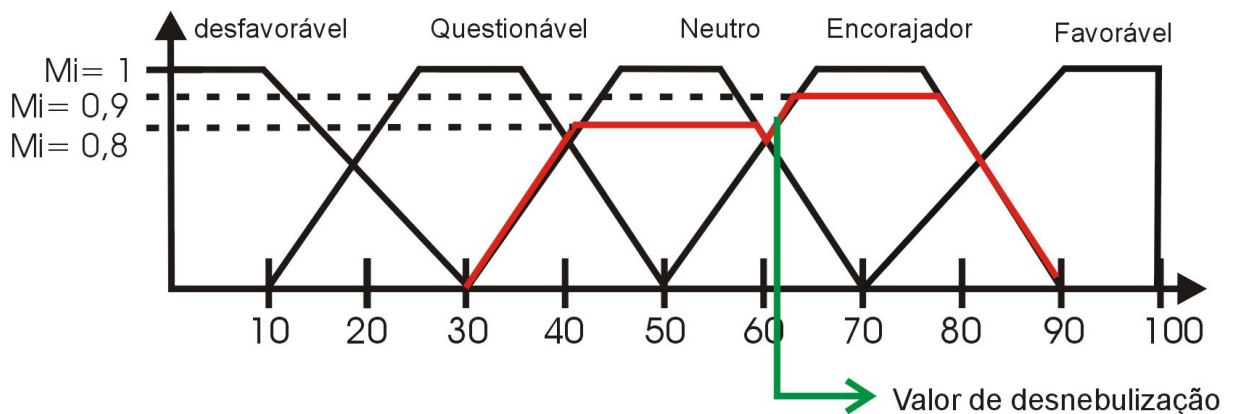


Figura 20: Representação das forças de ativação das regras de inferência Nebulosa.

Neste caso é necessário um processo de “interpolação” para a criação de um conjunto Nebuloso de favorabilidade que poderia ser chamado neste caso de

“NEUTRO/ENCORAJADOR” e cujo valor médio do trapézio é, portanto, o valor desnebulizado. O método para geração deste novo conjunto Nebuloso é o seguinte:

- calcula-se um valor  $\alpha$  conforme o método anterior.
- calculam-se os valores de a,b,c,d do novo conjunto Nebuloso conforme o método anterior.
- calcula-se a média das forças de ativação dos conjuntos Nebuloso ativados da seguinte maneira:

$$\frac{\sum_{i=1}^n \mu(x)}{n}$$

Onde:

n: número de conjuntos Nebulosos ativados.

$\mu(x)$ : força de Ativação de cada conjunto Nebuloso

Desta forma temos o novo conjunto Nebuloso e a sua “força de ativação”. A partir deste passo determina-se o intervalo de probabilidade da mesma forma que no primeiro caso, ou seja, o intervalo é obtido pelos pontos de intersecção entre a área de representação do novo trapézio e a força de ativação para este novo conjunto, conforme Figura 21.

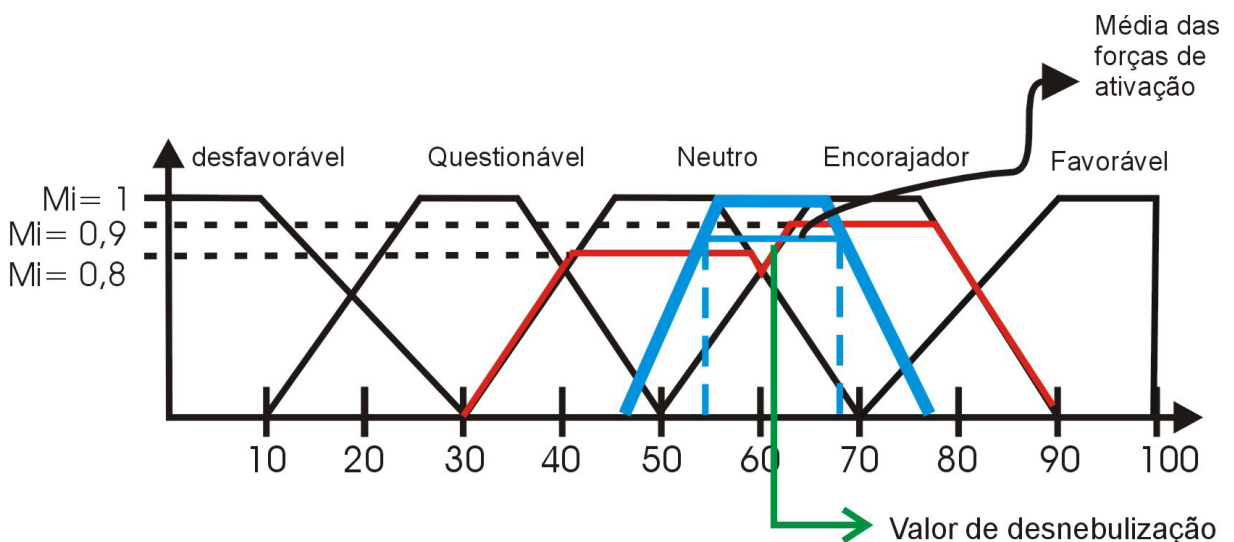


Figura 21: Ilustração da obtenção do intervalo probabilístico quando mais de uma variável linguística está ativada.

Quarto passo: Simulação de amostra com possíveis medidas de (certeza) probabilidades de sucesso do prospecto

Este passo é igual ao quarto passo proposto no método anterior.

Quinto passo: Ajuste de uma Distribuição Beta de probabilidades à amostra simulada

Este passo é igual ao quinto passo proposto no método anterior.

#### **5.4 Função de Saída representa a Função de Distribuição de Probabilidade**

Este terceiro método difere dos métodos anteriores basicamente no sentido de suprimir a etapa de desnebulização e de não considerar uma distribuição uniforme de probabilidade para cada fator do Sistema Petrolífero. A idéia central deste terceiro método é interpretar o conjunto formado pelas saídas ativadas de cada fator como sendo a distribuição de probabilidade específica daquele fator.

Os passos para implementação deste método são as seguintes:

Primeiro passo: Construção dos conjuntos Nebulosos de saída

Este passo é igual ao primeiro passo proposto no método anterior.

Segundo passo: Normalização da distribuição de probabilidade do fator

Ao serem executadas as regras de inferência Nebulosa, vários conjuntos de saída representando a favorabilidade do fator são ativados com diferentes valores de ativação. Isso gera uma área conforme a Figura 22, representada por uma função particular. Esta função pode ser interpretada como sendo a distribuição de probabilidade específica do fator. É necessário, entretanto, “normalizar” a função para que esta seja adequada aos requisitos de uma distribuição de probabilidade, ou seja:

$$F(x) = \int_0^1 f(x) dx = 1$$

Terceiro passo: Geração da distribuição de probabilidade do fator

Neste passo aplica-se o Teorema da Probabilidade Integral para, a partir de uma Distribuição Uniforme em [0-1], obterem-se valores de favorabilidade segundo a distribuição de probabilidade obtida no segundo passo, conforme ilustrado na Figura 22.

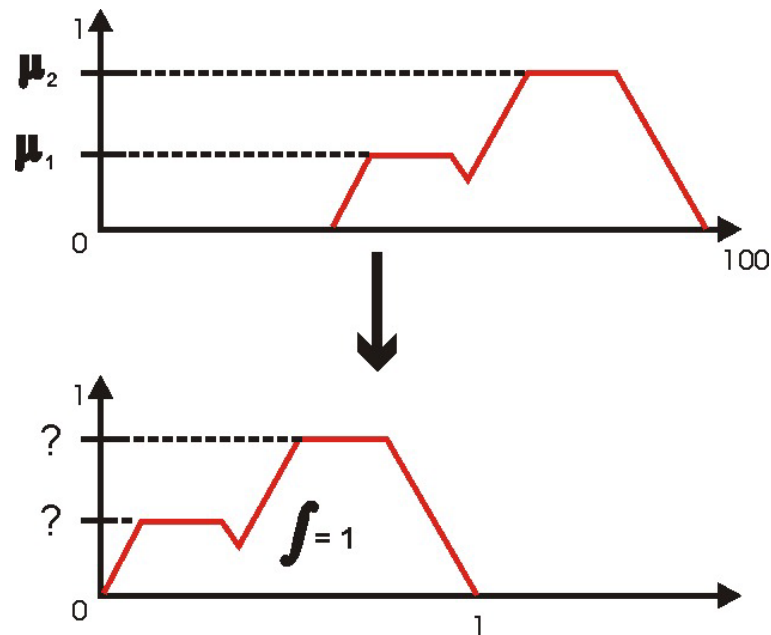


Figura 22: Ilustra o segundo passo onde se trata da Normalização da distribuição de probabilidade do fator.

Quarto passo: Simulação de amostra com possíveis medidas de (certeza) probabilidades de sucesso do prospecto

Este passo é igual ao quarto passo proposto no método anterior.

Quinto passo: Ajuste de uma Distribuição Beta de probabilidades à amostra simulada

Este passo é igual ao quinto passo proposto no método anterior.

### 5.5 Justificativas da escolha da conexão

Durante o projeto RCSUEX, as três metodologias propostas foram implementadas, testadas e avaliadas. Os resultados obtidos foram os seguintes:

- O valor médio da certeza do sucesso de um prospecto exploratório, apresentado pelos três métodos, foi o mesmo para exemplos fictícios submetidos ao Sistema RCSUEX.
- Como a variância do intervalo é que difere para os três métodos, observou-se que o segundo método – “Cálculo de A e B pelo Núcleo da Função de Pertinência” \_ apresentou freqüentemente valores de variância menor quando comparado com o método do “Cálculo de A e B pelo Suporte da Função de Pertinência“, enquanto que o terceiro método apresentava intervalo de variância maior que os outros dois métodos.
- A variância pequena do segundo método implicou a sua inflexibilidade para incorporar dados históricos. Em contrapartida, o terceiro método, “Função de

Saída Representa a Função de Distribuição de Probabilidade”, incorpora adequadamente os dados históricos.

- Para todos os métodos, o quinto passo, referente ao ajuste de uma Distribuição Beta de probabilidades à amostra simulada, é garantido teoricamente em Rohatgi (1976) pelo fato de as distribuições uniformes serem casos particulares da Distribuição Beta, o que permite que a distribuição resultante da sua multiplicação (quarto passo) possa ser ajustada a uma Distribuição Beta. Entretanto, na terceira metodologia não temos distribuições uniformes, e sim distribuições de probabilidade quaisquer. Assim, o ajuste de uma Distribuição Beta de probabilidades é realizado sem base formal.
- Optou-se pelo método do “Cálculo de A e B pelo Suporte da Função de Pertinência”, porque este possui um formalismo já defendido e aprovado, leva em consideração a força de ativação das regras selecionadas pela inferência nebulosa e é sensível à incorporação de dados históricos.

## 6 O Sistema RCSUEX

### 6.1 Introdução

A partir do momento em que o Projeto definiu a metodologia de tratamento dos dados subjetivos, iniciou-se a construção do Sistema RCSUEX. O Sistema RCSUEX é um sistema especialista que atende aos objetivos do projeto: a estruturação do conhecimento, representação computacional de um método para integrar argumentação geológica (lingüística), informações visuais e estatísticas exploratórias.

Os conceitos vistos no projeto RCSUEX foram implementados em Linguagem de Programação Builder C++, versão 5.0, seguindo padrão windows<sup>15</sup>.

### 6.2 O Sistema RCSUEX

O Sistema RCSUEX é uma ferramenta para o desenvolvimento, edição e avaliação de modelos nebulosos, isto é, existe uma ferramenta de construção de variáveis Nebulosas funções de pertinências e regras. Outro motivo para permitir criar vários modelos Nebulosos foi a constatação de que cada bacia teria suas próprias regras, isto é, cada bacia apresenta características diferentes para a presença do hidrocarboneto. Por exemplo, se para a bacia hipotética “XYZ” o tipo de rocha que forma o selo do reservatório possui uma porosidade de “x%” como ideal para a presença do selo para outra bacia, como a rocha selante é uma outra formação rochosa, o percentual da porosidade ideal para presença de selo pode ser outro valor diferente de “x”. Desta maneira, seria necessária a criação de um modelo Nebuloso (definição de funções de pertinências e regras) diferente para cada bacia. Além disso, o Sistema RCSUEX estará disponível para a construção de modelos Nebulosos para outros processos da Petrobras que não o exploratório, devido a esta abertura na construção de vários modelos Nebulosos. Com esta permissão de construção de modelos Nebulosos, fez-se necessária a criação de níveis de acesso ao Sistema RCSUEX: Administrador e Usuário.

O acesso em nível de “Administrador” pertence às pessoas que gerenciam o Sistema, ou melhor, que conhecem as metodologias empregadas no Sistema e definem modelos Nebulosos. O administrador deve ter conhecimentos de Lógica Nebulosa e estatística. Outras funções básicas do administrador seriam:

- Concessão de autorização para acessar o Sistema;

---

<sup>15</sup> aproveito aqui para agradecer, mais uma vez, ao colega Dennis Kerr Coelho que implementou o Sistema RCSUEX e tornou concretas as idéias aqui apresentadas. Muito Obrigada.

- Especificação dos modelos Nebulosos;
- Especificação do Cadastro de DHI;
- Demais funções do usuário.

Já o acesso em nível de “Usuário” seria adequado aos geólogos, que submeteriam seus prospectos ao Sistema, como forma de auxílio para expressar a probabilidade de sucesso. Para os acessos em nível de “Usuário” não são necessários conhecimentos profundos de Lógica Nebulosa, porque neste nível a metodologia está escondida do usuário do Sistema. As Funções básicas do acesso em nível de “Usuário” seriam:

- abrir modelos Nebulosos;
- submeter um prospecto ao modelo Nebuloso;
- consultar e imprimir dados referentes à avaliação do prospecto.

### **6.3 Representação de dados Subjetivos**

Enquanto se estudava o domínio do problema de exploração petrolífera constatou-se que as variáveis dos fatores geológicos possuíam domínios diferentes. Algumas variáveis poderiam ser quantificadas, isto é, representadas por um número; outras eram puramente lingüísticas ou nominais.

Na Lógica Nebulosa tradicional, a primeira etapa refere-se a nebulizar, ou seja, transformar as entradas nítidas em conjuntos Nebulosos. Mas uma vez que tendo-se variáveis que não são quantificáveis, não se pode aplicar simplesmente a nebulização direta. Surgiu, assim, a necessidade de classificar as variáveis por tipos. Algumas informações, como, por exemplo, a espessura do selo, são medidas em metros ou quilômetros; já o tipo de trapa é tratado de forma nominal: “mista”, “estratigráfica”, entre outras. Por esse motivo, as informações de entrada para o sistema especialista foram classificadas em:

- Variável Nominal: é aquela representada por uma palavra ou nome, por exemplo, preto/branco;
- Variáveis Ordinais: são aquelas que representam escalas de ordem que nem sempre diferem da mesma unidade; por exemplo, escala de dureza (1, 2, 3, ..., sendo que a diferença entre 1 e 2 não é a mesma de 2 a 3, ou ainda, o valor numérico 2 não é 2 x 1);
- Variável Racional: é aquela que representa números racionais; por exemplo, uma distância de 1 km (1 é um número racional);

Para cada uma destas classificações opta-se por um tratamento antes de se aplicar a nebulização.



No caso de variáveis nominais, como, por exemplo, o tipo de trapa, além do usuário responder qual o tipo da trapa, ele ainda deverá responder qual o fator de certeza da sua escolha. Este fator de certeza é um número entre zero e cem por cento (0% - 100%), que justificará sua confiança na resposta nominal dada. O Sistema RCSUEX aplica o fator de certeza fornecido pelo usuário como sendo a entrada para a nebulização daquela resposta. Caso o fator de certeza na resposta seja inferior a 100%, o Sistema rateia o restante do valor para fechar os 100% nas demais respostas possíveis. Dessa forma o fator de certeza já é o valor nebulizado da variável. Variáveis do tipo nominal são valores singulares (“Singleton”), isto é, são valores individuais que indicam uma única coisa ou nome, por isso não precisam ser definidas funções de pertinência para este tipo de variável. Basta definir-se a variável que o fator de certeza funcionará como valor de nebulização.

Já as variáveis do tipo ordinal e racional são tratadas pela forma tradicional da Lógica Nebulosa através da etapa de nebulização, onde o usuário informa uma entrada nítida e o Sistema RCSUEX nebuliza para conjuntos Nebulosos. A importância da distinção entre variáveis ordinais e racionais está na forma como o usuário entra com a variável. Assim, o sistema consegue validar se a variável é nominal e permite que o usuário informe um número racional e, caso a variável seja ordinal, o usuário utiliza uma barra de rolagem (“scroll-bar”) para informar o valor da variável. Isso se faz necessário porque variáveis do tipo ordinal são valores entre dois outros valores (idéia de seqüência, ordem), mas não se tem uma precisão de onde exatamente entre os valores.

#### **6.4 Os Módulos do Sistema**

O Sistema RCSUEX possui três módulos distintos:

- Probabilidade de Sucesso Subjetiva (a priori);
- Indicação de Sucesso a Posteriori (dados objetivos);
- DHI.

A Figura 23 ilustra o menu principal do Sistema RCSUEX com os dados de entrada, possibilidade de até seis fatores geológicos, e os módulos de Probabilidade Subjetiva, Indicação de Sucesso a Posteriori e DHI. A saída gerada pelo Módulo de Probabilidade Subjetiva é uma das entradas para os módulos de Indicação de Sucesso a Posteriori e Módulo de DHI. A Indicação de Sucesso a Posteriori pode ser feita sobre os dados subjetivos sem obrigatoriedade de aplicação da indicação de DHI, e da mesma forma o Sistema RCSUEX permite que seja incorporado o DHI sem a necessidade de incorporar sucesso a posteriori.

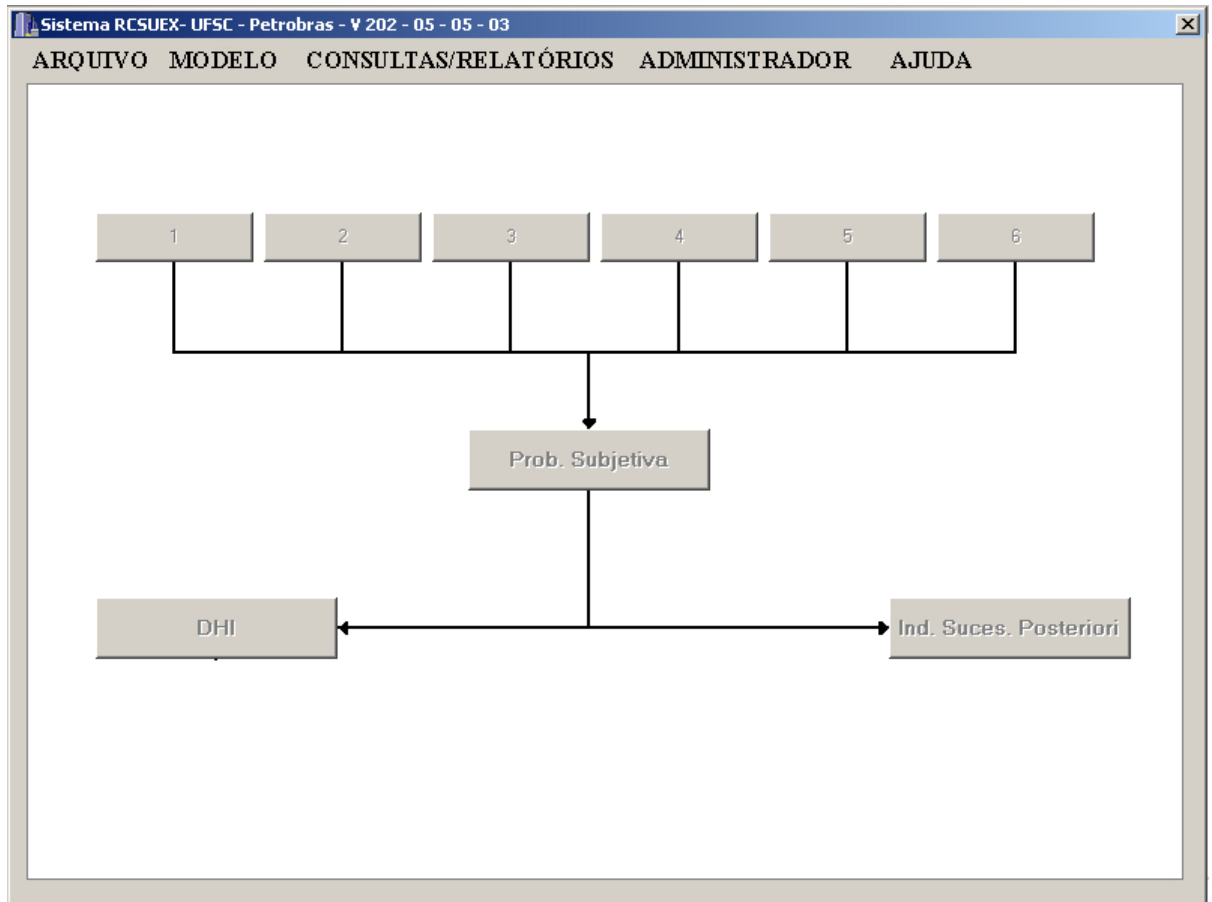


Figura 23: Ilustração do menu principal do Sistema RCSUEX.

O Módulo da Probabilidade Subjetiva trata da representação do conhecimento geológico incerto, através da Lógica Nebulosa. Nesta etapa, os dados subjetivos são observados através de perguntas e respostas. Cada pergunta refere-se a uma variável Nebulosa, enquanto que as respostas são mapeadas para conjuntos Nebulosos por intermédio da etapa de nebulização. Por exemplo, uma pergunta onde a variável é do tipo racional: “Qual a espessura do selo ?” para o fator Rocha Selante. O usuário então responderia um valor numérico que seria nebulizado com graus de pertinências a conjuntos nebulosos. De forma semelhante trabalhar-se-ia com variáveis do tipo nominal, como, por exemplo, “Qual o tipo de trapa ?” para o fator Trapa. A resposta do usuário seria um nome, por exemplo, “estratigráfica”, e informaria qual a sua certeza nesta resposta, por exemplo, 80%. Desta forma, o Sistema rateia os outros 20% para os demais conjuntos nebulosos da variável. Com esta etapa de nebulização das variáveis de entrada encerrada, a Máquina de Inferência Nebulosa, a partir da base de regras, seleciona as regras mais adequadas às entradas fornecidas. Assim, as regras definem a força de ativação de cada conjunto nebuloso que forma

a variável de saída, onde se aplica o último processo da Lógica Nebulosa: a desnebulização, gerando uma solução.

Ao final da primeira etapa do Sistema RCSUEX, obtêm-se seis números probabilísticos que correspondem à favorabilidade na ocorrência da acumulação de petróleo para os seis fatores geológicos do sistema petrolífero.

Contudo, esta saída numérica não é exata, pois está carregada de subjetividade, por isso associa-se a cada um destes números um intervalo  $[a,b]$ , que representa a variância possível deste valor numérico. Por fim, aplica-se a Simulação de Monte Carlo, que é uma probabilidade conjunta, isto é, o produto das probabilidades individuais dos conjuntos de combinações possíveis para posterior ajuste a uma função de Distribuição Beta.

Portanto, o Módulo da Probabilidade Subjetiva, ilustrado pela Figura 24, gera como saída os seguintes itens:

- Solução lingüística para a probabilidade do sucesso de cada fator geológico;
- Associação de cada solução lingüística com intervalos de probabilidade  $[a,b]$  (ver explicação no capítulo 5);
- Ajuste de uma Distribuição Beta de probabilidade à amostra simulada (média e variância);

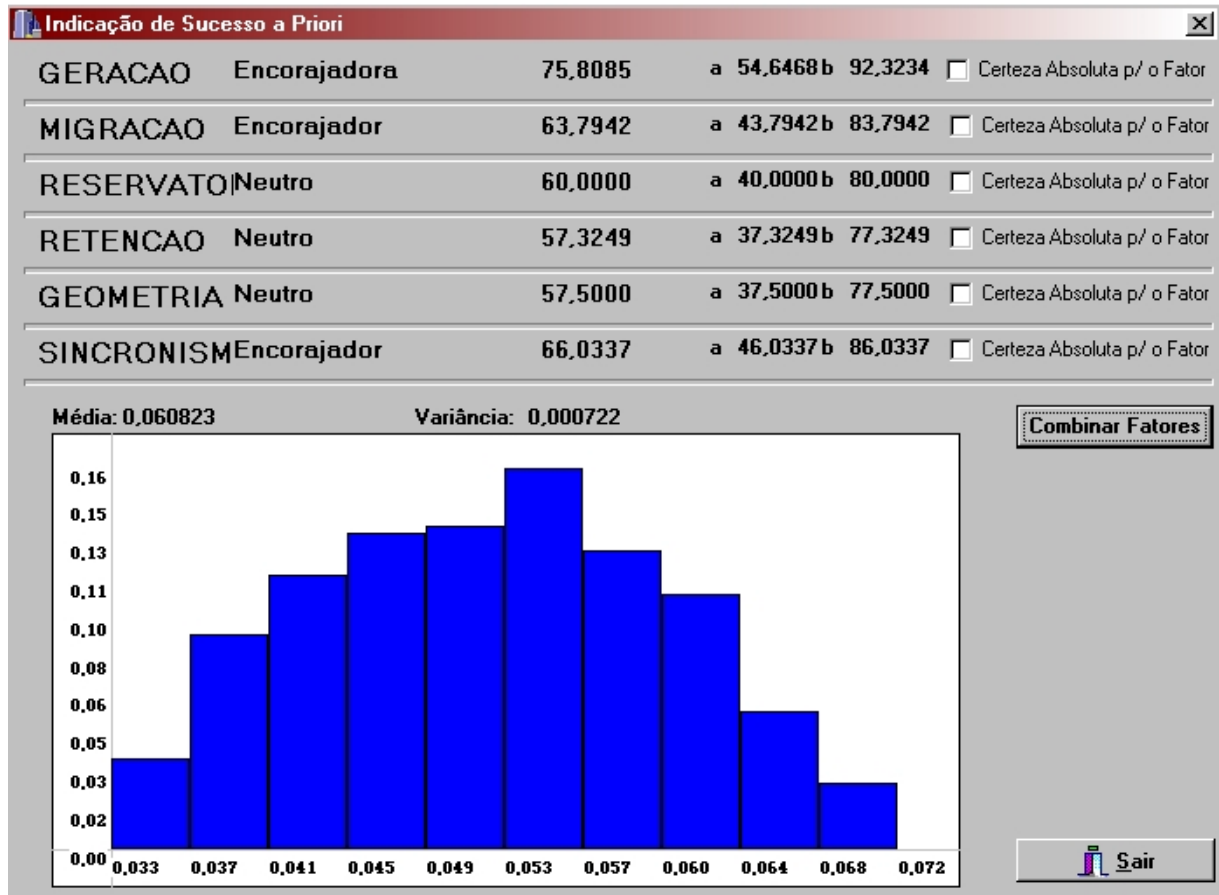


Figura 24: Ilustra a tela de saída da Probabilidade Subjetiva do Sistema RCSEUX

O Módulo de Indicação de Sucesso a Posteriori se refere à incorporação de dados históricos. Só será possível indicar Sucesso a Posteriori depois de gerada a Probabilidade Subjetiva. As saídas geradas no Módulo de Probabilidade Subjetiva são as entradas para o Módulo de Sucesso a Posteriori.

Desta forma, a distribuição a posteriori, ilustrada pela Figura 25, representa o conhecimento subjetivo combinado com o conhecimento objetivo. Ao final desta etapa tem-se:

- um histograma para os dados históricos informados (número de poços pioneiros e número de descobertas) com os valores de média e variância. Ilustrado pela Figura 25 (a);
- um histograma com a aplicação da posteriori com a incorporação dos dados históricos, isto é, probabilidade subjetiva conjugada com objetiva (dados históricos informados), com média e variância. Ilustrado pela Figura 25 (b);

- um histograma com aplicação da posteriori com a incorporação da tecnologia de DHI, isto é, probabilidade subjetiva conjugada com objetiva (dados históricos) e conjugada com a tecnologia de DHI. Ilustrado pela Figura 25 (c).

Por fim tem-se o módulo de incorporação da tecnologia de DHI, ilustrado pela Figura 26. Quando existe uma indicação de DHI muito forte, o prospecto como um todo fica com uma favorabilidade grande para acumulação de hidrocarboneto. Assim, incorpora-se a tecnologia de DHI de forma semelhante aos dados históricos, ou seja, alteram-se os valores de alfa e beta conforme a intensidade do DHI associado ao prospecto. É importante lembrar que a tecnologia de DHI possui cinco graus de intensidade (“Muito Forte”, “Forte”, “Moderado”, “Fraco” e “Muito Fraco”) e que para cada valor lingüístico de intensidade existe um valor numérico que representa a quantidade de poços descobertos com aquela respectiva indicação de DHI. Desta forma, os novos valores de alfa e beta são:

Alfa = alfa (estimado pelo Módulo da Probabilidade Subjetiva) + número de sucessos de descobertas ou valor da respectiva intensidade do DHI.

Beta = beta (estimado pelo Módulo da Probabilidade Subjetiva) + (número de poços pioneiros – número de sucessos de descobertas ou valor da respectiva intensidade do DHI).

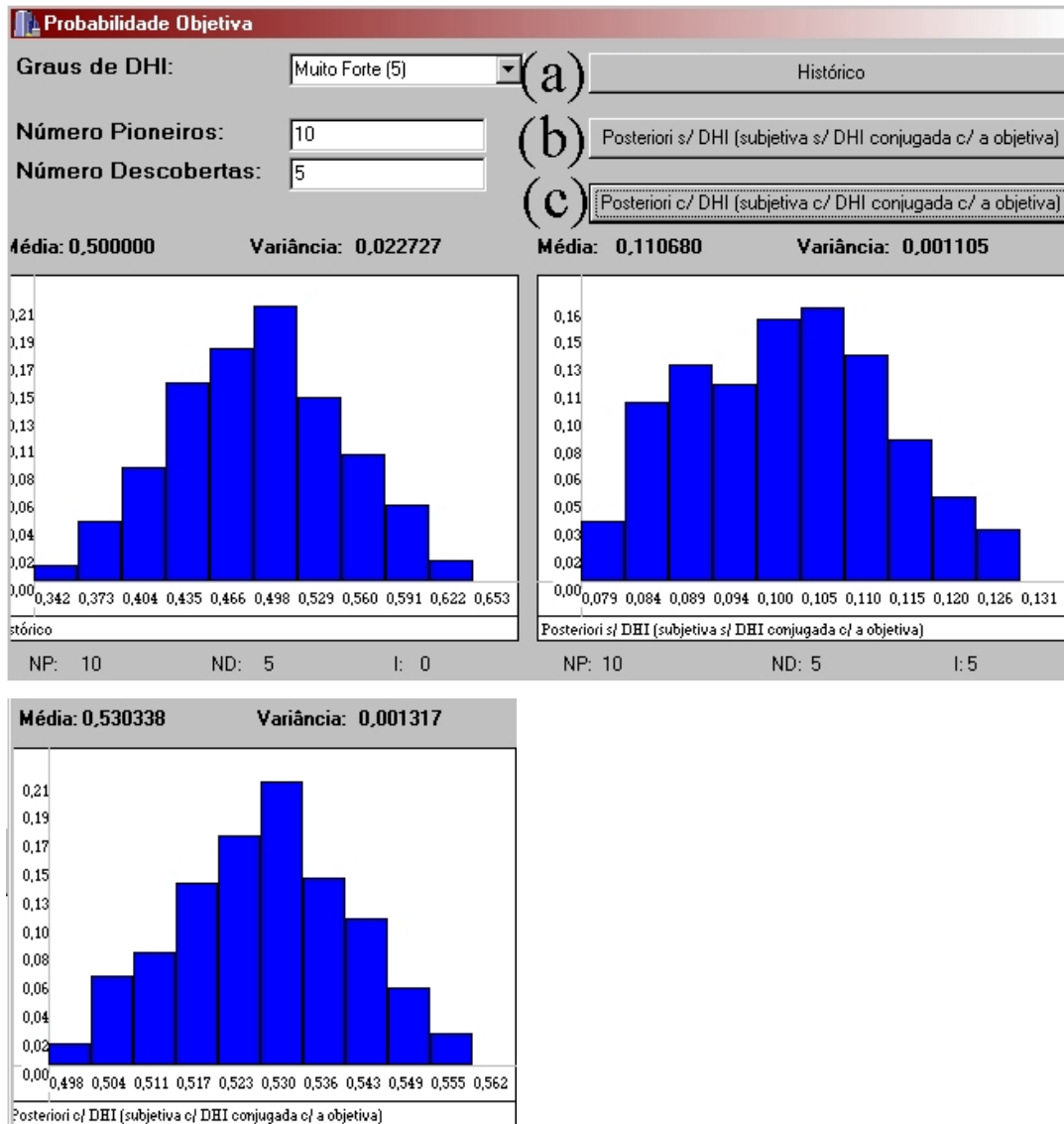


Figura 25: Representa a saída do Módulo de Indicação de Sucesso a Posteriori: (a) o botão histórico gera um gráfico sobre os valores de poços pioneiros e descobertos; (b) incorpora dados históricos sobre os dados subjetivos, não aplica DHI; (c) incorpora dados históricos à saída do Módulo Subjetivo e DHI.

Estes valores de intensidade do DHI são cadastrados pelo Administrador do Sistema RCSUEX e podem ser alterados a qualquer momento. Deve-se lembrar também que o Sistema RCSUEX permite que a tecnologia de DHI seja aplicada tanto a priori quanto a posteriori da incorporação dos dados históricos.

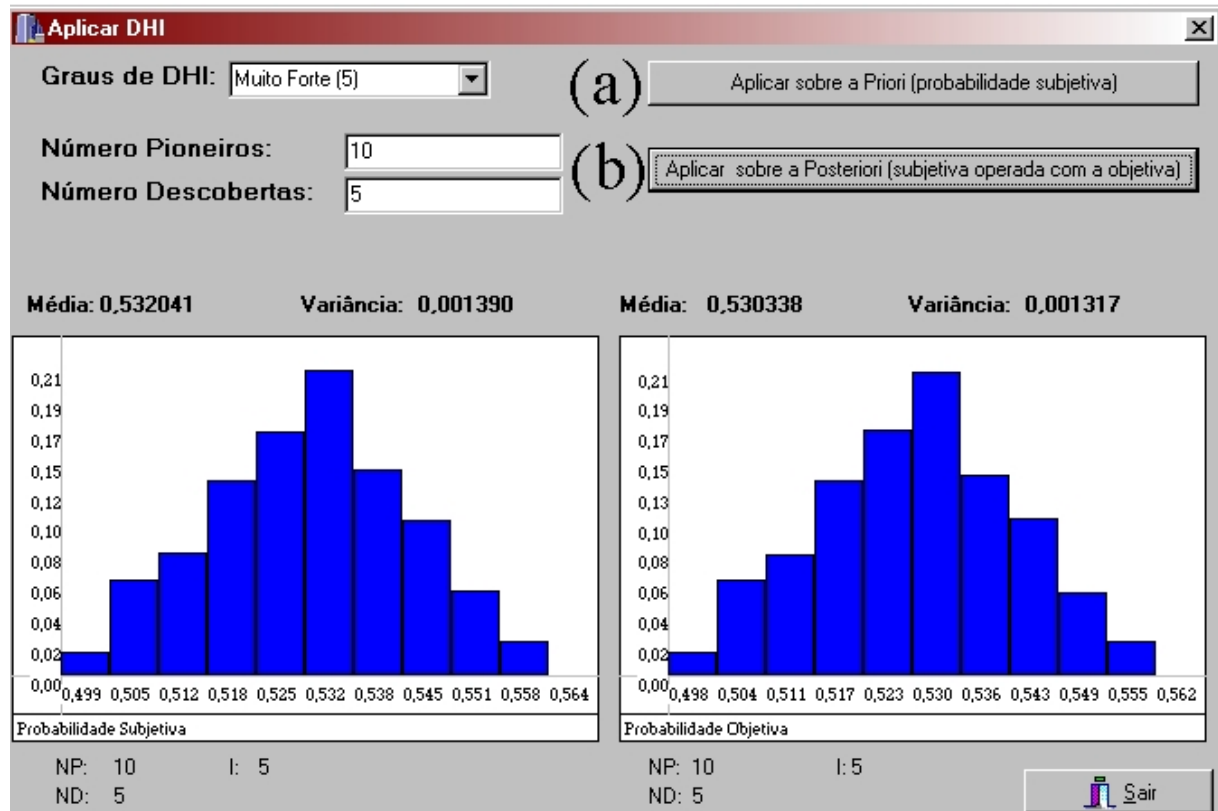


Figura 26: Ilustra a saída do módulo de incorporação de DHI: (a) incorpora a DHI sobre os dados subjetivos; (b) incorpora DHI sobre os dados subjetivos operados com dados objetivos.

## 7 Conclusão

A exploração petrolífera é uma atividade repetitiva, repleta de informações imperfeitas e que envolve um grande investimento financeiro envolvendo a análise de fatores geológicos e requerendo estudos de viabilidade econômica.

Analisando o domínio do problema petrolífero, descobriu-se que o tipo de imperfeição mais aparente é a incerteza. Em muitas variáveis o geólogo chega a uma conclusão, mas, mesmo assim, não descarta por completo as outras possibilidades, demonstrando incerteza na argumentação. Essas variáveis incertas costumam ser perfeitamente estimadas somente depois que o poço foi perfurado e os resultados reais disponibilizados. A indústria do petróleo avançou muito em nível tecnológico, tentando minimizar suas incertezas, mas ainda não as eliminou.

O objetivo principal deste trabalho, modelar o conhecimento através de uma representação estruturada, foi atingido em sua totalidade. O modelo nebuloso definido no Sistema RCSUEX auxilia o geólogo a identificar as principais características a serem avaliadas, gerando assim uma lista de informações (“chek-list”) a ser observada antes de perfurar-se um novo prospecto. Ele auxilia na comparação entre prospectos de um portfólio, mensurados sob as mesmas variáveis. O resultado deste trabalho, o Sistema RCSUEX, é uma representação computacional que integra argumentação geológica (lingüística) e estatísticas exploratórias na avaliação de prospectos.

Levar-se em conta apenas a argumentação geológica e não se levar em consideração o valor das freqüências de sucesso observadas, seria ignorar o histórico da área em estudo, de grande valia na representação da certeza do sucesso de prospectos que tenham concepção semelhante. Da mesma forma, apenas avaliar pela perspectiva objetiva, ignorando a argumentação geológica, indicações em mapas, dados sísmicos entre outros, seria erro grave. Portanto, propôs-se uma metodologia que se conecta informações subjetiva e objetiva através de um modelo matemático válido.

Uma etapa crítica do Projeto RCSUEX foi a calibração do sistema nebuloso. Calibrar é ajustar o Sistema de forma que a resposta fornecida corresponda a uma resposta realista. Calibrar o Sistema RCSUEX foi um processo bastante complexo e envolveu algumas alterações, como:

- formatos das funções de pertinência das variáveis do sistema no modelo nebuloso;
- inserção e alteração de regras no modelo;



- alteração nos valores dos pesos das regras de inferência nebulosa.

A calibração foi feita juntamente com especialistas que submetiam prospectos para os quais já possuíam uma argumentação e uma probabilidade de sucesso, esperando que o sistema RCSUEX respondesse de forma consistente.

Observou-se nesta etapa que o processo de construção do Modelo Nebuloso Petrolífero, quanto ao conjunto de regras para a base de conhecimento e definição das funções de pertinência para os atributos, era crítico para o funcionamento adequado do Sistema RCSUEX.

Os vários especialistas da Petrobras (geólogos, interpretes, etc. ) a quem se apresentou o Sistema RCSUEX, após a etapa de calibração, e que submeteram alguns prospectos ao Sistema, ficaram satisfeitos com a respostas fornecida pela ferramenta de suporte ao processo de análise de risco exploratório.

Uma grande vantagem do Sistema RCSUEX é a sua flexibilidade, pois permite a definição de Modelos Nebulosos, para outras aplicações, fazendo uso da mesma máquina de inferência.

Quanto à metodologia de Lógica Nebulosa, alguns itens devem ser esclarecidos, entre eles:

- Existem outros operadores não clássicos da Lógica Nebulosa, além do mínimo e do máximo. Contudo, preferiu-se trabalhar com os operadores clássicos, mínimo e máximo, visto que se comportavam de maneira satisfatória.
- O projeto RCSUEX estudou e apresentou à Petrobras um estudo sobre os diferentes métodos de desnebulização, de forma a justificar a sua escolha pelo método do centróide para o domínio do problema petrolífero.

Concluindo, este trabalho atendeu a todos objetivos propostos em termos de Sistema:

- apresentar uma representação computacional estruturada para o conhecimento no processo de exploração petrolífera;
- integrar argumentação geológica (dados subjetivos) e dados históricos;
- apresentar uma distribuição de probabilidades de sucesso para o prospecto exploratório.

### **7.1 Trabalhos Futuros**

Este trabalho foi uma etapa para estruturação e representação computacional na Exploração Petrolífera. A partir dele, vislumbraram-se outros trabalhos a serem testados e desenvolvidos. Sugere-se como trabalho futuro a geração automática de regras, já que no Sistema RCSUEX as regras são incluídas de forma manual pelo administrador do sistema. Além disso, sugere-se que a própria calibração do Modelo Nebuloso possa ser feita de forma automática, utilizando-se alguma metodologia de IA.

Acredita-se que à medida que o Sistema for sendo utilizado, este será uma boa forma de estudo de post mortem dos prospectos que não obtiveram sucesso, e até mesmo para os prospectos com sucesso exploratório pode-se realizar uma Mineração de Dados (“Data Mining”) para identificar modelos de prospectos.

## Anexo A

### Lógica Nebulosa

“Lógica Nebulosa é basicamente uma lógica multivalorada que permite definir valores intermediários entre avaliações convencionais como Sim/Não, Verdadeiro/Falso, Preto/Branco etc. Noções como *bastante quente* ou *bastante frio* podem ser formuladas matematicamente e processadas através de computadores”  
Definição de Lógica Nebulosa por Bauer *et al.*

#### Introdução

O conceito de Lógica Nebulosa foi concebido por Lotfi Zadeh, professor da Universidade de Berkley, na Califórnia, em 1965. A Lógica Nebulosa é uma extensão da Lógica Clássica. Nesta uma sentença pode assumir só dois valores: *Verdadeiro* ou *Falso*, o que se caracteriza como uma lógica dicotômica. Na Lógica Clássica, cada fato ou sentença, como 'choverá amanhã' deve ser verdadeiro ou falso.

Contudo, o professor Zadeh argumentou que as pessoas não trabalham sempre com informações precisas, e ainda assim são capazes de tomar decisões altamente adaptáveis. Muitas das informações que as pessoas usam sobre o mundo envolvem algum grau de incerteza.

O professor Zadeh, então, desejou inserir nos sistemas computacionais a idéia de que uma sentença ou fato não precisa necessariamente pertencer a um único conjunto, mas ter pertinências aos conjuntos. Assim, uma sentença não precisa ser verdadeira ou falsa, mas terá uma pertinência no conjunto verdade e outra pertinência no conjunto falso. A dicotomia entre pertencer e não pertencer não existe, mas existe uma gradação de pertinência da sentença aos conjuntos.

Desta forma, a Lógica Nebulosa provê uma maneira simples de chegar-se a uma conclusão definitiva utilizando-se informações vagas, ambíguas, imprecisas, ruidosas, ou, até mesmo, a perda de informação. A Lógica Nebulosa imita o raciocínio humano na tomada de decisões, pois é uma técnica para representar e manipular informação imprecisa.

Este anexo tem por objetivo explicar a Lógica Nebulosa, definindo seus conceitos e apresentando exemplos que ilustrem estas definições.

## Conjuntos Nebulosos

Conjunto significa uma coleção de coisas do mesmo tipo, como, por exemplo, um conjunto de letras, números, formas geométricas, frutas, cores, etc. Na Lógica Clássica, um elemento pertence (é verdadeiro para o conjunto) ou não pertence (é falso ao conjunto); por exemplo, o conjunto das frutas poderia ser {laranja, maçã, pêsego, banana} e o elemento {triângulo} não pertence ao conjunto das frutas, e sim ao conjunto das formas geométricas {quadrado, triângulo, retângulo, trapézio}. Assim, pode-se definir a pertinência de um elemento particular  $X$  de um conjunto  $A$  em um determinado universo  $U$  através da função característica (TRANSCHEIT, 1996):

$$\mu_A(x): U \rightarrow \{0,1\}$$

$$\mu_A(x) = 1 \quad \text{se } x \in A$$

$$\mu_A(x) = 0 \quad \text{se } x \notin A$$

Já os conjuntos Nebulosos, que fazem parte da Lógica Nebulosa, possuem uma diferença para os conjuntos Clássicos: o termo conjunto continua sendo uma coleção de coisas, mas agora seus elementos podem pertencer a mais de um conjunto ao mesmo tempo, com um certo nível de pertinência. Por exemplo, a temperatura da água de um chuveiro poderia ser classificada conforme cinco níveis de temperatura: muito fria, fria, normal, quente ou muito quente. Um conjunto clássico possui um limite rígido entre estes cinco conjuntos, sendo que a temperatura da água do chuveiro poderá pertencer a apenas um desses conjuntos, conforme a Figura 27(a). Um conjunto Nebuloso não possui uma fronteira (um limite) precisa entre seus conjuntos, de forma que a temperatura da água pode pertencer ao conjunto normal com um certo grau de pertinência e ao conjunto quente com outro grau de pertinência (Figura 27(b)).

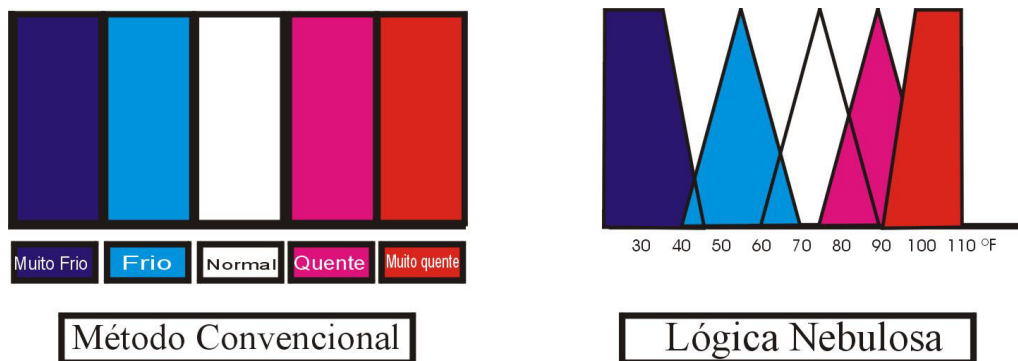


Figura 27: (a) Conjuntos Tradicionais

(b) Conjuntos Nebulosos (FUZZY,2002)

Do ponto de vista térmico, se a temperatura da água estiver em 40°C ou 41°C é indiferente, pois a esta temperatura a água está quente para um banho. Considere-se o que se faz no chuveiro se a temperatura está muito quente: far-se-á que a água fique mais para morno. Lógica Nebulosa é capaz de imitar este tipo de comportamento.

Zadeh propôs uma caracterização mais ampla, generalizando a função característica, de modo que ela pudesse assumir um número infinito de valores no intervalo  $[0,1]$ . Um conjunto Nebuloso  $A$  em um universo  $U$  é definido por uma função de pertinência  $\mu_A(x):U \rightarrow [0,1]$ , e representado por um conjunto de pares ordenados

$$A = \{ \mu_A(x)/x \}, x \in U.$$

Um conjunto Nebuloso  $A$  é o subconjunto de pontos  $x$  de  $U$  tal que  $\mu_A(x) > 0$ , sendo que se este conjunto possuir apenas um único ponto  $x$  e este tiver um  $\mu_A(x) = 1$  é chamado de conjunto unitário Nebulosa ou singleton (TRANSCHEIT, 1996).

### Variáveis Nebulosas

Uma variável é como um objeto lingüístico utilizado para representar os conjuntos Nebulosos. Ao introduzir uma informação, por exemplo, “temperatura”, associam-se variáveis para modificar o valor de entrada desta informação (por exemplo, “temperatura” está associada às variáveis “quente”, “normal” e “fria”).

### Operadores dos Conjuntos Nebulosos

A Teoria da Lógica Nebulosa está baseada na Lógica Clássica e, por isso, possui propriedades e teoremas matemáticos análogos. A teoria dos conjuntos Nebulosos possui os operadores de Complemento, União e Intersecção.

Sejam dois conjuntos Nebulosos  $A$  e  $B$  em um universo  $U$  e para qualquer  $x \in U$  tem-se:

- Complemento de  $A$  – O Complemento de um conjunto contém todos os elementos que não estão no conjunto. A definição formal de complemento é:

$$\mu_{\overline{A}}(x) = 1 - \mu_A(x).$$

- Intersecção – Na teoria clássica, intersecção de dois conjuntos são todos os elementos que pertencem aos dois conjuntos. Na Teoria Nebulosa, o elemento possui um grau de pertinência em cada conjunto, e a intersecção fica sendo o equivalente ao conectivo “e”. A definição formal de intersecção é:

$$\mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x))$$

- União – Na União Nebulosa se quer ressaltar que o elemento pertence (isto é, possui um  $\mu(x) > 0$ ) a qualquer um dos conjuntos Nebulosos e equivale ao conectivo “ou”. A definição formal da União é:

$$\mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x))$$

Na teoria dos conjuntos Nebulosos, a Intersecção é implementada por uma família de operações chamadas de t-normas, e a União é implementada por uma família de operações chamadas de t-conormas. O conjunto das t-normas e t-conormas forma as normas triangulares. Cada norma triangular  $\nabla : [0,1]^2 \rightarrow [0,1]$  verifica para todos os a,b,c,d em [0,1] as seguintes propriedades:

- Comutatividade:  $\nabla(a,b) = \nabla(b,a)$ ;
- Associatividade:  $\nabla(a, \nabla(b,c)) = \nabla(\nabla(a,b), c)$
- Monotonicidade:  $\nabla(a,b) \leq \nabla(c,d)$  se  $a \leq c$  e  $b \leq d$

Usando-se os operadores de Intersecção, União e Complemento na Lógica Nebulosa agregam-se as seguintes propriedades:

- Involução:  $\overline{\overline{a}} = a$
- Comutativa:  $a + b = b + a$
- Associativa:  $(a + b) + c = a + (b + c)$
- Distributiva:  $a * (b + c) = (a * b) + (a * c)$
- Idempotência  $a + a = a$
- Absorção:  $a + (a * b) = a$
- Identidade:  $a + \phi = a$
- Leis de Morgan

Desta forma, justifica-se uma das grandes vantagens da Lógica Nebulosa: seu embasamento matemático.

### Funções de Pertinência

A Função de pertinência é uma função matemática que dá o significado numérico a um conjunto Nebuloso, isto é, ela mapeia variáveis de entrada rígida de um domínio específico para níveis de pertinências entre 0-1.

Na Lógica Clássica aplica-se o conceito de completamente verdade (tendo grau de verdade um (1.0)), ou completamente falso (tendo grau de verdade zero (0.0)). Já a Lógica

Nebulosa associa um valor numérico entre zero(0) e um(1) para cada sentença representando incerteza, isto é, a Lógica Nebulosa mede o grau para o qual a sentença está correta.

Função de Pertinência é uma representação gráfica do valor de cada entrada. Cada função de pertinência especifica um intervalo válido para uma variável Nebulosa . Na Figura 28, nota-se que a função de pertinência não define limites bruscos onde a variável é aplicada de maneira completa, pois existe uma região onde o valor da variável é mudado de forma gradual.

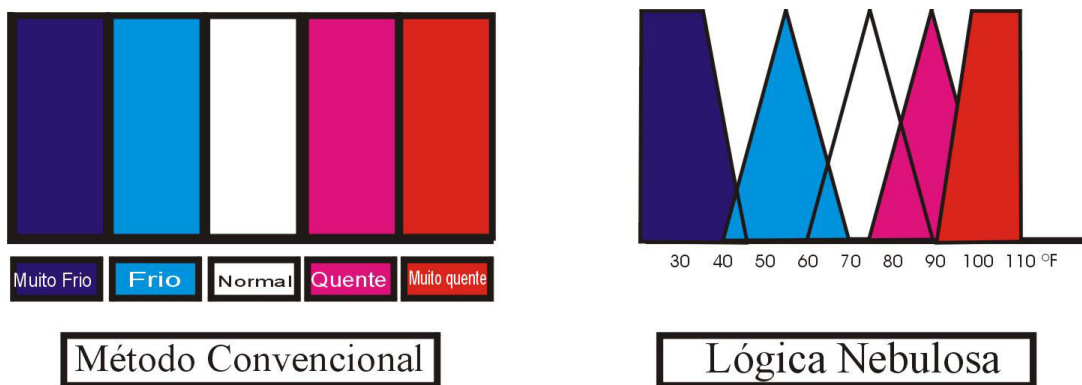


Figura 28: Conjuntos para a variável Temperatura (FUZZY,2002).

As funções de pertinência possuem diferentes formas, sendo que a trapezoidal e triangular são as mais comuns (Figura 29).

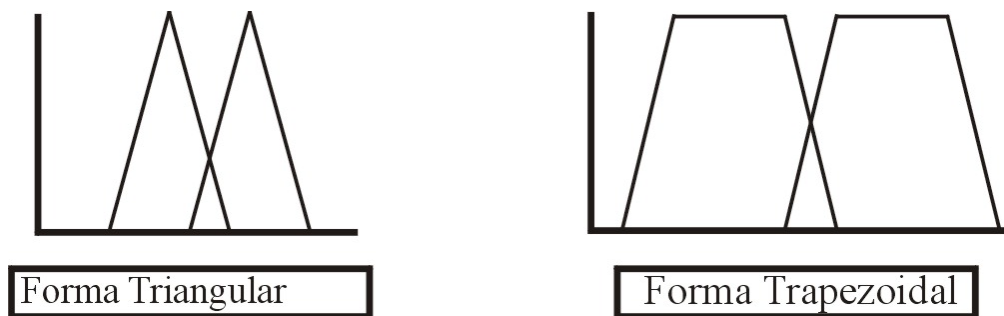


Figura 29: Formas mais comuns da função de pertinência.

A forma triangular é um caso especial da forma trapezoidal, porque nesta são necessários quatro pontos escalares para definir-se a função  $(x_1, x_2, x_3, x_4)$  e basta fazer-se  $x_2 = x_3$  para se passar da forma trapezoidal para a forma triangular. Na forma triangular possui-se apenas um ponto onde a função possui valor Máximo de um(1), enquanto que a função trapezoidal possui um intervalo de valores onde a função possui pertinência de um(1). Ambas as funções são de fácil computação e requerem pouca memória.

### Regras Nebulosas

Regras Nebulosas são cláusulas do tipo SE-ENTÃO que descreve uma ação a ser tomada dependendo dos valores de entrada na Nebulização. A sintaxe das regras é:

SE <antecedente> ENTÃO <conseqüente>

Onde:

<Antecedente> se refere a uma variável lingüística associada a uma função de pertinência, por exemplo, “temperatura da água = quente”.

<Conseqüente> se refere a uma variável lingüística de saída associada a uma função de pertinência, por exemplo, “abrir torneira = água fria”.

Desta forma, uma regra para o sistema do chuveiro seria:

SE temperatura da água = quente ENTÃO abrir torneira = água fria

Para se agrupar em uma regra mais de um antecedente, utilizam-se operadores básicos da Lógica Nebulosa: Complemento (não), União (ou) e Intersecção (e). Desta maneira, ao se desejar construir uma regra basta adicionar-se o conectivo de intersecção ou união entre os antecedentes da seguinte forma:

SE temperatura da água = quente E quantidade de água = pequena ENTÃO abrir torneira = água fria.

### Sistema da Lógica Nebulosa

A Figura 30 ilustra o Sistema de Lógica Nebulosa explicado até aqui, que consiste de:

- Nebulizar: mapeia um conjunto de entradas nítidas para um conjunto de entradas Nebulosas
- Base de Regras: contém uma coleção de regras SE-ENTÃO.
- Máquina de Inferência Nebulosa: mapeia o conjunto Nebuloso para outro conjunto Nebuloso de acordo com a base de regras e funções de pertinências.
- Desnebulizar: mapeia um conjunto de saída Nebulosa para um conjunto de saídas nítidas.



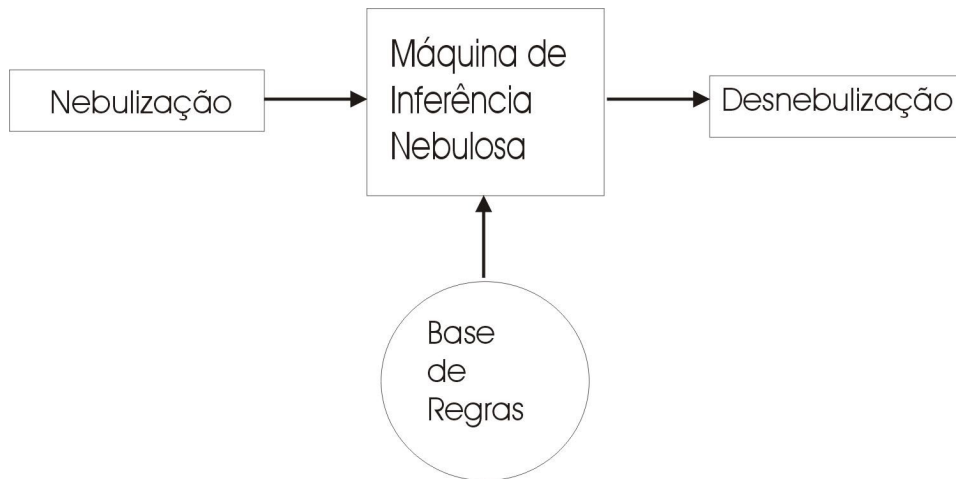


Figura 30: Sistema de Lógica Nebulosa.

### **Nebulizar**

É o primeiro processo da Lógica Nebulosa que transforma as entradas nítidas em conjuntos Nebulosos. Por exemplo, se a entrada nítida for a temperatura da água de 95°C, ela será transformada pela Nebulização em termos de conjuntos Nebulosos para “quente” com um certo grau de pertinência e “muito quente” com outro grau de pertinência. Isto é, o valor 95°C passa agora a descrever um valor lingüístico. Para transformar uma entrada nítida em um conjunto Nebuloso é necessário determinar a função de pertinência para cada entrada.

Voltando ao exemplo da temperatura da água (variável Nebulosa) com as seguintes funções de pertinências: “muito frio”, “frio”, “normal”, “quente” e “muito quente”. A entrada nítida da variável de temperatura da água será de 67°C que, aplicando o processo de Nebulização, corresponderá a um grau de pertinências de 0.15 ao conjunto “frio” e 0.65 ao conjunto “normal”. O grau de pertinência é representado por  $\mu(x)$  onde x equivale ao valor 67°C. A Figura 31 ilustra a Nebulização da variável temperatura.

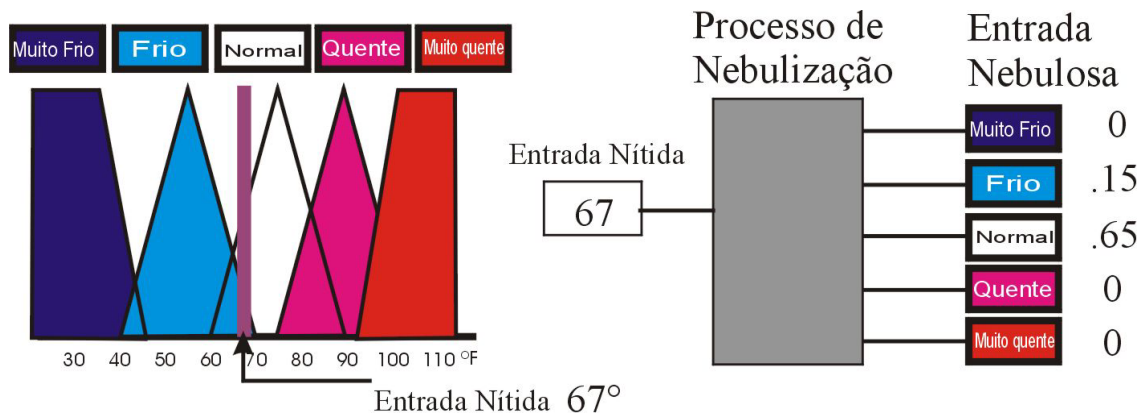


Figura 31: Nebulização da variável Temperatura(FUZZY,2002).

### Inferência

A Nebulização fornece o valor de  $\mu(x)$  para cada antecedente de uma regra, de forma que estes  $\mu$ 's(x) serão tratados dependendo do tipo de operador da regra. Quando se trabalha com o operador de União (ou) o mais comum é que se utilize a função Máximo do grau de pertinência, isto é, o grau de pertinência da regra será o antecedente da regra de maior grau de pertinência; portanto, a união de conjuntos Nebulosos é determinado pelo maior valor de função de pertinência máxima.

No operador de intersecção (e) a inferência da regra é o valor mínimo do grau de pertinências dos antecedentes da regra. Analisando a Tabela 3, onde se tem uma regra com o conectivo “E” e outra com o conectivo “OU” e seus antecedentes com os respectivos valor de  $\mu(x)$ .

Regras	Força Regra
1. SE temperatura água = muito quente(0.46) E quantidade de água = pequena (0.25) ENTÃO abrir=água fria.	0.25
2. SE temperatura água = quente(0.20) OU quantidade de água=pequena(0.75) ENTÃO abrir = água fria.	0.75

Tabela 3: Inferência de Regras Nebulosas .

Com a força de cada regra definida, deve-se identificar todas as regras ativadas com a mesma função de pertinência para determinar-se um único valor de ativação para esta saída.

Supondo-se que a Tabela 4 seja das regras ativadas pelas entradas Nebulosas do nosso sistema do chuveiro. Nota-se que existem valores de ativação diferentes para o mesmo conjunto Nebuloso, como, por exemplo, abrir água quente com valor de pertinência de 0.20 e

com valor de pertinência de 0.46. A Lógica Nebulosa Tradicional sempre seleciona a regra com maior valor de ativação para o mesmo conjunto Nebuloso (água quente). No caso entre as regras 3 e 4, a selecionada seria a de número 4, devido ao seu valor de ativação ser maior.

Regras	Força Regra	Selecionada
1. SE temperatura = muito quente (0.46) E quantidade de água = pequena (0.25) ENTÃO abrir=água fria	0.25	Sim
2. SE temperatura = muito quente (0.20) E quantidade água = normal (0.75) ENTÃO abrir = água fria.	0.20	Não
3. SE temperatura = fria (0.20) E quantidade de água = grande(0.25) ENTÃO abrir = água quente.	0.20	Não
4. SE temperatura = muito quente (0.46) E quantidade de água = grande (0.75) ENTÃO abrir = água quente	0.46	Sim

Tabela 4: Inferência de Regras Nebulosas: selecionar o valor do conjunto Nebuloso de saída.

### Desnebulização

A Desnebulização é a última etapa da Lógica Nebulosa que mapeia conjuntos Nebulosos em números nítidos, isto é, para um elemento do conjunto Universo. Existem vários métodos que transformam saídas Nebulosas em números nítidos, contudo o método mais utilizado é o Centróide. Neste método é feito o cálculo da área dos conjuntos Nebulosos ativados pelas regras, e o valor do centro desta área será o valor de saída. A fórmula genérica para o cálculo do Centróide é:

$$\text{Centróide} = \frac{\sum_{i=1}^n \mu(x) * x}{\sum_{i=1}^n \mu(x)}$$

Onde,

N é o número de pontos

$\mu(x)$  é o valor da função de pertinência do i-ésimo ponto da região

x é o valor da Função de pertinência do i-ésimo ponto da região.

A Figura 32 ilustra uma saída Nebulosa na qual se vai aplicar o método do Centróide. A área em amarelo refere-se à área de saída dos conjuntos Nebulosos.

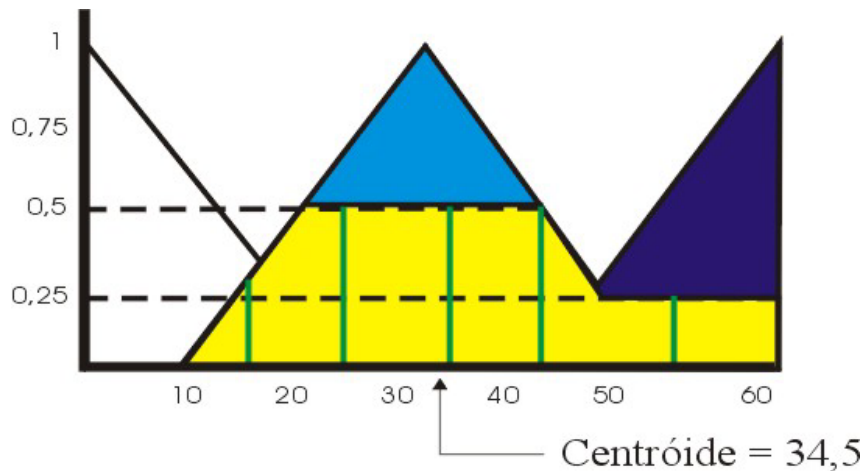


Figura 32: Aplicação do método do centróide(FUZZY,2002).

Se o conjunto  $x=\{15,25,35,45,55\}$  está representado como uma linha vertical verde, o valor de cada  $\mu(x)$  correspondente aos valores de  $x$  é  $\mu(x)=\{0.3,0.50,0.50,0.50,0.25\}$ . Desta forma basta aplicar-se a fórmula do método Centróide:

Valor da Desnebulização = 34.5

$$\frac{(15*0.3)+(25*0.50)+(35*0.50)+(45*0.50)+(55*0.25)}{0.3+0.50+0.50+0.50+0.25} = \frac{70.75}{2.05} = 34.5$$

### Aplicações de Lógica Nebulosa: Decisão e Controle

A área de maior atuação da Lógica Nebulosa são os Sistemas de Controle. Estes são aplicações que possuem variáveis quantitativas, isto é, são sistemas que a partir das variáveis de entrada quantitativas geram uma variável de saída também quantitativa. Lógica Nebulosa aplicado a controle de circuitos eletrônicos é um exemplo típico de aplicação de Sistema de Controle.

Contudo, tem havido um crescente interesse na teoria de Lógica Nebulosa aplicada a Sistemas de Decisão, os quais são aplicações que auxiliam os especialistas a tomarem uma decisão. Sistemas de Decisão envolvem a seleção ou escolha de alternativas, que, geralmente, está baseada em uma combinação de informação empírica e julgamento do ser humano. Para tomar-se uma decisão, tende-se a ver semanticamente o problema, e tenta-se automatizar o processo, ou melhor, traduzir o que se entende semanticamente do problema em um

algoritmo quantificável. O uso de Lógica Nebulosa permite tomar decisões fazendo-se regras para representar-se o domínio do problema. Isto torna o desenvolvimento de Sistemas de Apoio de Decisão uma ferramenta que aproxima especialistas novatos a especialistas experientes, e disponibiliza uma metodologia uniforme para justificativas de decisões e, conseqüentemente, compreensível a todas as pessoas envolvidas no Sistema.

O projeto RCSUEX está classificado como uma aplicação de decisão que se caracteriza em construir a semântica do problema da exploração petrolífera, buscando automatizar o processo e aproximar as opiniões entre especialistas novatos e experientes, disponibilizando uma justificativa mais estruturada para que o fórum responsável pela decisão de perfurar ou não um novo prospecto se sinta mais seguro.

### Referências Bibliográficas

AZEVEDO, F.M., BRASIL, L.M., OLIVEIRA, R.C.L. **Redes Neurais com aplicações em controles e em sistemas especialistas.** Florianópolis : Bookstore, 2000, 401 p.

BARRETO, J. M. **Inteligência Artificial do século XXI.** Florianópolis : Duplic, 2001. 379 p.

BITTENCOURT, G. **Inteligência Artificial: ferramentas e teorias.** Florianópolis : Editora da UFSC, 1998. 362 p.

BUCHANAN, B.; SHORTLIFFE, E. H. **Rule-Based Expert System:** The MYCIN experiments of the Stanford heuristic programming project. Mass.: Addison-Wesley Publishing Company, 1984. 748 p.

BUENO, F.S.. **Dicionário Escolar da Língua Portuguesa.** Rio de Janeiro : Faname, 1982. 1263 p.

CHEN, H. C., FANG J.H. **A New Method for Prospect Appraisal.** AAPG Bull. 77 N° 1, (January) : 9-18

CLARK, P. **Exemplar-based Reasoning in Geological Prospect Appraisal.** 1989

DACHS, N. **Estatística Computacional.** Rio de Janeiro : Livros técnicos e científicos editora, 1988. 86 p.

DUBOIS, D., PRADE, H. **Fuzzy Sets And Systems: Theory and Applications.** New York : Academic Press, 1980. 393 p.

DUCLÓS, M. L. **Consciência Home-Page.** Disponível em:

<<http://www.consciencia.org/antiga/arist.shtml>>. Acesso em Fevereiro de 2003.

**FUZZY Expert Exploration Tool (FEE Tool)** da Petroleum Recovery Research Center (PRRC) – New Mexico Tech. Disponível em: <<http://baervan.nmt.edu>>. Acesso em outubro de 2002

**FUZZY Tech Home Page** . Disponível por www em:  
<[http://www.fuzzytech.com/e/e\\_ft4bf4.html](http://www.fuzzytech.com/e/e_ft4bf4.html)>. Acesso em dezembro de 2002.

GAINES, B.R.; BOOSE, J. H.; **Machine Learning and Uncertain Reasoning**. Academic Press Limited, 1990. 3 vol.

GRUBER, T. **The acquisition of strategic knowledge: perspectives in artificial intelligence**. Boston : Academic Press, 1989. 311 p.

GUPTA, M. M.; YAMAKAWA, T. **Fuzzy Logic in Knowledge-Based Systems, Decision and Control**. Amsterdam : Elsevier Science Publishers, 1988. 410 p.

KARBAUGH, J. W., DAVIS, J. C., WENDEBOURG. **Computing Risk for Oil Prospects: Principles and Programs**. Florence : Pergamon, 1995. 452 p.

KLIR, G. J., FOLGER, T. A. **Fuzzy Sets, Uncertainty, and Information**. New Jersey : Prentice-Hall, 1988. 352 p.

KOVÁCS, Z. L. **Redes Neurais Artificiais: Fundamentos e aplicações**. 2 ed. São Paulo : Edição Acadêmica, 1996. 174 p.

LADEIRA, M.; VICARRI, R. M.; COELHO, H. **Redes Bayesianas Multiagentes**. Rio de Janeiro: Tutorial apresentado no ENIA 99 – Encontro Nacional de Inteligência Artificial, 1999. 28 p.

MAGOON, L.B., DOW, W. G.; **The Petroleum System-From Source to Trap**. Tulsa, E.U.A : Published by The American Association of Petroleum Geologists – AAPG –1994. 654 p.

MOTOROLA. **Fuzzy Logic Program 2.0** produzido por Motorola by Cotex Communications, Incs. Austin, Texas 1992-1994.

NEWENDDORP, P. D., SCHUYLER, J. **Decision Analysis for Petroleum Exploration**. Aurora : Decision Precision, 2000. 606 p.

NORVING, P.; RUSSELL, S. **Artificial Intelligence: A Modern Approach**. New Jersey : Editora Prentice-Hall, 1995. 932 p.

OTIS, R. M., SCHNEIDERMAN, N.; **A Process for Evaluating Exploration Prospect**. AAPG Bull. v .81, No. 7 (July) : 1087-1109

REID, C.A., FUNG, R.M., TONG R.M., TSE E. **A Knowledge Representation for Reasoning About Petroleum Geology** 2<sup>nd</sup> IEEE Computer Soc. Artif. Intel. Appl. Conf. 1985

ROSE, P. R. **Risk Analysis and Management of Petroleum Exploration Ventures**. Tulsa, Oklahoma: publicado por *The American Association of Petroleum Geologists Tulsa, Oklahoma U.S.A., 2001*. AAPG Methods in Exploration Series, Nro 12. 164 p.

ROHATGI, V. K. **An Introduction to Probability Theory and Mathematical Statistics**. New York: Wiley-Interscience publication, 1976.

SCHWADE, I.; **Geologic Quantification: Description → numbers → Success Ratio**. Los Angeles: Then American Association of Petroleum Geologists Bulletin, July, 1967, v51 nro: 7, p 1225-1239.

SILVA, R. R. **Explorator: Protótipo de Sistema Holístico em Exploração de Petróleo**. Rio de Janeiro, 2000. Tese – Universidade Federal do Rio de Janeiro, realizada no Instituto de Geociências.

STEVEN, D. K.; **Fuzzy Logic – An Introduction** disponível em: <<http://www.seattlerobotics.org/encoder/library.html>>. Acesso em julho de 2002.



SEMINÁRIO de **Matemática e Lógica Fuzzy**. Capítulo 1, disponível em:  
<<http://members.tripod.com/geloneze/fuzzy.htm> >. Acesso em junho de 2002.

SUSLICK, S.B. **Conhecer as incertezas: o desafio da indústria de petróleo**. Disponível em: <<http://www.comciencia.br/reportagens/petroleo/pet20.shtm> >. Acesso em junho de 2003.

TANSCHKEIT, R. **Lógica Fuzzy, Raciocínio Aproximado e Mecanismos de Inferência**. DEE-PUC-Rio, CP 38.063 22.452-970, Rio de Janeiro, 1996.

UCHÔA, J. Q.; PANONTIM, S. M.; NICOLETTI, M. **Elementos da Teoria de Evidencia de Dempster-Shafer**. UFSCar - São Carlos, Universidade Federal de São Carlos – SP, departamento de Computação, 1997, 29 p.