



NÚMERO: 444/2012
UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE GEOCIÊNCIAS

MICHELLE CHAVES KURODA

**REDES NEURAS APLICADAS AO ESTUDO DE ROCHAS
RESERVATÓRIO**

DISSERTAÇÃO DE MESTRADO APRESENTADA
AO INSTITUTO DE GEOCIÊNCIAS DA UNICAMP
PARA OBTENÇÃO DO TÍTULO DE MESTRE EM
GEOCIÊNCIAS, NA ÁREA DE GEOLOGIA E
RECURSOS NATURAIS.

ORIENTADOR: PROF. DR. ALEXANDRE CAMPANE VIDAL

CO-ORIENTADOR: PROF. DR. EMILSON PEREIRA LEITE

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA
DISSERTAÇÃO DEFENDIDA PELA ALUNA, E ORIENTADA
PELO PROF. DR. ALEXANDRE CAMPANE VIDAL

Campinas, 2012

FICHA CATALOGRÁFICA ELABORADA POR
CÁSSIA RAQUEL DA SILVA – CRB8/5752 – BIBLIOTECA “CONRADO PASCHOALE” DO
INSTITUTO DE GEOCIÊNCIAS
UNICAMP

Unidade 18.1
TA/UNICAMP
Cutter _____
V. _____ Ed. _____
Tombo BC 753.28
Proc. 16.10.10
C. _____
Preço 18.11.12
Data 18.11.12
Cód. lit. 850378

K966r Kuroda, Michelle Chaves, 1984-
Redes neurais aplicadas ao estudo de rochas
reservatório / Michelle Chaves Kuroda--
Campinas,SP.: [s.n.], 2012.

Orientador: Alexandre Campanhe Vidal.
Dissertação (mestrado) - Universidade Estadual de
Campinas, Instituto de Geociências.

1. Mapas auto-organizáveis. 2. Reservatórios. 3.
Análise multivariada. I. Vidal, Alexandre Campanhe,
1969- II. Universidade Estadual de Campinas, Instituto de
Geociências. III. Título.

Informações para a Biblioteca Digital

Título em inglês: Neural networks applied to the study of reservoir rocks.

Palavras-chaves em inglês:

Self-organizing maps

Reservoirs

Multivariate analysis

Área de concentração: Geologia e Recursos Naturais

Titulação: Mestre em Geociências.

Banca examinadora:

Alexandre Campanhe Vidal (Presidente)

Jurandir Zullo Filho

Marcelo Monteiro da Rocha

Data da defesa: 27-02-2012

Programa de Pós-graduação em Geociências



UNICAMP

**UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE GEOCIÊNCIAS
PÓS-GRADUAÇÃO EM GEOCIÊNCIAS NA
ÁREA DE GEOLOGIA E RECURSOS NATURAIS**

AUTORA: Michelle Chaves Kuroda

"Redes Neurais Aplicadas ao Estudo de Rochas Reservatório."

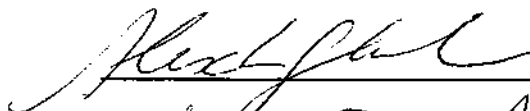
ORIENTADOR: Prof. Dr. Alexandre Campanhe Vidal

CO-ORIENTADOR: Prof. Dr. Emilson Pereira Leite

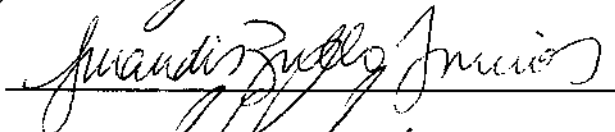
Aprovada em: 27 / 02 / 2012

EXAMINADORES:

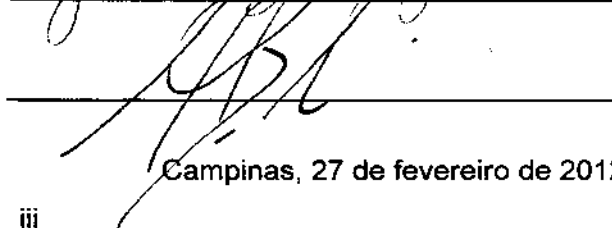
Prof. Dr. Alexandre Campanhe Vidal

 - Presidente

Prof. Dr. Jurandir Zullo Júnior



Prof. Dr. Marcelo Monteiro da Rocha



Campinas, 27 de fevereiro de 2012

DEDICATÓRIA

*Aos avós, Deuclécio e Eurides,
Mitytoshi e Tuiyuko (in memoriam),
aos pais, Oscar e Teresa
e à irmã, Marcela
- com amor.*

AGRADECIMENTOS

Agradeço ao prof. Dr. Rodrigo Portugal por ter encaminhado minha solicitação de orientação de mestrado ao prof. Dr. Vidal e ao prof. Dr. Emilson Leite, pela co-orientação.

Ao meu orientador, Vidal, por ter me acolhido em seu grupo de pesquisa, sem me conhecer, e por ter tornado esse projeto realidade. Obrigada pelo incentivo, pela motivação, pelas conversas ao final do dia, pela paciência em ensinar, ainda que em férias, uma matemática a pensar geologicamente. Certamente foi um grande desafio. Obrigada por estar presente em praticamente todos os dias de trabalho, é um privilégio tê-lo tido como meu orientador.

Obrigada aos meus colegas de trabalho: Rodrigo Drummond, Alexandre Cruz e Bruno Honório, e às minhas amigas geólogas: Ana Carolina Stevanato e Aline Belila – pela paciência e bom humor em me ensinar, pelas conversas engraçadas, pelos conselhos, pela instrução e por tantas conversas na cozinha.

Muito obrigada à Ancilla Carvalho, minha companheira de viagens. Obrigada pelos passos de salsa – ainda que já não os saiba mais, pela paciência em me ouvir, por me guiar nas cidades desconhecidas, e por ter se tornado uma grande amiga.

À Juliana Bueno e Paola Faccini pela companhia, pelo carinho, pelos almoços e desabafos.

Agradeço aos pesquisadores prof. Dr. Alvaro Penteado Crósta, que participou da banca de qualificação, prof. Dr. Jurandir Zullo Júnior, que fez parte das bancas de qualificação e defesa e ao prof. Dr. Marcelo Monteiro da Rocha, pelo auxílio no desenvolvimento da pesquisa, com críticas e sugestões construtivas, bem acolhidas.

Às meninas da secretaria de pós-graduação, em especial, Val e Gorete, sempre muito atenciosas, prestativas e eficientes!

Obrigada à minha família: meus pais, Oscar Kuroda e Teresa C. Kuroda, à minha irmã, Marcela e aos meus avós, Mititoshi e Tuiyuko (*in memorian*) e Deoclécio e Eurides; pelo

companheirismo, pelo amor, paciência, bondade e compreensão. Acima de tudo, pelos ensinamentos do Senhor, pela dedicação e carinho. A partir do exemplo de vocês, construí o meu caráter, aprendi a amar, a respeitar e ajudar ao próximo.

Em especial, obrigada à minha mãe, com quem tenho um estreito laço de amizade e amor, que torna meus dias muito melhores. Obrigada pela vida de dedicação exclusiva à nossa família. Pelo incentivo diário, por todas as injeções de ânimo, pelas palavras de conforto, por sua alegria contagiante, pela beleza, riso fácil e bom humor.

Agradeço aos amigos Flávio Ribeiro, Francisco Fransa, Fabio Ayres, Juliane Paixão, Daniele Costa, Ricardo Morimoto, Kênia Pereira, Vanessa Araújo, Renata Novais, Maiara Assis, André Felipe, Paulo Hino, Lucas Loureiro, Cleyton Carneiro, Ivan Mingireanov Filho, Igor Leite, Lucas Seiti Ando e Leonardo Noshi; que de longe ou de perto serão sempre importantes para mim. A todos os amigos que fiz em Coimbra, Portugal, dos quais jamais esquecerei.

Obrigada aos meus tios, tias e primos.

E aquele que me deu a vida e me sustenta, que está comigo ainda que me sinta só, que é fiel e nunca me abandonará, à Deus, obrigada! Obrigada pela salvação, pela sabedoria, pelas conquistas e por ter colocado em minha vida pessoas tão especiais! Te louvo pelo teu amor!

BIOGRAFIA

Michelle Chaves Kuroda fez a graduação em Licenciatura em Matemática pela Universidade Estadual de Campinas (UNICAMP), de 2004 a 2007. Realizou duas iniciações científicas no Centro de Tecnologia e Informação de Campinas, sob orientação do professor Marcius Fabios Henriques de Carvalho, na área de Logística, com os temas: Análise Comparativa de Diferentes Formas de Gestão e Visibilidade da Informação em Cadeias de Suprimentos (2005-2006) e Importância do Estudo da Informação no Desempenho da Cadeia de Suprimento (2006-2007).

Nesse período foi *trainee* da empresa BSH Continental®, responsável pela construção de nova disposição de máquinas e simulação de cenários para otimização de tempo de produção.

A aluna concluiu seu curso de graduação como terceira melhor colocada e reingressou em Matemática Aplicada e Computacional, na mesma universidade, ocasião em que foi contemplada com bolsa de intercâmbio Santander para estudar em Portugal, no período de Setembro/2008 a Fevereiro/2009. Fez a disciplina Análise Real Avançada na Universidade de Coimbra, pré-requisito para ingresso no curso de Mestrado em Matemática Aplicada e Computacional na Unicamp, obtendo conceito A.

Em Março de 2010 ingressou no curso de Mestrado em Geociências, orientada pelo professor Alexandre Campanhe Vidal, que é coordenador do grupo de pesquisa vinculado à Rede de Caracterização e Modelagem de Reservatórios da Petrobras, onde trabalha como pesquisadora plena, com enfoque em aplicações de Redes Neurais Artificiais, desde Outubro de 2010.

SUMÁRIO

ÍNDICE DE FIGURAS.....	xiii
ÍNDICE DE TABELAS.....	xvii
RESUMO.....	xix
ABSTRACT.....	xxi
INTRODUÇÃO.....	1
CONCLUSÃO.....	133

CARACTERIZAÇÃO DE ELETROFÁCIES POR MEIO DE MAPAS AUTO-ORGANIZÁVEIS.....9

1.1. INTRODUÇÃO.....	13
1.2. CAMPO DE NAMORADO.....	14
1.3. MÉTODOS.....	17
1.3.1. SOM.....	17
1.3.2. K -MEANS.....	21
1.3.3. ANOVA.....	22
1.3.4. ANÁLISE DE DADOS.....	23
1.3.5. TREINAMENTO.....	26
1.4. RESULTADOS.....	27
1.5. CONCLUSÃO.....	38
AGRADECIMENTOS.....	38
REFERÊNCIAS.....	39

CLASSIFICATION AND CORRELATION OF SEDIMENTARY FACIES IN THE TAUBATÉ BASIN, SÃO PAULO STATE, BRAZIL.....45

2.1. INTRODUCTION.....	49
2.2. GEOLOGY.....	51
2.3. KOHONEN'S SELF-ORGANIZING MAPS.....	56
2.4. WORKING METHOD.....	57
2.5. RESULTS AND DISCUSSIONS.....	60
2.6. CONCLUSIONS.....	72

<i>REFERENCES</i>	74
INTERPRETATION OF SEISMIC MULTIATTRIBUTES USING A NEURAL NETWORK	79
3.1. <i>INTRODUCTION</i>	83
3.2. <i>FIELD DESCRIPTION</i>	85
3.2.1. <i>GEOLOGICAL SETTING</i>	85
3.2.2. <i>THE DATABASE</i>	88
3.3. <i>METHOD</i>	90
3.4. <i>RESULTS AND DISCUSSION</i>	92
3.5. <i>CONCLUSIONS</i>	96
<i>ACKNOWLEDGEMENTS</i>	96
<i>APPENDIX A. SELF-ORGANIZING MAPS</i>	96
<i>REFERENCES</i>	100
CARACTERIZAÇÃO DE RESERVATÓRIOS E PREDIÇÃO DE PERMEABILIDADE ATRAVÉS DE REDES NEURAS SUPERVISIONADAS	105
4.1. <i>INTRODUÇÃO</i>	109
4.2. <i>REDES NEURAS ARTIFICIAIS (RETROPROPAGAÇÃO DE ERROS)</i>	111
4.3. <i>CAMPO DE NAMORADO</i>	117
4.3.1 <i>CONJUNTO DE DADOS</i>	120
4.4. <i>MÉTODOS</i>	120
4.5. <i>RESULTADOS E DISCUSSÕES</i>	123
4.6. <i>CONCLUSÕES</i>	128
<i>REFERÊNCIAS</i>	129

ÍNDICE DAS FIGURAS

ANEXO 1.”Kuroda, M.C.; Vidal, A.C.; Leite, E.P.; Drummond, R.D. Caracterização de eletrofácies por meio de Mapas Auto-Organizáveis (aceito pela Revista Brasileira de Geofísica).”

Figura 1. 1: Localização da Bacia de Campos com os principais campos petrolíferos (adaptado de Milani <i>et al.</i> 2000).....	16
Figura 1. 2: Duas configurações de grid e níveis de vizinhança do BMU vencedor aos circunvizinhos. (A) representa o grid hexagonal e (B) o grid retangular. Adaptado de Costa (1999).	17
Figura 1. 3: Esquema da metodologia de trabalho.	25
Figura 1. 4: Planos de componentes de cada variável analisada.	28
Figura 1. 5: Imagem gerada pela ferramenta de Vesanto <i>et al.</i> (1999). (A) Escolha de número de clusters do método de agrupamento k-means, (B) U-matriz reagrupada com o método mencionado.	28
Figura 1. 6: Boxplots das litologias analisadas. 1 – Arenito Médio Gradado ou Maciço, 2 – Arenito/Folhelho Finamente Interestratificado, 3 – Arenito/Folhelho Interestratificado, 4 – Arenito Grosso Amalgamado, 5 – Conglomerado Residual, 6 – Arenito Médio Fino Laminado.....	34
Figura 1. 7: Resultado de classificação litológica dos poços testemunhados.	37

ANEXO 2. “Carvalho, A.M.A; Kuroda, M.C.; Vidal, A.C. Classification and Correlation of Sedimentary Facies in the Taubaté Basin, São Paulo State, Brazil (submetido aos Anais da Academia Brasileira de Ciências).”

Figure 2. 1: Location of the study area. In detail the Taubaté Basin in São Paulo State.	51
Figure 2. 2: Schematic geological map for the Taubaté Basin. (1) basement rocks; (2) Resende Formation; (3) Tremembé Formation; (4) São Paulo Formation; (5) Pindamonhangaba Formation; (6) Quaternary sediments (Modified after Riccomini, 1989).	53
Figure 2. 3: Location of the study area. In detail the Taubaté Basin in São Paulo State.	58
Figure 2. 4: Location of correlated wells and interpreted seismic profiles.	60

Figure 2. 5: Example of classification in reservoir (R – yellow) and non-reservoir (NR – green) facies for well 3ca24.	61
Figure 2. 6: Box-plots of the variables GR, ILD and DT for the classes established for wells 4ca25, 43cpq1,167 and 109sj160. The facies 1 is the reservoir and 2 is the no reservoir.	63
Figure 2. 7: Correlation of the reservoir facies in wells close to Caçapava city.	67
Figure 2. 8: Geophysical log for well 167, exemplifying the standard response of the unit C RG and ILD variables for the analyzed wells.	68
Figure 2. 9: Correlation of the reservoir facies and distribution of the sedimentary units interpreted in the seismic section A and in the electric logs.	69
Figure 2. 10: Correlation of the reservoir facies and distribution of the sedimentary units interpreted in seismic section B and in the electric logs.	70
Figure 2. 11: Representation of the sediment transport direction in the São José dos Campos Compartment based on the analysis of wells and seismic profiles	72

ANEXO 3. “Kuroda, M. C.; Carvalho, A.M.A; Vidal, A.C.; Leite E.P.; Drummond R.D. Interpretation of Seismic Multiattributes Using a Neural Network (submetido ao Journal of Applied Geophysics).”

Figure 3. 1: Location of seismic sections in the Taubaté Basin.	86
Figure 3. 2: Stratigraphic units of the Taubaté Basin. Modified from Fernandes (2003).	87
Figure 3. 3: Seismic Line "C". Picture taken from OpendTect ® software.	91
Figure 3. 4: Transformation matrices of seismic attributes for set 1. Adapted from Moraes (2004).	92
Figure 3. 5: Set 1 classified into 3 (A), 4 (B) and 5 (C) clusters.	93
Figure 3. 6: Set 2 classified into 3 (A), 4 (B) and 5 (C) clusters.	94

ANEXO 4. “Kuroda, M. C.; Vidal, A.C. Caracterização de Reservatórios e Predição de Permeabilidade Através de Redes Neurais Supervisionadas (para submissão)”

Figura 4. 1: Arquitetura da rede neural Back Propagation. O número de neurônios e camadas pode ser diferente.	111
--	-----

Figura 4. 2: Diagrama de um neurônio. Fonte: Onerous, 2002.....	112
Figura 4. 3: Localização da Bacia de Campos com os principais campos petrolíferos. Adaptado de Milani <i>et al.</i> (2000).....	119
Figura 4. 4: Fluxograma da classificação automática dos dados com Retropropagação de Erros.	121
Figura 4. 5: Disposição dos poços no campo e sua divisão em 3 blocos, enumerados da esquerda para a direita. Em vermelho os poços com descrição de valores de litologia e permeabilidade, calculada em laboratório.....	122
Figura 4. 6: Matriz de confusão para o treinamento de litofácies para os blocos 1, 2 e 3, respectivamente. As classes de entrada são referentes aos dados de testemunhos e as classes de saída referem-se à classificação encontrada com o algoritmo utilizado.....	125
Figura 4. 7: Resultados de classificação de litologia e permeabilidade para o poço NA02, com os perfis GR, RHOB, DT e NPHI. Os pontos em vermelho do gráfico de permeabilidade são referentes aos valores em logaritmo calculados em laboratório e a linha contínua representa a permeabilidade para o intervalo classificado como reservatório.	126
Figura 4. 8: Os valores de permeabilidade preditos com o método de redes neurais e em vermelho os dados de permeabilidade adquiridos a partir de amostras de testemunho.	127
Figura 4. 9: Os valores de permeabilidade estimados versus a permeabilidade calculada a partir do testemunho.	128

ÍNDICE DAS TABELAS

ANEXO 1.”Kuroda, M.C.; Vidal, A.C.; Leite, E.P.; Drummond, R.D. Caracterização de eletrofácies por meio de Mapas Auto-Organizáveis (submetido à Revista Brasileira de Geofísica).”

Tabela 1. 1: Critérios utilizados para rejeição da hipótese H0. Fonte: Shimakura (2008).....	23
Tabela 1. 2: Descrição das fácies do Campo de Namorado.	24
Tabela 1. 3: Estatística básica dos valores de densidade (RHOB), tempo de trânsito (DT), porosidade neutrônica (NPHI) e raios gama (GR) para os quatro grupos.	29
Tabela 1. 4: Classificação dos dados supervisionados, em porcentagem.	30
Tabela 1. 5: Classificação litológica dos dados supervisionados.....	31
Tabela 1. 6: P-valores.....	32
Tabela 1. 7: Intervalos de confiança.....	33

ANEXO 2. “Carvalho, A.M.A; Kuroda, M.C.; Vidal, A.C. Classification and Correlation of Sedimentary Facies in the Taubaté Basin, São Paulo State, Brazil (submetido aos Anais da Academia Brasileira de Ciências).”

Table 2. 1: Lithofacies identified by Riccomini (1989) in the Taubaté Basin.	65
---	----

ANEXO 4. “Kuroda, M. C.; Vidal, A.C. Caracterização de Reservatórios e Predição de Permeabilidade Através de Redes Neurais Supervisionadas (para submissão)”

Tabela 4. 1: Principais resultados encontrados na classificação de litotipos. O número de neurônios na tabela representa a melhor estrutura da rede Retropropagação de Erros encontrada e a porcentagem de acerto calculada a partir da comparação de classes.	124
---	-----



UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE GEOCIÊNCIAS
PÓS-GRADUAÇÃO EM GEOCIÊNCIAS
ÁREA DE GEOLOGIA E RECURSOS NATURAIS

REDES NEURAI APLICADAS AO ESTUDO DE ROCHAS RESERVATÓRIO

RESUMO

Dissertação de Mestrado

Michelle Chaves Kuroda

A caracterização de reservatórios é um trabalho complexo, que envolve muitas variáveis com informações em diferentes escalas. Para minimizar incertezas, este trabalho propõe a utilização de redes neurais artificiais, algoritmo computacional inspirado no funcionamento cerebral que mapeia, agrupa e prevê informações a partir de reconhecimento de padrões supervisionados ou não. Neste trabalho foram aplicados dois métodos: Mapas Auto-Organizáveis e *Backpropagation*. O objetivo da aplicação da ferramenta é o melhor entendimento dos reservatórios estudados, a partir da identificação litológica e previsão de características petrofísicas em dados de poços e a melhoria de visualização sísmica realizada a partir do estudo de multiatributos sísmicos. Através dos resultados é possível delimitar a geometria dos reservatórios possibilitando ajustes e tomadas de decisões que aperfeiçoam o processo de exploração. Com este propósito foram analisados duas áreas de estudo: a bacia de Taubaté localizada na porção leste do estado de São Paulo, que ocupa uma área aproximada de 2400 km², estendendo-se ao longo do Vale do Rio Paraíba, desde Jacareí até a cidade de Cruzeiro; e o campo de Namorado, na bacia de Campos, localizado na parte central norte na zona de acumulação de hidrocarbonetos da Bacia de Campos, a 80 km da costa. Na bacia de Taubaté foram realizados estudos de caracterização de eletrofácies em dados de perfis de poços, posteriormente utilizados no ajuste da análise sísmica das linhas 2D da mesma bacia. No campo de Namorado foram estudadas as fácies reservatório e possíveis reservatórios para identificação e classificação de qualidade, além da predição de permeabilidade nos intervalos de reservatório.

Palavras chaves: Mapas Auto-Organizáveis, Retropropagação de Erros, Caracterização de Reservatórios.



UNIVERSITY OF CAMPINAS
INSTITUTE OF GEOSCIENCE

NEURAL NETWORKS APPLIED TO THE STUDY OF RESERVOIR ROCKS

ABSTRACT

Master Degree

Michelle Chaves Kuroda

The reservoir characterization is complex, involving many variables with information on different scales. To minimize uncertainties, this paper proposes the use of artificial neural networks, computational algorithm inspired on the brain function, which maps, groups, and provides information based on supervised pattern recognition or not. In this work we applied two methods: Self-Organizing Maps, and Backpropagation. The purpose of the application of the tool is a better understanding of the reservoirs, by identifying lithologic and predicting of petrophysical characteristics on data from well logs and improving seismic preview done from the multi-attribute seismic study. Through the results it is possible to define the boundaries of reservoirs, allowing adjustments and making decisions that enhance the exploration process. For this purpose we analyzed two study areas: the Taubaté basin located in the southeastern portion of the São Paulo state, encompasses an area of approximately 2400 km², stretching along the Paraíba River Valley, from Jacareí to the Cruzeiro city; and Namorado field in the Campos Basin, located in the central zone of hydrocarbon accumulation in the Campos Basin, 80 km from the coast. In Taubaté basin the studies have been performed on well log characterization, they were used to adjust the seismic analysis of 2D lines in the same basin, through multi-attributes analysis. In Namorado field the facies of reservoir and possible reservoir were studied for identify e classify the rock types, besides to predict the permeability in reservoir intervals.

Key words: Self-Organizing Maps, Backpropagation, Reservoir Characterization.

INTRODUÇÃO

A caracterização de reservatórios é de fundamental importância para o melhor entendimento das heterogeneidades presentes em um campo petrolífero. A análise geológica (modelos deposicionais, diagenéticos e estruturais), a engenharia de reservatórios (modelos de fluxo), os métodos estatísticos (clássicos e geoestatística), os métodos de interpretação de dados sísmicos (modelos elásticos), métodos numéricos e de computação gráfica constituem metodologias importantes que ao serem combinadas permitem a melhor compreensão dos reservatórios e suas heterogeneidades (Baldissera, 1992).

Neste contexto, a aplicação de métodos matemáticos em geociências tem como objetivo a redução de incertezas e obtenção de resultados mais confiáveis que minimizem os riscos, os custos da exploração e o tempo gasto na interpretação das informações.

Atualmente, muitos estudos demonstram o bom desempenho de técnicas de análise estatística multivariada no processamento e interpretação de dados geofísicos. Nesses trabalhos, a construção de modelos de reservatórios de hidrocarbonetos é desenvolvida com base em estimativas das propriedades de reservatório, como litologia, porosidade, permeabilidade, tipo do fluido e na atenuação de ruídos em dados sísmicos (Alhoniemi *et al.*, 1999; Poulton, 2001; Bhatt & Helle, 2002; Stundner & Oberwinkler, 2004; Singh, 2005; Chopra & Marfurt, 2006; Madji *et al.*, 2010).

Nesta área, o método de análise multivariada que mais recebe destaque são as Redes Neurais Artificiais (RNAs) (Haykin, 1999; Poulton, 2001; Madji *et al.*, 2010). As características destes algoritmos que permitem sua evidência são: o estabelecimento de relações não-lineares entre as variáveis envolvidas, a alta adaptação ao conjunto de dados de entrada e a não exigência do conhecimento prévio das distribuições estatísticas do conjunto analisado, que elimina possíveis condições de existência do conjunto de dados analisado para a aplicação do método. Além disso, pode ser utilizado em dados incompletos e com presença de ruídos (Haykin, 1999; Kohonen, 2001; Stunder & Oberwinkler, 2004).

As redes neurais surgiram na década de 1940, mas os indícios da utilização de ideias de funcionamento do método podem ser diagnosticados há séculos atrás, com os trabalhos do psicólogo William James, datados de 1890, nos quais indaga sobre os princípios da memorização

e de estruturas fundamentalmente simples. Apesar de não apresentar abordagem matemática, tal linha de pesquisa foi precursora de diversos estudos e diferentes enfoques sobre o tema, tratado de forma menos abstrata.

Os primeiros artigos sobre redes neurais são atribuídos a McCulloch e Pitts, em 1943, com a criação de uma máquina inspirada no sistema nervoso. As redes neurais não foram tema de muitas pesquisas até 1949. Neste ano, o interesse aumentou a partir da publicação do livro “The Organization of Behavior”, escrito por Donald Hebb. Neste trabalho, o autor propõe uma lei para aprendizagem, que serviu de inspiração para muitas outras publicações na década de 1950.

Em 1951, Marvin Minsky desenvolveu Snark, o primeiro neuro computador, que apesar de nunca ter realizado grandes processamentos, foi utilizado como base para desenvolvimento de outras idéias.

Depois deste, o Mark I Perceptron foi o primeiro neuro computador que obteve destaque. Criado por um grupo liderado por Rosenblatt e Wightman, entre 1957 e 1958, com o objetivo de reconhecer padrões, o método é reconhecido por muitos como pioneiro, pela inovação de funcionamento.

Entretanto, depois dessa publicação uma grande quantidade de estudos foi realizada com muito entusiasmo, mas sem grandes contribuições para a área de pesquisa, tornando o método pouco confiável. Então, no período de 1967 a 1982 poucos foram os pesquisadores que se dedicaram a estudar a respeito.

Nos anos 80, contrário à corrente tradicional de classificação de dados, a DARPA (Defence Advanced Research Projects Agency) iniciou a análise com neuro computação, sendo a precursora que estimulou o renascimento da técnica.

Neste período, receberam destaque os pesquisadores John Hopfield, David Rumerlhart e James McClelland. A partir de 1987 muitas conferências de redes neurais surgiram, tornando esta técnica popular até os dias atuais.

As redes neurais se consolidaram nas últimas décadas com diversas aplicações em inúmeras áreas, seja em controle de processos ou qualidade, manufatura, design de produtos, análise financeira, detecção de fraudes, aprovações de crédito, reconhecimento de voz e face e

análise de dados. A adaptação da ferramenta possibilitou seu maior desenvolvimento e popularização. Poulton (2001) elucida diferentes utilizações de redes neurais, como no uso de softwares antivírus, capazes de reconhecer padrões de algoritmos referentes a vírus, ou em sites de busca.

Diferentes de técnicas como PCA (Análise de Componentes Principais) ou ICA (Análise de Componentes Independentes), baseadas em funções de otimização, as redes neurais artificiais (RNA) têm funcionamento baseado no sistema nervoso humano. Formadas por unidades computacionais, denominadas neurônios, as redes são “treinadas” segundo associação de memória, que reconhecem padrões e os associa a diferentes classes. Poulton (2001) exemplifica o processo de reconhecimento do algoritmo com a identificação de pessoas, animais e figuras. A cada etapa ocorre a associação de características chaves para identificar as propriedades designadas.

Em caracterização de reservatórios a aplicação de RNAs é realizada em seções sísmicas e análise de perfis geofísicos. A interpretação das linhas sísmicas tem como fundamento a análise da geometria das maiores estruturas, através da identificação dos refletores e definição do padrão de empilhamento dos pacotes sedimentares, delimitando corpos potenciais de rochas-reservatório, evidenciados com o uso de diversos atributos (traços derivados de cálculos da sísmica original). Os dados de poços permitem a correlação lateral das fácies, a determinação da geometria dos depósitos e o conhecimento das propriedades petrofísicas (porosidade e permeabilidade) das fácies reservatórios.

Nesta pesquisa tais conjuntos de dados são utilizados para a caracterização de reservatórios fluviais da Bacia de Taubaté e marítimos do Campo de Namorado, situado na Bacia de Campos.

A Bacia de Campos apresenta área de aproximadamente 100.000 km² e está localizada no litoral do Estado do Rio de Janeiro e Espírito Santo, entre o Alto de Vitória ao norte e Alto de Cabo Frio ao sul. Com mais de 1.600 poços perfurados desde a década de 70, a bacia possui mais de 90% das reservas petrolíferas brasileiras atuais, sendo o campo de Namorado um de seus reservatórios mais expressivos (Winter *et al.*, 2007).

Para análise desse campo foram utilizados dados de poço. Com eles foram encontradas as fácies de rochas reservatório, através do método k-nn (*k-nearest neighbors*), que realiza o

reconhecimento de padrões a partir da distribuição espacial dos dados. Depois de encontradas as litofácies, as rochas reservatório foram novamente analisadas, mas dessa vez com o algoritmo de Redes Neurais Artificiais Mapas Auto-Organizáveis, para classificação de litotipos que caracterizassem a qualidade dos reservatórios. Este estudo é apresentado no Anexo 1.

Posteriormente foi analisado o comportamento petrofísico do Campo de Namorado, utilizando-se a metodologia descrita por Stundner & Oberwinkler (2004) e Cozzi *et al.* (2006), que primeiramente encontram as fácies reservatório e posteriormente utilizam esse conjunto de dados para previsão de permeabilidade, a partir da utilização de métodos de Redes Neurais Artificiais supervisionadas de Retropropagação de erros. Com os resultados obtidos foi possível realizar a distribuição dessa variável no campo. Este estudo está presente no Anexo 4.

A segunda área de estudo é referente ao sistema Aquífero Taubaté, representado pelos sedimentos terciários da Bacia de Taubaté, corresponde a uma importante fonte de exploração de água do Estado de São Paulo. A bacia está situada no principal eixo de ligação entre os centros urbanos de São Paulo e Rio de Janeiro, e apresenta grande número de poços tubulares perfurados para captação de água subterrânea, além de 11 linhas sísmicas 2D.

Para esta bacia foi criado um modelo de fácies sedimentares através do método de Redes Neurais Artificiais Mapas Auto-Organizáveis. Uma vez obtida a classificação, a correlação entre poços e seções sísmicas permitiram o maior entendimento do sistema deposicional e da distribuição de fácies reservatório no Compartimento São José dos Campos. Esse estudo é apresentado no Anexo 2.

O dado sísmico da Bacia de Taubaté foi analisado posteriormente, com o mesmo método, através do uso de diferentes atributos, para realçar a continuidade lateral e as unidades sísmicas destacadas por Fernandes (1993). Este estudo está presente no Anexo 3.

O presente trabalho apresenta capítulos em forma de artigos que foram dispostos em ordem cronológica, por isso podem ocorrer repetições de itens relacionados. Todos os artigos foram submetidos a revistas da área de Geociências.

REFERÊNCIAS

- Alhoniemi E., Hollmén J., Simula O., Vesanto J. 1999. Process Monitoring and Modeling using the Self-Organizing Map. Integrated Computer Aided Engineering. Disponível em: <<http://citeseerx.ist.psu.edu>>. Acesso em: 17 janeiro 2010.
- Bhatt & Helle B. 2002. Determination of facies from well logs using modular neural networks. Petroleum Geoscience, 8:217-228.
- Baldissera, P.R. 1992. Modelagem Estocástica de Fácies para Caracterização da Distribuição Espacial das Heterogeneidades de Reservatório. Dissertação de Mestrado. Instituto de Geociências. Universidade Estadual de Campinas, Campinas, 200p.
- Chopra S. & Marfurt K, 2006. Seismic Attributes – a promising aid for geologic prediction. CSEG Recorder, Special Edition, 111-121p.
- Cozzi M., Ruvo L., Scaglioni P., Lyne A.M. 2006. Core data preprocessing to improve permeability log estimation. SPE Annual Technical Conference and Exhibition being held in San Antonio.
- Fernandes F.L. 1993. Arcabouço estrutural e evolução da Bacia de Taubaté - SP. Dissertação de Mestrado, Universidade Federal de Ouro Preto, Ouro Preto, MG.
- Haykin S. 1999. Neural Networks(2) Prentice Hall, Virginia, USA, 842 pp.
- Kohonen T. 2001. Self-Organizing Maps. Springer-Verlag, New York, 501pp.
- Majdi A., Beiki M., Pirayehgar A., Hosseinyar G. 2010. Identification of Well logs with significant impact on prediction of oil and gas reservoirs permeability using statistical analysis of RSE values. Journal of Petroleum Science and Engineering. 75: 91-99.
- Poulton M.M. 2001. Multi-Layer Perceptrons and Back-Propagation Learning. Handbook of Geophysical Exploration, Seismic Exploration. Volume 30. Computational Neural Networks for Geophysical Data Processing. Pergamon, Tucson, USA. 27-53pp.
- Singh S. 2005. Permeability Prediction Using Artificial Neural Network (ANN): A Case of Uinta Basin. SPE Annual Technical Conference and Exhibition being held in Dallas. 9-12.

Stundner M. & Oberwinker C. 2004. Self-Organizing Maps for Lithofacies Identification and Permeability Prediction. In: SPE Annual Technical Conference and Exhibition held in Houston, p.8.

Winter W.R., Jahnert R.J., França A.B. 2007. Bacia de Campos. B.Geoci. Petrobras, Rio de Janeiro, v.15, p.511-529.

ANEXO 1

“Kuroda, M. C.; Vidal, A.C.; Leite E.P.; Drummond R.D.
Caracterização de eletrofácies por meio de Mapas Auto-
Organizáveis (aceito pela Revista Brasileira de Geofísica).”

CARACTERIZAÇÃO DE ELETROFÁCIES POR MEIO DE MAPAS AUTO-ORGANIZÁVEIS

Michelle Chaves KURODA*, Alexandre Campana VIDAL, Emilson Pereira LEITE, Rodrigo
Duarte DRUMMOND

Departamento de Geologia e Recursos Naturais, Universidade Estadual de Campinas
(UNICAMP). R. Pandiá Calógeras, 51. CEP: 13083-970. Campinas, SP. Caixa Postal: 6152.

Endereços eletrônicos: mckuroda@ige.unicamp.br; vidal@ige.unicamp.br;
emilson@ige.unicamp.br; rdrummond@gmail.com

RESUMO – A caracterização de eletrofácies é essencial para a modelagem de reservatórios. Todavia, esse é um processo dependente de muitas variáveis, com erros e ruídos associados que dificultam a interpretação visual das mesmas. Para minimizar incertezas, este trabalho propõe a utilização da rede neural artificial Mapas Auto-Organizáveis, um algoritmo computacional inspirado no funcionamento cerebral que mapeia e agrupa informações semelhantes. Os dados petrofísicos utilizados são referentes ao Campo de Namorado, da Bacia de Campos, RJ, dos quais foram estudados os perfis porosidade neutrônica, raio gama, densidade e sônico, com o objetivo de classificar a litologia dos reservatórios do campo. Através dos resultados foi possível delimitar a geometria dos reservatórios e detalhar suas características de forma mais precisa.

Palavras-Chave: Redes Neurais, Mapas Auto-Organizáveis, Reconhecimento de Litofácies.

ABSTRACT – The reservoir characterization is a complex process that involves many variables with information at different scales. To minimize these uncertainties, this work proposes the use of a mathematical tool for automatic clustering, based on unsupervised pattern recognition. This is an artificial intelligence method, called Self Organizing Maps, a computational algorithm inspired by the brain that maps and groups similar information. The characteristics of this method that highlights of the others are its capacity of decision-making, no-linearity of the information clustering and sensitivity to environmental interference. Through the results is possible define the geometry of the reservoirs and determine the petrophysical properties of the same, allowing adjustments and making decisions that enhance the exploration process. Thus we can find shells in a faster and more accurate.

Keywords: Neural Networks, Self-Organizing Maps, Recognition of Lithofacies.

1.1. INTRODUÇÃO

Os modelos geológicos de reservatórios de hidrocarbonetos são desenvolvidos com base em estimativas das propriedades de reservatório, como litologia, porosidade, permeabilidade e tipo do fluido (e.g. Alhoniemi *et al.*, 1999; Bhatt & Helle, 2002; Matos *et al.*, 2004; Stundner & Oberwinkler, 2004). A aplicação de métodos matemáticos robustos tem como objetivo a redução de incertezas na geração de tais modelos, para obtenção de resultados mais confiáveis que minimizem os riscos e os custos da exploração.

Dentre estes, as tecnologias de inteligência artificial são os métodos de maior aplicação em geociências, com destaque às Redes Neurais Artificiais (RNAs) (Haykin, 1999). As RNAs possuem a capacidade de modelar relações não-lineares entre as variáveis envolvidas, são altamente adaptáveis ao conjunto de dados de entrada e não exigem o conhecimento prévio das distribuições estatísticas desses dados. Estas características fundamentais situaram as RNAs entre os métodos de classificação e agrupamento mais importantes da última década (Haykin, 1999; Stunder & Oberwinkler, 2004), permitindo grande aplicação na área de geologia de reservatórios, que envolve heterogeneidades geológicas em diversas escalas e o tratamento de um conjunto de dados obtido a partir de várias fontes (Kohonen, 2001; Bhatt & Helle, 2002; Stundner & Oberwinkler, 2004; Esposito *et al.*, 2008).

Como exemplos de aplicações no estudo de reservatórios, Stundner & Oberwinkler (2004) citam três tipos básicos de utilização de Redes Neurais com base em dados de poços: predição de permeabilidade e/ou porosidade; geração de poços sintéticos e determinação de eletrofácies.

Este trabalho propõe a utilização de Redes Neurais para análise de eletrofácies com a finalidade de identificar litologias a partir de um conjunto de perfis geofísicos do reservatório de petróleo do Campo de Namorado, na Bacia de Campos. A escolha para o estudo da litologia deve-se à sua importância na construção do modelo geológico de reservatórios, cujas principais fontes de informação são os dados de poços (Matos *et al.*, 2004).

O método principal aplicado neste trabalho é uma RNA não-supervisionada, no qual as informações *a priori* sobre a litologia não são utilizadas para classificação das eletrofácies (Johann, 2004). Para este tipo de amostra, recebe destaque a RNA chamada de Mapas Auto-Organizáveis de Kohonen (*Self-Organizing Maps* - SOM) (Kohonen, 2001; Coléou *et al.*, 2003).

Coléou *et al.* (2003) defendem o uso de SOM por apresentar resultados mais precisos em detrimento de outras técnicas estatísticas multivariadas, tais como PCA, ICA e K-means, utilizando critérios de comparação como redundância, ruído e continuidade. Segundo Stundner & Oberwinkler (2004), SOM ainda possibilita a mescla de métodos de previsão estatística para obtenção de informações de poços, combinando visualização de dados multivariados e capacidade de agrupamento.

O propósito da pesquisa é o detalhamento das rochas consideradas como reservatório pelo algoritmo de agrupamento *k-NN* (*k nearest neighbor*) dos dados de perfis de poços, do Campo de Namorado, no qual são descartadas as outras fácies identificadas pelo método. Desta forma é possível identificar assinaturas litológicas das fácies reservatório e entender o comportamento dos reservatórios da área estudada.

Na maioria dos casos, os métodos tradicionais são eficientes na separação de fácies claramente distintas, como reservatório e não reservatório. Porém, o detalhamento de uma única fácies apresenta como limitação a pequena variação de valores de perfis e escassez de informação de testemunhos, que dificultam a aplicação de um método supervisionado. Além disso, as medidas petrofísicas podem estar vinculadas às mudanças de propriedades físicas e não às variações litológicas. Entretanto a aplicação do SOM possibilita a identificação de grupos que favorecem a melhor caracterização das rochas reservatório.

1.2. CAMPO DE NAMORADO

Descoberto em 1975, o campo gigante de Namorado está localizado na parte central norte na zona de acumulação de hidrocarbonetos da Bacia de Campos, a 80 km da costa (Fig. 1.1). O reservatório é limitado por fechamentos estratigráficos e estruturais, que delimitam o campo à área aproximada de 21 km². Em profundidade, o reservatório situa-se entre -2900 e -3400 m, sendo que os limites de topo e base, claramente reconhecidos por perfis geofísicos, totalizam espessuras que variam entre 90 m e 180m (Johann, 1997; Souza Jr, 1997)

O reservatório foi depositado em um complexo de canais, junto a sistemas de frentes deltaicas (Barboza, 2005), denominados de arenitos turbidíticos de Namorado. Dentre as 21 fácies descritas na análise de testemunho, as principais rochas reservatórios são referentes a duas

fácies de arenitos, que apresentam espessuras métricas nos testemunhos. A fácies de maior ocorrência nos testemunhos corresponde a arenito médio maciço, arcoseano e bem selecionado.

A segunda fácies são arenitos grossos amalgamados, com gradação da fração areia grossa-conglomerática na base para fração grossa no topo. As porosidades e permeabilidades médias são respectivamente 26% e 400mD (Bacoccoli *et al.*, 1980). As fácies arenitos e folhelhos interestratificados são consideradas como reservatórios secundários.

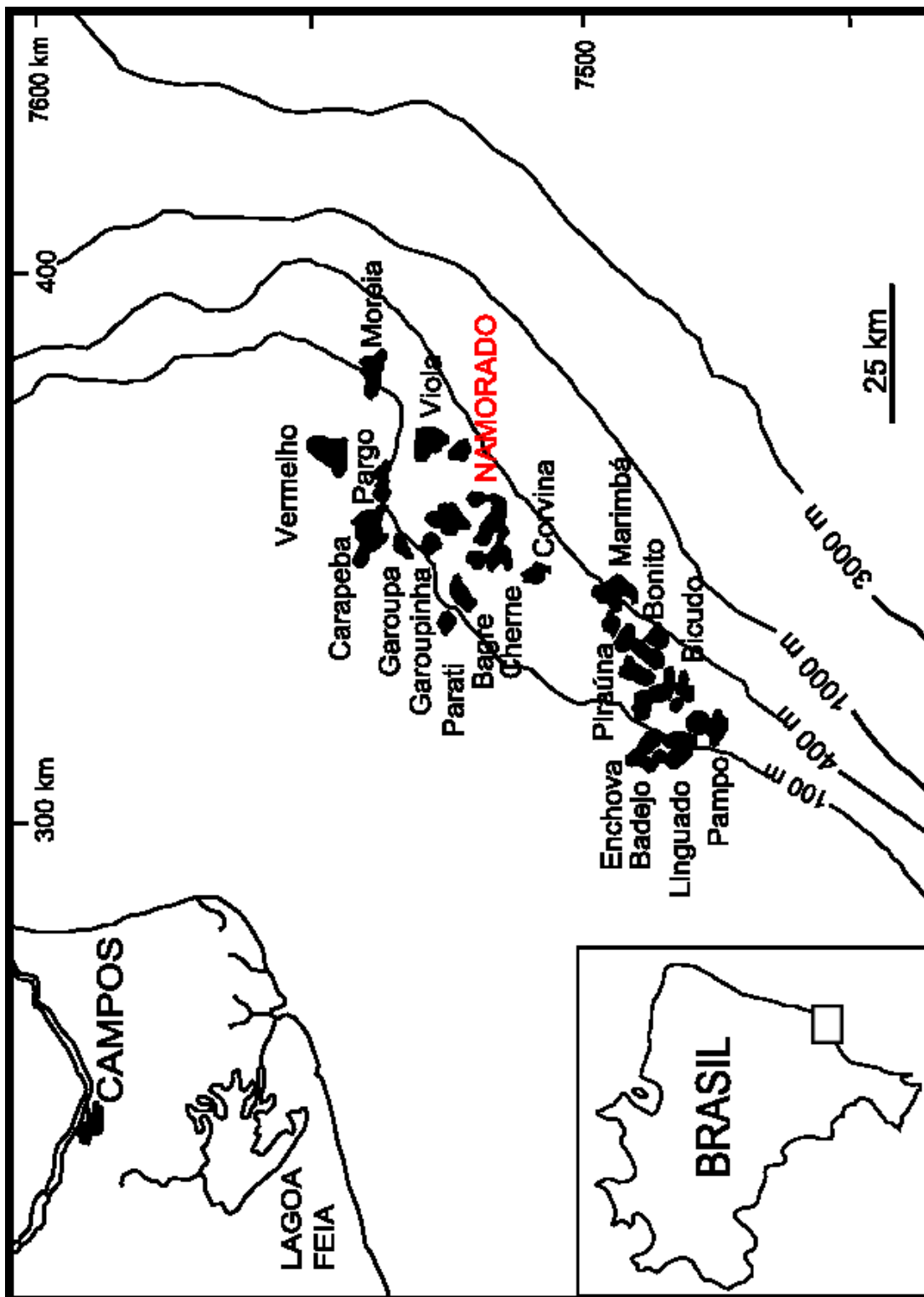


Figura 1. 1: Localização da Bacia de Campos com os principais campos petrolíferos (adaptado de Milani *et al.* 2000).

1.3. MÉTODOS

1.3.1. SOM

O método SOM, desenvolvido por Kohonen no final da década de 70, é uma RNA para amostras não supervisionadas que se baseia no processo de aprendizagem competitiva, na qual somente uma unidade computacional (também chamada de BMU – *Best Matching Unit*) de saída (neurônios) ou grupo local de unidades computacionais fornece uma resposta ativa a um sinal de entrada corrente (Gonçalves *et al.*, 1996).

O objetivo do algoritmo é mapear o conjunto de dados de entrada multidimensionais buscando similaridade em um espaço de dimensão inferior, 2-D, chamado de mapa auto-organizável. As unidades do SOM são conectadas às suas vizinhas, formando mapas com estruturas, normalmente hexagonais ou retangulares, que determinam sua vizinhança, como mostra a Fig. 1.2, no qual cada nó representa um neurônio.

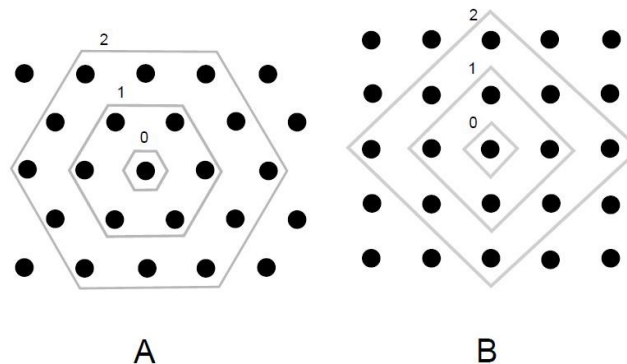


Figura 1. 2: Duas configurações de grid e níveis de vizinhança do BMU vencedor aos circunvizinhos. (A) representa o grid hexagonal e (B) o grid retangular. Adaptado de Costa (1999).

Com isso, SOM pode ser interpretado como uma forma de preservação de relações métricas e topológicas do espaço de entrada em uma rede de duas dimensões, usada como uma ferramenta de visualização, que exhibe diferentes características dos dados, como sua possível estruturação em grupos (Matos *et al.*, 2004; Leite & Filho, 2010).

Com esta lógica, dados próximos em um espaço de entrada terão representações próximas no espaço de saída e assim serão classificados como pertencentes ao mesmo *cluster* (grupo) ou em *clusters* próximos.

A estrutura do algoritmo é constituída por duas camadas: a camada de entrada e a camada de saída, ou de Kohonen (U). Na camada de entrada, cada amostra do poço é um vetor-linha dessa matriz de dados (sendo cada variável considerada uma dimensão no conjunto) representado pelo vetor $x = [x_1, x_2, \dots, x_n] \in R^n$, que será associado a uma única unidade que representa o centro do seu conjunto, ou seja, a BMU será sempre o ponto central, ou centróide do conjunto ao qual pertence o ponto analisado da matriz.

Cada unidade i da camada de Kohonen possui um vetor de peso associado: $m_i = [m_{i1}, m_{i2}, \dots, m_{in}]^T$. As BMUs nesta camada são conectadas às vizinhas pelo *grid* hexagonal, que considera mais unidades vizinhas (6-conectados) que o *grid* retangular (4-conectados), como mostra a Fig. 1.2.

Neste estudo, os pesos têm inicialmente valores aleatórios e ao decorrer do treinamento os valores são adaptados, convergindo para valores mais adequados independentemente de seu valor inicial (Stundner & Oberwinkler, 2004). A função de ativação, que determina o BMU associado a cada vetor de entrada, utilizada nos testes é a Distância Euclidiana, dada pela seguinte equação:

$$d_i = \sum_{j=1}^N (x_j(t) - w_{ij}(t))^2, \quad (1.1)$$

no qual $x_j(t)$ é a j -ésima entrada em uma dada iteração e $w_{ij}(t)$ é o peso do BMU j da camada de entrada conectado ao neurônio i da camada de saída.

Como cada ponto vetorial da camada de entrada possui apenas um neurônio associado a ele, o algoritmo representa a regra através da seguinte equação, no qual a unidade escolhida (chamada de neurônio vencedor) é representada pelo índice c :

$$\|x - m_c\| = \min_i \{\|x - m_i\|\}, \quad (1.2)$$

no qual $\|\cdot\|$ é a métrica de distância, no caso analisado, Euclidiana.

Depois de encontrada a BMU cuja distância é a menor do ponto analisado, a unidade e suas vizinhas (determinadas pelo *grid* anteriormente escolhido) são atualizadas, ou seja, são movidas em direção ao vetor de entrada no espaço de saída.

Dessa forma a regra de atualização para a unidade i desloca o BMU vencedor em direção ao vetor x , de acordo com a função densidade de probabilidade, dada por:

$$\begin{cases} m_i = m_i(t) + \alpha(t)h_{bi}(t)[x - m_i(t)], & \text{se } m_i \in \text{vizinhança de } c \\ m_i = m_i, & \text{caso contrário} \end{cases}, \quad (1.3)$$

no qual t denota tempo, $\alpha(t)$ é a taxa de aprendizado, que varia de 0 a 1 (e decresce ao longo das iterações até que seu valor não seja mais significativo, ou seja, até que ocorra convergência) e $h_{bi}(t)$ representa o tamanho da vizinhança, descrito por:

$$h_{bi}(t) = e^{-\frac{\|r_b - r_i\|^2}{2\sigma^2(t)}}, \quad (1.4)$$

no qual r_b e r_i são as posições dos neurônios b e i no reticulado do SOM.

A intensidade de atração do vetor de entrada é controlada pela média de aprendizagem, e o número de vetores atraídos pelo BMU vencedor depende do raio da vizinhança.

Depois do treinamento, as BMUs representam o mapeamento dos dados de entrada no SOM e a aproximação deles com os vetores de entrada.

Como o principal objetivo deste trabalho é a identificação de agrupamentos de dados, serão utilizadas, para visualização do SOM, técnicas que enfatizem as correlações entre os pontos do mapa.

Para visualizar as similaridades entre os elementos do SOM é interessante observar as distâncias entre as BMUs, especificamente para o mapa da matriz unificada de distância, a Matriz-U, que é uma das formas de representação delas (Ultsch, 1992; 1993).

A Matriz-U permite a visualização da distância entre as BMUs vizinhas e a média entre elas. Dessa forma, a imagem releva possíveis agrupamentos formados pelos atributos

representados pelo SOM sugerindo, inclusive, o número de grupos existentes no dado através de análise visual.

Mas nem sempre essa análise é eficaz ou suficiente. Matos *et al.* (2004) revelam as dificuldades para escolha do número de *clusters*, que são realizadas normalmente por sistemas heurísticos pelos sistemas atualmente existentes.

Mediante essas dificuldades, o agrupamento semi-automático dos dados da Matriz-U é realizado através de um método tradicional, chamado K-means (Khedairia & Khadir, 2008).

O número de unidades no mapa (número de BMUs) é escolhido *a priori* e determina a eficácia e a capacidade de generalização do modelo, pois durante o treinamento, o SOM forma uma rede elástica que se adapta aos dados de entrada.

Para escolha do tamanho da Matriz-U, Vesanto & Alhoniemi (2000) sugerem a seguinte expressão:

$$m = 5 * \sqrt{n}, \quad (1.5)$$

no qual m é o tamanho da matriz e n é o número de amostras treinadas.

Todavia, a escolha do número de grupos pode ser pré-determinada pelo intérprete, por exemplo, utilizando um tamanho menor, caso o interesse seja a identificação de padrões predominantes, e um tamanho maior se o objetivo é detectar todos os padrões da imagem, inclusive aqueles com baixa probabilidade de ocorrência (Matos *et al.*, 2003; Gonçalves *et al.*, 2007).

Posteriormente, as BMUs podem ser agrupadas através do método K-means, que além de fornecer um melhor entendimento sobre a formação dos grupos, tem como vantagem a redução drástica do tempo computacional (Vesanto *et al.*, 1999).

1.3.2. K -MEANS

Um dos algoritmos mais populares para identificação de clusters (Khedairia & Khadir 2008), que agrupa em classes um conjunto de pontos de entrada sem informações *a priori*, é o método K-means.

Neste trabalho, o K-means foi utilizado para identificação de clusters da Matriz-U, após o treinamento com o SOM, sendo essa uma maneira eficiente de caracterização de grupos da matriz especificada (Rocha & Sousa, 2003).

As BMUs são classificadas por pertencer a k grupos, onde o valor de k é escolhido *a priori*, que serão divididas entre estes grupos, mediante o treinamento que minimiza a seguinte função de erros:

$$E = \sum_{n=1}^k \sum_{x \in q_n} \|x - k_n\|^2, \quad (1.6)$$

no qual k é o número de clusters, k_n é o centro do cluster n . Note que SOM e K-means apresentam o mesmo algoritmo quando o raio da função vizinhança de SOM é igual a zero. Nesse caso o ajuste é realizado apenas na unidade vencedora.

A vantagem do SOM é que ele possibilita a visualização de qualquer conjunto de entrada em um conjunto de saída em duas dimensões e, K-means não preserva topologia e é de difícil visualização. Todavia não há a exigência de um método tão sofisticado e custoso para a segunda separação de grupos, uma vez que os dados já foram classificados pelo SOM.

Feita a escolha do algoritmo menos robusto, é imprescindível a determinação do número de *clusters* do algoritmo K-means. Para tal foi escolhido o índice Davies-Bouldin que auxilia e automatiza o processo de classificação, gerenciado pela Eq.(1.7):

$$\frac{1}{C} \sum_{k=1}^C \max_{l \neq k} \left\{ \frac{S_c(Q_k) + S_c(Q_l)}{d_{ce}(Q_k, Q_{kl})} \right\}, \quad (1.7)$$

no qual C é o número de grupos, S_c é a distância entre as amostras dos grupos Q_k e Q_{kl} e d_{ce} a distância entre os centróides dos respectivos grupos (calculada através da localização das unidades do mapa). Portanto, valores do índice de Davies e Bouldin menores que 1 representam

agrupamentos separados, enquanto valores maiores que 1 representam grupos possivelmente cruzados.

1.3.3. ANOVA

Como o conjunto de dados utilizado para estudo apresentava apenas 2,5% de dados de testemunho, foi necessário um estudo estatístico dos dados dos perfis de poço em cada ponto, para inferir se o comportamento do restante do conjunto (97,5%) era similar ao dos pontos testemunhados e assim extrapolar tais características para todo o grupo.

O método escolhido para a validação dos grupos foi a análise de variância multivariada utilizada para comparar vetores de médias, cujos dados normalmente são provenientes de parâmetros estatísticos. A formulação de um teste estatístico para comparar vetores de médias depende da partição do total da variância subdividida em: variância devido ao efeito de tratamentos e variância devido ao erro. O método estatístico utilizado para avaliar a partição da variância total é denominada de ANOVA (Johnson & Wichern, 1999).

Em experimentos que envolvem variáveis aleatórias contínuas, como no caso analisado, pode-se pressupor a multinormalidade que permite a análise multivariada. Um ponto relevante da análise multivariada é o aproveitamento em conjunto das variáveis envolvidas (Regazzi, 2012).

Na técnica ANOVA, trabalha-se com duas hipóteses: H_0 , ou hipótese nula, em que não há variância da média das variáveis analisadas; H_1 , ou hipótese alternativa, na qual há variância da média das variáveis.

As hipóteses são testadas segundo F de Levene, que consiste em obtenção de um critério de evidência de rejeição da hipótese H_0 , dada pelo p-valor, segundo a Tab. 1.1.

Tabela 1. 1: Critérios utilizados para rejeição da hipótese H_0 . Fonte: Shimakura (2008).

$P \geq 0,10$	Não existe evidência contra H_0
$P < 0,10$	Fraca evidência contra H_0
$P < 0,05$	Evidência significativa contra H_0
$P < 0,01$	Evidência altamente significativa contra H_0
$P < 0,001$	Evidência muito altamente significativa contra H_0

Da Tab. 1.1 é possível inferir que para valores de p maiores que 0,1 não há evidências estatísticas contra a variação dos valores de média da variável analisada, ao passo que a medida que tal valor diminui, a evidência de que há variação destas médias aumenta. É relevante ressaltar o cuidado da definição dos critérios de evidências, que salienta a não existência de uma certeza absoluta, enfatizada principalmente pelo estudo da contrariedade e não da afirmação direta da ocorrência do evento.

A partir deste critério, é possível inferir se os valores testemunhados de cada grupo, formulado a partir do método matemático SOM, são representativos do restante da amostra não testemunhada.

Para os casos em que há variação da média amostral, ou seja, quando o valor de p for baixo, é necessária uma segunda abordagem estatística, para determinação de intervalos de confiança que avalie se tal variação é significativa.

1.3.4. ANÁLISE DE DADOS

Os dados de poços utilizados do Campo de Namorado, disponibilizados pela Agência Nacional de Petróleo (ANP), são referentes a um conjunto de 17 poços, dos quais foram analisados os perfis sônico (DT), porosidade neutrônica (NPHI), densidade (RHOB) e raios gama (GR). Dos poços analisados, seis contém dados de testemunho, com 29 litofácies.

Os dados de eletrofácies foram previamente classificados utilizando o método supervisionado k-NN (*k-Nearest Neighbor*) (Drummond & Vidal, 2011), que apresenta 80% de acerto na validação cruzada. Por meio deste método, foram identificados três grupos de eletrofácies: reservatório, possível reservatório e não-reservatório. Apenas os intervalos

classificados como reservatório e possível reservatório foram utilizados neste trabalho. Esse conjunto é composto por 11.600 amostras das quais apenas 2,5% apresentam testemunho.

O banco de dados apresenta seis litotipos, dois de rochas reservatório e quatro de rochas possível reservatório (Tab. 1.2).

A partir da análise do conjunto objetivou-se encontrar subdivisões conforme as características dos perfis geofísicos de poços e avaliar a distribuição dos diferentes tipos definidos no Campo de Namorado.

Tabela 1. 2: Descrição das fácies do Campo de Namorado.

Fácies	Classificação	Descrição	Número de amostras
4	Possível reservatório	Conglomerados Residuais	6
6	Reservatório	Arenito Grosso, Amalgamado	90
7	Possível reservatório	Arenito Médio Fino Laminado	4
8	Reservatório	Arenito Médio Gradado ou Maciço	115
10	Possível reservatório	Arenito/Folhelho Interestratificado	35
11	Possível reservatório	Arenito/Folhelho Finamente Interestratificado	55

Para a utilização do SOM na caracterização de fácies reservatório, o trabalho foi dividido em três etapas: Análise dos dados, Treinamento e Identificação de Litologia (Fig. 1.3).

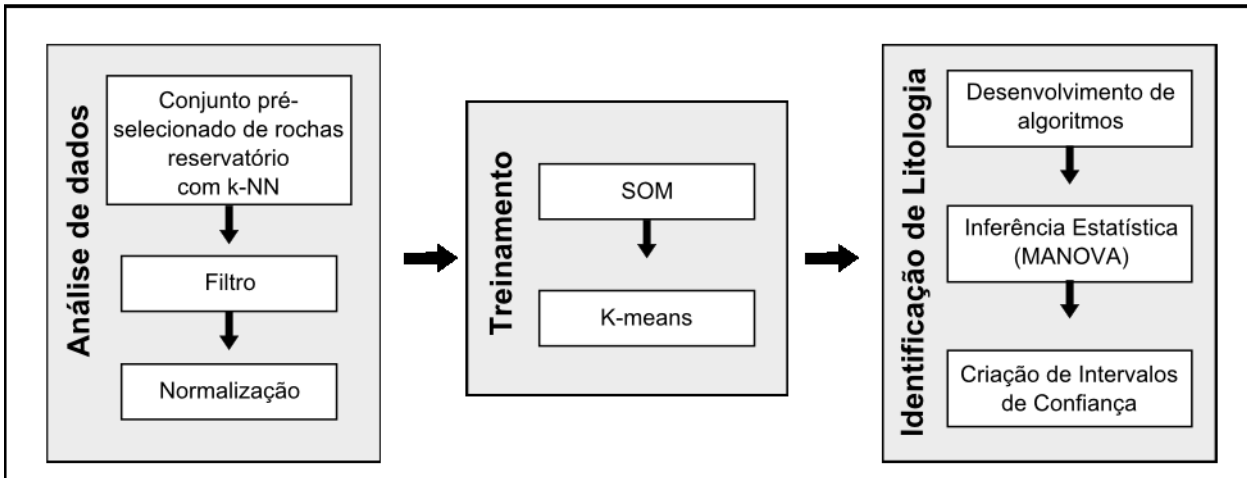


Figura 1. 3: Esquema da metodologia de trabalho.

O desempenho do modelo de Redes Neurais é baseado no princípio descrito por Alhoniemi *et al.* (1999), que enfatizam que os resultados são condicionados principalmente pela qualidade do conjunto de dados utilizados.

Seguindo o método proposto pelos mesmos autores, os dados foram processados segundo cinco etapas:

a) Aquisição de dados: os dados originais são analisados, suas variáveis medidas e as *strings* que representam a nomenclatura dos litotipos foram transformadas em entradas numéricas;

b) Pré-processamento de dados: nesta etapa os dados incorretos são removidos ou corrigidos. Geralmente o filtro é realizado mediante condições especiais ou fixadas. Neste caso, todas as entradas, mesmo com falhas de dados de algumas variáveis, foram consideradas para treinamento, pois SOM apresenta bons resultados mesmo com a ausência de algumas entradas;

c) Segmentação: divide o conjunto de dados em subgrupos de acordo com critérios, que são determinados usando informações *a priori*. Na base de dados analisada foi considerado o domínio não-supervisionado e desta forma não foram realizadas subdivisões;

d) Extração de Características: transforma os vetores de entrada de modo que descrevam o problema de acordo com a análise de “ponto de vista”. Modificações matemáticas, como por exemplo, decomposição de vetores com transformadas de Fourier, acréscimo de variáveis ou descon siderações de outras que prejudicam o treinamento da ferramenta. Nesta etapa do trabalho, definimos que o perfil ILD (resistividade) era redundante para o conjunto de dados a ser treinado, mediante testes preliminares com o método escolhido. A sua ausência não interferiu na qualidade de resultados obtidos, além disso, foi descon siderada por não caracterizar informações petrofísicas (como as outras variáveis), mas de fluido.

e) Normalização: oferece aos vetores de entrada o peso correto para o treinamento do SOM, assegurando que cada componente terá influência equivalente durante o treinamento. No caso deste trabalho, os dados foram previamente submetidos à fórmula:

$$Valor\ Normalizado = \frac{x_i - \min x}{\max x - \min x}, \quad (1.8)$$

em que x_i é o vetor analisado, $\min x$ é o menor valor de perfil e $\max x$ o maior.

Posteriormente ao agrupamento da base de dados com a técnica de RNA foram desenvolvidos algoritmos para avaliação das respostas obtidas, mediante estudo de ANOVA, descrito anteriormente.

1.3.5. TREINAMENTO

Para os diversos treinamentos dos perfis foi utilizada a ferramenta desenvolvida por Vesanto & Alhoniemi (2000), intitulada “Self-Organizing Map in Matlab”, que permite o uso de SOM e o reagrupamento posterior com o método K-means (Vesanto *et al.*, 1999).

Este pacote oferece a possibilidade de tomadas de decisão *a priori* pelo intérprete e utiliza fórmulas empíricas para escolha de valores, como o índice de Davies-Bouldin, considerado bom indicador para análises não-supervisionadas (e.g. Matos *et al.*, 2004).

Os dados referentes aos 17 poços foram treinados juntos, considerando a ausência de importantes efeitos diagenéticos (Flexa *et al.*, 2004).

1.4. RESULTADOS

Para o treinamento do conjunto de dados, o tamanho da matriz-U foi calculado pela Eq.(1.5), sendo dimensionada em 308 unidades (14x22).

Na Fig. 1.4, a contribuição de cada variável de entrada é representada nos planos das componentes e a Matriz-U obtida. Os menores valores indicam as maiores similaridades entre os vetores de entrada e os nós, representados por cores “frias”, enquanto as cores “quentes” representam as menores similaridades. Dessa forma, os grupos são caracterizados por regiões azuladas e homogêneas separadas por regiões de cores quentes.

A Fig. 1.5A refere-se ao diagrama Davies-Bouldin. O ponto em que a função apresenta menor valor é em $x=4$ e por isso este foi considerado o melhor número de grupos para agrupamento da Matriz-U, uma vez que minimiza a função descrita pela Eq.(1.7).

Na Fig.1.5B encontra-se a Matriz-U reagrupada por K-means. Nota-se que as regiões entre grupos apresentam unidades de BMU menores, o que indica menor concentração de amostras pertencentes a elas e, portanto, representam uma boa separação entre os quatro grupos.

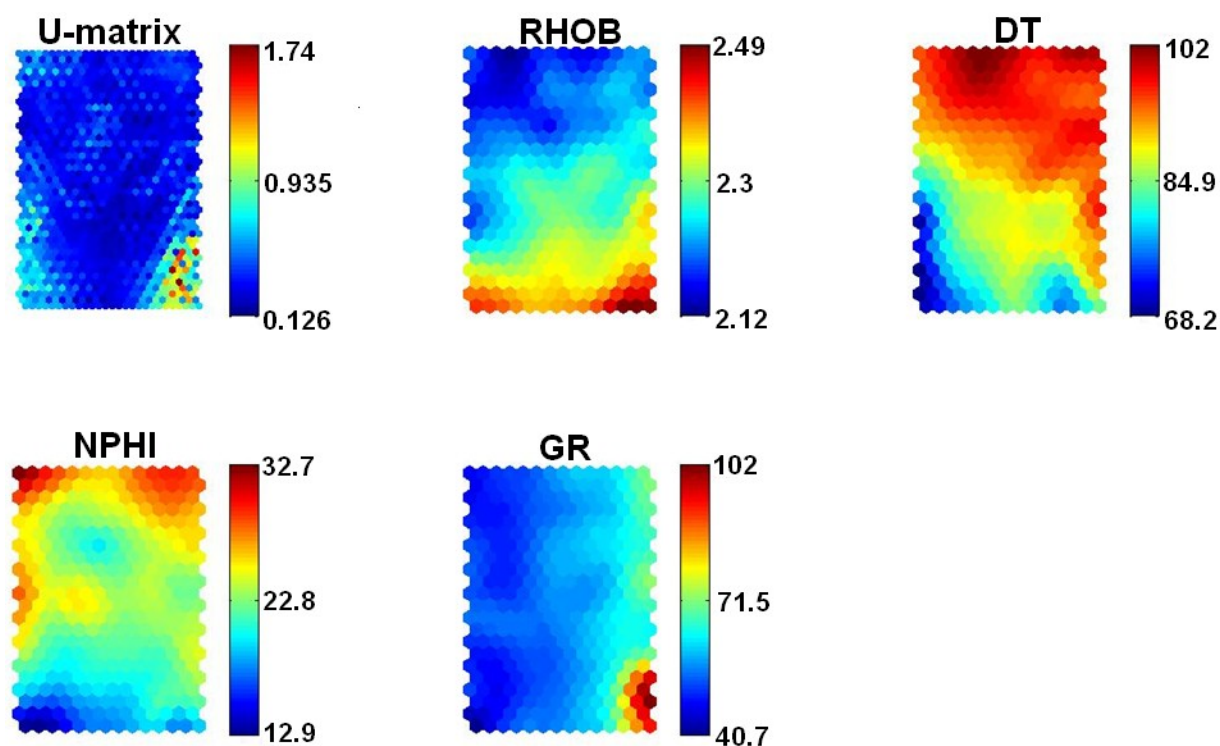


Figura 1. 4: Planos de componentes de cada variável analisada.

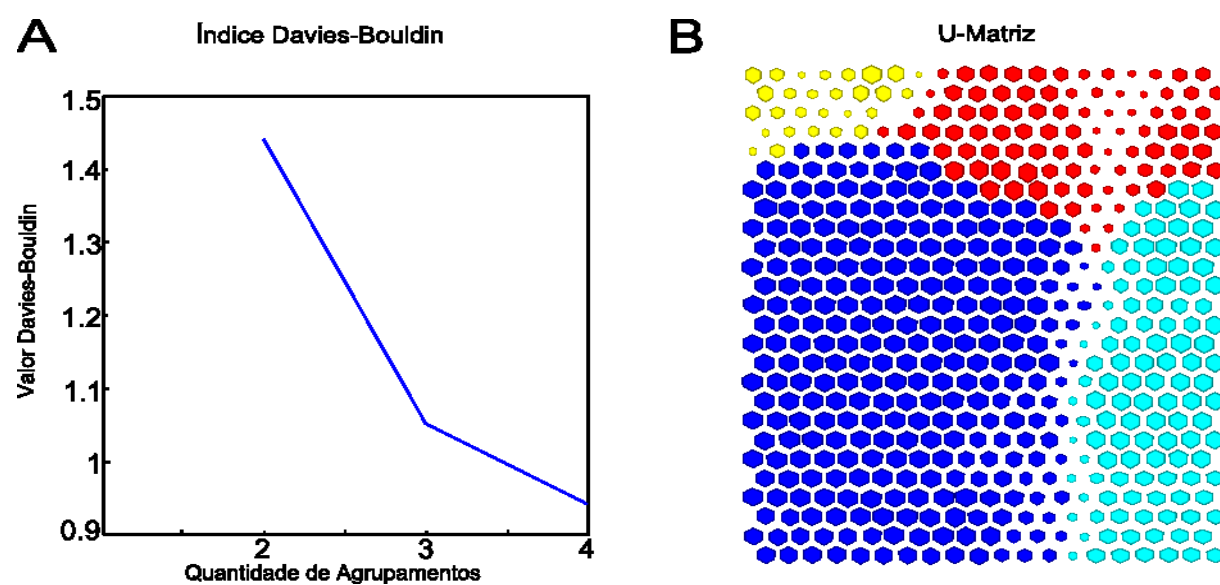


Figura 1. 5: Imagem gerada pela ferramenta de Vesanto *et al.* (1999). (A) Escolha de número de clusters do método de agrupamento k-means, (B) U-matriz reagrupada com o método mencionado.

Ao analisar os valores dos perfis para cada grupo definido após o treinamento é possível observar que a base de dados apresenta maior variação de valores nos perfis sônicos e raios gama, com exceção do primeiro grupo. Os valores de perfis de densidade quase não apresentam variação, uma vez que o intervalo do conjunto de dados é extremamente pequeno. Avaliando apenas as médias é possível verificar que os grupos apresentam valores distintos para a maior parte dos perfis, porém esta análise não é suficiente para inferir a classificação litológica e, portanto, a qualidade dos reservatórios que representam.

Tabela 1. 3: Estatística básica dos valores de densidade (RHOB), tempo de trânsito (DT), porosidade neutrônica (NPHI) e raios gama (GR) para os quatro grupos.

Grupo 1					Grupo 2			
	RHOB	DT	NPHI	GR	RHOB	DT	NPHI	GR
Medianas	2,621725	97,14255	29,1523	85,87885	2,34975	80,1862	18,6719	53,457
Desvio Padrão	42,961	42,80318	27,23086	20,78267	0,059747	7,086432	3,279699	7,937084
Variância	1845,648	1832,113	741,5198	431,9192	0,00357	50,21751	10,75642	62,99731
Máximo	88,7639	110,1523	90,94525	122,9446	2,6243	94,6484	27,4766	75,9258
Mínimo	2,3984	18,6719	1	75,9258	2,1721	55,8873	0	18,2422
Médias	2,271066	88,24874	23,28761	56,77831	2,352741	78,78757	18,26767	52,44084

Grupo 3					Grupo 4			
	RHOB	DT	NPHI	GR	RHOB	DT	NPHI	GR
Medianas	2,427	87,3359	21,0664	90,94525	2,1936	96,0391	25,7053	55,3828
Desvio Padrão	0,084509	7,691089	3,68205	14,00786	0,050729	4,33935	3,489456	7,227869
Variância	0,007142	59,15284	13,55749	196,22	0,002573	18,82995	12,1763	52,24209
Máximo	2,614	104,625	29,1523	122,9446	2,3984	110,1523	37,2344	87,6953

Mínimo	2,1129	64,3125	11,69788	65,266	2,0079	78,2865	11,4778	37,1641
Médias	2,429389	86,51059	20,5286	91,67676	2,191262	96,12575	25,88543	56,26466

Os valores de desvio padrão dos perfis nos grupos são muito pequenos, com exceção do grupo 1, o que mostra relativa concentração dos grupos e ocorrência de poucos pontos dispersos. Assim, é possível concluir que os grupos estão bem separados.

A partir da Tab. 1.3, nota-se que as variáveis com maior variabilidade foram separadas de acordo com intervalos bem definidos, demonstrando que o método é sensível a pequenas variações, como é o caso dos dados de perfis densidade, que mesmo apresentando escala de valores menor, foram separados de maneira satisfatória.

Foram realizadas comparações dos grupos com as informações de testemunhos associados a eles. Mediante esse estudo, destacam-se os seguintes resultados da Tab. 1.4.

Tabela 1. 4: Classificação dos dados supervisionados, em porcentagem.

Grupo	Dados Testemunhados (%)	Reservatório (%)	Possíveis Reservatórios (%)
1	15,75	78,57	21,43
2	19,48	90,32	9,68
3	30,94	14,29	85,71
4	17,37	83,86	16,44

Da Tab. 1.4 é possível verificar a baixa representatividade dos testemunhos para cada um dos grupos. Comparados com testemunho, os dados apresentaram três grupos com alta porcentagem de rochas reservatório, respectivamente 78%, 90% e 83% e um grupo com alta concentração litológica de possíveis reservatórios, 85%.

Encontra-se, portanto, três grupos para rochas reservatório, com destaque para o segundo que agrupa maior quantidade dessa litologia. Vale destacar que a descrição de testemunho classificou duas fácies para esses intervalos e pela técnica SOM ocorreu uma terceira subdivisão.

O grupo 3 é composto por rochas que são possíveis reservatórios, com valores bem distintos em relação aos demais grupos.

A Tab. 1.5 apresenta a comparação com dados de testemunho para cada grupo formado com a técnica SOM. É possível identificar que o grupo três, representante de rochas possíveis reservatórios, é constituído por 81% de Arenito e Folhelho Finamente Interestratificado, enquanto o grupo quatro tem concentração de aproximadamente 93% de Arenito Médio Gradado ou Maciço.

Os grupos 1 e 2 apresentam maior variação de litologias de rochas reservatórios. O grupo 1 possui cerca de 23% de Arenito Grosso e Amalgamado e 52% de Arenito Médio Gradado ou Maciço, enquanto o grupo 2 constitui-se de 30% e 54% das litologias mencionadas respectivamente.

Tabela 1. 5: Classificação litológica dos dados supervisionados.

Grupo	Arenito Grosso, Amalgamado	Arenito Médio Gradado ou Maciço	Conglomerados Residuais	Arenito Médio Fino Laminado	Arenito / Folhelho Interestratificado	Arenito / Folhelho Finamente Interestratificado
1	0,2262	0,5204	0,0271	0,0090	0,0769	0,0905
2	0,2985	0,5373	0,0224	0,0075	0,0224	0,0373
3	0,1163	0,1667	0	0	0,0233	0,8140
4	0,0553	0,9290	0	0,0383	0,0766	0,0383

Diante da complexidade dos dados, torna-se imprescindível a análise estatística do p-valor para diagnóstico do comportamento dos perfis observados. Através do teste F de Levene, relacionados aos critérios utilizados na Tab. 1.2, os p-valores encontrados para os perfis densidade (RHOB) e sônico (DT) são muito baixos, revelando evidência muito significativa da variação da média amostral referentes aos conjuntos 1, 2 e 4 (Tab. 1.6). Com isso os perfis densidade e sônico apresentam, para os grupos referidos, alta possibilidade de possuírem pontos com valores muito distintos e, por isso, comprometem a unicidade da média de cada grupo e necessitam de um estudo estatístico posterior.

Entretanto, os perfis porosidade neutrônica (NPHI) e raios gama (GR), do mesmo conjunto de grupos, não apresentam evidências de variação da média amostral do conjunto analisado e por isso, não se pode rejeitar a hipótese nula. Conclui-se, então, que estas duas variáveis não apresentam variações de média amostral estatisticamente significativa. Assim, o conjunto de dados destas duas variáveis pode ser representado através das características testemunhadas, ou seja, os resultados obtidos indicam grande evidência de que os grupos apresentam comportamento semelhante aos dados de testemunho para os perfis mencionados.

Analisando os p-valores das variáveis do grupo 3, nota-se que a primeira variável, perfil de densidade (RHOB), apresenta valor em que há fraca evidência de variação da média amostral e para as demais variáveis não existem evidências contra a hipótese de que a variação da média seja nula. Dessa maneira é possível inferir que os valores de testemunho analisados para o grupo três estão altamente correlacionados aos outros dados não testemunhados do grupo, e assim, tal grupo pode ser representado por seus dados testemunhados.

Tabela 1. 6: P-valores.

Variável	Grupo 1	Grupo 2	Grupo 3	Grupo 4
RHOB	0,0004	0,0000	0,0171	0,0000
DT	0,0095	0,0103	0,2068	0,0147
NPHI	0,3094	0,2583	0,4041	0,6124
GR	0,4650	0,4196	0,3595	0,7736

Todavia, por causa da variabilidade encontrada para os dois primeiros perfis, é necessário analisar os intervalos de confiança, calculados segundo análise de inferência estatística para obtenção de valores que identifiquem e confirmem suas características geológicas.

A partir da Tab. 1.7 é possível inferir que escolhidos, ao acaso, qualquer ponto do conjunto de dados, a probabilidade deste ponto pertencer aos intervalos de perfis descritos acima é de 95%, considerado um índice de elevada confiabilidade.

Tabela 1. 7: Intervalos de confiança.

	Grupo 1		Grupo 2		Grupo 3		Grupo 4	
	Valor Mín.	Valor Max	Valor Mín.	Valor Max	Valor Mín.	Valor Max	Valor Mín.	Valor Max
RHOB	2,268383	2,273749	2,348273	2,357208	2,415289	2,443489	2,188558	2,193966
DT	87,90005	88,59742	78,25766	79,31748	85,22736	87,79382	95,89444	96,35706
NPHI	23,16557	23,40965	18,02242	18,51292	19,91426	21,14294	25,69942	26,07144
GR	56,44568	57,11095	51,84732	53,03437	89,3396	94,01392	55,87937	56,64994

É possível verificar que os valores de perfil raio gama dos grupos 1, 2 e 4, estão entre 51 e 57, caracterizando bons reservatórios, constituídos de arenitos com baixa concentração de argila, conforme descrito nas amostras de testemunho e evidenciados pelos baixos valores de densidade em relação ao grupo 3.

Por outro lado, o grupo 3 é marcado pelos valores elevados de raios gama, superiores a 89, representativo de possíveis reservatórios, e elevada densidade, compatíveis com folhelhos e arenitos finos descritos para as litologia desse grupo.

Para comparar os valores de intervalos dos perfis nos grupos foram construídos *boxplots* associados à litologias testemunhadas no conjunto de dados, como mostra a Fig. 1.6.

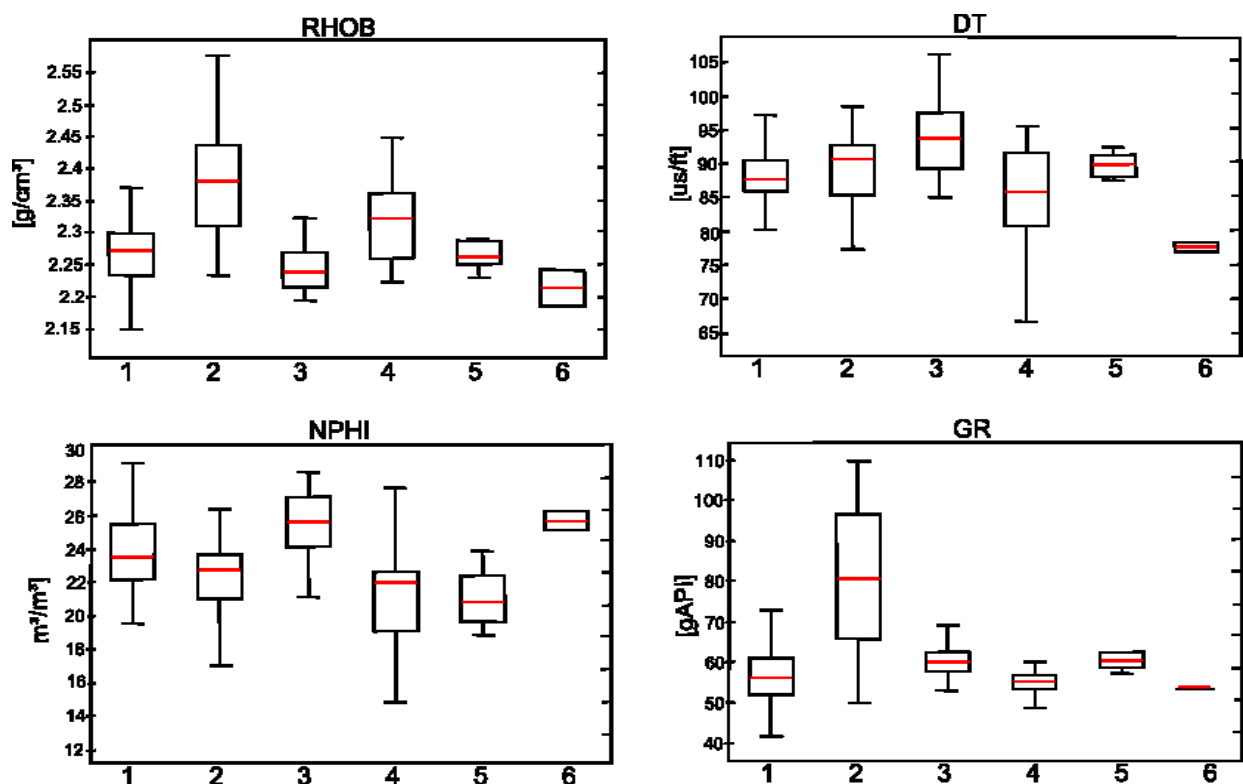


Figura 1. 6: Boxplots das litologias analisadas. 1 – Arenito Médio Gradado ou Maciço, 2 – Arenito/Folhelho Finamente Interestratificado, 3 – Arenito/Folhelho Interestratificado, 4 – Arenito Grosso Amalgamado, 5 – Conglomerado Residual, 6 – Arenito Médio Fino Laminado.

A partir da Fig.1.6 nota-se que os perfis possuem pequenas variações e valores muito semelhantes entre as litologias estudadas. Dentre as variáveis analisadas, o perfil de densidade apresenta maior similaridade nos seis grupos, com pouca variação para Arenitos médio gradado ou maciço, arenito/folhelho interestratificado, conglomerado residual e arenito médio fino laminado.

Dentre as litologias analisadas, a única variável que se destaca das demais é o arenito/folhelho finamente interestratificado (#2), que possui maior intervalo de valores, porém o arenito médio gradado ou maciço (#1), do grupo reservatório, apresenta intervalo de densidade muito semelhante a de litologias de possíveis reservatórios, arenito/folhelho interestratificado, conglomerados residuais e arenito médio fino laminado.

Do grupo de reservatórios, o litotipo arenito grosso amalgamado (#4) se assemelha muito ao litotipo arenito/folhelho finamente interestratificado (#2). Para estas duas classes os valores de perfis sônico e raios gama também são muito semelhantes, com intervalos contidos, mas são

diferenciados pelo perfil porosidade, em que a rocha possível reservatório apresenta valores maiores.

No perfil sônico destacam-se os valores do arenito médio fino laminado, que tem intervalo abaixo das outras classes. Todavia essa variável apresenta apenas quatro amostras e por isso pode-se notar a pequena variação de valores para todos os perfis. Com exceção deste litotipo, todos os outros apresentam médias muito próximas de densidade, o que dificulta a separação de classes.

A classe de arenito/folhelho interestratificado possui intervalo de perfil de porosidade que engloba os valores de arenito médio fino laminado, e representam as duas litologias com maiores valores.

No perfil de raios gama o intervalo do arenito médio gradado ou maciço (#1) compreende os valores encontrados nos grupos de possíveis reservatórios. Entretanto, o arenito/folhelho finamente interestratificado (#2) apresenta números muito maiores que os demais. Essa diferença de valores de raios gama para esta classe facilita o reconhecimento de padrão desse litotipo.

É possível observar que as litologias pouco se diferem entre si, dificultando a separação entre grupos litológicos, principalmente entre reservatórios e possíveis reservatórios, que não possuem assinaturas distintas entre os perfis.

O arenito médio gradado ou maciço (#1) não apresenta diferenças significativas que possam distingui-lo de litotipos de possíveis reservatórios. Em todos os perfis analisados, seus valores se encontram muito próximos dos valores de conglomerados residuais, salvo pela porosidade, que a rocha possível reservatório apresenta média relativamente menor.

Para a comparação dos resultados obtidos com as informações dos perfis de poços, foram utilizados os perfis RG com valor de corte de 60 °API, indicado pela cor verde. Os outros três perfis (DT, RHOB e NPHI) também estão representados, com destaque para o cruzamento dos perfis NPHI e RHOB, indicando rocha reservatórios (em amarelo).

Nos quatro poços presentes na Fig. 1.7, é possível verificar que os melhores intervalos para rocha reservatório indicados pelos perfis foram selecionados como fácies arenito grosso e arenito

médio pelo método SOM. Os arenitos intercalados com folhelhos foram definidos corretamente na maioria dos casos. Os resultados honram os valores de testemunho e classificam satisfatoriamente os poços nas diferentes litologias.

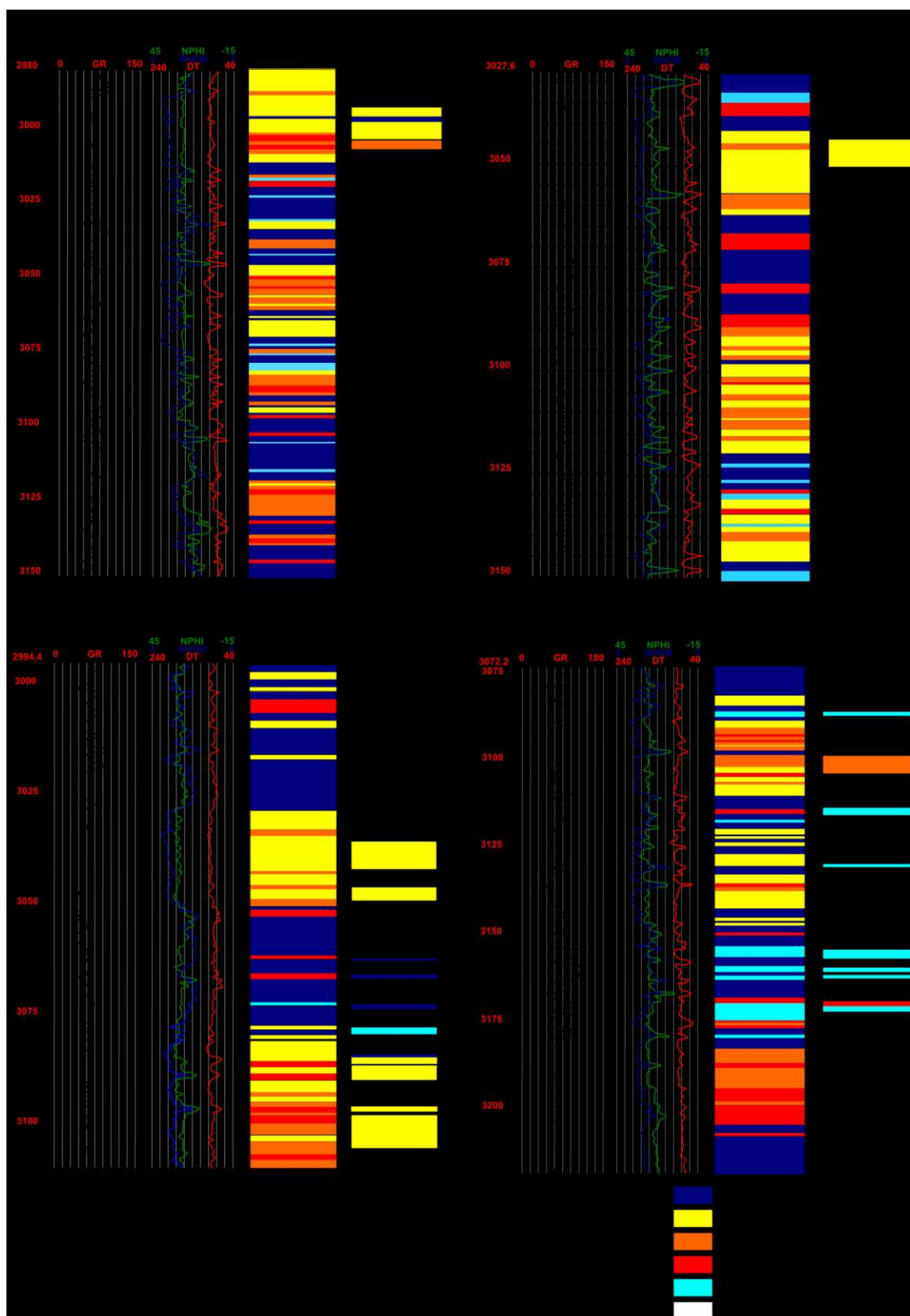


Figura 1. 7: Resultado de classificação litológica dos poços testemunhados.

1.5. CONCLUSÃO

Os dados geofísicos analisados neste trabalho (densidade, porosidade neutrônica, raios gama e tempo de trânsito) permitiram o melhor detalhamento da fácies reservatório, em uma abordagem automática realizada em duas etapas: a caracterização de eletrofácies com o método k-nn, que classificou o conjunto de dados em três classes (rochas reservatório, possível reservatório e não reservatório) e em uma segunda classificação, através de redes neurais artificiais, na qual o método SOM encontrou, sem a utilização de dados *a priori*, quatro litologias das fácies reservatório definida na primeira etapa (Arenito grosso amalgamado, arenito médio gradado ou maciço, arenito/folhelho finamente interestratificado e arenito/folhelho interestratificado).

Na análise básica dos dados estatísticos observa-se que os valores de perfis, de acordo com seus grupos, são constantes e similares nos quatro cenários encontrados com a aplicação de SOM. Isso implica que o comportamento dos grupos nos diferentes ambientes permanece quase inalterado. Por isso a análise mais detalhada dos grupos foi imprescindível.

Através do diagnóstico com os dados de testemunho é possível concluir que três dos quatro grupos apresentam comportamento semelhante, associado à composição de rochas reservatório. Enquanto apenas um grupo se destaca dos demais, constituído de rochas possíveis reservatório. Com análise dos dados utilizando SOM, foi possível obter um zoneamento dos poços e assim destacar e identificar características reservatório de cada grupo, com base nos perfis de poços. A eficácia do zoneamento foi comprovada também a partir do confronto com informações adicionais sobre o Campo de Namorado e sua extrapolação através de inferência estatística.

AGRADECIMENTOS

Os autores agradecem à Agência Nacional de Petróleo - ANP e à PETROBRAS pela disponibilidade de dados do Campo de Namorado.

REFERÊNCIAS

- ALHONIEMI E., HOLLMÉN J., SIMULA O., VESANTO J. -1999- Process Monitoring and Modeling using the Self-Organizing Map. Integrated Computer Aided Engineering. Disponível em: <<http://citeseerx.ist.psu.edu>>. Acesso em: 17 janeiro 2010.
- BACOCOLI G., MORALES R.G., CAMPOS O.J. -1980-. The Namorado Oil Field: A Major Discovery in the Campos Basin, Brazil. AAPG Bulletin, 63(3):329-338.
- BARBOZA E.G.- 2004-. Análise Estratigráfica do Campo de Namorado (Bacia de Campos) com base na Interpretação Sísmica Tridimensional. Tese de Doutorado, Universidade Federal do Rio Grande do Sul, 235p.
- BHATT A & HELLE H. B. -2002- Determination of facies from well logs using modular neural networks. Petroleum Geoscience, 8:217-228.
- COLÉOU T., POUPON M., AZBEL K. -2003- Unsupervised seismic facies classification: A review and comparison of techniques and implementation, Journal The Leading Edge, 22:942-953.
- COSTA J.A.F. -1999- Classificação Automática e Análise de dados por Redes Neurais Auto-Organizáveis. Dissertação de doutorado, FEEC/Unicamp, 345p.
- DIAS J.L., SCARTON J.C. ESTEVES F.R., CARMINATTI M, GUARDADO L.R. -1990-. Aspectos da Evolução Tectono-Sedimentar e a Ocorrência de Hidrocarbonetos na Bacia de Campos. In: Gabaglia, G.P.R. & Milani, E.j. (Eds.) Origem e Evolução das Bacias Sedimentares Brasileiras. Rio de Janeiro, Petrobras, p. 333-360.
- DRUMMOND R.D. & VIDAL A.C. -2011-. Comparação entre as técnicas multivariadas MAF e PCA aplicadas na classificação de eletrofácies. Revista Brasileira de Geofísica (aceito para impressão).
- ESPOSITO, A.M., GIUDICEPIETRO, F., D'AURIA, L., SCARPETTA S., MARTINI M.G., COLTELLI M., MARINARO M. -2008- Unsupervised Neural Analysis of Very-Long-Period Events at Stromboli Volcano Using the Self-Organizing Maps”, Boletim of the Seismological Society of America, 98:2449-2459.

- FLEXA R. T., ANDRADE A., CARRASQUILLA A. -2004-. Identificação de litotipos nos perfis de poço do campo de namorado (Bacia de Campos, Brasil) e do Lago Maracaibo (Venezuela) aplicando estatística multivariada. Rev. Bras. de Geoci. 34(4):571-578.
- GONÇALVES, M.L., NETTO, M.L.A., COSTA, J.A.F. -2007- Explorando as Propriedades do mapa Auto-Organizável de Kohonen na Classificação de Imagens de Satélite. Anais XXVII Congr. SBC, 1:1132-1141.
- GONÇALVES M.L., NETTO M.L. de A., ZULLO, J.J. -1996- Um Sistema Neural Modular para Classificação de Imagens Utilizando Mapas de Kohonen. Anais VIII Simp. Bras. de Sensoriamento Remoto 6º Encontro Nacional de Inteligência Artificial, 1:845-849, Salvador, BA.
- HAYKIN S. -1999- Neural Networks(2) Prentice Hall, Virginia, USA, 842 pp.
- JOHANN P.R.S. -1997-. Inversion Sismostratigraphique et Simulations Stochastiques en 3D: Réservoir Turbiditique Offshore du Brésil, Intégration Géologique, Géophysique et Géostatistique, Ph.D. Thesis, Académie de Paris, Université Marie Curie. 352 pp.
- JOHANN, P.R.S. -2004- Fácies Sísmicas. Boletim de Geociências da Petrobrás(12), 2:317-355. Rio de Janeiro, RJ.
- JOHNSON R. A., WICHERN D. W. -1999- Applied multivariate statistical analysis(4), Upper Saddle River, New Jersey: Prentice-Hall, 815pp.
- KHEDAIRIA S. & KHADIR M.T. -2008- Self-Organizing Map and K-Means for Meteorological Day Type Identification for the Region of Annaba-Algeria. 7th Computer Information Systems and Industrial Management Applications, p. 91-96.
- KOHONEN, T. -2001- Self-Organizing Maps. Springer-Verlag, New York, 501pp.
- LEITE E.P., FILHO C.R.S., -2010-. Mapas Auto-Organizáveis aplicados ao mapeamento do potencial mineral na região de Serra Leste, Província Mineral de Carajás, Pará. Rev. Brasileira de Geofísica. 28(3):397-409.
- MAJDI A., BEIKI M., PIRAYEHGAR A., HOSSEINYAR G. -2010-. Identification of Well logs with significant impact on prediction of oil and gas reservoirs permeability using

- statistical analysis of RSE values. Journal of Petroleum Science and Engineering. 75: 91-99.
- MATOS M.C., OSÓRIO P.L.M., JOHANN P.R.S. -2004- Agrupamento de Mapas Auto Organizáveis de Kohonen Aplicado a Análise de Fácies Sísmicas. XV Brazilian Automation Congress, 1:1-6.
- MILANI E.J., BRANDÃO J.A.S.L., ZALÁN P.V., GAMBOA L.A.P. -2000-. Petróleo na Margem Continental Brasileira: Geologia, Exploração, Resultados e Perspectivas, Rio de Janeiro. Rev. Bras. de Geofísica, 18:3:352-396.
- RAUBER A. -1999- LabelSOM: On the Labeling of Self-Organizing Maps. Citação de referencias e documentos eletrônicos. Disponível em: <<http://ieeexplore.ieee.org>>. Acesso em: 15 abril 2010.
- REGAZZI, A.J. -2002- Análise multivariada, notas de aula INF 766, Departamento de Informática da Universidade Federal de Viçosa.
- ROCHA, A.C.B. & SOUZA, F.A.M. -2003- Aplicação do K-means cluster a dados de perfis de poços petrolíferos. 2º Congresso Brasileiro de P&D em Petróleo e Gás, Rio de Janeiro.
- STUNDNER M. & OBERWINKLER C. -2004- Self-Organizing Maps for Lithofacies Identification and Permeability Prediction. In: SPE Annual Technical Conference and Exhibition held in Houston, p.8.
- SHIMAKURA, S. -2008- Citação de referências e documentos eletrônicos. Disponível em <<http://leg.ufpr.br/~silvia>>. Acesso em: 5 janeiro 2010.
- SOUZA JUNIOR OG. 1997. Stratigraphie Séquentielle et Modélisation Probabiliste des Réservoirs d'un Cône Sous-marin Profond (Champ de Namorado, Brésil). Intégration des Données Géologiques. Thèse de Doctorat. Université Paris. 128 pp.
- ULTSCH A. -1992- Knowledge acquisition with self-organizing neural networks. In: Aleksander, I. and Taylor, J. (eds.), Artificial Neural Networks, Elsevier Science Publishers B.V., 735-740.
- ULTSCH A. -1993- U*-matrix: a tool to visualize clusters in high dimensional data. University of Marburg, Department of Computer Science, Technical Report, 36, 12 pp.

- VESANTO J. & ALHONIEMI E. -2000- Clustering of the Self-Organizing Map. IEEE Transactions on Neural Networks, 11:586-600.
- VESANTO J., HIMBERG J., ALHONIEMI E., PARHANKANGAS J. -1999- Self-Organizing Map in Matlab: The SOM Toolbox. Proceeding of the Matlab DSP Conference, Espoo, Finland, 59pp.
- WINTER W.R., JAHNERT R.J., FRANÇA A.B. -2007-. Bacia de Campos. B.Geoci. Petrobras, Rio de Janeiro, v.15, p.511-529.

ANEXO 2

“Carvalho, A.M.A.; Kuroda, M.C.; Vidal, A.C.
Classification and Correlation of Sedimentary Facies in the
Taubaté Basin, São Paulo State, Brazil (submetido aos Anais da
Academia Brasileira de Ciências).”

CLASSIFICATION AND CORRELATION OF SEDIMENTARY FACIES IN THE TAUBATÉ BASIN, SÃO PAULO STATE, BRAZIL

Ancilla Maria Almeida de CARVALHO, Michelle Chaves KURODA, Alexandre Campana
VIDAL

Institute of Geosciences, P.O.Box 6152, University of Campinas – UNICAMP, 13083-870,
Campinas, SP, Brazil. Tel: +55-19-3521-4659; Fax: +55-19-3289-1562. E-mails:
mckuroda@ige.unicamp.br, carvalho.ancilla@gmail.com, vidal@ige.unicamp.br

ABSTRACT – The facies classification is a fundamental step in the process of geological characterization of a sedimentary basin. Given the lack of core data, the integration of mathematical techniques and indirect geological data, such as geophysical logs and seismic sections, the estimate facies and petrophysical properties becomes essential. The aim of this work is applying the neural network technique, Kohonen Self-organizing Maps, to define electrofacies from the geophysical logs, in Taubate Basin. This classification allowed the identification of reservoir units, the correlation of these units in São José dos Campos and Caçapava areas and association with seismic units provided a better understanding of the distribution of sedimentary units in this portion of the basin.

Keywords: Taubaté Basin, neural network, geophysical well log.

2.1. INTRODUCTION

The definition of a geological model in reservoirs characterization is an important step that includes the definition of the faciological model, the petrophysical properties and the fluid dynamic properties, allowing a better understanding of the reservoirs and their heterogeneities (Consentino, 2001).

The subsurface facies characterization implies a lithofaciological study using direct geological data, such as core samples, and indirect data, such as seismic and geophysical well logs. Traditionally the lithological classification is more precise applying core analysis, but this type of information is not always available. Therefore, the analysis of geophysical logs, allied to techniques such as multivariate statistics and artificial intelligence, is fundamental to estimate lithological and petrophysical properties.

The different well logs provide information on some rock properties, making possible, with the help of multivariate statistical methods, the clustering of samples with similar characteristics. These are frequently named electrofacies.

The multivariate statistical methods include the discriminant analysis (Tang *et al.* 2004; Li and Anderson-Sprecher, 2006), cluster analysis, principal component analysis (PCA, Jolliffe, 2002), independent component analysis (ICA), among others. According to Tang and White (2008) these methods are flexible and furnish clear mathematical expressions of the facies-log relationship.

The ICA and PCA methods aim at data reduction by eliminating redundancy in each dimension, whereas the discriminant analysis searches for the best (also linear) function of a group of variables that reproduces *a priori* clustering of selected cases. However, due to growing complexity, the data may not yield the expected correlation using such methods, making the implementation impossible. To explore more complex and effective relationships and to take into account as many variables as possible, without eliminating any of them from the analysis, the most indicated methods, according to Bishop (1994), are the artificial neural networks.

Some advantages gained using these methods are the following: sensitivity to the presence of noise in the data set; no necessity of knowing the statistic distribution of the variables (no restrictions to data type), and modeling of non-linear relationships between variables (Bishop 1994, Costa 1999, Stundner & Oberwinkler 2004).

The artificial neural networks are methods typical of artificial intelligence (Baldwin *et al.* 1990; Bhatt and Helle, 2002). They are classified in two main types, non-supervised and supervised (on the basis of the existing information *a priori*). According to Chang *et al.* (2000), the most common supervised network is the back-propagation neural network (BPNN), and the most common non-supervised network is Kohonen's Self-Organizing Maps (SOM).

According to Coléou *et al.* (2003), among the available clustering techniques, SOM provides precise results and represents a less expensive method than the others. The main characteristic of the method is a convenient 2D display of the clustering of n-dimensional data (Kohonen, 2001). Being pattern recognition with no use of previous data information its essential characteristic, this method is an important tool to characterize wells devoid of core samples.

In this sense, this work aims at the application of the SOM tool to geophysical logs available for the Taubaté Basin for electrofacies characterization. The application of this tool is appropriate because gamma-ray and resistivity logs are available for a large number of wells, which are devoid of detailed geological information.

The definition of the main electrofacies and the correlation of the main logs in the regions of the São José dos Campos and Caçapava cities, areas with higher density of information, will provide a better understanding of the sedimentary units distribution in the Taubaté Basin and of its depositional model.

The Taubaté Basin, located in the axis linking the São Paulo and Rio de Janeiro urban centers, corresponds to one of the main aquifers of São Paulo State (Fig. 2.1). Because of the high density of wells, 2D seismic lines, good outcrops and a significant number of published works, allow it to be studied as an example analogous to fluvial reservoirs.

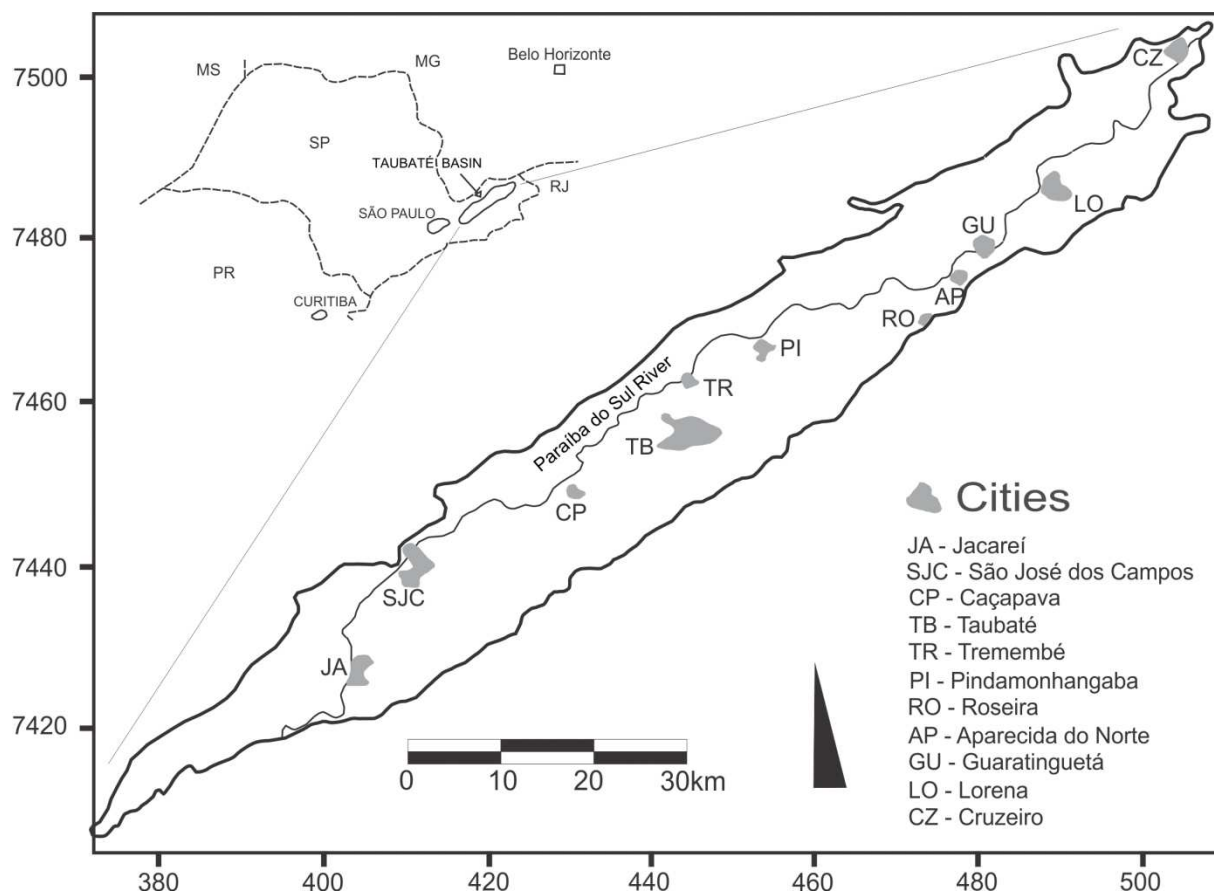


Figure 2. 1: Location of the study area. In detail the Taubaté Basin in São Paulo State.

2.2. GEOLOGY

The Taubaté Basin is one of the basins that compose the Southeastern Brazil Continental Rift – RCSB (Riccomini, 1989), a Cenozoic tectonic feature located between the Mar and Mantiqueira Ridges, in the eastern portion of São Paulo State.

The basin is structured in half-grabens, separated by transfer or accommodation zones, with reversed depocenters and alluvial fans associated with border faults, in a geometry typical of rift basins (Fernandes 1993). Fernandes and Chang *et al.* (2000) propose the subdivision of the Taubaté Basin in three compartments named, from southwest to northeast, São José dos Campos, Taubaté and Aparecida, separated by two internal structural highs, Caçapava and Pindamonhangaba. Carvalho *et al.* (2011), using geostatistical methods and seismic, gravimetric and well log data, propose another subdivision in four main depositional centers, corroborating with Marques (1990) for the Lorena sub-basin. The Lorena sub-basin is located in the

northeastern portion of the Taubaté Basin and is separated from the Aparecida Compartment by the Capela de Santa Luzia High (Fig. 2.2).

The sedimentation in the Taubaté Basin is typically continental and, according to Riccomini (1989), it can be divided in two phases: the first, syn-tectonic to the rift, corresponds to the deposition of the Taubaté Group, formed by the Resende, Tremembé and São Paulo formations. The second, posterior to the diastrophic tectonics, is represented by the Pindamonhangaba Formation and alluvial and colluvial deposits (Fig. 2.2).

Appi *et al.* (1986) and Chang *et al.* (2000), using sequence analysis, divide the Taubaté Basin sedimentation in three main evolutionary phases. Phase I corresponds to the development of an alluvial fan system, with the gravitational flow deposition, characterized by lamites and sandy lamites (Resende Formation). As subsidence progresses, phase II is installed, corresponding to a lacustrine system extending up to its apex during the deposition of the Tremembé Formation bituminous shales. In this phase sporadic floods occurred, leading to the deposition of inundite bodies at lake rims, and a desiccation event responsible for the formation of a caliche-type limestone layer. Phase III corresponds to the installation of a meandering fluvial system that persisted up to the Holocene (São Paulo and Pindamonhangaba formations).

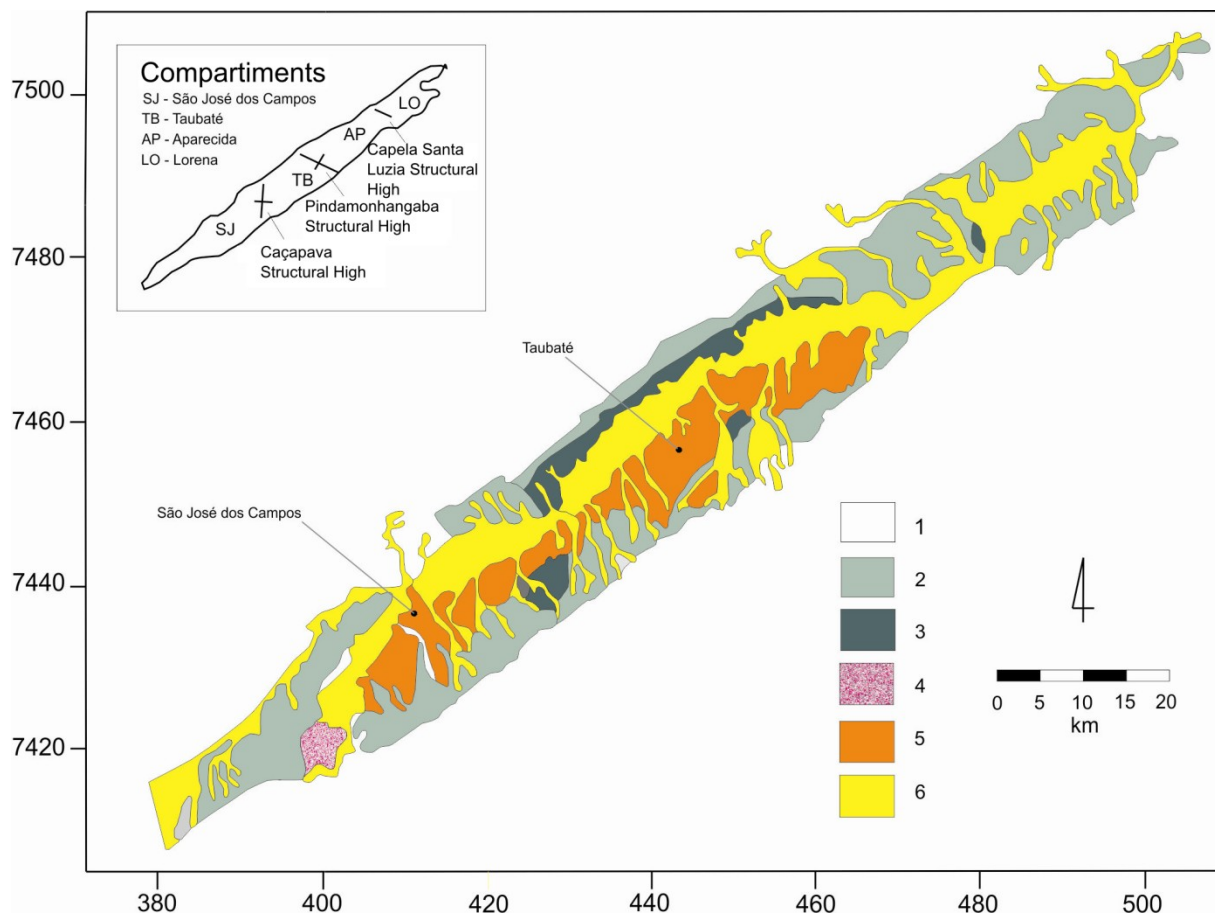


Figure 2. 2: Schematic geological map for the Taubaté Basin. (1) basement rocks; (2) Resende Formation; (3) Tremembé Formation; (4) São Paulo Formation; (5) Pindamonhangaba Formation; (6) Quaternary sediments (Modified after Riccomini, 1989).

In the study of Paleogene basins of the São Paulo and Rio de Janeiro states, Riccomini (1989) identifies 20 facies that were grouped in the corresponding depositional systems. According to the author, the main phase in the RCSB evolution occurred in the Eocene-Oligocene, with the syntectonic sedimentation represented by the Taubaté Group. Out of the 20 facies identified, 11 are well represented in the southern and central portions of the Taubaté Basin.

The basal and lateral unit of the Taubaté Group is the Resende Formation, constituted by alluvial fan systems associated with braided alluvial plain deposits. This unit is formed by ortho and paraconglomerates in proximal areas, lamites in the intermediate and distal portions of the alluvial fans, and fluvial sandstones in braided rivers axial to the basins (Riccomini *et al.* 2004).

The conglomeratic rocks predominate along the northern border of the basins of the RCSB central segment, which were deposited by subaerial gravitational flow processes. In the Taubaté Basin they occur close to the border faults, cropping out in the northern border of the Aparecida Compartment and south of the Taubaté Compartment. The intermediate and distal portions of the alluvial fans are represented by massive lamites and sandy, sometimes conglomeratic, lamites that laterally interpenetrate proximal alluvial fan deposits. The more distal portions of the alluvial fans can contain accumulations of centimeter- to decimeter-sized organic matter (lignite) (Riccomini *et al.* 2004).

Based on detailed descriptions of sedimentary sections located on the southeastern border of the basin close to Jacareí, Almeida (2007) defines nine sedimentary facies, including conglomerates, conglomeratic sandstones, medium- to coarse-grained sandstone and massive siltite. The outcrops represent an alluvial fan system established under arid climatic conditions and deposited at the flexural border of the RCSB southern compartment.

In the central portion of the Taubaté basin, Almeida (2007) describes a segment composed of nine facies: conglomeratic sandstone, massive medium-grained sandstone, massive siltite, massive argillite, very fine-grained sandstone, fine-grained sandstone, siltite with bioturbations, diamictite and shale. This segment represents an alluvial fan sedimentation at the faulted border of the RCSB central compartment.

The Resende Formation alluvial fan system grades laterally and vertically to the Tremembé Formation *playa lake*-type lacustrine system, located in the Taubaté Basin central portion (Riccomini and Coimbra, 1992). From core and outcrop data, the main lithofacies recognized in the Tremembé Formation encompass: massive green argillites, shale and marl rhythmities, dolomites and sandstones (Riccomini 1989; Riccomini & Coimbra 1992).

On top of the Taubaté Group, the São Paulo Formation is represented by sedimentary deposits related to meandering fluvial systems. The main São Paulo Formation lithofacies are coarse-grained, conglomeratic sandstones with cross stratifications, erosive base and clayey clasts representative of meandering channel deposits; siltites and laminated argillites, sometimes fossiliferous (lignites), deposited in abandoned meanders; and medium- to coarse-grained sandstones, with sediments representing the breakup of marginal dikes, grading to more fine-grained, flood plain rhythmic and laminated sediments (Riccomini *et al.* 2004).

The Pindamonhangaba Formation discordantly overlies the deposits of the Tremembé, Resende and São Paulo formations. It corresponds to a meandering fluvial system developed in the central and southwestern portion of the Taubaté Basin. This unit was subdivided by Mancini (1995) in two interpenetrated members: the Rio Pararangaba Member, representing the coarse-grained canal facies, and the Presidente Dutra Member, related to a fine-grained flood plain deposition of the meandering fluvial system.

By interpreting seismic profiles, Fernandes (1993) carries out a stratigraphic analysis of the Taubaté Basin sediments. Five seismic units are recognized in 11 seismic profiles along the basin. According to the author, the units constitute reflectors with similar characteristics. The author stresses out that the division was based on the vertical variations of the seismic reflection, so that the same seismic unit can contain different seismic characteristics, according to its location in the basin.

Seismic unit A is the oldest, overlies the basement, and is limited to the northeastern portion of the basin. It has very irregular wedged shapes interpenetrating fanglomeratic deposits close to border faults. According to the author, the progradational seismic facies can characterize tractive processes, characteristic of conglomerate and sandstone deposition.

Seismic unit B is extensively distributed in all compartments and overlies unit A. In the northwestern and central portions of the basin, unit B directly overlies the basement and is in contact with fanglomerates close to the border fault. This unit is irregularly tabular or wedged. Its subparallel to chaotic seismic facies may be associated with a fluvial or fluvial-lacustrine system.

Seismic unit C covers unit B and its reflectors present excellent lateral continuity, being traced all over the basin. Its seismic facies vary from parallel to slightly continuous parallel, which are characteristic of suspension processes, according to Fernandes (1993). In the central portions of the basin this interpretation is corroborated by the lithologic description of a well drilled in the Taubaté city, in which a thick sequence of shales and lamites with intercalations of coarse-grained facies is interpreted as inundite deposits by Riccomini (1989).

Concordantly with unit C, unit D is the expression of the most recent sedimentation in the Taubaté Basin, with a lensoidal shape. Its major development took place in the central portion of the basin, where the seismic facies is predominantly parallel; in other portions its characteristics are masked by the effect of seismic processing. According to Fernandes (1993), due to the

characteristics of the Taubaté Basin central portion, the depositional environment of unit D varied a little in relation to unit C.

Unit E is characterized close to the border faults, advancing on the more continuous reflectors. The seismic facies is chaotic, formed by weakly ordered reflectors, which represent the Resende Formation proximal alluvial fan deposits.

2.3. KOHONEN'S SELF-ORGANIZING MAPS

The SOM technique developed by Kohonen (1982) yields thematic maps ordered from input data (Freeman and Skapura, 1991; Ripley, 1996). This technique is based on the competitive learning process, where only one output computational unit (also called best-matching unit or BMU) gives an active response to a current input signal (Gonçalves *et al.* 1996).

The algorithm maps input data in a lower-dimensional space, delineated by a geometric structure called self-organizing map (Kohonen, 2001). The tool produces similarity of the input data and generates maps for the visualization of multidimensional clusterings, transforming statistical relationships of high-dimensional data in simple geometric relationships of the corresponding points in the lower-dimensional space. A set of close data in an input space will have close representations in the output space and will be classified as pertaining to the same cluster or to close clusters.

The competitive process occurs according to the criterion of greatest similarity given by the Euclidian distance or metric, measured from the input vector to each unit of the map. This value is weighed on the basis of the weight vector associated with each neuron, thus defining the location of the neuron in the map.

Once the closest BMU is found, the algorithm updates the weight values of the winning BMU and its neighbors, according to a function inversely proportional to the distance between them; for this reason, the resulting map is also called elastic net. This type of adjustment warrants the preservation of the important characteristic named topology.

The SOM unsupervised approach allows a self-adaptation of the algorithm to any input database without *a priori* knowledge. However, because no markers are used during the analysis, the clusters found by the algorithm need to be evaluated by an interpreter, in order to identify each one of the clusters.

In general, the number of units in the map is extremely high and a second reclustering of these BMUs is necessary, which frequently represent clusters with the same characteristics of their neighbors. This reclustering is commonly carried out by the k-Means method. It is a traditional, simpler clustering tool than Kohonen's, because the topology preservation is not taken into account. Because the topology has already been preserved (Kohonen, 2001), this feature is not necessary in reclustering once this second analysis is performed in each BMU centroid.

The data analysis using SOM and k-means shows sensitivity and performance directly associated with the choice of input attributes, such as the number of interactions or size of the adjustment radius of the net, but it is also tied to the complexity of the data (continuous or discrete), number of variables involved and sample size. The decision of the input parameters, according to Kohonen (2001), does not depend on logic but on the analyst's experience and sensibility, and therefore a series of tests is necessary.

2.4. WORKING METHOD

The work was developed in two steps: the first comprising electrofacies classification using geophysical logs available for the southwestern and northeastern portions of the basin, and the second for correlation of sedimentary units and association with seismic profiles in the São José dos Campos and Caçapava regions.

Information from 33 deep tubular wells located in the regions of the Jacareí, São José dos Campos, Caçapava and Lorena cities (Fig. 2.3) was used in the first step. The geophysical well logs were available by the *Companhia de Saneamento Básico do Estado de São Paulo* (SABESP) and contain gamma-ray (GR), resistivity (ILD), sonic (DT) and spontaneous potential (SP) curves.

In this study the GR, ILD and DT curves were used because they give a better response to the lithological variation described in the well. The SP profiles presented low resolution for the majority of the wells. The GR profiles record the rock natural radioactivity and it is related to the clay content. The DT profile records the time it takes for a sonic pulse to cross an interval of the formation with values directly related to the rock porosity. The resistivity profile indicates the presence of fluids in the rock using the resistance offered to the electric current (Thomas 2004).

The well data were digitized and adapted to the application of the Neural Networks method to determine the electrofacies. The classification carried out based on the well logs was unsupervised, in other words, not guided by core data. Thus the Artificial Neural Network adequate to this approach is SOM.

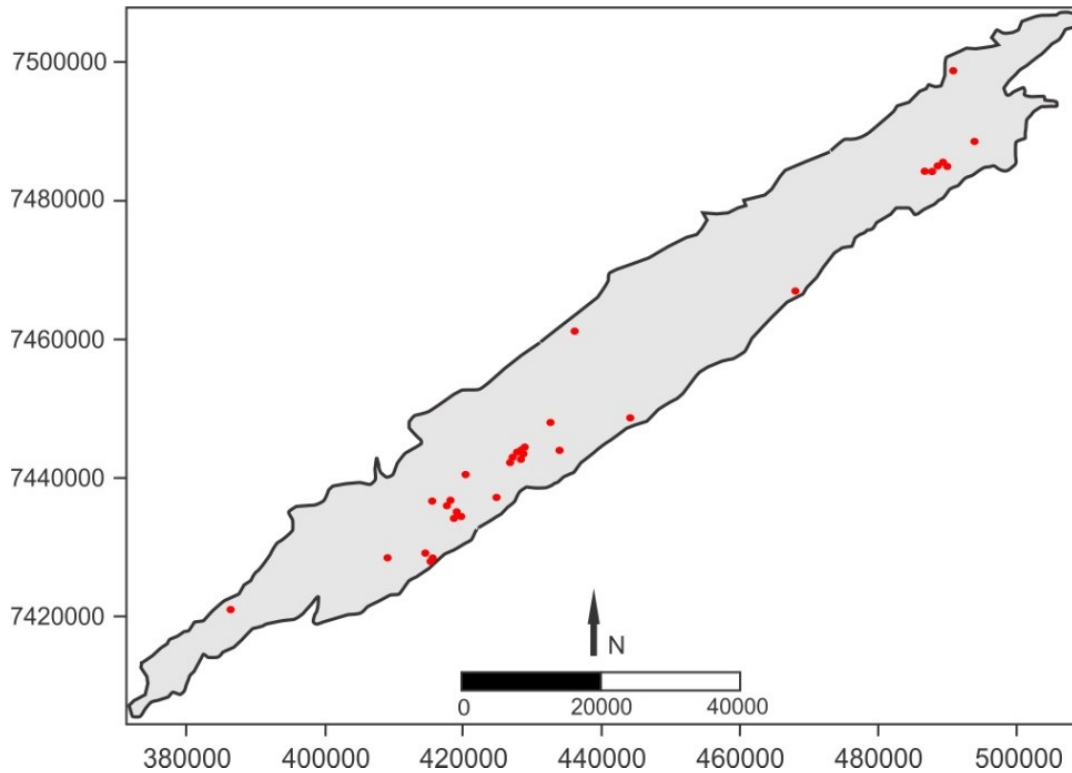


Figure 2. 3: Location of the study area. In detail the Taubaté Basin in São Paulo State.

The proposed classifications were obtained applying the software *Matlab*®, version 7.6.0. For several profile training events the tool developed by Vesanto *et al.* (1999) named “Self-Organizing Map in Matlab” was used.

After the application of SOM, the results obtained were analyzed by the ANOVA (Variance Analysis) statistical analysis. This algorithm is used to diagnose the average variation within or between clusters, using the hypothesis testing and F distribution. From this analysis it is possible to infer whether there is interaction between variables or clusters, constructed by SOM. Therefore the ANOVA analysis was also carried out to validate the clusters found using the study of the variation of the mean values of the variables involved, diagnosing the possibility of interaction between them.

To illustrate the values of the variables for each cluster box plot-type graphs were generated, checking the behavior of the profile attributes.

The resulting classifications were also analyzed according to geological parameters. Firstly an analysis in view of the values of GR, ILD and DT variables was carried out, seeking compatibility between the classification and the expected well log behavior. To reach this goal, the log data and the electrofacies defined were imported to the geological modeling software Petrel®.

Then the electrofacies were analyzed in relation to the well cuttings descriptions. Out of the 33 wells used in the classification, 27 had such descriptions, thus allowing evaluate the correlation between the results obtained by SOM and the well cuttings description.

The results were analyzed in greater detail for the wells of the São José dos Campos and Caçapava regions. The main goal of this step was to correlate the facies found in the classification and associate them with the seismic sections available for this portion of the basin.

Thirteen wells, grouped according to their location, and two 2D seismic sections, acquired by PETROBRAS in 1988 and made available by ANP – *Agência Nacional de Petróleo*, were used (Fig. 2.4). Regarding the geophysical well logs, the electrofacies were digitized and their lateral continuity was interpreted according to similarity patterns of the curves between different wells.

The analysis of the seismic sections was performed using the OpendTect® software, which allows the visualization and interpretation of 2D and 3D seismic data. The seismic expression of the crystalline basement was assumed, between reflections of disordered character and horizontally parallel reflections, these representing the sedimentation of the Taubaté Basin.

The division of the sedimentary column in seismic units was based on Fernandes (1993). Four seismic units were recognized in the sections used in the present work, named after Fernandes (1993) B, C, D and E. These units were associated with the patterns depicted during the interpretation of the geophysical logs.

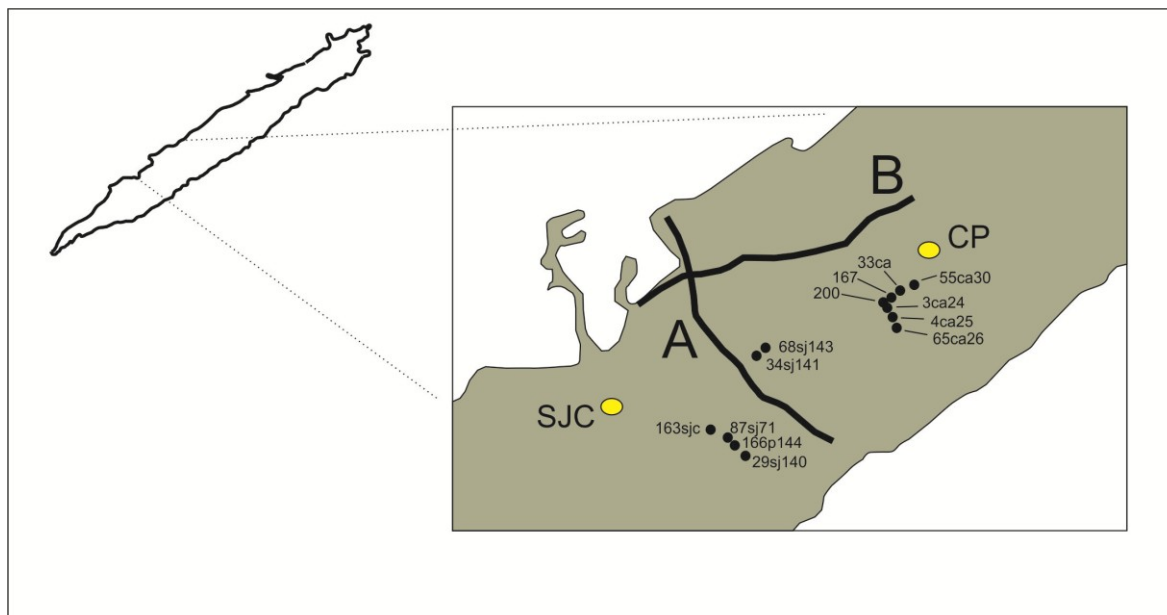


Figure 2. 4: Location of correlated wells and interpreted seismic profiles.

2.5. RESULTS AND DISCUSSIONS

Tests with SOM were carried out, considering two, three and four facies. The least mean square error of the simulations was found for two facies, with values between 11% and 5.4%, whereas for three and four facies the error intervals were: 15% to 16% and 19% to 23%, respectively.

The best classification found involved two distinct electrofacies that, after the analysis of the results, were named reservoir and non-reservoir (Fig. 2.5). The reservoir electrofacies is characterized by low gamma-ray and high resistivity values, relative to intervals of low-salinity (potable) water. As the majority of the deposits is composed of siliciclastic sediments, the low GR values represent clean sandstones of aquifer intervals, by the association with the resistivity profile.

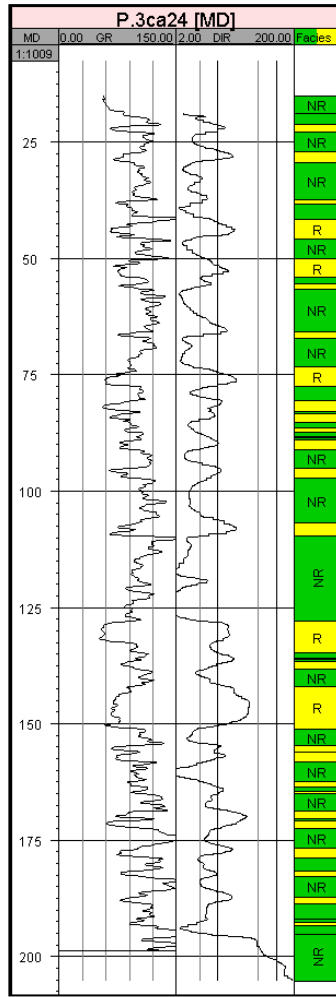


Figure 2. 5: Example of classification in reservoir (R – yellow) and non-reservoir (NR – green) facies for well 3ca24.

The application of the ANOVA method yielded expected results. The p-value found for the variables was zero, strongly suggesting variation of mean values. The same result was observed for the behavior of the mean values between the groups found with SOM. There is strong evidence of variation of the mean values between the reservoir and non-reservoir groups, with a null chance of being equal, confirming the good performance of the clustering algorithm chosen for the implementation of this work.

The results of the statistical analysis showed strong indication of interaction between the variables and the groups, suggesting that reservoir and non-reservoir have alterations of the values of the DT, ILD and GR variables distinctly, which is clearly observed in the box-plots analysis.

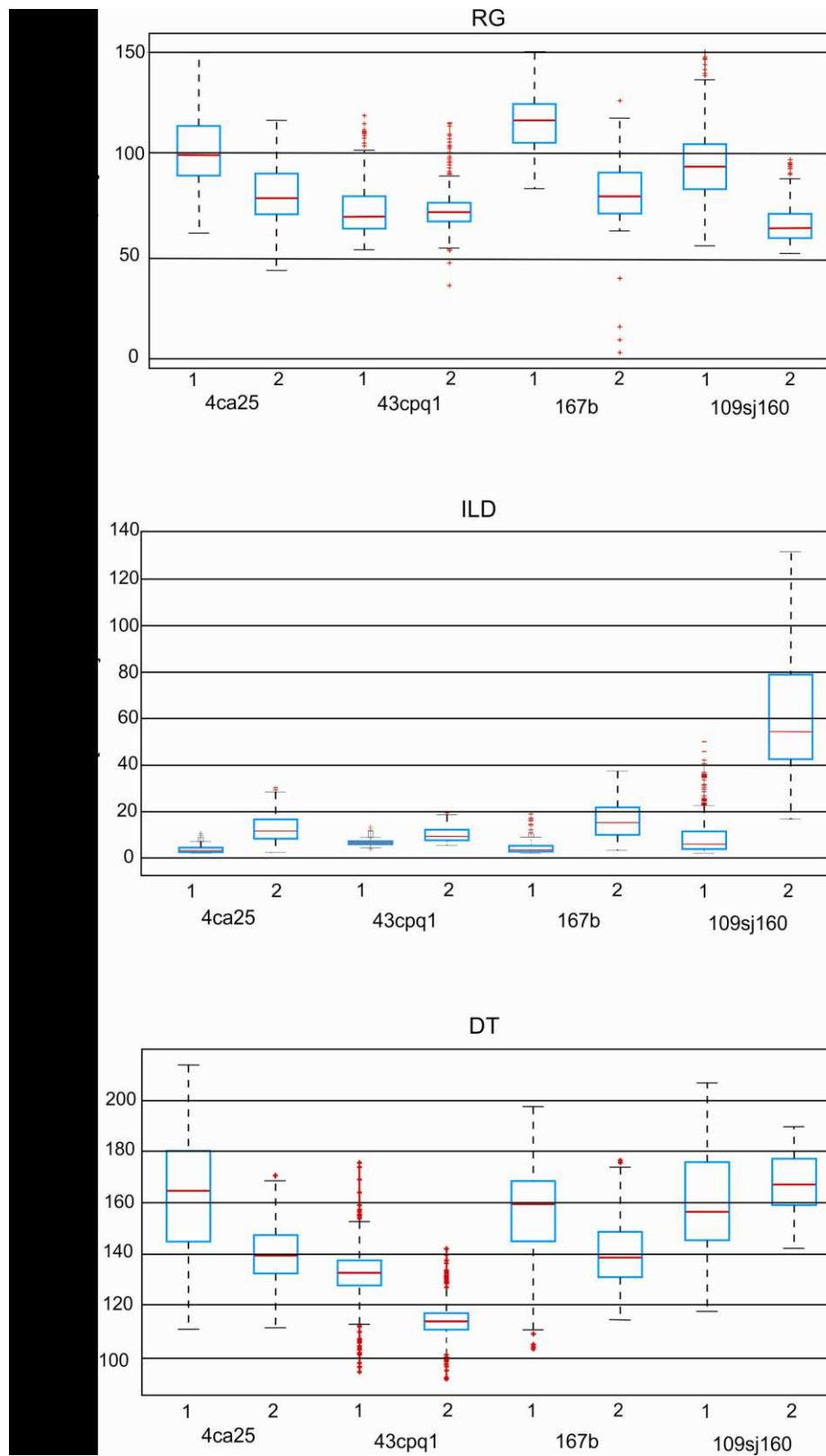
The box-plots diagrams present the values for reservoir and non-reservoir GR, ILD and DT variables. The amplitudes of variation between different attributes are observed in these graphs and the influence of each variable in the decision of separation by groups can be defined.

Fig. 2.6a presents the box-plots for GR for wells 4ca25, 43cpq1, 167b and 109sj160. Wells 4ca25 and 167 are located close to Caçapava, well 109sj160 close to São José dos Campos, and well 43cpq1 close to Lorena.

Excepting well 43cpq1, the wells presented well-discriminated GR values between the classes. It is possible to attest that for well 43cpq1 the dominant variable for the two-class clustering is profile DT, the only dimension where there is a clear difference between the classes (Fig. 2.6c). It is also observed that among the variables of this study, GR is that of largest influence for the separation of wells 4ca25 and 167b.

Regarding resistivity (Fig. 2.6b), a significant variation between classes was observed only for well 109sj160, representing a very particular case in which the variation of points is extremely high, even causing many points to fall outside the expected confidence interval, which is rare in other cases.

Although resistivity was the variable that yielded the largest significant difference between classes for well 109sj160, this well was the one that presented the least variability when analyzing DT graphs. On the other hand, the other three wells presented very distinct values between classes.



It is clear in these examples that each variable presented a different and particular behavior in each well, making the visual classification difficult and the use of an automatic classification tool essential.

As the visual classification, the statistical classification has certain limitations, because many variables present a normal classic behavior, making the definition of representative value groups for the lithological attributes impossible. Therefore, by applying multivariate techniques it is possible to define the correct influences of each variable of the study.

Compared to the well cuttings descriptions, the electrofacies presented satisfactory results. Despite the imprecise descriptions, this analysis helped check the existence of lithofacies characterizing the classification in two electrofacies, attesting the efficiency of the proposed method. The analysis also helped check the lithofacies that represent the reservoir facies.

On the basis of the well cuttings descriptions, of the lithofacies described by Riccomini (1989) and (Table 2.1), and associating permeability-porosity characteristics with grain size and selection, the reservoir facies can be represented by the conglomeratic sandstones, coarse- and medium-grained sandstones, greenish sandstones and fine-grained sandstones. The non-reservoir facies is represented by conglomerates, diamictites, siltites, lamites, greenish argillites, rhythmities and shales.

The sandy lamites, sandstones with abundant clayey matrix and fine-grained sandstones intercalated with lamites may represent transitional lithofacies between reservoir and non-reservoir electrofacies.

It is worth mentioning that the classification of the electrofacies is made difficult by the resolution of profiles with sampling points every 20 cm and intercalations within thinner intervals.

Table 2. 1: Lithofacies identified by Riccomini (1989) in the Taubaté Basin.

CODE	FACIES
A	Diamictites and conglomerates with mud- and arcosean sandy-matrix
B	Sandy mudstones and greenish sandstones (interdigitated with facies A)
C	Massive green claystone, locally with carbonated concretions
D	Dolomitic limestones intercalated with clays of facies C
E	Ritmites alternated with dark brown shale and marl
F	Medium to coarse sandstone with cross bedding, in the basal layers may occur conglomerates
G	Arcosean coarse conglomeratic sandstones ,graded to fine sandstones and at the top bioturbated siltstones
H	Coarse conglomeratic sandstones,with fining upward to siltstones and mudstones
I	Conglomeratic sandstones with abundant clay matrix at the base, up to siltstones and mudstones at the top
J	Sandstone, medium- to coarse-grained, with fining upward to silt- and mudstone
K	Calcified crusts associated with facies B
Q	Basal conglomerate with sandy matrix and fining upward

Out of the 33 wells characterized by SOM, 13 were selected for facies correlation, according to the location and the distance between them (Fig. 2.4). The reservoir facies lateral continuity was interpreted according to similarity patterns recognized in the GR, ILD and DT curves. The profile DT curve was essential to identify similar patterns between the wells.

Due to the lack of core data and the imprecise well cuttings descriptions, the correlation was restricted to the two defined facies instead of a detailed identification of faciological variations.

The wells are 500 m and 800 m far from each other and the correlated facies do not represent tabular and continuous layers, but intervals where canal deposits predominate, which cannot be characterized in detail because of lack of information and the large distances between wells. The main objective of this type of interpretation was to find similar patterns between depositional units.

Three main sedimentary units are observed in the resulting correlation (Fig. 2.7). In the lower sedimentary unit the geophysical logs show alternating reservoir and non-reservoir layers, with very expressive (high) ILD and (low) RG peak values. Sometimes representative, more argillaceous levels occur, but the percentage of reservoir rocks is more expressive. The mean thickness is *ca.* 100 m, and in wells that reach the basement (29sj140 and 166p144) this unit is from 40 m- to 100 m-thick.

In the intermediate unit the resistivity profile records a large decrease in resistivity, coupled with an increase in gamma-ray values (Fig. 2.8). In some profiles it is possible to distinguish a specific pattern of resistivity drop, reaching minimum values, as well as similar patterns in the sonic profile curve. This unit, present in all profiles, is interpreted as a clayey level, with thickness between 20 m and 50 m.

In the upper unit D, the profiles show again intercalations of reservoir and non-reservoir facies with thickness varying between 100 m and 150 m. In the Caçapava region the reservoir facies is thinner, having a higher concentration of clayey sediments. On the other hand, wells located in the flexural margin present a higher percentage of the reservoir facies.

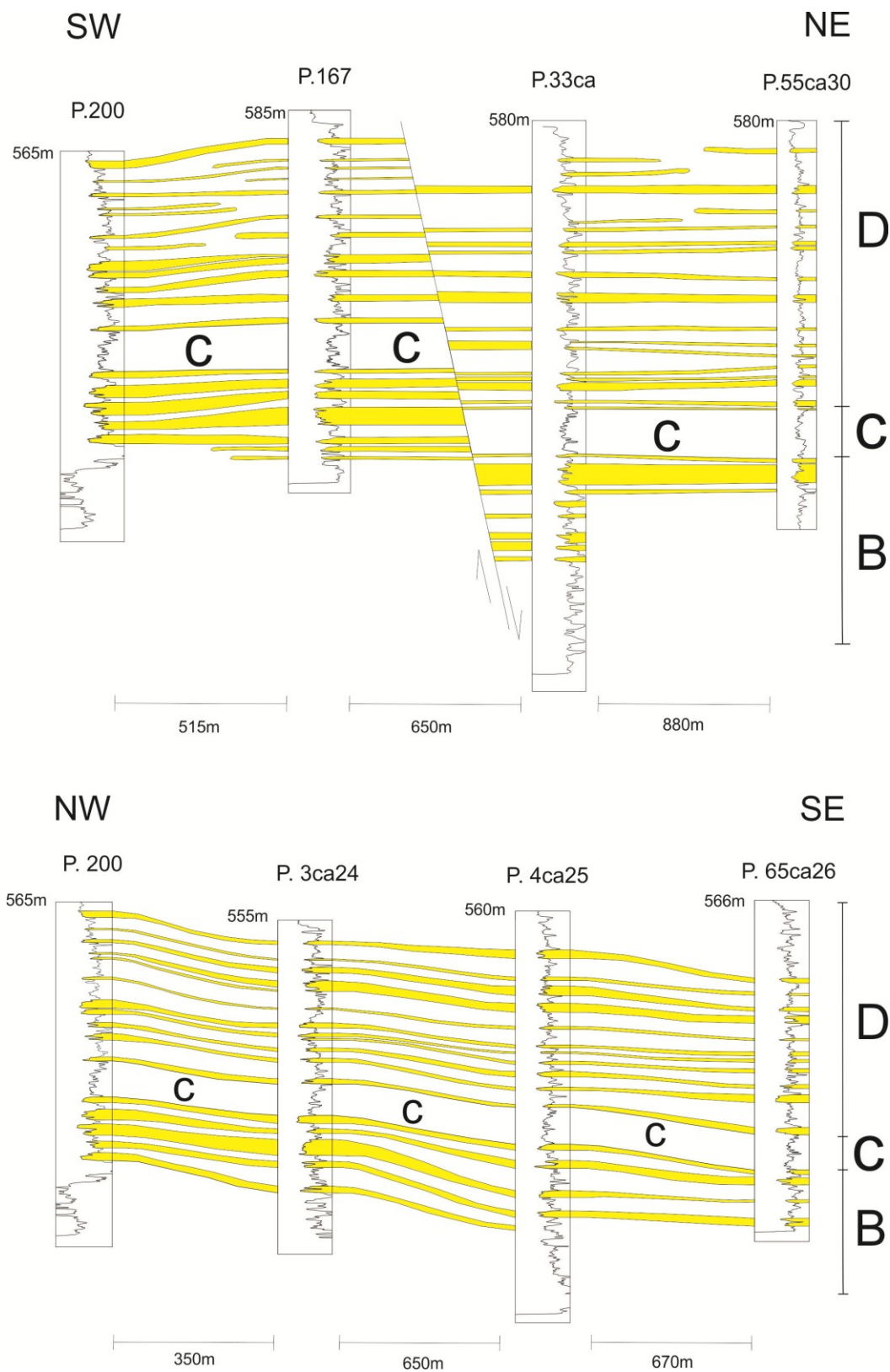


Figure 2. 7: Correlation of the reservoir facies in wells close to Caçapava city.

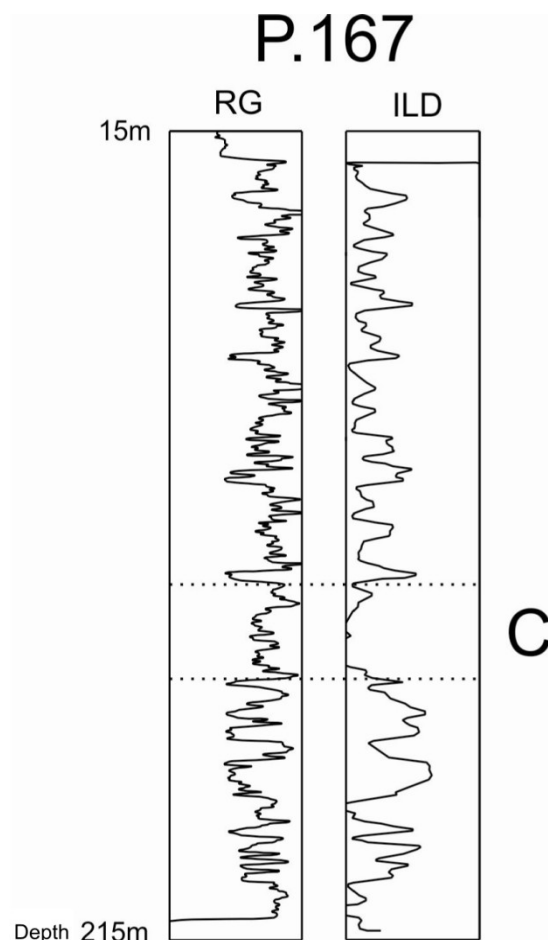


Figure 2. 8: Geophysical log for well 167, exemplifying the standard response of the unit C RG and ILD variables for the analyzed wells.

In view of the location of wells respective to the seismic profiles of the São José dos Campos Compartment, a correlation of the sedimentary units defined in the profiles with the seismic units was possible.

For the interpretation of the seismic profiles, the sedimentary units defined in the geophysical logs and the Fernandes's (1993) proposal were taken into account. In general the seismic units defined by the author coincide with the units defined here, so that the same nomenclature was used. Additionally, the horizons of each unit were reinterpreted, and therefore the thickness obtained is different from those of Fernandes (1993). The interpreted seismic units and the correlated wells are presented in Fig. 2.9 and Fig. 2.10.

The lower sedimentary unit defined in the wells for compartment São José dos Campos was associated with seismic unit B, which overlies the basement in this portion of the basin. The

characteristics of the geophysical logs and seismic profiles for this unit match the Fernandes's (1993) interpretation, who associates it with fluvial system and fluvial-lacustrine deposits.

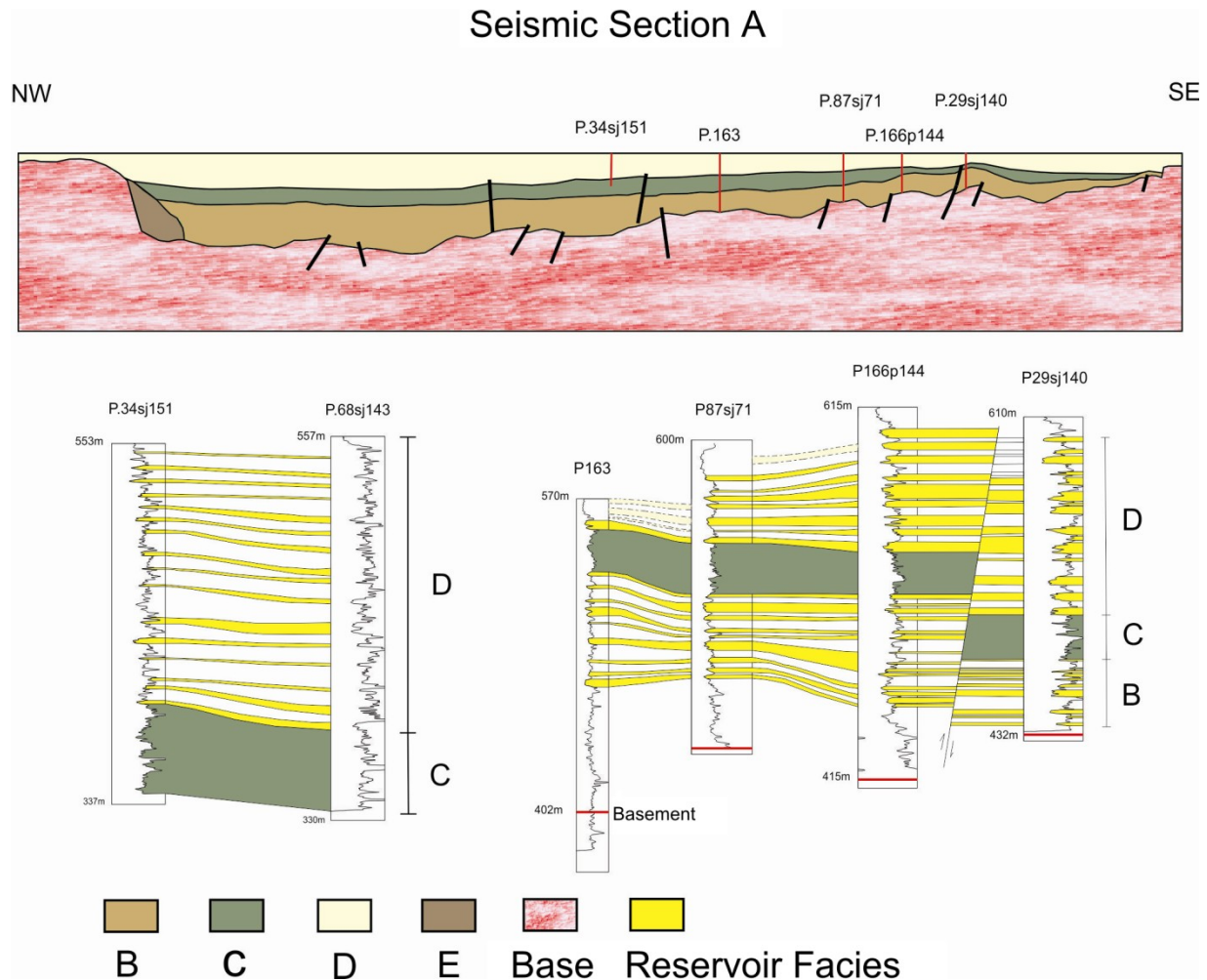


Figure 2. 9: Correlation of the reservoir facies and distribution of the sedimentary units interpreted in the seismic section A and in the electric logs.

The intermediate unit was associated with seismic unit C. In the São José dos Campos Compartment, seismic unit C presents two types of reflectors where subsidence is stronger, close to the border fault. There the seismic facies are parallel and well pronounced, but because of the increasing distance towards the flexural margin, the seismic facies are sub-parallel to chaotic. These characteristics are associated with a less marked deposition of clayey sediments towards the flexural margin, which is expected because of a less expressive subsidence. This fact is also evidenced by the thinner unit C in the analyzed wells.

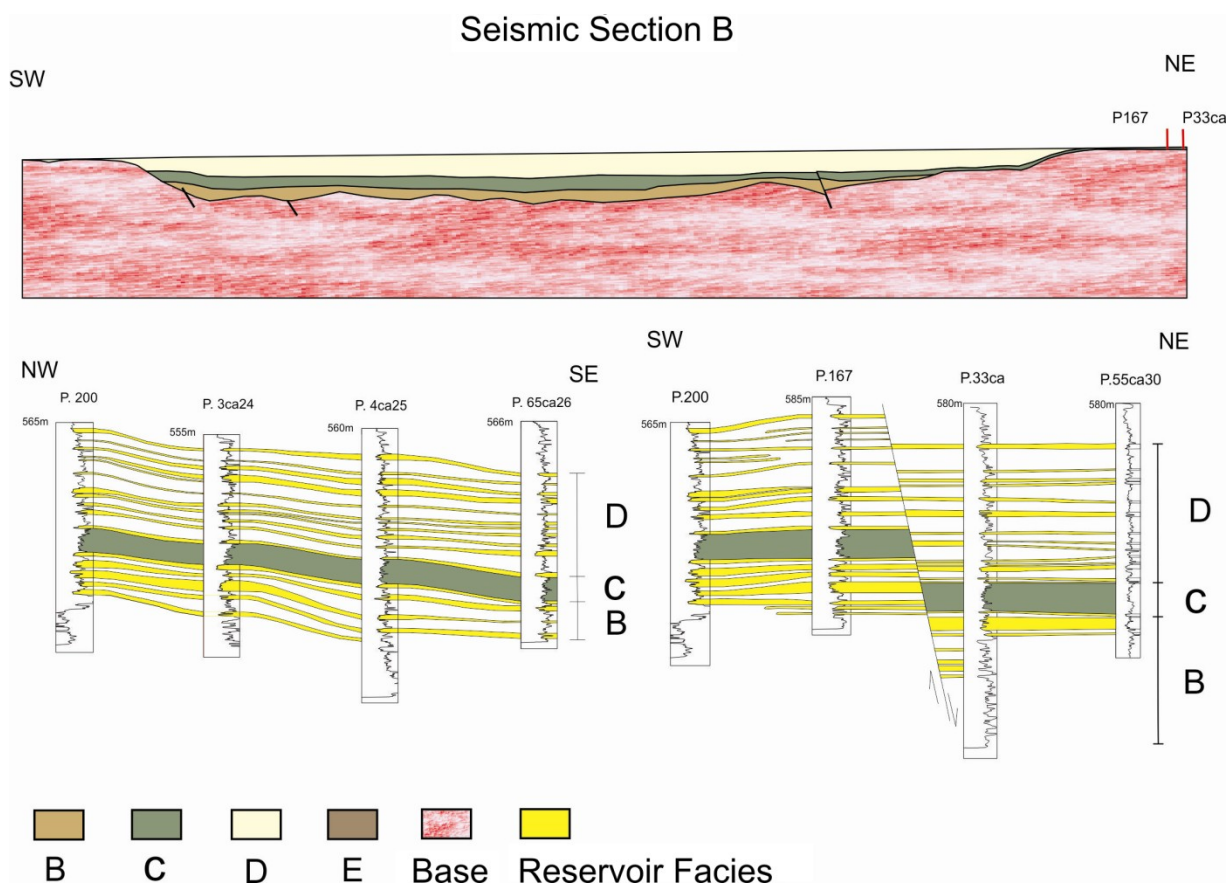


Figure 2. 10: Correlation of the reservoir facies and distribution of the sedimentary units interpreted in seismic section B and in the electric logs.

According to Fernandes (1993) the unit C seismic pattern northeast of the Caçapava High, with excellent lateral continuity and well marked reflectors, is compatible with rocks having a very continuous lateral bedding, characterizing sedimentation in lacustrine environment. Fluvial plain deposits must have occurred in the southwestern portion of the Caçapava High, as indicated by the seismic pattern with less lateral continuity. Still according to Fernandes (1993), in view of the seismic response, during the deposition of unit C the sedimentation spread out throughout the compartments, overlying the Pindamonhangaba High and part of the Caçapava High.

According to the geophysical logs, this unit surpasses the Caçapava High reaching portions close to the São José dos Campos city. In the São José dos Campos Compartment unit C appears as a thin package of clayey sediments, which is more expressive close to the border fault, associated with a lower-energy depositional environment.

In this portion of the basin, these deposits can be associated with a smaller lacustrine environment, contemporaneous to the deposition of the Tremembé Formation lacustrine sediments, which are conspicuous in the central portion of the basin (Taubaté Compartment).

The upper unit is associated with seismic unit D. In the São José dos Campos Compartment the seismic facies of this unit are sub-parallel to chaotic, which is consistent with the profile response, with expressive reservoir facies and non-reservoir intercalations. This pattern suggests a higher-energy depositional environment similar to the unit B depositional environment. The flexural margin deposits contain thicker reservoir facies layers, as shown by wells 166p144 and 29sj140, whereas the deposits closer to the border fault contain finer layers, as in well 34sj151.

Analyzing the relationship between the units established in seismic profiles and geophysical logs, as well as the location of wells *versus* seismic profiles, it is possible to infer the preferential direction of the sediment deposition in the São José dos Campos Compartment.

The São José dos Campos Compartment is a half-graben with the floor inclined northwestwards, against the São José master fault (Riccomini *et al.* 1989; Marques 1990), being well characterized by seismic profiles A and B. In view of its characteristics, the expected general sediment transport direction is to the basin axis (NW-SE). However, the units identified in the wells close to Caçapava city (Fig. 2.4) correlate with the units of the half-graben central portion, indicating that the sedimentation occurred concomitantly towards this portion of the basin, also characterizing transport from west to east (Fig. 2.11).

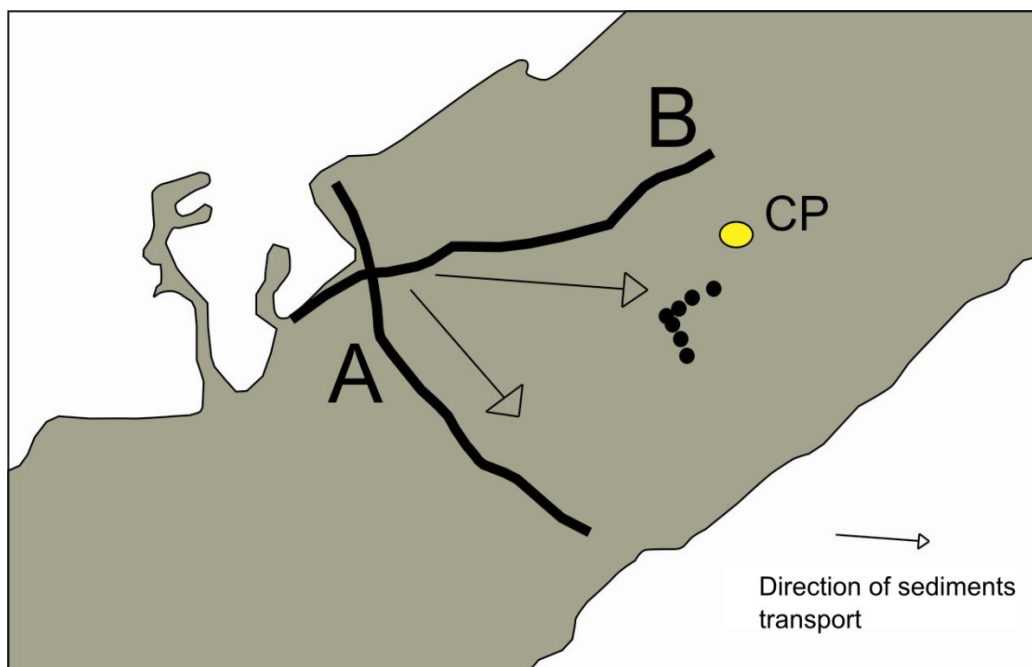


Figure 2. 11: Representation of the sediment transport direction in the São José dos Campos Compartment based on the analysis of wells and seismic profiles

2.6. CONCLUSIONS

As shown by the results and analyses, the SOM method was satisfactory in the characterization of facies from geophysical log data. Due to the lack of significant information, such as core data, this method is an effective alternative, also aiding a qualitative evaluation of the resulting clusters. The geophysical log data were grouped in two distinct classes, in agreement with the response expected for variables RG, ILD and DT for reservoir and non-reservoir rock intervals.

The resulting classification enabled a detailed faciological analysis of the São José dos Campos Compartment, integrating correlation of wells with seismic profiles interpreted in the region. The correlation of wells defined the existence of three distinct sedimentary units, seismic units interpreted, consistent with previous studies, reinforce the proposal of three main depositional units for that portion of the basin.

The lower unit (B) overlies the basement and contains intercalations of reservoir and non-reservoir facies, either with more argillaceous levels or with expressive presence of the reservoir facies. The characteristics of the profiles and seismic units suggest association of fluvial plain and

alluvial deposits, which concords with Riccomini(1989) proposal for the distal alluvial fan deposits of the Taubaté Group basal unit.

The intermediate unit (C) stands out for a sudden drop of ILD values followed by higher RG values, characterizing a sedimentary package with high proportion of non-reservoir facies. The response from electric logs agrees with the seismic response, which mostly presents parallel and continuous reflectors. This unit was interpreted as a lower-energy sedimentary deposit contemporaneous with the lacustrine deposits of the central portion of the basin, associated with its major subsidence phase.

The upper unit (D), which is the most recent expression of sedimentation, again presents reservoir and non-reservoir facies intercalations, with higher proportions of the reservoir facies towards the flexural margin of the half-graben that defines this compartment. This unit is associated with fluvial deposits, which is more conspicuous towards the southeastern and southern border of the basin, the latter in the region of the Jacareí city. These results agree with Riccomini(1989) proposal for the transition from fluvial deposits associated with alluvial fans to meandering fluvial deposits of the São Paulo and Pindamonhangaba formations, axial to the basin.

The correlation of wells indicates in the São José dos Campos Compartment sediment transport towards the basin axis and of the central portion of the half-graben eastwards, towards Caçapava. Besides, it was possible to characterize the main faults of the São José dos Campos Compartment, with indication in the wells for the extension of the Quiririm and Bom Retiro faults, as well as evidences of faults defined by Riccomini (1989).

REFERENCES

- ALMEIDA, T. C. 2007. Depositional architecture of deposits of alluvial fans in an intracontinental rift, Taubaté Basin, São Paulo. Rio de Janeiro, Msc Thesis, Universidade Federal do Rio de Janeiro.
- APPI, C.J., FREITAS, E.L., CASTRO, J. C. 1986. Faciology and stratigraphy of the Taubaté Basin. Internal Report. CENPES/PETROBRAS.
- BALDWIN, J.L., BATERMAN, R.M., WHEATLEY, C.L. 1990. Application of a neural network to the problem of mineral identification from well logs. *Log Anal.* 3, 279–293.
- BHATT, A.; HELLE, H. 2002. Determination of facies from well logs using modular neural networks. *Pet. Geosci.* 8 (3), 217–228.
- BISHOP, C. M. 1994. Neural networks and their applications. *Rev. Sci. Instrum.* 65 (6), 1803-1832.
- CARVALHO, A.M.A., VIDAL, A.C., CHANG, H.K. 2011. Definition of the Basis of Taubaté Basin. *Journal of Geosciences - USP*, v. 11, n.1, p. 19-32.
- CHANG, H.C., KOPASKA-MERKEL, D.C., CHEN, H.C., DURRANS, S.R. 2000. Lithofacies identification using multiple adaptive resonance theory neural networks and group decision expert system. *Computers & Geosciences* 26(5), 591–601.
- COLÉOU, T., POUPON, M., AZBEL, K. 2003. Unsupervised seismic facies classification: A review and comparison of techniques and implementation. *Journal The Leading Edge*, 22:942-953.
- CONSENTINO, L. 2001. Integrated Reservoir Studies. Paris: Technip, 310p
- COSTA J.A.F. -1999- Classificação Automática e Análise de dados por Redes Neurais Auto-Organizáveis. Dissertação de doutoramento, FEEC/Unicamp, 345p.
- FERNANDES, F.L. 1993. Structural framework and evolution of the Taubaté Basin – SP. Thesis, Universidade Federal de Ouro Preto, MG.

- FERNANDES, F.L. & CHANG, H.K. 2001. Gravimetric modeling of Taubaté Basin- Vale do Paraíba do Sul River, easts of São Paulo. *Journal of Geophysics*, 19:131-144.
- FREEMAN, J.A., SKAPURA, D.M. 1991. *Neural Networks: Algorithms, Applications, and Programming Techniques*. Addison-Wesley, Reading, MA, 401 p.
- GONÇALVES, M.L., NETTO, M.L. DE A., JÚNIOR, J.Z. 1996. A Modular Neural System for Classification of Images Using Kohonen Maps, In: VIII BRAZILIAN SYMPOSIUM ON REMOTE SENSING, 6th National Meeting on Artificial Intelligence, p. 845-849.
- JOLLIFE, I.T. 2002. *Principal Components Analysis*. 2nd ed., New Yotk, Springer Series in Statistics, 487 p.
- KOHONEN, T. 1982. Analysis of a simple self-organizing process. *Biol. Cybern.*, 44:135-140.
- KOHONEN, T. 2001. *Self-Organizing Maps*. Springer-Verlag, New York, 501 P.
- LI, Y.; ANDERSON-SPRECHER, R. 2006. Facies indentification from well logs: a comparasion of discriminant analysis and Naïve Bayes classifier. *Journal of Petroleum Science and Engineering*, v.56, p. 149-157.
- MANCINI, F. 1995. Stratigraphy and tectonic aspects of deforming Pindamonhangaba Formation, Taubaté Basin, São Paulo, São Paulo. Thesis, Institute of Geosciences, Universidade de São Paulo, São Paulo.
- MARQUES, A. 1990. Tectono-sedimentary evolucion and exploration prospects of the Taubaté Basin, São Paulo, Brazil. *Bulletin of Geosciences, Petrobras*, v. 4, n. 3, p. 253-262.
- RICCOMINI, C. 1989. The continental rift of southeastern Brazil. São Paulo. PHD Dissertation, Institute of Geosciences, Universidade de São Paulo, São Paulo.
- RICCOMINI, C., COIMBRA, A.M. 1992. Geology of the sedimentary basin. In: Negro, J.R.A., Ferreira, A.A.; Alonso U.R; Luz, P.A. (eds). *Soil of São Paulo City*. São Paulo: ABMS/ABEF, p. 37-94.

- RICCOMINI, C., SANT'ANNA, L. G., FERRARI, A. L. 2004. Geological evolution of the continental rift of southeastern Brazil. In: Mantesso-Neto, V., Bartorelli, A., Carneiro, C.D.R. *et al.*(eds). Geology of the South American continent: the evolution of the work of Gernando Marques de Almeida, 1st ed, São Paulo, Beca Cultural Productions Ltd., chapter 23
- RIPLEY, B.D. 1996. Pattern recognition and neural networks. Cambridge University Press, Cambridge, 403 p.
- STUNDNER, M.; OBERWINKLER, C. 2004. Self-Organizing Maps for Lithofacies Identification and Permeability Prediction. In: SPE ANNUAL TECHNICAL CONFERENCE AND EXHIBITION, Houston, p.8.
- TANG, H., WHITE, C.; ZENG, X.. GANI, M.; BHATTACHARYA, J. 2004. Comparison of multivariate statistical algorithms for wireline log facies classification. AAPG ANNUAL MEETING ABSTRACT, vol. 88, p. 13.
- TANG, H., WHITE, C. 2008. Multivariate statistical log log-facies classification on a shallow marine reservoir. J. Pet. Sci. Eng., 61: 88-93.
- THOMAS, J. E. 2004. Fundamentals of Petroleum Engineering. 2. Ed. Rio de Janeiro: Interciência: Petrobras.
- VESANTO, J., HIMBERG, J., ALHONIEMU, E.; PARHANKANGAS, J. 1999. Self-Organizing Map in Matlab: The SOM Toolbox. In: PROCEEDING OF THE MATLAB DSP CONFERENCE, p.59.

ANEXO 3

“Kuroda, M. C.; Vidal, A.C.; Carvalho, A.M.A. Interpretation of Seismic Multiattributes Using a Neural Network (submetido ao Journal of Applied Geophysics).”

INTERPRETATION OF SEISMIC MULTIATTRIBUTES USING A NEURAL NETWORK

Michelle Chaves Kuroda*, Alexandre Campana Vidal, Ancilla Maria Almeida de Carvalho

*Institute of Geosciences, P.O.Box 6152, University of Campinas – UNICAMP, 13083-870,
Campinas, SP, Brazil. Tel: +55-19-3521-4659; Fax: +55-19-3289-1562. E-mails:
mckuroda@ige.unicamp.br, vidal@ige.unicamp.br, carvalho.ancilla@gmail.com,*

ABSTRACT - Bodies in 2D seismic section are characterized by differences from the surrounding response. These differences can be highlighted by attributes that are sensitive to the desired feature. In this paper the attributes were carefully chosen and trained by a neural network. These groups are transformed into a new attribute that allows the optimal view of the seismic lines. The database used for this study is 2D seismic lines of the Taubaté Basin, São Paulo, Brazil. Two seismic groups were analyzed and the results reveal the lateral continuity and the boundary between seismic units, which helps a better understanding of the evolution of the Taubaté sedimentary basin.

Keywords: Seismic attributes, neural network, Self-Organizing Map, Taubaté basin.

3.1. INTRODUCTION

In the first stage of reservoir characterization, the structural model is predominantly based on the 2D and 3D seismic reflections. Since the first digital seismic data acquisition, the amount of new information from the original signals, called seismic attributes, has considerably increased, allied with the development of computational and statistical techniques for evaluating multiple variables (Chopra & Marfurt, 2006; Taner 2001).

Such data can represent geometric content, kinematic, dynamic or statistical characteristics, therefore having different sensitivities to events, making the description and characterization of specific data signals possible, and enabling visual alternatives to the interpreter (Chen & Sidney, 1997; Taner, 1992).

Today the success of the seismic interpretation is directly linked to combinations of attributes enhancing features of different signals, which allows the grouping of similar lithologies and petrophysical proprieties, making the interpretation simpler than using only the original data (Ashida, 1996; Chen & Sidney, 1997; Chopra & Marfurt, 2006; Forte *et al.*, 2011; Meldahl *et al.*, 2001; Miller & Powell, 2001; Rezvandehy *et al.*, 2011; Taner, 2001).

For these reasons the seismic reflection method currently represents more than 90% of the investments in exploration. The creation of new features makes a better understanding of the geology possible and assists in minimizing uncertainty through a more detailed signal representative of structural, lithological and stratigraphic properties (Chopra & Marfurt, 2006; Taner, 1992; Taner, 2001).

On the other hand, the process of seismic interpretation is still a time consuming manual task and because it is based on natural data, this area of engineering is concerned not only with calculations and numerical analysis, but also with concepts, perception, judgment, empirical relationships and employment of experience that cannot be represented strictly numerically (Ferentinou & Sakellariou, 2007). For these reasons the interpretation carries a high level of uncertainty.

The seismic attribute data are extremely complex, extensive and normally do not follow a normal distribution. These characteristics do not allow the application of the traditional multivariate statistical analysis (Sheline, 2005).

According to Miller & Powell (2001) the computational neural networking applications are promising for automation of some labor-intensive tasks in seismic data processing, such as trace editing, travel-time picking, and velocity analysis as well as applications in seismic interpretation. Poulton (2001) describe the use of self-organizing neural networks for facies mapping, which has been successfully applied in several fields, such as facies mapping, time-lapse seismic interpretation, log property prediction and reservoir characterization. Furthermore Zhang *et al.* (2001) relate this method for tracking horizons and classifying seismic traces by the analysis of the average amplitude waveforms and distribution of the total variability of the seismic response into different classes.

The proposal of this work is to apply the multivariate analysis of neural networks, Self-Organizing Map (SOM), to improve the process of seismic interpretation by assisting the interpreter in image enhancement with the use of highlighting, automatic identification of geological features and event correlation, which is achieved through a combination of different seismic attributes.

The amplitudes of the considered set of attributes are separated into groups or classes, according to the similarity of the signs. The advantage of this method is that it is sensitive to noise and therefore such interference can be included in a class representing geologic features.

Chopra & Marfurt (2006) describe the application of seismic attributes on automated identification of seismic objects and stratigraphic features. For this purpose, different attributes are combined to enhance the contrast between features of interest and their surroundings.

Meldahl *et al.* (2001) developed a workflow for the use of neural network. Their study combines multiple attributes into a new attribute that gives the optimal view of the targeted object, which indicates the probability of the object at a seismic position. The authors demonstrated how to classify the entire data set into fault and non-fault samples with a neural network.

In this current case study, 2D seismic lines along the Taubaté Basin, located in São Paulo State, Brazil, were used to be trained by SOM. Although some authors have already investigated some of the geological seismic features of this basin (Marques, 1990; Fernandes, 1993), this research aims to advance the understanding of image enhancement techniques and give more information about geology of the basin. The seismic data were obtained in 1994 in the continental area, thus including a high noise content. The use of SOM helped the interpretation of this data. The additional information from well logs was a useful guide for interpretation and validation of the results using SOM.

3.2. FIELD DESCRIPTION

3.2.1. GEOLOGICAL SETTING

The Taubaté Basin is one of the basins that composes the Continental Rift of Southeastern Brazil - CRSB - (Riccomini, 1989), a Cenozoic tectonic feature located between Curitiba city in the Paraná State and Barra de São João city in the Rio de Janeiro State. Positioned in the central segment of the CRSB, the basin is related to the Paleogene distensive tectonics that characterizes it as a rift basin.

The sedimentation in the Taubaté Basin is typically continental and, according to Riccomini (1989), it can be divided in two phases: the first, syn-tectonic to the rift, corresponds to the deposition of the Taubaté Group, formed by the Resende, Tremembé and São Paulo Formations. The second, posterior to the diastrophic tectonics, is represented by the Pindamonhangaba Formation and alluvial and colluvial deposits.

This basin presents ten seismic sections distributed axially and transversely to the main axis of the basin (Fig. 3.1). Marques (1990) shows the first analysis of the seismic sections by defining the main depocenters and faults, thereby generating a structural contour map of the basement. The author proposes the subdivision of the basin in four sub-basins named from southwest to northeast Eugenio de Melo, Quiririm, Roseira and Lorena. In the southern portion of the basin, not sampled by seismics, two sub-basins are recognized, Parateí and Jacareí with not very significant depocenters. Also three internal structural highs were recognized: Caçapava, Capela de Nossa Senhora do Socorro and Capela de Santa Luzia.

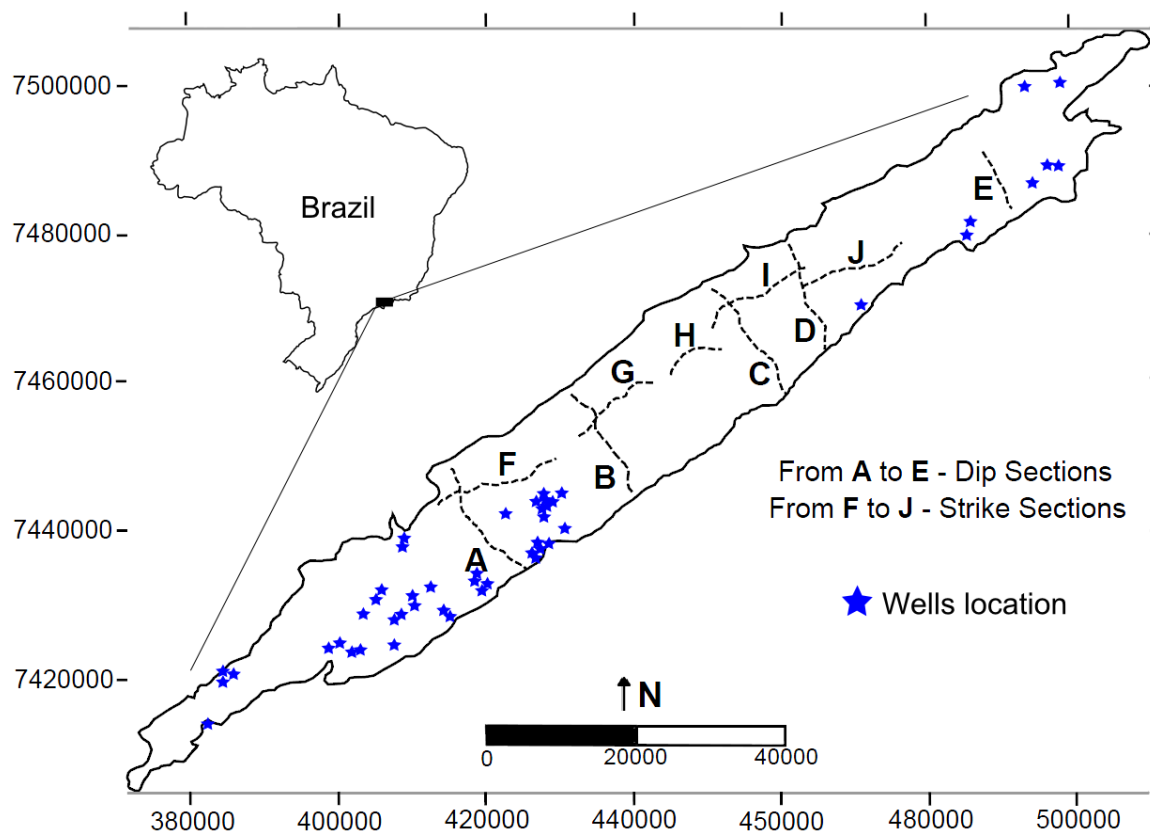


Figure 3. 1: Location of seismic sections in the Taubaté Basin.

Fernandes (1993) based on seismic and gravimetric data identifies four sub-basins, named from southwest to northeast Parateí, Jacareí-São José dos Campos, Taubaté-Quiririm and Aparecida-Lorena and transfer zones named São José Fault, Alto de Caçapava and Alto de Pindamonhangaba.

From the interpretation of sections, Fernandes (1993) also conducts a stratigraphic analysis of the sediments of the Taubaté Basin. Five seismic units are recognized and according to the author they are sets of reflectors with similar characteristics among themselves. These units allowed a better understanding of the sedimentary basin evolution (Fig. 3.2).

Unit A is an older unit that overlies basement rocks and is limited to the northeastern portion of the basin. It has an irregular wedge shape and interpenetrates fanglomeratic deposits close to border faults. According to Fernandes (1993), the progradational sismofacies can characterize tractive processes, characteristic of conglomerate and sandstone deposition.

Unit B overlies the unit A, has extensive distribution and is present in all compartments of the basin. In the northwestern and central portion of the basin, unit B directly overlies the basement and is in contact with fanglomerates close to the border fault. This unit is irregularly tabular or wedged. Its subparallel to chaotic seismic facies may be associated with a fluvial or fluvial-lacustrine system.

Unit C overlies unit B and its reflectors present excellent lateral continuity, being traced all over the basin. Its seismic facies vary from continuously parallel to slightly parallel and according to Fernandes (1993) they are characteristic of lower energy events in which suspension is predominant.

Unit D is the most recent expression of sedimentation in the basin. It is lensoidal in shape and its lower limit is with unit C. Its major development occurs in the central portion of the basin, where parallel seismic facies predominate; in other portions, the features of this unit are masked by the effect of seismic processing. According to Fernandes (1993), given the characteristics of the central portion the basin, the depositional environment of this unit varied slightly in relation to that of unit C.

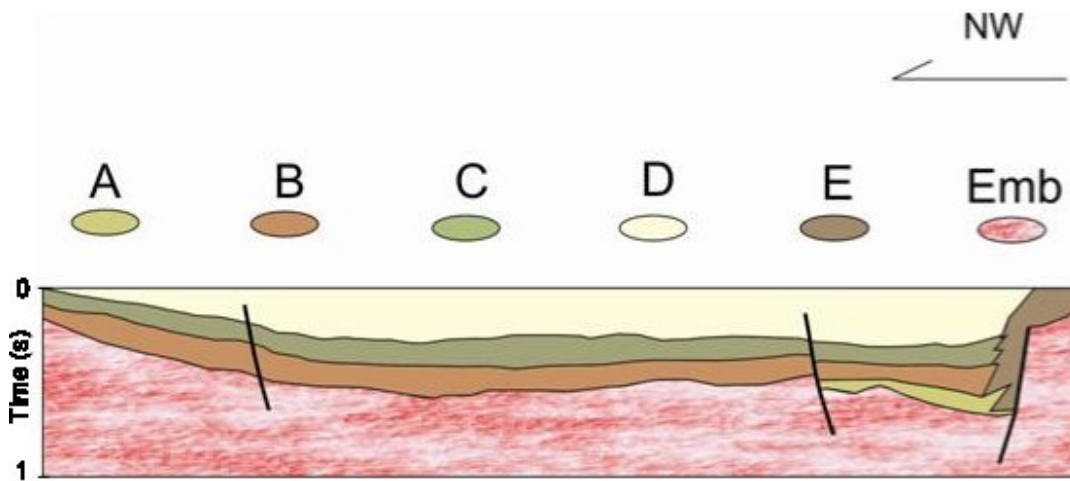


Figure 3. 2: Stratigraphic units of the Taubaté Basin. Modified from Fernandes (2003).

3.2.2. THE DATABASE

The Taubaté basin has 11 seismic lines in SEG-Y format that were analyzed in this study. These data were acquired by Petrobras in 1988 and issued by the ANP - National Petroleum Agency. From this data set various seismic attributes were calculated.

The seismic acquisition of Taubaté Basin, realized in July, 1988, was provided by Agência Nacional de Petróleo (ANP). There are eleven 2D lines, of which ten are available. In Brazil the South American Geodesic System, known as SAD69 is adopted. The projection used in this acquisition is the UTM – Transverse Universe of Mercator, utilized in the production of topographical charts and cadastre plans.

According to Marques (1990) the 178 km of seismic lines show that the framework of Taubaté Basin is defined by two major classes of structural features; dextrorotatory transcurrent, faults, reaching the entire sedimentary section, and oriented in the longitudinal direction of the basin and normal faults cross the basin, controlling horts and asymmetric grabens that define the six sub-basins.

Besides this information, 33 wells distributed throughout the basin were used. Through the use of these wells it was possible to more precisely define the horizons identified in seismic sections, in particular interval C. Gamma ray and resistivity logs allowed the identification of the intervals consisting of reservoir rocks and the main horizons identified by seismic.

The well data were underused because the lines are far apart and have a few profiles. The acquisition of 2D seismic data and the irregular distribution of the wells does not allow a structural interpretation of the basin, thus the correlation of wells were undermined (Fig.3.1).

The seismic data were loaded into the Opendtect ® DGB program, version 4.0.1 (<http://www.dgb.nl>), which is an interactive software for data analysis and loading, calculations and visualization of seismic data (Fig. 3.3). The data were imported and the different attributes were generated, and exported for multivariate analysis.

Multivariate analysis of the chosen attributes was performed with Matlab®, version R2010a, developed by MathWorks®. However, for loading and for viewing the results in SEG-Y

files, it is necessary to use two packages of functions: SeisLab ®, version 3.0 provided by Matlab®, and the free software SegyMat®.

The best combination of attributes depends on several parameters, such as the characteristics of the reservoir, the reliability of the data and especially the experience of the interpreter (Sheline, 2005). To guide the interpreter, the classification and definition of seismic attributes assist in the choice of the best combination to be used to refine the analysis. The original seismic data, as well as the seismic attributes, are related to the geological context of the study area, and therefore must be used with caution, because the wrong choice of attributes may produce results different from those expected (Taner, 2001).

The attributes used in this study are wavelet and instantaneous attributes. The instantaneous attributes are calculated sample to sample. They represent instantaneous variations of some parameters that can be derived from complex features, such as envelope trace, its derivatives, frequency and phase (Taner, 1992).

The instantaneous attributes are calculated at the peak of the envelope and are directly related to the Fourier transform, as the instantaneous frequency and instantaneous phase called Wavelet attributes (Taner, 2001).

According Taner (2001) these geological attributes can be categorized as Physical or Geometric. The Physical attributes are related to physical quantities and qualities and, therefore, often used for rating lithological and reservoir characterization. Geometric attributes describe spatial and temporal relations, used to obtain depositional information for stratigraphic interpretation.

After some tests of attribute combinations, we decided to use the combination of five attributes: Instantaneous Amplitude, Instantaneous Phase, Medium Frequency, Envelope Weighted Frequency and Envelope Weighted Phase.

The Instantaneous Amplitude is traditionally used for structural and stratigraphic interpretations and usually used as basis to others attributes derived from the Amplitude to isolate areas of high and low amplitudes as bright spots. Because it is a physical attribute, it gives

information as acoustic impedance, reflection coefficients, effects of absorption and speed, with content on the structural geology and direct indicator of reservoirs (Chen & Sydney, 1997).

The Instantaneous Phase is used to minimize noise and enhance intra-event signal reservoirs. The attributes derived from the phase are used to highlight the geometry and continuity of seismic reflectors, stratigraphic interpretation and depositional systems (Chen & Sydney, 1997).

The Instantaneous Frequency is a physical attribute used to discriminate reservoirs, accounting for the effects of wave propagation and depositional features. Low frequency values are associated with the fluid present in the porous regions, as well as possible fracture zones and thinner layers (Taner, 2001).

The combination of attributes related to frequency, phase and amplitude will be studied in order to improve the visualization of seismic lines from the Taubaté Basin, to define the main horizons by noise reduction and improvement of lateral continuity.

3.3. METHOD

The choice of seismic attributes was based on the Meldahl et al. (2001) workflow, ensuring that only one type of geologic object is targeted at the time and the selection of attributes with the potential to enhance the objects is then made. The trained network is applied to the entire seismic line to produce a new line with high values at positions where the object is recognized.

After loading all attributes, two data sets were created for testing and comparison:

Set 1: formed by the attributes Instantaneous Amplitude, Instantaneous Phase, Medium Frequency and Envelope Weighted Frequency.

Set 2: comprising attributes Instantaneous Amplitude, Instantaneous Phase, Envelope Weighted Frequency and Envelope Weighted Phase.

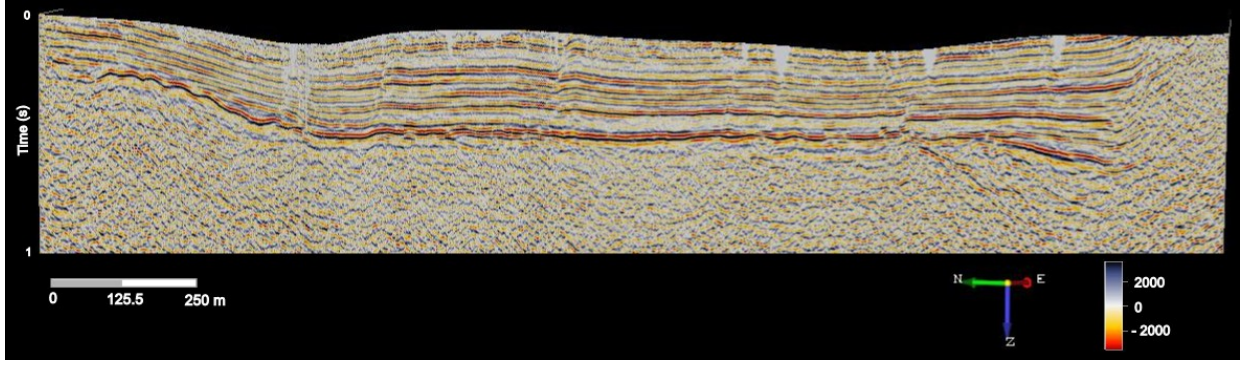


Figure 3. 3: Seismic Line "C". Picture taken from OpendText © software.

The seismic lines show a great deal of information about the basin basement bellow 1000 m, all seismic information being related to Precambrian rocks (Fig. 3.3). The area of interest for classification is restricted to the area above this timeline, so the entire region related to basement data was removed from the data set for analysis by SOM. Thus, the calculated seismic attributes line was reduced to the initial 1000 m, significantly reducing the size of the matrix.

In order to train the data set using the package SOM Toolbox, the attribute data had to be adapted for loading and training on the algorithm. For this, we used the technique described by Moraes (2004) and displayed in Fig. 3.4. The left part of the figure represents the matrix of seismic attributes used in this work and, the right side the matrix with relocation of these data for training by SOM.

Each of the colored circles represents the first element of each column, i.e., the first point of the seismic trace. These elements will be the first line of the new array. The second line is represented by the second point of the first column of each attribute matrix, and so on. Seismic section C, for example, contains 1445 traces and 751 points, of which 251 were analyzed. Thus the resulting matrix dimension for classification has $1445 * 251$ rows and four columns, in which each one represents an attribute, as shown in the equation below:

$$R((i-1)*1445+j,k) = ak(i,j), \text{ for } i = 1:251, j = 1:1445 \text{ and } k = 1:4, \quad (3.1)$$

where i is the row and j is the column, R is the new matrix and ak is the matrix of the k attribute, which could be one of the attributes described before.

For training, the data were normalized and the initial synaptic weight values were set between $[0,1]$, according to standard uniform distribution (Hasan & Tahar, 2006).

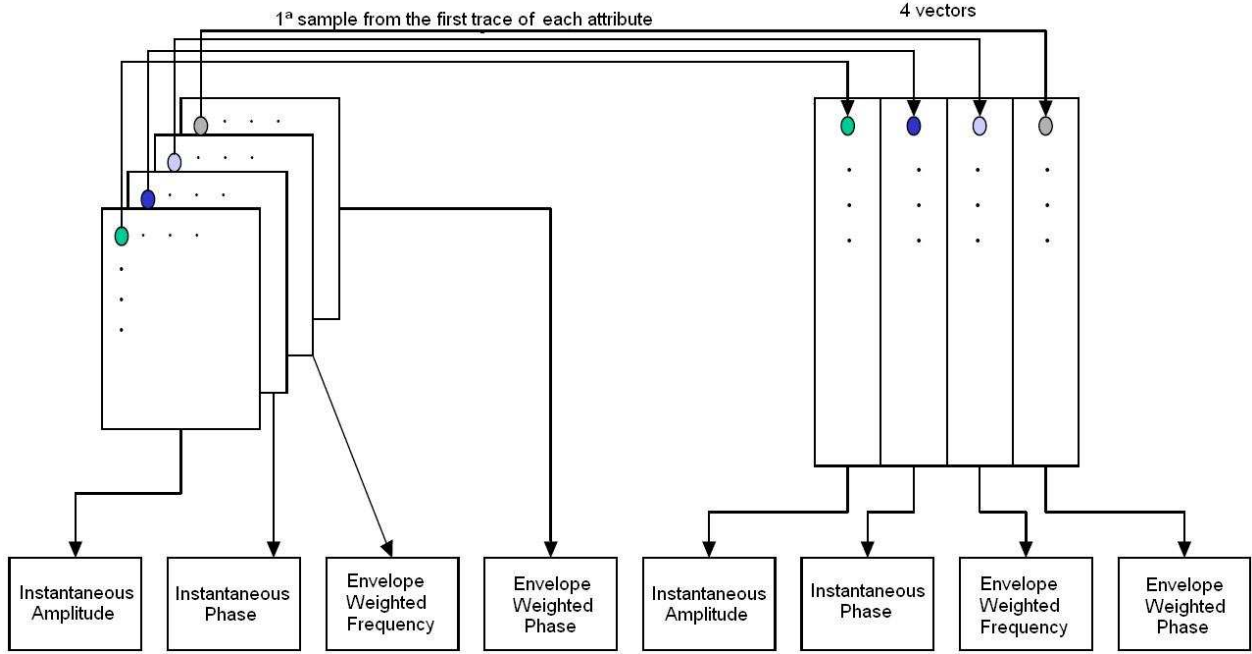


Figure 3. 4: Transformation matrices of seismic attributes for set 1. Adapted from Moraes (2004).

3.4. RESULTS AND DISCUSSION

This work shows the results of the seismic section C, located in the central portion of the basin. The first data set was trained with the SOM algorithm, for 3, 4 and 5 classes, defined by Daives-Bouldin (and explained in Appendix A).

In the package developed by Vesanto *et al.* (2000), training is conducted in two stages. In the first step $T = 1000$ (the number of iterations) and $\varepsilon = 10^{-3}$ (the stopping criterion), $\alpha_0 = 0.9$ (the rate of learning), $\sigma = \frac{\max(nx, ny)}{4} = 16.75$ (for groups 3 and 5) and 15.75 for group 4 and to the second step $\alpha_0 = 0.1$ and $\sigma = 1$. The parameters T , ε and α are default of the SOM toolbox used in this work.

The U-matrix created for a better view of the groups is 66×46 and 67×45 for sets 1 and 2, respectively. So i in Eq. (A – 2) is 3036 (the result of 66×46) for the groups of the first set and 3015 (the result of 67×45) for groups of the second set, described in Appendix A.

The results obtained with the classification for the various classes mentioned are shown below.

- Results for Set 1:

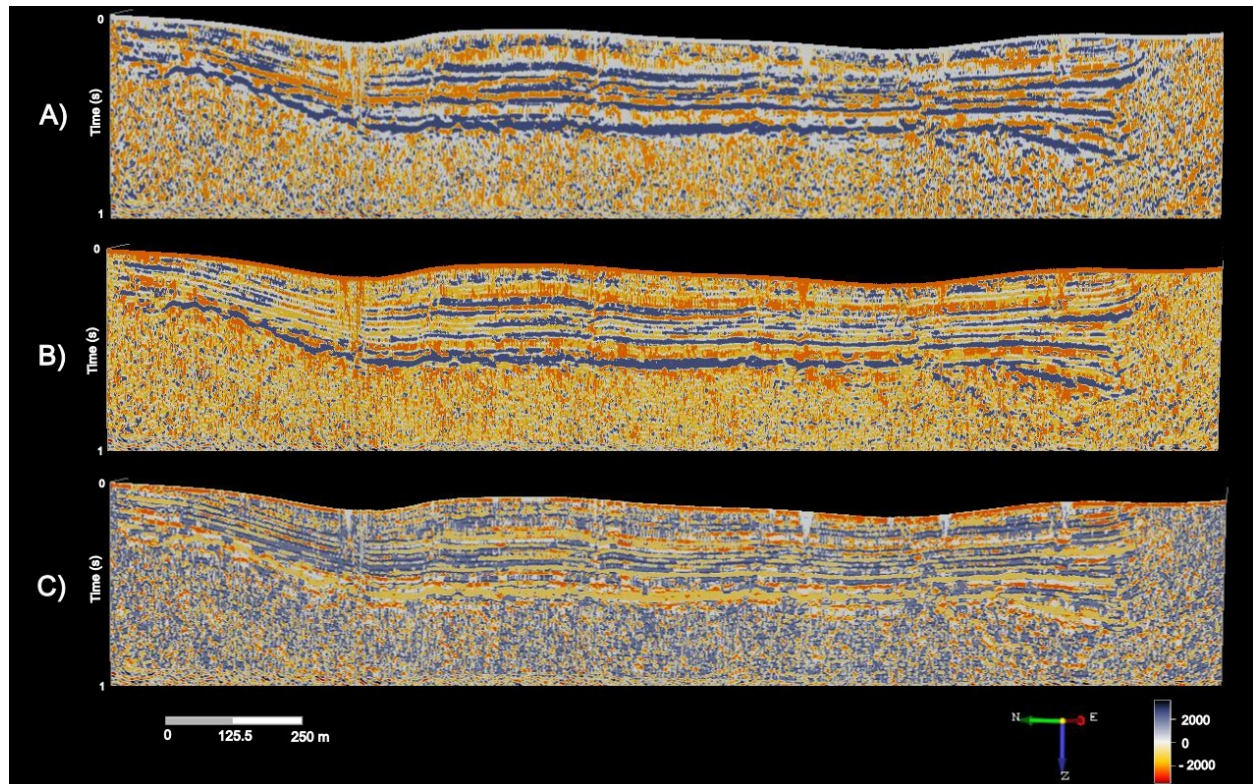


Figure 3. 5: Set 1 classified into 3 (A), 4 (B) and 5 (C) clusters.

- Results for Set 2:

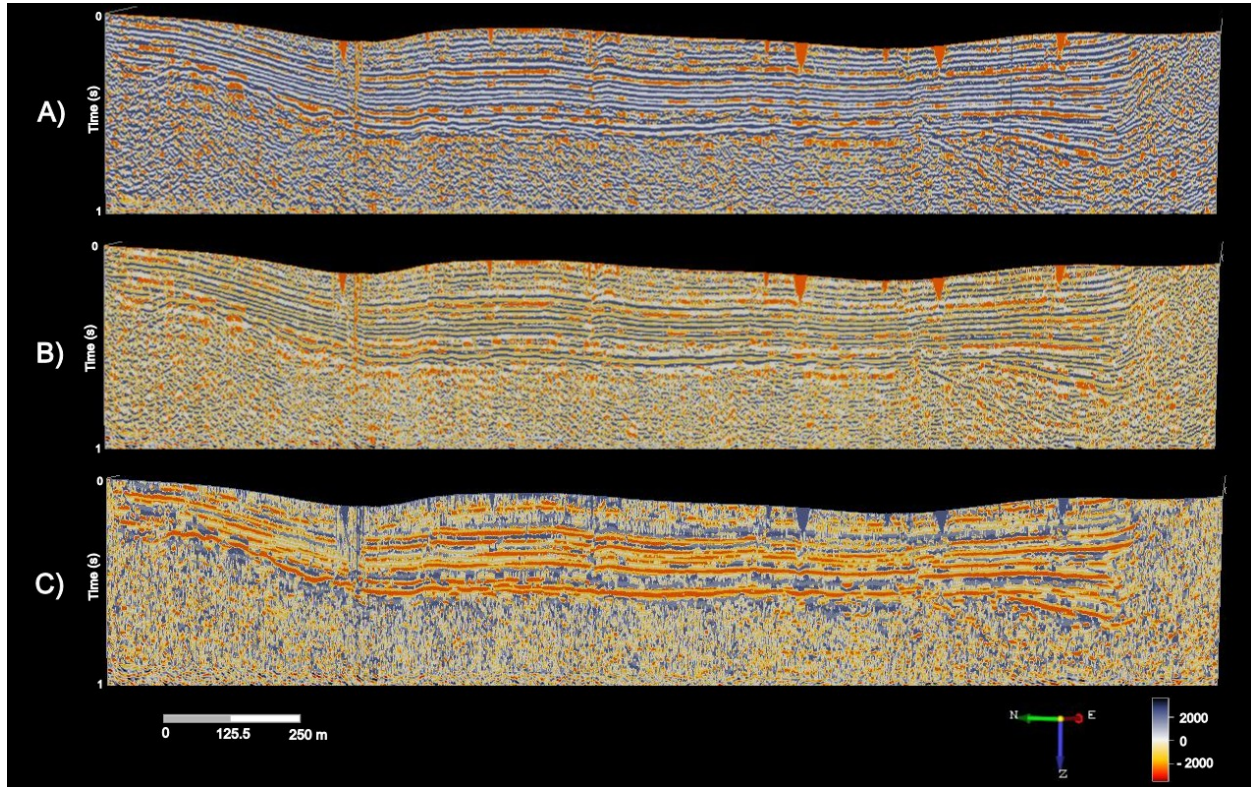


Figure 3. 6: Set 2 classified into 3 (A), 4 (B) and 5 (C) clusters.

The first set was trained 194 times and presented a 66 x 46 U-matrix and quantization error equal to 0.686. The second set was trained 113 times, yielding a 67 x 45 U-matrix and quantization error equal to 0.6. However, for the different groupings, the results of error distribution were similar for the same data sets.

From the images obtained it is possible to see that the training for the first set of attributes is a rough classification of the different units, with strong presence of noise in certain portions. The classification in both groups highlights the existence of four seismic units; however, the lateral continuity of reflectors is not well marked for classification into three groups (Fig. 3.5A). In Fig. 3.5B and Fig. 3.5C there is a greater emphasis on the seismic unit in the middle portion of the sedimentary package, with the lateral continuity of reflectors being more pronounced.

Note that the best classification for the first set of attributes occurs for four distinct clusters (Fig. 3.5B) with lower noise and greater concentration of lateral continuity of reflectors. The red

color highlights a seismic unit that covers the basement. Covering this unit, a second unit of similar thickness is highlighted in dark blue and light blue. In the intermediate portion there is a thicker third unit highlighted by parallel and continuous reflectors in white. The upper (fourth) unit is a mixture of continuous reflectors (white) on the edge and more chaotic in the central portion (blue).

Training for the second set of attributes shows an enhancement for the best set of lights, highlighting their lateral continuity throughout the portion of the section. The second set of attributes was more efficient in the definition of horizons because of the lower amount of noise. As the classification of the first set, there is a highlight of four major seismic units, better visualized in the classification into five groups (Fig. 3.6C). In Figs. 3.6A and Fig. 3.6B it is possible to observe that the lower unit (basement horizon) and the intermediate unit are similar, both having continuous and parallel reflectors highlighted by the same colors, dark blue and light blue. The transition from the lower unit to the intermediate seismic unit is highlighted in red and shows much noise and less emphasis on the lights. A similar characteristic appears in the upper seismic unit that also has portions with some greater emphasis on continuous reflectors (blue).

All the images obtained from the combination of attributes have improved resolution, but the best image quality corresponds to the second group, divided in five clusters. It is because the data set had the best arrangement and the best classes separation by the map. The MANOVA analysis for this group indicated that the classes are very well separated. The variables that better separated the four classes were the Envelope Weighted Frequency, Instantaneous Phase, Instantaneous Frequency and Envelope Weighted Phase.

The results show that the seismic units resemble in part the Fernandes's (1993) proposal, differing in the discrimination of a fourth unit that covers the basement seismics across the portion of the section. Furthermore, there is no highlight near the fault seismofacies edge, interpreted in Fernandes (op. cit.) as characteristic of proximal alluvial fan deposits. So, the combination of these attributes is not indicated for the definition of faults, but just for the desired lateral continuous feature.

From the analysis of the classes found, it can be inferred that the lower seismic unit and the intermediate unit have a similar pattern, with continuous and parallel reflectors. The lower units

with parallel reflectors indicate a lower energy depositional environment, whereas for the other units, highlighted in red, a higher energy depositional environment is suggested.

3.5. CONCLUSIONS

For the reservoir characterization, the identification of geological features and highlighting continuities and discontinuities of sedimentary bodies in seismic sections are very important, and the use of additional tools for interpreting such data is of great interest.

The use of seismic attributes, combined with the technique of neural networks, has increased the signal-noise ratio to detect discontinuities and enhance continuity of seismic reflectors, which would not be easily interpreted in the original seismic. Thus, the method allows the generation of seismic imaging with higher resolution based on the groups of seismic data identification.

Of the two sets of seismic attributes analyzed, it is possible to see less noise and greater lateral continuity in the images generated by the second set, in which the variable Envelope Weighted Phase corroborates with the other chosen attributes for sharper results and better definition than the variable Medium Frequency, which highlights the presence of fluids and formation of thin layers.

The seismic units are defined here according to previously published works on the Taubaté Basin and their definition in the remaining seismic sections of the basin will enable a better understanding of the tectono-sedimentary evolution of the different compartments.

ACKNOWLEDGEMENTS

The authors acknowledge the financial support of Petrobras, and also thank ANP - National Petroleum Agency for providing the data.

APPENDIX A. SELF-ORGANIZING MAPS

Self-Organizing Map (SOM) is a technique of clustering that performs pattern recognition based on the human nervous system (Kohonen, 2001).

Considered a neural network method, SOM is the only technique in this group that works without using *a priori* information. Therefore this characteristic makes it highly adaptable to different types of data and an important aid in the characterization of geological data without or inadequate identification. In addition, the multivariate technique is able to quickly resolve problems by clustering several different attributes, which would be very difficult for a human interpreter to perform, and is highly sensitive to noise and redundancy, common in the seismic data analyzed in this study (Bhatt & Helle, 2002; Kohonen, 2001; Stundner & Oberwinkler, 2004; Esposito *et al.*, 2008).

For classification with SOM, the seismic attributes are represented by input vectors in space R^n , being the purpose of this to organize these attributes in a geometric structure called the self-organizing map (Kohonen, 2001).

The method consists of computational units called neurons, which are interconnected forming an elastic mesh. The number of map units that have the ability to determine effectiveness and generalization of the method (Kohonen, 2001) will be defined in this work by the following heuristic formula described in the software Somtoolbox, used in this research, develop by Vesanto *et al.* (2000):

$$N = 5 \times \sqrt{n} \quad (\text{A} - 1)$$

where N is the number of nodes, n is the number of reflected waves.

The algorithm readapts the location of the nodes, through adjustment of their synaptic weights, in an amount equal to the number of nodes defined by Equation (A – 1). Only the nearest neuron and its neighborhood have its value updated in a process called competitive (Kohonen, 2001).

The Update method is described by Equation (A – 2).

$$\begin{cases} m_i = m_i(t) + \alpha(t) h_{bi}(t) [x - m_i(t)], & \text{se } m_i \in \text{neighborhood of } c \\ m_i = m_i, & \text{otherwise,} \end{cases} \quad (\text{A} - 2)$$

where t denotes time, it is the step index of the sequential process, x the input vector analysis – in this work represented by the seismic attribute, m_i the weight associated with each

neuron, represented by the index i , which is also called prototype vector, $\alpha(t)$ the learning rate, which varies from 0 to 1 (and decreases along the iterations until its value is no longer significant, that is, until there is convergence), and $h_{bi}(t)$, which represents the neighborhood size, described by:

$$h_{bi}(t) = e^{-\frac{\|r_b - r_i\|^2}{2\sigma^2(t)}}, \quad (\text{A - 3})$$

where r_b and r_i are the positions of the b and i neurons in the SOM lattice, σ is the standard deviation and below the rate of learning:

$$\alpha(t) = \alpha_0 / \left(1 + 100 \frac{t}{T}\right), \quad (\text{A - 4})$$

where T is the number of iterations and α_0 the initial learning ratio.

At the end of iterations, or training, each node represents a group of seismic attributes. Such training can be determined by convergence or stopping by a preset number of iterations.

In general the number of neurons is extremely high and a reunification of these nodes is necessary, because groups of neighboring neurons have similar characteristics. The most common method of reunification is k-Means (Khadir & Khedairia, 2008), developed by MacQueen (1967), in which the number of groups (K) is chosen by the Davies-Bouldin index (1997):

$$\frac{1}{K} \sum_{C=1}^K \max_{l \neq k} \left\{ \frac{S_K(Q_C) + S_K(Q_l)}{d_{Ke}(Q_C, Q_{Cl})} \right\}, \quad (\text{A - 5})$$

where K is the number of neurons clusters, S_k is the distance between them and groups Q_c e Q_{cl} and d_{Ke} the distance between the centroids of the respective groups (calculated using the location of map units). In this function the values of the Davies-Bouldin group index smaller than one represent a separate group, while values greater than 1 represent possibly crossed groups.

During the training, the k-Means units belonged to group K , which minimized the following objective function:

$$E = \sum_{k=1}^c \sum_{x \in q_k} \|x - c_k\|^2, \quad (\text{A - 6})$$

where c is the number of clusters and c_k is the center of cluster k . Note that SOM and k-Means algorithm have the same algorithm when the radius of the SOM neighborhood function is equal to zero. In this case the adjustment is made only in the winning unit.

The images obtained from the method were quantitatively evaluated by criteria from the package SOM toolbox of Vesanto *et al.* (2000). The analysis was based on the calculation of the final quantization error (qe), error of the topographic map (te).

The parameters adopted are dependent, *i.e.*, the analysis is performed by the input pixels, and not according to the training data, since the training is unsupervised and there is no exact values of pixels related to the stratigraphy and lithology that enables correlation values to specific classes (Vesanto *et al.* 2000).

The quantization error is associated with the final resolution of the map, on the average distance between each input pixel and the computer unit associated with it. The topographic error calculates the proportion of all pixels which have the second and third best unit different from the adjacent units with which they are associated with the preservation of topology (Vesanto *et al.* 2000).

Bellow the equations defining the errors used (Kohonen, 2001):

$$qe = \frac{1}{M} \sum_{k=1}^M \|x_k - m_{bk}\| \quad \text{and} \quad te = \frac{1}{M} \sum_{k=1}^M f(x_k) \quad (\text{A - 7})$$

where M is the amount of analyzed input vectors x . The function $f(x_k)$ is equal to 1 if the first and second x_k BMUs are adjacent, and 0 (zero) if otherwise.

To view the groups formed by the method, we chose to use the U-matrix (Ultsch, 1992), which computes the distance between the vectors in which the size of the nodes is proportional to the amount of input data belonging to them. The number of nodes in the U-matrix is defined by Eq. (A – 1) and the number of rows and columns is estimated so that their product is as close to N . The arrangement of the U-matrix is $N = nx \times ny$, where nx and ny are the dimensions of the matrix.

REFERENCES

- Ashida, Y., 1996. Data processing of reflection seismic data by use of neural network. *Journal of Applied Geophysics*, vol. 35, 89-98p.
- Bhatt, A. & Helle, H. B.. 2002. Determination of facies from well logs using modular neural networks. *Petroleum Geoscience*, 8:217-228.
- Chen, Q. & Sidney, S., 1997. Seismic attribute technology for reservoir forecasting and monitoring. *The Leading Edge*. Western Atlas International Inc., Houston, Texas, 1-8p.
- Chopra, S. & Marfurt, K., 2006. Seismic Attributes – a promising aid for geologic prediction. *CSEG Recorder*, Special Edition, 111-121p.
- Davies, D. L. & Bouldin, D. W., 1997. A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI, vol. 1(2), pp. 224-227.
- Esposito, A.M., Giudicepietro, F., D'Auria L., Scarpetta S., Martini M.G., Coltelli M., Marinaro M. 2008. Unsupervised Neural Analysis of Very-Long-Period Events at Stromboli Volcano Using the Self-Organizing Maps, In: *Boletin of the Seismological Society of America*, 98:2449-2459.
- Ferentinou, M.D. & Sakellariou, M.G., 2007. Computational intelligence tools for the prediction of slope performance. *Computers and Geotechnics*, vol.34, 362-384p.
- Fernandes, F. L. 1993. Arcabouço estrutural e evolução da Bacia de Taubaté - SP. 1993. 147 f. Tese (Mestrado) - Universidade Federal de Ouro Preto, Ouro Preto, MG.

- Forte, E., Pipan, M., Casabianca, D., Di Cuia, R., Riva, A. 2011. Imaging and characterization of a carbonate hydrocarbon reservoir analogue using GPR attributes, *Journal of Applied Geophysics*, doi: 10.1016/j.jappgeo.2011.09.009
- Hasan, O. & Tahar, S. 2006. Formalization of Standard Uniform Random Variable. Technical Report, Concordia University, Montreal, Canada.
- Khadir, M.T. & Khedairia, S., 2008. Self-Organizing Map and K-Means for Meteorological Day Type Identification for the Region of Annaba-Algeria- In: *7th Computer Information Systems and Industrial Management Applications*, p. 91-96.
- Kohonen, T. 2001. *Self-Organizing Maps*. Springer-Verlag, New York, 501pp.
- Marques, A. 1990. Evolução tectono-sedimentar e perspectivas exploratórias da Bacia de Taubaté, São Paulo, Brasil. *Boletim de Geociências da Petrobrás*, v. 4, n. 3, p. 253-262.
- Meldahl, P., Heggland, R, Bril B., Groot, P., 2001. Identifying faults and gas chimneys using multiattributes and neural networks. *The Leading Edge*, May, 474-482p.
- Miller, M. S. & Powell, K. S. 2001. Seismic Interpretation and Processing Applications. *Computational Neural Networks for Geophysical Data Processing. Handbook of Geophysical Exploration*, vol. 30, chapter 7.
- Moraes, D.R. deS., 2004. *Inteligência Computacional na Classificação Litológica*. Tese de Doutorado em Ciências em Engenharia Civil, UFRJ, 82p.
- Poulton, M. M. 2001. A Brief History. *Computational Neural Networks for Geophysical Data Processing. Handbook of Geophysical Exploration*, vol. 30, chapter 1.

- Rezvandehy, M., Aghababaei, H., Raissi, S.H.T., 2011. Integrating Seismic Attributes in the accurate modeling of geological structures and determining the storage of the gas reservoir in Gorgan Plain (North of Iran). *Journal of Applied Geophysics*, 73: 187-195.
- Riccomin,i C., 1989. O rift continental do sudeste do Brasil. Tese de doutoramento, Instituto de Geociências – USP, São Paulo, SP, Brasil.
- Sheline, H.E., 2005. The use and abuse of seismic attributes. Disponível em <http://www.searchanddiscovery.net/>.
- Stundner, M. & Oberwinkler, C. 2004. Self-Organizing Maps for Lithofacies Identification and Permeability Prediction. In: SPE Annual Technical Conference and Exhibition held in Houston, p.8.
- Taner, M.T., 1992. Attributes Revisited. Rock Solid Images, Houston, Texas. 38pp.
- Taner, M.T., 2001. Seismic Attributes. CSEG Recorder. Rock Solid Images, Houston, USA, 89-56p.
- Ulsch, A. 1992. Knowledge acquisition with self-organizing neural networks. In: Aleksander I & Taylor J (Eds). *Artificial Neural Networks*. Elsevier Science Publishers, B.V., 735-740.
- Vesanto, J., Himberg, J., Alhoniemi, E., Parhankangas, J. 2000. Self-Organizing Map in Matlab: The SOM Toolbox. In: *Proceeding of the Matlab DSP Conference*, p.59.
- Zhang, L., Quijren, J., Schuelke J. 2001. Self-Organizing Map (SOM) Network for Tracking and Classifying Seismic Traces. *Computational Neural Networks for Geophysical Data Processing*. Handbook of Geophysical Exploration, vol. 30, chapter 10.

ANEXO 4

“Kuroda, M. C.; Vidal, A.C.; Bueno, J.F., Caracterização de Reservatórios e Predição de Permeabilidade Através de Redes Neurais Supervisionadas (para submissão)”

CARACTERIZAÇÃO DE RESERVATÓRIOS E PREDIÇÃO DE PERMEABILIDADE ATRAVÉS DE REDES NEURAIIS SUPERVISIONADAS

Michelle Chaves Kuroda*, Alexandre Campana Vidal, Juliana Finoto Bueno

Institute of Geosciences, P.O.Box 6152, University of Campinas – UNICAMP, 13083-870,

Campinas, SP, Brazil. Tel: +55-19-3521-4659; Fax: +55-19-3289-1562. E-mails:

mckuroda@ige.unicamp.br, vidal@ige.unicamp.br, juliana.bueno@ige.unicamp.br

RESUMO - Um dos fatores responsáveis pelo melhor desenvolvimento de um campo de petróleo está relacionado à compreensão da distribuição da permeabilidade na rocha reservatório. Entretanto devido ao alto custo para a estimativa dessa variável, obtida com precisão apenas em laboratório, esses dados são muito escassos. Este trabalho propõe a utilização de redes neurais de Retropropagação de Erros para caracterização de eletrofácies e posterior previsão de valores de permeabilidade em amostras classificadas como reservatório. O método é aplicado para dados do Campo de Namorado, da Bacia de Campos, que foi dividido em três blocos, para que a classificação dos poços fosse realizada em regiões com heterogeneidades semelhantes. O total de 44 poços tiveram a litologia determinada em três classes: reservatório, não reservatório e possível reservatório. Posteriormente, para a fácies reservatório foram estimados valores de permeabilidade com um segundo treinamento com Retropropagação de Erros. Os resultados são consistentes com os dados geológicos de testemunho, demonstrando que o método proposto pode ser utilizado como uma ferramenta complementar na caracterização de reservatórios e para a extrapolação de dados de permeabilidade.

Keywords: permeabilidade, neural network, Campo de Namorado.

4.1. INTRODUÇÃO

A permeabilidade é um parâmetro fundamental para a análise de reservatórios de hidrocarbonetos. Seu conhecimento é de extrema importância para o entendimento do comportamento do fluxo de fluidos e está diretamente ligada ao sucesso na etapa de desenvolvimento de um campo de petróleo (Wong *et al.* 2000; Singh, 2005).

Os resultados mais precisos em geofísica são encontrados com o auxílio de dados de testemunho. A análise deve ser realizada em muitos poços para a obtenção de valores de distribuição que representem a heterogeneidade do campo, o que torna o processo caro. Os valores de permeabilidade encontrados em laboratório, a partir de amostras de testemunho, são considerados reais e extensamente utilizados para inferir permeabilidade ao longo dos poços a partir de correlação com dados de perfis (Cozzi *et al.*, 2006).

Apesar de não ser possível medir diretamente a permeabilidade a partir de dados de perfis geofísicos, a correlação entre dados de laboratório e perfis é uma prática comum na indústria. Assim, métodos que utilizam modelos de correlação são utilizados para prever permeabilidade em intervalos não testemunhados, gerando estimativas satisfatórias (Wong *et al.* 2000; Singh, 2005).

Com esse objetivo muitas equações empíricas são formuladas para transformar dados de perfis geofísicos em valores de permeabilidade, como a teoria de Kozeny-Carmen, que encontra permeabilidade a partir de dados de porosidade em condições idealizadas (Wong *et al.* 2000, Onerous, 2002). Logo, os parâmetros e constantes são ajustados para casos particulares, mas devido à dispersão de dados nem sempre é possível fazer ajustes satisfatórios.

Paralelamente aos métodos empíricos, as aproximações estatísticas são mais versáteis para a aproximação de valores de permeabilidade. Entretanto, requerem comportamento linear e multi-normal, restringindo consideravelmente a possibilidade de aplicação do método.

Atualmente as redes neurais artificiais (RNA) são amplamente empregadas para determinar a distribuição de permeabilidade em reservatórios de óleo e gás, eliminando muitos dos problemas associados a custo e generalização de valores de modelos de predição (Stundner & Oberwinkler, 2004).

Muitos autores, como Cozzi *et al.* (2006), Singh (2005) e Majdi *et al.* (2010), demonstraram o uso de redes neurais na estimativa de informações como porosidade e permeabilidade, a partir de dados convencionais de poços, gerando resultados consistentes em reservatórios heterogêneos.

As vantagens da aplicação de redes neurais para a predição de dados petrofísicos são intrínsecas à rapidez e confiabilidade dos resultados. Além disso, as redes neurais não requerem o conhecimento estatístico dos dados para que sejam aplicadas, pois o algoritmo “aprende” as relações lineares e não lineares entre os dados automaticamente. O método é sensível a ruído, permite o uso de dados incompletos e preserva o intervalo original e a variabilidade das variáveis de saída (Olson, 1998; Alhoniemi *et al.*, 1999; Bhatt & Helle, 2002; Matos *et al.*, 2004; Stundner & Oberwinkler, 2004).

Entretanto para o uso do método de forma eficiente, Olson (1998) salienta que os dados treinados em uma análise supervisionada precisam conter características da heterogeneidade presentes nos poços a serem correlacionados. Faz-se necessária a concordância entre os dados de treino e o que se objetiva classificar, analisando os dados de perfis previamente à aplicação do método de redes neurais para a obtenção de resultados satisfatórios.

O principal objetivo deste trabalho é estimar satisfatoriamente valores de permeabilidade para poços do Campo de Namorado, através da análise multivariada de redes neurais Retropropagação de Erros.

Devido à grande heterogeneidade do conjunto de dados e a escassez de valores de permeabilidade para fácies não reservatório, os dados foram classificados com o mesmo método Retropropagação de Erros, em três litofácies (reservatório, não reservatório e possível reservatório) e a permeabilidade foi estimada para a fácies reservatório. Segundo Olson (1998), Lee & Datta-Gupta (1999), Stundner & Oberwinkler (2004) e Cozzi *et al.* (2006) esse método aumenta consideravelmente a precisão dos resultados e apresenta menor tempo de processamento.

4.2. REDES NEURAIS ARTIFICIAIS (RETROPROPAGAÇÃO DE ERROS)

As redes neurais são modelos computacionais inspirados no funcionamento do cérebro humano. Diferente dos métodos tradicionais de classificação, as redes neurais são guiadas pelos dados e durante seu treinamento são capazes de se adaptar e desenvolver associações, transformações ou mapas entre os dados. Dessa forma as redes neurais são utilizadas para reconhecimento, classificação e generalização.

Durante o treinamento do método, as unidades computacionais associam padrões diferentes de cada dado de entrada. Apesar de não ser necessário o conhecimento dos dados, os erros associados aos dados de entrada afetam a generalização dos resultados. Geralmente as redes possuem três camadas: a camada de entrada, a camada oculta e a camada de saída; como mostra a Fig. 4.1.

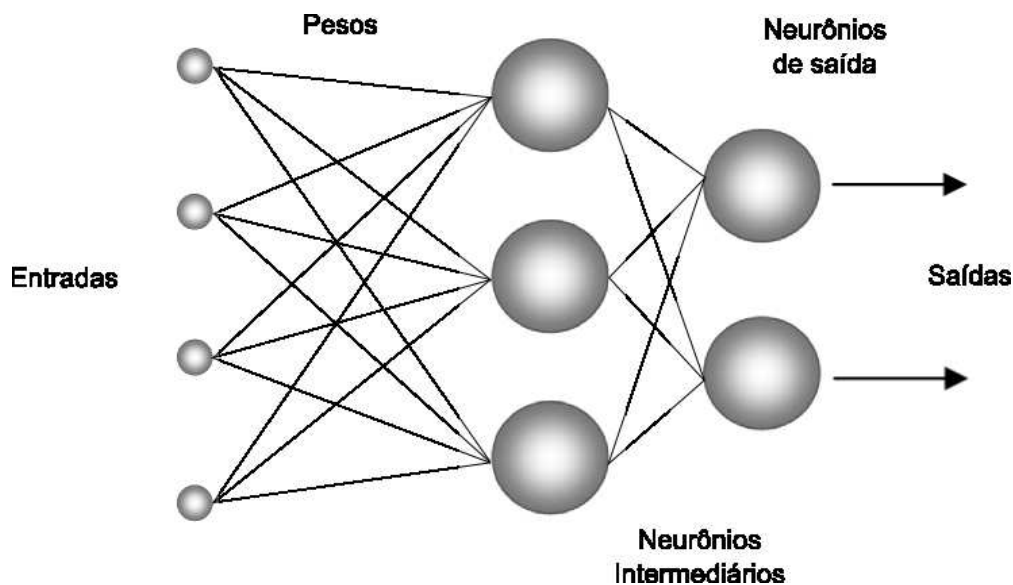


Figura 4. 1: Arquitetura da rede neural Back Propagation. O número de neurônios e camadas pode ser diferente.

Os neurônios processam e analisam as informações de predição e generalização do método. Eles atuam em paralelo e reúnem informações do neurônio anterior a ele, que são utilizadas como entrada pela camada seguinte.

A Fig.4.2 mostra o diagrama de um neurônio. A soma do dado de entrada (x_i) e seu peso correspondente (w_i) são recebidos por um neurônio da camada seguinte. Esse neurônio

processará a informação através da função de ativação, gerando uma saída, que será utilizada por outro neurônio.

Os conectores promovem a comunicação entre dois ou mais neurônios, representando o fluxo de informações na estrutura da rede. Já os pesos armazenam as informações processadas e são atualizados a cada iteração.

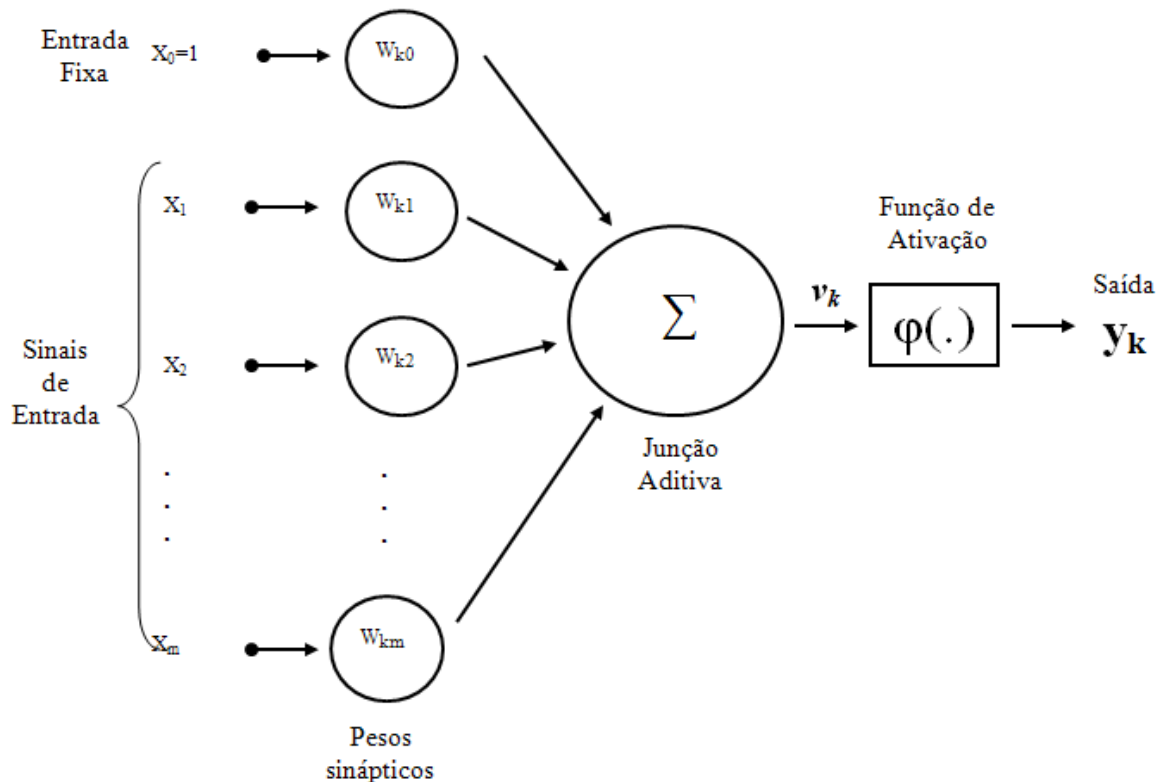


Figura 4. 2: Diagrama de um neurônio. Fonte: Onerous, 2002.

A função de ativação é responsável por transformar as informações de fluxo. As mais comuns são as funções Linear, Binária, Probabilística e Sigmóide.

A função sigmóide é a função mais utilizada pelas redes supervisionadas para predição e generalização. Trata-se de uma função unipolar, que produz saídas normalizadas. É uma função contínua e, portanto, diferenciável e positiva. É representada matematicamente como:

$$f(x) = \frac{1}{1+e^{-x^2}} \quad (4.1)$$

A rede alcança o aprendizado desejado ajustando os pesos dos conectores continuamente até que a diferença entre a saída e o valor do neurônio seja satisfatória.

A rede neural Retropropagação de Erros propaga as respostas dos dados de entrada para frente, enquanto envia os erros para trás para ajustar os pesos, a fim de melhorar o desempenho da rede. A minimização do erro é realizada pela execução do *gradiente decrescente* na superfície de erros do espaço de pesos, no qual a altura desse espaço corresponde à medida do erro. O ajuste inicia-se nas unidades de saída, onde está disponível e é retropropagado entre as camadas, ajustando os pesos até que a camada de entrada tenha sido processada (Mueller, 1996; Poulton, 2001).

Há então duas fases no processo de aprendizagem do método: a fase da propagação das entradas pelas camadas e a fase de fluxo contrário do ajuste dos erros, alterando valores de pesos sinápticos.

Na segunda fase, o sinal de erro na saída do neurônio j , na iteração n é dado segundo a equação abaixo:

$$e_j(n) = d_j(n) - y_j(n), \quad (4.2)$$

Para j sendo um nó de saída, d é a resposta desejada e y é o sinal de entrada que aparece na saída do neurônio j .

Para cada neurônio j o valor instantâneo da energia do erro é: $\frac{1}{2}e_j^2(n)$, logo a energia total do erro $\varepsilon(n)$ instantâneo é:

$$\varepsilon(n) = \frac{1}{2} \sum_{j \in C} e_j^2(n), \quad (4.3)$$

no qual C é o conjunto dos neurônios da camada de saída da rede, que são os únicos neurônios visíveis, para os quais os sinais de erro poder ser calculados diretamente. Logo a energia média do erro quadrado (MSE – *mean squared error*) é obtida somando-se os $\varepsilon(n)$ para todos os n e então normalizando em relação ao tamanho do conjunto de exemplos do conjunto de treinamento, N :

$$\varepsilon_{med} = \frac{1}{2N} \sum_{n=1}^N \sum_{j \in C} (d_j(n) - y_j(n))^2. \quad (4.4)$$

Essa energia é uma função cujos valores de pesos sinápticos e níveis de bias são parâmetros livres, e representa a função de custo como medida do desempenho de aprendizagem, que busca ajustar tais parâmetros e minimizar ε_{med} .

Considere um neurônio j sendo alimentado por um conjunto de sinais de entrada produzidos por uma camada de neurônios à sua esquerda. O campo local induzido $v_j(n)$ produzido na entrada da função de ativação associada ao neurônio j é:

$$v_j(n) = \sum_{i=0}^m w_{ji}(n) y_i(n), \quad (4.5)$$

no qual m é o número total de entradas aplicadas ao neurônio j . O peso sináptico $w_{ji}(n)$ conecta a saída do neurônio i à entrada do neurônio j , na iteração n . E y_i é o sinal funcional, ou sinal de entrada (estímulo). Assim, o peso w_{j0} corresponde à entrada fixa $y_0 = +1$ é igual ao bias b_j aplicado ao neurônio j .

O sinal funcional $y_j(n)$ que aparece na saída do neurônio j na iteração n é:

$$y_j(n) = \varphi_j(v_j(n)), \quad (4.6)$$

no qual φ_j é a função de ativação.

O algoritmo de reropropagação aplica uma correção Δw_{ji} ao peso w_{ji} proporcional à derivada parcial $\partial \varepsilon(n) / \partial w_{ji}(n)$, que de acordo com a regra da cadeia do cálculo, pode ser expresso como:

$$\frac{\partial \varepsilon(n)}{\partial w_{ji}(n)} = \frac{\partial \varepsilon(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial w_{ji}(n)}, \quad (4.7)$$

tal gradiente representa um fator de sensibilidade determinando a direção de busca no espaço de pesos para o peso sináptico w_{ij} .

Calculando as diferenciais parciais da Eq. 4.7:

Diferenciando ambos os lados da Eq. 4.3 em relação à $e_j(n)$:

$$\frac{\partial \varepsilon(n)}{\partial e_j(n)} = e_j(n), \quad (4.8)$$

Diferenciando a Eq. 4.2 em relação à $y_j(n)$:

$$\frac{\partial e_j(n)}{\partial y_j(n)} = -1, \quad (4.9)$$

Diferenciando a Eq. 4.6 em relação à $v_j(n)$:

$$\frac{\partial y_j(n)}{\partial v_j(n)} = \varphi'_j(v_j(n)), \quad (4.10)$$

Diferenciando a Eq. 4.5 em relação à $w_{ji}(n)$:

$$\frac{\partial v_j(n)}{\partial w_{ji}(n)} = y_j(n), \quad (4.11)$$

Substituindo tudo na Eq. 4.7:

$$\frac{\partial \varepsilon(n)}{\partial w_{ji}(n)} = -e_j(n) \varphi'_j(v_j(n)) y_j(n), \quad (4.12)$$

A correção $\Delta w_{ji}(n)$ aplicada à $w_{ji}(n)$ é definida pela regra delta:

$$\Delta w_{ji}(n) = -\eta \frac{\partial \varepsilon(n)}{\partial w_{ji}(n)}, \quad (4.13)$$

Onde η é o parâmetro da taxa de aprendizagem.

O sinal negativo indica a busca por uma direção para a mudança de peso que reduza o valor de $\varepsilon(n)$.

$$\Delta w_{ji}(n) = \eta \delta_j(n) y_i(n), \quad (4.14)$$

Onde $\delta_j(n)$ é o gradiente local, definido por:

$$\delta_j(n) = -\frac{\partial \varepsilon(n)}{\partial v_j(n)} = e_j(n) \varphi'_j(v_j(n)), \quad (4.15)$$

Esse gradiente aponta para as modificações necessárias aos pesos.

Assim:

$$\Delta w_{ji}(n) = (\text{taxa de Aprendizagem}) \left(\frac{\eta}{\text{Gradiente Local}} \right) (\text{Sinal de entrada do neurônio } j) \quad (4.16)$$

Da Eq. 4.16 é possível concluir que:

- 1) Se j é um nó de saída: $\delta_j(n)$ é igual ao produto de $\varphi'(v_j(n))$ pelo sinal do erro, ambos associados ao neurônio j .
- 2) Se j é um nó oculto: $\delta_j(n)$ é igual ao produto de $\varphi'(v_j(n))$ pela soma ponderada dos δ s calculados para os neurônios na próxima camada oculta ou camada de saída que estão conectados ao neurônio j .

O maior desafio da rede é saber estipular o fim do treinamento. Para resolver esse problema, a rede é calibrada por um número de iterações, entre de 20 mil a 40 mil vezes, para as redes neurais Retropropagação de Erros (Onerous, 2002), ou por um valor de teste que controla a convergência do processo de iteração.

Caso a rede tenha um processo de treinamento muito longo, os neurônios passam a “memorizar” ao invés de aprender, prejudicando a eficiência do método. Há duas maneiras de estipular um valor para o melhor número de iterações, realizado através de um conjunto de 10 a 40% dos dados que são separados antes do treinamento para validação dos resultados (nesse caso as iterações ficam entre 50 e 400) ou através de MSE, que requer 20.000 a 40.000 iterações.

Depois de treinados os resultados são verificados. A rede neural é considerada confiável se os dados preditos são próximos ou iguais aos valores dos dados de entrada. Caso o resultado seja satisfatório, a rede pode ser aplicada para qualquer dado sem testemunho no campo, considerando que os dados treinados apresentam a heterogeneidade dos intervalos a serem estimados (Onerous, 2002).

Para cada treinamento são calculados valores de MSE, que procuram encontrar o melhor ajuste para um conjunto de dados, tentando minimizar a soma dos quadrados das diferenças entre o valor estimado e os dados observados (tais diferenças são chamadas resíduos). Estas informações foram utilizadas como decisão de parada do treinamento. Os resultados são

apresentados de maneira estatística, por meio da análise de regressão para quantificar o grau de identificação entre os resultados obtidos e os desejados. O valor R^2 (coeficiente de determinação) entre 0 e 1 indica se o modelo está perfeitamente ajustado ($R^2=1$) ou se está incorreto ($R^2=0$).

$$R^2 = 1 - \frac{\sum_{i=1}^n (atual_i - predito_i)^2}{\sum_{i=1}^n (atual_i - média)^2} \quad (4.17)$$

Existem muitas possibilidades de arquiteturas para diferentes tipos de dados e problemas envolvidos. A quantidade de padrões conectivos entre os neurônios os torna úteis para objetivos diferentes.

4.3. CAMPO DE NAMORADO

O conjunto de dados analisado é proveniente da Bacia de Campos, que apresenta área de aproximadamente 100.000 km² e está localizada no litoral do Estado do Rio de Janeiro e Espírito Santo, entre o Alto de Vitória ao norte e Alto de Cabo Frio ao sul. Com mais de 1.600 poços perfurados desde a década de 70, a bacia possui mais de 90% das reservas petrolíferas brasileiras (Winter *et al.*, 2007).

A Bacia de Campos é de margem passiva, relacionada à fragmentação do supercontinente Gondwana durante o Cretáceo. Tectonicamente, insere-se na fase de pós-rifte na qual se depositaram os sedimentos da Megasequência Marinha, especificamente na Sequência Ocêânica Progradante, durante o Eoceno Médio/Recente (Dias *et al.*, 1990)

Nesta fase, o intenso tectonismo formou a Serra do Mar, resultando em uma área-fonte de sedimentos para a bacia. O grande volume de corpos arenosos turbidíticos, denominado informalmente de Arenito Namorado, provenientes das áreas de plataforma foram transportados e depositados nas porções mais profundas da bacia. Os arenitos turbidíticos são os reservatórios de hidrocarbonetos mais importantes da bacia, com idades variáveis do Meso-Cretáceo ao Neo-Terciário (Barboza, 2005) e foco de estudo neste trabalho.

O reservatório situa-se entre -2900 e -3400 m, sendo que os limites de topo e base apresentam espessuras que variam entre 90 m e 180m (Johann, 2004; Souza Jr, 1997)

O reservatório possui 21 fácies testemunhadas, das quais as principais rochas reservatórios são referentes a duas fácies de arenitos, que apresentam espessuras métricas nos testemunhos. A fácies de maior ocorrência nos testemunhos corresponde a arenito médio maciço, arcoseano e bem selecionado. A segunda fácies é composta por arenitos grossos amalgamados, com graduação da fração areia grossa-conglomerática na base para fração grossa no topo. As porosidades e permeabilidades médias são respectivamente 26% e 400mD (Bacoccoli *et al.*, 1980). As fácies arenitos e folhelhos interestratificados são consideradas como reservatórios secundários.

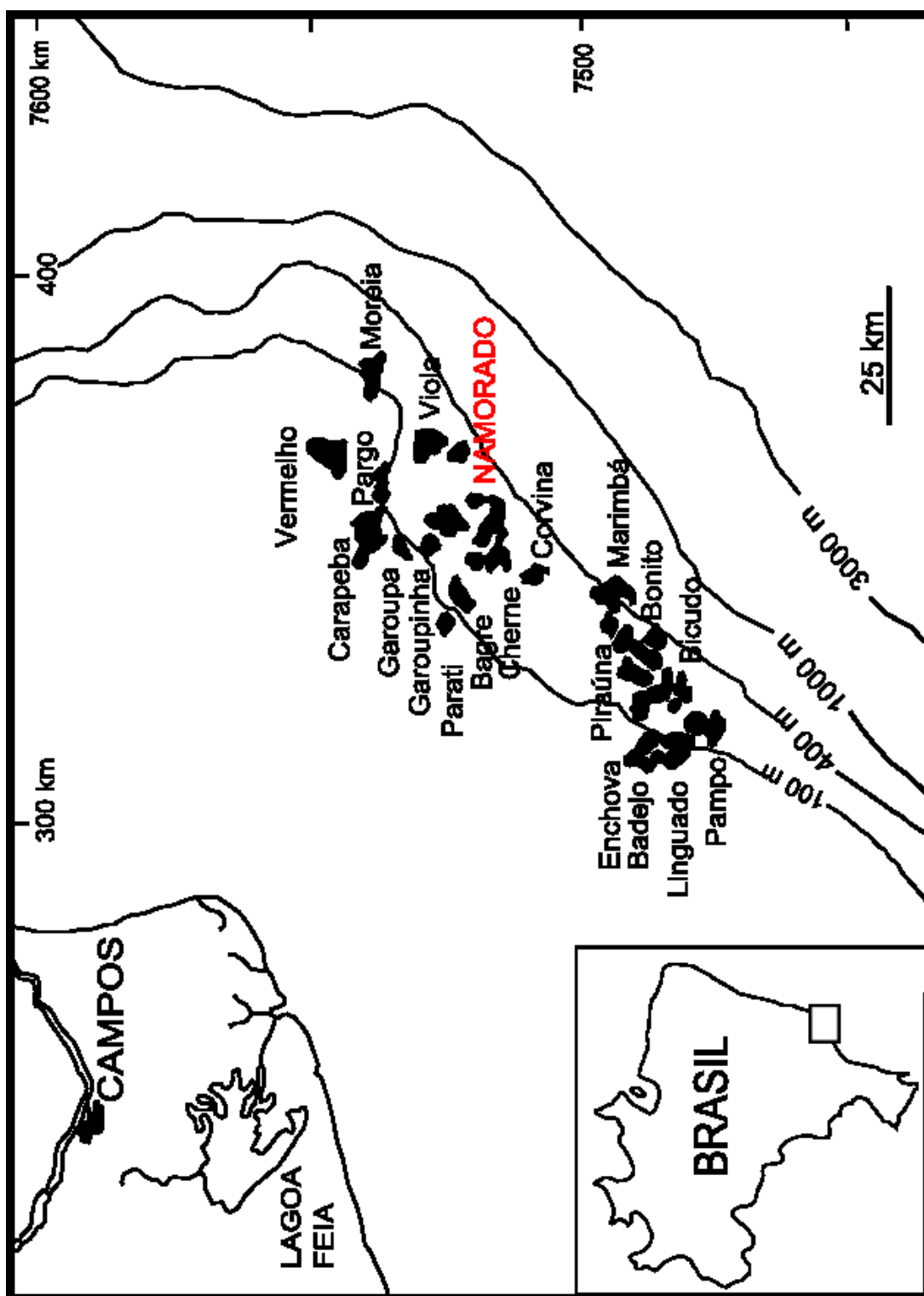


Figura 4. 3: Localização da Bacia de Campos com os principais campos petrolíferos. Adaptado de Milani et al.(2000).

4.3.1 CONJUNTO DE DADOS

Os dados de poços analisados são referentes a 44 poços disponibilizados pela ANP (Agência Nacional de Petróleo), dos quais sete (NA01A, NA02, NA04, NA07, NA11A, NA22 e RJS042) apresentam dados de testemunho que foram utilizados para validar o método e avaliar o desempenho da ferramenta escolhida para análise e predição de propriedades petrofísicas.

As 29 litologias descritas nos testemunhos foram divididas em três grupos:

- Arenito Grosso, Amalgamado e Arenito Médio Gradado ou Maciço, representam fácies **reservatório**;

- Arenito Médio Fino Laminado, Arenito/Folhelho Interestratificado, Arenito/Folhelho Finamente Interestratificado e Arenito Médio/Fino Laminado Cimentado representam as fácies de **possíveis reservatórios**; e

- As litologias restantes representam **não reservatório**.

Para classificação de litofácies e predição de permeabilidade, foram utilizados os seguintes dados: GR, RHOB, NPHI, DT, porosidade calculada em laboratório e os litotipos testemunhados, totalizando 3735 amostras, das quais 264 apresentam dados de porosidade e permeabilidade calculados em laboratório, utilizados para treinamento da rede neural e são provenientes das litofácies reservatório.

4.4. MÉTODOS

O método utilizado neste trabalho é similar ao realizado por Stundner & Oberwinkler (2004) e Cozzi *et al.* (2006), que obtiveram resultados satisfatórios na predição de permeabilidade através de dados de poços.

Os passos para o processo de aplicação de redes neurais, descritos por Olson (1998), foram utilizados e consistem em:

1. Definição do problema: determinação de qual será o resultado esperado, escolha cuidadosa dos dados de entrada, utilizados como guia na classificação ou predição dos dados durante o treinamento;
2. Tratamento de dados: análise dos dados para retirada de valores anômalos e correção de dados de poços;
3. Treinamento da rede neural: escolha da arquitetura da rede, assim como os parâmetros de treinamento, definição da função objetivo e decisão de parada das iterações;
4. Teste dos resultados: os valores de saída podem ser simulados com os dados treinados para verificação da precisão de resultados.
5. Aplicação da rede neural nos dados a serem classificados.

Os dados descritos foram aplicados primeiramente com o objetivo de classificação litológica. Com esse propósito, a classificação de fácies descritas por testemunhos foi utilizada para treinamento de redes neurais. Posteriormente à classificação, os dados de reservatórios foram separados dos demais para predição de permeabilidade, em uma segunda etapa de treinamento com Retropropagação de Erros, descritos na Fig. 4.4.

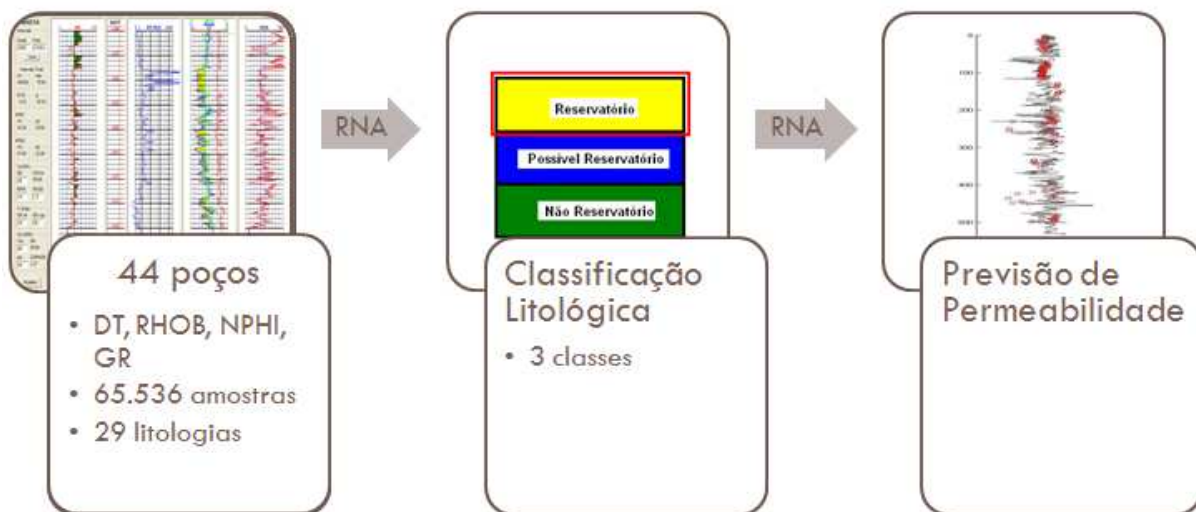


Figura 4. 4: Fluxograma da classificação automática dos dados com Retropropagação de Erros.

O número de iterações do treinamento foi decidido segundo valores de erros MSE e o número de neurônios na camada oculta foi testado para 2 a 20. Os testes foram repetidos 100 vezes na classificação de litotipos e 50 vezes na predição de permeabilidade.

Diante da complexidade dos dados, da heterogeneidade e grande dimensão do campo, a área estudada foi dividida em três grandes blocos, de acordo com a disposição dos poços e as falhas identificadas, como ilustra a Fig. 4.5. Cada bloco foi classificado separadamente. O Bloco 1 à esquerda, com 23 poços, foram classificados segundo litologia de acordo com os valores de testemunho dos poços RJS42, NA7 e NA02. O segundo bloco, central, contendo 18 poços foram classificados com valores de testemunho dos poços NA1A e NA4. E o último bloco, com 13 poços foram rotulados em 3 classes, segundo treinamento dos poços NA22 e NA11.

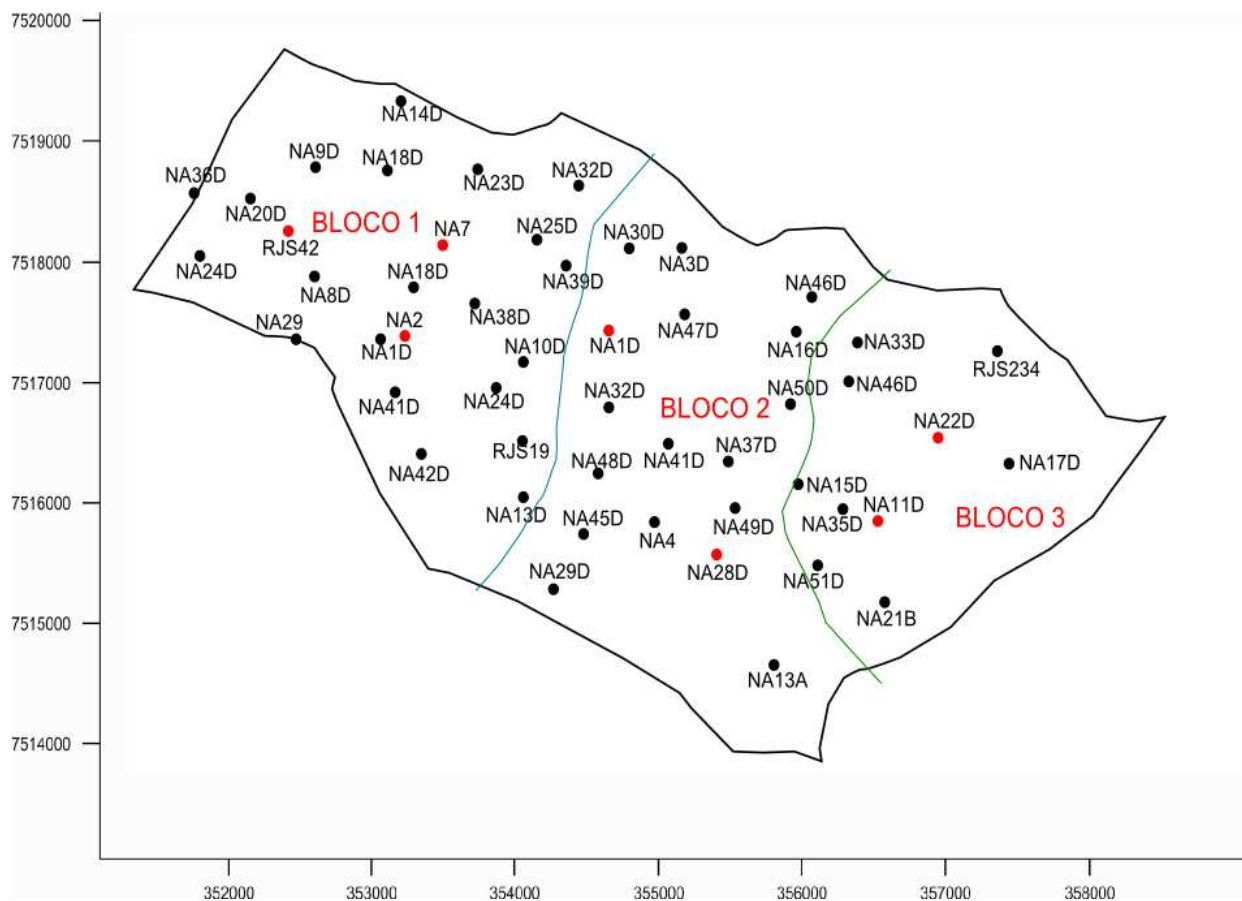


Figura 4. 5: Disposição dos poços no campo e sua divisão em 3 blocos, enumerados da esquerda para a direita. Em vermelho os poços com descrição de valores de litologia e permeabilidade, calculada em laboratório.

Para o treinamento de predição de permeabilidade, os dados de testemunho de todos os poços foram treinados e classificados juntos, devido à escassez de amostras testemunhadas. Os valores de permeabilidade foram utilizados em escala logarítmica, a fim de reduzir o erro associado na predição de valores, segundo estudo de Cozzi *et al.* (2006).

4.5. RESULTADOS E DISCUSSÕES

Os dados descritos foram analisados com o método Retropropagação de Erros em diferentes arquiteturas.

Os valores utilizados para classificação foram: 1 para reservatórios, 2 para possíveis reservatórios e 3 para não reservatório. Os resultados de acerto calculados a partir da comparação das classes encontradas e testemunhadas foram utilizados para definir a melhor estrutura da rede por blocos e a melhor iteração encontrada. A Tab. 4.1 mostra a melhor estrutura da rede e sua respectiva porcentagem de acerto e erros quadrados médios (MSE).

Os dados de possíveis reservatórios foram a classe com maior erro associado durante o treinamento, pois são pouco amostradas em testemunhos, por não serem alvo principal de prospecção e pequena representatividade volumétrica. Além disso, esse agrupamento representa grande número de litologias que, conseqüentemente, apresentam um leque de valores que se sobrepõem aos valores representativos de reservatório e não-reservatório. Com isso, os poucos dados e a grande variabilidade dos valores prejudicaram a eficiência do treinamento desse grupo.

Entretanto, como se observa na Tab. 4.1, os valores de erros e acertos obtidos com o treinamento do método são satisfatórios. A porcentagem de acerto dos poços é maior que 70%, valor satisfatório, mas como estão restritos aos dados de testemunho que representam muito pouco do conjunto de dados analisados, os dados de MSE tornam-se os principais indicadores do desempenho dos treinamentos, já que são intrínsecos aos dados estimados e observados.

Tabela 4. 1: Principais resultados encontrados na classificação de litotipos. O número de neurônios na tabela representa a melhor estrutura da rede Retropropagação de Erros encontrada e a porcentagem de acerto calculada a partir da comparação de classes.

Bloco	MSE	Melhor Iteração	Número de Neurônios	Acerto
1	8,54	9	4	0,704
2	4,45	5	20	0,786
3	6,47	8	17	0,730

O melhor resultado foi encontrado para o bloco 2, que também apresenta o menor erro associado. O menor desempenho, referente ao bloco 1, está associado aos poços utilizados para treino: RJS42, NA07 e NA02. Tais poços se encontram em uma área heterogênea do campo e apresentam maior número de intercalações de folhelhos e arenitos, que interferem nos valores de perfis e confundem a classificação.

A matriz de confusão revela os erros cometidos na classificação de litofácies por bloco. Nos três casos, o maior erro está associado à classificação de rochas possível reservatório, que como descrito anteriormente, não apresenta um padrão específico de valores de perfis.

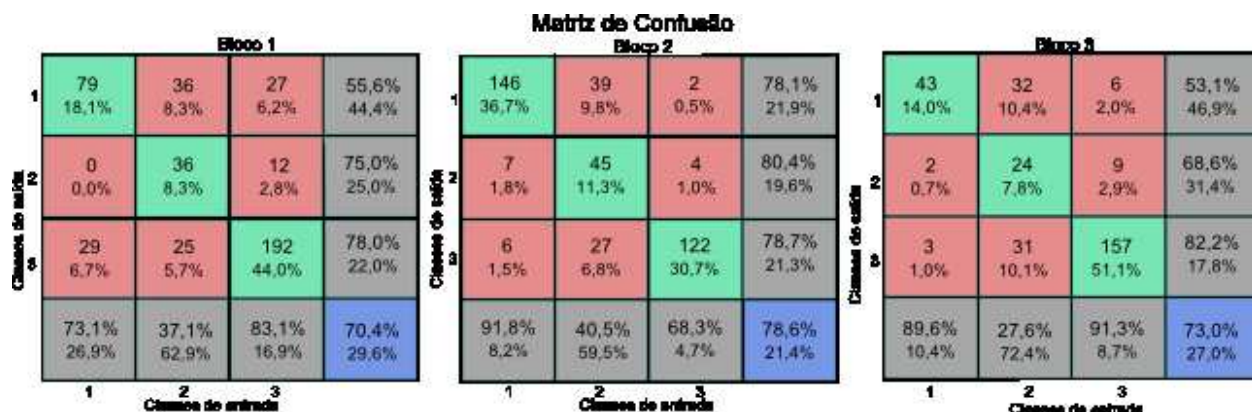


Figura 4. 6: Matriz de confusão para o treinamento de litofácies para os blocos 1, 2 e 3, respectivamente. As classes de entrada são referentes aos dados de testemunhos e as classes de saída referem-se à classificação encontrada com o algoritmo utilizado.

A segunda fase da classificação é referente à predição de valores de permeabilidade para a fácies classificada como reservatório.

Todos os dados de permeabilidade disponíveis foram utilizados para treinamento do método. O MSE foi utilizado como parâmetro para escolha do melhor treinamento do método, semelhante à primeira classificação.

A melhor estrutura da rede neural foi encontrada para 13 neurônios na camada oculta, na décima primeira iteração e erro de 0,265.

Como exemplo dos resultados obtidos é apresentada a classificação de litofácies e permeabilidade para o poço NA02 (Fig. 4.7). Na imagem pode ser observada a boa correlação entre as informações de testemunho e a classificação, além do grande número de intercalações das litologias que caracterizam esse reservatório.

Os valores de permeabilidade obtidos por redes neurais para os intervalos classificados como reservatórios são coerentes com os dados obtidos de laboratório, evidenciando a qualidade do fluxo das fácies descritas como reservatório nos intervalos que ocorrem no poço.

NA02

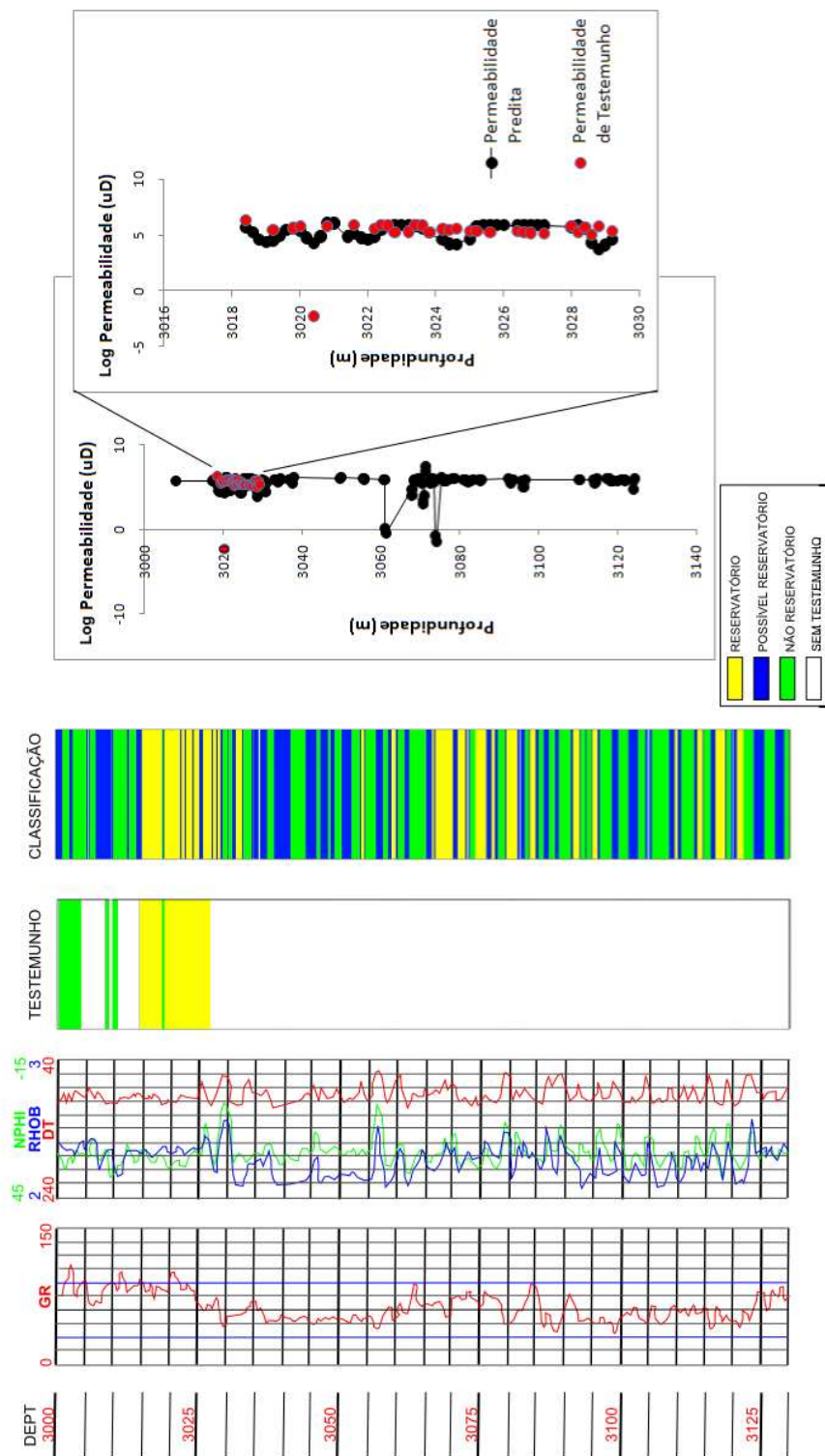


Figura 4. 7: Resultados de classificação de litologia e permeabilidade para o poço NA02, com os perfis GR, RHOB, DT e NPHI. Os pontos em vermelho do gráfico de permeabilidade são referentes aos valores em logaritmo

calculados em laboratório e a linha contínua representa a permeabilidade para o intervalo classificado como reservatório.

Na Fig. (4.8) encontram-se os valores estimados a partir do método e em vermelho os valores das amostras analisadas para os sete poços, em escala logarítmica. O resultado de R^2 foi de 0,8468. Próximo de 1 o valor garante o bom desempenho do método (Fig. 4.9).

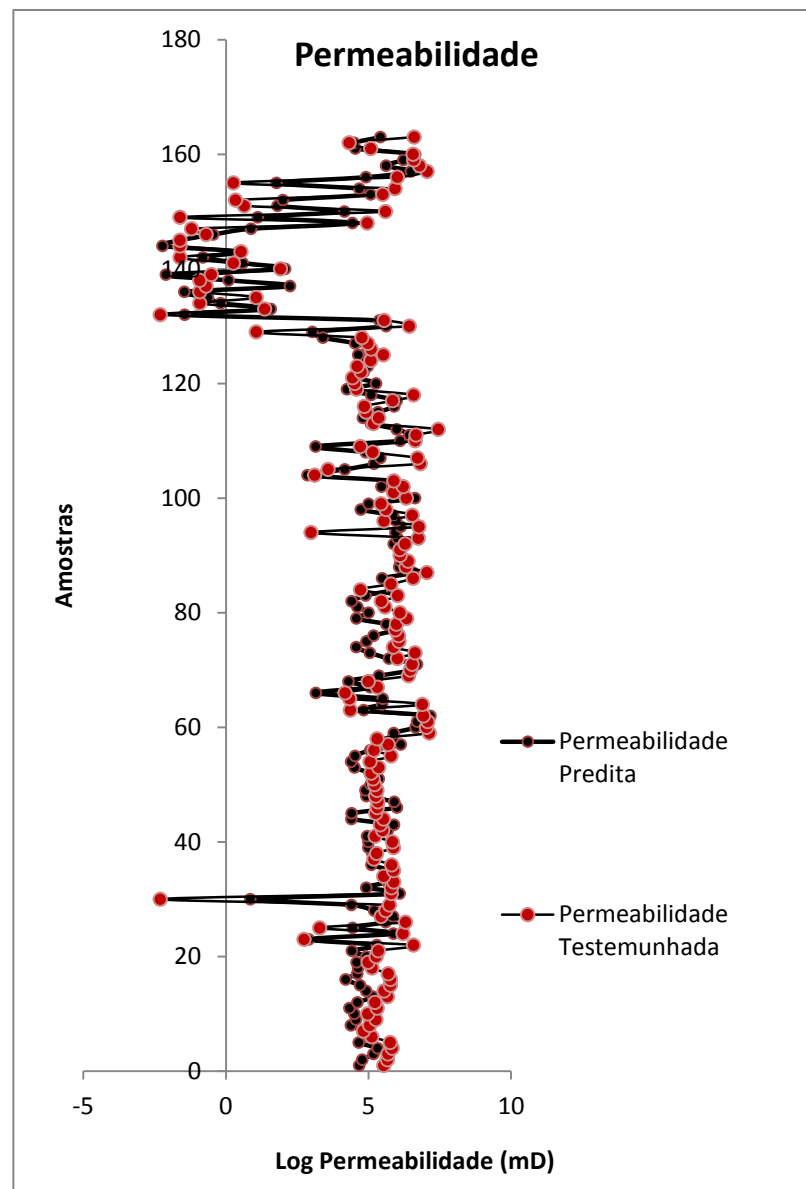


Figura 4. 8: Os valores de permeabilidade preditos com o método de redes neurais e em vermelho os dados de permeabilidade adquiridos a partir de amostras de testemunho.

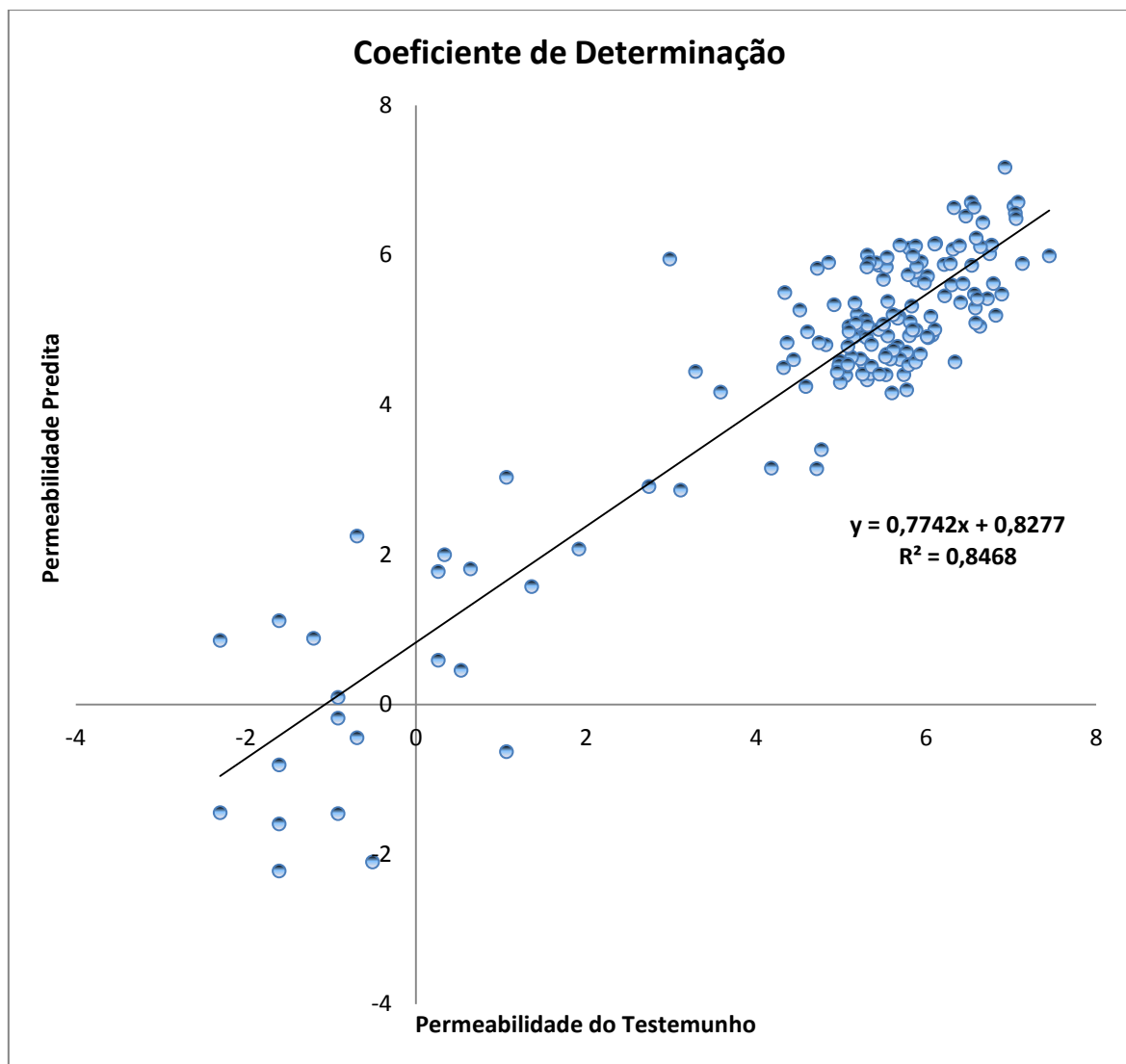


Figura 4. 9: Os valores de permeabilidade estimados versus a permeabilidade calculada a partir do testemunho.

4.6. CONCLUSÕES

A estimativa de permeabilidade em laboratório é um processo extremamente custoso e demorado e, portanto, inviável economicamente. Embora não seja conhecida a relação dos seus valores com a resposta de perfis geofísicos, a grande importância dessa estimativa para o melhor entendimento da heterogeneidade do campo e identificação de reservatórios torna o conhecimento dessa variável um grande desafio.

As redes neurais artificiais Retropropagação de Erros demonstram ser uma ferramenta útil para predição de propriedades de reservatórios, a partir de dados de poços. É possível aplicá-la

com sucesso em casos em que há pouca informação sobre a relação entre os dados de entrada e as saídas desejadas, ou em casos em que não seja conhecida uma relação entre os dados utilizados e os que se deseja estimar.

Neste trabalho, os resultados obtidos na identificação de reservatórios e estimação de valores de perfis foram expressivos e demonstraram correlação acima de 70% com os valores utilizados como alvo. O bom desempenho da ferramenta também está relacionado à qualidade dos dados analisados e ao cuidado com a extrapolação dos dados treinados aos dados a serem simulados.

REFERÊNCIAS

- ALHONIEMI E., HOLLMÉN J., SIMULA O., VESANTO J. -1999- Process Monitoring and Modeling using the Self-Organizing Map. Integrated Computer Aided Engineering. Disponível em: <<http://citeseerx.ist.psu.edu>>. Acesso em: 17 janeiro 2010.
- BACOCOLI G., MORALES R.G., CAMPOS O.J. 1980. The Namorado Oil Field: A Major Discovery in the Campos Basin, Brazil. AAPG Bulletin, 63(3):329-338.
- BARBOZA E.G. 2005. Análise Estratigráfica do Campo de Namorado (Bacia de Campos) com base na Interpretação Sísmica Tridimensional. Tese de Doutorado, Universidade Federal do Rio Grande do Sul, 235p.
- BHATT A. & HELLE H. B. 2002 Determination of facies from well logs using modular neural networks. Petroleum Geoscience, 8:217-228.
- COZZI, M., RUVO L., SCAGLIONI P., LYNE A.M. 2006. Core data preprocessing to improve permeability log estimation. SPE Annual Technical Conference and Exhibition being held in San Antonio.
- DIAS J.L., SCARTON J.C. ESTEVES F.R., CARMINATTI M, GUARDADO L.R. 1990. Aspectos da Evolução Tectono-Sedimentar e a Ocorrência de Hidrocarbonetos na Bacia de Campos. In: Gabaglia, G.P.R. & Milani, E.j. (Eds.) Origem e Evolução das Bacias Sedimentares Brasileiras. Rio de Janeiro, Petrobras, p. 333-360.

- JOHANN, P.R.S. -2004- Fácies Sísmicas. Boletim de Geociências da Petrobrás(12), 2:317-355. Rio de Janeiro, RJ.
- LEE S.H. & DATTA-GUPTA. 1999. Eletrofácies characterization and permeability predictions in carbonate reservoirs: Role of Multivariate Analysis and Nonparametric Regression. SPE Annual Technical Conference and Exhibition being held in Houston. 3-6.
- MAJDI A., BEIKI M., PIRAYEHGAR A., HOSSEINYAR G. 2010. Identification of Well logs with significant impact on prediction of oil and gas reservoirs permeability using statistical analysis of *RSE* values. Journal of Petroleum Science and Engineering. 75: 91-99.
- MATOS M.C., OSÓRIO P.L.M., JOHANN P.R.S. 2004 Agrupamento de Mapas Auto Organizáveis de Kohonen Aplicado a Análise de Fácies Sísmicas. XV Brazilian Automation Congress, 1:1-6.
- MILANI E.J., BRANDÃO J.A.S.L., ZALÁN P.V., GAMBOA L.A.P. 2000. Petróleo na Margem Continental Brasileira: Geologia, Exploração, Resultados e Perspectivas, Rio de Janeiro. Rev. Bras. de Geofísica, 18:3:352-396.
- MUELLER, A. 1996. Uma aplicação de redes neurais artificiais na previsão do mercado acionário. Tese de mestrado da Universidade Federal de Santa Catarina, Departamento de Pós-Graduação de Engenharia do Petróleo.
- OLSON, T.M. 1998. Porosity and Permeability Prediction in Low-Permeability Gas Reservoirs from Well Logs using Neural Networks, Paper SPE 39964, April 5 – 8.
- POULTON, M.M. 2001. Multi-Layer Perceptrons and Back-Propagation Learning. Handbook of Geophysical Exploration, Seismic Exploration. Volume 30. Computational Neural Networks for Geophysical Data Processing. Pergamon, Tucson, USA. 27-53pp.
- SOUZA JUNIOR, O.G. 1997. Stratigraphie Séquentielle et Modélisation Probabiliste des Réservoirs d'un Cône Sous-marin Profond (Champ de Namorado, Brésil). Intégration des Données Géologiques. Thèse de Doctorat. Université Paris. 128 pp.
- SINGH S. 2005. Permeability Prediction Using Artificial Neural Network (ANN): A Case of Uinta Basin. SPE Annual Technical Conference and Exhibition being held in Dallas. 9-12.

- STUNDNER M. & OBERWINKLER C. 2004. Self-Organizing Maps for Lithofacies Identification and Permeability Prediction. *In: SPE Annual Technical Conference and Exhibition held in Houston*, p.8.
- WINTER W.R., JAHNERT R.J., FRANÇA A.B. 2007. Bacia de Campos. B.Geoci. Petrobras, Rio de Janeiro, v.15, p.511-529.
- WONG P.M., JANG M., CHO S., GEODEON T.D. 2000. Multiple permeability predictions using an observational learning algorithm. *Computers & Geosciences*. 26: 907-913.

CONSIDERAÇÕES FINAIS

A caracterização de reservatórios é uma tarefa complexa que requer a integração de diversas ferramentas e experiência de intérpretes para garantir o sucesso do desenvolvimento do campo de petróleo.

Devido à grande quantidade de dados envolvidos nesta análise, as diferentes escalas de informações entre eles e a heterogeneidade em relação à distribuição espacial, a caracterização de reservatórios exige a utilização de métodos alternativos para a otimização do processo e minimização de riscos e tempo de interpretação.

Este trabalho aplicou diferentes métodos de redes neurais artificiais, em dados de poços – para caracterização de eletrofácies e previsão de propriedades petrofísicas – e em dados sísmicos, buscando a atenuação de ruídos e destaque de continuidade lateral, para realce de horizontes.

Como exemplos de aplicação no estudo de reservatórios, Stundner & Oberwinkler (2004) citam três tipos básicos de utilização de Redes Neurais com base em dados de poços: predição de permeabilidade e/ou porosidade; geração de poços sintéticos e determinação de eletrofácies.

Dispostos em anexos, os artigos presentes nesta dissertação são resultados de aplicações realizados em duas áreas distintas: Campo de Namorado, na Bacia de Campos e a Bacia de Taubaté e estão ordenados em ordem cronológica de submissão às revistas geofísicas.

O primeiro artigo, “Caracterização de eletrofácies por meio de Mapas Auto-Organizáveis”, submetido à Revista Brasileira de Geofísica, destaca a aplicação das redes neurais de Kohonen para o melhor detalhamento de fácies reservatório, posterior à classificação de litofácies com o método k-nn. Neste trabalho, encontra-se em detalhes a descrição do método utilizado e a análise estatística realizada, que corrobora para a validação da ferramenta escolhida.

Neste artigo foram utilizados os perfis elétricos: densidade, porosidade neutrônica, tempo de trânsito e raios gama, para o treinamento da ferramenta de redes neurais artificiais proposta. A eficiência do método RNA pode ser comprovada pela validação por métodos estatísticos como MANOVA e validação cruzada, descritos no texto. A partir desses resultados foi possível identificar quatro diferentes litotipos, que especificam a qualidade do reservatório e sua localização nos poços do Campo de Namorado, essencial para o melhor entendimento das heterogeneidades do campo. O segundo artigo, intitulado “Classification and Correlation of

Sedimentary Facies in the Taubaté Basin, São Paulo State, Brazil”, foi submetido aos Anais da Academia Brasileira de Ciências. Este trabalho contém resultados inéditos de classificação de eletrofácies da Bacia de Taubaté, possíveis apenas com a utilização do método de redes neurais para a classificação automática, tarefa que não seria viável visualmente devido à heterogeneidade geológica na bacia. Com tal análise foi possível realizar posteriormente a correlação lateral das unidades reservatórios, que corroboram com o estudo prévio realizado pelo pesquisador Fernandes (1993) na interpretação de linhas sísmicas 2D.

O terceiro anexo é referente ao artigo “Interpretation of Seismic Multiattributes Using a Neural Network”, submetido ao Journal of Applied Geophysics. Neste trabalho está presente a aplicação inédita de redes neurais de Kohonen para a classificação de multiatributos sísmicos, para realce da continuidade lateral das linhas sísmicas da Bacia de Taubaté.

O artigo detalha a metodologia utilizada para a criação de pseudos-atributos, criados com a análise de quatro atributos, escolhidos cuidadosamente, que destacam as unidades sedimentares analisadas por Fernandes (1993) e realçam corpos sísmicos não visíveis na imagem sísmica original. São analisados dois grupos de atributos sísmicos físicos, para os quais é aplicado o método sugerido e os resultados são ilustrados para a linha sísmica “0062”.

Os resultados positivos do método podem ser verificados visualmente com o significativo ganho para a interpretação de dois horizontes que destacam a unidade C mais argilosa e contínua em toda a porção do Compartimento São José dos Campos.

O quarto artigo, “Caracterização de Reservatórios e Predição de Permeabilidade Através de Redes Neurais Supervisionadas” apresenta a aplicação do algoritmo Retropropagação de Erros, em 44 poços do Campo de Namorado, disposta em duas etapas: a caracterização de eletrofácies (reservatório, possível reservatório e não reservatório) e posteriormente a predição de permeabilidade para a fácies reservatório de cada poço. Os resultados obtidos com essa análise permitiram a modelagem do campo para delimitar a distribuição da permeabilidade e, com isso, o melhor entendimento do fluxo no reservatório.

Os quatro artigos compõem o estudo realizado no Instituto de Geociências, Departamento de Geologia e Recursos Naturais, como parte dos requisitos para obtenção do título de Mestre em Geociências.

Com enfoque em caracterização de reservatórios, os artigos dispostos em anexos, apresentam diferentes aplicações de redes neurais artificiais que auxiliam o intérprete com idéias simples e resultados úteis para tal finalidade.

REFERÊNCIAS

- FERNANDES, F. L. 1993. Structural framework and evolution of the Taubaté Basin – SP. Thesis, Universidade Federal de Ouro Preto, MG.
- STUNDNER M. & OBERWINKLER C. 2004. Self-Organizing Maps for Lithofacies Identification and Permeability Prediction. In: SPE Annual Technical Conference and Exhibition held in Houston, p.8.