EDGAR DAVANÇO BRONIZESKI

Classificação de sismofácies carbonáticas
a partir da técnica *Self-Organizing Maps* (SOM)

Versão corrigida

Dissertação apresentada ao Departamento de
Engenharia de Minas e de Petróleo da Escola
Politécnica da Universidade de São Paulo para
obtenção do título de Mestre em Ciências

Área de Concentração: Engenharia de Petróleo

Orientador: Prof. Dr. Cleyton de Carvalho Carneiro

Santos

2018

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, 18 de Dezembro de 2018

Assinatura do autor: _____

Assinatura do orientador: _____

Catálogo-na-publicação
Serviço de Biblioteca e Documentação
Escola Politécnica da Universidade de São Paulo

Bronizeski, Edgar Davanço

Classificação de sismofácies carbonáticas a partir da técnica Self-Organizing Maps (SOM) / E. D. Bronizeski, C. C. Carneiro – versão corr. – São Paulo, 2018.

111 p.

Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Minas e Petróleo.

1.Mapas Auto-organizáveis 2.SOM 3.Classificação de fácies
4.Sismofácies 5.Plataforma carbonática I.Universidade de São Paulo.
Escola Politécnica. Departamento de Engenharia de Minas e Petróleo II.t.
III.Carneiro, Cleyton de Carvalho

Nome: BRONIZESKI, Edgar Davanço

Título: Classificação de sismofácies carbonáticas a partir da técnica *Self-Organizing Maps* (SOM)

Dissertação apresentada ao Departamento de Engenharia de Minas e de Petróleo da Escola Politécnica da Universidade de São Paulo para obtenção do título de Mestre em Ciências

Aprovado em:

Banca Examinadora

Prof. Dr. _____
Instituição: _____
Julgamento: _____

Prof. Dr. _____
Instituição: _____
Julgamento: _____

Prof. Dr. _____
Instituição: _____
Julgamento: _____

DEDICATÓRIA

A minha esposa Janielli que esteve sempre presente e compreensiva durante todo a jornada deste curso de mestrado. Obrigado pelo seu amor e pelo seu apoio.

AGRADECIMENTOS

À Etienne Maurice Gonthier, que solicitamente me liberou das tarefas rotineiras do trabalho para poder me dedicar neste curso de mestrado.

À Escola Politécnica, departamento de Engenharia de Minas e de Petróleo que ofereceu a oportunidade de realizar o curso de mestrado. Agradeço também aos professores do departamento de engenharia de Petróleo, que sempre foram prestativos durante todo o curso.

Aos amigos Reinaldo Machado Tavares, Georges Waisman, Hugo Saar, Leandro Vasconcelos, João Paulo Picolini pelas discussões e ensinamentos que contribuíram de maneira expressiva para a realização deste trabalho.

Não se pode descobrir novas terras
sem aceitar perder de vista a costa
por um longo período.
(André Gide)

RESUMO

BRONIZESKI, Edgar Davanço. **Classificação de sismofácies carbonáticas a partir da técnica *Self-Organizing Maps* (SOM)**, 2018, 111 f. Dissertação (Mestrado) – Escola Politécnica, Universidade de São Paulo, Santos, 2018.

O desenvolvimento de soluções para a classificação de ambientes de sedimentação ainda é um desafio, sobretudo em ambientes de deposição carbonáticos. Em reservatórios de óleo e gás offshore, onde a geometria dos ambientes deposicionais é fundamental para a estimativa de viabilidade técnica e econômica de exploração e produção, a discriminação destes ambientes depende de dados indiretos. Nesse sentido, a sísmica de reflexão e seus respectivos atributos funcionam como robusta ferramenta interpretativa. Grandes reservas de petróleo offshore se encontram em rochas carbonáticas constituídas, sobretudo, por calcita e dolomita. Estes minerais, no entanto, não apresentam resposta sísmica contrastante, já que a diferença de velocidade entre tais minerais é inexpressiva. A utilização da amplitude sísmica, unicamente, não possibilita diferenciar propriedades das rochas carbonáticas, tal como ocorre com rochas siliciclásticas. Compactação e hidrotermalismo são fatores pós-deposicionais que influenciam na alteração das características de rochas carbonáticas, e apresentam difícil detecção nas seções sísmicas. Tais fatores necessitam de uma abordagem que envolva as propriedades dos múltiplos atributos sísmicos para a respectiva caracterização. A técnica *Self-Organizing Maps* (SOM) consiste em ferramenta de análise e visualização de dados vetoriais dentro do espaço dimensional definido pelas variáveis. Trata-se de abordagem multi-atributos em que a identificação e classificação das amostras ocorre de maneira não supervisionada. Neste trabalho são utilizados dados sísmicos dos quais foi extraída a amplitude sobre um horizonte interpretado. Foram utilizados também os atributos do traço sísmico complexo, tais como: envelope, fase instantânea e frequência instantânea, além da impedância acústica e derivada espacial. Estes atributos foram escolhidos em função das características geológicas que representam. Além disso, tais atributos têm sido utilizados em outros trabalhos de classificação de fácies sísmicas. Após uma primeira análise SOM, tendo em vista o realce dos atributos de entrada, foi aplicada a Análise por Principais Componentes (APC). Com base nos componentes obtidos, foi realizada uma segunda análise SOM, obtendo-se assim os

resultados que mais corresponderam ao sistema deposicional analisado. Foram identificados também clinoformas nos dados sísmicos, que apresentam correlação com a classificação dos resultados assim como um mapa de distribuição de rochas carbonáticas foi elaborado pela interpretação das classes identificadas correlacionando-as ao poço 1-SPS-0029 associando-o ao modelo de deposição de uma plataforma carbonática.

Palavras-chave: SOM. Mapas auto organizáveis. Classificação de fácies. Sismofácies. Atributos sísmicos. Plataforma carbonática

ABSTRACT

BRONIZESKI, Edgar Davanço. **Classification of carbonate sismofacies applying the Self-Organizing Maps technique (SOM)**, 2018, 111 f. Dissertação (Mestrado) – Escola Politécnica, Universidade de São Paulo, Santos, 2018.

Developing solutions for classifying sedimentation environments is still quite a challenge, particularly in carbonate sites. It's crucial to use indirect data to differentiate the geometry of depositional environment of oil and gas offshore reservoirs when estimating the productions under technical and economical sights. Therefore, the seismic and its attributes work like a powerful interpretative tool. A great amount of oil reservoirs can be found in carbonate rocks, whose main constitution minerals are calcite and dolomite. Those minerals, however, don't offer a significantly contrasting response to seismic. The seismic amplitude, by itself, isn't a good property estimation to be used on carbonate rocks, as it is on the siliciclastic rocks. Compactation and hydrothermalism are post-depositional influencers on the alteration of carbonate rocks features, and they present a difficult detection on seismic sections. Those factors will need multiple seismic attributes to respective characterization. Self-Organizing Maps (SOM) appear as a tool for analysis and visualization of vectorial data within the dimensional space defined by the variables. That's a multi-featured approach whose samples identification and classification will occur in a non-supervised way. The use of SOM in this paper do not intend to substitute the interpretation of the results, but aims to pop up the intrinsic relations between used attributes. It has been interpreted a horizon, from which it was possible to take the amplitude on seismic as a first attribute. There have also been used acoustic impedance, spatial derivative and complex seismic trace attributes, like reflection strength, instant phase and instant frequency. Such attributes have been chosen by reflecting geological characteristics. After a previous SOM analysis, aiming to highlight raw attributes, there was applied the Principal Components Analysis (PCA). Based on obtained components, a second SOM was performed, leading to results that best corresponded to the depositional system analyzed. Besides, there were identified clinoforms on seismic data, which keep relation with classification of the result.

Keywords: SOM. Self Organizing Maps. Facies Classification. Sismofacies. Seismic Attributes. Carbonate shelf.

Índice de figuras

Figura 1.1 - Localização da bacia de Santos na costa brasileira.....	23
Figura 2.1 - Mapa do embasamento da bacia de Santos e zonas de transferência Fonte: Pereira e Macedo (1989).....	31
Figura 2.2 - Área de estudo de Severino (2000) e em colorido o carbonato miocênico interpretado.....	33
Figura 2.3 - Separação da área da fig. 2.2 em 3 regiões e também a posição das linhas sísmicas utilizadas.....	34
Figura 2.4 - Linha sísmica com as 4 supersequências interpretadas Fonte: Severino (2000).....	35
Figura 2.5 - Mapa do topo da sequência Laranja interpretada por Severino (2000), em tempo, com o retângulo indicando a área de estudo desta dissertação.....	36
Figura 2.6 - Ambientes de sedimentação carbonático.....	38
Figura 2.7 - Diferentes classificações de ambientes de sedimentação carbonático marinho raso.....	40
Figura 2.8 - Classificações e perfis dos ambientes de sedimentação carbonático marinho raso.....	40
Figura 2.9 - Exemplo de disposição de fácies em uma plataforma carbonática.....	41
Figura 2.10 - Exemplo das diferentes condições de deposição de fácies em uma plataforma carbonática com borda.....	41
Figura 2.11 - Deposição de fácies em um ambiente de sedimentação em rampa....	43
Figura 2.12 - Esquema codificador - decodificador utilizado na técnica Vector Quantization.....	46
Figura 2.13 - Vizinhança dos neurônios e raio de atualização.....	48
Figura 2.14 - Exemplo de transformação PCA com os eixos da nova base α_1 e α_2 .	50
Figura 3.1 - Sequência de etapas e análises desenvolvidas nesta dissertação.....	60
Figura 3.2 - Detalhe entre 0 e 1,7s, onde ocorre o evento sísmico mapeado e analisado nesta dissertação, indicado pela seta vermelha.....	62
Figura 3.3 - Mapa estrutural em tempo do topo da seq. Laranja na área de estudo.	63
Figura 3.4 - Mapa das condições de deposição previstas para a área de estudo.....	64

Figura 3.5 - Mapas de Amplitude (a), envelope da amplitude (b), fase instantânea (c) e frequência instantânea (d).....	66
Figura 3.6 - Mapas de impedância (a), primeira derivada (b), isópaca (c) e tempo duplo de reflexão (d) – estrutural em tempo.....	67
Figura 3.7 - Mapas das PC's 1 a 4, respectivamente os itens a, b, c, e d, obtidas apartir dos atributos originais.....	69
Figura 3.8 - Mapas das PC's 5 à 8, respectivamente os ítems a, b, c e d.....	70
Figura 4.1 - Histograma do atributo fase instantânea.....	71
Figura 4.2 - Histograma do atributo amplitude.....	72
Figura 4.3 - Gráfico dos erros de quantização e topográfico por análise.....	74
Figura 4.4 - Mapa de resultado da classificação executada na análise 7.....	75
Figura 5.1 - Matriz-U, evidenciando regiões de alta similaridade de neurônios separadas por alta dissimilaridades, relativas à análise 11.....	82
Figura 5.2 - As principais componentes no domínio do SOM na análise 2.....	83
Figura 5.3 - Divisão da matriz-U em 7 classes.....	84
Figura 5.4 - Sobreposição da divisão de classes aos mapas das principais componentes.....	85
Figura 6.1 - Mapa classificado das amostras no domínio espacial original correspondente à análise 11.....	88
Figura 6.2 - Horizonte interpretado e influências de eventos assinalados em vermelho que levaram a erros na interpretação analisado na seção A-A' da fig. 6.3.....	90
Figura 6.3 - Mapa de localização das seções A-A' e B-B' das figuras 6.2 e 6.4 respectivamente.....	91
Figura 6.4 - Seção sísmica arbitrária e enfoque na distribuição da classe 7.....	92
Figura 6.5 - Mapas de distribuição de rochas carbonáticas e mapa das zonas de deposição.....	93
Figura 9.1- Mapa de distribuição de classes para as análises 1 (a) 2 (b).....	106
Figura 9.2- Mapa de distribuição de classes para as análises 6 (a) 9 (b).....	107
Figura 9.3- Mapa de distribuição de classes para as análises 10 (a) 11 (b).....	108
Figura 10.1 - Exemplo 1 de reconstrução de uma imagem a partir de PC's com perda desprezível de características com 150 PC's ou aproximadamente 96% da variância total.....	110
Figura 10.2 - Exemplo 2 de reconstrução de uma imagem a partir de PC's com perda de características com 150 PC's ou aproximadamente 96% da variância total.....	110

Figura 10.3 - Gráfico de variância cumulativa explicada de acordo com o número de PC's utilizada.....	111
---	-----

Índice de tabelas

Tabela 2.1 - Principais ambientes de sedimentação carbonática.....	39
Tabela 3.1 - Parâmetros de aquisição e processamento dos dados sísmicos.....	61
Tabela 3.2 - Parâmetros para o treinamento do SOM.....	65
Tabela 4.1 - Corelação entre os atributos utilizados nas análises.....	73
Tabela 4.2 - Análises realizadas e respectivos atributos utilizados.....	73
Tabela 4.3 - Combinação linear dos principais componentes a partir dos atributos de entrada.....	76
Tabela 4.4 - Análises realizadas e componentes utilizados.....	77
Tabela 4.5 - Curtose para os principais componentes com distribuição unimodal.....	78
Tabela 4.6 - Componente e percentagem da variância explicada.....	79
Tabela 5.1 - Matriz de correlação entre os componentes para a análise 11.....	81

Índice de quadros

Quadro 2.1 - Classificação das rochas carbonáticas.....	37
---	----

Índice

1 INTRODUÇÃO.....	23
1.1 Apresentação.....	23
1.2 Objetivos.....	26
1.3 Justificativa.....	26
2 REVISÃO BIBLIOGRÁFICA.....	29
2.1 Geologia da área Bacia de Santos.....	29
2.1.1 Evolução tectono-estratigráfica.....	29
2.1.2 Formação Iguape.....	32
2.2 Rochas carbonáticas.....	37
2.2.1 Classificação de carbonatos.....	37
2.2.2 Ambientes de sedimentação carbonáticos.....	38
2.2.3 Separação de fácies carbonáticas.....	43
2.3 Self organizing maps.....	45
2.3.1 Introdução.....	45
2.3.2 Teoria.....	46
2.3.3 Características e utilização.....	49
2.4 Análise por principais componentes (Principal Component Analysis – PCA). .	50
2.4.1 Introdução.....	50
2.4.2 Teoria.....	51
2.4.3 Características e utilização.....	53
2.5 Índice Davies e Bouldin (DBI).....	54
2.5.1 Introdução.....	54
2.5.2 Teoria.....	54
2.5.3 Características e utilização.....	55
2.6 Particionamento K-means.....	56
2.6.1 Introdução.....	56
2.6.2 Teoria.....	56
2.6.3 Características e aplicações.....	57
3 MATERIAIS E MÉTODOS.....	59
3.1 Estrutura da pesquisa.....	59

3.2 Aquisição e característica dos dados.....	60
3.3 Processamento e obtenção dos atributos sísmicos.....	63
3.4 Análises dos SOM e PCA.....	65
4 ESCOLHA DO MELHOR TREINAMENTO DO SOM.....	71
5 ANÁLISES POR SELF-ORGANIZING MAPS UTILIZANDO PCA DOS ATRIBUTOS SÍSMICOS.....	81
5.1 Análise exploratória dos principais componentes.....	81
5.2 Leitura e análise dos resultados.....	82
6 INTERPRETAÇÃO DOS AMBIENTES DE DEPOSIÇÃO CARBONÁTICA.....	87
6.1 Distribuição de classes no domínio espacial original.....	87
7 CONCLUSÕES.....	95
7.1 Utilização da PCA como pré-condicionamento aos SOM.....	95
7.2 Caracterização do ambiente carbonático.....	96
8 BIBLIOGRAFIA.....	99
9 APÊNDICE A.....	105
10 APÊNDICE B.....	109

1 INTRODUÇÃO

1.1 Apresentação

A bacia de Santos abrange uma área de 350.000km² e é limitada a norte pelo Alto de Cabo Frio, ao sul pela plataforma de Pelotas e, na direção E-W, se estende desde a linha da costa continental a oeste até o platô de São Paulo a leste (MOREIRA et al., 2007). Em setembro de 2017 haviam mais de 600 poços já perfurados na bacia de Santos e aproximadamente 40 campanhas de aquisição sísmica perfazendo uma área de aproximadamente 127.000km² (ANP, 2018).

Figura 1.1 - Localização da bacia de Santos na costa brasileira

Fonte: Bronizeski (2018).

O primeiro poço na busca por hidrocarbonetos na Bacia de Santos foi perfurado em 1970, poço 1-PRS-0001 da Petrobras que não encontrou reservas. A produção de hidrocarbonetos foi iniciada após a perfuração do 1-SPS-0011 em 1979 da empresa Pecten, descobridor do campo de gás de Merluza. No entanto, a sua relevância mudou após a descoberta de grande quantidade de petróleo nas

camadas pré-sal em 2007, após a descoberta do campo Tupi. Muitos esforços têm sido feito para a identificação e caracterização dos campos “*offshore*” das camadas do pré-sal. Nesse sentido, o desenvolvimento de estudos que exploram a integração de informações geofísicas auxilia a tomada de decisão, o que torna possível a redução de riscos exploratórios. Dentre as unidades litológicas encontradas na Bacia de Santos, os carbonatos denotam atenção peculiar, devido à ampla variação das características de permeabilidade e porosidade dentro de uma mesma camada, bem como às alterações diagenéticas, sejam estas localizadas ou de pequenas extensões (VON HARTMANN et al., 2012).

Na bacia de Santos são conhecidas 3 formações contendo rochas carbonáticas a saber: formação Barra Velha, formação Guarujá e formação Iguape. Os ambientes de sedimentação destas três formações são diferentes, refletindo a evolução tectônica da Bacia de Santos. Iniciando com o ambiente lagunar – restrito ocorre a deposição da formação Barra Velha, no ambiente plataformar-raso durante a deposição da formação Guarujá e por fim plataformar costeiro durante a deposição da formação Iguape (MOREIRA et al., 2007).

A formação Iguape, objeto de estudo desta dissertação, ocorre no sul da Bacia de Santos. A deposição desta formação ocorreu a partir da discordância do Oligoceno na forma de plataformas carbonáticas, condicionada pelos ciclos de transgressão e regressão marinha (SEVERINO, 2000).

As imagens sísmicas são obtidas pela propagação de ondas elásticas e apresentam relações com informações geológicas e estratigráficas. Os produtos resultantes dos levantamentos sísmicos e suas respectivas propriedades possibilitam informações não apenas sobre a disposição geométrica das camadas geológicas, mas também possibilitam estimativas de propriedades, tais como porosidade, velocidade sísmica e possivelmente inferir o tipo de fluido contido em seus poros. A interpretação destas propriedades é comumente feita utilizando atributos sísmicos aliados à correlação com as informações obtidas a partir de poços (CHOPRA; MARFURT, 2005).

Vale ressaltar que as áreas de exploração e produção não possuem cobertura regular de poços, ou mesmo os dados de poços não amostram todas as fácies existentes. Além disso, as campanhas de perfuração são planejadas para identificar

a existência do reservatório e presença de hidrocarboneto, terminam assim não coletando amostras de fácies em uma malha regular.

Na ausência de poços, ou ainda para estender a abrangência das informações advindas destes, os atributos sísmicos são utilizados pois permitem analisar e identificar sismofácies de uma maneira robusta. No entanto, a utilização de diversos atributos conjuntamente não é simples nem mesmo corriqueira. Cada atributo sísmico possibilita o realce de uma característica específica, por exemplo litologia, indicação de fluído nos poros e etc (BROWN, 1996). A utilização de técnicas de aprendizado de máquina ou “machine learning”, redes neurais e mineração de dados possibilitam a integração de diversos atributos que proporcionam informações importantes para a análise das sismofácies (COLÉOU, POUPON; AZBEL, 2003; GUO; MARFURT; LIU, 2009; KURODA, 2016; SAYAGO et al., 2012; SKIRIUS et al., 1999). Dessa maneira, ao se explorar a utilização dos atributos sísmicos com estas técnicas seria possível obter classificações de sismofácies com maior correspondência às fácies litológicas e, consequente redução do risco exploratório.

Dentre as diversas técnicas de redes neurais artificiais (RNA), há uma série de vantagens na aplicação da rede Self-Organizing Maps (SOM) (COLÉOU, POUPON; AZBEL, 2003). A RNA “Self-Organizing Maps” foi desenvolvida por Kohonen (1990) e é um tipo de rede neural não supervisionada, ou seja, seu treinamento é realizado apenas com os dados de entrada. Uma das principais características do SOM é a redução de dimensionalidade, o treinamento da rede resulta em um mapa bidimensional representando os dados de entrada agrupados por similaridade. Esta técnica é utilizada em tarefas de identificação de agrupamentos e também redução de dimensionalidade (JOLLIFFE, 2002).

Desta forma, este trabalho aplica a técnica SOM a atributos de dados sísmicos com o intuito de obter uma classificação de sismofácies que reflita características de um ambiente deposicional de plataforma carbonática. Tais resultados poderão colaborar para o desenvolvimento de mecanismos e rotinas mais ágeis e simples para classificar tais ambientes carbonáticos.

1.2 Objetivos

Como objetivo principal, este trabalho se propõe a obter um mapa de distribuição de sismofácies em uma plataforma carbonática depositada durante o Mioceno, e localizada no sul da bacia de Santos. Ainda dentro do escopo de estudo, serão identificadas as diferenças ao aplicar a Análise por Principais Componentes (PCA), como etapa prévia ao SOM e os atributos que melhor possuem correlação com os ambientes de deposição como a isópaca e tempo duplo de reflexão (mapa estrutural em tempo).

A compreensão de como a APC altera as características dos dados de entrada, quais características ressalta ou atenua, é de extrema importância para o resultado final.

1.3 Justificativa

A classificação de dados, recorrente nas áreas das geociências, incluindo a classificação de sismofácies, vêm sendo executada por diversos autores em diferentes contextos (CARNEIRO et al., 2012; DUMAY e FOURNIER, 1988; KURODA, 2012; LAI et al., 2016). Diversas técnicas foram empregadas com o intuito de identificar conjuntos representativos de rochas, ou propriedades destas utilizando os dados sísmicos e, notadamente, essas técnicas são aplicadas para identificar e delimitar características bastante específicas, seja pela pequena ocorrência em área ou contrastes sutis (RAFIQ et al., 2016; SKIRIUS et al., 1999; VON HARTMANN et al., 2012).

Estes processos utilizam das mais diversas técnicas relacionadas à classificação e predição de informações geológicas, a saber: regressão linear multi-atributos, análise por principais componentes (PCA), k-means, lógica fuzzy, RNA's supervisionadas e não supervisionadas dentre as principais (RODEN; SMITH; SACREY, 2015; ZHAO et al., 2015; SHAKIBA et al., 2015).

Após comparações entre regressão linear multi-atributos e RNA's para a predição de perfis de poços, foi demonstrada a vantagem nos resultados obtidos pela rede neural probabilística (PNN) sobre a minimização por erro quadrático médio

e também multilayer feedforward network (MLFN) (HAMPSON; SCHUELKE; QUIREIN, 2001). Ao aplicar a PNN, os autores puderam utilizar atributos que não eram apenas oriundos do volume sísmico empilhado. Além disso, foi possível prever perfis que não estavam limitados à impedância acústica ou porosidade, devido às relações não lineares identificadas pela PNN, o que é uma das características das RNA's.

A aplicação da técnica PCA, por exemplo em Linari et al. (2003), mostra resultados em que foi possível a discriminação em zonas de ocorrência de óleo e água. Neste exemplo foram descartadas as principais componentes de menor contribuição à variabilidade dos dados, atribuindo-as ao ruído.

Outras técnicas supervisionadas, como a lógica Fuzzy, necessitam de padrões que descrevam o comportamento dos atributos. Algumas aplicações com esta abordagem são bastante recorrentes, como em Laasri et al. (2015) e Liu et al. (2016), as quais utilizam processos para minimizar os efeitos de características intrínsecas a este método. Incluem-se aqui as regras de fuzzificação das variáveis de entrada dependentes de análise e inicialização adequadas para o sucesso da técnica e que, caso contrário, pode sofrer com erros sistemáticos.

Uma comparação entre diversas técnicas de classificação de sismofácies demonstrou que os melhores resultados de classificação foram obtidos pela rede neural não supervisionada SOM (COLÉOU; POUPON; AZBEL, 2003). Já Sayago et al. (2012) não corrobora este resultado para a rede neural não supervisionada aplicada a um ambiente de paleo karst embora não haja uma discussão sobre os tipos de rede utilizada. Já em outra aplicação, utilizando os SOM e transformada de wavelet como atributo, De Matos, Osório e Johan (2007) identificam um canal formado por arenito, reservatório de hidrocarbonetos, com sucesso dentre as demais sismofácies.

Por se tratar de uma abordagem não supervisionada, a utilização dos SOM faz com que recaia grande responsabilidade à escolha dos atributos utilizados para o treinamento da rede. Este fato cria expectativas quanto ao método utilizado para escolher os atributos a serem analisados pela técnica SOM. Uma das vertentes realiza a escolha de atributos pela predição de um perfil de poço, onde o melhor conjunto de atributos irá prever o perfil com o menor erro, como em Dorrington e

Link (2004) e Kuroda, (2012, 2016). Outra possibilidade é a análise por principais componentes de um conjunto de atributos (RODEN; SMITH; SACREY, 2015; SAYAGO et al., 2012) e por último ocorre também a aplicação de atributos cuja resposta reflita alguma característica das rochas, como: parâmetros elásticos, porosidade ou tipo de fluído (HASHEMI et al., 2014; ITURRARÁN-VIVEROS; PARRA, 2014; MA; GOMEZ; LUNEAU, 2017; MOQBEL; WANG, 2011; SKIRIUS et al., 1999). Dentre estes grupos repete-se a utilização de atributos como amplitude, impedância e aqueles obtidos pela transformada de Hilbert (envelope, fase e frequência instantânea por exemplo). A utilização de atributos que refletem o ambiente de deposição mostra que, somados à usual aplicação do SOM, trazem benefícios para a interpretação final dos resultados como a definição dos ambientes de deposição (MA; GOMEZ; LUNEAU, 2017).

Assim, pretende-se identificar um ambiente de deposição carbonático pela aplicação do SOM em áreas com pouco ou nenhum poço, o que será possível desde que tenhamos um conjunto de atributos relevantes para a classificação. Neste trabalho, tais atributos são representados pela amplitude, frequência e fase instantânea, envelope da amplitude e impedância acústica. Além destes, foram utilizados atributos relacionados ao ambiente deposicional que são: o mapa de isópacas o mapa estrutural em tempo (apresentado como tempo duplo de reflexão). Será aplicada a técnica APC com o intuito do aumento da variância dos atributos escolhidos e assim realçar os contrastes internos, realizando assim uma classificação que reflita estruturas e variações internas a uma área de deposição carbonática marinha rasa.

2 REVISÃO BIBLIOGRÁFICA

2.1 Geologia da área Bacia de Santos

2.1.1 Evolução tectono-estratigráfica

O desenvolvimento de tensores extensionais que propiciaram a fase rifte iniciada no Jurássico deu origem ao processo de separação entre os continentes sul-americano e africano. Foram instalados três tratos de oceanos: Atlântico Equatorial, Atlântico Central e Atlântico Austral com evolução diferenciada (BIZZI et al., 2003).

Localizada no Atlântico Austral, a Bacia de Santos teve os processos que a moldaram iniciados durante a ruptura do Gondwana (SEVERINO, 2000) e sua evolução dividida em 3 estágios: (i) rifte, (ii) transicional e (iii) drifte ou margem passiva. Estes estágios implicam em características tectônicas diferentes entre si, as quais serão abordadas a seguir.

O início da bacia começa com a fase rifte (i) que ocorreu deformando as rochas supracrustais do Cinturão Ribeira (ALMEIDA, 1976). Nesse momento as falhas pré-existentes passaram a ser as condutoras dos esforços. Durante a fase rifte ocorre a distensão da crosta continental, com seu consequente adelgaçamento variando entre 20km junto à charneira, até 10km próximo à borda da plataforma (CHANG; KOWSMANN, 1984), o que implicou na elevação da descontinuidade de Moho.

Durante o Eo-cretáceo, fase transicional (ii), estes esforços distensivos culminam em falhas de transferência que se refletem na Bacia de Santos nas zonas de transferência do Rio de Janeiro, Curitiba e Florianópolis segundo Pereira e Macedo (1989). O mapa do embasamento da bacia de Santos na figura 2.1, apresenta as zonas de transferências e os principais elementos estruturais.

A margem passiva ou distencional (iii) não é marcada por intensas alterações do embasamento. Esta última fase é caracterizada por um período de quiescência tectônica onde ocorre o processo de subsidência térmica da bacia de Santos.

O arcabouço tectônico atual da Bacia de Santos é composto por uma direção principal de estruturas NW-SE, conforme visto na figura 2.1, bem como destacado por diversos outros autores, a saber: Garcia (2013); Meisling, Cobbold e Mount (2001); Modica e Brush (2004); Moreira et al. (2007); Picolini (2017). As zonas de transferência e estruturas do embasamento são apresentadas na fig. 2.1.

Estas zonas de transferência formam divisões internas à bacia e criam barreiras ao influxo de sedimentos. A diferenciação na dispersão dos sedimentos permite a separação entre os embaixamentos norte e sul (PEREIRA et al., 1986).

O extremo leste da Bacia de Santos é dado pelo Platô de São Paulo e este alto do embasamento é limitado pela isóbata de 3500 m (GARCIA, 2013). Relacionado ao processo de rompimento do Gondwana é o resultado do estiramento da crosta continental (MEISLING; COBBOLD; MOUNT, 2001). Deste mesmo processo encontramos também a origem do gráben de Merluza (Mohriak e Paula, 2005; apud GARCIA, 2013).

Uma sequência de vulcões no embasamento com orientação NNE formam a Cadeia de Avedis. A cadeia aparece como uma formação triangular localizada no extremo sul da bacia de Santos entre a dorsal de São Paulo e a plataforma de Florianópolis. Segundo Demercian, Szatmari e Cobbold (1993) esta cadeia vulcânica foi um centro abortado de espalhamento de crosta oceânica. Do mesmo modo, a Dorsal de São Paulo, é uma cadeia vulcânica orientada em E-W com aproximadamente 300km de extensão (AMORIM, 2013). Segundo a mesma autora, a dorsal de São Paulo é também coincidente com a Zona de fratura do Rio de Grande (ZFRG).

O Alto de Florianópolis é uma estrutura importante ao sul da Bacia de Santos, delimitando-a da Bacia de Pelotas. A Charneira de Santos tem idade cretácea e é caracterizada por separar os mergulhos suaves do embasamento, a noroeste, dos mais acentuados, a sudeste (MOREIRA et al., 2007).

O evento de soerguimento da Serra do Mar, que Modica e Brush (2004) determinam entre 87 e 90 Ma, modifica a região emersa da bacia de Santos que, embora não afete o embasamento, altera o padrão de deposição na fase de margem passiva.

Figura 2.1 - Mapa do embasamento da bacia de Santos e zonas de transferência

Fonte: Pereira e Macedo (1989).

Seguindo com os eventos que modificam a dinâmica do aporte sedimentar, ocorreu a reorganização do padrão de drenagem costeiro, onde o rio Paraíba do Sul, principal fornecedor de sedimentos para a Bacia de Santos, mudou seu ponto de encontro com o mar e passou a levar sedimentos para a Bacia de Campos. Segundo Modica e Brush (2004) este processo teve início no neo-Cretáceo e se estendeu até o Paleógeno e apontam como evidência a deposição de sedimentos mais finos ao sul em comparação com o norte da Bacia de Santos.

Ocorre entre o Oligoceno e o Mioceno uma grande subida do nível do mar propiciando um período retrogradacional (MODICA; BRUSH, 2004) onde foram depositadas as sequências E70 a N50-N60. Durante a sequência E70 ocorre o início da deposição da formação Iguape, de especial interesse para este trabalho.

2.1.2 Formação Iguape

A deposição da Formação Iguape iniciou-se durante após a discordância do Oligoceno devido a características particulares. O afastamento das fontes siliciclásticas tornou a Bacia de Santos, em sua parte central e sul, um ambiente de pouca disponibilidade de sedimentos (MODICA; BRUSH, 2004) e permitiu o desenvolvimento da sedimentação carbonática, concretizada por calciruditos e calcarenitos (MOREIRA et al., 2007).

Na carta estratigráfica da Bacia de Campos, ocorrem as sequências E80 à N50-N60, correlatas às da Bacia de Santos e também a deposição de rochas carbonáticas na formação Ubatuba com descrição próxima a da formação Iguape na Bacia de Santos.

As interpretações de Severino (2000) sobre a deposição carbonática ao sul da bacia de Santos está resumida na interpretação de 4 supersequências na área localizada na figura 2.2, em frente ao litoral do Paraná e Santa Catarina. A referida autora identificou 3 domínios distintos: norte, central e sul, que serão mais expressivos a partir da deposição da segunda sequência, a figura 2.3 mostra os diferentes domínio além da localização dos poços utilizados, vale ressaltar que em comum com este trabalho, temos o poço 1-SPS-0029 (P1) e o poço 3-TB-0001 (P3). As sequências são correlacionadas com a glacio-eustasia, associada às mudanças paleoclimáticas globais, corroborados por análise isotópica de oxigênio em um dos poços analisados (SEVERINO, 2000).

A análise destas sequências foi realizada pela interpretação de linhas sísmicas 2-D que recobrem parte da área em destaque no mapa da fig. 2.2, uma destas linhas está reproduzida na figura 2.4.

As sequências estratigráficas identificadas apresentam elementos bem definidos para sua individualização nas seções sísmicas. A discordância entre as sequências é formada por uma reflexão entre arenito (marcador do fim de uma sequência), o desenvolvimento de um ambiente carbonático acompanhando a subida do nível do mar até a ocorrência da superfície de inundação máxima e posterior queda do nível do mar encerrando a sequência.

Figura 2.2 - Área de estudo de Severino (2000) e em colorido o carbonato miocênico interpretado

Fonte: Bronizeski (2018).

A sedimentação predominante nesta sequência é siliciclástica durante o trato de sistema de mar baixo e transgressivo. A deposição carbonática predomina durante o trato de sistema de mar alto nas regiões norte e central quando passa a ser uma rampa intermediária no final deste estágio. A deposição siliciclástica perdura na região sul durante o trato de sistema de mar baixo.

Figura 2.3 - Separação da área da fig. 2.2 em 3 regiões e também a posição das linhas sísmicas utilizadas.

Fonte: Severino (2000).

Figura 2.4 - Linha sísmica com as 4 supersequências interpretadas

Legenda

— Limite superior da Ss Azul (Mioceno Superior)	— Limite superior da Ss Lilás (Mioceno Inferior)	— SIM das Ss Azul, Laranja, Lilás e Marrom
— Limite superior da Ss Laranja (Mioceno Médio)	— Limite superior da Ss Marrom (Mioceno Inferior)	— Topo da Regressão forçada da Ss Azul

Fonte: Severino (2000).

A sedimentação da supersequência Laranja tem pouca expressão durante o trato de sistema de mar baixo. Durante o trato de sistema de mar alto (quando ocorrem as maiores taxas de sedimentação carbonática) a diferenciação entre as regiões se intensifica. Ao término do trato de mar alto, no compartimento norte, temos uma plataforma carbonática agradacional e no compartimento central, dois altos isolados. O mapa da fig. 2.5 apresenta o topo da sequência Laranja e foi digitalizado de Severino (2000). Ainda na fig. 2.5 temos o retângulo em preto indicando a área de estudos desta dissertação, na porção norte definida na fig. 2.3.

A última sequência, Azul, é marcada por sucessivos eventos de rebaixamento do nível do mar, caracterizados por cunhas de regressão forçada na área plataformal e cunhas de mar baixo progradantes na região de talude/bacia. Novamente durante o trato de sistema de mar alto predomina a sedimentação carbonática nas regiões norte e central.

Figura 2.5 - Mapa do topo da sequência Laranja interpretada por Severino (2000), em tempo, com o retângulo indicando a área de estudo desta dissertação

Fonte: Bronizeski (2018).

2.2 Rochas carbonáticas

2.2.1 Classificação de carbonatos

A caracterização das rochas carbonáticas, neste trabalho, utiliza a classificação textural definida por Dunham (1962) e revista por Embry e Klovan, (1971) Embry e Klovan (1971). Nesta abordagem, a primeira grande diferenciação ocorre pela presença ou ausência de lama carbonática. Nesta classificação é importante distinguir entre lama, definida como partículas menores do que 20 microns, e grão, definido como partículas maiores do que 20 microns (DUNHAM, 1962). Já na revisão de Embry e Klovan (1971) para a escala de classificação de rochas carbonáticas temos a inclusão da interação de organismos durante a sedimentação. Esta classificação está apresentada no quadro 2.1.

Quadro 2.1 - Classificação das rochas carbonáticas

Calcário alóctones Componentes originais não organicamente ligados durante a deposição							Calcário autóctone Componentes originais organicamente ligados durante a deposição		
Menos do que 10% de partículas > 2mm				Textura deposicional não reconhecível	Maior que 10% de componentes > 2mm		Por organismos que atuam como obstáculo ("baffle")	Por organismos que encrustam e grudam ("bind")	Por organismos que constroem uma estrutura rígida
Contém matriz com partículas < .03 mm			Não contém matriz		Suportado pela matriz	Suportado pelos componentes > 2mm			
Suportado pela matriz		Suportado pelo grão							
Menos que 10% de grãos (entre 0,03mm e 2mm)	Mais do que 10% de grãos								
Mudstone	Wackestone	Packstone	Grainstone	Cristalino	Floatstone	Rudstone	Bafflestone	Bindstone	Framestone

Fonte: Terra et al. (2010).

2.2.2 Ambientes de sedimentação carbonáticos

As rochas carbonáticas guardam a particularidade de que se formam no local de deposição, seja pela ação de organismos ou pela precipitação química (FLÜGEL, 2004). Segundo Henry Clifton (apud FLÜGEL, 2004)), os fatores que controlam a produção dos carbonatos produzidos organicamente são: luminosidade, temperatura da água e aporte sedimentar. A dependência da biota é também um fator a ser considerado e que está ligado à luminosidade e temperatura da água.

A classificação dos ambientes carbonáticos, sobretudo os marinhos, é alvo de estudo por diversos autores, embora as classificações sejam levemente diferentes, as características básicas de cada ambiente são homogêneas (POMAR, 2001). Utilizando estas características, encontramos em Flügel (2004) os seguintes ambientes de deposição: não marinho, transicional, marinho raso e marinho profundo (fig. 2.6).

Figura 2.6 - Ambientes de sedimentação carbonático

Fonte: Adaptado de Flügel (2004).

Para cada um dos ambientes temos as principais subdivisões, de acordo com Flügel (2004), mostradas na tabela 2.1. A área de estudo deste trabalho é caracterizada como uma plataforma carbonática distalmente escarpada em Severino (2000), que faz parte de um ambiente marinho raso.

Tabela 2.1 - Principais ambientes de sedimentação carbonática

Ambiente	Subdivisões
Não marinho	Exposição sub-aérea Aquático
Transicional	Praia, barreira e lagoas costeiras
Marinho raso	Maré Rampa Plataforma
Marinho Profundo	Altos isolados Sedimentação pelágica Ressedimentação, correntes de fundo e fluxo gravitacional Corais

Fonte: Bronizeski (2018).

A divisão do ambiente marinho raso em 3 categorias ocorre pela declividade da superfície, ligação ou não ao continente, e existência de borda. Embora não tenhamos citado anteriormente, a biota é responsável pelo tamanho do grão de carbonato e taxa de produção. O seu comportamento ao longo do tempo promove a alteração entre os tipos de ambiente marinho. A figura 2.7 mostra os tipos de deposição carbonática do ambiente marinho raso na classificação de três autores, diferentes entre si. Iremos adotar a classificação dos tipos de plataforma descritas em Pomar (2001) (fig. 2.8), pois é adequado para a descrição do ambiente de deposição da área de estudo e também à sua evolução com o passar do tempo.

As plataformas são caracterizadas por um topo com declividade próxima a zero, uma lagoa (ambiente de baixa energia) em direção à terra, uma zona de deposição afetada pela alta energia de ondas ou afetada pela erosão, um recife ou banco de areia ao longo da margem que restringe a circulação, um aumento pronunciado da declividade de poucos graus para 60° ou mais formando um talude e por fim depósitos abundantes de fluxos de massa no sopé do talude vindo da borda da plataforma contendo material oriundo da própria plataforma. Ainda na margem

externa da plataforma pode ocorrer ou não um banco. Na figura 2.9 temos a distribuição de fácies para uma plataforma carbonática sem borda, e na figura 2.10 o modelo para as fácies de deposição de uma plataforma carbonática com borda.

Figura 2.7 - Diferentes classificações de ambientes de sedimentação carbonático marinho raso

Fonte: Adaptado de Pomar (2001).

Figura 2.8 - Classificações e perfis dos ambientes de sedimentação carbonático marinho raso

Fonte: Adaptado de Pomar (2001).

Figura 2.9 - Exemplo de disposição de fácies em uma plataforma carbonática

Fonte: Adaptado de Tinker (1998, apud. Pomar, 2001).

Figura 2.10 - Exemplo das diferentes condições de deposição de fácies em uma plataforma carbonática com borda

Fonte: Adaptado de Flügel (2004).

Na fig. 2.10 temos as litofácies mais comuns descritas abaixo, de acordo com Flügel (2004):

1 – Litofácies comum – mudstone pelágico e wackestone, marga, packstone alóctone, grainstone e brecha;

2 – Litofácies comum – wackestone, margas e folhelhos, ocasionalmente grainstones;

3 – Litofácies comum – mudstone calcário, packstone alóctone e grainstone

4 – Litofácies comum – Mudstone, packstone e grainstone alócone, rudstone e floatstone, brecha;

5 – Litofácies comum – Framestone, baffestone, bindstone, wackestone, floatstone, grainstone e rudstone;

6 – Litofácies comum – Grainstone e packstone;

7 – Litofácies comum – mudstone calcário; wackestone, floatstone, packstone e grainstone;

8 – Litofácies comum – mudstone calcário e dolomítico, wackestone, grainstone, bindstone e brecha sedimentar;

9 – Litofácies comum – mudstone dolomítico, bindstone, camadas alternadas de gipsium e anidrita;

10 – Litofácies comum – paleokarst.

As mesmas características citadas anteriormente são compartilhadas pelo ambiente de alto isolado sendo que a diferença reside na ligação com o continente que inexistente para o alto isolado.

Uma rampa carbonática é uma superfície com caimento suave no fundo do mar, o gradiente vertical da parte rasa à profunda é em geral menor do que 1°, o que representa poucos metros verticais por quilômetro percorrido em direção à parte profunda (fig. 2.11), conforme descrito por Flügel (2004). Diferente da plataforma, não há uma descontinuidade bem marcada limítrofe dos depósitos de baixa energia. No ambiente raso ocorre alta energia pela atuação de ondas vindas do oceano, além de sedimentos oriundos de água rasa, que são transportados para profundidades maiores.

O ambiente com variações suaves de topografia favorece também ao aumento da complexidade e diferenciação das fácies proximais de praia, bem como a produção de carbonatos, que ocorre tanto na parte proximal como média da rampa. Já na parte distal, nas maiores profundidades da rampa, ocorre a produção carbonática autóctone por organismos bentônicos. Podem ou não ocorrer “*buildups*” ainda nesta zona (FLÜGEL, 2004).

Figura 2.11 - Deposição de fácies em um ambiente de sedimentação em rampa

Fonte: Adaptado de Flügel (2004).

As fácies mais comuns para cada divisão da figura 2.11 são:

- 1 – Zona de influência de maré – Evaporitos laminares, mudstone, wackestone, bindstone fenestral, mudstone calcário e dolomítico, packstone e floatstone;
- 2 – Lagoa – Packstone e wackestone, mudstone calcário, floatstone;
- 3 – Banco de areia – grainstone e packstone bioclástico, grainstone e packstone com oóides;
- 4 – Marinho restrito – Mudstone, wackestone bioclástico, packstone, grainstone, floatstone;
- 5 – Marinho franco – Packstone bioclástico, wackestone, packstone, grainstone, floatstone bioclástico e boundstone;
- 6 – Rampa média – Mudstone ou wackestone argiloso, wacke/packstone bioclástico, mudstone pelágico, conglomerado calcário, margas com intraclastos, boundstone, margas;
- 7 – Rampa externa - Mudstone ou wackestone argiloso, mudstone peloidal com espículas, wacke/packstone bioclástico, mudstone pelágico, grainstone laminado, conglomerado calcário, margas;
- 8 – Bacia – Mudstone peloidal com espículas, mudstone ou wackestone argiloso e mudstone pelágico.

2.2.3 Separação de fácies carbonáticas

Conhecer os modelos de deposição ajudam na interpretação dos dados sísmicos, mas como vimos anteriormente as fácies deposicionais não estão restritas e delimitadas. Um método que possa ajudar no detalhamento das propriedades e

sua variação nos ambientes de deposição carbonático é de extrema importância, já que diminui a incerteza do modelo obtido, o que é feito não apenas utilizando os dados sísmicos, mas também seus respectivos atributos (CHOPRA; MARFURT, 2005; SKIRIUS et al., 1999; VON HARTMANN et al., 2012).

Os atributos sísmicos são provenientes dos dados sísmicos, o que os tornam diferentes meios de apresentação e estudo de um conjunto de dados sísmicos (BROWN, 1996). A finalidade de utilizar atributos é obter uma visualização destacada das camadas litológicas, propriedades do reservatório, bem como de realçar a estrutura ou ambiente deposicional (CHOPRA; MARFURT, 2005). A lista de atributos sísmicos disponível é bastante extensa, uma revisão histórica dos atributos sísmicos é apresentada em Chopra e Marfurt (2005), assim como propostas de classificação dos atributos em categorias por Brown (1996) e Barnes (2007) .

A utilização de atributos nos trabalhos que abordam a classificação de fácies mostra que não existe um conjunto padrão. Entretanto, a impedância acústica é apresentada como um atributo bastante frequente além dos atributos complexos do traço sísmico (HASHEMI et al., 2014; ITURRARÁN-VIVEROS; PARRA, 2014; KURODA, 2012; LINARI et al., 2003; MA; GOMEZ; LUNEAU, 2017; MOQBEL; WANG, 2011; RODEN; SMITH; SACREY, 2015; SAYAGO et al., 2012)). Alguns autores executam uma seleção prévia entre os atributos, utilizando algoritmo genético ou evolutivo (DORRINGTON; LINK, 2004; KURODA, 2012), análise por principais componentes RODEN; SMITH; SACREY, 2015; SAYAGO et al., 2012 , algoritmo imunológico (SARASWAT; SEN, 2012) enquanto outros autores não realizam uma etapa de pré-seleção entre os atributos disponíveis para a análise (MA; GOMEZ; LUNEAU, 2017; MOQBEL; WANG, 2011; VON HARTMANN et al., 2012).

A análise de diversos atributos se apresenta com um problema de difícil resolução. Por isso a aplicação de algoritmos capazes de identificar as particularidades intrínsecas entre os atributos é altamente necessária. As redes neurais artificiais (RNA's) são ferramentas capazes de capturar estas relações entre os atributos inclusive relações não lineares (HAGAN et al., 2002). As comparações entre a aplicação das RNA's supervisionadas e não supervisionadas nas geociências mostram que a escolha é dependente da análise em cada caso, não existe uma regra a ser aplicada a todos os casos (BHATTACHARYA; CARR; PAL, 2016; SAYAGO et al., 2012). No caso específico deste trabalho, optamos por utilizar

a técnica SOM. A escolha foi feita pela indisponibilidade de amostras de rochas, além de que os poucos perfis de poços não permitem o treinamento supervisionado, como em muitos dos trabalhos citados anteriormente. A pouca disponibilidade de dados impediu utilizar uma abordagem mais completa, que seria aplicar duas etapas de classificação: uma não-supervisionada e uma supervisionada, como em Moqbel e Wang (2011), o que possibilitaria a correlação final entre fácies encontradas e tipos de rocha.

2.3 Self organizing maps

2.3.1 Introdução

O estudo das RNA's começou entre os séculos XIX e XX em trabalhos multidisciplinares de física, psicologia e neurofilosofia pelos cientistas Hermann von Helmholtz, Ernst Mach e Ivan Pavlov (HAGAN et al., 2002). A pesquisa destes cientistas mostra que a conexão entre os neurônios no cérebro humano é, em sua maioria, desenvolvida durante a vida a partir de eventos. Assim, é dito que as conexões entre os neurônios são desenvolvidas ao longo da vida, ou seja, a rede é treinada pelo aprendizado do indivíduo durante as experiências vivenciadas.

De uma forma descritiva, McCulloch e Pitts (1943) descrevem as iterações de uma RNA's como um caso de lógica proposicional, onde a excitação dos neurônios ocorre por semelhança “tudo-ou-nada”. A evolução das RNA's a partir de então avançou e passou por refinamentos, iniciando com a regra de aprendizado de Hebb (Hebb, 1949 apud Hagan et al., 2002), multi layer perceptron, o algoritmo de treinamento por backpropagation (Rumelhart e McClelland, 1986 apud Hagan et al., 2002), self-organizing maps (KOHONEN, 1990) e a resolução do problema fundamental do deep-learning .

Pode ser feita uma distinção entre as RNA's pelo seu respectivo tipo de treinamento, classificando-as em supervisionadas ou não-supervisionadas. As redes supervisionadas necessitam de valores de entrada e valores de saída para realizar o treinamento. O treinamento ocorre pela aplicação dos valores de entrada e

observação dos resultados, qualquer diferença entre o valor obtido e o esperado indica um erro e necessidade de novo treinamento da rede. As redes não supervisionadas necessitam apenas dos dados de entrada para treinamento.

2.3.2 Teoria

Os SOM representam um tipo de rede neural não supervisionada desenvolvida por Kohonen (KOHONEN, 1990). Esta técnica realiza uma redução de dimensionalidade pois representa em um mapa (2-D) as informações de similaridade existentes entre amostras compostas de diversos atributos e é utilizado para a classificação e análise de dados (Vesanto e Alhoniemi, 2000).

A técnica SOM teve início com o reconhecimento de padrões para transmissão de sinais conhecido como “Vector quantization” (VQ), e foi desenvolvida para a transmissão de informação focada em velocidade de transmissão e minimização da necessidade de armazenamento do sinal preservando a qualidade de reprodução (GRAY e OLSHEN, 1997).

O reconhecimento dos padrões, mencionado acima, depende da identificação de vetores codificadores, referidos como “codebooks” em inglês. O princípio de funcionamento do VQ é de um conjunto codificador – decodificador, exemplificado na figura 2.12, em que é feita a codificação de um sinal de entrada com distorção mínima, seguido pela transmissão deste sinal amostrado, e uma posterior decodificação que recupera o sinal original (GRAY, 1984).

Sendo x o sinal de entrada e x' o sinal recuperado, a distorção (D) representa a fidelidade do vetor reconstruído e é calculada pela equação 1 (GRAY, 1984). O decodificador é escolhido para que minimize a distorção, a r -ésima potência do erro de reconstrução.

Figura 2.12 - Esquema codificador - decodificador utilizado na técnica Vector Quantization

Fonte: Adaptado de Grey (1984)

$$D = \sum_x \|x - x'(c(x))\|^r = \int \|x - x'(c(x))\|^r p(x) dx \quad (1)$$

O termo $x'(c(x))$ é usualmente referenciado na literatura como m_c , o vetor de referência de menor distância, que pertence ao conjunto de “codebooks” descrito como $\{m_i(t): m_i \in \mathbb{R}^N, i=1,2,\dots,k\}$. Utilizando esta informação é reescrita a eq. 1 na sua forma mais comum (eq. 2) nas referências do SOM.

$$D = \sum_x \|x - m_c\|^r = \sum \|x - m_c\|^r p(x) dx \quad (2)$$

A escolha de $r = 2$ na eq. 2 é usual e desta maneira a VQ aproxima a função densidade de probabilidade $p(x)$ (KOHONEN, 1990). O vetor m_c é escolhido pela comparação simultânea entre o vetor x e cada um dos vetores de referência (“codebooks”) m_i segundo a distância Euclidiana (eq. 3).

$$\|x - m_c\| = \min_i \{\|x - m_i\|\} \quad (3)$$

Esta competição entre os vetores m_i como apresentado pela eq. 3 e a escolha de m_c caracteriza este algoritmo como “winner-take-all” (GRAY 1984). Cada um dos “codebooks” m_c encontrados representa um dos conjuntos identificados nos dados de entrada, e são denominados como “*best matching units*” (BMU). A cada época de treinamento da rede neural em que é identificado o m_c é executada uma atualização cuja finalidade é diminuir a distância entre $x(t)$ e $m_c(t)$, conforme mostrado na eq. 4.

$$m_c(t+1) = m_c(t) + \alpha(t)[x(t) - m_c(t)] \quad (4)$$

Onde $\alpha(t)$ é a taxa de aprendizado, uma função decrescente no intervalo $[0,1]$.

Os SOM têm como característica uma conexão entre os neurônios m_i vizinhos e que é inserida por um raio de atualização (N_c) centrado em m_c (eq. 5) tal como apresentado na figura 2.13.

$$m_i(t+1) = \begin{cases} m_c(t) + \alpha[x(t) - m_i(t)] & \text{p/ } i = N_c(t) \\ m_i(t) & \text{p/ } i \neq N_c(t) \end{cases} \quad (5)$$

A conexão entre os neurônios é aumentada quando os neurônios são representados por hexagonos nos quais uma superfície toroidal pode ser formada. A

superfície toroidal garante que não existirão bordas, permitindo a identificação de similaridade de uma forma contínua entre as celas da matriz-U (KOHONEN, 1990).

Figura 2.13 - Vizinhança dos neurônios e raio de atualização

Fonte: Bronizeski (2018).

É esperado que, mesmo após as iterações de treinamento do SOM não tenhamos uma correspondência exata entre x e m_c , portanto teremos ainda algum erro. É natural a obtenção destas diferenças, bem como o fato de não ser necessário um número de iterações de treinamento excessiva, afim de evitar o “*overfitting*”, estado que a RNA torna-se capaz de reproduzir apenas os dados de treinamento.

Após o treinamento da RNA é avaliado o erro de quantização (QE), o erro na representação de um grupo de vetores de entrada pelo seu respectivo BMU correspondente (KOHONEN, 1990), é obtido pela média das distâncias entre os vetores de entrada e o centróide da classe que os amostra (B.M.U.) e expresso pela equação 6.

$$QE = \frac{1}{N} \sum_{i=1}^N \|x_i - B.M.U._i\| \quad (6)$$

Outra característica avaliada ao término do treinamento é o erro topográfico, proporção de todos os vetores de entrada (x) para os quais o primeiro e segundo BMU's não são adjacentes. O erro topográfico mede a preservação de topologia (Kohonen, 1990).

A qualidade de treinamento da rede do SOM é influenciada pela quantidade de células utilizadas e vemos dois tipos diferentes de abordagem. A escolha do tamanho da rede que obtenha o menor erro topográfico (KASKI; KOHONEN, 1996) é uma das opções de abordagem. A segunda abordagem utiliza a regra empírica $5\sqrt{N}$, onde N é o número de pontos para análise, focado na redução do tempo computacional (VESANTO e ALHONIEMI, 2000) e utilizado em De Matos, Osório e Johan (2009), Kuroda (2012), Ceréghino e Park (2009) e Carneiro et al. (2012).

O meio mais comum de visualização do treinamento dos SOM é pela matriz de distâncias unificada (matriz-U) (ULTSCH; MÖRCHEN, 2005). A matriz-U torna possível a representação de cada um dos BMU's colorido pela distância de seus vizinhos e normalizados pela maior distância (ULTSCH; MÖRCHEN, 2005). Caso a distância seja representada como altura e visualizada tridimensionalmente, criaria picos e vales identificando as compartimentações em classes (LÖTSCH e ULTSCH, 2004).

Conhecendo o BMU é possível representar a variação dos N atributos de entrada no domínio do SOM e a estes mapas denominamos “*component plots*”. Tais mapas auxiliam a identificar: áreas na matriz-U de maior contribuição individual dos atributos; correlação entre os atributos e os BMU's na matriz-U; além de variações nos atributos relacionadas às barreiras de dissimilaridade na matriz-U. Exemplos desta correlação e interpretações são encontradas em Saski e Kohonen (1996).

2.3.3 Características e utilização

A utilização do SOM necessita de alguns passos prévios para garantir que os resultados sejam confiáveis:

- Uma boa técnica é realizar múltiplas passagens de treinamento da rede, por exemplo em duas etapas. A primeira etapa de treinamento, realizada com uma taxa de aprendizado alta para determinar cada um dos vetores de referência (“*codebooks*”), e uma segunda etapa com taxa de aprendizado padrão e inferior à primeira, para ajustar mais precisamente os vetores de referência. O intuito é que o processo de aprendizado da RNA se torne mais eficiente (JOLLIFFE, 2002).

- Normalização das entradas entre valores 0 e 1. Este procedimento evita que um atributo seja dominante nas medidas de distância. Da mesma forma diminui o

número de vetores de referência necessários pois a variação entre máximo e mínimo de todos os atributos estará limitada ao mesmo intervalo (JOLLIFFE, 2002).

2.4 Análise por principais componentes (Principal Component Analysis – PCA)

2.4.1 Introdução

A PCA é uma técnica de transformação ortogonal linear que leva os dados n -dimensionais para um novo sistema de coordenadas. Neste novo sistema de coordenadas, os dados são decompostos em componentes linearmente independentes entre si, em que a ordem dos componentes nos resultados indica a variância indo da maior para a menor. A figura 2.14 mostra de maneira gráfica um exemplo desta transformação.

Figura 2.14 - Exemplo de transformação PCA com os eixos da nova base α_1 e α_2

Fonte: Bronizeski (2018).

A PCA foi desenvolvida por Pearson (1901) e Hotelling (1933), que desenvolveram esta metodologia independentemente. Temos em Pearson (1901) a motivação de encontrar um plano ou linha que defina os dados, não sendo feita nenhuma suposição para a escolha de uma das variáveis do problema estudado como independente. A reta de melhor ajuste passa pelo centroide dos dados, ou seja seu centro geométrico, e possui o menor momento possível para o conjunto de pontos estudado. Este método descrito em Pearson (1901) depende do

conhecimento apenas da média, desvio padrão e correlação entre as variáveis. É também demonstrado que as retas de melhor e pior ajuste podem ser descritas como os eixos maior e menor de uma elipse, respectivamente.

Em Hotelling (1933) as premissas para os componentes estudados são de que temos um sistema cujas componentes possuem correlação zero e variância unitária, estas componentes serão responsáveis por determinar o valor da variável x e são caracterizadas (as componentes de entrada) pela média, variância e covariância. Também aqui ocorre a identificação dos eixos de uma hipérbole com os eixos das principais componentes e inclui a prova matemática. O desenvolvimento deste autor utiliza os multiplicadores de Lagrange para encontrar os autovetores e autovalores das principais componentes. Além da dedução algébrica da técnica, o autor fornece uma resolução iterativa.

2.4.2 Teoria

Reproduzimos aqui a dedução matemática desta técnica como é encontrada em Jolliffe (2002). Sendo x um conjunto de dados com a matriz de covariância conhecida Σ com as seguintes características : $\alpha_1^T x$ é a projeção do conjunto de dados x na direção α_1 ; $\alpha_1^T \Sigma \alpha_1$ é a variância do conjunto de dados x na direção α_1 ; o vetor α_1 deve possuir norma 1 e que implica a relação $\alpha_1^T \alpha_1 = 1$. O objetivo é obter o vetor α_1 que possua a maior variância possível, o que é expresso por $\text{var}[\alpha_1^T x] = \alpha_1^T \Sigma \alpha_1$. Tem-se então o problema de maximizar $\alpha_1^T \Sigma \alpha_1$ sujeito a $\alpha_1^T \alpha_1 = 1$ e neste tipo de problema aplicaremos a técnica dos multiplicadores de Lagrange (eq. 7) e o termo λ é o multiplicador de Lagrange.

$$\alpha_1^T \Sigma \alpha_1 - \lambda (\alpha_1^T \alpha_1 - 1) \quad (7)$$

A diferenciação da eq 7 com respeito a α_1^T leva às eq. 8 e consequentemente à equação 9.

$$\Sigma \alpha_1 - \lambda \alpha_1 = 0 \quad (8)$$

$$\sum \alpha_1 = \lambda_1 \alpha_1 \quad (9)$$

Portanto a maximização da variância (eq. 7) leva a equação 9, de onde obtemos que o parâmetro α_1 é o autovetor e λ_1 o seu autovalor associado. Para a matriz de variância Σ o par α_1 e λ_1 formam a primeira principal componente. De uma forma geral, a n-ésima principal componente de x é $\alpha_n^T x$ em que α_n é o n-ésimo autovetor e $\text{var}(\alpha_n^T x) = \lambda_n$ é o n-ésimo autovalor.

Como prova da afirmação acima, temos o cálculo da segunda principal componente a seguir. Os autovetores devem ser não correlacionados e, portanto, a covariância entre eles será zero. Dessa maneira temos que, para o conjunto de dados x decomposto em duas principais componentes α_1 e α_2 , a covariância $[\alpha_1^T x, \alpha_2^T x] = 0$, o que implica nas relações da eq. 09 abaixo:

$$\text{cov}[\alpha_1^T x, \alpha_2^T x] = \alpha_1^T \sum \alpha_2 = \alpha_2^T \sum \alpha_1 = \alpha_2^T \lambda_1 \alpha_1 = \lambda_1 \alpha_2^T \alpha_1 = \lambda_1 \alpha_1^T \alpha_2 \quad (10)$$

A igualdade da eq. 10 permite escrever as relações (eq. 11) indicativas da falta de correlação entre as principais componentes $\alpha_1^T x$ e $\alpha_2^T x$.

$$\begin{aligned} \alpha_1^T \sum \alpha_2 &= 0, & \alpha_2^T \sum \alpha_1 &= 0 \\ \alpha_1^T \alpha_2 &= 0, & \alpha_2^T \alpha_1 &= 0 \end{aligned} \quad (11)$$

Novamente aplicando a técnica dos multiplicadores de Lagrange para maximizar a função $\alpha_2^T \sum \alpha_2$ sujeito às limitações $\alpha_2^T \alpha_2 = 1$ e uma das relações da eq. 11, que indicam o fato das direções das componentes α_1 e α_2 serem ortogonais entre sí, obtemos a eq. 12.

$$\frac{\partial \alpha_2^T \sum \alpha_2}{\partial \alpha_2^T} = \frac{\lambda_2 \partial (\alpha_2^T \alpha_2 - 1)}{\partial \alpha_2^T} + \frac{\theta \partial (\alpha_2^T \alpha_1)}{\partial \alpha_2^T} \quad (12)$$

Executando a diferenciação proposta e re organizando os termos obtemos (eq. 13):

$$\sum \alpha_2 - \lambda_2 \alpha_2 - \theta \alpha_1 = 0 \quad (13)$$

A multiplicação da eq. 13 por α_1^T nos leva à eq. 14:

$$\alpha_1^T \sum \alpha_2 - \lambda_2 \alpha_1^T \alpha_2 - \theta \alpha_1^T \alpha_1 = 0 \quad (14)$$

Segundo a eq. 11 os dois primeiros termos da eq. 14 são zero e como $\alpha_1^T \alpha_1 = 1$, temos que $\theta = 0$. Substituindo este resultado na equação 13 chegamos às equações 15 e 16 em que α_2 é o autovetor e λ_2 é o seu respectivo autovalor. Dessa forma fica comprovada a forma geral das principais componentes como queríamos demonstrar.

$$\sum \alpha_2 - \lambda_2 \alpha_2 = 0 \quad (15)$$

$$\sum \alpha_2 = \lambda_2 \alpha_2 \quad (16)$$

2.4.3 Características e utilização

A transformação do conjunto de dados x na nova base composta pelos n -principais componentes ocorre com a máxima variância possível em cada componente. Este fato permite agrupar apenas as componentes com variância considerável, ignorando as componentes com pequena variância. Uma aplicação desta técnica voltada às ciências da terra é encontrada em Linari et al. (2003); Roden, Smith e Sacrey, (2015) e Unglert, Radić e Jellinek (2016).

A PCA permite também que sejam estimadas variáveis aleatórias no conjunto de dados x pela transformação linear do conjunto de observações (y) e como descrito em Jolliffe (2002), a variância residual nesta predição será minimizada pela utilização das primeiras principais componentes (as de maior variância).

No apêndice B são apresentadas outras características da técnica PCA que inclusive corroboram o fato de utilizarmos todas os componentes sem executar a redução de dimensionalidade.

2.5 Índice Davies e Bouldin (DBI)

2.5.1 Introdução

O índice de Davies e Bouldin (DBI) é uma métrica para a avaliação da partição de um conjunto de dados em classes (Davies e Bouldin, 1979). Dentre as proposições para a criação do DBI, estavam a definição do número de classes sem a intervenção humana, deveria ser possível a aplicação a conjuntos de dados hierárquicos, ser computacionalmente factível para grandes conjuntos de dados, deveria ser capaz de tratar dados de dimensionalidade arbitrária.

O DBI necessita do conjunto de dados de entrada e não necessita de nenhum tipo de relação a priori entre os dados e agrupamentos. O algoritmo se utiliza de características do próprio conjunto de dados para avaliar o número de divisões e obtém esta divisão ótima de classes, pela relação entre a distância intra-classe e inter-classes.

2.5.2 Teoria

Seja o agrupamento C cujos membros sejam $X_1, X_2, \dots, X_m \in E_p$. A definição de uma métrica de separação entre agrupamentos $R(S_i, S_j, M_{ij})$ permitirá o cálculo da similaridade entre dois agrupamentos. O termo M_{ij} é a distância entre os vetores característicos dos agrupamentos i e j cuja dispersão é S_i e S_j respectivamente. A equação 17 é a medida de separação entre os dois agrupamentos i e j .

$$R_{ij} = \left\{ \frac{S_i - S_j}{M_{ij}} \right\} \quad (17)$$

A dispersão S_i é obtida pela equação 18 Sendo S a medida de espalhamento dos valores do conjunto de dados x , pertencentes ao agrupamento i cujo valor central é z .

$$S_i = \left(\frac{1}{T_i} \sum_{j=1}^{T_i} \|X_j - A_i\|^q \right)^{1/q} \quad (18)$$

Onde: T_i é o número de vetores no agrupamento i e A_i é o centróide do agrupamento i .

A distância intra-classes é obtida pela equação 19, a soma das distâncias entre os vetores que compoem os centróides dos agrupamentos i e j .

$$M_{ij} = \left(\sum_{k=1}^N |a_{ki} - a_{kj}|^p \right)^{1/p} \quad (19)$$

A escolha de $p = 2$ na equação 19 torna M_{ij} a distância Euclidiana entre os centróides. O mesmo acontece pela escolha de $q = 1$ na equação 18, tornando S_i a distância média Euclidiana entre os vetores do cluster i e o centróide do cluster i .

A similaridade média do sistema de agrupamentos é $\bar{R}$, definida pela equação 20. A escolha ótima do número de agrupamentos será aquela cuja similaridade média seja a menor.

$$\bar{R} \equiv \frac{1}{N} \sum_i R_i \quad (20)$$

Onde $R_i \equiv \text{máximo de } R_{ij}, i \neq j$.

2.5.3 Características e utilização

O DBI é utilizado como ferramenta para os problemas de partição de dados não supervisionado como indicativo do número de agrupamentos ótimos e para casos de aplicação à classificação de fácies este índice representa a quantidade de litologias diferentes identificadas (DE MATOS; OSORIO; JOHANN, 2007; KOURKI; ALI RIAHI, 2014; SARASWAT; SEN, 2012).

2.6 Particionamento K-means

2.6.1 Introdução

O algoritmo k-means é uma técnica de particionamento de um conjunto de dados em um número k de grupos. Não necessita de anotação dos dados de entrada, sem a definição de categorias ou grupos, é incluído como uma técnica de aprendizado não supervisionado. Ao término da classificação serão obtidos os centróides dos conjuntos formados e a associação de cada um dos dados de entrada a uma classe de tal forma que o centróide que defina um grupo tenha a menor distância possível para os dados que representa.

Estes resultados vindos do k-means são importantes pois ao realizar o agrupamento pode-se: identificar características comuns a cada um dos grupos, permitir a compressão de dados ou ainda identificar pontos anômalos (MACKAY, 2003).

O algoritmo k-means foi apresentado em dois diferentes artigos: Forgy (1965) (apud MACQUEEN, 1967)) e Macqueen (1967).

2.6.2 Teoria

O algoritmo k-means particiona o grupo de N dados, denotado por $x^{(n)}$ em k grupos caracterizados pela média $m^{(k)}$, apresentamos a seguir a descrição deste algoritmo presente em Mackay (2003). Cada ponto pertence ao seu centro próximo tornando-se parte de um dos k grupos, vale ressaltar que o centróide ou média $m^{(k)}$ não necessita ser um dos pontos de N (PEDREGOSA et al., 2011). A identificação do grupo que cada um dos pontos pertence é definida pela equação 21.

$$\hat{k}^{(n)} = \underset{k}{\operatorname{argmin}} \{ d(m^{(k)}, x^{(n)}) \} \quad (21)$$

A métrica que define a distância entre os pontos, d , expressa pela equação 22, avalia cada um dos n pontos aplicando a norma L2, a distância Euclidiana em relação a cada um dos centróides.

$$d(x, y) = \sum_i \|x_i - y_i\|_2 \quad (22)$$

Uma representação alternativa desta associação do ponto $x(n)$ ao agrupamento k é a variável indicadora, conhecida como responsabilidade, $r_k^{(n)}$ (eq. 23) associando 1 se a média k é aquela próxima ao ponto $x(n)$, ao contrário $r_k(n)$ será zero.

$$r_k^{(n)} = \begin{cases} 1 & \text{se } \hat{k}^{(n)} = k \\ 0 & \text{se } \hat{k}^{(n)} \neq k \end{cases} \quad (23)$$

A cada iteração a média do agrupamento é ajustada tal que reflita a média dos pontos que fazem parte, segundo a equação 24.

$$m^{(k)} = \frac{\sum_n r_k^{(n)} x^{(n)}}{R^{(k)}} \quad (24)$$

Onde $R(k)$ é a responsabilidade total do grupo k (eq. 25).

$$R^{(k)} = \sum_n r_k^{(n)} \quad (25)$$

2.6.3 Características e aplicações

Ao utilizar a implementação do algoritmo k-means pela biblioteca scikit-learn (em Python) é possível escolher os pontos iniciais dos centróides de duas maneiras:

- Associação aleatória dos pontos a um dos agrupamentos e cálculo da média;
- k-means++, em que os centróides de cada um dos agrupamentos é escolhido aleatoriamente mas garantindo que os centróides estejam distantes entre si.

A técnica k-means é de fácil compreensão e sempre converge, nem sempre para o mínimo global (Pedregosa et al, 2011). Algumas das desvantagens está na

classificação equivocada em conjuntos de dados de formato alongado, ou formas irregulares (Pedregosa et al., 2011).

3 MATERIAIS E MÉTODOS

3.1 Estrutura da pesquisa

No intuito de desenvolver uma rotina de classificação em ambiente carbonático com maior eficiência, a pesquisa aqui proposta irá investigar o pré condicionamento dos atributos de entrada para as análises por SOM. Os diversos treinamentos da rede do SOM serão analisados quanto aos erros de quantização e topográfico utilizando combinações diferentes de atributos . Em outra abordagem, as comparações se darão entre treinamentos com quantidades diferentes dos principais componentes.

As análises também incorporam o impacto de dois atributos não tão comuns em trabalhos de classificação de fácies: (i) a isópaca; (ii) o tempo duplo de reflexão (mapa estrutural em tempo). Atributos como estes estão relacionados ao ambiente de deposição e são importantes para a delimitação da distribuição de fácies. Encontramos um paralelo a esta abordagem com atributos relacionados à deposição em Ma, Gomez e Luneau (2017) que utiliza a distância à costa durante o processo de classificação de fácies.

Na figura 3.1 temos uma indicação das etapas de análise desta dissertação. Nesta área de estudo a correlação entre a classificação da melhor análise acontecerá com o mapa dos ambientes mais prováveis (fig. 3.4). Vale ressaltar que mesmo este mapa é apenas uma possibilidade para os resultados, uma vez que o mesmo foi obtido pela interpretação do mapa estrutural.

Figura 3.1 - Sequência de etapas e análises desenvolvidas nesta dissertação

Fonte: Bronizeski (2018).

3.2 Aquisição e característica dos dados

A fonte para os dados sísmicos e de poços foi a Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP) disponibilizados segundo protocolo 61298 de Junho de 2016. Foram disponibilizados 100km² do processamento 0268_Santos_W_1 e os poços 1-SPS-0029, 1-SCS-0026, 3-TB-0001 e 3-TB-0002.

Os dados sísmicos utilizados neste trabalho fazem parte do levantamento sísmico 0268_Santos_W_1, adquirido e processado pela PGS em 2010, a campanha de aquisição e processamento possuem o mesmo nome no banco de dados da ANP. Esta aquisição sísmica foi realizada em 4 partes com área total de 4906km². A direção das linhas de aquisição foi N-S para as fases I, II e III, enquanto para a fase IV foi E-W. Os dados utilizados nesta dissertação pertencem à fase I, portanto, adquiridos na direção N-S e a área adquirida é de aproximadamente 705km².

A aquisição foi realizado pelo navio sísmico Ramform Valiant, em que eram rebocados 10 cabos “*streamers*”, com 480 canais cada, à profundidade de 9m e comprimento de 8100m. A cobertura de aquisição nominal é 81, separação entre in-lines de 25m e entre cross-lines de 12,5m. Foram utilizados dois conjuntos de fontes na configuração flip-flop com separação de 50m, cada conjunto possuía pressão de 2000 psi e volume de 3090 in³ distribuído em um arranjo de 3 canhões.

A aquisição dos dados ocorreu com taxa de amostragem de 2ms e foi utilizado o sistema de registro Syntrak 960-24. O tempo total de registro foi de 9216ms com filtro corta baixas 3Hz/ 12dB por oitava e filtro corta altas 206Hz/276dB por oitava. Foi utilizado o datum SAD-69 e projeção UTM nas malhas de aquisição e processamento.

O processamento sísmico foi realizado sobre uma malha 12,5x25m e foi utilizado o algoritmo de migração Kirchhoff em tempo, mas os dados disponibilizados estavam decimados para uma malha 25x25m. Um resumo das principais informações é apresentado na tabela 3.1, na figura 3.2 é apresentada a in line 9928 em detalhe com o tempo máximo de 1,7 segundos e o horizonte interpretado que foi objeto de estudo deste trabalho.

Tabela 3.1 - Parâmetros de aquisição e processamento dos dados sísmicos

Parâmetro de aquisição	Valor
Espaçamento entre linhas de registro	25
Espaçamento entre tiros	25 flip-flop
Profundidade da fonte	7m +/- 1m
Quantidade de cabos	10
Comprimento dos cabos	8100m
Quantidade de canais por cabo	480
Intervalo entre grupos de hidrofones	12,5m
Direção de aquisição	N-S (fases I, II e III) e E-W (fase IV)
Cobertura nominal	81
Equipamento de registro	Syntrak 960-24
Tempo de registro	9216ms
Razão de amostragem	2ms
Filtro de aquisição	3Hz / 12dB – 206Hz/ 276dB
Processamento	
Algoritmo de migração	Kirchhoff em tempo
Intervalo entre in-lines	25,0m
Intervalo entre cross-lines	12,5m
Dado disponibilizado	
Intervalo entre in-lines	25,0m
Intervalo entre cross-lines	25,0m

Fonte: Bronizeski (2018).

Figura 3.2 - Detalhe entre 0 e 1,7s, onde ocorre o evento sísmico mapeado e analisado nesta dissertação, indicado pela seta vermelha

Fonte: Bronizeski (2018).

O horizonte interpretado (fig. 3.3), o limite superior da sequência Laranja que foi depositado durante o Mioceno, corrobora algumas das feições estruturais identificadas e interpretadas por Severino (2000), que classifica a área como uma rampa carbonática distalmente escarpada. Surgem também as feições de um banco carbonático próximo ao talude e uma pequena depressão próxima a este, alinhando estes fatos tem-se o perfil de uma plataforma carbonática com borda. A interpretação possui detalhes que não estavam presentes nas linhas 2-D interpretadas em Severino (2001) pois a estrutura do banco possui pequenas dimensões.

Ainda que os dados tenham dimensão em tempo duplo de reflexão e não profundidade, as estruturas permaneceriam inalteradas. As camadas acima do topo da sequência Laranja têm deposição plano paralela o que não criará falsas estruturas em tempo. Dessa forma é possível associar o mapa estrutural em tempo ao ambiente deposicional carbonático de uma plataforma com borda e também identificar as zonas de deposição da fig. 2.5, embora não identifiquemos a base das ondas normais e de tempestade. Esta previsão de ambientes é apresentada na fig. 3.4 com 5 divisões maiores: interior da plataforma, banco, talude e plataforma profunda.

Figura 3.3 - Mapa estrutural em tempo do topo da seq. Laranja na área de estudo

Fonte: Bronizeski (2018).

3.3 Processamento e obtenção dos atributos sísmicos

Os atributos complexos do dado sísmico foram obtidos a partir da transformada de Hilbert. Cada um dos atributos, a saber: envelope da amplitude, fase instantânea e frequência instantânea, são os definidos em Taner, Koehler e Sheriff (1979). Esta tarefa foi realizada pela suíte de processamento sísmico Geovation.

Figura 3.4 - Mapa das condições de deposição previstas para a área de estudo

Fonte: Bronizeski (2018).

Para todos os atributos foram calculados cubos e, para a análise em mapa, foram retiradas as amostras sobre o horizonte interpretado. Dessa forma, foram obtidos os mapas que servirão de entrada para as classificações posteriores. Na figura 3.5 estão apresentados os mapas de amplitude sísmica, envelope da amplitude, fase e frequência instantânea.

O cubo de impedância acústica foi obtido pela metodologia “*Constrained Sparse Spike*”, como exemplificado em Pendrel (2001), e este processo foi executado utilizando o programa Jason. Nesta tarefa foram utilizados os poços 1-SCS-0006, 1-SPS-0029, 3-TB-0001 e 3-TB-0002. No intervalo de estudo (entre aproximadamente 600 a 900m de profundidade) os perfis disponíveis são o sônico acústico, raios gama

e resistividade. Utilizando a relação de Gardner foi obtido o perfil de densidade, sendo $\alpha=0,31$ e $\beta=0,25$, para que fosse possível calcular a impedância acústica no poço. Na metodologia em questão os poços são utilizados como controle de qualidade da inversão feita. Para este trabalho a correlação entre a impedância do poço e inversão acústica variou entre 85% e 94%.

O mapa de isópaca, que representa a variação de espessura de uma camada, foi obtido após a interpretação do limite da sequência Lilás (definida em Severino, 2000), que fica abaixo da sequência Laranja. O valor do mapa de isópaca é, então, a diferença em tempo entre os dois horizontes interpretados. O mapa de tempo de reflexão apresenta a variação do tempo de reflexão do evento sísmico interpretado. A partir deste mapa, foi criado o mapa da primeira derivada horizontal na direção das “*in-lines*” que estão paralelas ao eixo norte-sul. A figura 3.6 apresenta os mapas de impedância acústica, primeira derivada, isópaca e tempo de reflexão ou mapa estrutural (em tempo).

3.4 Análises dos SOM e PCA

As análises dos SOM foram executadas pelo programa SiroSOM®. Foi gerada uma malha de 72x72 nós, projetada em superfície toroidal e células hexagonais. O treinamento da rede SOM foi repetido em todas as análises com os parâmetros apresentados na tabela 3.2.

Tabela 3.2 - Parâmetros para o treinamento do SOM

Etapa de treinamento	Quantidade de neurônios na direção x	Quantidade de neurônios na direção y	Tipo de cela	Raio de vizinhança inicial	Raio de vizinhança final	Épocas de treinamento
Grosseiro	72	72	Hexagonal	102	26	100
Fino	72	72	Hexagonal	26	1	200

Fonte: Bronizeski (2018).

Figura 3.5 - Mapas de Amplitude (a), envelope da amplitude (b), fase instantânea (c) e frequência instantânea (d)

Fonte: Bronizeski (2018).

Figura 3.6 - Mapas de impedância (a), primeira derivada (b), isópaca (c) e tempo duplo de reflexão (d) – estrutural em tempo

Fonte: Bronizeski (2018).

Para as PCA foi utilizada a biblioteca SciKit Learn disponível para Python a qual é disponibilizada no sub-pacote “*decomposition*”. A PCA utilizada executa a resolução da matriz inversa pela decomposição por valores singulares (SVD) e os mapas dos PC's estão apresentados nas figuras 3.7 e 3.8. O único parâmetro necessário durante a utilização da função PCA foi o número de componentes, que variou entre 4 e 8.

A classificação de fácies foi executada pelo algoritmo k-means, implementado no programa SiroSOM[®]. A seleção do número de classes foi realizada pelo índice Davis e Bouldin (DBI) (DAVIES; BOULDIN, 1979), também implementado no SiroSOM[®], e o resultado de 7 classes foi indicado como ótimo para este problema analisado.

Figura 3.7 - Mapas das PC's 1 a 4, respectivamente os itens a, b, c, e d, obtidas apartir dos atributos originais

Fonte: Bronizeski(2018).

Figura 3.8 - Mapas das PC's 5 à 8, respectivamente os itens a, b, c e d

Fonte: Bronizeski (2018)

4 ESCOLHA DO MELHOR TREINAMENTO DO SOM

O treinamento desenvolvido pelos SOM, como mostrado na figura 3.1, ocorrerá utilizando mapas de atributos sísmicos e os principais componentes em análises diferentes. Os atributos utilizados neste trabalho possuem características diferentes quando analisada a distribuição de valores nos histogramas. Os valores das amostras sobre o horizonte sísmico interpretado identificam características de variação do coeficiente de reflexão entre as camadas, além de evidenciar diferenças internas como as propriedades acústicas das rochas e continuidade do refletor, por exemplo. Os atributos estruturais identificam relações alheias ao dado sísmico, relacionadas às estruturas formadas durante a deposição da rocha. A análise dos atributos fase instantânea (fig. 4.1), frequência instantânea e impedância acústica em histogramas indica que se assemelham a distribuição normal. Os atributos tempo duplo de reflexão, isópaca, amplitude (fig. 4.2) e envelope possuem distribuições bimodais. Os histogramas de cada um dos atributos com distribuição unimodal apresentam valores de curtose que indicam distribuições leptocúrticas, equivalente a dizer que os valores da cauda da distribuição são poucos e distribuídos entre os valores máximo e mínimo mas não ocorrem com grande frequência.

Figura 4.1 - Histograma do atributo fase instantânea

Fonte: Bronizeski (2018)

Figura 4.2 - Histograma do atributo amplitude

Fonte: Bronizeski (2018).

O treinamento da rede utilizou os atributos indicados pela tabela 4.2, em que para cada análise existe o par correspondente dos valores de erro de quantização e erro topográfico. Os resultados estão ordenados pelo erro de quantização crescente. É apresentado na tabela 4.1, a correlação entre cada um dos atributos utilizados durante as 7 primeiras análises.

A figura 4.3 apresenta o gráfico do erro de quantização e o número de atributos das análises. Ao comparar as análises de 1 a 7 é observada relação direta entre o aumento do erro de quantização e o aumento do número de atributos. As análises 1 e 2 permitem as comparações entre utilizar apenas atributos do traço sísmico e atributos estruturais, respectivamente. O erro de quantização associado é menor para a análise 1 (utilizando atributos estruturais) do que para a análise 2 (utilizando atributos derivados do traço sísmico).

Para as análises 1 e 2, os resultados obtidos não são abrangentes pois levam em consideração apenas as características estruturais ou os atributos do traço sísmico. Embora estas análises tenham pequenos erros de quantização, a interpretação dos resultados é limitado. Análises que utilizam apenas um conjunto de características torna a interpretação final com significado limitado. O resultado da classificação não será um mapa que reflita feições estruturais e litológicas de forma conjunta.

Tabela 4.1 - Corelação entre os atributos utilizados nas análises

	Amplitude	Envelope da amplitude	Fase instantânea	Frequência instantânea	Impedância	Primeira derivada	Tempo duplo de reflexão	Isópaca
Amplitude	1,000	0,988	0,034	0,149	0,585	0,112	0,175	0,023
Env. da amplitude		1,000	0,035	0,118	0,584	0,080	0,173	0,026
Fase inst.			1,000	0,118	0,197	0,006	0,004	0,016
Freq. inst				1,000	0,109	0,037	0,005	0,018
Imp. acústica					1,000	0,079	0,208	0,001
Prim. derivada						1,000	0,201	0,056
Tempo duplo de reflexão							1,000	0,482
Isópaca								1,000

Tabela 4.2 - Análises realizadas e respectivos atributos utilizados

Análises	Amplitude	Envelope da amplitude	Fase instantânea	Frequência instantânea	Impedância	Primeira derivada	Tempo duplo de reflexão	Isópaca	Número de atributos	Erro de quantização	Erro topográfico
1						X	X	X	3	0,0412	0,0971
2	X		X	X	X				4	0,195	0,0733
3	X		X	X	X		X		5	0,247	0,0711
4	X		X	X	X			X	5	0,268	0,0766
5	X		X	X	X		X	X	6	0,297	0,0754
6	X		X	X	X	X	X	X	7	0,353	0,0720
7	X	X	X	X	X	X	X	X	8	0,380	0,0719

Fonte: Bronizeski (2018).

Figura 4.3 - Gráfico dos erros de quantização e topográfico por análise

Fonte: Bronizeski (2018).

O resultado da análise 7 é a de maior erro de quantização e portanto seria o melhor resultado (fig. 4.4). Foi constatado que existe uma correlação entre os resultados classificados no domínio das coordenadas espaciais e os mapas dos atributos amplitude (fig. 3.5 a) e impedância acústica (fig. 3.6 a). A classificação refletiu algumas das características presentes também no mapa de isópacas, não permitiu observar feições propícias para identificar os ambientes de deposição.

O maior valor do erro de quantização das RNA's após o treinamento pode indicar, para os dados geológicos, bordas ou contatos litológicos (FRASER; DICKSON, 2007). o treinamento das RNA's é beneficiado pela utilização de componentes com a máxima variância possível. A obtenção de um SOM melhor treinado implicará que a cada atributo inserido na análise com a máxima variância possível só levará a um melhor treinamento se o erro de quantização aumentar comparado com o treinamento anterior, neste caso ocorrerá uma melhor classificação e identificação de contatos litológicos. No apêndice A são demonstrados os resultados de diversas análises executadas neste trabalho.

A transformação realizada pela PCA obtém uma nova base ortogonal cujas direções possuem a máxima variância possível. A combinação linear que forma este componentes é apresentado na tabela 4.3, e embora não sejam interpretáveis geologicamente, favorecem a identificação de anomalias, isto é, dados afastados do

Figura 4.4 - Mapa de resultado da classificação executada na análise 7

Fonte: Bronizeski (2018).

comportamento padrão ou valor médio. Neste trabalho a PCA tornou os eventos rotineiros próximos à média, e evidenciou os eventos anômalos distantes da média, obtendo, produzindo, em sua maioria, componentes com uma distribuição unimodal. As distribuições com muitos pontos anômalos tendem a gerar alto valor de desvio padrão e curtose positiva. Também é correto associar esta distribuição mais ampla a valores maiores de entropia, embora não quantificada.

Estas condições de distribuição leptocúrtica para os componentes da PCA se resumem em um treinamento da rede dos SOM com maior erro de quantização, pois embora a variância seja a mesma, as combinações lineares para formar os novos vetores da base, alteram a distribuição do conjunto de dados e faz com que a quantidade de pontos na cauda destas distribuições sejam pequenos porém espalhados, com a máxima variância possível. Assim, ao invés da escolha de atributos com distribuição leptocúrtica pode-se aplicar a PCA e utilizar seus componentes para treinamento da RNA. Este procedimento foi aplicado em 4 análises que diferem pelo número de componentes utilizados e estão resumidas na tabela 4, ordenadas pelo valor do erro de quantização. As distribuições dos principais componentes 1, 2, 3, 4, 7 e 8 são leptocúrticas e os valores de curtose são apresentados na tabela 4.5. Para os componentes 5 e 6 a distribuição é bimodal ou multimodal, portanto, o valor de curtose não tem significado.

Tabela 4.3 - Combinação linear dos principais componentes a partir dos atributos de entrada

Componente	Amplitude	Envelope da amplitude	Fase instantânea	Frequência instantânea	Impedância	Primeira derivada	Tempo duplo de reflexão	Isópaca
1	-2.6 e-03	-2.6 e-03	1.2 e-05	-1.4 e-02	9.9 e-01	6.8 e-07	1.8 e-04	-1.2 e-05
2	-9.2 e-03	-6.4 e-03	1.2 e-04	-9.9 e-01	-1.4 e-02	5.6 e-06	-5.7 e-04	3.3 e-04
3	6.9 e-01	7.2 e-01	1.5 e-03	-1.1 e-02	3.5 e-03	-1.0 e-04	-1.6 e-02	-7.5 e-03
4	-7.2 e-01	6.9 e-01	-8.3 e-03	2.1 e-03	-1.4 e-05	1.6 e-03	7.6 e-02	-4.2 e-02
5	-7.1 e-02	5.0 e-02	6.5 e-03	9.6 e-04	1.4 e-04	-2.1 e-03	-9.5 e-01	3.1 e-01
6	-3.9 e-03	2.0 e-02	5.8 e-02	4.4 e-05	-2.6 e-06	1.1 e-03	3.1 e-01	9.5 e-01
7	6.3 e-03	-3.1 e-03	-9.9 e-01	-1.4 e-04	1.6 e-05	-1.4 e-04	1.1 e-02	5.7 e-02
8	1.1 e-03	-9.8 e-04	-1.3 e-03	2.8 e-06	9.8 e-08	9.9 e-01	-2.4 e-03	-2.4 e-04

Fonte: Bronizeski (2018).

Cabe aqui uma ressalva quanto à maneira de utilização da PCA. Muitos autores utilizam a técnica PCA como ferramenta para redução de dimensionalidade. Com este intuito são escolhidos os componentes em ordem decrescente de variância até que a soma acumulada da variância seja próxima de 95% da variância total. A tabela 4.6 mostra a percentagem da variância explicada por cada um dos componentes utilizados neste trabalho. Neste sentido, observa-se que o primeiro dos componentes é responsável por 99,7% da variância total. O mapa do componente 1 é muito semelhante ao mapa de impedância acústica, amplitude e consequentemente aos resultados da análise 7.

A PCA pode também ser uma ferramenta para remoção de ruídos, e neste caso os componentes de maior ordem são usualmente, associados aos ruídos. No entanto, neste estudo foi decidido incluí-los nas análises desenvolvidas, uma vez que foram identificadas características dos atributos originais (fig. 3.5 e 3.6) nos componentes 5, 6, 7 e 8 (fig. 3.8), além de entender-se que informações relevantes estejam contidas entre o que tradicionalmente seria considerado como ruído. As diferenças entre as fácies carbonáticas aqui estudadas são sutis, e dessa maneira, podem ter pequena variância e estar próxima aos níveis do ruído. Neste contexto, descartar uma das componentes significaria remover informações sutis de diferenças entre fácies.

Tabela 4.4 - Análises realizadas e componentes utilizados

Análises	Principal Componente 1	Principal Componente 2	Principal Componente 3	Principal Componente 4	Principal Componente 5	Principal Componente 6	Principal Componente 7	Principal Componente 8	Número de atributos	Erro de quantização	Erro topográfico
8	X	X	X	X					4	0,21	0,0602
9	X	X	X	X	X	X			6	0,472	0,0782
10	X	X	X	X	X	X	X		7	0,546	0,0799
11	X	X	X	X	X	X	X	X	8	0,553	0,0744

Fonte Bronizeski (2018).

As análises que utilizam os principais componentes apresentam erro de quantização superior àquelas que utilizaram os atributos diretamente. O erro topográfico não parece sofrer alterações devido às transformações executadas pela PCA. A análise escolhida ao final será a de maior erro de quantização, que foi a de número 11 e utiliza os 8 principais componentes.

Tabela 4.5 - Curtose para os principais componentes com distribuição unimodal

	Componente 1	Componente 2	Componente 3	Componente 4	Componente 7	Componente 8
Curtose	2,0666	48,6888	1,3811	120,6090	37,6808	31,8658

Fonte: Bronizeski (2018).

A escolha da análise 11 é adequada ao propósito de classificar um ambiente de sedimentação carbonático. As diferenças entre os tipos de rocha carbonáticas se manifestam com pequena expressividade, quando observadas por atributos sísmicos podem até não ser detectadas. Dessa forma os métodos aqui empregados e analisados se mostram adequados para o propósito de identificar diferentes tipos de rochas carbonáticas. O principal quesito identificado é a alteração realizada nas distribuições dos atributos proporcionado pela PCA, possibilitando identificar os contrastes ao utilizar os SOM.

Tabela 4.6 - Componente e percentagem da variância explicada

Componente	Variância
PC1	99,7523%
PC2	0,24644%
PC3	$1,21918 \times 10^{-3} \%$
PC4	$1,64988 \times 10^{-5} \%$
PC5	$1,05567 \times 10^{-5} \%$
PC6	$5,14457 \times 10^{-7} \%$
PC7	$6,24883 \times 10^{-8} \%$
PC8	$4,67655 \times 10^{-10} \%$

Fonte: Bronizeski (2018).

5 ANÁLISES POR SELF-ORGANIZING MAPS UTILIZANDO PCA DOS ATRIBUTOS SÍSMICOS

A análise do treinamento da RNA SOM e posterior classificação executada será avaliada com critérios específicos. As oito componentes principais oriundas da PCA foram desenvolvidas com o intuito de caracterizar o ambiente de deposição carbonático e identificar as divisões criadas pelos atributos analisados.

5.1 Análise exploratória dos principais componentes

As 8 componentes principais (PC's) selecionadas serão aqui analisadas quanto à correlação entre si aplicando o teste R^2 . Os resultados estão apresentados na tabela 5.1.

Tabela 5.1 - Matriz de correlação entre os componentes para a análise 11

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
PC1	1	0,094	0,009	0,001	-0,041	0,001	0,042	0,055
PC2		1	0,041	-0,057	-0,006	0,020	-0,014	-0,013
PC3			1	-0,112	0,006	0,047	-0,065	-0,009
PC4				1	-0,042	0,002	-0,045	-0,014
PC5					1	0,035	-0,011	0,055
PC6						1	-0,025	-0,039
PC7							1	0,001
PC8								1

Fonte: Bronizeski (2018).

A análise da matriz permitiu observar que a correlação entre os componentes é baixa, com valores em módulo, entre 0,001 e 0,112. Tais valores indicam que todas as componentes possuem baixa correlação entre si. Este fato decorre da aplicação da transformação por PC's que, aplicada aos oito atributos sísmicos, diminui a correlação e, como explicitado no capítulo 4, aumenta o erro de treinamento e os contrastes entre grupos de rochas na área de análise.

Os procedimentos da PCA identificam relações lineares entre os atributos de entrada, criando componentes ortogonais entre si. A composição linear de cada um dos vetores desta nova base realiza uma combinação entre atributos diferentes e não permitiria uma interpretação de características individuais de cada um dos componentes, assim como é realizado com os atributos sísmicos.

5.2 Leitura e análise dos resultados

Ao término da fase de treinamento do SOM, utilizando os 8 componentes, o resultado relativo à similaridade dos neurônios projetados no mapa auto-organizado foi observado a partir da matriz-U, apresentada na fig. 5.1. A escala de cores representa as dissimilaridades entre os neurônios, com valores compreendidos entre 0,102 e 8,8.

Figura 5.1 - Matriz-U, evidenciando regiões de alta similaridade de neurônios separadas por alta dissimilaridades, relativas à análise 11.

São nítidos alguns contrastes de dissimilaridade, mas o valor extremo da escala de cores dificulta observar com maior detalhe todas as variações na fig. 5.1. Ainda assim, vale ressaltar a região central da matriz-U (fig. 5.1) onde existe uma população expressiva de BMU's com baixas dissimilaridades, sugerindo que entre estes pontos não existem diferenças expressivas. Os valores de alta dissimilaridade foram agrupados isoladamente, quase em sua totalidade.

As interpretações utilizando os PC's no domínio dos SOM através dos “*component plots*” apresentados na figura 5.2 compara as variações dos valores dos atributos distribuídos nos mapas indicando quais características são as mais influentes em cada uma das regiões da matriz-U e, por estas correlações, poderia facilitar a identificação de rochas e agrupamentos. Nesta análise não será possível este tipo de correlação, pois ao utilizar os PC's não há significação geológica em cada componente, mas sim uma combinação linear das propriedades das rochas. Desta forma, não será abordado neste estudo os pesos de cada atributo em cada componente, ou suas respectivas relações.

Figura 5.2 - As principais componentes no domínio do SOM na análise 2

Fonte: Bronizeski (2018).

A matriz-U foi classificada em 7 grupos pelo algoritmo k-means. O resultado desta classificação é apresentado na figura 5.3, e a superposição desta divisão em classes sobre os principais componentes é apresentado na figura 5.4. Tem-se que 6 das 7 classes no domínio dos SOM formaram grupos contínuos, ressaltando que as amostras com características semelhantes devem ser representadas por BMU's próximos. A matriz-U é apresentada em um mapa 2-D decorrente da planificação de

uma hipersuperfície toroidal, onde os neurônios treinados no espaço n-dimensional foram projetados. Esse formato de projeção favorece a preservação de relações topológicas.

Figura 5.3 - Divisão da matriz-U em 7 classes

Fonte: Bronizeski(2018).

É indicado pela figura 5.4 a correlação entre a classe 1 e a aglomeração dos valores extremos negativos do PC 4, assim como a classe 2 e os maiores valores positivos do PC 2. As classes 3, 4 e 5 sugerem uma combinação entre vários PC's. Não há evidentes fronteiras de dissimilaridade entre as classes 4 e 5 na matriz-U (fig. 5.1).

Figura 5.4 - Sobreposição da divisão de classes aos mapas das principais componentes

Fonte: Bronizeski (2018).

A classe 6 é correlacionada com os maiores valores do PC 5. Já a classe 7 é correlacionada aos maiores valores do PC 8, e formou o único grupo de amostras com agrupamentos isolados.

Ainda que 4 classes tenham sido associadas a apenas um componente, estas recebem a influência de diversos atributos devido às transformações aplicadas pela PCA, e sabemos que cada PC é criada a partir da combinação linear dos atributos de entrada. Os resultados descritos e também os que destes derivam, são portanto o resultado de uma classificação multi-atributos.

6 INTERPRETAÇÃO DOS AMBIENTES DE DEPOSIÇÃO CARBONÁTICA

As diversas análises sobre o treinamento da RNA e classificação dos resultados indica que, utilizando os atributos apresentados neste trabalho, e com o pré-condicionamento pela PCA ocorre a melhor distribuição das classes dentro da área de estudo. A interpretação do ambiente de deposição é então realizada e comparada com o modelo inicial interpretado (figura 3.5) e também com o modelo das zonas de deposição de uma plataforma carbonática com borda (figura 2.6).

6.1 Distribuição de classes no domínio espacial original

As características de construção dos PC's não possibilitam correlações entre os “*component plots*”, zonas de deposição ou tipo de fácies. Assim, é necessário interpretar os resultados analíticos obtidos pelos SOM (Fig. 5.2 e 5.3), referente à análise 11 apresentada no capítulo 5, no domínio espacial original das amostras (fig. 6.1).

O modelo interpretado inicialmente a partir do mapa estrutural (fig. 3.5) é de uma plataforma carbonática com borda. Diferentemente deste modelo de deposição, a distribuição das classes mostrada na figura 6.1 é próxima a uma plataforma carbonática sem borda, uma vez que as zonas da lagoa e banco de areia não foram destacadas na classificação efetuada. Provavelmente a lagoa e o banco não apresentavam características suficientes para serem discretizados pelos atributos presentes neste trabalho, ou ainda, a influência destas zonas foi pouco pronunciada durante a época de deposição.

O ambiente não se assemelha a uma rampa carbonática, pois segundo Flügel (2004) em uma rampa devem existir depósitos de alta energia próximos à linha de costa, além da ausência de recifes marginais. Também é característico de rampa carbonática uma declividade suave, sem a presença de talude pronunciado, condições que não ocorrem no caso estudado neste trabalho.

Figura 6.1 - Mapa classificado das amostras no domínio espacial original correspondente à análise 11.

Fonte: Bronizeski (2018).

Ainda em Flügel (2004) é mencionado que o modelo simplificado de Wilson, proposto para uma plataforma carbonática com borda, é bastante completo e, desse modo, é comum encontrar as plataformas com um número reduzido de fácies. Portanto, esta distribuição de classes da fig. 6.1 pode ser interpretada como uma plataforma carbonática com borda (fig. 2.6) com três zonas principais: planície (classes 4 e 5), plataforma profunda ou sopé do talude (classe 6) e talude (classe 7).

O grupo com maior quantidade de amostras é a classe 4, seguido pela classe 5 que unidos representam a totalidade da planície no modelo proposto (fig. 2.6). A

separação entre as classes 4 e 5 não parece ser expressiva, fato este evidenciado pela pequena dissimilaridade na matriz-U.

A rocha referente à superfície laranja no poço 1-SPS-0029 é descrita em Severino (2000) como packstone híbrido, depositado sobre wackestone/packstone bioclástico. Considerando o mapa estrutural e o modelo da figura 2.6, as classes 4 e 5 podem ser inseridas na região de deposição plataforma interior, entre as zonas deposicionais 8 e 7, pois não foi mencionado intemperismo e nem mesmo a presença da deposição de sais, o que exclui a possibilidade da existência das zonas deposicionais 9 (zona evaporítica ou salmoura) e 10 (zona de rochas carbonáticas intemperizadas) do modelo citado.

Como não foi obtida uma classe individual para o banco, existe a incerteza para associar uma zona da figura 2.6 às classes 4 e 5. As características de baixa dissimilaridade entre as classes 4 e 5, sugerem a ocorrência de uma gradação contínua nas propriedades das rochas entre as zonas do interior da plataforma: restrito, marinho aberto e margem da plataforma. Dessa forma, a classe 5 pode ser associada ao packstone e a classe 4 pode ser associada ao grainstone, ambos os tipos de rocha (respectivamente, packstone e grainstone), são suportados pelo grão e diferem em conter ou não matriz.

Exclui-se nesta análise os efeitos diagenéticos sobre as classes 4 e 5, pois é descrito em Severino (2000) uma pequena exposição da área da plataforma durante o rebaixamento do nível do mar e, provavelmente, sem a passagem de sedimentos por esta área. A existência destas condições indicam que as influências pós deposicionais não existiram ou foram pouco pronunciadas, limitando as diferenças existentes entre as classes 4 e 5 à época da deposição.

A classe 2 ocorre de forma discordante, próximo ao término da zona exterior da plataforma, limitada pela curva de 950ms. A distribuição das amostras desta classe possui orientação NW-SE e coincide com as correntes oceânicas durante o Mioceno inferior (~17Ma), as quais tinham direção SE-NW. Pode-se imaginar que ocorria circulação de água influenciada pelas correntes oceânicas atrás da borda da plataforma com sentido SW, causada pela entrada de água pela região em que ocorre a classe 2. Assim, este cenário inclui a classe 2 como indicativa de retrabalhamento de sedimentos, ou ainda de um canal.

A classe 2 ocorre também em algumas amostras sem estrutura definida e distantes do limite da planície exterior. O refletor interpretado, em quase sua totalidade é contínuo, mas possui algumas marcas de progradação, que induziu a alguns erros de rastreamento, como é visto no círculo em vermelho na figura 6.2. Estes pontos, embora incluídos na classe 2, não representam uma zona de deposição, tendo sido associados a erros de interpretação. A localização das seções apresentadas nas figuras 6.2 (seção A-A') e 6.4 (B-B') é apresentada na figura 6.3.

Figura 6.2 - Horizonte interpretado e influências de eventos assinalados em vermelho que levaram a erros na interpretação analisado na seção A-A' da fig. 6.3

Fonte: Bronizeski (2018).

A classe 1 ocorre próximo à classe 2 e em alguns pontos esparsos, todos estes na borda da plataforma. Não foi possível correlacionar uma das zonas da figura 2.6 (modelo de Wilson simplificado) à classe 2. Na matriz-U (fig. 5.1) esta classe reúne os valores extremos de dissimilaridade, indicando que não existe unidade na representação da classe.

Assim que se tornam conhecidas as distribuições das classes voltamos também aos mapas dos PC's 1 a 8, em busca de informações sobre as classes 1 e 2. Nos PC's 1 a 6 existem as feições das classes 1 e 2, bem como os diferentes valores que criam o contraste entre estas duas classes.

A classe 1 é formada por um grupo de rochas com características que não se destacam nos atributos sísmicos ou estruturais. As características se sobressaem pelo uso da PCA, mas impossibilita a identificação de litotipo correspondente.

Figura 6.3 - Mapa de localização das seções A-A' e B-B' das figuras 6.2 e 6.4 respectivamente

Fonte: Bronizeski (2018)

Apenas considerando a disposição das amostras da classe 1 no mapa da figura 6.1 próximas às amostras da classe 2 sugerem se tratar de sedimentos retrabalhados. Ocorre que a presença de amostras desta mesma classe 1 em outras áreas (como o talude e sopé do talude, fig. 6.4) reforçam a dificuldade de interpretação. Talvez esta classe represente uma região que foi erodida e que não puderam ser descritas pelos atributos utilizados neste trabalho.

A classe 7 delimita a margem da plataforma pelo seu limite superior. Indica o término da plataforma e início do talude, pois ocorre a mudança de declividade (fig. 6.4), sendo que a fácies mais provável é framestone. Esta zona de deposição recebe influência direta das ondas. As rochas neste ambiente são formadas no local de deposição sob a influência da alta energia, e necessita de coesão durante a deposição para permanência no local. Pode-se interpretar a grande contribuição da

declividade, ressaltada pelo atributo primeira derivada (fig. 3.8), mas outros atributos foram considerados para a identificação das amostras desta classe, já que não foram todos os pontos anômalos da primeira derivada inclusos na classe 7.

Figura 6.4 - Seção sísmica arbitrária e enfoque na distribuição da classe 7

Fonte: Bronizeski (2018).

A zona de baixa energia no sopé do talude é identificada na classe 6. As facies previstas são margas, packstone alóctone e grainstones. Não ocorrem divisões internas nesta região, o que leva à identificação de um grupo de rochas homogêneo.

A classe 3 ocorre sempre entre as classes 2 e 4. Além disso, a matriz-U indica que entre as classes 3 e 4 existem barreiras de dissimilaridade dos BMU's, assim como uma barreira de dissimilaridade maior entre as classes 2 e 3. Assim, a classe 3 pode indicar rochas carbonáticas erodidas pela ação da circulação de água, que não são rochas retrabalhadas (classe 2) nem grainstones (classe 4).

As informações de poços podem alterar as interpretações sobre os tipos de rocha aqui interpretados, mas as informações disponíveis dos poços são escassas. Não existem testemunhos de rochas nos poços na área de estudo. Desse modo, as

amostras de rocha disponíveis são amostras de calha e, dentre todas, são conhecidas apenas as lâminas petrográficas referentes ao poço 1-SPS-0029.

Unindo todas as associações entre as classes e rochas pode-se construir o mapa de distribuições de classes na plataforma carbonática com borda (fig. 6.5), o qual está colocado ao lado do mapa das zonas de deposição interpretadas.

Figura 6.5 - Mapas de distribuição de rochas carbonáticas e mapa das zonas de deposição

Fonte: Bronizeski (2018).

7 CONCLUSÕES

7.1 Utilização da PCA como pré-condicionamento aos SOM

A utilização da PCA como etapa de pré-condicionamento permitiu ampliar a discriminação do SOM;

A PCA transformou as distribuições dos atributos de entrada, foram obtidas distribuições leptocúrticas em sua maioria;

A utilização da PCA como condicionamento para o SOM resultou em melhor classificação pois permitiu a correlação entre os ambientes do modelo deposicionais de Wilson (simplificado) e a distribuição de 5 das 7 classes encontradas;

O erro de quantização da RNA SOM aumentou com o aumento do número de atributos utilizados para a classificação, ao utilizar os PC's o mesmo padrão se repetiu;

Em todas as análises, as classes 4 e 5 não possuem uma fronteira de dissimilaridade expressiva entre si. Conclui-se que as rochas destas classes não possuem grandes diferenças quando observadas pelos dados sísmicos. O mesmo foi observado entre as classes 3 e 4;

A transformação realizada pela PCA não altera características intrínsecas aos dados mas ressalta os menores contrastes. Um exemplo foi a classe 2, mais expressiva nos PC's do que nos atributos originais;

A classificação do mapa do SOM pelo algoritmo k-means identificou a classe 7 com 3 agrupamentos distantes na matriz-U, o que não é esperado. As demais classes identificadas formam agrupamentos contínuos no domínio do SOM ;

Concluimos também deste trabalho que os contrastes ente algumas das fácies foi melhor evidenciado pela utilização das RNA's, pois permitiu a integração de atributos tanto relacionados ao traço sísmico quanto ao ambiente de deposição;

7.2 Caracterização do ambiente carbonático

A identificação de classes em carbonatos recai sobre características tênues. A existência dos poços nem sempre é capaz de amostrar todos os tipos de fácies, portanto, utilizar um modelo para o ambiente de deposição mostrou-se necessário;

Com a utilização de todas as componentes da PCA foi possível incorporar feições relacionadas às variações sutis entre diferentes texturas de rochas. Caso tivessem sido desprezadas as componentes de maior grau (PC's 5 à 8), algumas das feições provavelmente teriam sido removidas;

Embora o mapa estrutural em tempo indique um ambiente de deposição carbonático em plataforma com borda os resultados da classificação desenvolvida neste trabalho não identificaram diferenciações entre a lagoa e banco carbonático. Tais resultados se assemelham a uma plataforma distalmente escarpada, conforme já indicado por Serverino (2000);

As pequenas diferenças na planície de maré, classes 4 e 5, podem indicar uma gradação entre os tipos de rocha, onde a interpretação pode ser desenvolvida com o auxílio de descrição de rochas e o modelo de deposição de plataforma carbonática com borda;

A classificação de fácies realizada neste trabalho é iniciada pela plataforma, identificada pelas classes 4 e 5, indica uma possível diferenciação gradacional, amostrada pelos atributos utilizados, entre plataforma, lagoa e banco. A plataforma começa com packstone e termina na sua borda com grainstones de acordo com o modelo utilizado. O talude foi corretamente identificado e deve ser formado por rochas do tipo framestone. O sopé do talude também foi identificado e delimitado, segundo o previsto pelo modelo de Wilson (1975, apud Flügel, 2001), e este grupo formado pela classe 6 é interpretado como margas e packstones alóctones;

A interpretação das zonas de deposição pelo mapa estrutural é um ponto de partida para a interpretação que ajuda a escolher um modelo de deposição. A interpretação da textura das rochas deve levar em consideração os dados sísmicos, atributos sísmicos e informações de descrição de rochas. O método que possibilita esta integração de informações é o SOM tal qual foi aqui apresentado;

Aspectos futuros incluem a busca e descrição de amostras de poços na região de estudo, incluindo a descrição petrográfica, que possibilitará rever a associação de rochas com as classes realizada neste trabalho . Além disso, há a sugestão de investigação de outros métodos estatísticos, além da curtose, para identificar o impacto de cada atributo em relação à classificação realizada pelo SOM. Essa abordagem talvez possa ser feita a partir do cálculo de entropia do sistema das variáveis.

8 BIBLIOGRAFIA

- ALMEIDA, F. F. M. DE. The system of continental rifts bordering the Santos basin, Brazil. **Anais da Academia Brasileira de Ciências**, v. 48(Suplem), p. 15–26, 1976.
- AMORIM, N. R. D. E. Caracterização geofísica da dorsal de são paulo e seu significado evolutivo na bacia de santos. 2013.
- ANP. **Webmaps**. Disponível em:
<webmaps.anp.gov.br/mapas/Lists/DSAppPages/MapasBrasil.aspx>. Acesso em: 29 maio. 2018.
- BARNES, A. E. A tutorial on complex seismic trace analysis. **Geophysics**, v. 72, n. 6, p. 33–43, 2007.
- BHATTACHARYA, S.; CARR, T. R.; PAL, M. Comparison of supervised and unsupervised approaches for mudstone lithofacies classification: Case studies from the Bakken and Mahantango-Marcellus Shale, USA. **Journal of Natural Gas Science and Engineering**, v. 33, p. 1119–1133, 2016.
- BIZZI, L. A. et al. (EDS.). **Geologia, Tectônica e Recursos Minerais Do Brasil**. 1. ed. Brasília: CPRM - Serviço Geológico do Brasil, 2003. v. 1
- BROWN, A. R. Seismic attributes and their classification. **The Leading Edge**, v. 15, n. October, p. 1090–1090, 1996.
- CARNEIRO, C. D. C. et al. Semiautomated geologic mapping using self-organizing maps and airborne geophysics in the Brazilian Amazon. **GEOPHYSICS**, v. 77, n. 4, p. 17–24, 2012.
- CÉRÉGHINO, R.; PARK, Y. S. Review of the Self-Organizing Map (SOM) approach in water resources: Commentary. **Environmental Modelling and Software**, v. 24, n. 8, p. 945–947, 2009.
- CHANG, H. K.; KOWSMANN, R. O. **Subsidência térmica e espessura crustal na Bacia de Santos**. XXXIII Congr. Bras. Geol. **Anais...**Rio de Janeiro: 1984
- CHANG, H. K.; KOWSMANN, R. O.; FIGUEIREDO, A. M. F. Novos conceitos sobre o desenvolvimento das bacias marginais do leste brasileiro. In: **Origem e Evolução das Bacias Sedimentares**. [s.l.: s.n.]. p. 269–289.
- CHOPRA, S.; MARFURT, K. J. Seismic attributes - A historical perspective. **Geophysics**, v. 70, n. 5, p. 3SO–28SO, 2005.
- COLÉOU, T. (CGG); POUPON, M.; AZBEL, K. Unsupervised seismic facies classification : A review. **The Leading Edge**, 2003.

DAVIES, D. L.; BOULDIN, D. W. A cluster separation measure. **IEEE transactions on pattern analysis and machine intelligence**, v. 1, n. 2, p. 224–227, 1979.

DE MATOS, M. C.; OSORIO, P. L.; JOHANN, P. R. Unsupervised seismic facies analysis using wavelet transform and self-organizing maps. **Geophysics**, v. 72, n. 1, p. P9–P21, 2007.

DEMERCIAN, S.; SZATMARI, P.; COBBOLD, P. R. Style and pattern of salt diapirs due to thin-skinned gravitational gliding, Campos and Santos basins, offshore Brazil. **Tectonophysics**, v. 228, n. 3–4, 1993.

DORRINGTON, K. P.; LINK, C. A. Genetic-algorithm/neural-network approach to seismic attribute selection for well-log prediction. **Geophysics**, v. 69, n. 1, p. 212, 2004.

DUMAY, J.; FOURNIER, F. Multivariate statistical analyses applied to seismic facies recognition. **GEOPHYSICS**, v. 53, n. 9, p. 1151–1159, 1988.

DUNHAM, R. J. Classification of carbonate rocks according the depositional structure. In: HAM, W. E. (Ed.). . **Classification of carbonate rocks**. 1. ed. [s.l.] American Association of Petroleum Geologists, 1962. v. 1p. 108–121.

EMBRY, A. F.; KLOVAN, E. A late devonian r e e f tract on northeastern banks island, n.w.t.' ashton. **Bulletin of Canadian Petroleum Geology**, v. 19, n. 4, p. 730–781, 1971.

FLÜGEL, E. Carbonate Depositional Environments. In: **Microfacies of Carbonate Rocks**. 2. ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. p. 7–52.

FRASER, S. J.; DICKSON, B. L. A New Method for Data Integration and Integrated Data Interpretation: Self-Organising Maps. **Proceedings of Exploration 07: Fifth Decennial International Conference on Mineral Exploration**, p. 907–910, 2007.

FRERY, A. C.; PERCIANO, T. Contrast Enhancement and Dimensionality Reduction by Principal Components. In: **Introduction to Image Processing Using R: Learning by Examples**. [s.l.: s.n.]. p. 77–84.

GARCIA, S. F. DE M. **Restauração Estrutural da halotectônica na porção central da bacia de Santos e implicações para os sistemas petrolíferos**. [s.l.] Universidade Federal de Ouro Preto, 2013.

GRAY, R. M. Vector quantization. **IEEE ASSP Magazine**, v. 1, n. 2, p. 4–29, 1984.

GRAY, R. M.; OLSHEN, R. **Vector quantization and density estimation**. Proceedings. Compression and Complexity of SEQUENCES 1997.

Anais...1997Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=666914>>

GUO, H.; MARFURT, K. J.; LIU, J. Principal component spectral analysis. **Geophysics**, v. 74, n. 4, p. 9, 2009.

HAGAN, M. T. et al. **Neural Network Design**. 2nd. ed. [s.l.] Campus Pub. Service, University of Colorado Bookstore, 2002, 2002.

HAMPSON, D. P.; SCHUELKE, J. S.; QUIREIN, J. A. Use of multiattribute transforms to predict log properties from seismic data. **Geophysics**, v. 66, n. 1, p. 220–236, 2001.

HASHEMI, S. et al. Channel characterization using multiple-point geostatistics, neural network, and modern analogy: A case study from a carbonate reservoir, southwest Iran. **Journal of Applied Geophysics**, v. 111, p. 47–58, 2014.

HOTELLING, H. Analysis of a complex of statistical variables into principal components. **Journal of Educational Psychology**, v. 24, n. 7, p. 498–520, 1933.

ITURRARÁN-VIVEROS, U.; PARRA, J. O. Artificial Neural Networks applied to estimate permeability, porosity and intrinsic attenuation using seismic attributes and well-log data. **Journal of Applied Geophysics**, v. 107, p. 45–54, 2014.

JOLLIFFE, I. T. **Principal Component Analysis**. [s.l.: s.n.]. v. 103

KASKI, S.; KOHONEN, T. Exploratory Data Analysis By The Self-Organizing Map: Structures Of Welfare And Poverty In The World (1996). **Neural Networks in Financial Engineering. Proceedings of the Third International Conference on Neural Networks in the Capital Markets**, p. 498–507, 1996.

KOHONEN, T. The self-organizing map. **Proceedings of the IEEE**, v. 78, n. 9, p. 1464–1480, 1990.

KOURKI, M.; ALI RIAHI, M. Seismic facies analysis from pre-stack data using self-organizing maps. **Journal of Geophysics and Engineering**, v. 11, n. 6, p. 65005, 2014.

KURODA, M. C. **Redes Neurais Aplicadas ao Estudo de Rochas Reservatório**. [s.l.] Universidade Estadual de Campinas, 2012.

KURODA, M. C. **Técnicas de aprendizagem de máquina Bio-Inspiradas aplicadas ao estudo de rochas reservatório**. [s.l.] Universidade Estadual de Campinas, 2016.

LAASRI, E. H. A. et al. A fuzzy expert system for automatic seismic signal classification. **Expert Systems with Applications**, v. 42, n. 3, p. 1013–1027, 2015.

LAI, J. et al. Prediction of Diagenetic Facies using Well Logs: Evidences from Upper Triassic Yanchang Formation Chang 8 Sandstones in Jiyuan Region, Ordos Basin, China. **Oil & Gas Science and Technology – Revue d'IFP Energies nouvelles**, v. 71, n. 3, p. 34, 2016.

LINARI, V. et al. Seismic facies analysis based on 3D multiattribute volume classification, La Palma Field, Maracaibo, Venezuela. **The Leading Edge**, v. 22, n. 1, p. 32, 2003.

LIU, L. et al. A modified Fuzzy C-Means (FCM) Clustering algorithm and its application on carbonate fluid identification. **Journal of Applied Geophysics**, v. 129, p. 28–35, 2016.

LÖTSCH, J.; ULTSCH, A. Exploring the structures of the U-matrix. In: VILLMANN, T. et al. (Eds.). . **Advances in Self-Organizing Maps and Learning Vector Quantization**. Mittweida: Springer, 2004. v. 1p. 249–257.

MA, Y. Z.; GOMEZ, E.; LUNEAU, B. Integration of seismic and well-log data using statistical and neural network methods. **Leading Edge**, v. 36, n. 4, p. 324–329, 2017.

MACKAY, D. J. C. An Example Inference Task: Clustering. In: **Information Theory, Inference and Learning Algorithms**. 1. ed. [s.l.] Cambridge University Press, 2003. p. 284–292.

MACQUEEN, J. B. Kmeans Some Methods for classification and Analysis of Multivariate Observations. **5th Berkeley Symposium on Mathematical Statistics and Probability 1967**, v. 1, n. 233, p. 281–297, 1967.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The Bulletin of Mathematical Biophysics**, v. 5, n. 4, p. 115–133, 1943.

MEISLING, K. E.; COBBOLD, P. R.; MOUNT, V. S. Segmentaion of an obliquely rifted margin , Campos southeastern Brazil. v. 11, n. 11, p. 1903–1924, 2001.

MODICA, C. J.; BRUSH, E. R. Postrift sequence stratigraphy, paleogeography, and fill history of the deep-water Santos Basin, offshore southeast Brazil. **AAPG Bulletin**, v. 88, n. 7, p. 923–945, 2004.

MOQBEL, A. AL; WANG, Y. Carbonate reservoir characterization with lithofacies clustering and porosity prediction. **Journal of Geophysics and Engineering**, v. 8, n. 4, p. 592, 2011.

MOREIRA, J. L. P. et al. Bacia de Santos. **Boletim de Geociências Petrobras**, v. 15, n. 2, p. 531–549, 2007.

PEARSON, K. On lines and planes of closest fit to systems of points in space. **Philosophical Magazine Series 6**, v. 2, n. 11, p. 559–572, 1901.

PEDREGOSA, F. et al. Scikit-learn: Machine Learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011.

PENDREL, J. Seismic Inversion- The Best Tool For Reservoir Characterization. **CSEG Recorder**, v. 26, n. 1, p. 18–24, 2001.

PEREIRA, M. J. et al. Estratigrafia da bacia de Santos: Análise das Sequências, sistemas deposicionais e revisão litoestratigráfica. **Anais do XXXIV Congresso Brasileiro de Geologia**, v. 1, n. 1, p. 65–79, 1986.

PEREIRA, M. J.; FEIJÓ, F. J. Bacia de Santos. **Boletim de Geociências da Petrobrás**, v. 8, n. 1, p. 219–234, 1994.

PEREIRA, M. J.; MACEDO, J. M. A Bacia de Santos: perspectivas de uma nova província petrolífera na plataforma continental sudeste brasileira. **Boletim Geociências da PETROBRAS**, v. 4, n. 1, p. 3–11, 1990.

PICOLINI, J. P. **Estudo das pressões de poros numa área central da bacia de Santos**. [s.l.] Universidade Estadual Paulista, 2017.

POMAR, L. Types of carbonate platforms: A genetic approach. **Basin Research**, v. 13, n. 3, p. 313–334, 2001.

RAFIQ, A. et al. Reservoir characterization using microseismic facies analysis integrated with surface seismic attributes. **Interpretation**, v. 4, n. 2, p. 1–15, 2016.

RODEN, R.; SMITH, T.; SACREY, D. Special section : Pattern recognition and machine learning Geologic pattern recognition from seismic attributes : Principal component analysis and self-organizing maps. **Interpretation**, v. 3, n. 4, p. 59–83, 2015.

SARASWAT, P.; SEN, M. K. Artificial immune-based self-organizing maps for seismic-facies analysis. **Geophysics**, v. 77, n. 4, p. 9, 2012.

SAYAGO, J. et al. Characterization of a deeply buried paleokarst terrain in the Loppa High using core data and multiattribute seismic fades classification. **AAPG Bulletin**, v. 96, n. 10, p. 1843–1866, 2012.

SEVERINO, M. DO C. G. **Evolução da sedimentação carbonática/siliciclástica miocênica na porção sul da bacia de Santos, margem sudeste do Brasil**. [s.l.] Universidade Federal do Rio Grande do Sul, 2000.

SHAKIBA, S. et al. Fault and non-fault areas detection based on seismic data through min/max autocorrelation factors and fuzzy classification. **Journal of Natural Gas Science and Engineering**, v. 26, p. 51–60, 2015.

SKIRIUS, C. et al. 3-D seismic attributes applied to carbonates. **The Leading Edge**, v. 18, n. 3, p. 384–393, 1999.

TANER, M. T.; KOEHLER, F.; SHERIFF, R. E. Complex seismic trace analysis. **GEOPHYSICS**, v. 44, n. 6, p. 1041–1063, jun. 1979.

ULTSCH, A.; MÖRCHEN, F. ESOM-Maps: tools for clustering, visualization, and classification with Emergent SOM. **Technical Report Dept. of Mathematics and Computer Science, University of Marburg, Germany**, p. 1–7, 2005.

UNGLERT, K.; RADÍĆ, V.; JELLINEK, A. M. Principal component analysis vs. self-organizing maps combined with hierarchical clustering for pattern recognition in volcano seismic spectra. **Journal of Volcanology and Geothermal Research**, v. 320, p. 58–74, 2016.

VANDERPLAS, J. **In Depth: Principal Components Analysis**. Disponível em: <<https://jakevdp.github.io/PythonDataScienceHandbook/05.09-principal-component-analysis.html>>. Acesso em: 16 dez. 2018.

VESANTO, J.; ALHONIEMI, E. Clustering of the self-organizing map. **IEEE Transactions on Neural Networks**, v. 11, n. 3, p. 586–600, 2000.

VON HARTMANN, H. et al. 3-D seismic analysis of a carbonate platform in the Molasse Basin - reef distribution and internal separation with seismic attributes. **Tectonophysics**, v. 572–573, p. 16–25, 2012.

ZHAO, T. et al. A comparison of classification techniques for seismic facies recognition. **Interpretation**, v. 3, n. 4, p. SAE29-SAE58, 2015.

9 APÊNDICE A

Uma alternativa à simples observação do valor dos erros de quantização é observar os resultados das classificações no domínio espacial original. Dentre as análises apresentadas a seguir, apenas foram alterados os atributos de entrada da RNA, os parâmetros de treinamento e quantidade de classes foi mantido.

A análise visa reforçar algumas das informações já apresentadas anteriormente, que são: utilizar o maior erro de quantização para identificar a melhor classificação e a utilização de todas as componentes da PCA resulta em uma melhor classificação pois contribui para o aumento do erro de quantização durante o treinamento dos SOM.

A análise 1 (fig. 9.1 a), quando observada no domínio espacial original, possui erro de quantização de 0,0412 e erro topográfico de 0,0971. A delimitação de classes pelas curvas de nível é, a princípio, um fator positivo embora só tenham sido utilizados atributos estruturais (mapa estrutural, mapa de isópacas e mapa de primeira derivada). Como foi mencionado anteriormente, é procurada uma integração das características dos atributos, não apenas características estruturais.

Figura 9.1- Mapa de distribuição de classes para as análises 1 (a) 2 (b).

Fonte: Bronizeski (2018).

A segunda análise (fig. 9.1 b) utiliza apenas atributos obtidos a partir do traço sísmico e os resultados são diferentes dos da fig. 9.1a). Os erros desta RNA foram: Erro de quantização – 0,195, Erro topográfico – 0,0733. As feições observadas no mapa de amplitude são repetidas nesta análise, provavelmente o mapa de amplitudes e de impedância acústica tiveram maior influência que os demais atributos. É esperado um comportamento como este pois, muitas das feições presentes no mapa de amplitude são repetidas pelo mapa de impedância acústica. Este fato é mais evidente pela distribuição das amostras da classe 1.

Avançaremos à análise 6 (fig 9.2 a), foram utilizados 6 atributos que são: amplitude, fase instantânea, frequência instantânea, impedância acústica e o mapa de isópaca. Nesta análise foi obtido erro de quantização de 0,0297 e erro topográfico de 0,0754. Características da análise 1, como a área restrita identificada como classe 6, e características da análise 2, como a disposição das classes 1 e 2,

indicam o benefício de incluir atributos com diferentes significados, pois o resultado da classificação reflete características dos atributos observados separadamente. Além deste fato, esta análise 6 nos indica que o valor maior do erro de quantização levou a uma classificação mais robusta, ou seja, com características oriundas de mais fontes ou atributos.

Fonte: Bronizeski (2018)

Ao realizar a análise 9 (fig. 9.2 b), foram utilizados apenas os 6 primeiros componentes obtidos pela PCA. Embora o erro de quantização seja maior do que na análise 6, os resultados não são melhores que os da análise 6. A delimitação da classe 6 desaparece e na região norte da área, ocorrem as classes 4, 5 e 7. Não se identifica uma melhor delimitação da disposição das amostras das classes, indicando contrastes geológicos.

O que a princípio poderia se tornar um conflito de informações, ou seja, o aumento do erro de quantização piorou os resultados classificados, é elucidado ao observar as análises 10 e 11 (fig. 9.3 a e b respectivamente). Ao serem inseridos os componentes com menor variância, ocorre melhora na definição das classes. Estes resultados indicam que as informações mesmo com baixa variância influenciam no resultado das classificações e devem ser incluídas e não desprezadas, evitando uma redução de dimensionalidade.

Fonte: Bronizeski (2018).

10 APÊNDICE B

A tarefa mais recorrente para aplicação da PCA é redução de dimensionalidade. Tal técnica possibilita que uma análise complexa de diversas componentes se torne uma tarefa simplificada, uma vez que reduz a quantidade de dados a ser analisada. O ganho almejado está na eficiência computacional com pequena perda de informação (FRERY; PERCIANO, 2013) ou na facilidade na interpretação de um menor número de variáveis. Outras aplicações que envolvem a PCA consistem na utilização desta como filtro de ruído aleatório, extração de características, dentre outras possibilidades (VANDERPLAS, 2016). A possível melhora de contraste, também foi destacada por Frery e Perciano (2013).

Seguindo os trabalhos de Vandeplas(2016), foram feitos dois exemplos sobre a reconstrução de uma imagem a partir dos principais componentes. Os exemplos utilizam a biblioteca Scikit-learn (PEDREGOSA et al., 2011) tanto para a transformação PCA quanto para o banco de dados de imagem. Os exemplos testam a transformação e reconstrução de imagens em tons de cinza com 2914 pixels, contendo 75, 150, 300 e 600 principais componentes. Estes números foram escolhidos a partir do exemplo citado dobrando o número de componentes a cada tentativa.

O primeiro exemplo (fig. 10.1) mostra um resultado satisfatório, sem haver perda de características da foto original a partir da reconstrução utilizando 150 principais componentes e cuja variância acumulada é próxima a 90% (0,9 na escala da fig. 10.3). Já no caso da figura 10.2, algumas características são perdidas ao se utilizar apenas os primeiros 150 componentes. Algumas características mais sutis não estão representadas nos primeiros 150 principais componentes como as rugas na testa e marcas de expressão ao redor da boca e nariz.

A variância acumulada demonstra variações sutis entre utilizar 150 ou 600 principais componentes. No entanto, são observadas diferenças consideráveis nos dados reconstruídos da expressão facial apresentada ao se utilizar um maior número de principais componentes no segundo exemplo (fig. 10.2), quando comparado ao primeiro exemplo (fig. 10.1).

Figura 10.1 - Exemplo 1 de reconstrução de uma imagem a partir de PC's com perda desprezível de características com 150 PC's ou aproximadamente 96% da variância total

Fonte: Bronizeski (2018).

Figura 10.2 - Exemplo 2 de reconstrução de uma imagem a partir de PC's com perda de características com 150 PC's ou aproximadamente 96% da variância total

Fonte: Bronizeski (2018).

Neste trabalho foram utilizadas apenas 8 variáveis para descrever o problema de classificação de fácies carbonática, caracterizadas por diferenças sutis e de pequena extensão. Além disso, vale ressaltar que o tempo de execução de cada

análise não foi um fator limitante. Portanto, ao se utilizar a PCA neste trabalho, sem a necessidade de redução de dimensionalidade, tem-se como premissa a obtenção de variáveis melhor condicionadas às entradas para a classificação por SOM. Entende-se que este formato é capaz de maximizar os contrastes geológicos sutis passíveis de identificação.

Figura 10.3 - Gráfico de variância cumulativa explicada de acordo com o número de PC's utilizada

Fonte: Bronizeski (2018).