



**UNIVERSIDADE FEDERAL DO CEARÁ
INSTITUTO DE CIÊNCIAS DO MAR - LABOMAR
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS MARINHAS TROPICAIS**

DAVID AURÉLIO LIMA SILVEIRA

**TÉCNICAS COMPUTACIONAIS APLICADAS NA ESTIMATIVA DA PRESSÃO
PARCIAL DE CO₂ EM ÁGUAS OCEÂNICAS**

FORTALEZA

2017

DAVID AURÉLIO LIMA SILVEIRA

TÉCNICAS COMPUTACIONAIS APLICADAS NA ESTIMATIVA DA PRESSÃO
PARCIAL DE CO₂ EM ÁGUAS OCEÂNICAS

Dissertação De Mestrado apresentada ao Programa de Pós-Graduação em Ciências Marinhas Tropicais do Instituto de Ciências do Mar (LABOMAR) da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Mestre em Ciências Marinhas Tropicais. Área de concentração: Utilização e Manejo de Ecossistemas Marinho e Estuarinos.

Orientadora: Prof^a. Dr^a. Rozane Valente Marins.
Coorientador: Dr. Francisco de Assis Tavares Ferreira da Silva.

FORTALEZA

2017

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- S587t Silveira, David Aurélio Lima.
Técnicas computacionais aplicadas na estimativa da pressão parcial de CO₂ em águas oceânicas /
David Aurélio Lima Silveira. – 2017.
81 f. : il. color.
- Dissertação (mestrado) – Universidade Federal do Ceará, Instituto de Ciências do Mar, Programa de Pós-
Graduação em Ciências Marinhas Tropicais, Fortaleza, 2017.
Orientação: Profa. Dra. Rozane Valente Marins.
Coorientação: Prof. Dr. Francisco de Assis Tavares Ferreira da Silva.
1. Inteligência computacional. 2. Estimativa. 3. pCO₂. I. Título.

CDD 551.46

DAVID AURÉLIO LIMA SILVEIRA

TÉCNICAS COMPUTACIONAIS APLICADAS NA ESTIMATIVA DA PRESSÃO
PARCIAL DE CO₂ EM ÁGUAS OCEÂNICAS

Dissertação De Mestrado apresentada ao Programa de Pós-Graduação em Ciências Marinhas Tropicais do Instituto de Ciências do Mar (LABOMAR) da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Mestre em Ciências Marinhas Tropicais. Área de concentração: Utilização e Manejo de Ecossistemas Marinho e Estuarinos.

Aprovada em: __/__/____.

BANCA EXAMINADORA

Prof^a. Dr^a. Rozane Valente Marins. (Orientadora)
Universidade Federal do Ceará (UFC)

Dr. Francisco de Assis Tavares Ferreira da Silva.
Instituto Nacional de Pesquisas Espaciais (INPE)

Prof. Dr. Carlos Eduardo Peres Teixeira
Universidade Federal do Ceará (UFC)

Dr. Antônio Macílio Pereira de Lucena
Instituto Nacional de Pesquisas Espaciais (INPE)

Dedico a meus pais, Pedro Aurélio e Maria das Graças, por toda dedicação para criar seus três filhos, deixando como herança a educação. Aos meus irmãos, Pedro e Marco, pelo apoio e incentivo de sempre. A minha esposa, Isabelle Caracas, por sua amizade, lealdade, fidelidade, paixão, compreensão, seu carinho, amor e respeito. A nossa amada filha, Luíza, por suportar minhas ausências durante longas jornadas de trabalho.

AGRADECIMENTOS

À Universidade Federal do Ceará em nome do seu Magnífico Reitor Henry de Holanda Campos, pela oportunidade.

Ao Instituto de Ciências do Mar em nome de sua Diretora Prof^a. Maria Ozilea Bezerra Menezes, pelo acolhimento.

Ao Laboratório de Biogeoquímica Costeira em nome de sua Coordenadora Prof^a. Dr^a. Rozane Valente Marins pelo fornecimento de dados costeiros valiosos que tornaram possível a execução deste trabalho, bem como aos projetos CNPq 306699/2009-5 e FUNCAP/CNPq/Pronex que tornaram viável essa pesquisa.

A Prof^a. Dr^a. Rozane Valente Marins pela excelente orientação e esclarecimentos biogeoquímicos pertinentes a execução de desenvolvimento deste trabalho.

Ao Instituto Nacional de Pesquisas Espaciais (INPE) em especial ao Dr. Francisco de Assis Ferreira Tavares da Silva, pelas contribuições decisivas referentes as técnicas de inteligência computacional.

Aos professores participantes da banca examinadora, Prof. Dr. Carlos Eduardo Peres Teixeira e Dr. Antônio Macílio Pereira de Lucena pelo tempo, pelas valiosas colaborações e sugestões, desde a qualificação deste trabalho.

Agradeço também a todos aqueles que, na ânsia de finalizar este trabalho eu possa ter esquecido, uma vez mais, obrigado.

“Você será feliz, mas antes a vida te ensinará a ser forte.”

Autor desconhecido.

RESUMO

As regiões costeiras apresentam poucos dados referentes a $p\text{CO}_2$ das águas marinhas. Em parte, isso se deve a grande dificuldade de modelar essas regiões biogeoquimicamente ativas, devido à complexidade de processos que ocorrem nessas áreas, às eventuais alterações ambientais e seus efeitos sobre os ecossistemas costeiros e aos elevados custos da logística necessária a estas medições no oceano. Portanto, as estimativas, eventualmente, tornam-se necessárias a partir de um grupo de variáveis de mensuração mais fáceis de serem executadas. A estimativa da $p\text{CO}_2$ encontra-se registrada em vários trabalhos, entretanto sua aplicação tem sido direcionada para oceano aberto e não para plataformas continentais. Técnicas de regressão linear múltipla (RLM) e de rede neural artificial (RNA) foram confrontadas diversas vezes, entretanto, a utilização do algoritmo de Levenberg-Marquardt, comumente utilizado na visão computacional, principalmente quando se deseja ajustar um modelo a um conjunto de dados experimentais, até o momento, não foi utilizado para a estimativa da $p\text{CO}_2$ na costa brasileira. Assim, neste estudo, os algoritmos de aprendizado de máquina foram utilizados, para estimar a $p\text{CO}_2$ na costa nordeste equatorial do Brasil, utilizando técnicas estatísticas e/ou de inteligência artificial. Para treinamento das técnicas computacionais foram utilizados os dados do navio MS Monte Olivia, tendo em vista a certificação destes em bancos de dados internacionais e, posteriormente aplicadas estas técnicas aos dados do navio hidro oceanográfico Cruzeiro do Sul (NHOCS) da Marinha do Brasil obtidos pelo Grupo de Pesquisa Biogeoquímica Costeira. Foram trabalhados dados de $p\text{CO}_2$, temperatura da superfície do mar (TSM) e salinidade da superfície do mar (SSM). Foram obtidos os Desvios Padrão Residual (RSD) para justificar a melhor técnica baseado no menor erro. A RNA apresentou melhores resultados em relação a RLM com a obtenção de RSDs equivalentes a $0,1034\mu\text{atm}$ para os dados totais do Monte Olívia e $0,8394\mu\text{atm}$ para os dados totais do Cruzeiro do Sul, ou seja, valores menores do que aqueles divulgados na literatura e, para o oceano aberto, as grandezas TSM e SSM são mais significativas e representam melhor a $p\text{CO}_2$, entretanto, na plataforma continental, por meio destas duas únicas variáveis, a estimativa da $p\text{CO}_2$ não foi satisfatória. Entretanto, mesmo com menor número de dados, acrescentando-se uma nova variável, clorofila *a*, a estimativa por RNA mostrou resultado satisfatório, com RSD equivalente a $0,5983\mu\text{atm}$. Quando combinou-se Salinidade à Clorofila *a* o RSD, via RNA, decresceu para $0,3514\mu\text{atm}$ após normalização dos dados, resultado que corroborou com estudo realizado na área com dados medidos.

Palavras-chave: Inteligência computacional. Estimativa. $p\text{CO}_2$.

ABSTRACT

Coastal regions have little data regarding pCO₂ in marine waters. In part, this is due to the great difficulty of modeling these biogeochemically active regions, due to the complexity of coastal processes, the possible environmental changes and their effects on ecosystems and the high logistics costs required for these measurements in the ocean. Therefore, estimates based on a group of measurement variables that are easier to perform are eventually necessary. The estimate of pCO₂ is recorded in several studies, however its application has been directed to the open ocean and not to continental shelves. Multiple linear regression (MLR) and artificial neural network (ANN) techniques have been confronted several times. However, the use of the Levenberg-Marquardt algorithm, commonly used in computational vision when it is desired to fit a model to a set of data, is to date, not used for the estimation of pCO₂ in the Brazilian coast. Thus, in this study, machine-learning algorithms were used to estimate pCO₂ on the equatorial northeast coast of Brazil, using statistical and/or artificial intelligence techniques. For the training of the computer techniques, the data of the MS Monte Olivia ship were used, in order to certify them in international databases, and later applied these techniques to the Cruzeiro do Sul (NHOCS) data of the Brazilian Navy obtained by Coastal Biogeochemistry Research Group. Data were analyzed for pCO₂, sea surface temperature (SST) and sea surface salinity (SSS). We obtained the Residual Standard Deviations (RSD) to justify the best technique based on the smallest error. The RNA presented better results in relation to MLR with the obtaining of RSDs equivalent to 0.1034μatm for the total data of Monte Olivia and 0,8394μatm for the total data of the Cruzeiro do Sul, that is, smaller values than those published in the literature. In addition, for the open ocean, the SST and SSS measures are more significant and represent better the pCO₂ than in the continental shelf. By means of these two only variables, the estimate of the pCO₂ was not satisfactory. However, even with a lower number of data, adding a new variable, chlorophyll a, ANN estimation showed a satisfactory result, with RSD equivalent to 0.5983μatm. When only Salinity and Chlorophyll are combined, the RSD via RNA drops to 0.3514μatm for the estimated data after normalization.

Keywords: Computational intelligence. Estimation. pCO₂.

LISTA DE FIGURAS

Figura 1	– Concentração de CO ₂ na atmosfera nos últimos séculos	14
Figura 2	– Média anual da temperatura global observada e projetada referente ao período de 1986-2005. As linhas pretas mostram a temperatura estimada de três conjuntos de dados. Linhas azuis, vermelhas e as sombreadas indicam o conjunto médio e $\pm 1,64$ a ordem do desvio padrão	15
Figura 3	– O pH das águas superficiais no Oceano Atlântico	19
Figura 4	– Perfil da profundidade e de pH no Atlântico	20
Figura 5	– Temperatura, Salinidade e Alcalinidade Total (AT) nas águas do Oceano Atlântico	21
Figura 6	– Alcalinidade Total em função da salinidade nas águas superficiais do Atlântico	22
Figura 7	– Ciclo Biogeoquímico do Carbono	23
Figura 8	– Distribuição global de pCO ₂ (μatm) para as águas superficiais dos oceanos do mundo	26
Figura 9	– Estrutura de um Neurônio Celular	31
Figura 10	– Esquema de neurônio artificial	32
Figura 11	– Representação gráfica de uma RNA com duas camadas intermediárias	32
Figura 12a	– Representação gráfica de problema não-linearmente separável	37
Figura 12b	– Representação gráfica de problema linearmente separável	37
Figura 13	– Representação gráfica de RNA com seus pesos	38
Figura 14	– Navio Mercante Monte Olivia	44
Figura 15	– Rota do Navio MS Monte Olivia. Dados utilizados neste estudo referem-se a área delimitada em amarelo, entre 1°S a 7°S	44
Figura 16	– Mapa de pontos da malha amostral da Campanha de Setembro de 2014 a bordo do NHO. Cruzeiro do Sul	45

Figura 17 – Navio Hidro Oceanográfico Cruzeiro do Sul (H-38S)	46
Figura 18 – Representação esquemática do treinamento de dados utilizando o algoritmo de Levenberg-Marquardt	51
Figura 19a – Dados de Temperatura em função do número da medição – Monte Olívia .	55
Figura 19b – Dados de Salinidade em função do número da medição – Monte Olívia	55
Figura 19c – Dados de pCO ₂ em função do número da medição – Monte Olívia	55
Figura 19d – Dados de temperatura, salinidade e pCO ₂ normalizados – Monte Olívia.....	56
Figura 20 – Dados ímpares e pares de pCO ₂ normalizados – Monte Olívia	56
Figura 21 – Estimativa de pCO ₂ pares normalizados via regressão obtida a partir de dados ímpares – Monte Olívia	57
Figura 22a – Representação dos dados medidos da pCO ₂ da campanha Monte Olívia normalizada entre 1 e -1	59
Figura 22b – Representação da estimativa da pCO ₂ da campanha Monte Olívia normalizada entre 1 e -1	59
Figura 22c – Representação dos dados reais da pCO ₂ da campanha Monte Olívia normalizada entre 1 e -1 utilizados para treinamento / obtenção da equação de regressão	59
Figura 23a – Dados pares relativos a pCO ₂ em relação ao número da medição – Monte Olívia	61
Figura 23b – Estimativa de pCO ₂ pares normalizados via regressão obtida a partir de dados ímpares – Monte Olívia	61
Figura 24 – Representação esquemática do treinamento, validação e teste da rede neural utilizada para estimar a pCO ₂ par nos dados da campanha do Navio Monte Olívia	62
Figura 25 – Representação esquemática do funcionamento da rede neural com suas principais informações	63
Figura 26a – Dados de Temperatura em função do número da medição – Cruzeiro do Sul	66

Figura 26 <i>b</i> – Dados de Salinidade em função do número da medição – Cruzeiro do Sul	66
Figura 26 <i>c</i> – Dados de Temperatura em função do número da medição – Cruzeiro do Sul	66
Figura 27 – Dados ímpares e pares de pCO ₂ normalizados sobrepostos - Cruzeiro do Sul.....	67
Figura 28 – Estimativa de pCO ₂ pares normalizados via regressão obtida a partir de dados ímpares – Cruzeiro do Sul	67
Figura 29 – Distribuição de Carbono 13 durante o Cruzeiro do Sul	69
Figura 30 <i>a</i> – Dados pares relativos a pCO ₂ em relação ao número da medição – Cruzeiro do Sul	70
Figura 30 <i>b</i> – Estimativa de pCO ₂ pares normalizados via regressão obtida a partir de dados ímpares – Cruzeiro do Sul	70
Figura 31 – Representação esquemática do treinamento, validação e teste da rede neural utilizada para estimar a pCO ₂ par nos dados da campanha do Navio Hidro-Oceanográfico Cruzeiro do Sul	71
Figura 32 – Representação esquemática do funcionamento da rede neural com suas principais informações	72

LISTA DE TABELAS

Tabela 1a – Exemplo da primeira etapa de separação de dados em planilha Excel® para isolamento de dados denominados ímpares e pares para uso estatístico	48
Tabela 1b – Exemplo da segunda etapa de separação de dados em planilha Excel® para isolamento de dados denominados ímpares (A) para uso estatístico	48
Tabela 1c – Exemplo da segunda etapa de separação de dados em planilha Excel® para isolamento de dados denominados pares (B) para uso estatístico	49
Tabela 2 – Comparações entre erro padrão, coeficiente de determinação e desvio padrão residual – Monte Olívia.....	58
Tabela 3 – Comparações entre erro padrão, coeficiente de determinação e desvio padrão residual utilizando somente a salinidade, somente a temperatura e utilizando salinidade e temperatura	60
Tabela 4 – RSD entre valores medidos e estimados nos dados salteados	63
Tabela 5 – RSD entre valores medidos e estimados nos dados salteados	72
Tabela 6 – RSD entre valores medidos e estimados nos dados salteados do Monte Olívia	73
Tabela 7 – RSD de variáveis utilizando RNA do Cruzeiro do Sul	75

LISTA DE ABREVIATURAS E SIGLAS

AT	Alcalinidade Total.
CDIAC	Carbon Dioxide Information Analysis Center
CID	Carbono Inorgânico Dissolvido.
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico.
COT	Carbono Orgânico Total.
IA	Inteligência Artificial.
IPCC	Painel Intergovernamental sobre Mudanças Climáticas.
LM	Levenberg-Marquardt.
pCO ₂	Pressão Parcial de CO ₂ .
pH	Potencial de Hidrogênio.
PROANTAR	Programa Antártico Brasileiro.
RLM	Regressão Linear Múltipla.
RNA	Rede Neural Artificial.
ROC	Radiações de ondas curtas.
ROL	Radiação de Ondas Longas.
RSD	Desvio Padrão Residual.
SSM	Salinidade da Superfície do Mar.
TSM	Temperatura da Superfície do Mar.

LISTA DE SÍMBOLOS

%	Porcentagem
©	Copyright
®	Marca Registrada
μm	Micrômetro
μatm	Microatmosfera
δ ¹³ C	Carbono 13

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Medidas dos parâmetros de sistema de CO₂ na água do mar	19
<i>1.1.1</i>	<i>pH</i>	<i>19</i>
<i>1.1.2</i>	<i>Alcalinidade Total (AT)</i>	<i>21</i>
<i>1.1.3</i>	<i>Carbono Orgânico Total (COT)</i>	<i>23</i>
<i>1.1.4</i>	<i>Pressão parcial de CO₂ – pCO₂</i>	<i>24</i>
<i>1.1.5</i>	<i>Estimativa da pCO₂</i>	<i>27</i>
2	REFERENCIAL TEÓRICO	28
2.1	Técnicas Computacionais de Estimativa	28
2.2	Regressão Linear Múltipla (RLM)	29
2.3	Rede Neural Artificial (RNA)	30
<i>2.3.1</i>	<i>Classificação das Redes Neurais Artificiais</i>	<i>34</i>
<i>2.3.1.1</i>	<i>Rede Neural do tipo Feed forward</i>	<i>34</i>
<i>2.3.1.2</i>	<i>Mapas auto-organizáveis (SOM)</i>	<i>35</i>
<i>2.3.1.3</i>	<i>Rede Neural Hopfield</i>	<i>35</i>
<i>2.3.1.4</i>	<i>Rede Neural recorrente simples (SRN) - Elman ou Jordan Style</i>	<i>35</i>
<i>2.3.1.5</i>	<i>Rede Neural Feed forward baseada em função de base Radial (RBF)</i>	<i>35</i>
<i>2.3.1.6</i>	<i>Algoritmo de Levenberg-Marquardt (LM)</i>	<i>36</i>
2.4	RNA com aplicação do algoritmo de Levenberg-Marquardt (LM)	36
2.5	Rede Neural Artificial vs Regressão Linear Múltipla	38
2.6	O uso de técnicas de Inteligência Artificial para estimativa da pCO₂	39
3	HIPÓTESE	42
4	OBJETIVOS	43
4.1	Objetivo Geral	43
4.2	Objetivos Específicos	43

5	MATERIAL E MÉTODOS	44
5.1	Área de Estudo	44
5.2	Treinamento e Aplicação de dados da campanha do Navio Mercante Monte Olívia em Técnicas Computacionais – Regressão Linear Múltipla	46
5.3	Treinamento e Aplicação de dados da campanha do Navio Mercante Monte Olívia em Técnicas Computacionais – Rede Neural Artificial	50
5.4	Treinamento e Aplicação de dados da campanha do Navio Hidro Oceanográfico Cruzeiro do Sul (H-38S) em Técnicas Computacionais – Regressão Linear Múltipla	52
5.5	Treinamento e Aplicação de dados da campanha do Navio Hidro Oceanográfico Cruzeiro do Sul (H-38S) em Técnicas Computacionais – Rede Neural Artificial	53
6	RESULTADOS E DISCUSSÃO	54
6.1	Treinamento e Aplicação de dados da campanha do Navio Mercante Monte Olívia em Técnicas Computacionais – Regressão Linear Múltipla	54
6.1.1	<i>Representação gráfica dos dados utilizados através de técnicas de Estatística Clássica</i>	54
6.1.1.1	<i>Gráficos dos dados totais por variável</i>	54
6.1.1.2	<i>Treinamento com dados ímpares e validação com dados pares</i>	56
6.1.1.3	<i>Treinamento e validação com dados salteados</i>	58
6.2	Treinamento e Aplicação de dados da campanha do Navio Mercante Monte Olívia em Técnicas Computacionais	60
6.2.1	<i>Representação gráfica dos dados utilizados através de Rede Neural Artificial</i>	61
6.2.1.1	<i>Treinamento com dados ímpares e validação com dados pares</i>	61
6.2.1.2	<i>Treinamento e validação com dados salteados</i>	63

6.3	Treinamento e Aplicação de dados da campanha do Navio Hidro Oceanográfico Cruzeiro do Sul em Técnicas Computacionais - Regressão Linear Múltipla	64
6.3.1	<i>Representação gráfica dos dados utilizados através de Regressão Linear Múltipla</i>	64
6.3.1.1	<i>Gráficos dos dados totais por variável</i>	64
6.3.1.2	<i>Treinamento com dados ímpares e validação com dados pares</i>	66
6.4	Treinamento e Aplicação de dados da campanha do Navio Hidro Oceanográfico Cruzeiro do Sul em Técnicas Computacionais – Rede Neural Artificial	69
6.4.1	<i>Representação gráfica dos dados utilizados através de Rede Neural Artificial</i>	69
6.4.1.1	<i>Treinamento com dados ímpares e validação com dados pares</i>	69
6.4.1.2	<i>Treinamento e validação com dados salteados</i>	72
6.5	Regressão Linear Múltipla vs Rede Neural Artificial	73
6.5.1	<i>A influência da Clorofila a na estimativa</i>	75
7	CONCLUSÃO	76
	REFERÊNCIAS	78

1 INTRODUÇÃO

O Sol, fonte primária de energia para o planeta, emite radiação eletromagnética principalmente nos comprimentos de onda que variam entre $0,1\mu\text{m}$ e $4,0\mu\text{m}$, que caracterizam as Radiações de Ondas Curtas (ROC). O percentual de ROC refletido para o espaço exterior é de aproximadamente 30% e é resultante, principalmente, do tipo de nuvens, da concentração de aerossóis, das partículas em suspensão no ar e das características da cobertura superficial tais como gelo/neve (90% de reflexão), florestas (12%) e oceanos/lagos (10%) (MOLION, 2008).

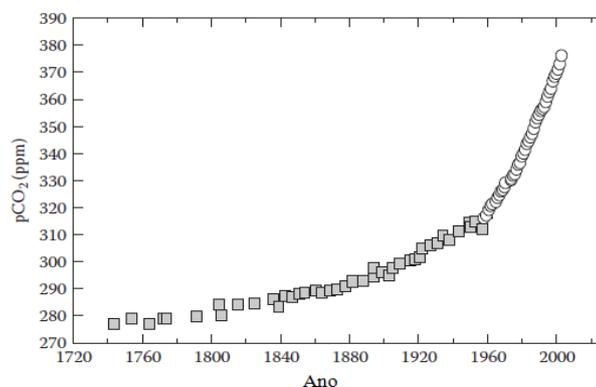
Para as temperaturas dos corpos encontrados tanto na superfície como na atmosfera terrestre, os comprimentos de onda emitidos estão entre $4,0\mu\text{m}$ e $50\mu\text{m}$, numa faixa espectral denominada Radiação de Ondas Longas (ROL).

O fluxo de ROL emitido pela superfície é absorvido por gases e pequenos constituintes, como o vapor d'água (H_2O), gás carbônico (CO_2), gás metano (CH_4), ozônio (O_3), óxido nitroso (N_2O) e os compostos de clorofluorcarbono (CFC), que por sua vez, emitem ROL em todas as direções, inclusive em direção à superfície e ao espaço exterior.

A absorção/emissão por estes gases nas várias camadas atmosféricas reduz a perda de ROL emitida pela superfície, que escaparia para o espaço exterior, constituindo o chamado efeito-estufa.

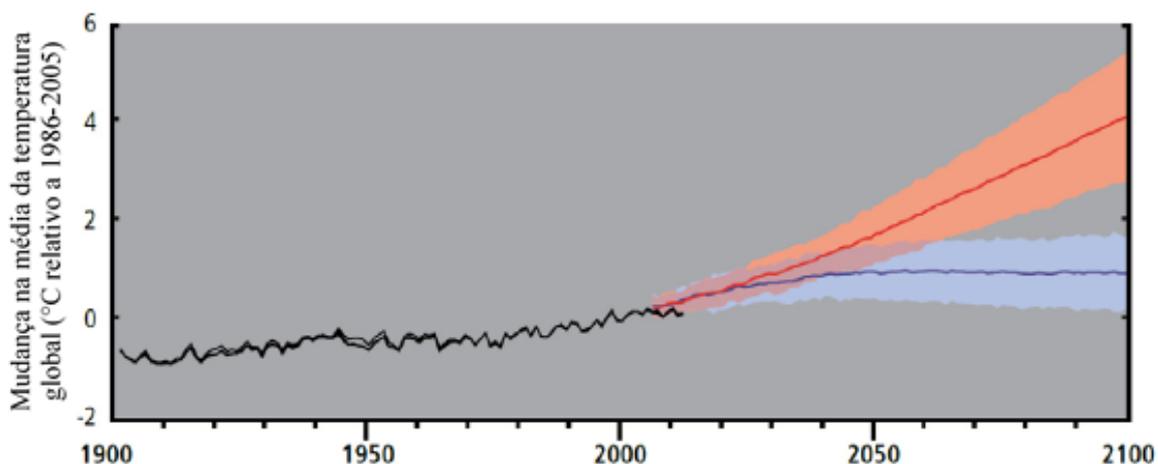
A concentração de CO_2 na atmosfera aumentou consideravelmente no século XX (Figura 1). Uma vez que a concentração de CO_2 esteja aumentada, há maior absorção de radiação infravermelha, o que está elevando a temperatura da Terra (Figura 2).

Figura 1 - Concentração de CO_2 na atmosfera nos últimos séculos.



Fonte: Adaptado de Millero (2013).

Figura 2 - Média anual da temperatura global observada e projetada referente ao período de 1986-2005. As linhas pretas mostram a temperatura estimada de três conjuntos de dados. Linhas azuis, vermelhas e as sombreadas indicam o conjunto médio $\pm 1,64$ a ordem do desvio padrão.



Fonte: Adaptado de IPCC (2015).

Nos últimos anos, o interesse em química atmosférica tem se mostrado cada vez mais crescente, provavelmente devido ao aumento da temperatura média global de aproximadamente $0,7^{\circ}\text{C}$ (MILLERO, 2013).

Esse aumento é atribuído à intensificação do efeito-estufa, devido às atividades antrópicas, classificadas aqui em dois tipos: a queima de combustíveis fósseis como o carvão, o petróleo e o gás natural e o desmatamento tropical pela queima da biomassa (MILLERO, 2013).

O aquecimento global é uma realidade e a contribuição do ser humano é significativa para a ocorrência de fenômenos ligados às mudanças climáticas. Por isso, ações imediatas, em escala global, são necessárias para reverter o que for possível, pois as magnitudes e taxas de mudança climática, neste século, apresentam alto risco de mudança abrupta e irreversível em escala regional na composição, estrutura e função dos ecossistemas terrestres e ecossistemas de água doce, incluindo as zonas úmidas. Impactos oriundos de eventos climáticos extremos como, por exemplo, ondas de calor, secas, inundações, ciclones, incêndios florestais e furacões (mais frequentes e mais fortes em águas mais quentes) revelam a significativa vulnerabilidade e a exposição de alguns ecossistemas – e de muitos sistemas humanos – à mudança climática atual (IPCC, 2014).

As interações oceano-atmosfera contribuem significativamente para a percepção do clima global tanto pela especificidade das correntes marinhas quanto pela natureza do ciclo

hidrológico (SADAVA *et al.* 2009).

Os seres vivos planctônicos, essencialmente os fitoplanctônicos são, dentre outros, os grandes responsáveis pela liberação de maior parte do O₂ disponível na atmosfera, justificando a importância do oceano para manutenção da vida aeróbica na Terra (SADAVA *et al.* 2009).

Esse mesmo fitoplâncton é o responsável pelo consumo de CO₂, que corrobora com a diminuição da taxa de crescimento de CO₂ atmosférico reduzindo assim o avanço das mudanças climáticas (FUNG *et al.*, 2005).

O CO₂ lidera os demais Gases Efeito Estufa contribuindo com 97% das emissões totais, o mesmo é proveniente de fontes móveis, admitindo-se uma incerteza de 5%, principalmente devido à operação do veículo, mais do que pela imprecisão do fator de emissão correspondente aos motores veiculares (IPCC, 2014).

A absorção do CO₂ atmosférico acrescido nestas emissões antrópicas acarreta, entretanto, em diminuição do pH da água do mar, o que eleva o potencial de decaimento do pH das águas oceânicas, processo denominado como acidificação dos oceanos, fragilizando os ecossistemas marinhos (FEELY *et al.*, 2004; ORR *et al.*, 2005; DONEY *et al.*, 2009; GRUBER, 2011).

O aquecimento da superfície das águas favorece a estratificação. Os resultados dessa estratificação podem ser traduzidos em cotas mais baixas de O₂ na superfície da água e essas alterações podem resultar em estresse sobre processos biológicos de organismos marinhos.

As alterações de umidade, as variações dos fluxos de calor e seus efeitos sobre o clima do planeta tem incentivado o desenvolvimento de estudos de interação entre o oceano e a atmosfera. Estes fluxos são de fundamental importância na manutenção e geração das circulações atmosférica e oceânica (DOYLE, 2002).

Essas circulações afetam e interagem localmente, alterando suas características e influenciando o desenvolvimento de fenômenos atmosféricos tais como ciclones tropicais e ciclones extratropicais e frentes costeiras (DOYLE, 2002). Por exemplo, quando sistemas de ar de baixa pressão encontram-se pairados sobre o mar das regiões tropicais temos um exemplo clássico de ciclones tropicais.

O sistema de ar pode ser definido como uma grande quantidade de ar que se move de forma organizada e controlada. O sistema de baixa pressão se refere à força que ar exerce sobre a superfície da Terra. Portanto, quando se afirma que uma área é de baixa pressão, isto quer dizer que o ar, naquela região, naquele momento, exerce menos força sobre a Terra do que em outras regiões. Regiões de baixa pressão acabam por atrair mais ventos promovendo um

maior equilíbrio da força que o ar exerce naquela região (FIOCRUZ, 2016).

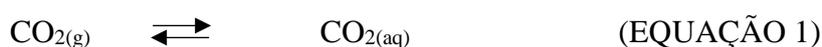
Um componente importante do sistema oceano-atmosfera é a interface atmosfera-oceano, onde pequenas variações espaciais do gradiente da Temperatura da Superfície do Mar (TSM) podem acarretar grandes variações nos fluxos dessa interface. Não obstante, a TSM também pode provocar impactos significativos no escoamento atmosférico e, por conseguinte, nos sistemas meteorológicos (LE QUERE, 2007).

O balanço de calor entre o oceano e a atmosfera tem-se demonstrado correlacionado com a TSM pois, através dela, a energia na forma de fluxos de calor é trocada. Pequenas variações de TSM podem acarretar grandes variações nos fluxos de calor na interface oceano-atmosfera e esses, por sua vez, tem papel preponderante na manutenção do clima do planeta.

Os trabalhos Pezzi e Cavalcanti (2001), por exemplo, demonstraram a influência da TSM do Oceano Atlântico tropical no clima da região nordeste do Brasil. Não obstante, a TSM também pode provocar impactos significativos no escoamento atmosférico e, por conseguinte, modular os sistemas meteorológicos na escala do tempo.

A maioria dos gases na água do mar é originária a partir de três fontes: a atmosfera da Terra, a atividade vulcânica no fundo do mar e, os processos químicos que ocorrem no mar tais como a fotossíntese e decomposição da matéria orgânica. No curso da circulação oceânica, os gases não-reativos são distribuídos ao longo da coluna d'água por convecção e difusão.

A maior parte do carbono nos oceanos ocorre como parte do sistema carbonato. Este sistema envolve os seguintes equilíbrios (Equações 1 a 4):



O sistema carbonato é muito importante, uma vez que regula o Potencial de Hidrogênio (pH) da água do mar e controla a circulação de CO_2 entre a biosfera, a litosfera, a atmosfera, e os oceanos. O interesse recente no sistema carbonato nos oceanos resultou do aumento do "efeito de estufa" ou "efeito de CO_2 ".

A entrada de $\text{CO}_{2(\text{g})}$ para o oceano ocorre através de processos físicos, as chamadas bombas de solubilidade e de processos biológicos chamados de bombas biológicas.

A bomba de solubilidade oceânica ocorre quando o CO_2 se dissolve nas águas geladas dos oceanos polares, sobretudo do Antártico. Quanto mais fria a água, maior é a

dissolução dos gases. Mas, ao contrário de outros gases atmosféricos, como o nitrogênio molecular e o próprio oxigênio, o gás carbônico reage com água e com íons de cálcio, formando carbonatos e bicarbonatos, ou seja, não é conservativo.

A captação de CO₂ dos oceanos é um processo que envolve a hidratação do CO₂, um processo lento em relação ao processo de ionização, entretanto, o aumento da adsorção de CO₂ pelos oceanos irá resultar numa diminuição no pH, processo que tem sido denominado de acidificação dos oceanos (LE QUERE, 2007).

A acidificação dos oceanos poderá causar graves consequências ambientais que afetarão principalmente a vida marinha. Prejudicará os corais e algas que formam os recifes calcários e as florações em massa das cocolitoforídeos e foraminíferos (representantes do Reino Protista que possuem uma carapaça externa formada por carbonato de cálcio), os quais dependem de pH básico para precipitação biogênica dos carbonatos em dissolução por esses organismos (SADAVA et al., 2009).

Castro et. al. (2010), afirma que em altas latitudes, o sal marinho é excluído do processo de congelamento da água de superfície, tornando as águas imediatamente abaixo do gelo mais salgadas e, portanto, mais densas. Com o aumento da densidade a água mais salgada submerge, “escorregando” pelo talude continental para camadas profundas dos oceanos Atlântico, Índico e Pacífico, levando consigo o CO₂ sob a forma de carbonatos.

Outro mecanismo importante de absorção de carbono atmosférico, antropogênico ou natural é a bomba biológica que se inicia com a absorção de CO₂ através dos organismos fotossintetizantes marinhos, geralmente fitoplânctons. Por meio do processo fotossintético, há transformação do carbono inorgânico dissolvido para carbono orgânico particulado, o qual flui constantemente através da teia alimentar gerando detritos que sedimentam e se acumulam no fundo marinho pela ação da gravidade (CASTRO *et. al.* 2010).

Uma vez que o CO₂ entra no oceano, ele se difunde em toda a interface atmosfera-oceano e participa dos processos do equilíbrio carbonato (Equações 1 a 4), além de poder ser utilizado pelas plantas na produtividade primária (EQUAÇÃO 5):



Segundo Millero (2013), estes processos estão longe de ser simples, uma vez que as taxas de circulação de CO₂ em toda a interface e desde a superfície até águas profundas variam com a latitude, o tempo, a estação, e os processos biológicos.

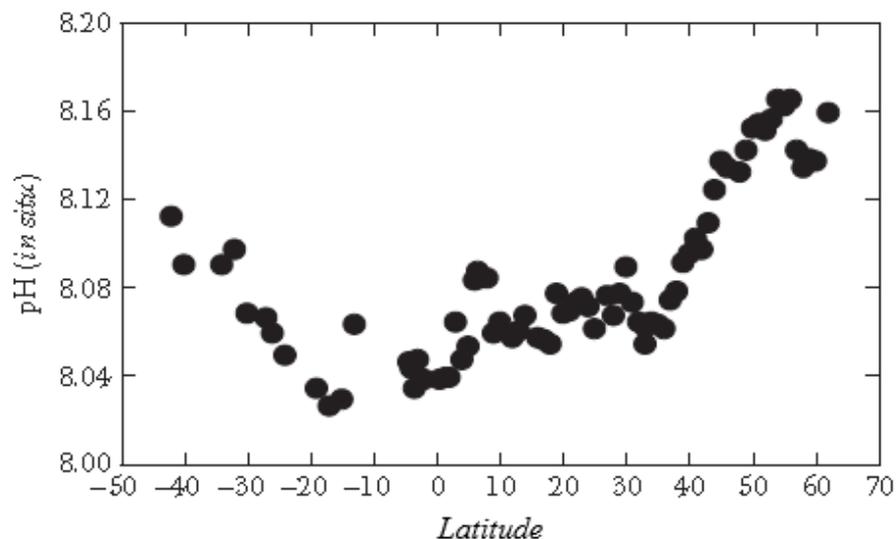
1.1 Medidas dos parâmetros do sistema de CO₂ na água do mar

Devido a capacidade do sistema carbonato controlar a circulação de CO₂ na interface oceano-atmosfera, faz-se necessário medir pelo menos dois dos quatro parâmetros mensuráveis: pH, Alcalinidade Total (AT), Carbono Orgânico Total (COT), Pressão Parcial de CO₂ (pCO₂), (MILLERO, 2013), a fim de se determinar a fugacidade de CO₂ e, conseqüentemente, avaliar a perda ou retenção de CO₂ para o meio aquático marinho.

1.1.1 pH

O pH da maior parte das águas superficiais oceânicas em equilíbrio com a atmosfera é de $8,2 \pm 0,1$. Recentes medições de pH nestas águas nos oceanos Atlântico e Pacífico mostram que estes valores tendem a diminuir a partir de 40°S de latitude até a latitude 20°S e, após um intervalo de 30°N, o pH aumenta até 60°N de latitude aproximadamente, como, por exemplo, pode ser observado na figura 3. As alterações no pH com a profundidade no Atlântico são mostradas na Figura 4.

Figura 3 - O pH das águas superficiais no Oceano Atlântico.



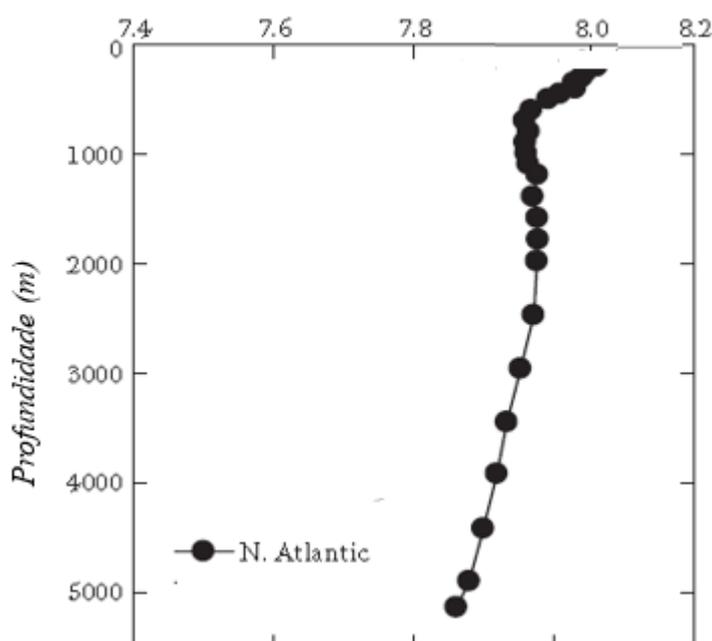
Fonte: Adaptado de Millero (2013).

Mudanças bruscas de pH ocorrem quando há um aumento da absorção de CO₂ pelas águas marinhas, ou seja, quando aumenta a pCO₂, menor é o pH. Menores valores de pH têm sido observados nas regiões equatoriais (Figura 3), em águas de ressurgência (MILLERO, 2013).

Em mares fechados ou em pequenas massas de água, o pH pode mostrar variações diurnas e seu ciclo fica entre 8,2 e 8,9. A redução ocorre no período noturno devido à respiração dos organismos e, o aumento ocorre no período vespertino devido a fotossíntese.

O pH atinge valores máximos em águas superficiais, devido a ocorrência da fotossíntese pelos organismos planctônicos, ou seja, devido a absorção de CO_2 pelos organismos fotossintetizantes. Por outro lado, a diminuição do pH ocorre devido a oxidação do material biogênico e atinge valores mínimos (cerca de 7,5) a cerca de 1 km de profundidade, enquanto que em águas profundas o aumento do pH é devido à dissolução de CaCO_3 e ao efeito da pressão sobre a ionização de ácido carbônico (MILLERO 2013).

Figura 4 - Perfil da profundidade e de pH no Atlântico.



Fonte: Adaptado de Millero (2013).

Águas costeiras apresentam geralmente elevada turbidez, dinâmica temporal e espacial, onde as trocas de CO_2 são frequentemente rápidas (COTOVICZ, 2016), entretanto, análises e estudos de alta qualidade do pH da água do mar em consonância aos ecossistemas fluviais e estuarinos não são triviais e medições de alta qualidade do pH e do sistema carbonato também são essenciais para o estudo dos efeitos da acidificação dos oceanos e regiões costeiras na biota marinha.

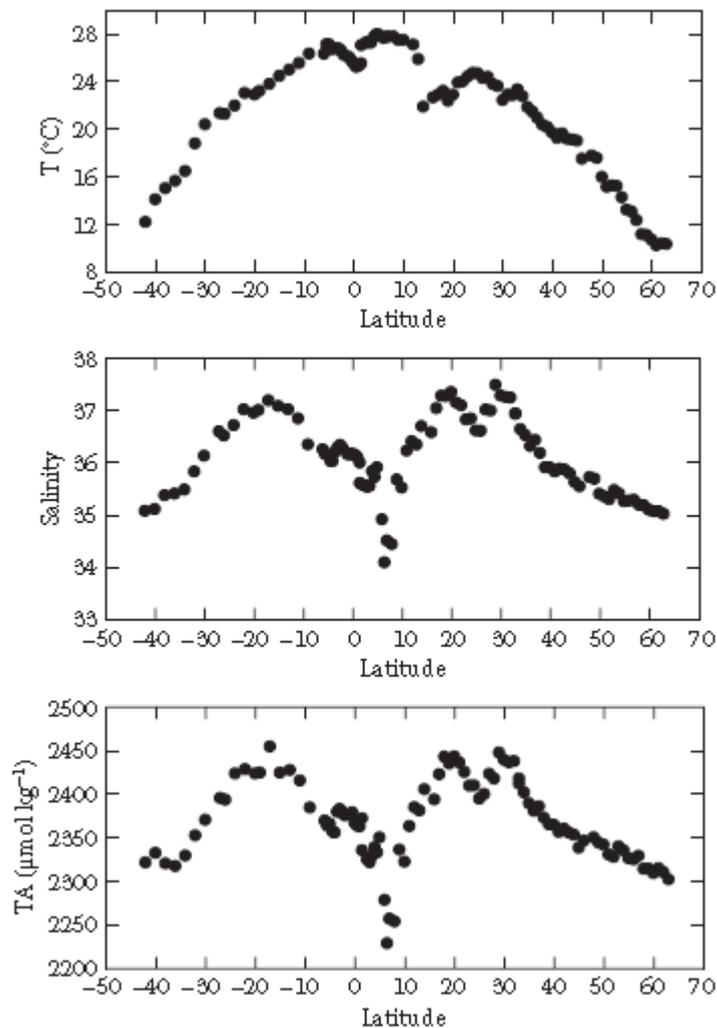
Segundo Cotovicz *et al.* (2016), o pH abaixo de 7,5, em regiões estuarinas, inviabiliza a estimativa da Pressão Parcial de CO_2 (pCO_2), em estudos comparativos realizados entre a baía de Guanabara e o Rio São Francisco.

Assim, estimativas da $p\text{CO}_2$ em regiões costeiras biogeoquimicamente ativas são dificultadas em relação às estimativas em oceano aberto.

1.1.2 Alcalinidade Total (AT)

Os valores da temperatura, salinidade e AT para as águas superficiais no oceano Atlântico encontram-se na Figura 5. Como é evidente a partir desses números, a AT das águas superficiais apresenta variações que são semelhantes à salinidade.

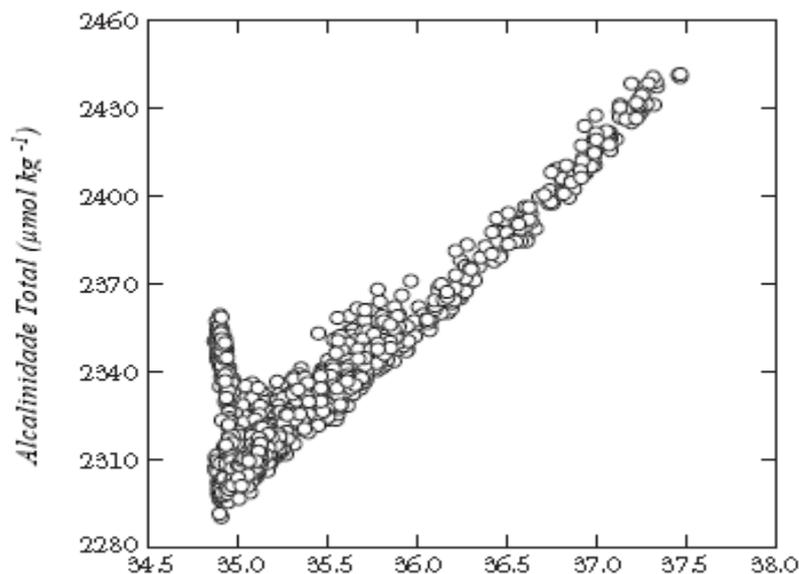
Figura 5 – Temperatura, Salinidade e Alcalinidade Total (AT) nas águas do Oceano Atlântico.



Adaptado de Millero (2013).

A relação entre AT e salinidade é melhor observada, por exemplo, quando AT está representada em função da salinidade do Atlântico Norte (Figura 6).

Figura 6 – Alcalinidade Total em função da salinidade nas águas superficiais do Atlântico.



Adaptado de Millero (2013).

Este comportamento linear ocorre porque o ânion HCO_3^- é um dos principais constituintes da água do mar e devido à salinidade ser “quase sempre constante”. Quando a salinidade muda significativamente, apresentando comportamento não-linear, com valores próximos de 35 é devido à ressurgência de águas profundas que apresentam uma alcalinidade maior devido à dissolução de CaCO_3 , neste caso.

A distribuição da Alcalinidade Total é muito semelhante às distribuições de salinidade que se seguem nas grandes massas de água. Os valores encontrados para alcalinidade total nos oceanos geralmente variam de 2000 a 2500 $\mu\text{mol.kg}^{-1}$, sendo os valores mais baixos encontrados geralmente nas águas superficiais de regiões costeiras e polares (DICKSON; SABINE; CHRISTIAN et al., 2007).

Cotovicz *et al.*, (2016), comparando as medições em tempo real da pCO_2 aquática com estimativas indiretas nos estuários da Baía de Guanabara (RJ) e do Rio São Francisco, atenta para o fato de se estabelecer estimativas mais precisas e confiáveis nos balanços de CO_2 e sua dinâmica na região costeira e oceânica, uma vez que o gradiente da pCO_2 na interface água-atmosfera é um dos componentes do cálculo dos fluxos. Seus resultados sugerem que a estimativa da pCO_2 não pode ser realizado para regiões de estuários com elevado percentual de água doce onde a alcalinidade total registra um valor abaixo de 700 $\mu\text{mol Kg}^{-1}$ sem que haja erro grosseiro desta estimativa.

1.1.3 Carbono Orgânico Total (COT)

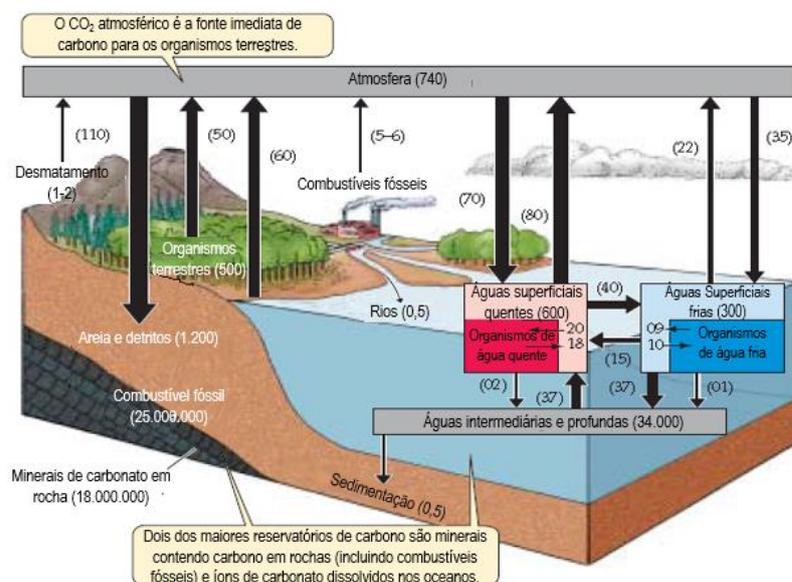
O Carbono Orgânico Total é uma variável utilizada para descrever a medição de substâncias orgânicas no sistema aquoso. Estas substâncias podem advir de uma variedade de fontes, já que podem ser de origem natural, como por exemplo açúcares, proteínas, ácidos húmicos e fúlvicos, lignina entre outras, ou ser de origem xenobiótica como derivados do petróleo, plásticos, pesticidas etc. (BAIRD; CANN, 2011).

Os compostos orgânicos podem estar presentes na água devido aos processos biogeoquímicos autóctones, ou pela lixiviação de solos, ou entrada de efluentes urbanos, industriais.

O COT é um importante componente do ciclo biogeoquímico do carbono. Os organismos para sobreviverem devem ter acesso aos átomos de carbono. Quase todo o carbono presente nos organismos é oriundo da atmosfera ou de íons carbonato dissolvidos (HCO_3^-) na água. O carbono é incorporado às moléculas orgânicas por fotossíntese nas células dos seres autótrofos. Todos os organismos heterotróficos obtêm seu carbono através do consumo de autótrofos ou de outros heterótrofos incluindo seus restos, ou os seus resíduos.

Os processos biológicos são responsáveis pela movimentação de carbono entre os meios atmosféricos e terrestres, removendo-os da atmosfera durante a fotossíntese e devolvendo-o para a atmosfera durante a respiração, (Figura 7).

Figura 7 – Ciclo Biogeoquímico do Carbono.



Fonte: Adaptado de Sadava, *et al.* (2009).

Os números representados na figura 7 demonstram as quantidades de carbono (expresso como unidades de 10^{15} gramas = 1 bilhão de toneladas), contidas ou trocadas anualmente através dos componentes do ecossistema. A largura das setas é proporcional ao tamanho dos fluxos. A maior parte do carbono da Terra está armazenado nos oceanos, entretanto a maioria do carbono que está disponível para os organismos está armazenado nos solos.

Estudos descritos por programas de pesquisa como o Land Ocean Interaction in the Coastal Zone (LOICZ, 1995), revelaram que há transporte do carbono orgânico total dos rios para o oceano, estando sujeito às modificações biogeoquímicas capazes de alterar a compartimentalização ambiental desse material.

Com o rápido desenvolvimento da economia e da população, acredita-se que as pressões antrópicas em pequenos rios, somadas às mudanças climáticas, têm modificado dramaticamente o fluxo de carbono em alguns rios. Os estuários são vias desse transporte, e são ambientes bastante dinâmicos e caracterizados por um forte gradiente físico-químico, atrelado às atividades biológicas, que dependendo das condições da maré, descarga fluvial e atmosfera sofrem variações no espaço e no tempo (CAVALCANTE, 2015).

1.1.4 Pressão Parcial de CO_2 – pCO_2

A pressão parcial do CO_2 (pCO_2) é a pressão que o gás em equilíbrio, na atmosfera, exerceria sobre uma superfície plana de água contendo CO_2 dissolvido. Em específico, segundo a lei das pressões parciais (Lei de Dalton) é uma função da pressão ambiente e da concentração do gás (fração molar), conforme a EQUAÇÃO 6 (FONSECA, 2011). Onde n_{CO_2} é a concentração do gás em moles e P é a pressão ambiente em hPa.

$$pCO_2 = n_{CO_2}P \quad \text{(EQUAÇÃO 6)}$$

A solubilidade de alguns gases, como o CO_2 , é complicada pela sua reação química com a água. No caso do CO_2 , essas reações tendem a atingir o equilíbrio rapidamente, fazendo com que a concentração de equilíbrio de CO_2 ocorra em função da temperatura, salinidade, pressão parcial atmosférica do CO_2 e pH. Os gases estão em constante troca na interface oceano-atmosfera e, quando essas taxas de trocas são iguais infere-se que o gás está em equilíbrio.

Alguns gases ainda podem aumentar sua solubilidade com a diminuição da temperatura e também com a diminuição da salinidade, entretanto essa não é uma relação linear pois pode ocorrer supersaturação de gases devido a mistura de águas de diferentes temperaturas.

Outrossim, a solubilidade aumenta com o aumento da pressão e, as taxas de difusão dos gases são as mesmas para o oceano e a atmosfera em determinadas condições de temperatura e salinidade (LIBES, 2009).

Como indicado por Takahashi *et al.* (2002), o oceano Atlântico equatorial é uma das fontes de CO₂ oceânico, junto com o oceano Índico e com a região equatorial do Pacífico. Por ser uma região de ressurgência, águas ricas em CO₂ são trazidas à superfície, alimentando o sistema. O valor do pCO₂ na superfície do oceano, para esta região, foi estimado por Takahashi *et al.* (2002) entre 350µatm e 450µatm.

Entretanto, na plataforma continental não tem sido observada a influência direta da ressurgência das águas de fundo do Atlântico Norte, que são observadas em profundidades muito superiores (>2.000 metros) do que as da plataforma continental desta região. Nesta plataforma as áreas mais profundas, próximo a quebra do talude, atingem profundidades de aproximadamente 100 metros.

A pCO₂ na equação 6 equivale a fugacidade, que difere da pressão parcial devido a interações entre as moléculas de CO₂ na fase gasosa e os demais componentes atmosféricos, por ser o CO₂ um gás não ideal. A diferença é bem pequena (aproximadamente 3µatm quando pCO₂ é 360µatm), os dois, entretanto, podem ser equiparados sem erros graves. Os valores de K₀ equivalem a salinidade (EQUAÇÃO 7).

$$p\text{CO}_2 = [\text{CO}_2]/K_0 \quad (\text{EQUAÇÃO 7})$$

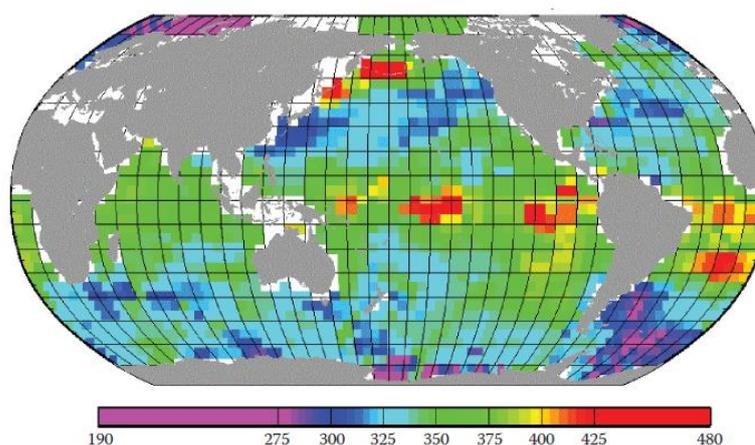
Alguns equipamentos foram desenvolvidos na perspectiva de determinar os dados de pCO₂, através do uso da EQUAÇÃO 7 onde a pressão parcial do CO₂ na água do mar é determinada pelo equilíbrio da amostra com atmosfera de ar ou de nitrogênio. A detecção do CO₂ em amostras individuais pode ser realizada por cromatografia gasosa, ou de forma contínua, realizando a separação do gás da coluna d'água em um sistema de difusão denominado equilibrador, e posterior detecção do gás na região do infravermelho (MILLERO, 2013; CARVALHO, 2015; COTOVICZ, 2015).

O monitoramento utilizando-se amostras individuais não permite a determinação em tempo real, trazendo em si erros amostrais significativos devido a própria mudança das condições originais da amostra causadas pela retirada do meio e estocagem em frascos. Por outro lado, monitoramentos contínuos que permitem medidas em tempo real exigem uma logística avançada de embarque em navios, nem sempre disponíveis, que elevam os custos analíticos.

Determinadas regiões dos oceanos, principalmente as regiões costeiras, apresentam poucos dados referentes a $p\text{CO}_2$. Em parte, isso se deve a grande dificuldade de modelar essas regiões biogeoquimicamente ativas, devido à complexidade de processos que ocorrem nessas áreas, às eventuais alterações ambientais e seus efeitos sobre os ecossistemas costeiros (CHEN; BORGES, 2009) e aos elevados custos da logística necessária a estas medições no oceano.

Os valores de $p\text{CO}_2$ na superfície das águas dos oceanos são apresentados na Figura 8. As águas frias são um sumidouro de CO_2 ($p\text{CO}_2$ inferior a 365ppm), enquanto que a ressurgência de águas são uma fonte de CO_2 ($p\text{CO}_2$ superior a 365ppm) para a atmosfera.

Figura 8 – Distribuição global de $p\text{CO}_2$ (μatm) para as águas superficiais dos oceanos do mundo

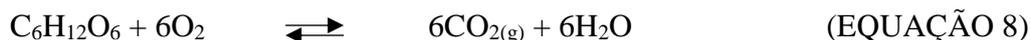


Fonte: Adaptado de Tsukuba, (1999)

Os valores mais elevados de $p\text{CO}_2$ próximo do Equador são o resultado de ressurgência equatorial, entretanto, ressalta-se as observações de Stramma & England (1999) sobre diferenças provocadas em escalas espaciais menores.

Os valores mais baixos de $p\text{CO}_2$ nas águas superficiais das regiões polares fazem com que essas águas capturem continuamente CO_2 . Como citado por Millero (2013), os níveis de $p\text{CO}_2$ e TCO_2 (Dióxido de Carbono Total) em águas superficiais estão relacionados com a troca de CO_2 através da interface atmosfera-oceano dependentes da SSM e da TSM.

A maior incidência de radiação solar próximo a linha do Equador estimula o aumento do metabolismo energético de organismos aeróbicos/fotossintetizantes no oceano, o que por sua vez determina uma maior produção de CO_2 como produto dessa reação (EQUAÇÃO 8). Dessa forma, a troca lenta de CO_2 na interface oceano-atmosfera será maior em regiões mais próximas da linha do Equador e menor em águas polares.



*Sentido direito: Respiração / Sentido inverso: Fotossíntese

Uma vez que as concentrações de pCO_2 na atmosfera continuam a aumentar ao longo dos anos, pode-se entender que os processos de absorção ou emissão do CO_2 pelas águas oceânicas tende a ser alterado. Assim, o monitoramento contínuo desses valores precisa ser realizado, bem como o incremento das medidas em regiões costeiras, que também estão sofrendo alterações significativas devido à grande concentração de populações nestas regiões ao longo do mundo, alterando as vazões naturais de rios e de materiais de origem continental, além de contribuir fortemente para alteração do ciclo natural do carbono (TELSZEWSKI et al., 2009).

1.1.5 Estimativas da pCO_2

A distribuição espacial georreferenciada nos oceanos de pCO_2 realizada a partir de medidas contínuas e intensas e, a necessidade da manutenção dos monitoramentos, bem como o conhecimento do sistema carbonato, levaram às pesquisas sobre a possibilidade de estimativas desses valores.

As estimativas, eventualmente, tornam-se/tornar-se-ão necessárias devido, em parte, a grande dificuldade de se explicar regiões consideradas biogeoquimicamente ativas, como as regiões costeiras, devido à heterogeneidade espaço-temporal dos fluxos de carbono e à complexidade de processos que são comuns nessas áreas (CHEN; BORGES, 2009).

Trabalhos científicos baseados em Redes Neurais Artificiais e Regressão Múltipla para a estimativa da pCO_2 são apresentados na literatura como, por exemplo, Landschützer *et al.* (2013); Lefèvre; Watson; Watson *et al.* (2005) e Telszewski *et al.* (2009). Entretanto estas estimativas foram realizadas para oceano aberto e não foram observadas para regiões de plataformas continentais.

2 REFERENCIAL TEÓRICO

2.1 Técnicas computacionais de estimativa

Dentre as várias técnicas de IA, pode-se observar algum destaque na aplicação de paradigmas de aprendizagem de máquina (do inglês *Machine Learning*) os quais têm por objetivo o desenvolvimento de métodos, técnicas e ferramentas aplicados na construção de máquinas inteligentes capazes de realizar tarefas de processamento da informação baseado, principalmente, em bases de dados já existentes e/ou modelos heurísticos.

De acordo com Mitchell (1997), essas máquinas podem ser treinadas sobre determinado conjunto de amostras e adquirir conhecimento de forma automática. As técnicas de aprendizado podem ser classificadas em supervisionadas (mecanismos de treinamento que usam dados de entrada associados aos dados de saída “domínio e contra-domínio” que possam caracterizar uma relação específica) e não-supervisionadas, por reforço ou supervisão fraca (mecanismos de treinamento que não usam conhecimento prévio baseado em dados de saída).

Os algoritmos de aprendizado de máquina são utilizados, principalmente, quando não existem modelos matemáticos bem definidos, ou ainda “procedimentos” ou algoritmos, para resolução de um problema específico ou uma determinada classe de problemas. Dentre os vários paradigmas, pode se citar aqueles inspirados em sistemas biológicos, como, por exemplo, Redes Neurais Artificiais – RNA’s (HAYKIN, 2001) e Algoritmos Genéticos – Ags (GOLDBERG, 1989), bem como modelos mais abstratos como, por exemplo, Support Vector Machines (SVMs), (VAPNIK, 1995), ou ainda aqueles relacionados com aprendizado estruturado como, por exemplo, Sistemas Baseados em Casos (do inglês *Case Based Reasoning*) (KOLODNER, 1993) ou Árvores de Decisão Adaptativas (do inglês *DecisionTree Learning*) (ROKACH, 2008).

Alguns sistemas de previsão estudados e desenvolvidos por pesquisadores de diversas áreas do conhecimento, incluindo a área da Inteligência Computacional/Inteligência Artificial, possui entre seus propósitos o suprimento de meios para a construção dos chamados métodos de apoio à decisão.

Qualquer instrumento que auxilie o homem no processo de tomada de decisão em qualquer área do conhecimento pode ser considerado um método de apoio à decisão. Esses instrumentos, geralmente baseiam-se em programas de computador capazes de aprender a partir da experiência, ou seja, a partir da alimentação de dados que serão “n” vezes lidos na perspectiva de fazer sentido em um conjunto de dados.

Como exemplo dessas técnicas pode-se citar: as Redes Neurais Artificiais (RNA's), os Algoritmos Genéticos, as Árvores de decisão, as Máquinas de Vetores de Suporte, o Sistema Baseado em Casos e as técnicas mais comuns, que são aquelas estudadas pela Estatística Clássica, como, por exemplo, a Regressão Múltipla (HEGER, 2016).

2.2 Regressão Linear Múltipla (RLM)

A Regressão Linear Múltipla é uma coleção de técnicas estatísticas utilizadas para construir modelos que descrevem de maneira razoável relações entre diversas variáveis explicativas de um processo determinado. A diferença entre a regressão linear simples e a múltipla é que na múltipla são tratadas duas ou mais variáveis explicativas.

É conhecida também como uma das formas estatísticas explanatórias causais referentes ao tratamento de séries temporais de dados. Sua base estatística origina-se da regressão linear, que se restringe a duas variáveis e a apenas uma equação funcional do primeiro grau de ajustamento (EQUAÇÃO 9) (KASZNAR; GONÇALVES, 2007):

$$(Y = a + bX) \quad (\text{EQUAÇÃO 9})$$

Onde:

Y: Variável Dependente;

a: Coeficiente fixo, valor-base a partir do qual começa *Y*;

b: Coeficiente atrelado à variável independente;

X: Variável dependente.

A análise de Regressão Múltipla é uma metodologia estatística de previsão de valores de uma ou mais variáveis de *resposta* (dependentes) através de um conjunto de variáveis *explicativas* (independentes). Esta metodologia também pode ser utilizada para a avaliação dos efeitos das variáveis independentes como *previsoras* das variáveis dependentes; isto é, útil para contribuir na obtenção de respostas a perguntas do tipo “Qual é o melhor estimador para ...?”.

Sua aplicação é de extrema importância, pois permite que estimemos o valor de uma variável com base no conjunto de outras variáveis. Quanto mais significativo for o peso de uma variável isolada, ou de um conjunto de variáveis independentes, afirmar-se-á com mais propriedade que alguns fatores afetam mais significativamente o comportamento de uma variável dependente especificamente procurada, em detrimento a outros.

Galton (1885), Goldberger (1964), Neter e Wasserman (1974), Draper e Smith (1981), Seber (1977) e Bowerman e O’Connel (1990), foram os primeiros pesquisadores a utilizarem o termo *regressão* e os mesmos estenderam a sua aplicabilidade e desenvolveram as hipóteses passíveis de regressão múltipla para inúmeras situações diferenciadas.

O formato geral da EQUAÇÃO de Regressão Linear Múltipla pode ser representado conforme a EQUAÇÃO 10:

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n \quad (\text{EQUAÇÃO 10})$$

Onde:

Y : Valor observado para a variável dependente;

a : Constante de regressão/Intercepto do eixo y ;

b_1, b_2, b_n : Coeficientes de regressão/coeficientes angulares atrelados às variáveis independentes – Representam a variação de Y em função da variável independente.

X_1, X_2, X_n : i -ésimo nível das variáveis Independentes X .

A $p\text{CO}_2$ é nossa variável dependente, enquanto que a TSM e SSM foram consideradas variáveis independentes.

Com o surgimento de sistemas operacionais cada vez mais sofisticados e elaborados, programas e/ou softwares como o “*R*”, “*R Studio*”, “*MatLab*”, “*Past*” e “*Excel*” passaram a tratar inúmeros dados de maneira rápida e eficiente.

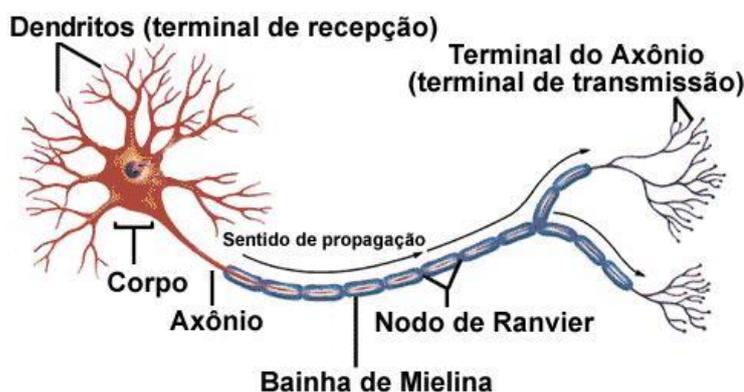
2.3 Redes Neurais Artificiais (RNA)

A observação do mecanismo de funcionamento do cérebro humano e sua comparação com o trabalho dos computadores foi a pedra fundamental para o surgimento das redes neurais artificiais. Pode-se inferir que o cérebro humano é uma estrutura não linear, complexa e paralela para o tratamento da informação/mensagem que armazena conhecimento em suas conexões que ocorrem entre as unidades de processamento (HAYKIN, 2001).

A maioria das RNA’s teve seu desenvolvimento com base na inspiração do Sistema Nervoso o qual é formado por um sistema interconectado de entidades biológicas complexas, os neurônios. Eles têm um papel essencial na determinação do funcionamento, comportamento

e raciocínio dos humanos. O neurônio é constituído por dendritos, que são um conjunto de terminais de entrada, pelo corpo celular, e pelo axônio, ou terminal de saída, conforme pode-se observar na figura 9.

Figura 9 – Estrutura de um Neurônio Celular.



Fonte: <http://www.artecor.com.br/blog/2o-cerebro/>, acessado em 24/12/16.

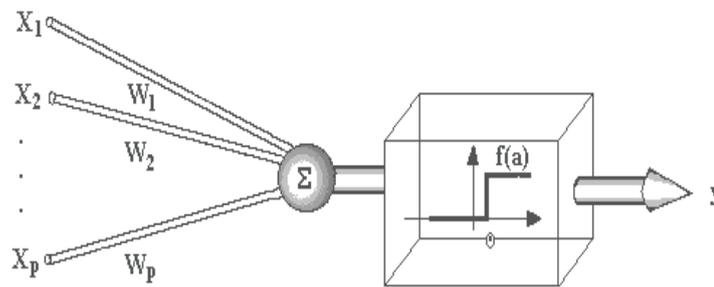
Uma rede neural artificial é composta por várias unidades de processamento, cujo funcionamento é bastante simples (GORNI, 2012). Essas unidades, geralmente são conectadas por canais de comunicação que estão associados a um determinado peso.

As unidades fazem operações apenas sobre seus dados locais, que são entradas recebidas pelas suas conexões de entrada, ou dendritos artificiais. O comportamento cognitivo de uma Rede Neural Artificial vem das interações entre as unidades de processamento da rede.

Segundo Gorni (2012), em 1943, Warren McCulloch e Walter Pitts foram os responsáveis pela criação do neurônio artificial. Posteriormente, em 1949, Donald Hebb, realizou a variação dos pesos de entrada (valores que atuam como fatores de correção) dos nós, conforme as figuras 10 e 11.

Ainda segundo Gorni (2012) o modelo de neurônio artificial proposto por McCulloch e Pitts em 1943, funciona da seguinte forma: os sinais são apresentados à entrada; cada sinal é multiplicado por um parâmetro, ou peso, que indica a sua influência na saída da unidade; É feita a soma ponderada dos sinais que produz um nível de atividade; Se este nível de atividade exceder certo limite (threshold) a unidade produz uma determinada resposta de saída.

Figura 10 – Esquema de neurônio artificial.



Fonte: <http://www.din.uem.br/ia/neurais/>, acessado em 24/12/16.

Onde:

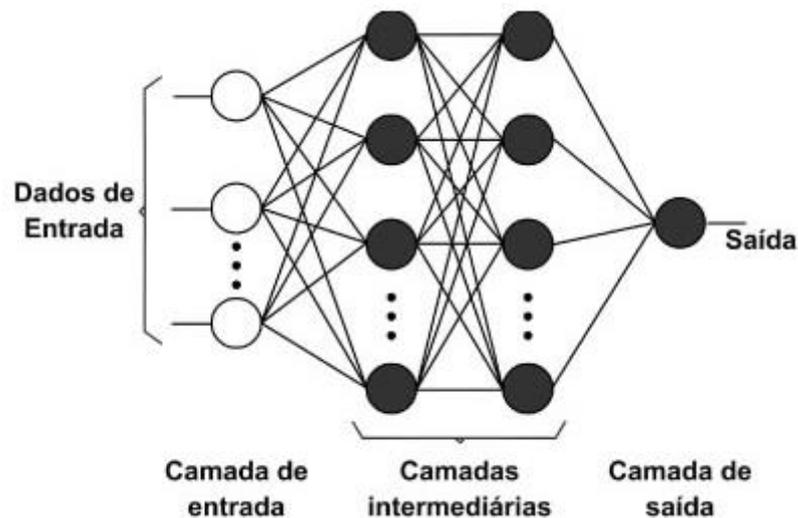
X_1, X_2, \dots, X_p : sinais de entrada: $x \in [-1, 1]$.

W_1, W_2, \dots, W_p : pesos, valores reais: $w \in [-1, 1]$.

t : Limiar

Neste modelo, o nível de atividade a é dado por: $a = W_1X_1 + W_2X_2 + \dots + W_pX_p$, já a saída y é dada por: $y = 1$, se $a \geq t$ ou $y = 0$, se $a < t$.

Figura 11 – Representação gráfica de uma RNA com duas camadas intermediárias.



Fonte: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1806-11172011000100009, acessado em 14/11/16.

A propriedade mais importante das redes neurais é a habilidade de apreender conhecimento sobre um determinado contexto, e, com isso, melhorar seu desempenho, similarmente ao comportamento, simplificado, observado nos neurônios biológicos. Isso é feito através de um processo iterativo de ajustes aplicado aos seus pesos, o treinamento. Portanto, a capacidade da máquina ou do programa em adaptar os parâmetros internos do sistema

conexionista, neurônio ou RNA's, a partir dos dados de treinamento que caracterizam totalmente ou parcialmente a solução do problema.

A partir desses dados de treinamento o aprendizado ocorre quando o sistema conexionista atinge uma solução generalizada para uma classe de problemas. No decorrer do processo de "aprendizagem", os referidos pesos vão se ajustando de forma a atingir o objetivo pretendido (BAPTISTELLA; STEINER; CHAVES NETO, 2006).

Uma RNA reflete um sistema que se baseia em operações de neurônios biológicos e, portanto, pode ser definida como uma emulação simplificada de sistemas neurais biológicos. As RNA'S estão no grupo de sistemas computacionais destinados a produzir, ou pelo menos imitar, o comportamento inteligente.

Considerando o exposto, as RNA's são sistemas paralelos de processamento de informação pertencente à classe de ferramentas computacionais construídas com base na analogia do sistema biológico de processamento de informação, portanto seus elementos estão conectados através de unidades de processamento simples, chamadas neurônios artificiais (SRIVASTAVA; TRIPATHI, 2012).

Essa rede tem propensão natural para armazenar conhecimento experiencial e torná-lo disponível para uso futuro, dessa forma as RNA's podem ser usadas para classificação, reconhecimento de padrões, aproximação de funções e previsão (SRIVASTAVA; TRIPATHI, 2012).

Ao contrário dos sistemas clássicos de Inteligência Artificial (IA), que são concebidos para imitar diretamente o raciocínio lógico e racional, as redes neurais têm como objetivo reproduzir os mecanismos de processamento subjacentes que dão origem a inteligência como uma propriedade emergente de sistemas complexos, adaptáveis (HEGER, 2016).

Antes do advento da RNA e outros modelos de IA, essas tarefas foram realizadas por métodos estatísticos como a regressão linear e não-linear, análise de componentes principais e os modelos de classificação de Baysian (HEGER, 2016). O domínio de aplicação é vasto e inclui diversas áreas como as finanças, vendas, economia, ciência forense e outros campos (SRIVASTAVA; TRIPATHI, 2012).

Sistemas de Redes Neurais foram desenvolvidos e implantados com sucesso na tentativa de solucionar o reconhecimento de padrões, a inteligência de negócios, a robótica ou problemas relacionados aos aspectos intuitivos (SRIVASTAVA; TRIPATHI, 2012).

Na ciência da computação, as redes neurais ganharam força ao longo dos últimos anos em áreas como reconhecimento de padrões, previsão, análise e prospecção de dados (HEGER, 2016). É importante salientar que as análises de dados são normalmente definidas

como a tarefa de examinar os dados brutos com a finalidade de tirar conclusões sobre essa informação. Essa análise é bastante usada em muitas indústrias para permitir às empresas tomarem as melhores decisões em seus negócios, ou na ciência, para verificar/refutar modelos ou teorias existentes (SPECHT, 1991).

Em muitas aplicações as RNA's são usadas como mecanismos de prospecção de dados, ou mineração de dados, os quais implementam o processo de descoberta de novos padrões de grandes conjuntos de dados, os quais, estavam implícitos ou ocultos quando tratados por técnicas da estatística clássica, de outros métodos de inteligência artificial ou da gestão de banco de dados (SPECHT, 1991).

A tarefa de mineração de dados real reflete uma análise automática (ou semi-automática) de grandes quantidades de dados e a meta é extrair padrões antes desconhecidos, tais como grupos de registros de dados (Análise de Cluster), registros incomuns (Detecção de Anomalias), ou dependências. Assim, a mineração de dados se concentra basicamente na triagem através de grandes conjuntos de dados para identificar padrões desconhecidos e estabelecer (explicitar ou ressaltar) relações ocultas (HEGER, 2016).

Para resumir, uma rede neural pode ser descrita como um sistema paralelo que é capaz de resolver problemas que a computação clássica ainda não teve êxito. A seguir, encontra-se um resumo das inúmeras classificações das redes neurais artificiais, proposta por Heger (2016).

2.3.1 Classificação das redes neurais artificiais

2.3.1.1 Rede Neural do tipo Feed forward

É um tipo de rede neural simples onde as conexões (sinapses) são feitas entre uma camada de entrada e uma ou mais camadas escondidas, por sua vez, conectada a uma camada de saída. É uma das redes neurais em uso mais comuns. É apropriada para muitos tipos de aplicações e são frequentemente treinadas através de técnicas de retro propagação do erro, e/ou usando algoritmos genéticos. Para ilustrar, pode-se fazer uma comparação com os ciclos de aquecimento/resfriamento em metalurgia: se um metal é aquecido a uma temperatura muito alta, os átomos se movem em alta velocidade. No entanto, se os átomos forem resfriados lentamente, é propiciado o estabelecimento de novos padrões e estruturas, o que, por sua vez, pode tornar o metal muito mais forte do que antes. Este princípio também pode ser utilizado como uma técnica de otimização em ciência da computação (HEGER, 2016).

2.3.1.2 Mapas auto-organizáveis (SOM)

Também chamados de *Self Organizing Maps* (SOM) ou Mapas de Kohonen em homenagem ao seu projetista ou desenvolvedor, Professor finlandês *Teuvo Kohonen*, os Mapas Auto-Organizados são redes competitivas que possuem a habilidade de formar mapeamentos que preservam a topologia entre os espaços de entrada e de saída. As redes SOM são utilizadas em muitos projetos industriais como ferramentas para resolver problemas práticos de difícil solução (HEGER, 2016).

2.3.1.3 Rede Neural Hopfield

É uma rede neural de camada única e recorrente. A rede neural *Hopfield* é treinada por meio de um algoritmo inspirado na teoria do caos, considerando os atratores, o que possibilita o aprendizado de classes de padrões. Na ativação, a rede de *Hopfield* irá indicar que o padrão é reconhecido ou associado a alguma classe previamente estabelecida durante o processo de aprendizagem. As redes neurais *Hopfield* também podem ser utilizadas para o reconhecimento de padrões. (HEGER, 2016).

2.3.1.4 Rede Neural recorrente simples (SRN) - Elman ou Jordan Style

É uma rede neural realimentada que tem uma camada de contexto. Essa camada de contexto é a saída da interação anterior de volta para a camada de entrada da rede neural. A camada oculta, em seguida, sempre recebe a entrada de saída de seu anterior. SRN são treinadas usando algoritmos genéticos ou uma das técnicas de propagação. São normalmente utilizados para problemas relacionados com previsão. (HEGER, 2016).

2.3.1.5 Rede Neural Feed forward baseada em função de base Radial (RBF)

É uma rede de propagação para frente com uma camada de entrada, uma camada de saída e uma camada oculta. A camada oculta é baseada em funções de base radial-RBF's. Várias RBF's na camada oculta permitem que a rede RBF aproxime uma função de ativação mais complexa quando comparada a capacidade de aproximação de uma única função de base radial. Redes RBF podem ser utilizadas para reconhecimento de padrões. Elas podem ser treinadas usando, por exemplo, algoritmos genéticos ou uma das técnicas de retro propagação

de erros. (HEGER, 2016).

2.3.1.6 Algoritmo de Levenberg-Marquardt (LM)

O algoritmo de Levenberg-Marquardt, desenvolvido de forma independente por Kenneth Levenberg e Donald Marquardt, fornece uma solução numérica para o problema de minimização de uma função não-linear, podendo ser classificado também como um método de otimização. É rápido e tem convergência estável, (YU; WILAMOWSKI, 2010).

2.4 RNA com aplicação do algoritmo de Levenberg-Marquardt (LM)

O método Levenberg-Marquardt é comumente utilizado na visão computacional quando da necessidade de se ajustar um modelo a um conjunto de dados experimentais. O algoritmo de Levenberg-Marquardt (LM) é uma técnica iterativa onde o mínimo de uma função multivariada é expresso como a soma dos quadrados de funções reais não-lineares. Tornou-se uma técnica padrão para problemas não-lineares (Figura 12a) de mínimos quadrados, amplamente adotada em um amplo espectro de experimentos.

O algoritmo de LM pode ser pensado como uma combinação de gradiente descendente com o método de Gauss-Newton. Quando a solução atual está longe da correta, o algoritmo se comporta como um método de gradiente descendente: lento, mas garantidamente convergente (GAVIN, 2017). Quando a solução atual está próxima da solução correta, ela se comporta como o método de Gauss-Newton (GAVIN, 2017). Este método, portanto, é uma derivação do método de Newton que é de convergência mais rápida.

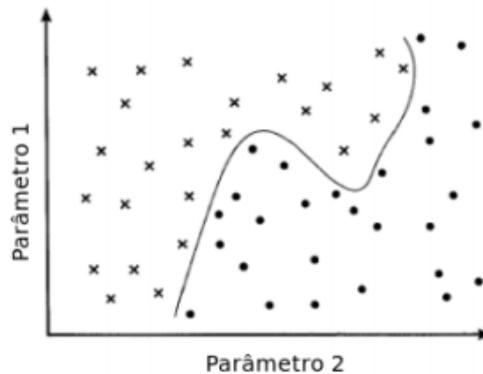
A popularidade do algoritmo Levenberg-Marquardt é tamanha devido a existência de diversas implementações disponíveis. Contudo, apenas uma implementação toma vantagem de uma estrutura, muito comum em problemas da visão computacional, que faz com que a minimização não-linear produza uma matriz Jacobiana esparsa. Se isso for levado em consideração, para a estimação de N parâmetros, a complexidade do algoritmo reduz-se de N^3 para apenas N (FRANÇA, 2009).

Métodos estatísticos mais sofisticados, como as redes neurais artificiais, conseguem encontrar uma relação entre os valores de entrada e saída de modo a se gerarem melhores previsões que os métodos clássicos de regressão linear (GIEBEL *et al.*, 2003). Em diversas áreas/subáreas, o algoritmo L.M. tem-se mostrado eficiente para sistemas de previsão em detrimento de técnicas de estatística clássica.

Um dos problemas mais comuns solucionados por redes neurais é a classificação de dados linearmente separáveis e/ou dados não linearmente separáveis, para cada tipo de dado tem-se um tipo de rede a ser utilizada.

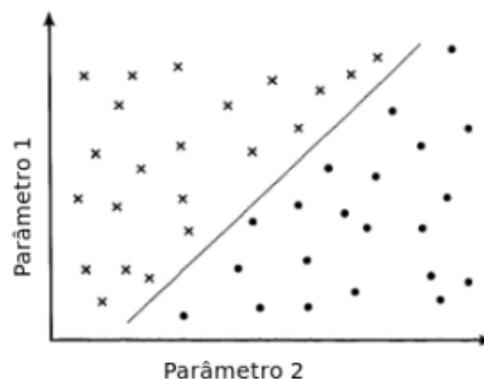
Percebe-se que não é possível separar os dois padrões apenas traçando uma reta (figura 12a). Situação não observada na figura 12b, problema linearmente separável. Isso significa que se objetivamos identificar se aquele objeto é redondo ou quadrado e em algum momento esses padrões apresentam a característica acima, algoritmos que não resolvem problemas desse tipo não serão capazes de distinguir redondo de quadrado.

Figura 12a - Representação gráfica de problema não-linearmente separável.



Fonte: <http://www-users.cs.umn.edu/~andre/arquivos/pdfs/svm.pdf>, acessado em 15/12/16.

Figura 12b - Representação gráfica de problema linearmente separável.



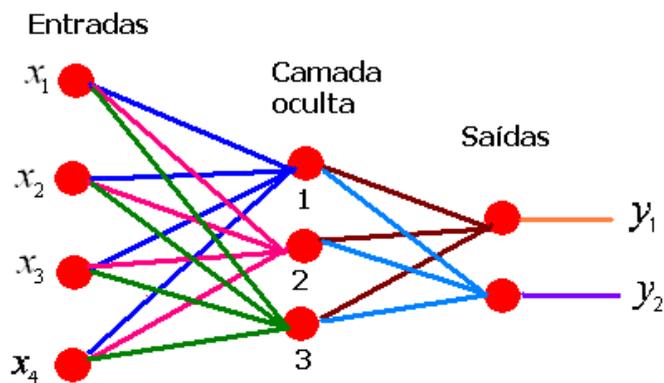
Fonte: <http://www-users.cs.umn.edu/~andre/arquivos/pdfs/svm.pdf>, acessado em 15/12/16.

As Máquinas de Vetor Suporte lineares são eficazes para classificação de um conjunto de dados linearmente separável ou que possuam uma distribuição aproximadamente linear, sendo que a versão de margens suaves tolera a presença de alguns ruídos e *outliers*. Entretanto, há casos em que não é possível dividir satisfatoriamente os dados de treinamento

por um hiperplano. Um exemplo é apresentado na Figura 12a, em que o uso de uma fronteira curva seria mais adequada na separação das classes.

Em qualquer RNA temos os seguintes itens: Neurônios artificiais e respectivos pesos. Os neurônios armazenam os valores dos pesos que serão usados para calcular suas respectivas saídas, baseada em uma determinada entrada de dados. Ou seja, no exemplo da figura 13, a interconexão dos vários neurônios, o treinamento para gerar os pesos apropriados para cada neurônio e, um algoritmo de ativação, funcionam como a “chave” para o funcionamento de toda RNA, permitindo-a diferenciar padrões como, por exemplo, redondos de quadrados.

Figura 13 - Representação gráfica de RNA com seus pesos.



Fonte: <http://www.devmedia.com.br/redes-neurais-artificiais-algoritmo-backpropagation/28559>

Da mesma forma, como ocorre em outros modelos não-lineares, as RNA's podem sofrer de *overfitting*, esta situação ocorre quando um número excessivo de dados é utilizado na camada oculta, figura 13, que levarão a uma perda da capacidade de generalização. Entretanto, se o número de neurônios em excesso for diminuído, ocorre a perda da capacidade de aproximar o processo gerador dos dados (MEDEIROS; PEDREIRA, 2001).

2.5 Regressão Linear Múltipla vs Rede Neural Artificial

RNAs estão sendo cada vez mais usadas para problemas de regressão e classificação não-lineares em meteorologia, devido à sua utilidade na análise de dados e previsão. As redes Neurais Artificiais (RNA) tem sido usadas para estudar vários fenômenos oceânicos e meteorológicos, tais como previsão, tempo de viagem no Oceano Índico, a previsibilidade da temperatura da superfície do mar, chuvas de monção de verão indiano, a classificação do gelo

do mar, a relação entre a pressão do nível do mar do Pacífico tropical e a temperatura da superfície do mar, a análise de componentes principais não-lineares do gelo do mar do Ártico e da Antártida (SRIVASTAVA; TRIPATHI, 2012).

Modelos estatísticos de regressão e Redes Neurais Artificiais têm sido empregados para prever as chuvas de monção de verão indiano usando relações defasadas entre as chuvas e várias combinações de índices do El Niño (SHUKLA *et al.*, 2011 *apud* SRIVASTAVA; TRIPATHI, 2012). Os autores usaram toda a precipitação indiana da temporada de monção para o desenvolvimento de um modelo de RNA e o aplicou para sistemas de previsão/predição.

A previsão de chuvas de monção de verão indiano, na escala de tempo sazonal, foi tentada de várias formas, por vários grupos de pesquisa utilizando diferentes técnicas, incluindo redes neurais artificiais. Ficou estabelecido que a previsão das chuvas em escalas mensais e sazonais não só possui modelagem matemática complexa, como também é importante para o planejamento e elaboração de estratégias agrícolas.

A técnica de RNA tem sido utilizada com sucesso para fornecer uma predição dessas chuvas de monção em escalas de tempo mensais e sazonais (SAHAI *et al.*, 2000 *apud* SRIVASTAVA & TRIPATHI, 2012).

Contudo, mesmo que se demonstre sucesso de aplicação de RNA's em alguns sistemas isolados, como nas chuvas de monção, é válido alertar que a generalização nem sempre é garantida, pois os resultados obtidos com aplicação de técnicas similares podem ser diferentes dos anteriores.

Os modelos descritos na Regressão Linear são frequentemente chamados de modelos empíricos, porque eles são baseados exclusivamente em dados observados. Os parâmetros do modelo normalmente não têm nenhuma relação com qualquer mecanismo de produção dos dados.

Para aumentar a precisão de um modelo linear dentro da gama de observações, o número de termos é simplesmente aumentado. Modelos não-lineares, por outro lado, envolvem tipicamente parâmetros com interpretações físicas específicas. (SAHAI *et al.*, 2000 *apud* SRIVASTAVA & TRIPATHI, 2012). Embora eles exijam, *a priori*, suposições sobre o processo de produção de dados, eles são muitas vezes mais parcimoniosos do que os modelos lineares, e mais preciso fora do intervalo de dados.

2.6 O uso de técnicas de Inteligência Artificial para estimativa da pCO₂

Landschützer *et al.* (2013) a fim de investigar a variabilidade da pCO₂ na superfície

do mar bem como o fluxo de carbono, desenvolveu uma abordagem baseada em rede neural. O método foi capaz de captar uma grande quantidade de variabilidade devido à relação não-linear de predição-observação em uma grade espacial de 1×1 .

O método desenvolvido pelos pesquisadores permitiu determinar as relações não-lineares entre as observações da $p\text{CO}_2$ na superfície do oceano e um conjunto de observações independentes para produzir mapas da $p\text{CO}_2$ na superfície do mar de toda a bacia mensalmente.

Telszewski *et al.* (2009) construíram mapas de periodicidade mensal de $p\text{CO}_2$ em uma bacia localizada a 10.5°N para 75.5°S e 9.5°E para 75.5°W , para três anos consecutivos. Eles utilizaram 137.000 medições de $p\text{CO}_2$ recolhidos *in situ* no Atlântico Norte nos anos de 2004, 2005 e 2006 como parte de um projeto integrado com o CarboOcean (<http://www.carboocean.org/>), financiado por um projeto da União Europeia e projetos paralelos dos EUA.

Os mapas foram calculados utilizando uma técnica de rede neural que reconstrói relações não-lineares entre três parâmetros biogeoquímicos (concentração de clorofila *a*, temperatura da superfície do mar e profundidade da camada de mistura) e a $p\text{CO}_2$ marinha. Um mapa de auto-organização (SOM) de rede neural foi treinado usando 389.000 dados.

Os pesquisadores mostraram a capacidade do método de sintetizar coerentes padrões de distribuição espacial e temporal de $p\text{CO}_2$ no Atlântico Norte, e propuseram o método para ser utilizado em conjunto com a coleta de dados *in situ* durante programas de monitorização da $p\text{CO}_2$ oceânica futuras.

É de grande importância a necessidade de se reconhecer e ser capaz de especificar o fluxo de CO_2 ar-mar em grandes áreas do oceano, baseados em épocas ou mensalmente (LEFÈVRE *et al.*, 2005). Tais estimativas são importantes para a validação de modelos de identificação de carbono no oceano.

Para alcançar estimativas precisas de $p\text{CO}_2$ sobre regiões oceânicas é necessária uma técnica para interpolar as medições relativamente escassas tanto no tempo quanto no espaço. Lefèvre e Colaboradores em 2005 apresentaram uma comparação de mapas de $p\text{CO}_2$ superficial mensal para o círculo subpolar (restrito a $50\text{-}70^\circ\text{N}$, $60\text{-}10^\circ\text{W}$) entre 1995 e 1997, obtidos por duas técnicas.

A primeira técnica utilizou regressão linear múltipla (RLM) de $p\text{CO}_2$ superficial sobre SST e posição, separando regressões para cada mês; a segunda técnica é uma rede neural auto-organizável tendo por base um algoritmo de rede. Ambas usaram como entrada um conjunto de dados de $p\text{CO}_2$, temperatura da superfície do mar (TSM), tempo e posição.

Bostock; Mikaloff Fletcher e Williams (2013), desenvolveram um algoritmo RLM

para estimar carbonato em águas intermediárias e profundas do Hemisfério Sul, usando dados hidrográficos (temperatura, salinidade e oxigênio dissolvido) e os dados de carbonato (alcalinidade e DIC – Carbono Inorgânico Dissolvido) recolhidos durante o Experimento de Circulação no Mundo Oceânico - WISE e campanhas do Programa Hidrográfico CLIVAR/CO₂ a partir do Oceano Antártico.

É possível realizar medições através de observações temporais por meio de instrumentos autônomos ancorados na base de navios onde um sistema complementar coleta amostras durante uma escala espacial ampla. Quando o transecto é longo, podem ocorrer falhas no processo de medição. Assim, a aplicação de técnicas de estimativa, RLM e RNA, no intuito de se estimar as medidas não realizadas, exigiriam menos do equipamento, evitando perda de resultados. Exemplo desta aplicação foi a aplicação de regressão linear para determinar a relação entre o fluorômetro utilizado na pesquisa e a taxa de clorofila para efeitos de calibração (SHADWICK *et al.*, 2011).

As redes neurais são, muitas vezes, bem adequadas para tarefas generalizadas. A vantagem de uma abordagem via rede neural é que ela pode reconhecer e explorar relações entre os dados que não são predefinidos (como em técnicas de regressão) e não precisam ser expressos por uma equação. Isto torna as redes neurais particularmente adequadas às relações de mapeamento que são empíricas e não-lineares, desde que sejam fornecidos dados suficientes para "treinar" a rede (LEFÈVRE *et al.* 2005).

No mesmo trabalho, os autores perceberam que, utilizando o mesmo subconjunto de dados, 69.611 no total, a rede neural determinou uma previsão de 66.182 dados restantes da pCO₂ com uma precisão melhor do que as regressões. O conjunto de dados de trabalho foi escolhido por amostragem aleatória e posteriormente divididos em dois conjuntos mais ou menos iguais.

Os coeficientes das regressões não foram muito diferentes dos obtidos no conjunto de dados total. Usando as novas equações de regressão, a pCO₂ foi calculada utilizando-se dados de temperatura, ano e localização dos dados restantes. A temperatura, o mês e a localização destes dados foram apresentados à rede neural para determinar a pCO₂. O desvio padrão residual (RSD) foi calculado para ambas as técnicas para estimar quão bem eles previram a pCO₂ dos dados restantes. Percebe-se ainda que as regressões são muito mais dependentes dos dados do que a rede neural devido a capacidade desta de auto ajustar-se durante o processo de treinamento, (LEFÈVRE *et al.* 2005).

3 HIPÓTESE

Estudos recentes indicam uma diminuição da eficiência dos oceanos em absorver CO_2 em várias regiões do planeta. A identificação da variabilidade do comportamento das áreas costeiras torna-se importante tendo em vista que os aportes continentais, as alterações do clima, as atividades antrópicas e os fenômenos oceânicos, que já espelham as mudanças do clima, possibilitariam mudanças nas características de absorção ou retenção do CO_2 pelas águas oceânicas.

A difusão do CO_2 ocorre em toda a interface atmosfera-oceano e participa dos processos complexos do equilíbrio carbonático. Esses processos dependem das taxas de circulação do CO_2 em toda a interface, desde a superfície até águas profundas, e podem variar também de acordo com a latitude, com o tempo, a estação do ano, além dos processos biológicos dependentes da disponibilidade de nutrientes.

O desenvolvimento de métodos analíticos padronizados para medições dos componentes principais do sistema carbono no oceano como carbono inorgânico dissolvido (DIC), alcalinidade total (AT), pH e pressão parcial do dióxido de carbono (pCO_2) ou fugacidade de CO_2 ($f\text{CO}_2$) tem contribuído grandemente para o desenvolvimento de estudos nesta área. Entretanto, cruzeiros oceanográficos demandam custos elevados e extenso trabalho laboratorial.

Desta forma, técnicas de inteligência computacional que permitem a estimativa da pCO_2 a partir de um menor número de variáveis de fácil mensuração facilitam a obtenção de dados em relação aos monitoramentos realizados em missões oceanográficas. Porém, modelos matemáticos para estas estimativas não são bem definidos, principalmente para as regiões da plataforma continental, tendo em vista que estas regiões do oceano são as que possuem maior variabilidade de parâmetros do ciclo do carbono nos oceanos devido à proximidade da costa e das interferências naturais associadas.

Assim, neste estudo, os algoritmos de aprendizado de máquina serão utilizados, para resolução da estimativa de pCO_2 na costa nordeste equatorial do Brasil, utilizando técnicas estatísticas e/ou de inteligência artificial e os parâmetros utilizados em oceano aberto para averiguação de sua aplicabilidade às regiões costeiras.

4 OBJETIVOS

4.1 Objetivo Geral

Este trabalho teve como principal objetivo relacionar variáveis disponíveis no banco de dados do Grupo de Pesquisa Biogeoquímica Costeira (cadastrado no CNPq) para estimar o comportamento do CO₂ em regiões costeiras tropicais utilizando-se para treinamento em técnicas computacionais os dados do navio MS Monte Olivia (2008 a 2012), certificados em bancos de dados internacionais.

4.2 Objetivos Específicos

- a) Utilizar o volume de dados disponíveis para obter prognósticos da pressão parcial do CO₂;
- b) Relacionar as variáveis salinidade e temperatura com a pressão parcial de CO₂;
- c) Comparar Técnicas Computacionais de Estimativa de pCO₂ no oceano;
- d) Investigar as variáveis de maior influência na pressão parcial do CO₂, através de métodos computacionais;
- e) Comparar estimativas da pCO₂ da Plataforma Continental com a pCO₂ do oceano aberto.
- f) Apresentar uma análise crítica dos resultados de estimativa via regressão por técnicas clássicas da estatística e por redes neurais.

5 MATERIAL E MÉTODOS

5.1 Área de estudo

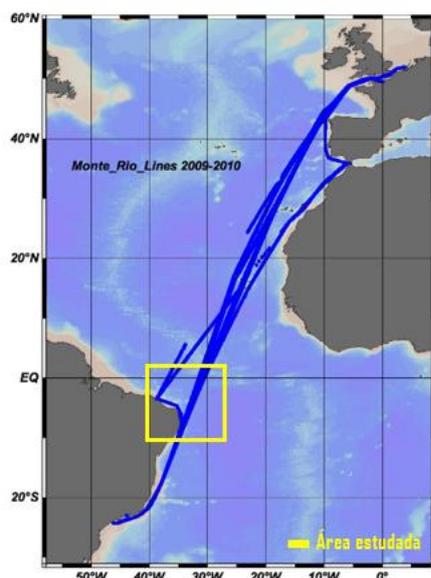
Inicialmente, foram trabalhados os dados de pCO₂, TSM e SSM da água do mar obtidos a bordo do navio mercante Monte Oliva (figura 14) usando 9 viagens realizadas entre 2008 e 2009 na região compreendida entre 1°S e 7°S. Esses dados coletados foram utilizados para treinamento nas técnicas computacionais. A rota dos navios está apresentada na Figura 15

Figura 14 – Navio Mercante Monte Oliva



Fonte: Adaptado de CDIAC http://cdiac.ornl.gov/oceans/VOSProgram/Monte_Olivia.jpg, acessado em 12/01/17.

Figura 15: Rota do Navio MS Monte Oliva. Dados utilizados neste estudo referem-se a área delimitada em amarelo, entre 1°S a 7°S.



Fonte: Adaptado de CDIAC http://cdiac.ornl.gov/oceans/VOS_Program/Monte_Rio.jpg, acessado em 12/01/17.

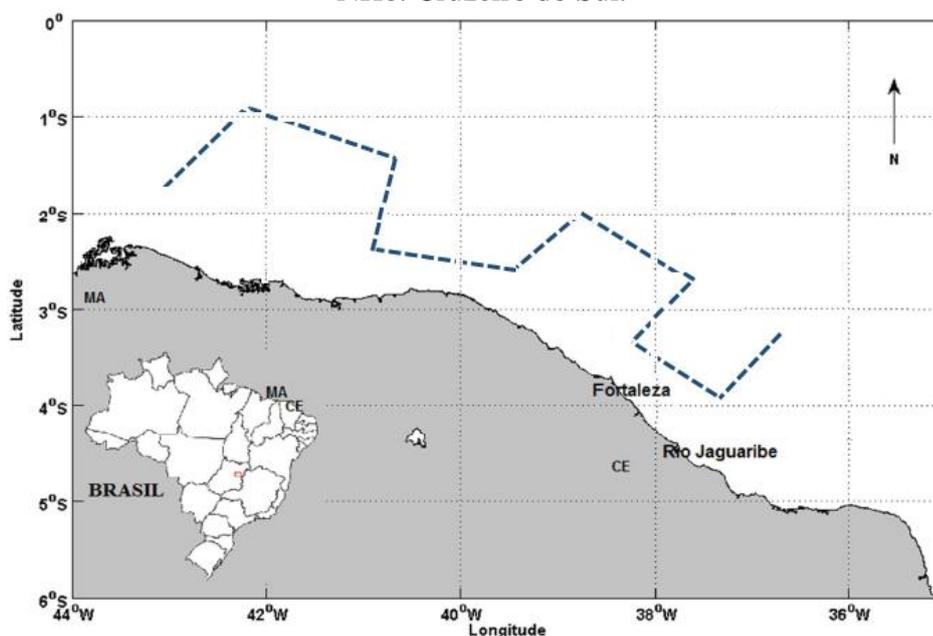
Os dados obtidos pelo Grupo de Pesquisa Biogeoquímica Costeira na plataforma continental da região nordeste equatorial constituíram parcela do Atlântico Equatorial, na área compreendida entre 1°S a 7°S, e entre 43°O a 35°O abrangendo as plataformas continentais do Rio Grande do Norte, do Ceará e Maranhão, através de uma missão oceanográfica, Cruzeiro do Sul do INCT-TMCOcean (Projeto CNPq 500573601/2008-9)

A malha amostral apresentada na Figura 16 representa o percurso do navio durante a missão oceanográfica realizada em setembro de 2014 a bordo do Navio Hidro Oceanográfico Cruzeiro do Sul (H-38S) figura 17.

A malha amostral contemplou a Plataforma Continental do Ceará nas profundidades das isóbatas de 10 a 100m, porção mais estreita em relação às demais margens continentais brasileiras, objetivando assim, observar possíveis variabilidades entre as plataformas interna e externa, além da possível influência de plumas de rios e descargas continentais.

Entretanto, deve-se ressaltar que as medidas foram obtidas em fluxo contínuo durante todo o trajeto dos navios na presença de uma equipe de pesquisadores a bordo, monitorando 24 horas o funcionamento do equipamento, assim como, realizando as demais atividades de coleta e análises de parâmetros necessários para obtenção de dados deste estudo; tal equipamento fora construído no Labomar, em parceria com o INPE/Rede Clima, também parte do INCT-TMCOcean.

Figura 16: Mapa de pontos da malha amostral da Campanha de Setembro de 2014 a bordo do NHo. Cruzeiro do Sul.



Fonte: Adaptado de Carvalho (2015)

Figura 17 – Navio Hidro Oceanográfico Cruzeiro do Sul (H-38S)



Fonte:

[https://www.google.com.br/search?q=Navio+Hidro+Oceanogr%C3%A1fico+Cruzeiro+do+Sul+\(H-38S\)&source=lnms&tbn=isch&sa=X&ved=0ahUKEwjtg8rK9oDTAhWJEZAKHZqXB8MQ_AUICCGB&biw=1366&bih=672#imgrc=rfsAxU6SQwehqM](https://www.google.com.br/search?q=Navio+Hidro+Oceanogr%C3%A1fico+Cruzeiro+do+Sul+(H-38S)&source=lnms&tbn=isch&sa=X&ved=0ahUKEwjtg8rK9oDTAhWJEZAKHZqXB8MQ_AUICCGB&biw=1366&bih=672#imgrc=rfsAxU6SQwehqM): acessado em 12/01/17.

A água do mar chegava ao equipamento a partir de bombeamento em um fluxo aproximado de 2L por minuto, diretamente para o equipamento. A qualidade da medição em fluxo contínuo foi garantida mediante a manutenção da calibração do equipamento, realizada a cada ciclo de 6 horas de leituras contínuas de água do mar durante todo o período amostral.

5.2 Treinamento e Aplicação de dados da campanha do Navio Mercante Monte Olívia em Técnicas Computacionais – Regressão Linear Múltipla

Os dados da campanha oceanográfica costeira disponíveis (Navio Mercante Monte Olívia), totalizando 3083 dados, pertencentes a uma base de dados robusta referendada mundialmente (CDIAC, 2017) foram utilizados para o treinamento inicial nas ferramentas de inteligência artificial. Estes dados foram tabelados para as variáveis independentes (temperatura e salinidade) e dependente ($p\text{CO}_2$) e, também, foi considerada a distribuição espacial destas variáveis.

Em seguida, esses dados foram organizados de acordo com o seu tipo: medidas da atmosfera, medidas da temperatura e salinidade da água do mar, medidas dos gases padrões. Posteriormente, estes dados foram tratados utilizando o software Microsoft® Excel® 2013 (Chave do Produto: 15.0.4885.1000) para organizá-los em colunas referentes a $p\text{CO}_2$,

temperatura e salinidade, para, em seguida, ser aplicada a ferramenta para análise de dados, a regressão múltipla.

A regressão múltipla encontrada inicialmente, refere-se à utilização das medidas de temperatura e salinidade dos dados totais como variáveis independentes e a $p\text{CO}_2$ como variável dependente. Em seguida, os dados de temperatura e salinidade foram substituídos na equação de regressão a fim de se obter a estimativa dos dados totais de $p\text{CO}_2$.

Após a elaboração da planilha do software Microsoft® Excel® 2013. Uma quarta coluna foi criada e em seguida, suas células foram nomeadas intercaladamente por A/B/A/B/A/B até o dado de número 3083.

A separação em pares e ímpares juntamente com a técnica de espaçamento foi utilizada como uma alternativa à retirada aleatória para permitir que futuros trabalhos possam comparar os resultados (e as feições de gráficos intermediários) numa forma padronizada (e sem tendências, pois mesmo numa escolha aleatória, não há como garantir que não houveram repetições). Por outro lado, ambos os métodos (Rede Neural e Regressão clássica) seguiram o mesmo procedimento. Portanto, não houve qualquer tendência sobre o conjunto de treinamento e o conjunto de validação, uma vez que o espaçamento em conjuntos pares e ímpares não determinam *a priori*, quais valores serão usados.

As quatro colunas foram selecionadas com todos os dados e foi ativado o comando “Classificar de A a Z”. Assim os dados ímpares (A) foram separados dos dados pares (B) e plotados conforme exemplificado na Tabela 1a, 1b e 1c.

Em seguida, foram construídos gráficos da distribuição espacial da $p\text{CO}_2$ dos dados ímpares, pares e dos dados totais para verificação da variabilidade dos resultados após este isolamento.

Tabela 1a – Exemplo da primeira etapa de separação de dados em planilha Excel® para isolamento de dados denominados ímpares e pares para uso estatístico.

	A	B	C	D
1	Temperatura	Salinidade	PCO2	Separação
2	27,33	35,544	383,1	A
3	27,33	35,544	383,1	B
4	27,33	35,544	383,1	A
5	27,33	35,544	383,1	B
6	27,33	35,544	383,1	A
7	27,33	35,544	383,1	B
8	27,33	35,544	383,1	A
9	27,33	35,544	383,1	B
10	27,33	35,544	383,1	A
11	27,33	35,544	383,1	B
12	27,33	35,544	383,1	A
13	27,33	35,544	383,1	B
14	27,33	35,544	383,1	A
15	27,33	35,544	383,1	B
16	27,33	35,544	383,1	A
17	27,33	35,544	383,1	B
18	27,33	35,544	383,1	A
19	27,33	35,544	383,1	B
20	27,33	35,544	383,1	A
21	27,33	35,544	383,1	B
22	27,33	35,547	382,4	B

Fonte: Elaborado pelo Autor.

Tabela 1b – Exemplo da segunda etapa de separação de dados em planilha Excel® para isolamento de dados denominados ímpares (A) para uso estatístico.

1	Temperatura	Salinidade	PCO2	Separação
2	27,33	35,544	383,1	A
3	27,33	35,544	383,1	A
4	27,33	35,544	383,1	A
5	27,33	35,544	383,1	A
6	27,33	35,544	383,1	A
7	27,33	35,544	383,1	A
8	27,33	35,544	383,1	A
9	27,33	35,544	383,1	A
10	27,33	35,544	383,1	A
11	27,33	35,544	383,1	A
12	27,33	35,547	382,4	A
13	27,34	35,544	382,3	A
14	27,34	35,546	382,2	A
15	27,36	35,544	382,3	A
16	27,36	35,545	382,2	A
17	27,36	35,546	382	A
18	27,36	35,546	381,9	A
19	27,38	35,552	382,8	A
20	27,38	35,555	382,8	A
21	27,38	35,556	382,8	A
22	27,38	35,555	382,9	A

Fonte: Elaborado pelo Autor.

Tabela 1c – Exemplo da segunda etapa de separação de dados em planilha Excel® para isolamento de dados denominados pares (B) para uso estatístico.

1527	28,72	35,106	392	A
1528	28,76	35,142	395,7	A
1529	28,69	35,179	394,7	A
1530	28,6	35,193	393,7	A
1531	28,66	35,237	394,7	A
1532	28,69	35,229	396	A
1533	28,73	35,247	397,9	A
1534	28,74	35,319	399,1	A
1535	28,48	35,365	395,2	A
1536	28,41	35,353	392,8	A
1537	28,51	35,354	395,1	A
1538	27,33	35,544	383,1	B
1539	27,33	35,544	383,1	B
1540	27,33	35,544	383,1	B
1541	27,33	35,544	383,1	B
1542	27,33	35,544	383,1	B
1543	27,33	35,544	383,1	B
1544	27,33	35,544	383,1	B
1545	27,33	35,544	383,1	B
1546	27,33	35,544	383,1	B
1547	27,33	35,544	383,1	B
1548	27,33	35,544	382,3	B

Fonte: Elaborado pelo Autor.

Outrossim, os dados totais foram isolados em outros dois conjuntos de dados, dessa vez, os mesmos, ordenados de 1 a 3083 foram salteados (espaçados) em dois: dados 1, dado 4, dado 7 e assim sucessivamente. Essa mesma metodologia foi utilizada salteando os dados em 5, 15, 30 e 50 dados. Ao término de cada separação em dois conjuntos de dados, foi encontrada a equação de regressão.

Para compatibilizar a regressão linear múltipla e rede neural, todos os dados utilizados foram previamente normalizados, no Excel®, através da aplicação da equação 11, incluindo as estimativas obtidas a partir da aplicação das duas técnicas.

$$VN = \frac{(2V_{med} - (V_{mín\ med})) - (V_{máx\ med})}{(V_{máx\ med} - V_{mín\ med})} \quad \text{Equação 11}$$

Onde:

VN = Valor Normalizado

V_{med} = Valor medido

$V_{mín\ med}$ = Valor mínimo medido

$V_{máx\ med}$ = Valor máximo medido

A fim de comparar o valor estimado com o valor real de pCO₂, foi aplicado o Desvio Padrão Residual (RSD), comumente utilizado nessas abordagens, especialmente na comparação entre técnicas de Inteligência computacional (LEFÈVRE; WATSON; WATSON *et al.*, 2005), conforme representado pela equação 12.

$$\text{RSD} = \sqrt{\frac{\sum i (y_i - x_i)^2}{D - 2}} \quad \text{Equação 12}$$

Onde:

RSD = Desvio Padrão Residual

y_i = valor estimado

x_i = valor medido

D = número de dados

Em virtude do grande volume de dados, foi possível encontrar relações entre os três parâmetros medidos: pCO₂, temperatura e salinidade para o desenvolvimento de um algoritmo regional (Alg_{Reg}).

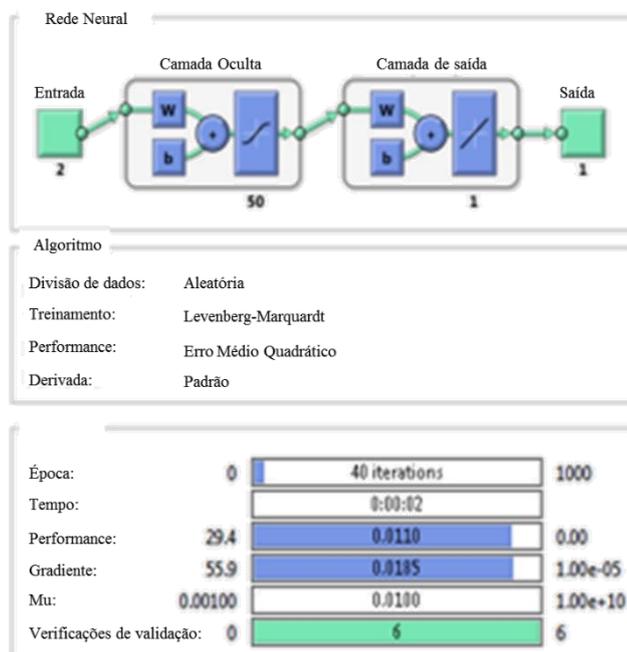
Foi calculado também o coeficiente de determinação, R², interpretado como uma medida de ajustamento de um modelo estatístico generalizado, como a regressão linear múltipla, em relação aos valores observados.

O R² varia entre 0 e 1, inferindo quanto o modelo consegue explicar os valores observados. Quanto maior o valor de R², maior a explicação do modelo e melhor ele se ajusta aos dados. Por exemplo, se o R² de um modelo é 0,9337, isto significa que 93,37% da variável dependente consegue ser explicada pelas variáveis independentes presentes no modelo.

5.3 Treinamento e Aplicação de dados da campanha do Navio Mercante Monte Olívia em Técnicas Computacionais – Rede Neural Artificial.

Os dados do navio Monte Oliva, divididos da mesma forma, descritos anteriormente com a utilização da ferramenta Excel® 2013, foram usados em uma RNA composta por 50 neurônios. Dos dados totais, separados em pares e ímpares, aqueles determinados como ímpares, foram utilizados para treinamento da rede utilizando o algoritmo de Levenberg-Marquardt conforme mostrado na figura 18.

Figura 18 – Representação esquemática do treinamento de dados utilizando o algoritmo de Levenberg-Marquardt.



Fonte: Elaborado pelo Autor

O código da rede neural foi adaptado dos scripts de treinamento e ativação de redes neurais encontrados no ambiente Matlab da Mathworks©.

Em seguida, os dados totais, já separados em outros dois conjuntos de dados, dessa vez, os mesmos, ordenados de 1 a 3083 foram espaçados em 2, 5, 15, 30 e 50 dados, no sentido de simular lacunas de aquisição de dados. Ao término de cada separação em dois conjuntos de dados, os de menor quantidade foram utilizados para treinar a rede e, os demais, utilizados para validá-la.

No ambiente do Matlab, foi criada uma programação para retornar, ao término de cada teste, com vários conjuntos de dados, o RSD, já calculado a partir da normalização dos dados.

O Alg_{Reg} obtido inicialmente, através da regressão linear múltipla, utilizando a pCO₂ como variável dependente e os principais fatores que a controlam, até então, temperatura e salinidade, como as variáveis independentes, foi testado e comparado com os resultados do tratamento dos mesmos dados, utilizando uma rede neural do tipo *feed forward*.

5.4 Treinamento e Aplicação de dados da campanha do Navio Hidro Oceanográfico Cruzeiro do Sul (H-38S) em Técnicas Computacionais – Regressão Linear Múltipla

Os dados da campanha oceanográfica costeira disponíveis (Navio Hidro Oceanográfico Cruzeiro do Sul), totalizando 1059 dados, pertencentes a uma base de dados do Laboratório de Biogeoquímica Costeira foram tratados de modo similar para o consequente treinamento inicial nas ferramentas de inteligência artificial. Estes dados foram tabelados para as variáveis independentes (temperatura e salinidade) e dependente ($p\text{CO}_2$) e, também, foi considerada a distribuição espacial destas variáveis.

Uma vez que os dados brutos foram gerados pelo programa em linguagem Labview, foi realizado o tratamento desses dados para retirar aqueles gerados em momentos em que o equipamento utilizado, durante a coleta de dados, não funcionou em condições ideais/ótimas, de acordo com as observações feitas pelos operadores do sistema.

Em seguida, esses dados foram organizados de acordo com o seu tipo: medidas da atmosfera, medidas da temperatura e salinidade da água do mar, medidas dos gases padrões. Posteriormente, estes dados foram tratados utilizando, mais uma vez, o software Microsoft® Excel® 2013 (Chave do Produto: 15.0.4885.1000) para organizá-los em colunas referentes a $p\text{CO}_2$, temperatura e salinidade, para, em seguida, ser aplicada a ferramenta para análise de dados, a regressão múltipla.

Feito isto, os dados totais foram isolados em outros dois conjuntos de dados, nessa segunda etapa, os mesmos, ordenados de 1 a 1059 foram salteados em 2, 5, 15, 30 e 50 dados. Ao término de cada separação em dois conjuntos de dados, foi encontrada a equação de regressão a partir dos dados em menor quantidade.

Na tentativa de minimizar os erros, todos os dados utilizados foram previamente normalizados (equação 11) e seus RSD's foram calculados (equação 12), relacionando as estimativas com os dados reais.

Para uma segunda aplicação, do conjunto de 1059 dados disponíveis obtivemos 44 pontos que representavam, temperatura, salinidade, clorofila *a* e fugacidade de CO_2 . Desse conjunto, realizou-se ensaios na perspectiva de se encontrar uma melhor relação entre a $p\text{CO}_2$ estimada e a medida, tendo em vista que a plataforma é uma região biogeoquimicamente ativa.

5.5 Treinamento e Aplicação de dados da campanha do Navio Hidro Oceanográfico Cruzeiro do Sul (H-38S) em Técnicas Computacionais – Rede Neural Artificial

Os dados do navio Hidro Oceanográfico Cruzeiro do Sul (H-38S), divididos nos mesmos conjuntos de dados, com a utilização da ferramenta Excel® 2013, foram utilizados para treinamento e ativação de uma RNA composta por 50 neurônios. Dos dados totais, separados em pares e ímpares, aqueles determinados como ímpares, foram utilizados para treinamento da rede utilizando o algoritmo de Levenberg-Marquardt.

Em seguida, os dados totais, já organizados em outros dois conjuntos de dados, dessa vez, os mesmos, ordenados de 1 a 1059 foram salteados em 2, 5, 15, 30 e 50 dados. Ao término de cada separação em dois conjuntos de dados, os de menor quantidade foram utilizados para treinar a rede e, os demais, utilizados para validá-la.

Ao término de cada teste obtivemos o RSD a partir dos valores já normalizados, modelo semelhante aquele descrito para os dados do Monte Olívia assim como a repetição dos os testes envolvendo clorofila *a*.

6. RESULTADOS E DISCUSSÃO

6.1 Treinamento e Aplicação de dados da campanha do Navio Mercante Monte Olívia em Técnicas Computacionais – Regressão Linear Múltipla

A equação da regressão normalizada (Equação 13) obtida a partir dos dados do Navio Mercante Monte Olívia mostrou um coeficiente de determinação (R^2) equivalente a 0,63, isto significa dizer que usando os valores das medições de temperatura e salinidade, a equação de regressão encontrada explica 63% dos dados, além de um erro padrão de 27,31% para as 3083 medições.

$$\text{pCO}_2(\pm 0,27) = -0,088(\pm 0,008) + 0,624(\pm 0,013)\text{TSM} + 0,765(\pm 0,015)\text{SSM}$$

(EQUAÇÃO 13)

No entanto, ainda restam 37% de variação não explicada por essa relação, possivelmente indicando que existem outras variáveis importantes para o estudo da variabilidade da pCO_2 que não foram incluídas. Os dados de temperatura e salinidade foram inseridos na equação 13 a fim de se estimar pCO_2 e, como resultado do teste, obteve-se um RSD (erro entre o valor medido e o valor estimado para um conjunto de dados) equivalente a $0,2753\mu\text{atm}$.

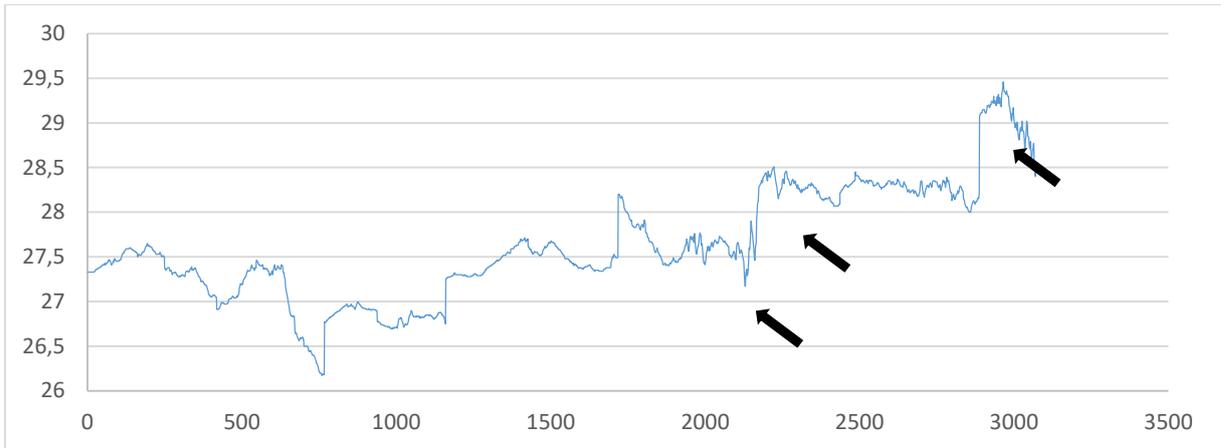
6.1.1 Representação gráfica dos dados utilizados através de técnicas de Estatística Clássica

6.1.1.1 Gráficos dos dados totais por variável

Observa-se, de uma forma geral, que em vários pontos, conforme apresentado nas figuras 19a, 19b e 19c, apontados em preto, a salinidade explica, com razoabilidade, a pCO_2 . Isso é claramente visível próximo ao dado de número 2000. De certa forma, pode-se afirmar que dados mínimos locais são explicados pela salinidade, enquanto que os máximos locais são explicados pela temperatura, conforme verificamos nas marcações indicadas pelas setas da cor verde.

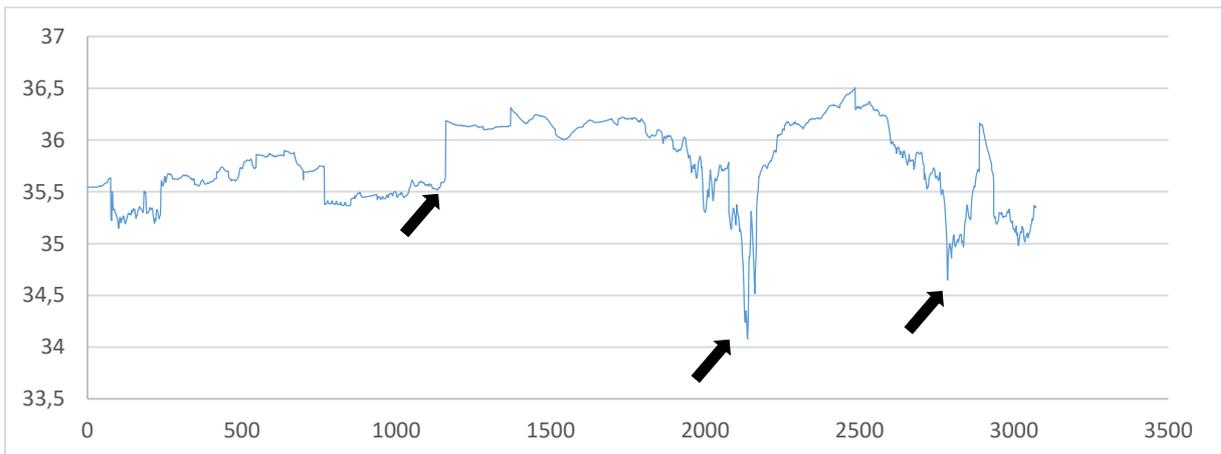
De maneira geral pode-se comparar as três medições na figura 19d, para melhor visualização.

Figura 19a – Dados de Temperatura em função do número da medição – Monte Olívia.

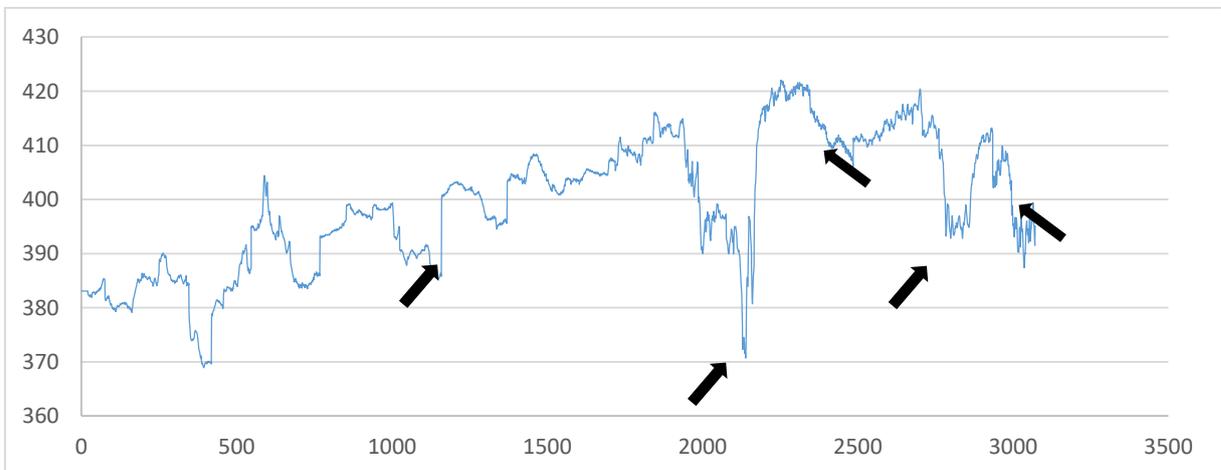


Fonte – Elaborado pelo autor

Figura 19b – Dados de Salinidade em função do número da medição – Monte Olívia.

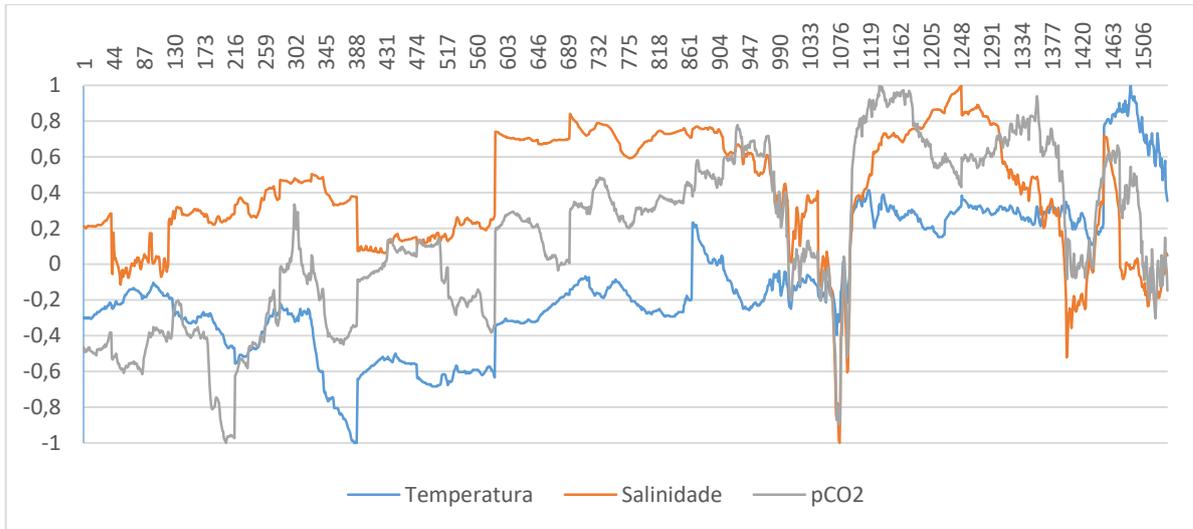


Fonte – Elaborado pelo autor

Figura 19c – Dados de pCO₂ em função do número da medição – Monte Olívia.

Fonte – Elaborado pelo autor

Figura 19d – Dados de temperatura, salinidade e pCO₂ normalizados – Monte Olívia.

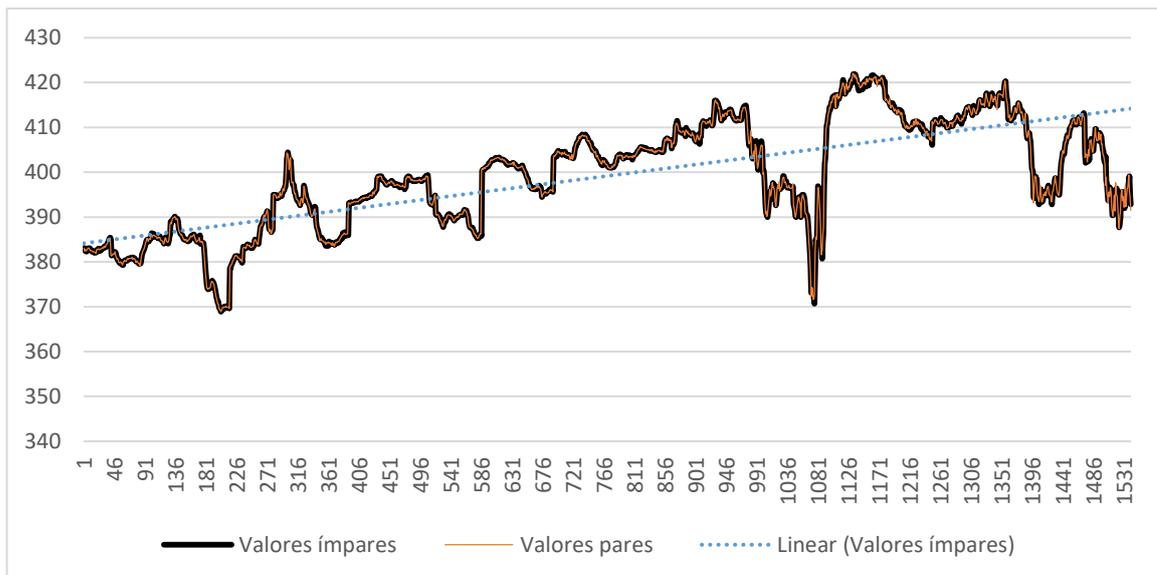


Fonte – Elaborado pelo autor

6.1.1.2 Treinamento com dados ímpares e validação com dados pares.

Com o objetivo de verificar se, a partir de um conjunto de dados, proveniente de pontos de amostragem próximos, consegue-se estimar os demais valores de dados, foi realizado o isolamento dos dados, como detalhado anteriormente e, suas representações encontra-se na figura 20.

Figura 20 – Dados ímpares e pares de pCO₂ normalizados sobrepostos – Monte Olívia.



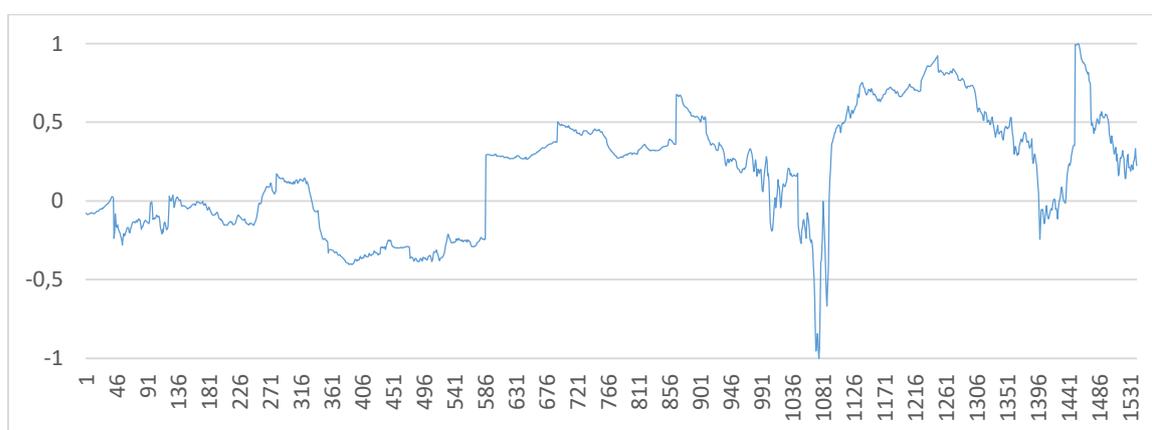
Fonte – Elaborado pelo autor

A observação visual da distribuição espacial dos dados não denotou diferença na distribuição espacial dos dados totais para os dados denominados ímpares e pares. A partir desta conclusão, pode-se utilizar um desses conjuntos de dados para obtenção de uma equação de regressão, para, em seguida aplicar o outro conjunto de dados e verificar se é possível estimar valores de um conjunto de dados a partir do outro.

A equação 14, representa a regressão normalizada encontrada a partir dos dados ímpares e a consequente aplicação dos valores pares, para se encontrar a estimativa dos valores pares, para um n amostral igual a 1541. A representação gráfica da estimativa está demonstrada na figura 21.

Equação normalizada de regressão dos dados ímpares – Monte Olívia
 $pCO_2(\pm 0,27) = -0,089(\pm 0,011) + 0,625(\pm 0,018)TSM + 0,765(\pm 0,022)SSM$
 (EQUAÇÃO 14)

Figura 21- Estimativa de pCO_2 pares normalizados de 1 a -1 via regressão obtida a partir de dados ímpares – Monte Olívia



Fonte: Elaborado pelo autor

O erro padrão da aplicação dos dados pares na equação de regressão proveniente dos dados ímpares foi de 27,29% e o R^2 , permaneceu dentro do esperado (63%), uma vez que o conjunto de dados, após a separação, permaneceu equivalente em relação ao número de medidas. Já o RSD, que representa o desvio padrão residual entre o dado estimado e o dado real foi de $0,2768\mu atm$, bem próximo do RSD total, justificando que a metade dos dados representa muito bem a outra metade.

6.1.1.3 Treinamento e validação com dados salteados.

Com os dados salteados em duas, cinco, quinze, trinta e cinquenta medidas, os erros das regressões, o R^2 e o RSD foram calculados e encontram-se na tabela 2. Mesmo com os dados espaçados, o comportamento da $p\text{CO}_2$, é explicado, na maior parte, pela variação da salinidade. As figuras 22a e 22b, provenientes da aplicação dos dados salteados em dois (de maior quantidade), na regressão obtida a partir dos dados em menor quantidade, mostram uma forte semelhança entre a $p\text{CO}_2$ medida e a $p\text{CO}_2$ estimada. Para fins comparativos representou-se também (Figura 22c) os dados da $p\text{CO}_2$ utilizados para treinamento (obtenção da regressão)

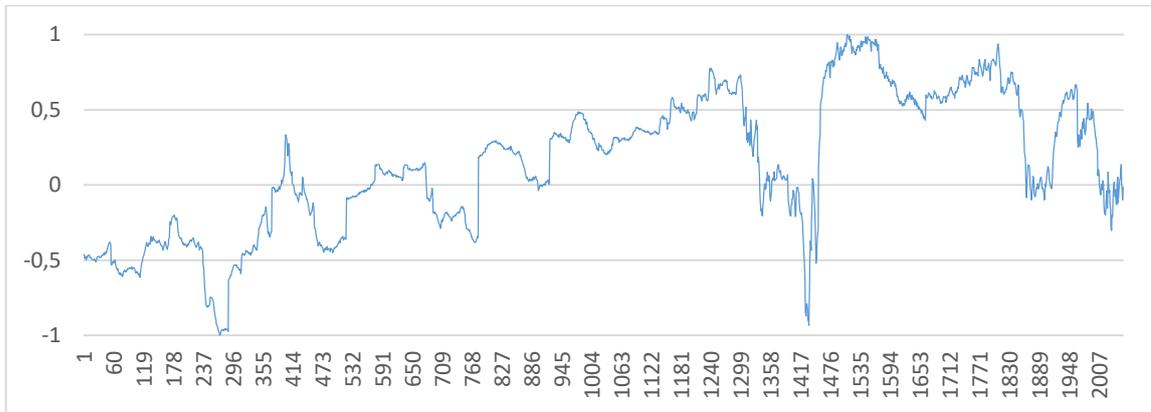
Tabela 2 – Comparações entre erro padrão, coeficiente de determinação e desvio padrão residual – Monte Olívia.

	n amostral	Erro Padrão	R^2	RSD
2 dados salteados	1028	27,29%	63%	0,2782 μatm
5 dados salteados	514	27,26%	63%	0,3093 μatm
15 dados salteados	193	27,28%	63%	0,2785 μatm
30 dados salteados	100	27,65%	63%	0,2761 μatm
50 dados salteados	61	28,38%	63%	0,2754 μatm

Fonte: Elaborada pelo autor

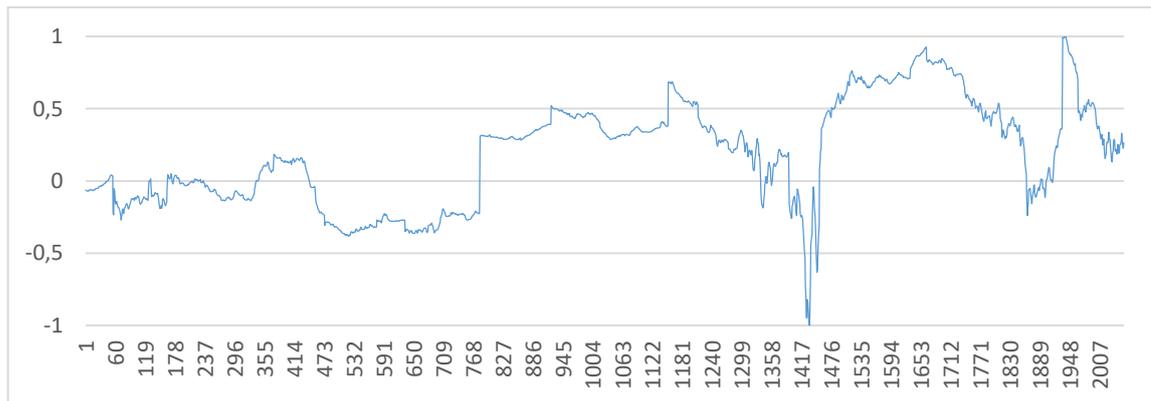
Observou-se que mudanças do número amostral de 1028 para 61 neste conjunto de dados aumentou em 1,07% o erro padrão, o R^2 se manteve o mesmo e igual a 63%, o RSD variou mas tornou-se menor com 61 dados, possivelmente por diminuições das variabilidades dos parâmetros reais que afetam a $p\text{CO}_2$.

Figura 22a – Representação dos dados medidos da pCO₂ da campanha Monte Olívia normalizada entre 1 e -1.



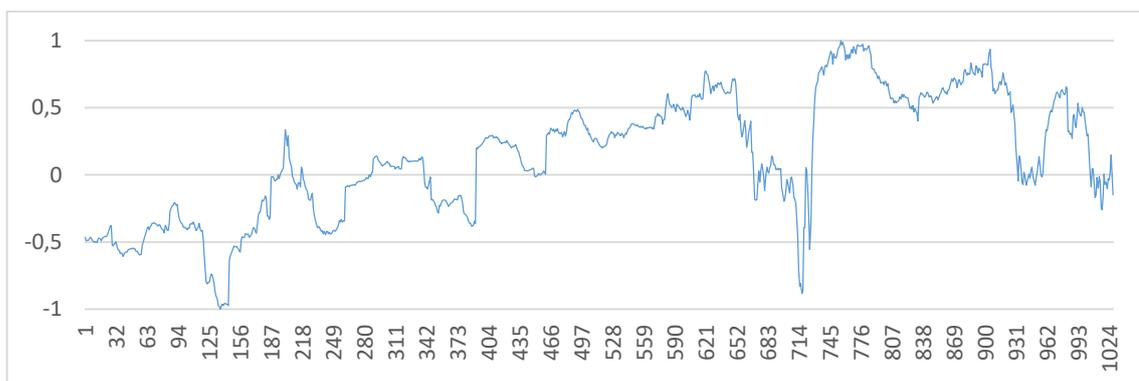
Fonte: Elaborado pelo autor

Figura 22b – Representação da estimativa da pCO₂ da campanha Monte Olívia normalizada entre 1 e -1.



Fonte: Elaborado pelo autor

Figura 22c – Representação dos dados reais da pCO₂ da campanha Monte Olívia normalizada entre 1 e -1 utilizados para treinamento / obtenção da equação de regressão.



Fonte: Elaborado pelo autor

Na região onde foram realizadas as medidas entre os pontos 498 – 853 e 1563 – 1776 (figuras 22a e 22b) tem-se um acompanhamento da $p\text{CO}_2$ pela salinidade (figura 22c), portanto, outras variáveis, diferentes da salinidade, podem estar influenciando a $p\text{CO}_2$ nesse conjunto de dados.

Para avaliar a influência da temperatura no comportamento da $p\text{CO}_2$ os dados do Navio Mercante Monte Olívia foram utilizados para a obtenção de uma equação de regressão utilizando apenas dados de salinidade, como variável independente e a $p\text{CO}_2$, como variável dependente.

A tabela 3 confronta os erros padrão, o R^2 e o RSD entre os dados totais ($n = 3083$) do Monte Olívia aplicando a temperatura e salinidade como variáveis independentes e os mesmos dados aplicando apenas a salinidade como variável independente.

Tabela 3 – Comparações entre erro padrão, coeficiente de determinação e desvio padrão residual utilizando somente a salinidade, somente a temperatura e utilizando salinidade e temperatura.

	Erro Padrão	R^2	RSD
Apenas salinidade	36,11%	35%	0,4436
Apenas Temperatura	36,50%	34%	0,4220
Salinidade e temperatura	27,31%	63%	0,2753

Fonte: Elaborado pelo Autor

A partir dessas informações, concluiu-se que a $p\text{CO}_2$, para este conjunto de dados, é melhor representada pelas relações com temperatura e salinidade, pois assim, obteve-se um menor erro padrão e um menor RSD.

6.2 Treinamento e Aplicação de dados da campanha do Navio Mercante Monte Olívia em Técnicas Computacionais – Rede Neural Artificial.

Após a inserção dos dados no ambiente do MATLAB contendo a linguagem de programação da RNA totalmente desenvolvida, o programa retornou as representações gráficas dos dados reais e da estimativa, bem como o RSD previamente normalizado.

Os RSD's obtidos aqui foram menores, em relação aqueles encontrados através da RLM e, a sensibilidade maior, além disso, observou-se que a rede neural foi mais sensível às

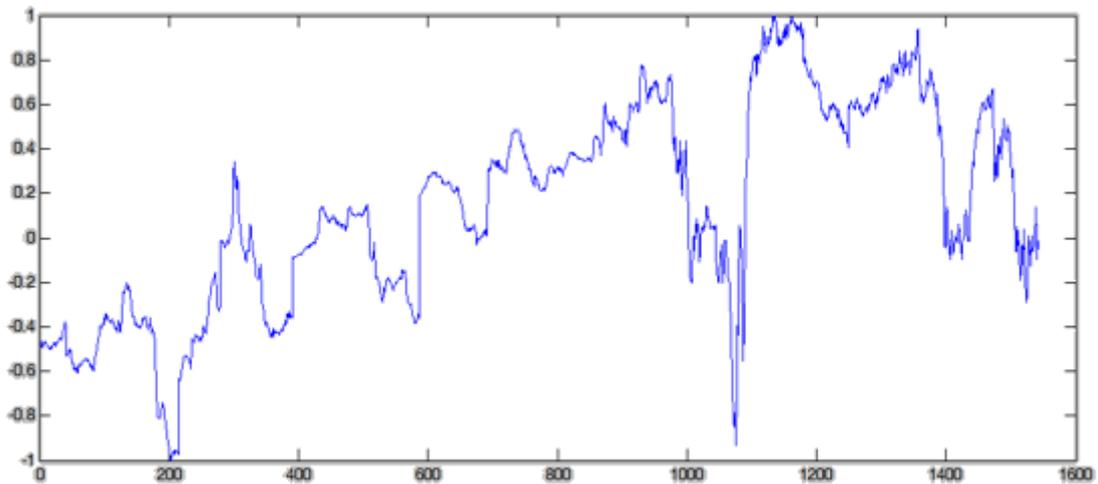
mudanças do n amostral.

6.2.1 Representação gráfica dos dados utilizados através de Rede Neural Artificial.

6.2.1.1 Treinamento com dados ímpares e validação com dados pares.

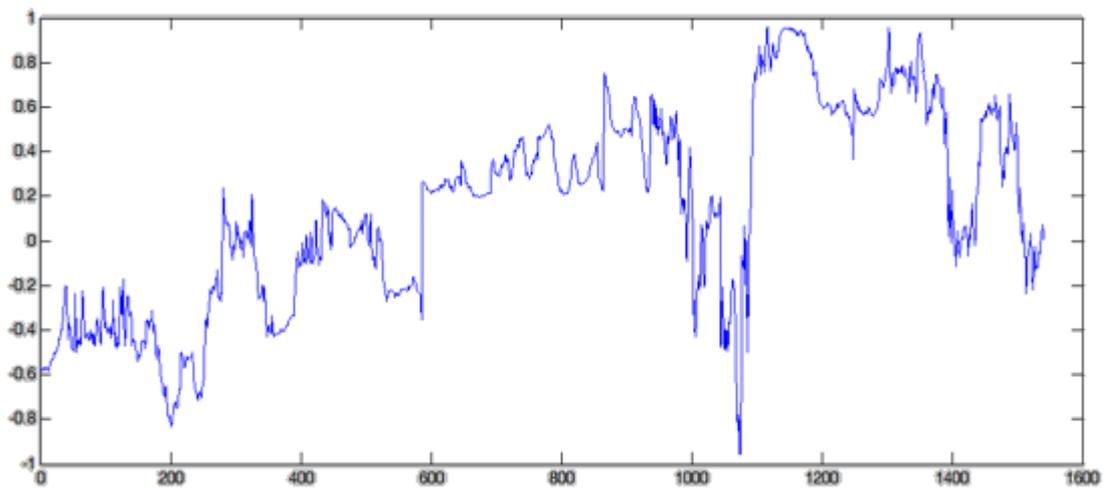
Os gráficos (figura 23a e 23b) referentes aos dados medidos e aos dados estimados apresentaram, visualmente, semelhança muito grande e a estimativa acompanha os dados reais. Obteve-se um RSD de $0,1150\mu\text{atm}$.

Figura 23a – Dados pares da pCO_2 em relação ao número da medição – Monte Olívia



Fonte: Elaborado pelo Autor

Figura 23b: Estimativa da pCO_2 pares normalizados via regressão obtida a partir de dados ímpares – Monte Olívia



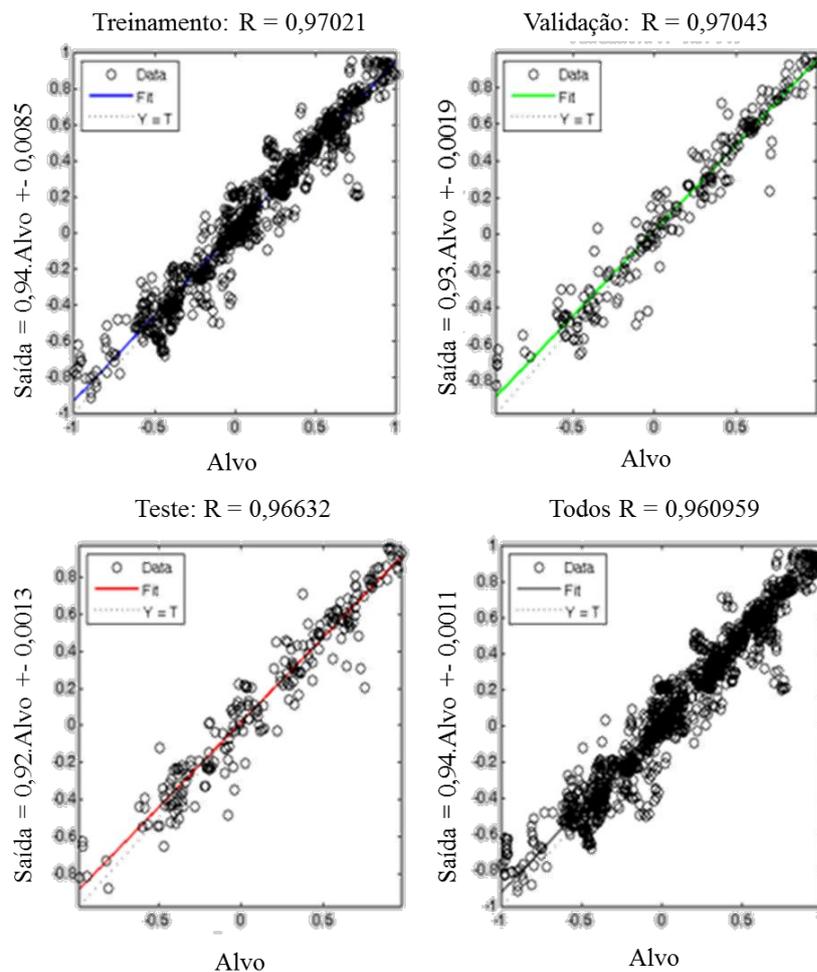
Fonte: Elaborado pelo Autor

A figura 24 contém todas as informações relacionadas ao treinamento em rede. As imagens representam as etapas de treinamento, validação e teste. A linha tracejada em cada gráfico representa o resultado perfeito subtraindo o desvio, indicando assim o alvo. A linha contínua representa a linha de regressão linear mais ajustada entre as saídas e os alvos.

O valor de R é uma indicação da relação entre saídas e os alvos. Se $R = 1$, isso indica que existe uma relação linear exata entre saídas e alvos. Se R é próximo de zero, então não há relação linear entre saídas e alvos.

Para este ensaio, os dados de treinamento indicam um bom ajuste. Os resultados de validação e teste também mostram valores R maiores que 0,9. O gráfico de dispersão é útil para mostrar que determinados pontos de dados têm pouca semelhança.

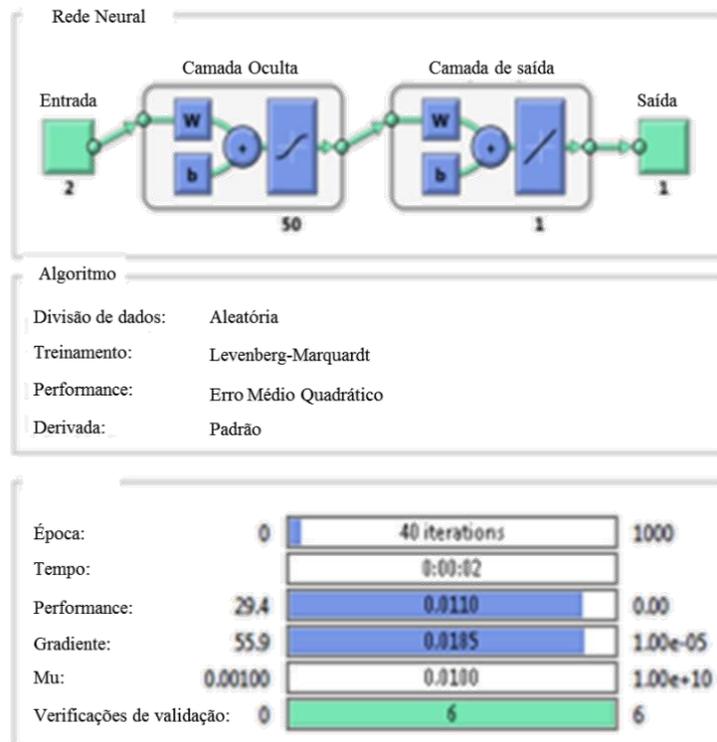
Figura 24 – Representação esquemática do treinamento, validação e teste da rede neural utilizada para estimar a pCO_2 par nos dados da campanha do Navio Monte Olívia.



Fonte: Elaborado pelo Autor

A Figura 25 representa esquematicamente a rede neural utilizada nos dados ímpares do Monte Olívia para treinamento e os dados pares para validá-la.

Figura 25: Representação esquemática do funcionamento da rede neural com suas principais informações.



Fonte: Elaborado pelo Autor

6.2.1.2 Treinamento e validação com dados salteados.

Com os dados salteados em duas, cinco, quinze, trinta e cinquenta medidas, o RSD foi calculado e apresenta-se na tabela 4.

Tabela 4 – RSD entre valores medidos e estimados nos dados salteados.

	n amostral	R²	RSD
2 dados salteados	1028	93,56%	0,1275 μ atm
5 dados salteados	514	89,82%	0,1487 μ atm
15 dados salteados	193	89,87%	0,1614 μ atm
30 dados salteados	100	89,31%	0,3971 μ atm
50 dados salteados	61	92,30%	0,2432 μ atm

Fonte: Elaborado pelo Autor

O esperado seria o RSD aumentar de acordo com a redução dos dados de treinamento, entretanto, eventualmente, a rede neural pode retornar resultados muito próximos a cada treinamento executado, o que justifica, parcialmente as diferenças pequenas entre essas medidas, porém a rede foi mais sensível ao n amostral do que a regressão.

Houve também pequenas variações em relação ao R^2 , entretanto o mesmo manteve-se bem acima daquele que fora encontrado pela RLM, ou seja, a RNA consegue, por meio do algoritmo de Levenberg-Marquardt, retornar um coeficiente de determinação maior do que aquele encontrado via RLM.

6.3 Treinamento e Aplicação de dados da campanha do Navio Hidro Oceanográfico Cruzeiro do Sul em Técnicas Computacionais – Regressão Linear Múltipla

A equação da regressão normalizada (Equação 15) obtida a partir dos dados do Navio Hidro Oceanográfico Cruzeiro do Sul mostrou um coeficiente de determinação (R^2) equivalente a 0,20, isto significa dizer que usando os valores das medições de temperatura e salinidade, a equação de regressão encontrada explica somente 20% dos dados, além de um erro padrão de 34,88% para as 1059 medições.

$$\text{Equação normalizada de regressão dos dados totais do Cruzeiro do Sul} \\ p\text{CO}_2(\pm 0,34) = -0,4377(\pm 0,06) + 0,3817(\pm 0,040)\text{TSM} + 0,49(\pm 0,083)\text{SSM} \\ \text{(EQUAÇÃO 15)}$$

No entanto, ainda restam 80% de variação não explicada por essa relação, o que implica que existem outras variáveis importantes para o estudo da variabilidade da $p\text{CO}_2$ que não foram incluídas. Os dados de temperatura e salinidade foram inseridos na equação 15 a fim de se estimar $p\text{CO}_2$ e, como resultado do teste, obteve-se um RSD equivalente a $0,6267\mu\text{atm}$.

6.3.1 Representação gráfica dos dados utilizados através de Regressão Linear Múltipla

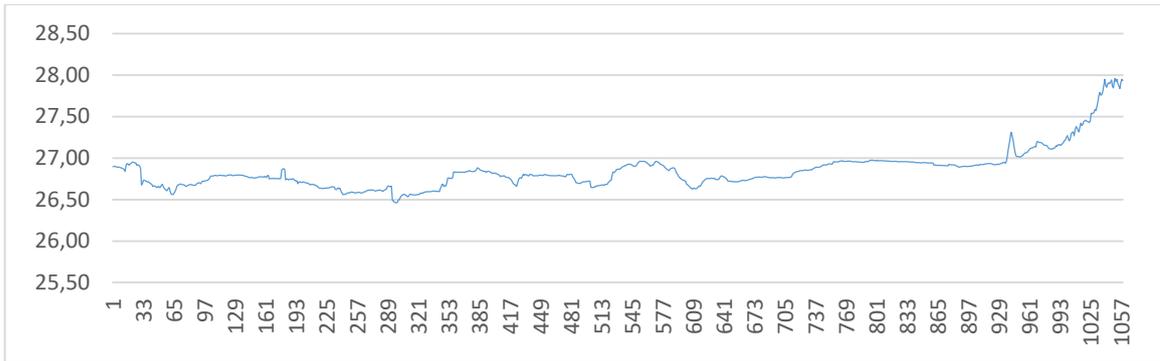
6.3.1.1 Gráficos dos dados totais por variável.

Conforme pode-se verificar nas figuras 26a, 26b e 26c, a salinidade, assim como a temperatura, não foi determinante na medição direta de $p\text{CO}_2$, fato este observado anteriormente nos dados do Monte Olívia, portanto observamos mudanças abruptas de perfil

tanto da salinidade quanto do $p\text{CO}_2$ e uma tendência crescente da temperatura, com alguns desvios maiores em alguns pontos, em direção aos números mais elevados das medições realizadas, entretanto, o aumento da salinidade e da temperatura, não refletem visualmente a grande amplitude de leituras da $p\text{CO}_2$.

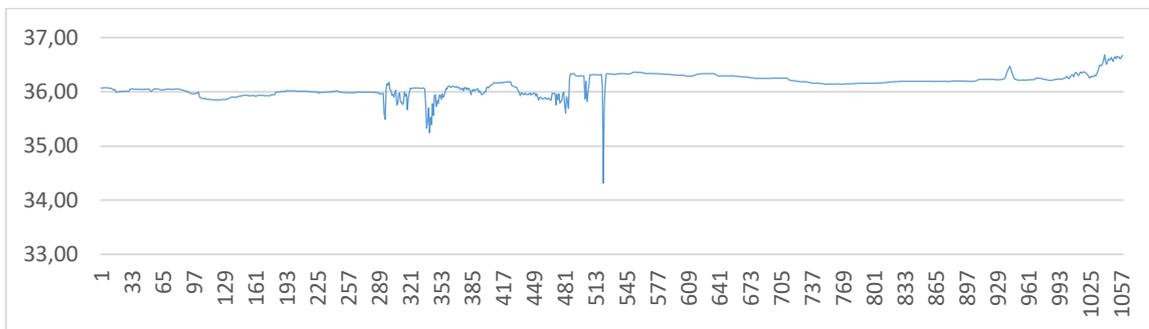
A variabilidade dos dados do Monte Olívia em relação aos obtidos ao longo da plataforma continental equatorial é relativamente pequena devido a poucas influências que as águas de mar aberto sofrem em relação à atividades biogeoquímicas da plataforma. Pequenas variações de temperatura e salinidade determinam alterações muito discretas na $p\text{CO}_2$. Situação bem diferente desta é a do Cruzeiro do Sul que sofre influência direta da plataforma continental e, o carreamento de materiais particulados e/ou componentes que alteram as aferições de diversos parâmetros, como a clorofila, compostos nitrogenados e outros podem determinar alterações significativas da $p\text{CO}_2$, conforme veremos no tópico 6.5.1.

Figura 26a – Dados de Temperatura em função do número da medição – Cruzeiro do Sul.



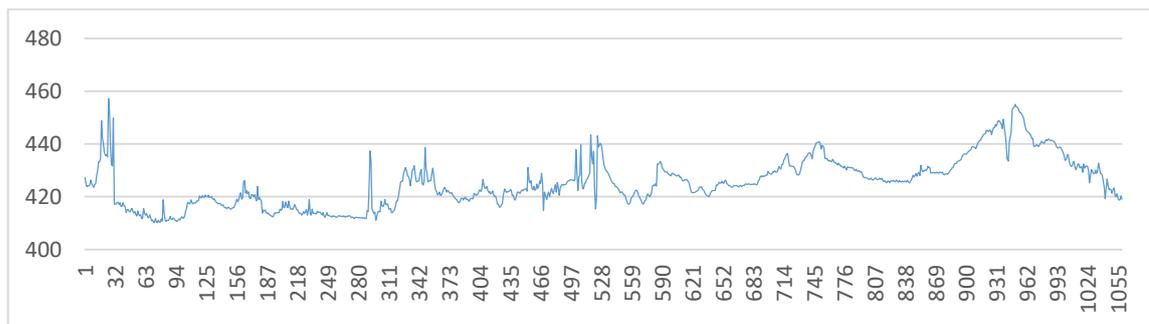
Fonte: Elaborado pelo autor

Figura 26b – Dados de Salinidade em função do número da medição – Cruzeiro do Sul.



Fonte: Elaborado pelo autor

Figura 26c – Dados da pCO₂ em função do número da medição – Cruzeiro do Sul.

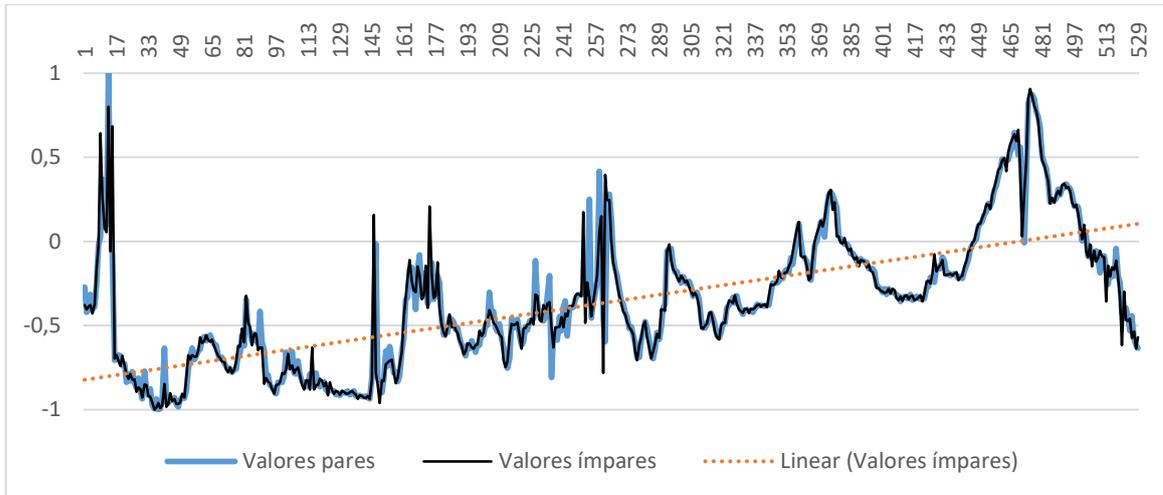


Fonte: Elaborado pelo autor

6.3.1.2 Treinamento com dados ímpares e validação com dados pares.

Com o objetivo de verificar se, a partir de um conjunto de dados, proveniente de pontos de amostragem próximos, consegue-se estimar os demais valores de dados, foi realizado o isolamento dos dados e suas representações encontram-se na figura 27.

Figura 27 – Dados ímpares e pares de pCO₂ normalizados sobrepostos – Cruzeiro do Sul.



Fonte: Elaborado pelo Autor.

A observação visual da distribuição espacial dos dados também não denotou diferença na distribuição espacial dos dados totais para os dados denominados ímpares e pares. A partir desta conclusão, uma vez mais, pode-se utilizar um desses conjuntos de dados (n=530) para obtenção de uma equação de regressão, para, em seguida aplicar o outro conjunto de dados e verificar se é possível estimar valores de um conjunto de dados a partir do outro.

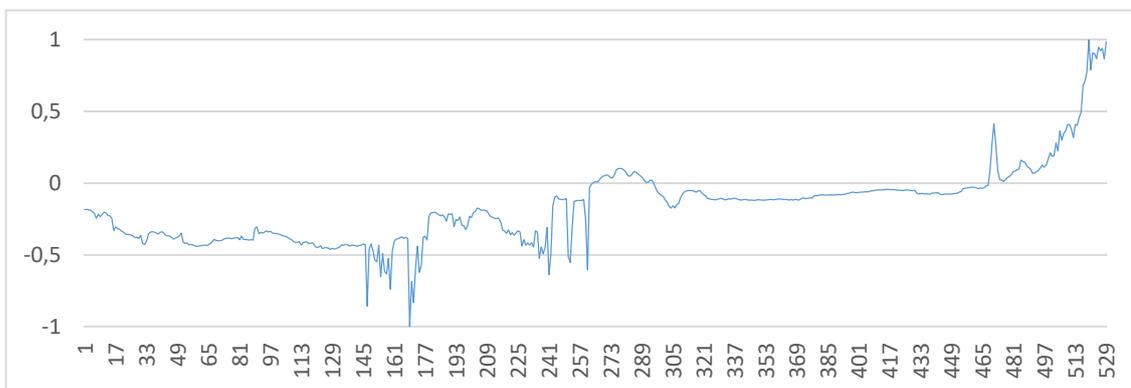
A equação 16, representa a regressão normalizada encontrada a partir dos dados ímpares e a consequente aplicação dos valores pares, para se encontrar a estimativa dos valores pares. A representação gráfica da estimativa está demonstrada na figura 28.

$$\text{Equação normalizada de regressão dos dados ímpares – Cruzeiro do Sul}$$

$$p\text{CO}_2(\pm 0,35) = -0,5951(\pm 0,088) + 0,3101(\pm 0,06)\text{TSM} + 0,71(\pm 0,127)\text{SSM}$$

(EQUAÇÃO 16)

Figura 28: Estimativa de pCO₂ pares normalizados via regressão obtida a partir de dados ímpares – Cruzeiro do Sul



Fonte: Elaborado pelo autor

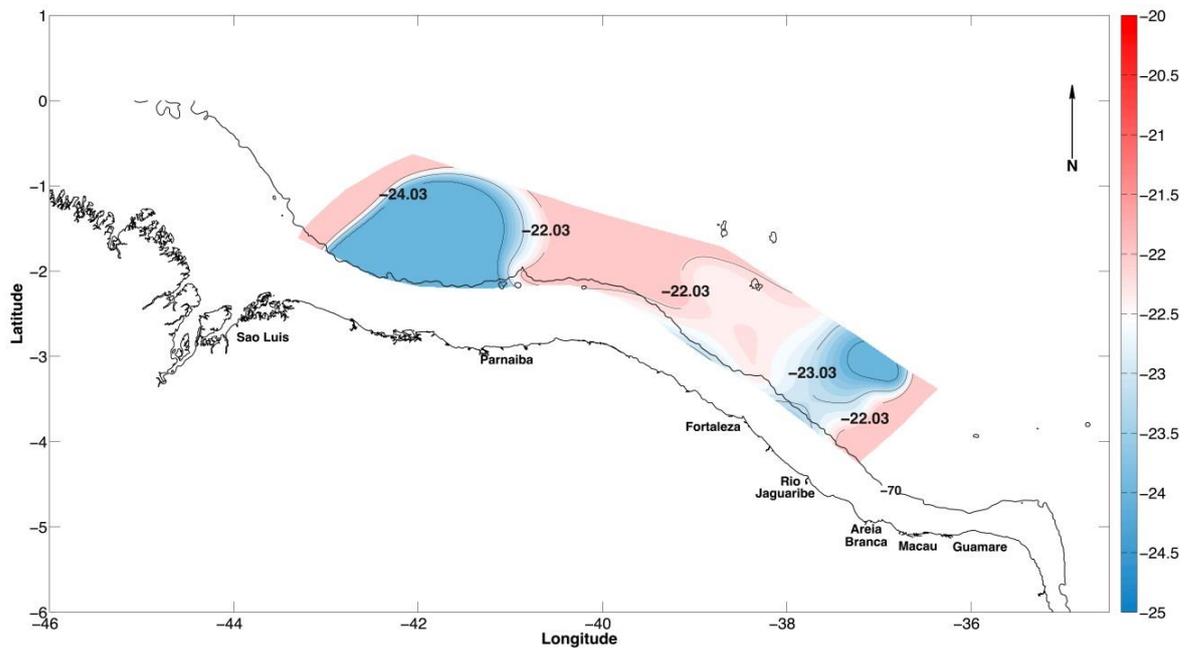
O erro padrão da aplicação dos dados pares na equação de regressão proveniente dos dados ímpares ($n=530$) foi de 35,78% e o R^2 , permaneceu dentro do esperado (20%), uma vez que o conjunto de dados, após a separação, permaneceu equivalente em relação ao número de medidas. Já o RSD, que representa o desvio padrão residual entre o dado estimado e o dado real foi de $0,4106\mu\text{atm}$, portanto, infere-se que não há provas estatísticas suficientes que valide a equação de regressão nos dados do Cruzeiro do Sul, a partir das variáveis temperatura e salinidade para estimativa da $p\text{CO}_2$.

Para o conjunto de dados do Cruzeiro do Sul optou-se por não realizar ensaios de regressão com dados salteados pois, conforme observamos a variabilidade real dos dados poderia ser removida aplicando-se tal distribuição.

Os dados do Cruzeiro do Sul comprovam que existe uma gama de variáveis que influenciam a $p\text{CO}_2$ na plataforma continental. A figura 29, por exemplo, mostra a grande variabilidade do $\delta^{13}\text{C}$, que é a razão entre $^{13}\text{C}/^{12}\text{C}$, indicativa da proporção de matéria orgânica terrestre que contribui para a matéria orgânica marinha, para determinações obtidas a partir de amostras coletadas em pontos fixos, durante o Cruzeiro do Sul. Os valores menores apontam a contribuição mais significativa de plantas terrestres (KURAMOTO; MINAGAWA, 2001). Através desta figura observou-se que outras variáveis merecem atenção durante a realização de experimentos que visem a mensuração de $p\text{CO}_2$ em regiões da plataforma equatorial brasileira e muito provavelmente em outras plataformas continentais também.

Durante o processo de decomposição de resíduos vegetais, bastante comum em regiões de manguezal, observa-se que mudanças de $\delta^{13}\text{C}$ e a matéria orgânica do solo são praticamente nulas. Portanto relações entre o solo e a decomposição de sua vegetação podem influenciar diretamente águas da plataforma continental.

Figura 29 – Distribuição de $\delta^{13}\text{C}$ durante o Cruzeiro do Sul



Fonte: Carvalho *et al.*, 2017.

6.4 Treinamento e Aplicação de dados da campanha do Navio Hidro Oceanográfico Cruzeiro do Sul em Técnicas Computacionais – Rede Neural Artificial

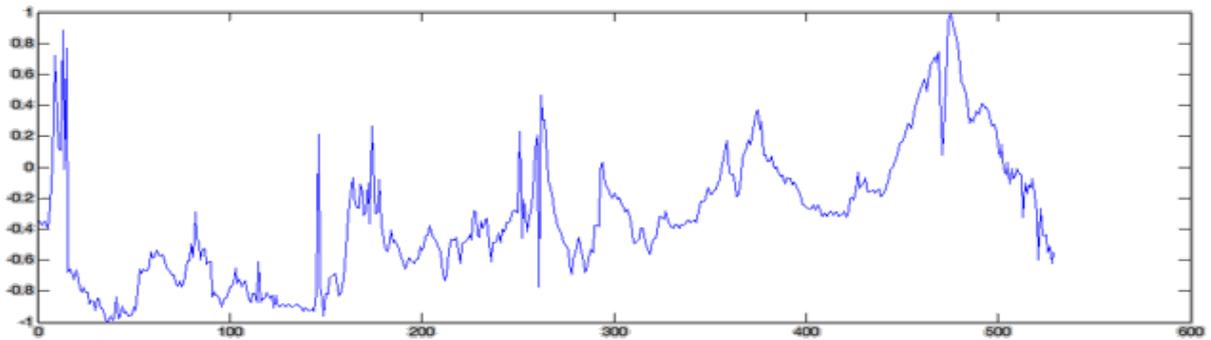
Com a linguagem de programação da rede neural totalmente desenvolvida, o retorno dos gráficos referentes a estimativa, aos dados reais, ao RSD e a normalização dos dados é bem mais fidedigno em relação ao Excel®, uma vez que os erros (Tabela 6) foram menores, em relação aqueles encontrados através da regressão linear múltipla e, a sensibilidade maior.

6.4.1 Representação gráfica dos dados utilizados através de Rede Neural Artificial

6.4.1.1 Treinamento com dados ímpares e validação com dados pares.

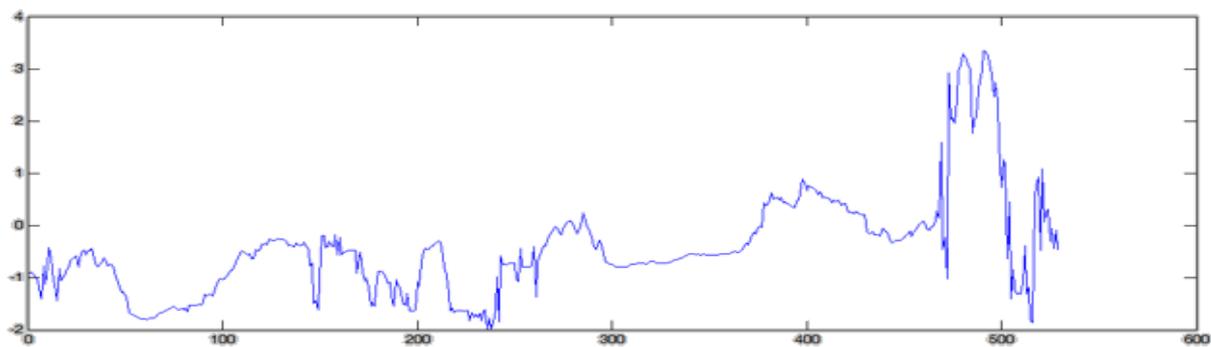
Apesar de visualmente os gráficos (figura 30a e 30b) referentes aos dados reais e aos dados estimados apresentarem semelhança muito grande obteve-se um RSD de $0,8394\mu\text{atm}$. Entretanto, as feições entre ambos se aproximam bastante e a estimativa acompanha os dados reais.

Figura 30a – Dados pares relativos a $p\text{CO}_2$ em relação ao número da medição – Cruzeiro do Sul



Fonte: Elaborado pelo Autor

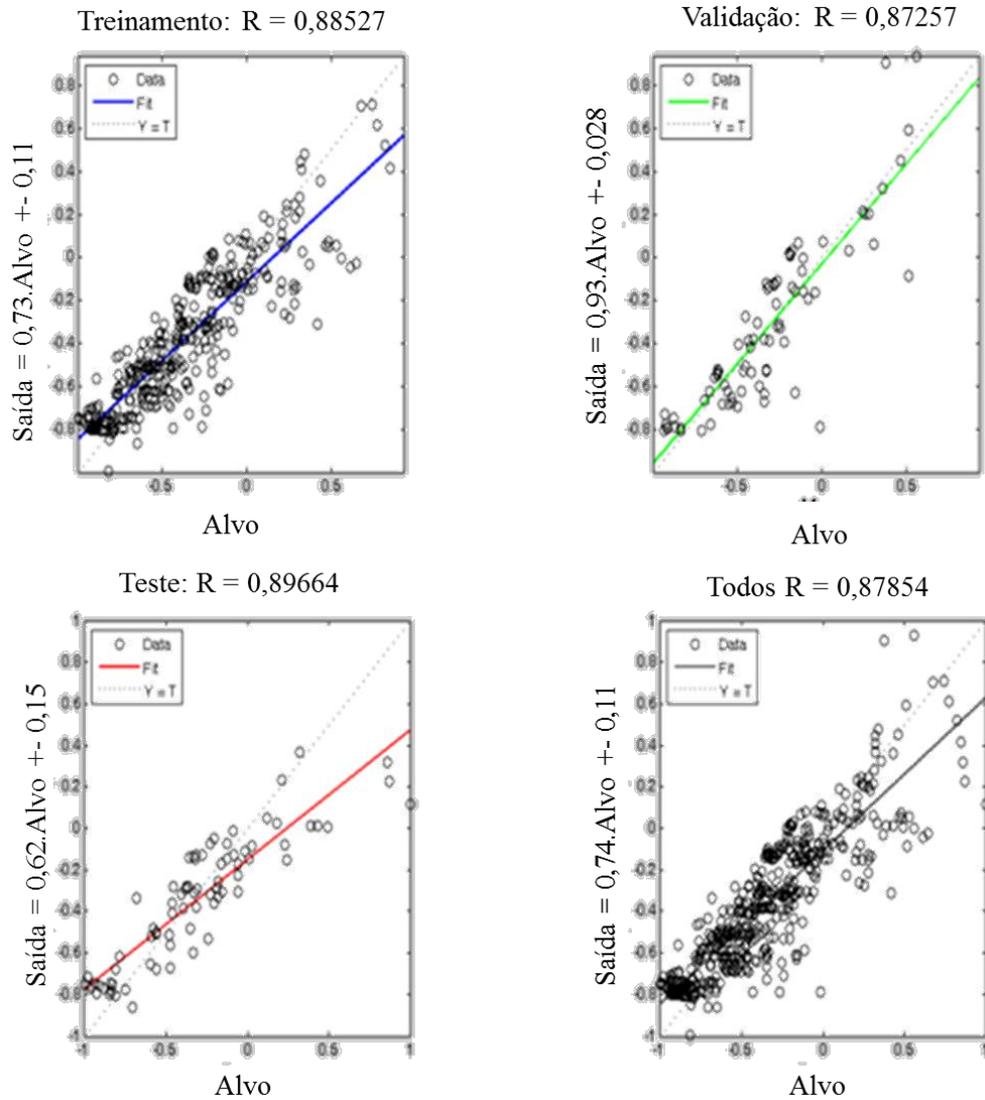
Figura 30b: Estimativa de $p\text{CO}_2$ pares normalizados via regressão obtida a partir de dados ímpares – Cruzeiro do Sul



Fonte: Elaborado pelo Autor

A figura 31 contém todas as informações relacionadas ao treinamento em rede. As imagens representam as etapas de treinamento, validação e teste. Neste ensaio, os dados de treinamento não indicam um bom ajuste devido aos resultados de validação e teste que mostram valores de R menores que 0,9. O gráfico de dispersão é útil para mostrar que determinados pontos de dados têm pouca semelhança.

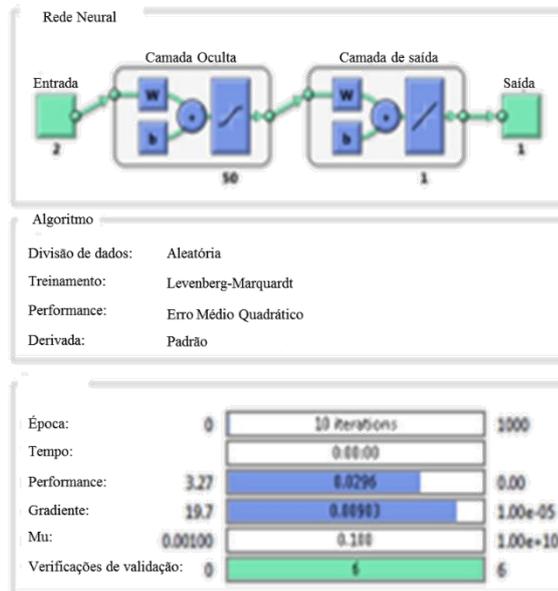
Figura 31 – Representação esquemática do treinamento, validação e teste da rede neural utilizada para estimar a $p\text{CO}_2$ par nos dados da campanha do Navio Hidro-Oceanográfico Cruzeiro do Sul



Fonte: Elaborado pelo Autor

A Figura 32 representa esquematicamente a rede neural utilizada nos dados ímpares do Cruzeiro do Sul para treinamento e os dados pares para validá-la.

Figura 32: Representação esquemática do funcionamento da rede neural com suas principais informações.



Fonte: Elaborado pelo Autor

6.4.1.2 Treinamento e validação com dados salteados – Cruzeiro do Sul.

Com os dados salteados em duas, cinco, quinze, trinta e cinquenta medidas, o RSD foi calculado e encontra-se na tabela 5.

Tabela 5 – RSD entre valores medidos e estimados nos dados salteados.

	n amostral	R²	RSD
2 dados salteados	353 dados	78,75%	0,8448 μ atm
5 dados salteados	177 dados	69,28%	0,8130 μ atm
15 dados salteados	67 dados	81,53%	0,6118 μ atm
30 dados salteados	35 dados	69,97%	1,1535 μ atm
50 dados salteados	21 dados	93,37%	0,6784 μ atm

Fonte: Elaborado pelo Autor

O esperado seria o RSD aumentar de acordo com a redução dos dados se houvesse um controle homogêneo da pCO₂ ao longo do trajeto de amostragem, mas não há, provavelmente em decorrência da influência direta da plataforma continental, como observado anteriormente na Figura 29. Com o treinamento, entretanto, eventualmente, a rede neural pode

retornar resultados muito próximos a cada treinamento executado, o que justifica, parcialmente as pequenas diferenças entre essas medidas.

Segundo Carvalho *et al.* (2017) há diferenças significativas entre as porções leste e oeste da plataforma relacionadas ao comportamento da pCO₂. A porção oeste da plataforma funciona de forma semelhante ao oceano aberto em relação ao pCO₂, influenciada por parâmetros físicos, principalmente a SSM, enquanto a parte ocidental da plataforma parece ter um comportamento mais complexo, influenciado também por parâmetros físicos como SST e SSM e ainda, por nutrientes e clorofila *a*, que mesmo em pequenas concentrações parece ter uma importância significativa na determinação da pCO₂.

6.5 Regressão Linear Múltipla vs Rede Neural Artificial

Os erros encontrados (Tabela 6) mostram que os gradientes utilizados pela RLM foram diferentes daqueles da rede neural, portanto devido à alta sensibilidade da rede neural, bem como ao menor erro encontrado, infere-se que a utilização do algoritmo de Levenberg-Marquardt em substituição a RLM apresenta estimativas mais fidedignas à pCO₂ medida.

O gradiente aqui citado fora utilizado para indicar o comportamento de pontos de inversão de cada grandeza com o intuito de exemplificar mudanças no gráfico de cada variável.

Tabela 6 – RSD entre valores medidos e estimados nos dados salteados do Monte Olívia.

	MONTE OLÍVIA		
	n amostral	RSD Excel	RSD Rede Neural
Dados Totais	3083 dados	0,2753μatm	0,1034μatm
Pares e ímpares	1542 dados	0,2768μatm	0,1150μatm
2 dados salteados	1028 dados	0,2782μatm	0,1275μatm
5 dados salteados	514 dados	0,3093μatm	0,1487μatm
15 dados salteados	193 dados	0,2785μatm	0,1614μatm
30 dados salteados	100 dados	0,2761μatm	0,3971μatm
50 dados salteados	61 dados	0,2754μatm	0,2432μatm

Fonte: Elaborado pelo Autor

A utilização do algoritmo de Levenberg-Marquardt, aplicado a rede neural artificial, mostrou-se bem mais eficiente do que a RLM, pois, apesar de ser um método de regressão/otimização aplicado ao Matlab, conseguiu obter um erro equivalente a 62,44%

menor, em relação ao RSD encontrado pela regressão linear múltipla nos dados totais da campanha Monte Olívia.

Conforme pode-se verificar na tabela 6, a comparação entre o desvio padrão residual do excel e a rede neural, referente ao Monte Olívia, apresenta uma eficiência da rede neural de no mínimo 11,7%, entretanto, em apenas um, dos sete ensaios propostos, a regressão linear se mostrou melhor do que a rede neural: o RSD obtido pela regressão foi 30,47% menor em relação aquele encontrado pela rede neural, isto justifica-se, provavelmente pela ocorrência de uma lacuna estatística ou até mesmo pela insuficiência de dados necessários para uma boa estimativa.

Mais uma vez reitera-se que não houve *offset* (medidas repetidas) nos dados analisados, ratifica-se ainda que as medidas obtidas estavam dentro do esperado e que a faixa analisada pela campanha Cruzeiro do Sul não pode ser representada unicamente pela temperatura e salinidade.

Os erros encontrados permitem concluir que a rede neural aplicada a estimativas de pCO₂ foi mais eficiente em relação as técnicas de estatística clássica via regressão linear múltipla. Para dados costeiros necessita-se de outras variáveis, além da TSM e SSM, pois a equação de regressão encontrada explica apenas 20% dos dados e, independentemente de ser RLM ou RNA, o modelo matemático não explica suficientemente os resultados, principalmente devido a argumentação de que o fenômeno sofre interferência biogeoquímica muito grande fugindo de padrões de natureza matemática.

A avaliação do RSD no uso de rede neural em estimativas de CO₂ no Atlântico Norte e Círculo Subpolar considerou que valores entre 2,6 a 15,8 eram aceitáveis (Lefèvre et al., 2005). Entretanto, a literatura internacional pode, por vezes, não se utilizar de dados normalizados para fins de cálculo do RSD, sendo necessário, portanto necessário a obtenção dos RSD deste trabalho com os dados não normalizados a fim de se preservar a comparação no pior cenário. A partir disso, o RSD da tabela 6, com 61 dados, dos 3083, do Monte Olívia, espaçados temporalmente, para os dados normalizados foi igual a 0,2432 μ atm, enquanto o RSD com os dados não normalizados foi igual a 6,18, na faixa inicial considerada aceitável pelos autores citados, embora haja diferença de métodos entre a rede usada naquele estudo e neste.

Para os dados totais do Monte Olívia, o RSD com dados não normalizados foi equivalente a 3,02 μ atm, enquanto na RLM foi equivalente a 7,35 μ atm. Reitera-se, portanto, que a RNA permanece apresentando RSD's menores do que aqueles encontrados pela RLM.

6.5.1 A influência da Clorofila *a* na estimativa

Corroborando com Carvalho *et al.* (2017), foi pesquisada a influência que a clorofila *a* exerce na estimativa de dados costeiros pois, tanto via estatística clássica quanto via rede neural ao incluirmos os dados de clorofila *a* obtidos pelos mesmos autores obteve-se RSD's menores em relação aqueles que não consideravam a clorofila *a*, tabela 7.

Estavam disponíveis 44 dados, dos 1059, com medidas de Clorofila *a*, salinidade, temperatura e fugacidade de CO₂.

Tabela 7 – RSD de variáveis utilizando RNA do Cruzeiro do Sul

	RSD Rede Neural
Temperatura	0,4608 μ atm
Salinidade	0,4046 μ atm
Clorofila	0,3700 μ atm
Temperatura e Salinidade	0,4116 μ atm
Temperatura e Clorofila	0,4435 μ atm
Salinidade e Clorofila	0,3514 μ atm
Temperatura, salinidade e clorofila	0,5983 μ atm

Conforme demonstrado neste trabalho, a rede neural apresenta resultados melhores do que RLM, entretanto, eventualmente, por alterações comuns no processo de inicialização da rede, onde roda o algoritmo LM, os erros podem variar a cada processo de inicialização, pois cada inicialização é um novo processo de aprendizagem da rede. Assim, em geral, as redes neurais respondem de maneira distinta sempre que são reiniciadas (retreinadas).

Diante disso, verificou-se que quando determina-se na RNA utilizada, apenas clorofila *a* como variável independente e a fugacidade como variável dependente, obtivemos um RSD de 0,37 μ atm, entretanto quando combinamos clorofila *a* com salinidade o RSD cai para 0,3514 μ atm, isto é, os dados do Cruzeiro do Sul, dados de plataforma continental, apresentam-se melhores estimados quando clorofila *a* e salinidade se combinam. Na tabela 7 encontram-se outros tipos de combinação de variáveis e ainda variáveis isoladas para a região estudada, durante o cruzeiro oceanográfico à bordo do NHOCS.

7 CONCLUSÃO

Este trabalho relacionou as variáveis disponíveis no banco de dados do Grupo de Pesquisa Biogeoquímica Costeira e estimou o comportamento do CO₂ em regiões costeiras tropicais, utilizando para treinamento dados referenciados internacionalmente via RLM e RNA.

As variáveis salinidade e temperatura foram relacionadas com a pCO₂, nos dados do Navio Mercante Monte Olívia, em região sem as variabilidades dos aportes continentais que afetam a pCO₂ em área da plataforma equatorial tropical, indicando que a salinidade é decisiva na determinação da pressão parcial, verificou-se também que a pCO₂ é explicada pela temperatura numa razão de 34%, pela salinidade numa razão de 35% e, pela temperatura e salinidade numa razão de 63% tomando como base o cálculo do R² e apenas as duas variáveis supracitadas.

A regressão linear múltipla e a rede neural artificial, técnicas computacionais utilizadas neste estudo, e seus poderes de análise, são ferramentas e instrumentações importantes para processos/estudos de estimativas da pCO₂. Essa aplicação é significativa pois, a obtenção de regressões, permite a representação gráfica para eventuais comparações visuais de feições.

Em oceano aberto, as grandezas salinidade e temperatura são mais significativas e representam melhor a variabilidade da pCO₂, entretanto, na plataforma continental, de acordo com os experimentos realizados neste trabalho, a heterogeneidade dos dados do Cruzeiro do Sul demonstrou que com $n \leq 1059$ não é possível estimar a pCO₂, utilizando-se a RLM.

A utilização do algoritmo de Levenberg-Marquardt, aplicado a rede neural artificial, mostrou-se mais eficiente do que as técnicas de estatística clássica, pois, apesar de ser um método de regressão/otimização aplicado ao Matlab, conseguiu obter um erro equivalente a 62% menor, em relação ao RSD encontrado pela regressão linear múltipla nos dados totais da campanha Monte Olívia.

Em relação aos dados do Cruzeiro do Sul, a rede neural obteve um erro 73,81% menor em relação ao RSD obtido pela regressão linear múltipla, tendo como base os dados totais.

Na tentativa de identificar um n mínimo para se ter uma boa previsão na campanha Cruzeiro do Sul, percebeu-se que com a utilização de todos os dados, 1059, obtivemos um RSD de 0,6267µatm, que, em relação a campanha Monte Olívia superou em muito seu valor, 0,2753µatm, embora com $n = 44$.

A clorofila *a*, quando presente, em ensaios via RNA nos dados do Cruzeiro do Sul, aumenta o RSD de $0,4116\mu\text{atm}$ para $0,5983\mu\text{atm}$, entretanto quando combina-se salinidade e clorofila *a*, como variáveis independentes, o RSD cai para $0,3514\mu\text{atm}$.

Sugerimos que em trabalhos futuros, Máquina de Vetores Suporte, do inglês, Support Vectors Machine – SVM, possam ser utilizadas para sistemas de previsão, entendendo que os erros encontrados sejam similares ou menores daqueles do algoritmo LM, pois a máquina é fundamentada na teoria da aprendizagem estatística, com o intuito de resolver problemas de classificação de padrões.

A técnica SVM objetiva construir um hiperplano como superfície de decisão, de tal forma que a separação entre exemplos seja máxima, isso se considerarmos padrões linearmente separáveis. Em situações onde os padrões não são linearmente separáveis buscar-se-á uma função de mapeamento apropriada para tornar o conjunto mapeado linearmente separável. Uma outra ferramenta importante para aplicação no oceano em um conjunto de dados hipervariáveis e vários parâmetros medidos.

REFERÊNCIAS

- BAIRD, C.; CANN, M. **Química Ambiental**. 4ª ed. Bookmann, Porto Alegre, 2011.
- BAPTISTELLA, M.; STEINER, M. T. A.; CHAVES NETO, A. **O uso de redes neurais e regressão linear múltipla na engenharia de avaliações: determinação dos valores venais de imóveis urbanos**. XXXVIII Simpósio Brasileiro de Pesquisa Operacional. Pesquisa Operacional na Sociedade: Educação, Meio Ambiente e Desenvolvimento. Goiânia: [s.n.], 2006.
- BOSTOCK, H. C.; MIKALOFF FLETCHER, S. E.; WILLIAMS, M. J. M. **Estimating carbonate parameters from hydrographic data for the intermediate and deep waters of the Southern Hemisphere oceans**. *Biogeosciences*, v. 10, p. 6199–6213, 2013.
- BOWERMAN, B.L.; O'CONNEL, R.T. **Linear Statistical Models: An Applied Approach**, 2. ed. Boston: PWS-Kent, 1990.
- CARVALHO, A.C.O. **Variabilidade da fugacidade do CO₂ (fCO₂) na interface oceano-atmosfera no Atlântico Equatorial**. Dissertação de Mestrado, Universidade Federal do Ceará, Instituto de Ciências do Mar, Fortaleza, 94 p., 2015.
- CARVALHO, A.C.O.; MARINS, R.V.; DIAS, F.J.S.; REZENDE, C.E.; LEFÈVRE, N.; CAVALCANTE, M.S.; ESCHRIQUE, S.A. **Air-sea CO₂ fluxes for the Brazilian northeast continental shelf in a climatic transition region**. *Journal of Marine Systems*. 2017.
- CASTRO, Belmiro Mendes; BRANDINI, Frederico P.; WAINER, Ilana E.K.C.; DOTTORI, Marcelo. **O mar de amanhã, com as mudanças climáticas de hoje**. *Cienc. Cult.* [online]. vol. 62, n. 3, p. 40-42, 2010.
- CAVALCANTE, M.S. **Transporte de carbono orgânico dissolvido no estuário do rio jaguaribe sob clima semiárido**. 2015.83p. Dissertação (mestrado) - Universidade Federal do Ceará. Fortaleza. 2015.
- CDIAC. Carbon Dioxide Information Analysis Center. Disponível em: http://cdiac.ornl.gov/ftp/oceans/VOS_Monte_Rio_Lines/MR_2009/. Acessos em 16/02/2017.
- CHEN, C. T. A.; BORGES, A. V. **Reconciling opposing views on carbon cycling in the coastal ocean: Continental shelves as sinks and near-shore ecosystems as sources of atmospheric CO₂**. *Deep-Sea Research II* v. 56, p. 578–590, 2009.
- COTOVICZ L. C. JR., KNOPPERS, B. A., BRANDINI, N., COSTA SANTOS, S. J., ABRIL, G. **A large CO₂ sink enhanced by eutrophication in a tropical coastal embayment (Guanabara Bay, Rio de Janeiro, Brazil)**. *Biogeosciences*. Discuss 12, 4671-4720, 2015.
- COTOVICZ L. C. JR.; LIBARDONI, B. G.; BRANDINI, N.; KNOPPERS, B. A.; ABRIL, G. **Comparações entre medições em tempo real da pCO₂ aquática com estimativas indiretas em dois estuários tropicais contrastantes: o estuário eutrofizado da Baía de Guanabara (RJ) e o estuário oligotrófico do Rio São Francisco (AL)**. *Química Nova*, Vol. 39, Nº 10, 1206-1214, 2016.

DICKSON, A.G., SABINE, C.L. e CHRISTIAN, J.R. **Guide to best practices for ocean CO₂ measurements**. PICES Special Publication 3, 191 p., 2007.

DONEY, S.C.; FABRY, V. J.; FEELY, R. A. e KLEYPAS, J.A. **Ocean Acidification: The other CO₂ problem**. Annual Review of Marine Science. v. 1, p.169-192, 2009.

DOYLE, J. D. **Coupled atmosphere-ocean wave simulations under high wind conditions**. Monthly Weather Review, v. 130, n. 12, p. 3087–3099, 2002.

DRAPER, N.R.; SMITH, H. **Applied Regression Analysis**. New York: John Wiley & Sons, 2. ed. 1981.

FEELY, R.A. *et al.* **Impact of Anthropogenic CO₂ on the CaCO₃ System in the Oceans**. Science v. 305, p. 362-366. 2004.

FONSECA, F. L. A. **Camada limite oceânica do Atlântico equatorial: experimentos numéricos**. 2011. 96 f. Dissertação (Mestrado em [Ciências Atmosféricas?]). Laboratório de Interação Ar-Mar. Grupo de Micrometeorologia. Instituto de Astronomia, Geofísica e Ciências Atmosféricas. Departamento de Ciências Atmosféricas. Universidade de São Paulo. 2011.

FRANÇA, J. A. de et al. **Uma implementação do algoritmo Levenberg-Marquardt dividido para aplicações em visão computacional**. Semina: Ciências Exatas e Tecnológicas, v. 30, n. 1, p. 51-62. Londrina, 2009.

FUNDAÇÃO OSWALDO CRUZ (FIOCRUZ). **O que são ciclones e tempestades tropicais?** <http://www.invivo.fiocruz.br/cgi/cgilua.exe/sys/start.htm?infoid=706&sid=9>. Acessado em: 09 jun. 2016.

FUNG, I. Y.; DONEY, S. C.; LINDSAY, K. e JONH, J. Evolution of carbon sinks in a changing climate. Proceedings of the National Academy of Sciences of the United States of America. V. 102, n. 32, p. 11201-11206, 2005.

GALTON, F. **Regression Toward Mediocrity in Heredity Stature**. Journal of Anthropological Institute. [S.l.]. v. 15, p. 246-263, 1885.

GAVIN, H. P. **The Levenberg-Marquardt method for nonlinear least squares curve-fitting problems**. Department of Civil and Environmental Engineering Duke University. 2017.

GIEBEL, G., LANDBERG, L.; KARIONIOTAKIS, G.; e BROWNSWORD R. **State-of-the-art on methods and software for short-term prediction of wind energy production**. European Wind Energy Conference and exhibition, CD-ROM, Madrid, Spain. 2003.

GOLDBERGER, A.S. **Econometric Theory**. New York. Ed. John Wiley & Sons, 1964.

GORNI, Antônio Augusto. **Redes Neurais Artificiais - Uma Abordagem revolucionária em Inteligência Artificial**. Revista MicroSistemas. [s.l.] ed. 133. p. 14-25 / ed. 134. p. 14-17, 2012.

GRUBER, N. **Warming up, Turning sour, Losing breath: ocean biogeochemistry under global change**. Philosophical Transactions of the Royal Society A. v. 369 p. 1943-1980, 2011.

- HAYKIN, Simon. **Redes neurais: princípios e prática**. 2.ed. Porto Alegre: Bookman, 2001.
- HEGER, Dominique A. **An Introduction to Artificial Neural Networks (ANN) - Methods, Abstraction, and Usage**. DHTechnologies, [s.l.]. Disponível em: <http://www.dhtusa.com/media/NeuralNetworkIntro.pdf>. Acesso em: 07 jun. 2016.
- IPCC. Painel Intergovernamental sobre Mudanças Climáticas. **Sumário do Relatório do IPCC para os Tomadores de Decisão do Quinto Relatório do Grupo de Trabalho II, 2014**. <http://www.ipcc.ch/report/ar5/wg2/>. Acessado em: 02 jun. 2016.
- KASZNAR, Istvan Karoly; GONÇALVES, Bento Mario Lages. **Regressão Múltipla: Uma digressão sobre seus usos**. Institutional Business Consultoria Internaciona (IBCI), Rio de Janeiro, 2007. Disponível em: http://www.ibci.com.br/Regressao_Multipla.pdf. Acesso em: 11 jun. 2016.
- KURAMOTO, T.; MINAGAWA, M. **Stable Carbon and Nitrogen Isotopic Characterization of Organic Matter in a Mangrove Ecosystem on the Southwestern Coast of Thailand**. *Journal of Oceanography*. v. 57, p. 421-431, 2001.
- KOLODNER, J. **Case-based reasoning**. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA. 1993.
- LANDSCHÜTZER, P. *et al.* **A neural network-based estimate of the seasonal to inter-annual variability of the Atlantic Ocean carbon sink**. *Biogeosciences*, v. 10, p. 7793-7815, 2013.
- LEFÈVRE, Nathalie; WATSON, Andrew J.; WATSON, Adam R. **A comparison of multiple regression and neural network techniques for mapping *in situ* pCO₂ data**. *University of East Anglia. Tellus*, v. 57B, p. 375–384, 2005.
- LE QUERE, C. *et al.* **Saturation of the Southern Ocean CO₂ sink due to recent climate change**. *Science* v. 316, p. 1735–1738, 2007.
- LIBES, Susan. **Introduction to Marine Biogeochemistry**. ed. 2, Elsevier, 2009.
- LOICZ. **Core project of the international geosphere-biosphere programme: a study of global change**. Quezon City: [s.n.]. Disponível em <http://projects.inweh.unu.edu/inweh/display.php?ID=752>. Acessos em 12 mar. 2016.
- MEDEIROS, M.C.; PEDREIRA, C.E. **What are the effects of forecasting linear time series with neural networks?** *Engineering Intelligent Systems*, v. 9, p. 237-242, 2001.
- MITCHELL, T.; HILL, M. **Machine Learning**. Simons Institute, Berkeley, 1997.
- MILLERO, Frank J. **Chemical Oceanography**. ed. 4. New York: CRC Press, 2013.
- MOLION, Luiz Carlos Baldicero. **Aquecimento Global: Uma visão crítica**. *Revista Brasileira de Climatologia*, 2008.

NETER, J.; WASSERMAN, W.; **Applied Linear Statistical Models**, Homewood III: Richard D. Irwin, 1974.

ORR *et al.* **Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms**. *Nature* n. 437, p. 681-686, 2005.

PEZZI, L. P.; CAVALCANTI, I.F.A. **The relative importance of ENSO and tropical Atlantic sea surface temperature anomalies for seasonal precipitation over South America: A numerical study**. *Climate Dynamics*, v. 17, p 205-212, 2001.

PEZZI, L. P.; SOUZA, R. B.; QUADRO, M. F. L. **Uma Revisão dos Processos de Interação Oceano-Atmosfera em Regiões de Intenso Gradiente Termal do Oceano Atlântico Sul Baseada em Dados Observacionais**. *Revista Brasileira de Meteorologia*, São Paulo, v. 31, n. 4, p. 428-453, dez. 2016 .

ROKACH, L. MAIMON, O. **Data Mining with Decision Trees**. Series in Machine Perception and Artificial Intelligence. v. 69. 2008.

SADAVA, David *et al.* **Vida: A Ciência da Biologia**. 8. ed. Porto Alegre: Artmed, 2009.

SHADWICK, E.H. *et al.* **Seasonal variability of dissolved inorganic carbon and surface water pCO₂ in the Scotian Shelf region of the Northwestern Atlantic**. *Marine Chemistry* v. 124, p 23–37, 2011.

SEBER, G.A.F. **Linear Regression Analysis**. New York: John Wiley & Sons, 1977.

SPECHT, Donald F. **A General Regression Neural Network**. *IEEE Transactions on Neural Networks*, vol. 2. ed. 6. 1991.

SRIVASTAVA, Shraddha; TRIPATHI, K.C. **Artificial Neural Network and Non-Linear Regression: A Comparative Study**. *International Journal of Scientific and Research Publications*, v. 2, ed. 12, 2012.

STRAMMA, Lothar.; ENGLAND, Matthew. **On the water masses and mean circulation of the South Atlantic Ocean**. *Journal of Geophysical Research*, v. 104, ed. c9, p. 20,863-20,883, 1999.

TAKAHASHI, Taro.; SUTHERLANDA, Stewart C.; SWEENEYA, Colm; *et al.* **Global sea-air CO₂ flux based on climatological surface ocean pCO₂, and seasonal biological and temperature effects**. *Deep-Sea Research II*. v. 49, p. 1601–1622, 2002.

TELSZEWSKI, M. *et al.* **Estimating the monthly pCO₂ distribution in the North Atlantic using a self-organizing neural network**. *Biogeosciences*. v. 6, p. 1405–1421, 2009.

VAPNIK, V. **Support-Vector Networks**. *Machine Learning*, v. 20, p.273-297, 1995.

YU, Hao; WILAMOWSKI, Bogdan M. **Intelligent Systems. Levenberg–Marquardt Training**. Cap. 12, p. 12-1, 12-2, 2010.