



## UM MÉTODO BASEADO EM MINERAÇÃO DE DADOS PARA A PREDIÇÃO DE VELOCIDADES E CONDIÇÕES EXTREMAS DE CORRENTES OCEÂNICAS

Katia Karinne de Oliveira Moraes

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia de Sistemas e Computação, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia de Sistemas e Computação.

Orientador(es): Geraldo Zimbrão da Silva  
Rodrigo Salvador Monteiro

Rio de Janeiro  
Março de 2014

UM MÉTODO BASEADO EM MINERAÇÃO DE DADOS PARA A PREDIÇÃO DE  
VELOCIDADES E CONDIÇÕES EXTREMAS DE CORRENTES OCEÂNICAS

Katia Karinne de Oliveira Moraes

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO  
LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA  
(COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE  
DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE  
EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Examinada por:

---

Prof. Geraldo Zimbrão da Silva, D.Sc.

---

Prof. Rodrigo Salvador Monteiro, D.Sc.

---

Prof. Geraldo Bonorino Xexéo, D.Sc.

---

Prof. Eduardo Soares Ogasawara, D.Sc.

RIO DE JANEIRO, RJ - BRASIL

MARÇO DE 2014

Moraes, Katia Karinne de Oliveira

Um método baseado em mineração de dados para a predição de velocidades e condições extremas de correntes oceânicas. – Rio de Janeiro: UFRJ/COPPE, 2014.

XV, 110 p.: il.; 29,7 cm.

Orientador: Geraldo Zimbrão da Silva

Rodrigo Salvador Monteiro

Dissertação (mestrado) – UFRJ/ COPPE/ Programa de Engenharia de Sistemas e Computação, 2014.

Referências Bibliográficas: p. 89-95.

1. Correntes Oceânicas. 2. Operações e Estruturas Offshore. 3. Mineração de Dados. 4. Predição de Condições Extremas. I. Silva, Geraldo Zimbrão, *et. al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III. Título.

À minha família.

## AGRADECIMENTOS

À Deus, por tudo o que fez em minha vida e por sempre iluminar e abençoar minha caminhada.

À minha família, que sempre me apoiou em todos os momentos e escolhas que fiz na vida, e pelo suporte durante os três anos de Mestrado.

Ao meu orientador, professor Geraldo Zimbrão, pela oportunidade e confiança em mim como aluna de regime parcial. Agradeço imensamente pela preciosa orientação, disponibilidade e atenção.

Ao professor Rodrigo Salvador, meu co-orientador, que me aceitou como orientanda e, sempre atencioso e comprometido, me guiou impecavelmente durante este trabalho. Meus sinceros agradecimentos pela dedicação, apoio, paciência e suas valiosas contribuições.

Ao professor Jano, que me orientou no início do Mestrado. Não poderia deixar de agradecer também pela oportunidade, confiança e por ter contribuído para a minha evolução durante este período.

À minha Gerente Ana Luzia, por todo empenho e esforços para que o Mestrado e minha liberação fossem aprovados. Obrigada pela oportunidade, generosidade, confiança e apoio.

Ao meu Coordenador Fabiano Sanches, que acompanhou minha caminhada nas fases finais da Dissertação e me apoiou a todo momento, sempre muita generosidade e atenção. Muito obrigada.

À Petrobras pela liberação para o Mestrado.

Ao amigo Elder Bonfim, que me apresentou o professor Rodrigo Salvador. Obrigada pelo incentivo, por torcer pelo meu sucesso e pela excelente indicação de Orientador.

A Francisco dos Santos e Bruna Cerrone, da empresa ProOceano, pela disponibilidade, contribuições na área de Oceanografia e valiosas orientações durante o projeto. As contribuições dos dois foram indispensáveis para este trabalho.

Aos professores Geraldo Xexéo e Eduardo Ogasawara, por participarem da minha banca de defesa de Mestrado.

Ao grupo de Computação para Multidões, em especial, Daniel e Gilda pela amizade, contribuições e incentivos.

Aos colegas que fiz no PESC, pelo companheirismo, amizade e pelos momentos agradáveis vividos no período.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

## UM MÉTODO BASEADO EM MINERAÇÃO DE DADOS PARA A PREDIÇÃO DE VELOCIDADES E CONDIÇÕES EXTREMAS DE CORRENTES OCEÂNICAS

Katia Karinne de Oliveira Moraes

Março/2014

Orientadores: Geraldo Zimbrão da Silva

Rodrigo Salvador Monteiro

Programa: Engenharia de Sistemas de Computação

Atividades offshore envolvem elementos estruturais com grande sensibilidade a carregamentos ambientais (ventos, correntes, ondas), que interferem nas operações, podendo resultar em danos ou falhas, com perdas financeiras. A observação e predição de condições meteo-oceanográficas futuras são requisitos altamente relevantes para o planejamento, projeto e operação. Na ausência de informações confiáveis sobre as condições futuras de operação, as mesmas podem ser interrompidas por períodos bem maiores do que o necessário por questões de segurança. O presente trabalho propõe uma metodologia baseada em Mineração de Dados para a geração de um preditor de intensidade de correntes oceânicas. Trata-se de um modelo de predição para um ponto específico do espaço geográfico, cuja abordagem é apoiada na influência de correntes medidas em pontos vizinhos para estimar o estado da corrente neste ponto em um momento futuro próximo. Os resultados mostraram que a série obtida com o preditor acompanha a série observada com um erro médio absoluto (MAE) de 0,09 m/s e raiz quadrada do erro médio (RMSE) de 0,12 m/s. Igualmente eficiente, o modelo classificador atingiu a taxa de acerto de 83% quando validado frente a série histórica. Estes resultados comprovam que a metodologia pode ser usada como uma ferramenta confiável para a previsão de intensidades de correntes.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

A METHOD BASED ON DATA MINING FOR PREDICTION OF OCEAN  
CURRENT'S SPEED AND EXTREME CONDITIONS

Katia Karinne de Oliveira Moraes

March/2014

Advisors: Geraldo Zimbrão da Silva  
Rodrigo Salvador Monteiro

Department: Systems Engineering and Computer Science

Offshore activities involve structural elements with high sensitivity to environmental loads (wind, currents, waves), which interfere on operations and may result in damage or failure, with financial losses. The observation and prediction of future meteocean conditions are strongly relevant requirements for planning, design and operation. In the absence of reliable information about future conditions of operation, the same may be interrupted by periods much larger than the necessary for security. This work proposes a methodology based on Data Mining for generating a predictor of ocean currents intensity. It is a prediction model to a specific point in geographical space, whose approach is supported by the influence of currents observed at neighboring points to estimate the state of the current at this point at a near future time. The results showed that the series obtained with the predictor followed the observed series with a (mean absolute error) MAE of 0.09 m/s and (root mean square error) RMSE of 0.12 m/s. Equally efficient, the classifier model achieved an accuracy rate of 83 %. These results prove that the method can be used as a reliable tool for predicting current intensities.

## SUMÁRIO

<b>CAPÍTULO 1: INTRODUÇÃO.....</b>	<b>1</b>
1.1 OBJETIVO .....	2
1.2 MOTIVAÇÕES .....	3
1.3 ESTRUTURA DO TRABALHO .....	6
<b>CAPITULO 2: EMBASAMENTO TEÓRICO.....</b>	<b>7</b>
2.1 OPERAÇÕES OFFSHORE .....	7
2.1.1 SISTEMAS DE POÇOS SUBMARINOS .....	11
2.1.2 CARREGAMENTOS .....	13
2.2 CORRENTES OCEÂNICAS .....	15
2.2.1 O EFEITO CORIOLIS.....	16
2.2.2 CIRCULAÇÃO TERMOALINA .....	17
2.3 CORRENTE DO BRASIL.....	18
2.3.1 ESTRUTURA VERTICAL DE MASSAS DE ÁGUA .....	20
<b>CAPITULO 3: MINERAÇÃO DE DADOS.....</b>	<b>23</b>
3.1 DECOMPOSIÇÃO EM VALORES SINGULARES .....	26
3.1.1 REDUÇÃO VIA SVD .....	28
3.2 TAREFAS DE MINERAÇÃO DE DADOS.....	29
3.3 MÉTODOS DE MINERAÇÃO DE DADOS.....	29
3.3.1 PREVISÃO .....	30
3.3.2 REGRESSÃO .....	30
3.3.3 CLASSIFICAÇÃO .....	30
3.3.4 AGRUPAMENTO .....	31
3.3.5 ASSOCIAÇÃO .....	31
3.3.6 ANÁLISE DE SEQUÊNCIA (OU SÉRIES TEMPORAIS) .....	32
3.4 MEDIÇÕES .....	32
3.5 REDES NEURAI.....	35
3.6 ÁRVORES DE DECISÃO .....	38
<b>CAPITULO 4: DEFINIÇÃO DO PROBLEMA.....</b>	<b>41</b>
4.1 CORRENTES OCEÂNICAS E REQUISITOS OPERACIONAIS .....	41
4.2 LIMITES OPERACIONAIS .....	45
4.3 TRABALHOS RELACIONADOS.....	46
<b>CAPITULO 5: PROPOSTA .....</b>	<b>52</b>

<b>5.1 ÁREA DE ESTUDO .....</b>	<b>52</b>
<b>5.2 METODOLOGIA DE TRABALHO .....</b>	<b>53</b>
5.2.1 COLETA DOS DADOS.....	53
5.2.2 ANÁLISE PRELIMINAR DOS DADOS .....	53
5.2.3 TRATAMENTOS DOS DADOS E GERAÇÃO DO CONJUNTO DE TREINAMENTO .....	54
5.2.4 CONSTRUÇÃO DO MODELO E EXPERIMENTOS.....	59
<b>5.3 RECURSOS E FERRAMENTAS .....</b>	<b>61</b>
5.3.1 BASES DE DADOS.....	61
5.3.2 NETCDF .....	65
5.3.3 FERRAMENTA R.....	66
5.3.4 WEKA .....	67
<b>6.1 RESULTADOS – REGRESSÃO.....</b>	<b>68</b>
<b>6.2 RESULTADOS – CLASSIFICAÇÃO ESCALA 20-20 (INTERVALOS DE 20 CM/S) .</b>	<b>76</b>
<b>6.3 RESULTADOS – CLASSIFICAÇÃO 3 ESCALAS .....</b>	<b>84</b>
<b>6.4 RESULTADOS – PREVISÃO 2 DIAS A FRENTE.....</b>	<b>89</b>
<b>6.5 AVALIAÇÃO CRÍTICA .....</b>	<b>95</b>
<b>CAPITULO 7: CONCLUSÃO.....</b>	<b>100</b>

## LISTA DE FIGURAS

Figura 1: Plataforma Fixa .....	8
Figura 2: Plataformas Auto-Eleváveis .....	9
Figura 3: Plataforma Tension Leg .....	10
Figura 4: Navio de Sonda .....	11
Figura 5: Plataforma Semi-submersível .....	11
Figura 6: Mapa dos cinco giros oceânicos. Fonte (UNIVERSITY OF WAIKATO, 2012).....	17
Figura 7: Circulação Termoalina .....	18
Figura 8: Correntes superficiais da Costa brasileira. FONTE: (MENDES, GOMES, 2007) .....	19
Figura 9: Representação esquemática da circulação superficial de larga escala do Oceano Atlântico Sul. (ASSIREU, 2003) .....	20
Figura 10. Representação esquemática do Sistema Corrente do Brasil. Fonte: (GODOI, 2005) .....	22
Figura 11: Processo KDD. FONTE: (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996a, 1996b) .....	25
Figura 12: SVD reduzida .....	28
Figura 13: Arquitetura de uma rede neural MLP .....	37
Figura 14: Árvore de Decisão para o Problema Jogar Tênis .....	39
Figura 15: Fontes de informações de correntes. FONTE: (SOS-OC, 2002) .....	50
Figura 16: Comparação das séries de intensidade em 2 pontos vizinhos .....	54
Figura 17: Grade 7x7 com o ponto central e vizinhança de 3 vizinhos .....	55
Figura 18: Exemplo de uma grade extraída da base do Hycom Consortium .....	56
Figura 19: Aplicação da SVD sobre a massa de dados para redução de dimensionalidade.....	57

Figura 20: Localização dos Fundeios de Medição de Corrente (Fonte: WOCE) .....	61
Figura 21: Fundeios de Medição Localizados na região da Bacia de Campos (Fonte: WOCE) .....	62
Figura 22: Grade 9x9 com o ponto central e vizinhança de até 4 pontos distantes .....	68
Figura 23: Resultado do Experimento REG1 - Gráfico Comparativo das séries observada e prevista .....	70
Figura 24: Resultado do Experimento REG2 - Gráfico Comparativo das séries observada e prevista .....	72
Figura 25: Grade 3x3 com o ponto central e vizinhança com os pontos mais próximos.....	73
Figura 26: Resultado do Experimento REG3 - Gráfico Comparativo das séries observada e prevista .....	74
Figura 27: Resultado do Experimento REG4 - Gráfico Comparativo das séries observada e prevista .....	76
Figura 28: Grade 13x13 com o ponto central e vizinhança com os pontos mais próximos.....	90
Figura 29: Resultado do Experimento REG5 - Gráfico Comparativo das séries observada e prevista .....	91

## LISTA DE TABELAS

Tabela 1: Matriz de Confusão .....	34
Tabela 2: Erros calculados para o modelo do Experimento REG1 .....	70
Tabela 3: Erros calculados para o modelo do Experimento REG2 .....	72
Tabela 4: Erros calculados para o modelo do Experimento REG3 .....	74
Tabela 5: Erros calculados para o modelo do Experimento REG4 .....	76
Tabela 6: Erros calculados para o modelo do Experimento REG5 .....	91
Tabela 7: Erros calculados por classe da matriz de confusão do modelo 7 – C1 .....	96

## LISTA DE QUADROS

Quadro 1: Formato dos conjuntos de dados .....	55
Quadro 2: Categorias de Intensidade por faixas de 20 m/s .....	58
Quadro 3: Categorias de Intensidade por faixas de segurança operacional .....	58
Quadro 4: Categorias de Intensidade por faixas adaptada .....	59
Quadro 5: Formato do conjunto de dados com SVD .....	60
Quadro 6: Formato do conjunto de dados sem SVD .....	61
Quadro 7: Exemplo de estrutura de um documento NetCDF .....	66
Quadro 8: Formato do conjunto de dados - Experimento REG1 .....	69
Quadro 9: Modelo selecionado para o Experimento REG1 .....	69
Quadro 10: Formato do conjunto de dados - Experimento REG2 .....	71
Quadro 11: Modelo selecionado para o Experimento REG2 .....	71
Quadro 12: Formato do conjunto de dados - Experimento REG3 .....	73
Quadro 13: Modelo selecionado para o Experimento REG3 .....	74
Quadro 14: Formato do conjunto de dados - Experimento REG4 .....	75
Quadro 15: Modelo selecionado para o Experimento REG4 .....	76
Quadro 16: Modelos de Rede Neural selecionados para o Experimento C1 .....	77
Quadro 17: Modelos de Árvore de Decisão selecionados para o Experimento C1 .....	78
Quadro 18: Modelos de Rede Neural selecionados para o Experimento C2 .....	80
Quadro 19: Modelos de Árvore de Decisão selecionados para o Experimento C2 .....	81
Quadro 20: Modelos de Rede Neural selecionados para o Experimento C3 .....	82
Quadro 21: Modelos de Árvore de Decisão selecionados para o Experimento C3 .....	84
Quadro 22: Modelo de Rede Neural selecionado para o Experimento C4 .....	85
Quadro 23: Modelo de Árvore de Decisão selecionado para o Experimento C4 .....	86
Quadro 24: Modelos de Rede Neural selecionados para o Experimento C5 .....	87
Quadro 25: Modelo de Árvore de Decisão selecionado para o Experimento C5 .....	87

Quadro 26: Modelo de Rede Neural selecionado para o Experimento C6 .....	88
Quadro 27: Modelos de Árvore de Decisão selecionados para o Experimento C6 .....	88
Quadro 28: Formato do conjunto de dados - Experimento REG5 .....	90
Quadro 29: Modelo selecionado para o Experimento REG5 .....	91
Quadro 30: Modelos de Rede Neural selecionados para o Experimento C7 .....	92
Quadro 31: Modelo de Árvore de Decisão selecionado para o Experimento C7 .....	93
Quadro 32: Modelo de Rede Neural selecionado para o Experimento C8 .....	94
Quadro 33: Modelo de Árvore de Decisão selecionado para o Experimento C8 .....	94
Quadro 34: Comparação do Modelo da Linha base para classes 20-20 com o Modelo 5 – C1 .....	97
Quadro 35: Comparação do Modelo da Linha base para classes por nível de segurança com o Modelo 4 – C3 .....	98

# Capítulo 1: Introdução

A crescente demanda por combustíveis fósseis e as especulações em torno da exploração da camada de pré-sal tem colocado em evidência a importância do setor petrolífero em âmbito mundial. A expansão da exploração off-shore estimulou empresas do ramo de petróleo a investir em pesquisa e tecnologia, visando maior excelência no processo de exploração e produção, porém, com uma maior preocupação com a segurança operacional e responsabilidade social.

As atividades petrolíferas envolvem, além de um capital alto, diversos impactos ambientais. Pequenas falhas podem causar grandes vazamentos com consequências graves ao meio ambiente e prejuízos financeiros de alta escala. Desta forma, torna-se essencial um planejamento cuidadoso das operações.

Na indústria de petróleo, condições ambientais como ondas, ventos e correntezas interferem nas operações de equipamentos e, por isso, são importantes fatores a serem analisados em locais onde plataformas são instaladas. BARBARIOLI et al. (2011) afirmam que informações ambientais precisas são fundamentais para segurança, economia e, conseqüentemente, para o sucesso das atividades exercidas. Ainda de acordo com BARBARIOLI et al. (2011), mais de 80% das atividades de exploração e produção de petróleo e gás são provenientes de regiões oceânicas.

Segundo Augusto, Andrade e Rossi (2002), as forças de correnteza podem agir no sentido de retirarem os cabos suspensos de seu plano vertical e essa limitação tem sido muito usada, pois permite o cálculo analítico das linhas de ancoragem, dentro das faixas de velocidades de correnteza encontradas na costa brasileira e dos materiais empregados para linhas de ancoragem.

Diante deste cenário, a análise e previsão de condições futuras é uma atividade que assume grande relevância e fornece informações estratégicas para o planejamento, projeto e operação em estruturas oceânicas. Com base nessas informações, engenheiros podem avaliar o risco de possíveis interferências nas operações dos equipamentos e tomar uma decisão com relação ao momento certo de operar ou suspender as operações.

(CAETANO NETO, et al., 1996) afirma que a previsão de tempo e ondas sobre áreas oceânicas é um problema de grande interesse para países com extensa faixa costeira, mas não há esforços no sentido de implantar ou desenvolver um sistema de

previsão com finalidade operacional no Brasil. Embora a utilização de produtos de modelos numéricos de previsão de tempo venha sendo feita já há várias décadas, os esforços no sentido da operacionalização desses modelos ainda é recente (CAETANO NETO, et al., 1996).

Mineração de Dados constitui-se em uma tecnologia capaz de automatizar processos de busca por informações preditivas que podem estar ocultas dentro de um amontoado de dados (BASTOS, 2011). Técnicas de Mineração de Dados possuem diversas aplicações em diferentes áreas com Medicina e Biologia, Marketing, Economia, Segurança, dentre outras.

A análise de dados referentes a fatores climáticos e meteorológicos se mostra como um interessante domínio para a aplicação de técnicas de mineração, tendo em vista que em geral as informações são frequentemente utilizadas para prever condições futuras. Isso é viabilizado pela disponibilidade de um grande volume de informações históricas que permite avaliar correlações, tendências, e descobrir padrões que auxiliem especialistas em suas previsões, de maneira automatizada e com alto nível de acurácia e confiabilidade.

## 1.1 OBJETIVO

O presente trabalho tem como objetivo avaliar os dados históricos de intensidade de correntes oceânicas observadas em um ponto específico e pontos de sua vizinhança, a fim de extrair alguma espécie de correlação ou influência que sirva como parâmetro para a definição de um modelo de predição. Como principal produto deste trabalho, é fornecido um modelo e a metodologia de geração de modelos semelhantes, que consigam aproximar condições futuras de correntes oceânicas para um ponto específico em função de valores históricos e atuais de intensidade deste e de seus vizinhos.

É importante enfatizar que este trabalho fornecerá um modelo aplicado a um ponto específico da Costa Brasileira, mais precisamente, na região da Bacia de Campos, que é uma área de enorme potencial e importância para a indústria nacional de petróleo. No entanto, a metodologia para a geração deste modelo pode ser aplicada a outras regiões para gerar novos modelos preditores para outros pontos específicos.

## 1.2 MOTIVAÇÕES

O processo de exploração de petróleo conta com uma ampla classe de estruturas offshore, como navios petroleiros, plataformas marítimas, tubulações e linhas de ancoragem.

As operações de perfuração em águas profundas envolvem elementos estruturais com grande sensibilidade a forças provocadas por processos oceanográficos a que estão expostos. Esses processos, conhecidos como carregamentos ambientais ou dinâmicos, causam movimentação dinâmica na estrutura e interferem nas operações de perfuração, podendo resultar em danos por excesso de carga ou falhas por fadiga. De acordo com (PICCININI, 2008), as ações resultantes do deslocamento de massas d'água, em especial, as provocadas pelas ondas e correntes, são as mais significativas na concepção do projeto. A interferência das condições ambientais nas operações offshore representam problemas tão significativos que tornaram-se requisito fundamental a ser considerado no projeto de instalações offshore. A DET NORSKE VERITAS, *Rules for Offshore Installations* (VERITAS, 1995), em suas recomendações deixa evidente tal relevância:

“Todos os fenômenos ambientais relevantes que podem influenciar no projeto devem ser considerados. Desse modo, como fenômenos devemos incluir:

- ventos; ondas; correntes; neve; temperatura, marés; incrustações de organismos marinhos; terremotos.”

Dentre os principais aspectos que afetam severamente as operações pode-se citar: fadiga devido a Vortex Induced Vibration (VIV), rompimento do *riser*, tempo de inatividade. De acordo com BARBOSA (2007), aproximadamente 10 milhões de dólares extras são gastos por ano pela indústria de petróleo com equipamentos que sofrem com o desgaste. Além disso, danos em estruturas como *riser* de perfuração geram paralisação das atividades para substituição ou reparos.

Em 2001, um levantamento realizado pelo Instituto Americano de Petróleo (AMERICAN PETROLEUM INSTITUTE, 2001) estimou que o custo anual total da atividade de perfuração offshore no Golfo do México ficava na faixa de 3 a 5 bilhões de dólares. Evidências empíricas sugerem que de 1% a 3% do custo total da perfuração são

devido à inatividade em função de condições climáticas e oceânicas adversas (JENKINS; CROCKFORD, 1975). Levando em conta que melhores sistemas de observação oceânica podem mitigar 1% destes custos, a expectativa anual de economia é estimada entre US\$ 300.000 e US\$ 1,5 milhões (KAISER; PULSIPHER, 2007). Atualmente, essa economia seria bem superior. O custo diário pelo aluguel de uma sonda de perfuração moderna é estimado em mais de US\$ 500.000 por dia. Em fevereiro de 2012, a Petrobras anunciou um acordo de contrato estimado em cerca de 75,938 bilhões de dólares para o aluguel de 26 navios-sonda de perfuração marinha em águas profundas, durante 15 anos. Do total de 26 sondas, 21 são de um contrato que prevê um aluguel diário de 530 mil dólares cada uma. As outras 5 são alugadas por um valor médio diário de 548 mil dólares, a unidade.

Conclui-se, portanto, que um dia de inatividade em uma única instalação pode ultrapassar 600 mil dólares, pois o prejuízo gerado também leva em consideração outros recursos envolvidos, como toda a equipe mobilizada.

Contribuindo para os valores acima, vale lembrar que, em caso de fraturas ou rompimentos, uma provável contaminação do meio ambiente pode ocorrer em função de vazamentos de fluídos transportados por tubulações. Este tipo de incidente gera altos custos que envolvem a limpeza da área impactada, multas por danos ambientais e prejuízos relacionados à imagem da empresa.

Diante do exposto, métodos acurados para estimar a intensidade de fenômenos extremos representam ferramentas úteis para o auxílio aos engenheiros na construção de estruturas oceânicas e no processo operacional, de forma mais econômica e confiável.

A enorme expansão, vivenciada nos últimos anos, da indústria do petróleo e derivados tem despertado atenção para as atividades de exploração e os desafios da camada pré-sal. A grande necessidade de estudos e pesquisas no ramo, em prol da excelência operacional, movimentou instituições e universidades não só para a formação de profissionais na área, mas também para o desenvolvimento de projetos em linhas de pesquisas relacionadas às atividades petrolíferas. Isso tem motivado o desenvolvimento de sistemas e modelos computacionais para apoiar em diferentes domínios de atividades como: simulação numérica do comportamento de estruturas, modelagem gráfica de estruturas, predição de condições ambientais, etc.

Sistemas como estes são amplamente utilizados por empresas de petróleo desde a concepção de projetos até a operação em estruturas, fazendo parte do cotidiano operacional.

Há alguns anos, vem sendo desenvolvida uma complexa linha de pesquisa com o intuito de compreender o comportamento de estruturas mediante efeito de correntes marítimas, análise de limites operacionais e previsões meteo-oceanográficas. Uma profunda investigação sobre natureza dos fenômenos, identificação dos efeitos sobre as instalações e a viabilização de soluções que garantam a segurança operacional (reduzindo efeitos danosos, aumentando a durabilidade dos equipamentos, antecipando condições críticas) são fundamentais.

Como parte dos esforços nessa linha, o presente trabalho está interessado em fornecer previsões de intensidade de correntes oceânicas, de maneira a antecipar condições extremas que inviabilizem a operação segura, apoiando engenheiros e operadores no planejamento e tomada de decisão. O objetivo não é criar um modelo de previsão genérico que seja utilizado em qualquer região. A proposta é focar em um ponto específico [latitude, longitude] (onde está localizada uma plataforma, por exemplo) para fornecer uma previsão mais aproximada. A abordagem consiste em avaliar o comportamento de correntes oceânicas medidas em pontos vizinhos para identificar a existência de influência destas correntes sobre o ponto de interesse, de modo que essa influência se torne um parâmetro para estimar o estado da corrente neste ponto em um momento futuro.

Vale ressaltar que a proposta não é fornecer um modelo baseado no fenômeno físico, utilizando-se de variáveis convencionais como ondas, marés, ventos, pressão atmosférica, dentre outros. Neste trabalho é utilizado o processo de Mineração de Dados para auxiliar na extração das características (vizinhança e instantes de tempo) que melhor contribuem para o ponto de interesse até a geração de um modelo capaz de prever condições futuras com base em informações históricas deste ponto e sua vizinhança. As principais características são:

- Utilização de uma quantidade mínima de variáveis: Não existem parâmetros de informação suficiente ou disponível para aplicar alguma técnica baseada no fenômeno físico. A única variável utilizada é o próprio valor da intensidade.

- Menor tempo de processamento: Preditores que modelam o fenômeno físico necessitam de dias para executar, enquanto que o modelo gerado prevê resultado em poucos minutos.

### 1.3 ESTRUTURA DO TRABALHO

O segundo capítulo caracteriza o embasamento teórico do trabalho, fazendo uma apresentação dos principais conceitos e domínios no qual o trabalho se fundamentou. É feita uma breve introdução às operações offshore, descrevendo as principais estruturas, carregamentos ambientais e seus efeitos. Em seguida, o trabalho discute as correntes oceânicas e no capítulo 3 aborda os conceitos relevantes de Mineração de Dados.

O capítulo 4 delimita o problema estudado, discutindo em mais detalhes a relação das correntes oceânicas no contexto de operações offshore e expondo os requisitos operacionais atrelados a esta relação. O capítulo relata a gravidade do problema, os riscos associados, bem como necessidade e os benefícios da implementação de técnicas que forneçam informações meteo-oceanográficas nas atividades offshore. O capítulo termina com um levantamento dos trabalhos relacionados.

A solução proposta é apresentada em seguida, no quinto capítulo, onde são descritas a metodologia do trabalho, a área do estudo, as ferramentas e recursos utilizados. O capítulo 6 expõe e discute os resultados obtidos, fazendo uma análise crítica dos mesmos. Por fim, o sétimo capítulo conclui o trabalho expondo os resultados atingidos, contribuições e perspectivas de trabalhos futuros.

## Capítulo 2: Embasamento Teórico

### 2.1 OPERAÇÕES OFFSHORE

As reservas petrolíferas no Brasil ficam em localizações marítimas (offshore) e terrestres (onshore). THOMAS (2001) afirma que as primeiras Unidades de Perfuração Marítimas (UPM) utilizavam-se das mesmas técnicas empregadas em terra, consistindo de sondas terrestres montadas sobre uma estrutura para perfurar em águas rasas. No entanto, as reservas marítimas brasileiras ocorrem em lâminas d'água rasas e profundas e, mesmo que no início da atividade esse processo adaptado funcionasse com sucesso, a necessidade de explorar em regiões de águas profundas movimentou investimentos e estudos que deram origem a novas técnicas e tipos de equipamentos voltados para a atividade marítima.

A valorização histórica e atual do petróleo, reforçada pelo alto preço do barril, fizeram com que empresas petrolíferas investissem cada vez mais em exploração e produção, direcionando os esforços para a exploração em profundidades maiores.

Existem vários tipos de estruturas offshore que podem ser utilizadas no processo de exploração de petróleo. Essas estruturas podem ser classificadas basicamente em dois tipos:

- unidades de perfuração marítima com Blowout Preventer (BOP) na superfície: nesta classe estão as plataformas fixas, as auto-eleváveis, as submersíveis e as *tension legs*.

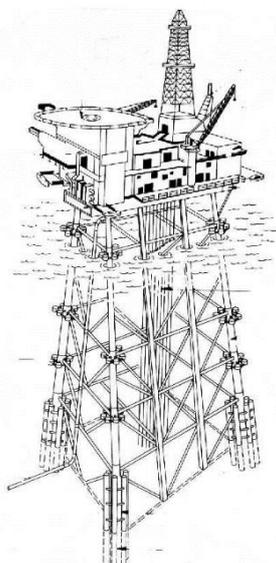
- unidades de perfuração marítima com BOP no fundo do mar: estes tipos de estruturas são conhecidos como unidades flutuantes, a exemplo das semi-submersíveis e navios sonda.

Dos equipamentos de segurança responsáveis pelo fechamento e controle do poço, o BOP é o mais importante, constituindo de um conjunto de válvulas conectado a plataforma por um *riser* de perfuração, que permite fechar o poço, controlar a pressão e monitorar o poço. Também conhecido como preventor, este equipamento é acionado sempre que ocorre o chamado *kick*, fluxo indesejável de fluido de formação para dentro do poço, de modo a impedir que esses fluidos atinjam a superfície de maneira

indesejável. *Kicks* podem levar a um evento potencialmente catastrófico, conhecido como *blowout*, que é o fluxo descontrolado de fluidos para a superfície, provocando a explosão do poço e, por consequência, causando danos graves ao reservatório, ao meio ambiente e colocando em risco a segurança da tripulação.

A escolha da unidade de perfuração é dependente de fatores como lâmina d'água (LAD), condições marítimas e relevo do fundo do mar.

**Plataformas fixas:** preferencialmente utilizadas nos campos caracterizados por lâminas d'água de até 300 metros, são estruturas construídas a partir de membros moduladas de aço e apoiadas no fundo do mar através de estacas cravadas no solo. Estes membros funcionam como uma pilastra, suportando o peso dos equipamentos de operação e as forças ambientais atuantes de ondas, ventos e correntes. Plataformas fixas possuem utilização restrita a campos de exploração comercial conhecida e já comprovada. Tal restrição se deve a elevados custos envolvidos no projeto.



**Figura 1: Plataforma Fixa**

**Plataformas auto-eleváveis (PAs):** Este tipo de estrutura constitui-se de uma balsa flutuadora apoiada por pernas que movimentam-se para baixo até encontrar o fundo do mar. A plataforma é então elevada a uma altura segura, acima do nível da água e sem a ação das ondas. Devido a esse comportamento, plataformas auto-eleváveis também são conhecidas como plataformas **Jack-ups**. Essas plataformas são voltadas principalmente para operações de perfuração exploratória e, em função disso, a característica de mobilidade torna-se um fator importante no projeto das estruturas. Segundo (THOMAS, 2001) trata-se de uma unidade de perfuração com maior histórico

estatístico de acidentes, em função da criticidade e grande influência de condições de tempo e mar sobre a estrutura durante as operações de elevação e abaixamento.



**Figura 2: Plataformas Auto-Eleváveis**

**Plataformas submersíveis:** constitui-se de uma estrutura montada sobre um flutuador. Devido à pequena capacidade de lâmina d'água, sua aplicação é limitada a locais de águas calmas, rios e baías.

**Plataformas Tension Leg (TLP):** este tipo de plataforma possui uma estrutura similar à plataforma semi-submersível, diferindo apenas em suas pernas principais, que são ancoradas no fundo do mar através de cabos tubulares. Estas pernas mantêm-se tracionadas, de modo que elas possuem relativamente elevada rigidez axial (baixa elasticidade), fazendo com que todo o movimento vertical da plataforma seja severamente eliminado. Isso permite que as operações de perfuração e completação funcionem de forma similar às operações realizadas em plataformas fixas.



**Figura 3: Plataforma Tension Leg**

**Plataformas flutuantes:** podem ser semi-submersíveis, compostas de estruturas de um ou mais conveses e apoiadas por colunas em flutuadores submersos; ou navios de sonda, projetados especialmente para perfuração. A ação de ondas, correntes e ventos provocam movimentações na estrutura que podem causar danos nos equipamentos de perfuração. Mesmo com a ação desses agentes, a plataforma mantém sua posição por meio de um sistema de ancoragem é capaz de restaurá-la. Pela grande capacidade de mobilidade este tipo de plataforma é preferido em casos de perfuração de poços exploratórios.

**Navios de Sonda:** constitui-se de uma embarcação projetada para a perfuração no solo marinho que suporta uma plataforma de perfuração em sua porção central, onde posiciona-se a torre e a abertura que permite a passagem da coluna de perfuração. Essas estruturas operam em águas muito profundas e, com frequência, sofrem rigorosas condições marítimas. O posicionamento, apesar de ser considerado bastante crítico (em função de sua natureza estrutural com uma área maior exposta à ação das ondas), dispensa a ancoragem durante a perfuração e possui um sistema composto de sensores acústicos, propulsores e computadores, que anula os efeitos causado por forças ambientais como ondas, ventos e correntes.



**Figura 4: Navio de Sonda**

**Plataformas semi-submersíveis:** são estruturas flutuantes apoiadas em múltiplas pernas e um grande convés localizado acima do nível do mar, de modo que a área operacional se mantém longe da ação das ondas. As pernas são interconectadas no fundo subaquático com membros horizontais flutuantes chamados de pontões. Essas plataformas podem ser movidas de um lugar para o outro, e lastradas para cima ou para baixo, aumentando a quantidade de alagamento nos tanques flutuantes. Vale destacar ainda que tais estruturas possuem boa estabilidade e características *seakeeping*, ou seja, é capaz de operar de maneira eficaz mesmo quando sujeita eventos indesejáveis devido a condições elevadas do mar no qual irá operar.



**Figura 5: Plataforma Semi-submersível**

### 2.1.1 SISTEMAS DE POÇOS SUBMARINOS

De acordo com o posicionamento do BOP, as operações de perfuração de

poços de petróleo no mar são caracterizadas por dois aspectos distintos: o primeiro consiste na perfuração com o BOP localizado na superfície, característica de estruturas como plataformas fixas e auto-eleváveis; o segundo caso compreende a perfuração com BOP no fundo do mar, como ocorre em plataformas flutuantes.

Em sistemas de cabeça de poço para plataformas fixas e auto-eleváveis existe a necessidade de um cabeçal de superfície com função de vedação secundária e de sustentação do peso dos tubos de revestimento localizados acima do fundo do mar.

Em plataformas flutuantes, todo o sistema de cabeça do poço fica situado no fundo do mar. As operações de perfuração por meio de estruturas fixas possuem alto grau de similaridade com a atividade realizada em estruturas terrestres, dado que a coluna de revestimento se estende do fundo do mar até a superfície.

Conforme já falado, as unidades flutuantes estão constantemente sujeitas a movimentações marítimas. Essas unidades são conectadas ao solo marinho através de elementos estruturais, como linhas de ancoragem, tendões, dutos ou umbilicais. Nesse caso, o sistema de cabeça de poço fica situado no fundo do mar e se conecta a superfície através de uma estrutura denominada *riser* de perfuração. O *riser* é um “tubo de aço que conecta uma plataforma de perfuração flutuante com a cabeça de poço, por onde todas as ferramentas de perfuração, revestimento, etc., sejam guiadas até poço e os fluidos de perfuração retornem do poço até a plataforma” (MACLACHLAN, 1999).

Essas tubulações permitem a circulação do óleo extraído, gases, água e detritos sólidos provenientes da perfuração. Os *risers* são um dos elementos mais importantes de qualquer operação de perfuração. Quanto ao tipo de material, os *risers* podem ser classificados como rígidos e flexíveis:

**Rígidos:** São formados, em geral, por sequências de tubos de aço (também são utilizados titânio e compósitos) cujo comprimento é de aproximadamente 12 metros e são unidos através de solda. Em geral, a opção de configuração vertical é a mais utilizada, por ser mais barata. No entanto, em águas profundas, a configuração catenária também é utilizada. Podem ser usados não apenas para a perfuração, como também em atividades de produção, em plataformas que apresentam baixos movimentos.

**Flexíveis:** São dutos compostos por várias camadas intercaladas de materiais e composições; além de uma armadura central espiralada. As diferentes camadas também proporcionam finalidades específicas como vedação, revestimento, estanqueidade e

proteção contra corrosão; sendo também responsáveis por resistir às forças ambientais sofridas. Estes *risers* conseguem contornar algumas limitações em relação aos *risers* rígidos, porém, são mais caros.

Os *risers* possuem uma forma cilíndrica com diâmetros entre 20 e 50 centímetros. Pela geometria esbelta e o grande comprimento, os *risers* se comportam como longas linhas e são continuamente sujeitos ao poder das forças de marés, ondas, correntes e movimentação das plataformas que atuam sobre ele. Essa condição é agravada pelo próprio peso quando ele está completamente preenchido pelo fluido e o comportamento pode ser influenciado pelas inúmeras submissões de operação. Devido à ação desses agentes, os *risers* respondem por meio de comportamentos dinâmicos diversos, podendo oscilar em diferentes frequências, amplitudes, direções e modos.

Diante deste cenário, essas estruturas devem ser resistentes e bem projetadas, especialmente em seus acoplamentos. Segundo MACLACHLAN (1999), o *riser* também deve ser capaz de resistir a grandes pressões externas no caso de o fluido de perfuração escapar completamente por qualquer razão; sendo, portanto, uma parte muito cara do equipamento de sonda que deve ser protegida e usada com muito cuidado. O processo de degradação decorrente de esforços repetitivos representa um dos principais responsáveis por danos em dutos e estruturas offshore, levando a sérias consequências como grandes prejuízos financeiros e/ou ambientais.

A seção seguinte irá tratar das forças atuando sobre as estruturas de perfuração.

### 2.1.2 CARREGAMENTOS

Estruturas offshore, montadas em regiões oceânicas para atividades de exploração e produção, estão sujeitas a condições ambientais impostas pela natureza, que constituem as principais causas para o tempo de inatividade, eficiência operacional reduzida e danos diversos. Essas condições enfrentadas por tais estruturas são as ondas, marés, correntes e ventos; elementos comumente referenciados como carregamentos ambientais. (CHAKRABART, 2005) enquadra esses elementos em uma classe mais específica, chamada de carregamentos meteo-oceanográficos.

Dependendo da função, localização e método de construção, uma plataforma offshore é sujeita a diversos tipos de carregamentos, que vão além das condições marítimas; incluindo carregamentos sísmicos, de convés e equipamentos (desgastes

operacionais), acidentais (erros humanos), explosões e incêndios, dentre outros. Vale destacar que fenômenos mais extremos como terremotos e Tsunami também podem ocorrer em certas regiões. Neste trabalho o foco está nos carregamentos meteo-oceanográficos, cujos efeitos e importância de suas investigações são mais bem discutidas no Capítulo 3.

Dentre os principais fenômenos de degradação dos *risers* e demais estruturas, estão a fadiga estrutural e a Vibração-Induzida por Vórtices (VIV). A fadiga é um fenômeno que pode acontecer quando a estrutura é submetida a carregamentos repetitivos de tensão ou deformação, de modo a surgirem pequenas trincas que podem evoluir culminando na ruptura da estrutura. A análise de fadiga é um assunto amplamente estudado, pois grande parte de falhas em serviço são decorrentes desse processo. Uma causa significativa da fadiga é a Vibração-Induzida por Vórtices devido a correntes marítimas. CUNFF et al. (2002) explica que se uma corrente impacta sobre um cilindro circular, forças flutuantes são criadas devido ao desprendimento do vórtice. Caso o cilindro seja montado sobre molas, as forças induzem a um deslocamento da estrutura na direção perpendicular ao fluxo e a própria estrutura. Vibração-Induzida por Vórtices é o nome dado a esse deslocamento do cilindro.

A intensidade dos carregamentos ambientais sobre as estruturas podem provocar movimentos tão complexos de modo a reduzir o tempo de vida útil, diante de danos provocados por choques entre tubos vizinhos, fraturas por excesso de carga ou fadiga estrutural. Dessa forma, aumenta-se a necessidade de prever com mais precisão possíveis condições ambientais consideradas agressivas para a operação, como também, quaisquer tipos de falhas estruturais e operacionais; visando evitar maiores impactos e prejuízos.

Com o intuito de preservar as estruturas, os efeitos desses carregamentos devem ser conhecidos pelos projetistas e considerados durante seus projetos. A escolha por sondas e equipamentos de manipulação, e o projeto de *risers* e sistemas de ancoragem são extremamente dependentes das condições meteo-oceanográficas previstas, específicas do local, uma vez que os processos oceanográficos variam de acordo com a localização. Entender o efeito desses agentes sobre as respostas das estruturas é de extrema relevância, pois fornece orientações para um melhor detalhamento do planejamento operacional.

Este trabalho foca especificamente nos carregamentos oriundos de correntes marítimas, buscando prever condições críticas de operação através de um modelo preditor de intensidade de corrente. A próxima seção irá discutir a importância das correntes para a indústria offshore.

## 2.2 CORRENTES OCEÂNICAS

As correntes oceânicas correspondem a movimentos horizontais de grandes massas de água que circulam através dos oceanos. Algumas correntes são pequenas e de curta duração, e outras são fluxos extensos que levam séculos para completar o circuito do globo, a exemplo da corrente do Golfo (NOAA OCEAN EXPLORER, 2012). Da mesma forma que a circulação dos ventos, as correntes exercem influência sobre o clima das regiões em que atuam, com intensidades, direções e constâncias bem definidas. Além do clima, a sua influência é considerável tanto para a navegação quanto para a economia geral dos mares.

A formação dessas correntes é resultado de diversos fatores como ventos, ondas, marés, gravidade, e as variações na densidade da água. Os movimentos terrestres, conforme será visto mais adiante, representam outro importante determinante na configuração das correntes.

As correntes oceânicas se dividem em dois sistemas distintos: as **correntes de superfície**, que movimentam uma camada relativamente fina na superfície do mar; e **correntes de águas profundas**, que movimentam ao longo do fundo do alto mar. O primeiro tipo representa cerca de 10% das águas oceânicas e são impulsionadas por grandes cinturões de vento. As correntes de superfície podem ainda ser divididas em **correntes costeiras** e **correntes oceânicas de superfície** (em oceano aberto). Os ventos superficiais são os responsáveis pela direção das correntes superficiais, de modo que as correntes costeiras são afetadas por ventos locais e as correntes oceânicas superficiais configuram-se pelo sistema global de ventos.

As águas oceânicas possuem um movimento contínuo que ocorre em função do movimento de rotação da Terra e da ação dos ventos sobre a superfície oceânica. Ao soprar sobre as marés, os ventos provocam, por meio do atrito, o arrastamento das águas superficiais, originando o surgimento de uma circulação típica, eminentemente horizontal e limitada às primeiras centenas de metros de profundidade (MIGUENS,

1993). Essa camada oceânica na qual os ventos exercem influência é conhecida como **Camada Ekman**.

Mudanças no campo de vento provocam alteração na direção, velocidade e extensão das correntes. A força de fricção que atua na superfície do mar é proporcional à velocidade do vento. Quanto maior a velocidade, maior a força de fricção. Essa força é conhecida como cisalhamento do vento e é resultado do campo do vento na superfície do mar. Apesar de o vento que sopra na superfície oceânica possuir maior efeito sobre a superfície, as camadas inferiores adjacentes também são afetadas, devido ao efeito do cisalhamento que é transmitido para tais camadas através da fricção interna entre elas. Sendo assim, a camada inferior se move mais lentamente em relação às camadas acima. Em função da força de Coriolis (próxima seção), a água é defletida em aproximadamente 45° da direção do vento na superfície com uma velocidade a menos de 3% da velocidade original do vento, aumentando conforme a profundidade é aumentada.

Correntes de águas profundas constituem o segundo sistema de circulação, as quais representam os 90% restantes das águas oceânicas. Conhecidas como correia transmissora mundial, essas correntes são ocasionadas pelas diferenças na densidade de água, em função de diferentes temperaturas e salinidade. Essas são correntes maiores e mais lentas do que as correntes de superfície. Fenômenos como resfriamento da água, excesso de evaporação sobre a precipitação pluvial, formação de gelo e aumento da salinidade das águas, resultam no aumento da densidade.

### **2.2.1 O Efeito Coriolis**

Os ventos não se movimentam em linhas retas. O movimento de rotação da Terra faz com que o ar seja forçado a se desviar para a direita no Hemisfério Norte (sentido horário) e para a esquerda no Hemisfério Sul (sentido anti-horário). Essa curvatura dos ventos denomina-se efeito de Coriolis.

Em ambos os hemisférios, a massa de ar vinda do Equador descende aos 30° de latitude, em função de seu esfriamento. Por outro lado, a camada de ar vinda dos pólos descende aos 60° de latitude, devido ao aquecimento. Originam-se, então, três sistemas de ventos:

- Ventos alísios: ocorrem entre 0° e 30° de latitude e sopram do leste para o oeste.

- Ventos do oeste: ocorrem entre 60° e 30° de latitude e sopram do oeste para o leste.

- Vente do leste: ocorrem nas regiões polares e sopram do leste para o oeste.

Esses sistemas de ventos alísios e do oeste possuem padrões circulares e originam correntes oceânicas conhecidas como **giros**. Existem cinco grandes giros que se estendem pelos oceanos e fluem tanto para o norte quanto para o sul do Equador: Atlântico Norte, Atlântico Sul, Pacífico Norte, Pacífico Sul, Oceano Índico.

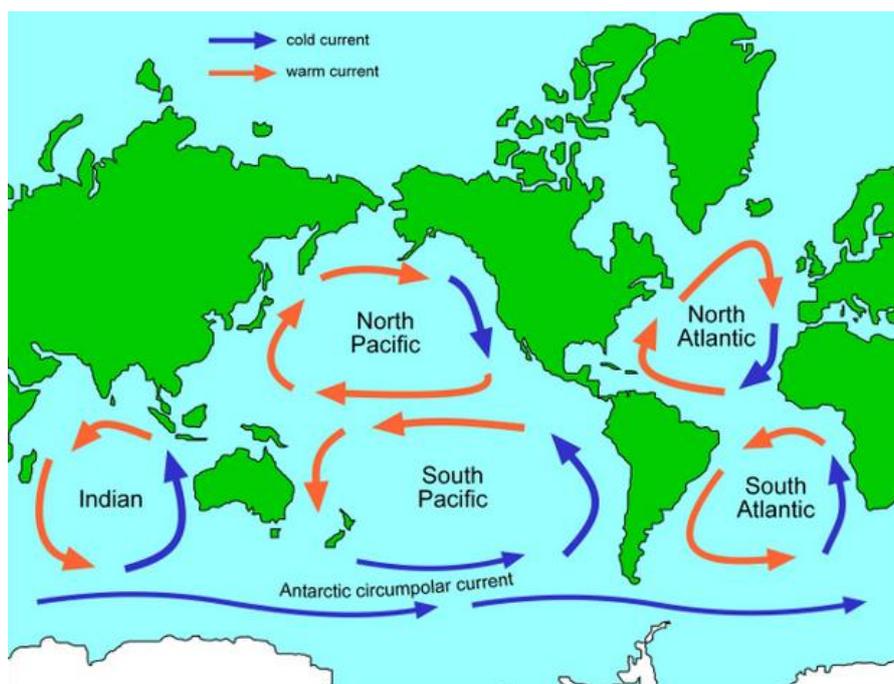


Figura 6: Mapa dos cinco giros oceânicos: Fonte (UNIVERSITY OF WAIKATO, 2012)

### 2.2.2 Circulação Termoalina

Nas regiões polares, gelo marinho descarta o sal durante o congelamento, formando um volume de água altamente concentrada, muito fria e salgada que escoo para o mar através de pequenos canais. Quando atinge o fundo do mar, essa água de elevada densidade é substituída por outras águas, criando uma corrente marinha profunda. A água fria e salgada é então aquecida e misturada gradativamente conforme a corrente se desloca, de modo que a densidade e a salinidade são reduzidas. Como consequência, origina-se uma corrente ascendente, aquecida e de baixa salinidade,

formando uma corrente superficial que completa o espaço deixado pelas águas frias e salgadas.

Esse processo é chamado de circulação termoalina e perpetua um ciclo que envolve todos os oceanos. Correntes dirigidas por circulação termoalina ocorrem nos níveis rasos e profundos dos oceanos e se movimentam mais lentamente que as correntes de superfície.

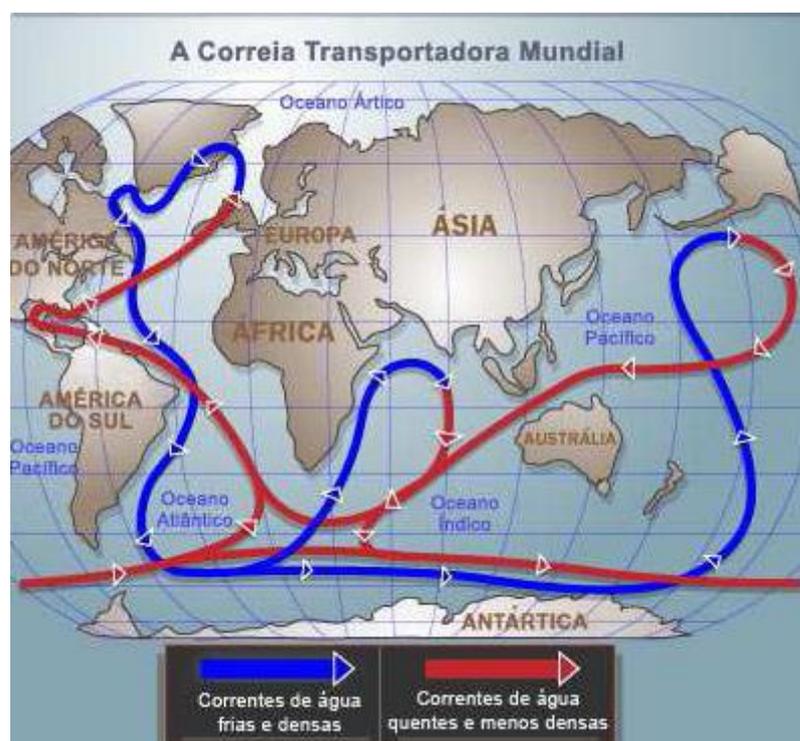


Figura 7: Circulação Termoalina

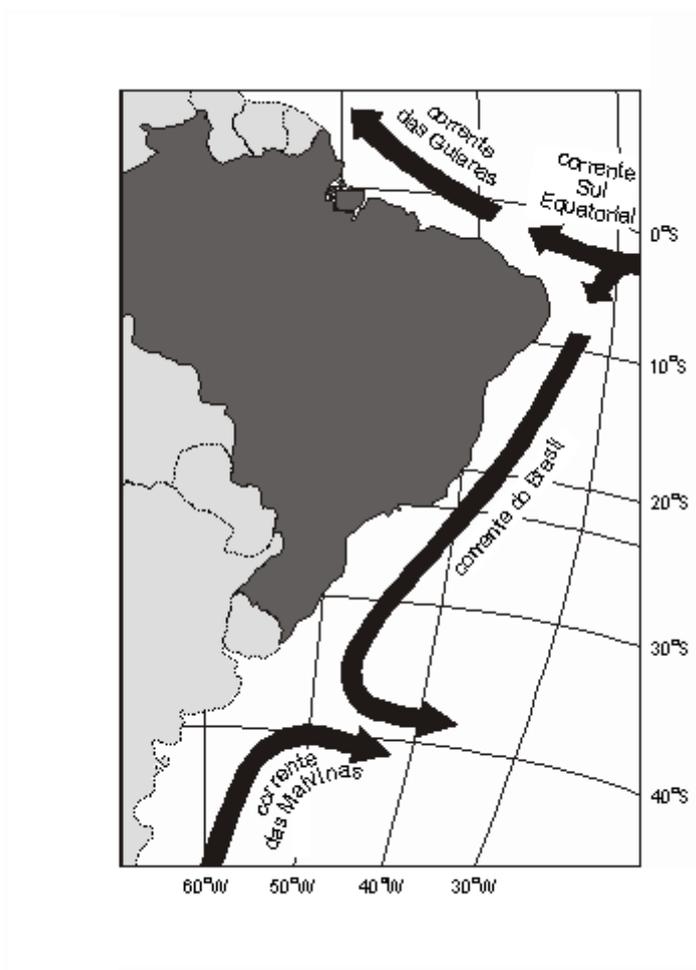
O Brasil sofre a influência de correntes marítimas quentes: a corrente do Brasil (no sentido sul) e corrente da Guiana (no sentido norte), que contribuem para a existência de climas quentes.

### 2.3 Corrente do Brasil

A quantidade de trabalhos realizados durante as últimas décadas em torno das características do Sistema de Correntes do Brasil revelou um aumento no interesse de pesquisadores e estudantes pelo assunto. Estes trabalhos buscam entender aspectos

como as características físicas, perfis de correntes, massas de águas em regiões específicas.

A Corrente Sul Equatorial (CSE) do Oceano Atlântico flui no sentido leste-oeste na altura do equador e, ao encontrar a costa do Nordeste do Brasil, bifurca-se, formando a Corrente do Brasil, conhecida como CB. Conforme ilustrada na Figura 8, a Corrente do Brasil então se volta para o sul, bordejando o continente sul-americano. Por outro lado, a corrente que se desvia para o noroeste, em direção ao Caribe, é denominada Corrente das Guianas.



**Figura 8: Correntes superficiais da Costa brasileira. FONTE: (MENDES, GOMES, 2007)**

A Corrente do Brasil, principal corrente superficial brasileira, é parte de um sistema de correntes de contorno Oeste de superfície que é associada ao Giro Subtropical do Atlântico Sul. Começando ao sul de 10°S, a CB é uma corrente oceânica rasa, quente e salina, do Oceano Atlântico Sul com direção paralela à costa leste a América do Sul. Após cruzar a altura do Rio da Prata, aproximadamente a 38° de

latitude sul, a CB flui para leste e encontra-se com a Corrente das Malvinas, distanciando-se da costa. A região de encontro entre a corrente do Brasil e a corrente das Malvinas é denominada Convergência Subtropical do Atlântico Sul (33-38°S).

A CB tem espessura característica de 500 m ao largo do sudeste brasileiro, ocupa pelo menos três quilômetros de coluna de água e flui no sentido sul, caracterizando sua **direção preferencial**. A Corrente de Contorno Intermediária (CCI) é subjacente a corrente do Brasil e flui para o norte, com uma extensão vertical de pelo menos 700 m. Por fim, a Corrente de Contorno Profunda (CCP), localizada abaixo da CCI, ocupa cerca de dois quilômetros de coluna de água e flui para o sul. Segundo (SILVEIRA, 2007), a CB é caracterizada por uma inversão de sentido das correntes na vertical, cuja região de transição fica entre CB e CCI.

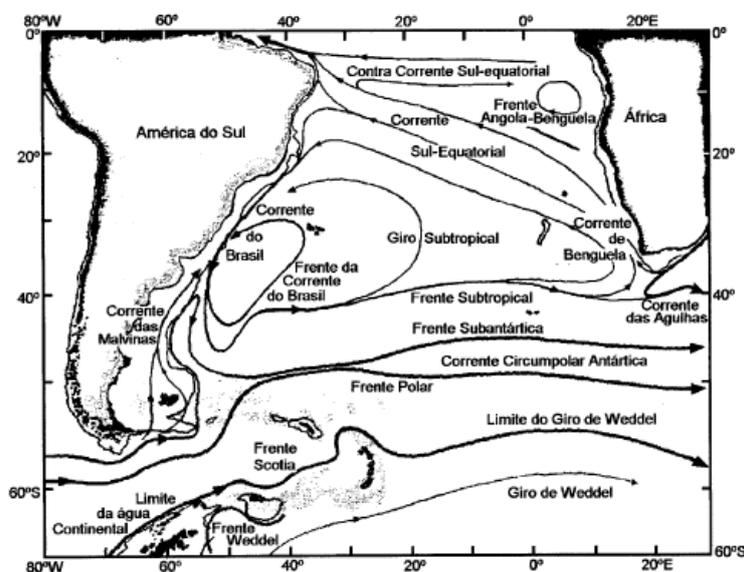


Figura 9: Representação esquemática da circulação superficial de larga escala do Oceano Atlântico Sul. (ASSIREU, 2003)

A região de estudo que ganha destaque é compreendida pela Bacia de Campos, com notável importância econômica em função de grandes descobertas que movimentam a indústria de petróleo brasileiro. Como já relatado anteriormente, as características oceânicas são componentes essenciais no projeto de plataformas e estruturas de extração de petróleo. Diante deste cenário, faz-se necessário um conhecimento da circulação nessa região.

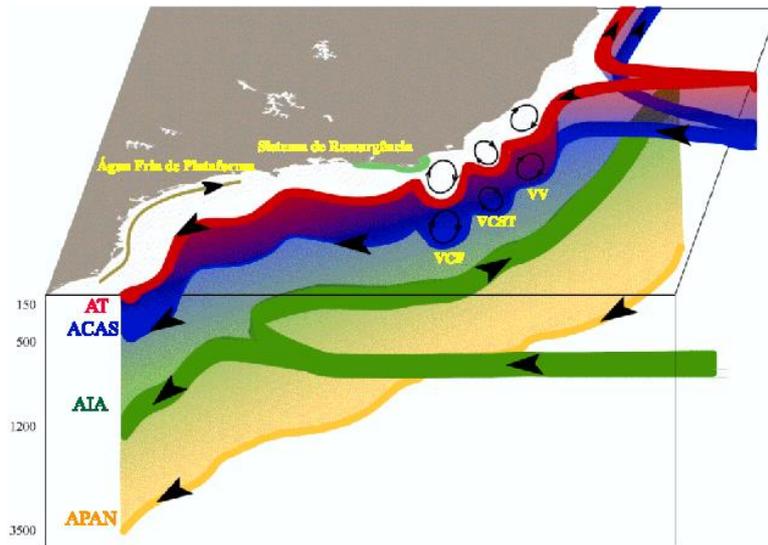
### 2.3.1 Estrutura Vertical de Massas de Água

A Corrente do Brasil é composta pelo empilhamento de massas de água que são características do Atlântico Sul. Nos primeiros 3500m encontram-se a Água Tropical (AT), Água Central do Atlântico Sul (ACAS), Água Intermediária Antártica (AIA), Água Circumpolar Superior (ACS), Água Profunda do Atlântico Norte (APAN) e Água Circumpolar Inferior (ACI) (SILVEIRA, 2007). Como a metodologia deste trabalho concentra dados de correntes localizadas em águas de superfície, os detalhes abordados aqui dizem respeito a AT.

De acordo com (GODOI, 2005), observações e experimentos revelam que a CB transporta a Água Tropical (AT) e a Água Central do Atlântico Sul (ACAS) entre a superfície e níveis picnoclínicos (região em que a densidade da água varia com a profundidade). Em seguida, a CCI transporta predominantemente a Água Intermediária Antártica (AIA), fluindo no sentido norte como uma corrente de contorno de nível intermediário.

A AT caracteriza correntes de águas aquecidas, entre 18°C e 28°C, com fluxo predominantemente para o sul. Em termos de profundidade, essa massa de água está localizada nos primeiros 100 m da coluna d'água. Silveira et.al descreve essa água de superfície como consequência da intensa radiação e excesso de evaporação em relação à precipitação, somando-se a mistura com águas de origem costeira mais frias e de baixa salinidade (fenômeno que se dá durante o trajeto no sentido sul).

(SILVEIRA, 2007) afirma que existe uma discordância na comunidade científica no que diz respeito à determinação dos limites verticais, dos núcleos e das características regionais dessas massas de água. Neste caso, propriedades como temperatura, salinidade, nutrientes (fosfatos, silicatos, nitrato, nitrito, amônio) e oxigênio dissolvido, são utilizados para separar e identificar a massa de água.



**Figura 10. Representação esquemática do Sistema Corrente do Brasil. Fonte: (GODOI, 2005)**

A figura 10, retirada de (GODOI, 2005) é uma representação esquemática do Sistema Corrente do Brasil que mostra o comportamento das correntes de contorno em função das massas de água transportadas por estas correntes.

## Capítulo 3: Mineração de Dados

Os constantes avanços tecnológicos, reforçados pela popularização dos computadores e a emergência dos métodos de gerenciamento de dados eletrônicos trouxeram à tona uma nova sociedade, a “Sociedade da Informação”, que vivencia uma realidade já conhecida por muitos como “Era da Informação”. As inovações que surgem nessa nova era, e mesmo, ao longo da evolução dos sistemas computacionais, afetam consideravelmente as organizações, os negócios e a forma como as pessoas trabalham; abrindo novos espaços, oportunidades e auxiliando na competitividade vantajosa.

Entidades de pequeno e grande porte investem esforços no sentido de capturar, organizar e armazenar dados; sendo apoiadas por sistemas informatizados e bancos de dados poderosos, capazes de lidar com enormes quantidades de informação. Com o passar dos tempos, mais operações são informatizadas e mais dados sobre operações, atividades e desempenho são acumulados, formando um ativo organizacional valioso o suficiente ao ponto de prover informações de tendências, padrões, e apoiar as decisões de negócio. Ainda assim, muitas organizações não usam esses ativos de forma adequada para que sejam extraídos conhecimentos que possam gerar valor agregado ao processo, seja comercial ou científico.

Diante da quantidade de dados armazenados, é praticamente impossível analisá-los manualmente. Muitas vezes a atividade requer a análise simultânea de dezenas de atributos independentes. Por isso, mais uma vez, faz-se necessário um suporte computacional para automatizar esse trabalho e gera-se um interesse crescente por técnicas que consigam analisar os dados, classificá-los, resumirá-los, descobrir tendências e anomalias.

Nesse contexto, Mineração de Dados tem ganhado bastante atenção por oferecer técnicas capazes de extrair informações estratégicas escondidas em grandes bancos de dados. Esse ramo teve início nos anos 80 e emergiu da interseção das áreas de Estatística Clássica, Inteligência Artificial e Aprendizado de Máquina.

Na literatura é possível encontrar diversas definições para Mineração de Dados. Neste trabalho, algumas definições foram destacadas. Friedman (1997), por exemplo, coletou e reuniu algumas definições da literatura: é o processo de extração de informação previamente desconhecida, compreensível e litigável, a partir de grandes

bases de dados e usando-as para tomada de decisões de negócio cruciais (Zeculin); é um conjunto de métodos usados no processo de descoberta de conhecimento para caracterizar relacionamentos e padrões previamente desconhecidos dentro dos dados (Ferruza); é o processo de descoberta de padrões vantajosos dentro dos dados (John); é um processo de suporte a decisão onde procuramos dentro de grandes bases de dados padrões de informações desconhecidos e inesperados (Parsaye).

Turban, Aronson e Liang (2005) definem Mineração de Dados como um processo que usa técnicas de estatística, matemática, inteligência artificial e aprendizado de máquina para extrair e identificar informações úteis e, por consequência, aprendizado a partir de grandes bancos de dados. (HAN; KAMBER; PEI, 2006) se refere a Mineração de Dados como simplesmente a extração ou mineração de conhecimento a partir de grandes volumes de dados.

Vale destacar que a Mineração de Dados é uma etapa particular de um processo maior conhecido como Descoberta de Conhecimento em Base de Dados (do inglês, *Knowledge Discovery in Databases*, KDD), embora os dois termos sejam frequentemente usados como sinônimos e, em alguns ramos, o termo Mineração de Dados seja mais popular.

O conceito de Descoberta de Conhecimento foi proposto por (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996a, 1996b) como “um processo não-trivial de identificação válida, nova, potencialmente útil e, finalmente, padrões compreensíveis nos dados”. Fayyad, Piatetsky-Shapiro e Smyth (1996b) diferencia KDD de Mineração de Dados, afirmando que este último “é um passo dentro do processo de KDD, consistindo da aplicação de análise de dados e algoritmos de descoberta que, sobre limitações de eficiência computacional aceitáveis, produz uma quantidade particular de padrões (ou modelos) sobre o dado.”

O processo KDD consiste de uma sequência de passos iterativos. Em (HAN; KAMBER; PEI, 2006), esse processo é composto por 7 etapas, como mostra a Figura 11, descritas a seguir:

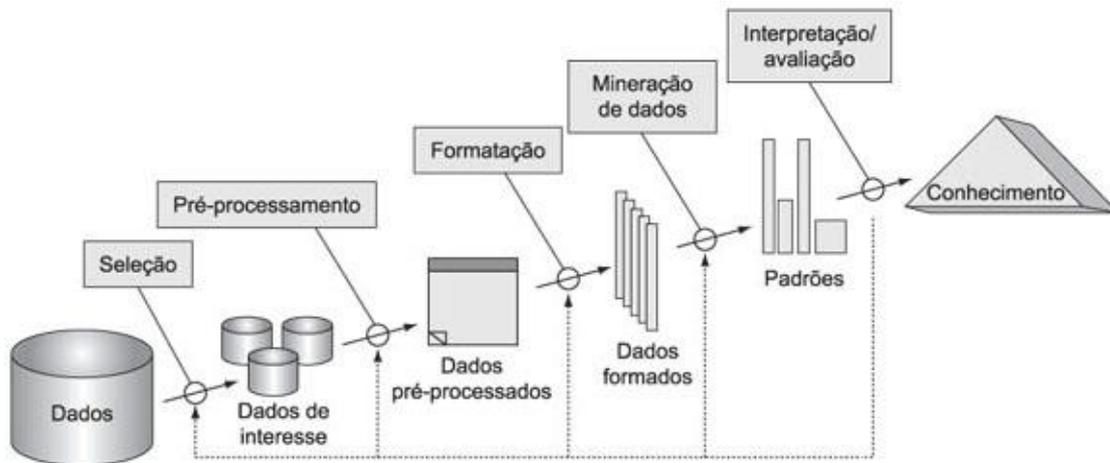


Figura 11: Processo KDD. FONTE: (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996a, 1996b)

1. **Limpeza de Dados:** etapa que consiste na eliminação de ruídos e inconsistência nos dados.
2. **Integração de Dados:** nesta etapa, os dados de múltiplas fontes de dados podem ser combinados.
3. **Seleção:** os dados relevantes para a tarefa de são capturados.
4. **Transformação:** essa etapa prevê a transformação ou consolidação dos dados para uma formatação apropriada para a técnica de mineração a ser adotada.
5. **Mineração:** trata-se da etapa essencial do processo, onde técnicas inteligentes são aplicadas com o objetivo de extrair padrões de interesse dentro dos dados.
6. **Avaliação:** essa etapa visa identificar os padrões realmente interessantes
7. **Apresentação:** por fim, essa é a etapa onde a visualização e as técnicas de representação do conhecimento podem ser usadas para apresentar o conhecimento extraído ao usuário.

Os dados sobre os quais as técnicas de mineração de dados são aplicadas derivam de repositórios de dados variados, incluindo bases de dados relacionais, transacionais e não-convencionais (temporal, geográfico, multimídia), data warehouses, arquivos de texto, web, e outros. Ainda de acordo com (HAN; KAMBER; PEI, 2006),

as etapas de 1 a 4 são diferentes formas de pré-processamento, onde o dado é preparado para a mineração.

O processo também inclui a decisão se um modelo reflete ou não um conhecimento útil. Essa tarefa geralmente requer o julgamento humano.

### **3.1 DECOMPOSIÇÃO EM VALORES SINGULARES**

Dentre os diversos métodos existentes para a manipulação de dados em formato matricial cujo objetivo é reduzir sua dimensão, pode-se destacar a Decomposição em Valores Singulares (SVD, Singular Value Decomposition). Assim como Mineração de Dados, outras áreas como Redes Sociais, Bioinformática e Recuperação de Informações possuem interesse nas técnicas de SVD como um meio de aprimorar seus resultados. No caso de Mineração de Dados, o SVD é amplamente aplicado para fins de redução de dimensionalidade.

Dimensionalidade refere-se ao número de atributos que compõem a representação de uma amostra. Técnicas de redução de dimensionalidade possuem como objetivo a representação de um conjunto de dados de dimensão  $D$  em um novo conjunto  $D'$  com dimensão menor, porém, com as características mais relevantes do conjunto original preservadas. Um conjunto de dimensões não muito grande proporciona maior rapidez nos algoritmos de mineração, menos consumo de memória, e melhor precisão.

O método SVD é uma transformação ortogonal que permite aproximar um conjunto de dados com tão poucos coeficientes possível, que são aqueles que mais influenciam dentro do domínio de aplicação. De acordo com (BAKER, 2005), SVD pode ser visto a partir de três pontos de vista, que são mutuamente compatíveis: por um lado, é visto como um método para transformar variáveis correlacionadas em um conjunto de variáveis não correlacionadas que melhor expõem as diversas relações entre os itens de dados originais; em uma segunda análise, o SVD permite identificar e ordenar as dimensões ao longo das quais os pontos de dados apresentam uma variação maior; este segundo ponto de vista concorda com o terceiro, em que uma vez que foi identificada a maior parte da variação, é possível encontrar a melhor aproximação dos pontos de dados originais usando menos dimensões. Essa análise confirma a aplicação do método como um meio para a redução de dados.

Formalmente, a decomposição de valores singulares de uma matriz  $A_{m \times n}$  (real ou complexa) é uma fatoração que consiste em representar esta mesma matriz na forma:

$$A = U\Sigma V^t \quad (1)$$

Onde,

- $U$  é uma matriz  $m \times m$  ortogonal, ou seja,  $UU^t = I_m$
- $V^t$  é uma matriz  $n \times n$  ortogonal, ou seja,  $VV^t = I_n$
- $\Sigma$  é uma matriz  $m \times n$  no formato  $\begin{pmatrix} S & 0 \\ 0 & 0 \end{pmatrix}$ , que contém em sua diagonal principal os *valores singulares* em ordem decrescente, onde  $S = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r)$  com  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$  para  $1 \leq r \leq \min(m, n)$
- $r$  é chamado de posto da matriz

Como exemplo de uma decomposição, considere a matriz  $4 \times 3$  a seguir:

$$A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -4 & 0 \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{pmatrix}$$

O cálculo da decomposição em valores singulares da matriz  $A$  resulta nas matrizes abaixo:

$$U = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

$$V^t = \begin{pmatrix} 0 & 0 & 1 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

### 3.1.1 REDUÇÃO VIA SVD

Considere a decomposição matricial  $A = U\Sigma V^t$ . Conforme visto anteriormente, a matriz diagonal  $\Sigma$  possui os valores singulares, que determinam o posto da matriz original. O conceito da redução via SVD consiste em obter o menor valor para o posto de modo a aproximar a matriz original. Isso permite excluir algumas dimensões da matriz original, transformando-a em uma matriz otimizada.

Os valores singulares expressam a relevância de uma determinada informação na matriz. Aqueles valores singulares que são muito pequenos não contribuem significativamente para a informação como um todo. Por isso, esses valores pequenos podem ser substituídos por zero. Ao fazer essa substituição, as últimas  $r-k$  linhas e colunas de  $\Sigma$  (a partir do primeiro valor nulo na diagonal) são eliminadas. Logo, a matriz  $\Sigma$  passa a ter posto  $k$ , sendo  $k < r$  o número de valores singulares que se deseja manter (quanto maior for  $k$ , melhor a qualidade, pois se aproxima da representação original).

Por consequência, quando a matriz diagonal é multiplicada por  $U$  e  $V^t$ , colunas de  $U$  e linhas de  $V^t$  também são eliminadas, devido à multiplicação por zero. O resultado final serão as matrizes  $U_k$ ,  $\Sigma_k$  e  $V_k^t$ . A multiplicação destas formará a matriz  $A_k$ , que é a melhor aproximação  $k$ -posto de  $A$ .

Desta forma, se  $A$  tem posto  $r$ , isto significa que  $r$  colunas de  $A$  são linearmente independentes. Sendo assim,  $U$  é  $m \times r$ ,  $\Sigma$  é uma matriz diagonal  $r \times r$  e  $V^t$  é  $r \times n$ , Figura 12.

$$\begin{matrix} & n \\ & \boxed{A} \\ m & \end{matrix} = \begin{matrix} & r \\ & \boxed{U} \\ m & \end{matrix} \begin{matrix} r \\ \boxed{\Sigma} \\ r \end{matrix} \begin{matrix} & n \\ & \boxed{V^T} \\ r & \end{matrix}$$

Figura 12: SVD reduzida

O objetivo dessa redução está na diminuição do custo computacional. Conforme será visto mais adiante, neste trabalho a SVD foi utilizada como técnica para

redução da dimensão do conjunto de dados.

### 3.2 TAREFAS DE MINERAÇÃO DE DADOS

Segundo (HAN; KAMBER; PEI, 2006), de modo geral, as tarefas de mineração de dados pode ser classificadas em duas categorias:

- **Tarefas Descritivas:** Caracterizam as propriedades gerais dos dados dentro de um banco de dados (HAN; KAMBER; PEI, 2006), gerando padrões compreensíveis pelo homem que descrevem esses dados. A investigação feita nesta categoria visa descrever fatos relevantes, não-triviais e desconhecidos. Métodos como *Associação*, *Agrupamento* e *Sumarização*, descritos no próximo tópico, se encontram nessa categoria.
- **Tarefas Preditivas:** Realizam inferência nos dados com o objetivo de fazer previsões. Neste caso, o sistema encontra padrões com o propósito de prever o comportamento futuro de algumas entidades (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996b). Esses padrões são obtidos a partir de algoritmos de mineração que aprendem a partir de exemplos. Nesta categoria estão os métodos de *Classificação*, *Predição* e *Regressão*, também detalhados mais a diante.

Ao desempenhar as tarefas descritas acima, a máquina realiza um processo de aprendizagem chamado Inferência Indutiva, no qual uma conclusão generalizada sobre os membros de uma classe é extraída a partir do exame de uma pequena amostra de membros dessa classe. De forma simplificada, a indução reflete um raciocínio que parte do particular para o geral.

### 3.3 MÉTODOS DE MINERAÇÃO DE DADOS

Um método de mineração refere-se ao que se pretende pesquisar nos dados, ou seja, qual tipo de padrão tem-se interesse em encontrar em uma tarefa de mineração. A literatura disponibiliza diversos métodos (ou classes) de mineração de dados, que podem possuir propósitos distintos, a depender do resultado pretendido. Há casos, porém, em que métodos com propósitos diferentes podem ser aplicados para alcançar resultado final comum. Para atingir o resultado desejado, ou seja, descobrir

conhecimento relevante, faz-se necessário que as metas sejam definidas a fim de direcionar o problema ao(s) método(s) mais adequado(s).

Diversos autores tratam os métodos de mineração de formas diferentes. É possível encontrar diferentes categorias de métodos, baseados nos diferentes tipos de conhecimento a serem minerados; e nomeações distintas para se referir ao mesmo método. Este trabalho lista abaixo os métodos mais tradicionais utilizados para abranger as metas de mineração.

### 3.3.1 PREVISÃO

Submetendo uma amostra a um modelo preditivo, baseado em padrões dentro de um grande conjunto de dados, prevê um valor futuro de um atributo específico (saída) dessa amostra. Trata-se de uma forma de estimativa.

### 3.3.2 REGRESSÃO

Consiste na análise de dependência de valores de atributos numéricos de saída de uma dada amostra, em função dos valores de outros atributos da mesma amostra, gerando um modelo de regressão com base na análise de várias amostras. Submetendo uma nova amostra a esse modelo, ele é capaz de realizar a previsão de valores numéricos contínuos para os atributos de saída. Assim como a previsão, também é uma forma de estimativa.

Neste trabalho, o método de regressão é utilizado para realizar as previsões de valores futuros de intensidade de corrente oceânica.

### 3.3.3 CLASSIFICAÇÃO

De posse de um conjunto de classes discretas pré-definidas, um modelo de classificação determina a qual das classes uma amostra desconhecida pertence. Métodos de classificação envolvem a avaliação de um conjunto de amostras para as quais os rótulos da classe associada são conhecidos. O objetivo é encontrar um modelo que seja capaz de inferir rótulos categóricos de novas amostras. Em um banco, por exemplo, um

modelo classificador pode inferir se um dado cliente oferece risco ou não para a obtenção de empréstimo pessoal.

Este método também é utilizado no presente trabalho no sentido de prever qual a classe futura de condição de corrente de acordo que as condições atuais e históricas.

### 3.3.4 AGRUPAMENTO

Métodos de agrupamento visam segmentar um conjunto de dados em subgrupos ou clusters, de modo que cada amostra é muito semelhante as demais dentro do mesmo subgrupo em relação a algum critério. As amostras são agrupadas em função da similaridade de seus atributos. Sendo assim, ao submeter um conjunto de amostras a um modelo, ele deve identificar grupos de amostras que compartilham uma determinada característica, sem consultar um rótulo de classe. Essa abordagem difere da classificação no sentido de não existirem classes pré-definidas caracterizando as amostras. Além de encontrar grupos de itens similares, agrupamentos também podem servir para gerar rótulos dessas classes desconhecidas. Em geral, depois que os clusters são identificados, uma análise deve ser feita para entender o significado de cada cluster no domínio do problema estudado.

### 3.3.5 ASSOCIAÇÃO

Identifica relacionamentos entre eventos que ocorrem em um dado momento (TURBAN; ARONSON; LIANG, 2005). Dado um conjunto de amostras, métodos de associação identificam relações entre atributos dentro de uma amostra e relações entre atributos e amostras, na forma de uma regra de associação do tipo *se X então Y*, ou *X implica Y* ( $X$  e  $Y$  são conjuntos de itens); de modo que é possível dizer que a presença de um padrão implica na presença de outro padrão. A ideia é derivar correlações que apoiem na tomada de decisão, satisfazendo medidas estatísticas de interesse. A esse tipo de abordagem são normalmente dirigidas as classes de problemas do tipo de análise de cestas de vendas, onde o foco está na identificação de preferências e afinidades para compor estrategicamente pacotes direcionados a consumidores e melhorar a disposição dos produtos. Um exemplo clássico é a associação entre produtos comprados por cliente numa mesma compra (dos clientes que compram leite, 80% também compram pão).

### 3.3.6 ANÁLISE DE SEQUÊNCIA (OU SÉRIES TEMPORAIS)

Este método tem por objetivo a modelagem do estado do processo para gerar a sequência ou para extrair e registrar desvios e tendências no tempo (FAYYAD; PIATETSKY-SHAPIO; SMYTH, 1996b). É um método similar à associação, diferenciando-se pelo fato de a investigação de relacionamento ocorrer sobre um período de tempo. O modelo é o resultado da identificação de padrões que indicam quais tipos de sequências podem ser formadas. Um padrão sequencial – representado por  $\langle I_1, \dots, I_n \rangle$  (sendo  $I_i$  um conjunto de itens, ordenados cronologicamente em relação ao acontecimento dos fatos representados por estes conjuntos) – pode ser exemplificado da seguinte forma: o cliente que compra um produto X, tempos depois compra o produto Y e Z.

Além dos 6 métodos descritos brevemente acima, existem ainda outros métodos baseados em técnicas avançadas de Inteligência Artificial como Algoritmos Genéticos, Programação Evolucionária, Lógica Fuzzy, e outros. Cada método de mineração carrega um conjunto de abordagens algorítmicas (técnicas) para a extração de padrões dentro dos dados.

## 3.4 MEDIÇÕES

Após a aplicação de um método de mineração a uma base de treinamento, o classificador (ou preditor) precisa ser submetido a um processo de avaliação cujo o objetivo é verificar o quão bem ele se desempenha em relação aos dados reais. Ou seja, deseja-se saber o quão preciso um modelo é ao prever o comportamento de dados reais. É importante que o modelo seja validado, analisando sua qualidade, antes da adoção do mesmo para que seja aplicado a novos dados em um domínio real.

Existe uma variedade de abordagens para avaliar a qualidade de um modelo. Em geral, a base de dados é separada em conjuntos de treinamento e testes para avaliar a precisão das previsões, com base na utilização de medidas estatísticas. Quando possível, é importante consultar os especialistas de negócio para que estes, através de suas experiências, avaliem os resultados e determinem o grau de qualidade das previsões ou se padrões descobertos possuem significado.

Nas próximas sessões serão apresentadas algumas medidas utilizadas para avaliar o desempenho de um modelo.

### 3.4.1 MEDIDAS DE PRECISÃO PARA CLASSIFICADORES

Modelos de classificação são validados através da sua aplicação aos dados de teste cujas classes são conhecidas. É realizada então a comparação entre as saídas previstas e as saídas conhecidas. A utilização da mesma massa de dados como conjunto de treinamento para gerar o classificador ou preditor e como conjunto de teste para estimar a precisão do modelo pode resultar em estimativas enganosas superestimadas, em função da superespecialização do algoritmo de aprendizagem (HAN; KAMBER; PEI, 2006).

Precisão (ou acurácia) é uma medida que refere-se ao quão bem o modelo correlaciona um resultado com os atributos dos dados que foram fornecidos. No caso de classificadores, precisão é a porcentagem de previsões corretas feitas pelo modelo, quando comparadas com as classificações reais dos dados de teste.

Uma medida comumente utilizada é a **taxa de erro**, também chamada de **taxa de classificação incorreta**. Sendo  $n$  o número de exemplos do conjunto de teste e  $nc$  o número de exemplos classificados corretamente pelo modelo indutor  $h$ , a taxa de erro é calculada pela equação (2). A precisão, portanto, é calculada simplesmente pelo complemento da taxa de erro, pela equação (3).

$$erro(h) = \frac{nc}{n} \quad (2)$$

$$precisão(h) = 1 - erro(h) \quad (3)$$

Apesar de muito utilizada, esta abordagem de estimativa de precisão é útil quando o domínio do problema trata de classes binárias (ex.: falso, verdadeiro). Quando o classificador é utilizado em um domínio com um número maior de classes, a abordagem denominada **matriz de confusão** torna-se uma ferramenta útil para avaliar a precisão. Sendo  $m$  o número de classes, a matriz de confusão consiste em uma tabela de dimensão  $m \times m$ , onde cada coluna representa as instâncias na classe prevista, enquanto que cada linha representa as instâncias na classe real.

A Tabela 1 demonstra a matriz de confusão para um problema de duas classes. As duas células diagonais fornecem as classificações corretas, ou seja, o valor da classe prevista foi igual ao valor observado. A célula superior direita consiste no número de amostras cuja classe é 1, porém foram classificadas incorretamente como 0. Finalmente, a célula inferior direita contém o número de amostras cuja classe é 0 classificadas incorretamente como 1.

**Tabela 1: Matriz de Confusão**

Classe Observada	Classe Prevista	
	C <sub>0</sub>	C <sub>1</sub>
C <sub>0</sub>	C <sub>0,0</sub> = Número de casos C <sub>0</sub> classificados corretamente	C <sub>0,1</sub> = Número de casos C <sub>0</sub> classificados incorretamente como C <sub>1</sub>
C <sub>1</sub>	C <sub>1,0</sub> = Número de casos C <sub>1</sub> classificados incorretamente como incorretamente C <sub>0</sub>	C <sub>1,1</sub> = Número de casos C <sub>1</sub> classificados corretamente

A partir da matriz de confusão é possível extrair a taxa de erro, aplicando a fórmula abaixo:

$$erro(h) = \frac{C_{0,1} + C_{1,0}}{n} \quad (4)$$

Quando problema envolve várias classes, a precisão pode ser medida aplicando o somatório dos valores da diagonal da matriz, dividido pelo número total de amostras.

### 3.4.2 MEDIDAS DE PRECISÃO PARA PREDITORES

No caso de modelos de predição, a saída retornada é um valor contínuo, de modo que dificilmente o valor predito será exatamente o valor esperado. (HAN; KAMBER; PEI, 2006) define **funções de perda** com sendo meios para a medição do erro entre a saída esperada e a saída prevista. As funções mais comuns são o **erro absoluto** e o **erro quadrado**, descritos pelas equações (5) e (6), respectivamente.

$$erro\ absoluto = |y_i - y'_i| \quad (5)$$

$$erro\ quadrado = (y_i - y'_i)^2 \quad (6)$$

onde  $y_i$  é o valor observado para a  $i$ -ésima amostra e  $y_i'$  é o valor previsto pelo modelo para a  $i$ -ésima amostra.

A taxa de erro é comumente medida pelo **erro quadrático médio** (do inglês, Mean Squared Error - MSE) e **erro médio absoluto** (do inglês, Mean Absolute Deviation - MAE). Esses erros são dados pelas equações abaixo:

$$MSE = \frac{\sum_{i=1}^n (y_i - y_i')^2}{n} \quad (7)$$

$$MAE = \frac{\sum_{i=1}^n |y_i - y_i'|}{n} \quad (8)$$

onde  $n$  é o número de amostras.

Diferentemente do MAE, o MSE exagera perante a presença de “outliers”. A raiz quadrada do erro quadrático médio define uma outra medida, chamada de **raiz do erro médio quadrático** (Root Mean Square Error - RMSE). Esta é uma medida das diferenças entre os valores previstos e os valores observados. Essas diferenças, chamadas de resíduos, são agregadas pelo RMSE em uma única medida. A aplicação da raiz quadrada é útil pois apresenta os valores do erro na mesma ordem de grandeza da variável analisada.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_i')^2} \quad (9)$$

### 3.5 REDES NEURAIS

Redes neurais artificiais são abstrações inspiradas nas compreensões iniciais sobre a estrutura e funcionamento do cérebro humano, implementadas computacionalmente com o objetivo de exercer a habilidade de aprendizado. Foram inicialmente propostas em 1943 por McCulloch e Pitts como um modelo computacional para o neurônio artificial – ainda sem capacidade de aprendizado – e evoluíram ao longo dos anos. Porém, somente nos anos 80 foi que o conceito se desenvolveu para o uso com computadores digitais, tornando-se popular entre pesquisadores e profissionais. A aplicação do conceito tem sido realizada com sucesso na resolução de problemas

reais em diversas áreas. Atualmente, as redes neurais são tratadas como uma ferramenta padrão para mineração de dados e são usadas em muitas tarefas de mineração, como classificações, previsões, análise de séries temporais e agrupamentos (MAIMON; ROKACH, 2010). Segundo (MAIMON; ROKACH, 2010), a maioria dos pacotes de software comerciais de mineração de dados incluem as redes neurais como o módulo núcleo.

Ainda que muito limitadas em tamanho, habilidade e potência em comparação às redes neurais biológicas, as redes neurais artificiais compartilham duas características bastante importantes: processamento paralelo da informação; aprendizado e generalização por experiência. Para (PYLE, 2003), mais do que algoritmos de regressão não-linear, as redes neurais, na verdade, não desempenham regressão da mesma forma que trabalham as técnicas de regressão desenvolvidas estatisticamente; e uma vez configuradas adequadamente, elas possuem o potencial para caracterizar relações não-lineares altamente complexas e complicadas, que são realmente difíceis de descobrir através de outras técnicas de regressão não-lineares.

Dentre as aplicações gerais pode-se citar o reconhecimento de padrões, classificação de padrões, previsão de séries temporais, aproximação de funções e suporte a decisão.

Nas redes neurais naturais, os neurônios recebem impulsos elétricos de outros neurônios. Esses impulsos são acumulados e processados até atingir um limiar de ação. Neste momento, o neurônio dispara esse impulso para o neurônio adjacente.

Uma rede neural artificial é composta por um conjunto de neurônios que são conectados uns aos outros formando uma arquitetura ou estrutura de processamento. Cada neurônio desempenha um processamento de informação, no qual as entradas recebidas são convertidas em saídas processadas. Essa rede possui variáveis de entrada ( $X_i$ ) que se conectam a neurônios vizinhos. Cada conexão entre neurônios tem um peso associado  $W_i$ , que é um valor numérico entre -1 e 1, e representa a força de interconexão entre 2 neurônios.

Cada neurônio pode processar uma entrada e transformá-la em uma saída. O valor da saída é calculado através da aplicação de uma função de limiar representada pela composição de dois módulos: a) junção aditiva, que é o somatório dos sinais de entrada multiplicados pelos pesos associados, e b) aplicação da função de ativação, que

definirá a saída de acordo no cálculo anterior. O processo de ativação é representado por uma função matemática, linear, não-linear ou logística. Mediante esse processo, o neurônio pode ser ativado e envia o sinal de saída que será propagado pelos demais neurônios.

A Figura 12, mostra a arquitetura de uma rede neural de três camadas: camada de entrada, camada escondida e camada de saída. A arquitetura forma uma rede na qual cada variável de entrada se conecta a uma ou mais saídas (modelo Single-layer Perceptron - SLP) ou, no caso mais interessante, uma camada intermediária (também conhecida como camada escondida) é intercalada entre os nós de entrada e saída (modelo Multi-layer Perceptron - MLP). Na camada de entrada estão os nós da primeira coluna, correspondendo aos atributos do conjunto de treinamento. Os neurônios na camada de entrada não processam informação, sendo usados para a recepção dos padrões de dados e transmissão para os neurônios da camada seguinte. A camada de saída armazena os nós que correspondem às classes do conjunto de dados. Na camada intermediária estão os neurônios essenciais para aprender os padrões nos dados e mapear as variáveis de entrada para as variáveis de saída (MAIMON; ROKACH, 2010).

A existência de nós em uma camada intermediária fornece ao modelo a capacidade de tratar problemas não-lineares, uma vez que é possível modelar relacionamentos não-lineares entre os nós de entrada e de saída. Quanto ao número de camadas intermediárias, dificilmente o número de camadas excede em duas. Modelos com duas camadas são usados em problemas muito complexos. É mais comum a utilização de apenas uma camada. De acordo com (CYBENKO, 1989), dado um número suficiente de nodos escondidos, uma MLP pode aproximar qualquer função contínua.

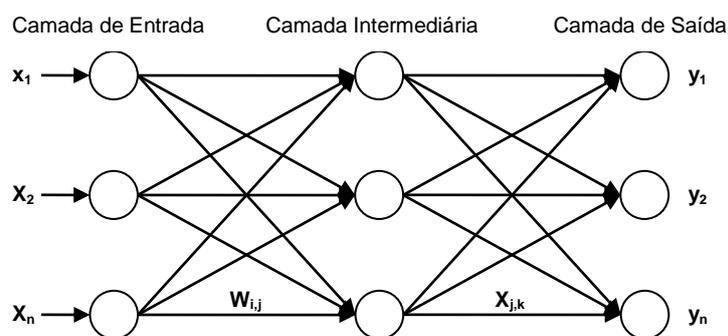


Figura 13: Arquitetura de uma rede neural MLP

A arquitetura pode conter apenas 1 nó de saída, configurando-se em uma regressão ou classificação binária, ou pode ser construída com múltiplos nós de saída. O processo de aprendizado consiste no treinamento da rede a partir de um conjunto de amostras reais já conhecidas, onde os pesos das conexões são ajustados repetidas vezes de acordo com os padrões apresentados. O processo mais comum, chamado *backpropagation*, faz o ajuste dos pesos para os casos classificados incorretamente, tendo como base a magnitude do erro de predição. Trata-se de um processo adaptativo que re-treina iterativamente o modelo, de forma a melhorar o desempenho e a acurácia na predição.

Os pesos das conexões são os parâmetros de uma rede neural e estes parâmetros precisam ser estimados antes que o modelo seja adotado para utilização posterior. Por isso, o treinamento de uma rede consiste em um processo onde os pesos são determinados e representa o mecanismo pelo qual a rede aprende (MAIMON; ROKACH, 2010). O processo de treinamento é descrito da seguinte maneira:

1. A rede é alimentada com os exemplos da base de treinamento
2. Os pesos para cada conexão são carregados aleatoriamente
3. Cada padrão é lido e os valores de cada nó são calculados como a soma das entradas multiplicadas pelos seus pesos.
4. Esses valores são transmitidos através da função de ativação especificada para a saída do nó, se tornando entrada para os nós da camada seguinte
5. Os valores de saída são comparados com os valores de saída desejados para determinar o erro
6. Os pesos das conexões são ajustados para que a rede produza uma melhor aproximação para a saída desejada

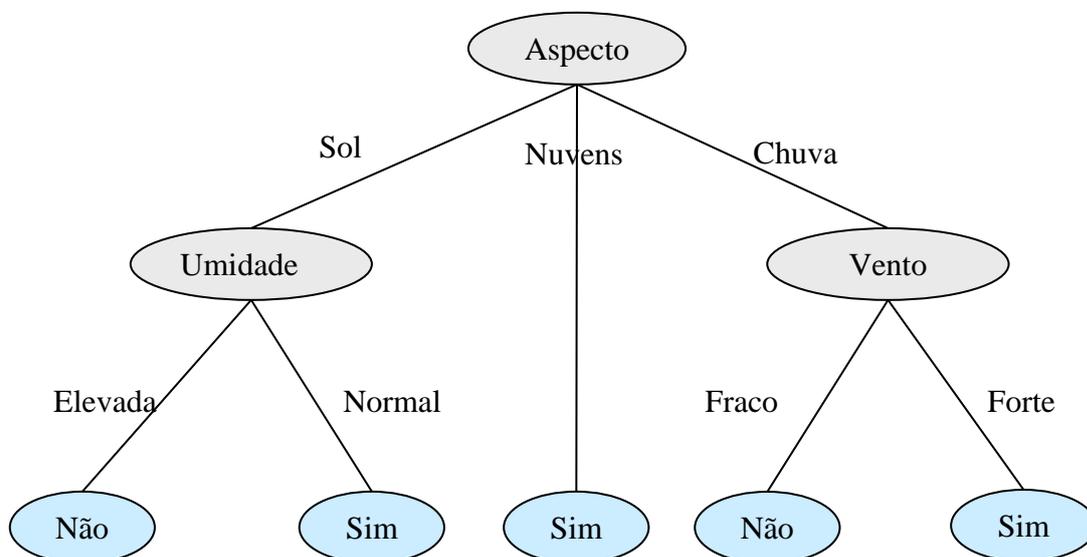
O processo se repete até que os erros calculados seja tão pequenos quanto possível.

### **3.6 ÁRVORES DE DECISÃO**

Árvores de Decisão é uma das técnicas de aprendizagem mais conhecidas, que permite representar um conjunto de regras que direcionam a uma classe ou valor. Trata-

se de uma árvore direcionada com um nó chamado de "raiz", sem arestas de entrada, seguido de nós internos interligados por arestas e os nós terminais, chamados de folhas. Ou seja, a árvore é formada por um conjunto de regras, onde cada nó não terminal representa um teste ou decisão sobre um atributo, as arestas cobrem todas as possíveis respostas para a decisão, e as folhas representam a classe ou rótulo. De acordo com (MAIMON; ROKACH, 2010), árvore de decisão consiste em um classificador expresso como uma partição recursiva do espaço instância. Cada nó interno divide a amostra em um ou mais sub-espacos, de acordo com a discretização de valores para o atributo (MAIMON; ROKACH, 2010).

Dada uma instância cujo rótulo é desconhecido, os valores dos atributos são testados sobre a árvore de decisão e a instância é classificada percorrendo a árvore em um caminho da raiz até a folha, que leva a classe para a tupla. No exemplo abaixo, cada nó é rotulado com o atributo e seus ramos são rotulados com seus valores correspondentes do atributo testado.



**Figura 14: Árvore de Decisão para o Problema Jogar Tênis**

É possível notar que cada caminho pode ser convertido em uma regra do tipo SE - ENTÃO, onde os nós internos e os valores formam o antecedente (SE), e a folha forma o conseqüente (ENTÃO). Desta forma, um dos caminhos representados pela árvore acima pode ser escrito pela regra:

SE Aspecto = Sol e Umidade = Elevada

ENTÃO JogarTênis = Não

A estrutura da árvore é determinada por meio do aprendizado. Algoritmos indutores de árvores de decisão constroem automaticamente uma árvore através de um conjunto de dados de treinamento. As regras são encontradas recursivamente através de valores de atributos. O objetivo é encontrar a árvore de decisão ótima que minimiza o erro de generalização. Existem diversos algoritmos como o ID3 (QUINLAN, 1986), C4.5 (QUINLAN, 1993), CART (BREIMAN, et al., 1984). O algoritmo J48 é uma implementação em Java do C4.5 desenvolvido no Weka (WAIKATO, 2013). Os algoritmos ID3, C4.5 e CART adotam uma abordagem gulosa, em que árvores de decisão são construídas recursivamente de maneira top-down, no método dividir para conquistar (HAN; KAMBER; PEI, 2006). O conjunto de treinamento é recursivamente dividido em partes menores conforme a árvore está sendo construída. Um algoritmo básico de árvore de decisão foi extraído de (HAN; KAMBER; PEI, 2006) e é resumido a seguir.

Seja,

- $D$  o conjunto de Treinamento;
- *lista\_atributos*, o conjunto de atributos candidatos;
- *método\_seleção\_atributos*, rotina que determina o critério de divisão que melhor particiona as tuplas de dados em classes individuais;

#### **Algoritmo: Gerar\_árvore\_decisão**

- (1) criar um nó  $N$ ;
- (2) se tuplas em  $D$  são todos da mesma classe  $C$ , então
- (3)     retornar  $N$  como um nó folha rotulado com a classe  $C$ ;
- (4) se *lista\_atributos* é vazia, então
- (5)     retornar  $N$  como um nó folha rotulado com a classe majoritária em  $D$
- (6) aplicar *método\_seleção\_atributos* ( $D$ , *lista\_atributos*) para encontrar o "melhor" *critério\_divisão*;
- (7) nó etiqueta  $N$  com *critério\_divisão*;
- (8) se *atributo\_divisão* é discreto de valor e divisões multiponto permitidas então
- (9)     *lista\_atributos*  $\leftarrow$  *lista\_atributos* - *atributo\_divisão*
- (10) para cada resultado  $j$  do *critério\_divisão*  
       // Particionar as tuplas e gerar sub-árvores para cada partição
- (11)    seja  $D_j$  o conjunto de tuplas de dados em  $D$  resultado satisfatório  $j$  // uma partição
- (12)    se  $D_j$  está vazia então
- (13)        anexar uma folha rotulada com a classe majoritária em  $D$  para o nó  $N$ ;
- (14)    senão anexar o nó retornado por *Gerar\_árvore\_decisão*( $D_j$ , *lista\_atributos*) para o nó  $N$ ;
- fim\_para
- (15) retornar  $N$ ;

## Capítulo 4: Definição do Problema

Nos capítulos anteriores, foi exposto todo o embasamento teórico que fundamentou a pesquisa. O presente capítulo visa descrever o problema no qual este trabalho foi motivado. Aspectos relativos aos impactos de correntes oceânicas sobre estruturas offshore e prejuízos para a indústria, limites operacionais, estimadores de condições futuras e seus benefícios serão abordados.

### 4.1 CORRENTES OCEÂNICAS E REQUISITOS OPERACIONAIS

No projeto de estruturas offshore as correntes oceânicas são geralmente consideradas. Requisitos por informações de correntes oceânicas são analisados em diferentes ramos, incluindo a indústria petrolífera, pesca, naval, serviços de busca e resgate, resposta à poluição. Pelo grande potencial e importância econômica, as indústrias de petróleo e de navegação ganham maior visibilidade e representam os principais motivadores para a construção de sistemas comerciais de medição de correntes.

As correntes oceânicas afetam as atividades offshore. Informações de correntes providas em tempo real suportam uma ampla gama de serviços operacionais e processos de tomada de decisão: orientação de plataforma, *stationkeeping* (capacidade de a embarcação manter-se na mesma posição perante ação de cargas ambientais), reboque, implantação e reparação de *risers* e outras estruturas cilíndricas, operações de ROV (Veículo Operado Remotamente, do inglês Remotely Operated Vehicle), dentre outras.

Não apenas as correntes superficiais são importantes para as atividades offshore. Os requisitos espaciais por informações mudam de acordo com a atividade e tendem a ser mais críticos para a indústria de petróleo e gás. Os efeitos de ambas as correntes superficiais e profundas devem ser considerados (SHILLING, BRENDLING, 2006). São necessárias informações de correntes em todas as profundidades e a relevância disso tem aumentado consideravelmente, na medida em que as pesquisas e trabalhos avançam para campos com profundidades cada vez maiores. Dado que esse tipo de medição não pode ser realizado por meio de satélites, torna-se necessária a construção de modelos numéricos em qualquer sistema para a monitoração de correntes.

Uma pesquisa realizada pela EuroGOOS (FISCHER; FLEMMING, 1999) sobre requisitos operacionais por informações oceanográficas revelou que, dentre um conjunto de variáveis analisadas, a intensidade e direção das correntes foram as mais frequentes.

A área geográfica é um fator importante para determinar se as correntes possuem impacto sobre as operações. Algumas áreas possuem correntes mais intensas e variantes, representando maior risco de impacto nas atividades marítimas. De modo geral, é pouco provável que as correntes sejam o principal fator de impacto em regiões onde a velocidade média não ultrapassa 0,5 m/s.

De acordo com (SOS-OC, 2002), várias atividades marítimas ocorrem a poucos quilômetros da costa, sendo que outras atividades como pescaria, navegação e prospecção de petróleo e derivados normalmente ocorrem em áreas específicas ou em rotas específicas. Por isso, é fundamental que haja um conhecimento e entendimento das condições ambientais e perfis de correntes nas quais as operações serão submetidas. No ramo offshore isso ganha bastante atenção principalmente pelo fato de a maioria das atividades possuírem instalações em locais fixos.

Dados de perfil de corrente provam ser particularmente úteis em alertas de correntes elevadas, permitindo operar de forma segura pelo tempo que for possível (OFFSHORE ENGINEER, 2002). Mesmo que a região apresente regimes de corrente extremos, sendo uma área favorável para a atividade fim, os trabalhos continuarão sendo desempenhados. Dessa forma, aumentam-se os cuidados e empregam-se métodos para mitigar possíveis acidentes em decorrência de condições críticas.

Ferramentas de medição *in situ*, dados de sensoriamento remoto e modelos numéricos estão sendo amplamente utilizados para determinar as condições de estado do mar em determinado instante de tempo. Tais ferramentas são conhecidas como sistemas de oceanografia operacional e são similares aos sistemas já conhecidos de previsão do tempo. As informações providas por estes sistemas são úteis pois orientam os planejamentos de operação e reduzem o impacto das correntes durante as operações.

A atividade de exploração e produção offshore é suportada por um processo rigoroso e extenso no qual um conjunto de atividades são envolvidas e organizadas em diferentes estágios. (SOS-OC, 2002) lista esses estágios e relaciona os principais requisitos por informações de correntes:

- A **Exploração** compreende atividades de busca por formações, prospecção e perfuração exploratória.
- A **Avaliação** é a fase que determina se o reservatório é viável economicamente para investir no desenvolvimento. São requisitos para esse estágio informações meteo-oceanográficas e critérios operacionais, de forma que o entendimento dos extremos e a variabilidade das correntes se fazem necessários para o projeto e planejamento, respectivamente.
- O **Desenvolvimento** ocorre após ter sido confirmado que o campo descoberto é economicamente viável e compreende atividades como construção e instalação de um ou mais poços. Neste momento, as informações de correntes são importantes para finalizar o projeto e apoiar as operações.
- A **Produção** é o processo de extração dos hidrocarbonetos, fazendo a separação dos líquidos, sólidos, gás e água, além de remover as substâncias não rentáveis. Dentre as principais atividades, está a perfuração, considerada uma operação crítica que requer dados meteo-oceanográficos em tempo real e previsões.
- No **Abandono** são feitas as desativações, remoções de estruturas e restauração do local. Dados previstos e em tempo reais também são necessários para apoiar essas operações.

(SOS-OC, 2002) ressalta ainda que a importância dos efeitos das informações de correntes sobre as atividades depende de fatores localização, tipo de estrutura e atividade, bem como os regimes de ondas e ventos (elementos ambientais que influenciam nas formações de correntes).

Cabe enfatizar que, assim como a intensidade das correntes é importante, o perfil e a direção também devem ser considerados. (SOS-OC, 2002) aponta que a variabilidade do perfil numa escala pequena de tempo (minutos) pode causar problemas estruturais devido às vibrações induzidas por vórtices. Outro exemplo está na direção das correntes, que influenciam o movimento dos gelos à deriva e estes pode impactar nas operações.

De acordo com (SOS-OC, 2002) não há dúvidas que o impacto das correntes sobre as operações offshore resultam em perdas financeiras. No entanto, não há como quantificar com precisão tais perdas pois essas informações não são divulgadas pelas

empresas operadoras e prestadoras de serviços. Pode-se, porém, estabelecer os tipos de impactos e os riscos; analisados sob três principais áreas:

- **Operações:** Os efeitos das correntes resultam em períodos de inatividade durante operações de perfuração e ROV, atrasos na descarga e instalação, danos ou perdas em tubulações. Por isso, as operações devem ser suspensas quando os valores ultrapassarem limites operacionais críticos. Correntes fortes em águas profundas podem produzir VIVs nas estruturas, causando danos por fadiga até a perda do *riser* e cabeça do poço. Esses danos podem levar a escapamentos e poluição que, por consequência, causam danos secundários para o ambiente.
- **Instalação e Construção:** As correntes afetam diretamente as atividades de construção e instalação. Por isso, é importante levá-las em consideração no momento do projeto e planejamento, visando reduzir o impacto. Vale lembrar que fatores como ondas e ventos não devem ser descartados. Dados históricos, conhecimento do regime meteo-oceanográfico do local, medições *in situ* e previsões devem ser observadas para um planejamento e programação eficaz das atividades (transporte de equipamentos, levantamento de pesos, operações de acoplamento e ROV, lançamentos de dutos, cabos e equipamentos de ancoragem). (SOS-OC, 2002) afirma ainda que, em muitos casos, há dados incompletos ou insuficientes para otimizar o cronograma e estimar o tempo de inatividade.
- **Projeto e Planejamento:** Informações meteo-oceanográficas e dados específicos do perfil do local são insumos que podem decidir a eficácia e um bom projeto ou planejamento. Quando essas informações são insuficientes os cuidados empregados nessas atividades podem ser excessivos o bastante, levando a investimentos elevados para a garantia da segurança operacional. Possuir uma orientação sobre as condições extremas ajudar a direcionar para um nível adequado de conservadorismo do projeto.

Diante da explanação acima, pode-se enumerar um conjunto de benefícios:

- Redução do tempo de inatividade
- Segurança operacional
- Redução dos custos de impactos

- Redução de perdas e danos
- Otimização e redução de custos associados aos projetos, especificação e planejamentos
- Desenvolvimento mais efetivo de recursos para a medição de correntes *in situ*

#### 4.2 LIMITES OPERACIONAIS

Medições de corrente é um requisito fortemente exigido pela indústria offshore para auxiliar suas atividades, seja em tempo real ou previsões para a tomada de ações preventivas em operações, ou através de informações de valores extremos para a definição de critérios de projeto, ou em medições para a análise de janelas operacionais no planejamento.

Nas operações são necessárias atualizações regulares de informações de correntes e a indicação é que essas atualizações ocorram diariamente (SOS-OC, 2002), com acurácia de 5-10 cm/s. A especificação da frequência requerida depende do conhecimento do regime de correntes e da natureza da atividade. Dependendo do caso, a frequência pode ser aumentada, ocorrendo a cada 12 horas e, às vezes, de 6 em 6 horas. As operações são sensivelmente afetadas por alterações bruscas nas correntes, o que justifica a exigência por avisos de alterações em janelas de tempo úteis, para que as devidas medidas sejam tomadas principalmente mediante atividades críticas.

Uma vez que os limites operacionais são ultrapassados, as operações devem ser temporariamente abandonadas. Somente quando as condições retornarem para dentro das capacidades operacionais, devem ser retomadas as operações. Pode-se especificar as condições seguras de trabalho conhecendo os valores críticos. (SHILLING, BRENDLING, 2006) define três etapas nas quais as atividades de perfuração podem se enquadrar:

- **Limites iniciais:** as condições devem ser inferiores a estes limites antes o início da operação, ou reinício após o abandono.
- **Limites de suspensão:** caso as condições ultrapassem esses limites, o trabalho deverá ser pausado e será retomado quando as condições caírem para um valor abaixo do limiar.

- **Limites de abandono:** A tarefa é abandonada quando os limites forem excedidos e o trabalho só será reiniciado quando as condições ficarem abaixo dos limites de início.

Segundo (SOS-OC, 2002), a implantação de tubulações e *risers* pode ser desempenhada com segurança em correntes de até 1,5 nós (aproximadamente 75 cm/s), operações de perfuração em águas profundas devem ser realizadas em correntes que não ultrapassem os 2 nós (1 m/s), e operações de mergulho são seguras em correntes de até meio nó (25 cm/s).

#### 4.3 TRABALHOS RELACIONADOS

A predição de condições meteo-oceanográficas é uma área de interesse para vários centros de pesquisa e indústria petrolífera e naval. Muitos trabalhos nesse sentido foram desenvolvidos e ainda existem muitos outros em andamento.

Existem trabalhos que buscam medir o impacto das condições oceânicas em estruturas offshore. Como parte do projeto do Laboratório Oceânico (LabOceano), foi inaugurado no Parque Tecnológico da UFRJ; o maior, mais profundo e mais moderno tanque de simulação para ensaios marítimos do mundo. O tanque foi projetado para realizar ensaios de modelos de estruturas e equipamentos usados nas atividades de exploração e produção de forma a reproduzir as mesmas condições de ondas, ventos e correntezas, que existem no local onde plataformas são instaladas (DEFESA BR, 2011). Como resultado, esses ensaios poderão definir o limite operacional da plataforma de modo que as condições ambientais não interfiram na operação dos equipamentos, além de qual a maior onda a plataforma pode suportar.

(SHIILING, BRENDLING, 2006), (FARRANT, JAVED), (PACHECO, 2004) e (SOUSA E SILVA, SILVA, FARIAS, 2005) abordam em seus trabalhos análises, resultados de estudos, simulações de operações offshore como perfuração e completação, tendo como objetivo observar e analisar o impacto de condições meteo-oceânicas durante essas operações evoluindo ainda para e previsão do tempo de vida útil das estruturas envolvidas. As análises buscam melhor quantificar o tempo, custo e risco associados ao planejamento das operações. De modo geral, estes trabalhos visam derivar guias de ajuda e recomendações para a operação e desenvolvimento de novas

tecnologias offshore, que irão gerar uma melhora na produtividade e aumentar o valor econômico do campo.

Outros trabalhos como (PICCININI, 2008), (VALADÃO, 2011) e (HARIHARAN, CERKOVNIK, THOMPSON) concentram os estudos na análise de impacto, fadiga e riscos de falha oriundos da ação de carregamentos ambientais (principalmente ondas e correntes) sobre estruturas offshore. (ARAÚJO, SILVA, FARIAS, 2007), por exemplo, apresenta resultados de ensaios em materiais de dutos com os quais foram extraídos parâmetros de fadiga, utilizados como base para estimativa da vida-útil de estruturas sujeitas às condições de corrosão-fadiga. Utilizando-se de simulação computacional, (TSUKADA, YAMAMOTO, MOROOKA, 2007) investiga o comportamento mecânico do *riser* de perfuração durante o processo de instalação do BOP em diferentes profundidades, através de análise dinâmica no domínio do tempo. Ewing (1990) descreve o ambiente oceânico, especialmente ondas, correntes, ventos e temperatura em águas rasas e profundas; para determinar como as forças ambientais devem ser consideradas no projeto de embarcações e estruturas offshore.

O trabalho de (SPRING et al., 1999) propõe um resumo dos dados de medições de corrente e o desenvolvimento de critérios a serem considerados na preparação para as atividades de perfuração em Tranche-6, região caracterizada por severas correntes de magnitude potencialmente alta. Dados coletados entre 1992 e 1994 serviram de fonte para a definição dos critérios baseados em perfis extremos e operacionais de correntes. Os resultados do estudo foram implementados e testados com sucesso durante a perfuração de dois poços em Tranche-6. A preparação incluiu a medição prévia da corrente, visando quantificar os desafios do projeto e a provisão para medições em tempo real durante a perfuração.

Em 2010, Hsiao e Hwang publicaram no Congresso Internacional em Computação Natural um estudo utilizando Redes Neurais Artificiais onde foi proposto um modelo de previsão de velocidade de correntes oceânicas no porto de Taiwan. O modelo tem como parâmetros de entrada as medições da altura, ciclo da onda e velocidade da corrente. Os resultados são então comparados com o modelo já existente AR(2). Os autores destacam que a abordagem baseada em redes neurais possui as vantagens de alta velocidade, habilidade de alto-aprendizado, alta capacidade de

memória e tolerância à falha; de forma que o erro pode ser corrigido pelo próprio modelo, aumentando a acurácia e estendendo o tempo de predição.

Na mesma linha, (AYDOGAN et al., 2010) propõe um preditor de velocidade de correntes no Estreito de Istanbul através da modelagem de uma rede neural parametrizada com informações de perfil de corrente, condições meteorológicas e elevações da superfície. O modelo proposto é capaz de obter previsões acuradas de até 12 horas no futuro com um erro médio de 0,16 m/s.

(SMIRREN, et. al, 1999) descreve a abordagem de empresas petrolíferas, uma companhia de serviços e um fabricante de instrumentos que se uniram para, de forma colaborativa, desenvolver uma tecnologia computacional capaz de otimizar dados disponíveis para engenheiros e operadores tomadores de decisão da indústria offshore. O trabalho discute a evolução e aplicação de um Perfilador de Correntes Marinhas (ADCP – Acoustic Doppler Current Profiler) utilizado para o monitoramento de correntes, ondas e marés por empresas petrolíferas em operações de perfuração em águas profundas.

Um caso de sucesso de parceria entre indústrias de petróleo é descrito no trabalho de (CROUT, et.al, 2006). O trabalho chama atenção para o impacto provocado por ciclos de correntes e vórtices sobre sondas de perfuração e plataformas de produção em operações de águas profundas no norte do Golfo do México. Com o objetivo de proporcionar melhores estimativas de correntes fortes, o Departamento do Serviço de Gerenciamento de Minerais do Interior (MMS - Department of Interior's Minerals Management Service) emitiu uma notificação, exigindo que os operadores do Golfo do México monitore e relate, através de site aberto, medições de correntes em plataformas de produção e perfuração que operam em profundidades superiores a 400 metros. A proposta consistiu em construir algoritmos de controle de qualidade, além de utilizar o modelo Perfilado Acústico de Correntes (Acoustic Doppler Current Profilers - ADCP), para processar os dados enviados e exibir o resultado em uma página web acessível ao público.

No trabalho de Shanks (1979) o foco está no planejamento e equipamentos necessários para perfuração em áreas caracterizadas por águas profundas e correntes elevadas. O autor afirma que a combinação dos dois requisitos exige que cada um tenha suas próprias considerações independentes, bem como uma fusão especial das duas

considerações. O trabalho aborda uma série de experimentos realizados pela companhia SEDCO para seus clientes e com base nesses experimentos, extrai as principais considerações exigidas no contexto analisado em relação à pré-planejamento, preparação técnica e obtenção de equipamento.

Visando o funcionamento otimizado de estruturas marítimas, Altunkaynak (2008) trata a abordagem em torno de parâmetros de ondas marítimas. O objetivo do trabalho é prever parâmetros de ondas com precisão, tendo com proposta o método chamado de filtragem Geno-Kalman, que consiste na junção de algoritmos genéticos e filtros de Kalman onde a base está no cálculo adaptativa para chegar à solução. O método é aplicado a dados de velocidade do vento e dados históricos e atuais de ondas.

Em (IQDOUR, ZEROUAL, 2006) técnicas de Mineração de Dados são utilizadas para gerar um modelo preditor de velocidade dos ventos. O trabalho propõe utilizar redes neurais Multi-Layered Perceptrons para prever a velocidade do vento atual através dos valores anteriores da mesma variável. Os resultados demonstram que o modelo pode ser usado para a predição com uma precisão de 90% ou acima.

Tendo como principal preocupação a segurança em operações marítimas, (GILBERT, 1998) destaca o papel vital de dados meteo-oceanográficos e modelos preditivos precisos e confiáveis na resposta a incidentes marítimos que envolvam derrames de petróleo e produtos químicos, operações de limpeza da faixa litorânea e salvamento de embarcações, bem como em pesquisa e planejamento de resgate marítimo. O artigo detalha também exemplos de uso de dados meteo-oceanográficos no suporte a decisão durante diversas situações de respostas a incidentes marítimos nas águas Australianas (área de interesse do artigo), e sugere métodos para melhorar a provisão de serviços meteo-oceanográficos em resposta aos incidentes.

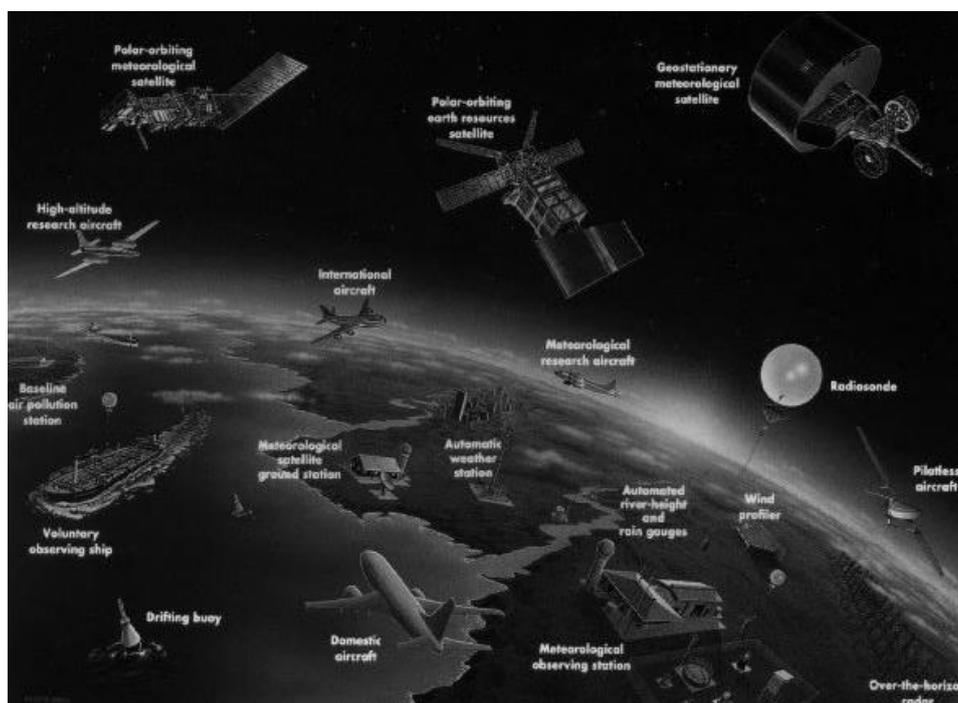
Kaiser e Pulsipher (2007) descrevem como a previsão oceânica e de tempo impactam nas atividades de produção offshore e gerenciamento de poluição no Golfo do México. No artigo são apresentados resultados físico e modelos de decisão no apoio a atividades de produção, desenvolvimento e gerenciamento de resposta a vazamentos. Casos reais são discutidos e quantificados, de modo que benefícios econômicos, oriundos de sistemas de observação e previsão oceânica, são demonstrados em números.

(JONATHAN, EWANS, 2013) defende técnicas estatísticas como abordagem eficaz na modelagem de condições oceânicas extremas para projetos marítimos. Os

autores propõem, em especial, a Inferência Bayesiana como um framework consistente para a modelagem e inferência de valores extremos.

Conforme observado na literatura estudada, os estimadores de condições de correntes desenvolvidos até então fazem a predição das condições futuras de uma determinada região, considerando os dados atuais e históricos de correntes desta mesma região, além de uma série de parâmetros (ventos, ondas, marés, etc) e do próprio conhecimento do fenômeno físico.

A coleta por informações meteo-oceanográficas é realizada com o apoio de diversas fontes: satélite (órbita / geoestacionária), estações costeiras e terrestres, bóias à deriva, bóias ancoradas, relatórios e observações de navios e estações automáticas, plataformas de petróleo, avião, radar meteorológico, balões meteorológicos.



**Figura 15: Fontes de informações de correntes. FONTE: (SOS-OC, 2002)**

As informações de correntes, em especial, são obtidas basicamente por meio das seguintes de medições in situ (plataformas marítimas, navios de perfuração e embarcações de pesquisa, bóias, correntômetros), modelos numéricos e sensoriamento remoto (satélites e aeronaves). Apesar da importância e da existência das fontes de informação listadas, ainda existe uma carência por informações precisas e perfiladas

para regiões específicas. Para o caso de modelos numéricos, alguns são muito genéricos e possuem processamento demorado.

Além de não existir um modelo de previsão implementado para a Costa do Brasil, não existe parâmetros de informação suficiente para aplicar alguma técnica baseada puramente no fenômeno físico, como existe para outras regiões. Isso justifica a busca por meios alternativos que viabilizem a utilização de menos parâmetros.

A possibilidade de uma estimativa baseada no estado de corrente em uma região vizinha representa uma alternativa cujo resultado pode ser igualmente eficiente aos métodos já conhecidos, porém, demandando de um conjunto menor de informações, conhecimento prévio e menor tempo de resposta.

Além disso, a probabilidade de um fenômeno "extremo" pode ser mais fácil prevista tendo como referência a observação da ocorrência na vizinhança.

## Capítulo 5: Proposta

Os trabalhos encontrados na literatura propõem preditores de condições de correntes para um determinado ponto de observação com base nas informações atuais e passadas deste mesmo ponto. Para alguns pesquisadores é possível que as condições observadas em regiões próximas sirvam de parâmetro para estimar as condições de um ponto alvo em um dado momento futuro. Nesse sentido, há uma suspeita da existência de algum tipo de influência entre diferentes pontos vizinhos.

A suspeita da influência de uma vizinhança sobre um ponto específico, gerou a seguinte questão de pesquisa: Em um dado momento e em uma determinada região oceânica foram observadas determinadas condições oceânicas. É possível formular de um modelo baseado em mineração de dados capaz de estimar a condição futura de corrente para um dado ponto interesse com base em observações atuais e históricas de sua vizinhança?

Para responder a questão, o presente trabalho propõe uma metodologia baseada em Mineração de Dados para a geração de modelos de predição e classificação de intensidade de correntes oceânicas. Trata-se de um modelo de previsão para um ponto específico baseado na influência de correntes medidas em pontos vizinhos, para estimar o estado futuro da corrente no ponto estudado. É importante acrescentar que o estudo pode ser aplicado em qualquer região. Porém, deve-se levar a consideração o perfil de correntes da região.

### 5.1 ÁREA DE ESTUDO

A área em que o estudo foi realizado localiza-se entre as latitudes -24.2790 e -22.0728, e longitudes -42.00000 e -40.00000, região inserida na Bacia de Campos.

Descoberta há 30 anos, a Bacia de Campos é uma bacia sedimentar com cerca de 100 mil quilômetros quadrados, localizada na costa norte do Estado do Rio de Janeiro (em Arraial do Cabo) e se estende até o sul do estado do Espírito Santo, nas proximidades da cidade de Vitória (PETROBRAS, 2013). Esta é uma das bacias mais atrativas do continente sul-americano, superando 1,7 milhão de barris de óleo (PETROBRAS, 2013), sendo responsável por cerca de 80% da produção nacional de

petróleo. De acordo com a prefeitura de Macaé, a Bacia de Campos possui 55 campos, com cerca de 2.350 poços perfurados e 45 plataformas marítimas.

Essa região foi escolhida devido ao grande potencial e importância para empresas de petróleo, bem como para a economia brasileira. Por possuir as maiores reservas comprovadas, é considerada a província petrolífera do país. O cenário de grandes descobertas representa um desafio para o setor petrolífero, concentra os principais projetos e investimentos para elevar o índice de eficiência operacional.

## 5.2 METODOLOGIA DE TRABALHO

O trabalho desenvolveu-se em cima de uma base histórica de dados de correntes oceânicas, compondo-se das 4 etapas detalhadas nas subseções a seguir.

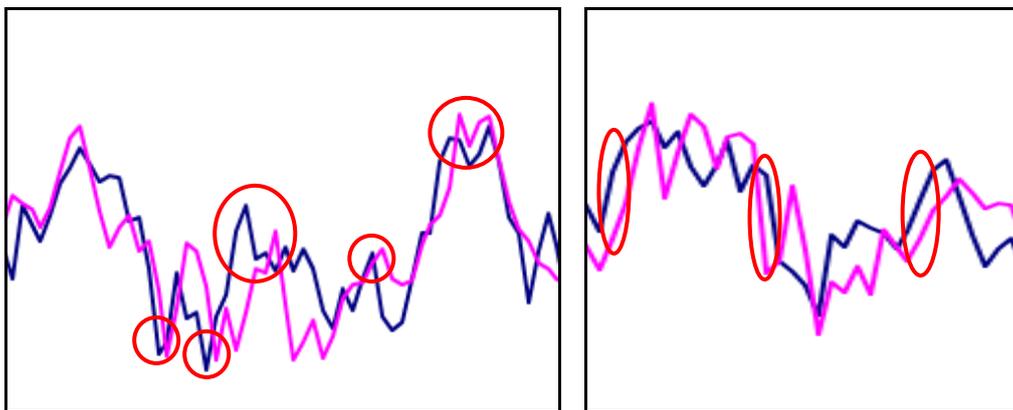
### 5.2.1 Coleta dos Dados

Os dados utilizados para o estudo foram coletados da base de dados disponibilizada pela Hycom Consortium (HYCOM, 2013), fonte de dados descrita na seção 5.3.1. A amostra extraída consiste em uma grade de pontos [lat,lon] com latitude variando entre -24.2790 e -22.0728, e longitude entre -42.00000 e -40.00000. O período da extração foi composto pelos anos de 2004 até 2010.

### 5.2.2 Análise Preliminar dos Dados

Nesta etapa, foram selecionados alguns pontos do espaço [lat,lon] da região estudada para analisar e entender o comportamento das correntes. Para cada um dos pontos selecionados, foram extraídas as séries temporais das intensidades e foram plotados gráficos de cada série individual, bem como gráficos com de séries agrupadas. Aspectos como faixas de intensidade predominantes, picos, frequência de picos e condições extremas foram investigados. Essa análise ajudou, por exemplo, a definir os 3 intervalos para as faixas de condições de intensidade (seção 5.2.3), já que a simples adoção dos limites operacionais descritos na seção 4.2 poderia resultar classes com um número insignificante de amostras, interferindo na qualidade dos resultados da aplicação de algoritmos de mineração.

A investigação comparativa das séries temporais das intensidades de correntes em pontos vizinhos foi realizada por meio de visualização de diversos gráficos. A análise revelou comportamentos similares em uma variação de tempo de até 3 dias. Ou seja, em um determinado ponto no espaço  $P1$ , a intensidade aumentou até um pico e depois sofreu uma queda. No dia seguinte (ou 2 dias depois), em um ponto vizinho no espaço  $P2$ , a intensidade também atingiu um pico e sofreu uma queda. Outros casos mostram que, se a intensidade da corrente em um ponto  $P1$  vem aumentando ou decrescendo com um delta  $d$ , é possível assumir que, no dia seguinte, a intensidade no ponto  $P2$  também vai aumentar (ou decrescer) com delta próximo de  $d$ . A Figura 16 mostra duas séries referentes a intensidade da corrente de dois pontos, exemplificando os comportamentos descritos.



**Figura 16: Comparação das séries de intensidade em 2 pontos vizinhos**

### 5.2.3 Tratamentos dos dados e Geração do Conjunto de Treinamento

O tratamento dos dados visa derivar um novo conjunto de dados a partir da base original, após um processo de limpeza e transformação dos dados para um formato que realmente agregue valor a fase posterior de implementação do modelo, se ajustando às técnicas de mineração utilizadas.

Correntes oceânicas podem ser expressas por vetores, que representam a velocidade e a direção dos movimentos da água. Sendo assim, oceanógrafos retratam as correntes oceânicas por vetores Euclidianos. Neste caso,  $u$  e  $v$  são os componentes do vetor de velocidade da corrente. O arquivo de dados extraído da base do Hycom Consortium representa uma grade de pontos  $[lat, lon]$ , onde cada ponto possui os valores das componentes  $u$  e  $v$ , em séries temporais distintas para cada nível de

profundidade. Como o interesse do estudo focou nas correntes de superfície, foi considerada apenas a primeira camada de profundidade.

As componentes  $u$  e  $v$  foram utilizadas para calcular a intensidade da corrente (Equação 9) que representou a única variável oceanográfica considerada no modelo.

$$\textit{intensidade da corrente} = \sqrt{v^2 + u^2} \quad (9)$$

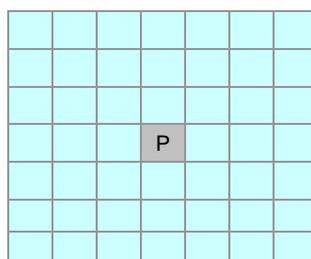
A estruturação do conjunto de treinamento seguiu o algoritmo abaixo:

- 1- Selecionar um ponto [lat,lon] da grade extraída, como alvo do modelo;
- 2- Definir um raio de pontos vizinhos nos sentidos norte, sul, leste e oeste; resultando em uma sub-grade;
- 3- Escolher um instante  $t$  da série temporal como início da geração de instâncias, onde cada instante  $t$  representa uma instância e o valor da intensidade do ponto alvo no instante  $t+1$  representa o atributo de saída da instância;
- 4- Para cada instante  $t$ , repetir o passo 5 até  $t =$  penúltimo instante da série;
- 5- Para cada ponto da sub-grade, incluindo o ponto central (alvo)
  - 5.1 - Selecionar as intensidades de corrente nos instantes  $t$ ,  $t-1$  e  $t-2$ , formando os atributos de entrada da instância;
  - 5.2 - Adicionar o atributo de saída = intensidade do ponto alvo no instante  $t+1$

O resultado da aplicação do algoritmo acima é uma matriz com o formato da Quadro 1:

**Quadro 1: Formato dos conjuntos de dados**

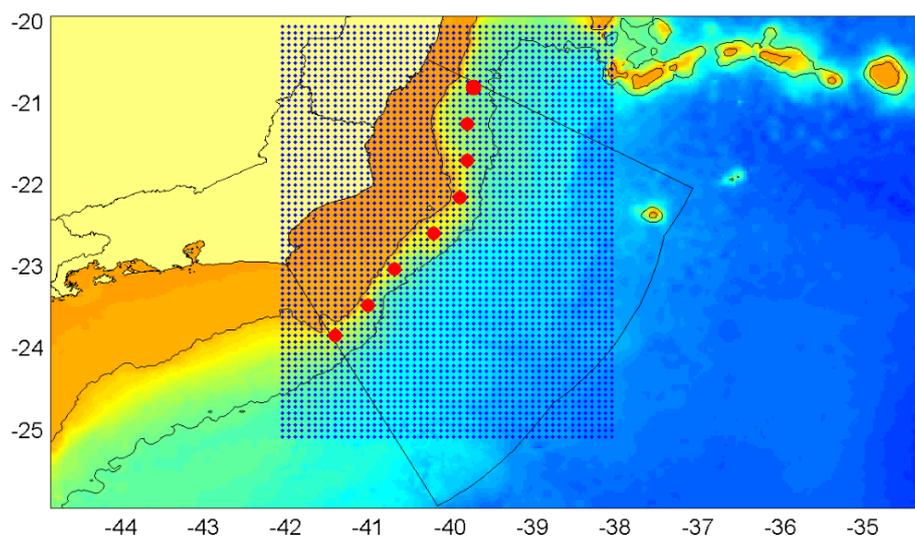
Entrada			Saída
hoje ( $t$ )	ontem ( $t-1$ )	ante-ontem ( $t-2$ )	Amanhã ( $t+1$ )



**Figura 17: Grade 7x7 com o ponto central e vizinhança de 3 vizinhos**

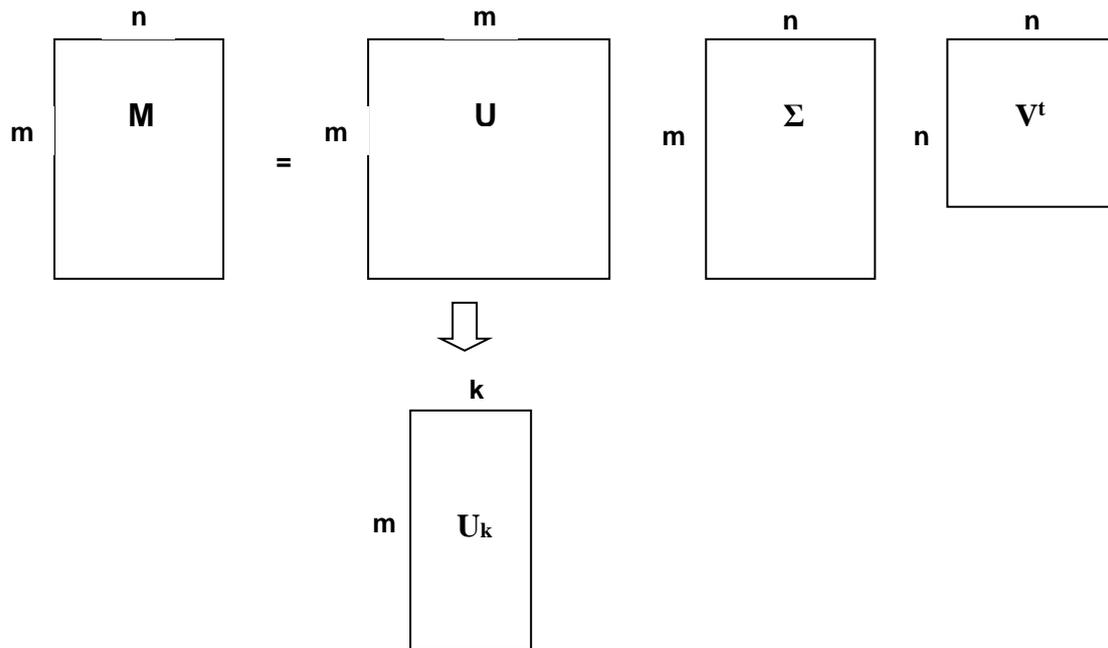
Levando em consideração um raio de 3 vizinhos (Figura 17), obtém-se uma grade de 7x7, ou seja, 49 pontos. Cada instância terá 147 atributos de entrada (49 pontos multiplicados por 3 instantes).

Sob orientação de especialistas, este trabalho se limitou preferencialmente a um raio de 30 km, que corresponde ao valor do raio de Rossby, pois a previsibilidade é melhorada dentro dos limites desse raio. De acordo com (HAZA, et. al., 2013) e (MOOERS, 1999), um modelo de predição é mais eficaz quando utilizado dentro de um círculo de raio R (raio de deformação de Rossby). A Figura 18 mostra um exemplo de uma grade extraída da base do Hycom Consortium. Cada ponto disposto no mapa possui as informações de corrente exatamente na posição [lat, lon] desenhada no mapa. Os pontos vizinhos são espaçados com uma distância de aproximadamente 6,5 km.



**Figura 18: Exemplo de uma grade extraída da base do Hycom Consortium**

Conforme o raio de vizinhos cresce, a quantidade de atributos também aumenta. A grande quantidade de atributos acarreta no problema conhecido como alta dimensionalidade dos dados. A alta dimensionalidade aumenta a complexidade e pode degradar o desempenho das técnicas de mineração, com impacto direto no custo e na precisão do modelo resultante. Por esse motivo, a matriz gerada foi submetida à técnica SVD de redução de dimensionalidade.



**Figura 19: Aplicação da SVD sobre a massa de dados para redução de dimensionalidade**

Seja  $M$ , a matriz gerada. A aplicação da SVD resulta nas matrizes  $U$ ,  $\Sigma$  e  $V^t$ . Os valores singulares muito pequenos são descartados e, por consequência, a matriz  $\Sigma$  passa a ter posto  $k$ , que é o número de valores singulares mantidos. O número  $k$  representa o número de colunas a serem consideradas na matriz  $U$ , que, por fim, formará o conjunto de treinamento utilizado para a aplicação dos algoritmos de mineração.

Para os dados de teste o conjunto foi gerado através do mesmo algoritmo aplicado aos dados de treinamento, com a aplicação de uma transformação diferente da SVD para reduzir a dimensão. A transformação aplicada possui a seguinte forma:

Seja  $X$  a entrada original, o valor  $Y$  requerido é representado pela equação (10) com SVD aplicado. Dado que  $X = U \Sigma V^t$ , então:

$$Y = (X * \text{Transposta}(V^t)) * \text{Inversa}(\Sigma) \quad (10)$$

Já que  $\Sigma$  é uma matriz diagonal, a inversa de  $\Sigma$  é  $1/d$ , onde  $d$  é o elemento da diagonal. Como o objetivo é truncar em  $k$  colunas, basta usar o valor 0 nos elementos da diagonal a partir de  $d[k][k]$ .

Alguns algoritmos de aprendizado de máquina manipulam apenas atributos com valores nominais. Também existem algoritmos que trabalham apenas com entradas

numéricas e, outros, aceitam ambos os tipos de dados. A variável “intensidade da corrente” compreende informações numéricas, medidas em m/s. Conforme foi descrito anteriormente, as técnicas de mineração adotadas neste trabalho foram Redes Neurais e Árvores de Decisão; ou seja, para ambos os casos, o tipo do atributo de entrada é adequado. Todavia, como o interesse do trabalho estava na construção tanto de um modelo de regressão quanto classificação, fez-se necessário aplicar uma transformação no atributo de saída para que o mesmo assumisse um formato categorizado. Desta forma, foram adotadas duas abordagens para a categorização. A primeira foi baseada na divisão do trabalho de (LIANG, et al., 2003) e consistiu na discretização do atributo de saída em intervalos de 0,2 m/s. Isso resultou em 6 classes:

**Quadro 2: Categorias de Intensidade por faixas de 20 m/s**

<b>20</b>	<b>40</b>	<b>60</b>	<b>80</b>	<b>100</b>	<b>120</b>
$0 \leq i \leq 0,2$	$0,2 < i \leq 0,4$	$0,4 < i \leq 0,6$	$0,6 < i \leq 0,8$	$0,8 < i \leq 1,0$	$i > 1,0$
m/s					

A segunda abordagem teve como base a definição dos limites operacionais para operações offshore, resultando nas 3 categorias a seguir:

**Quadro 3: Categorias de Intensidade por faixas de segurança operacional**

<b>Seguro</b>	<b>Atenção</b>	<b>Não operável</b>
$0 \leq i \leq 0,75$	$0,75 < i \leq 1,0$	$i > 1,0$
m/s		

Na região estudada neste trabalho, a frequência de ocorrências de medições de intensidade que ultrapassam o limite de 1 m/s é muito pequena em relação aos intervalos inferiores. Caso o trabalho adotasse fielmente os limites acima, isso implicaria em conjuntos de treinamento desbalanceados, provocando impacto na qualidade dos modelos resultantes, que tenderiam a acertar menos para a classe mais alta. Esse problema foi discutido com especialistas, que orientaram deslocar o limite não operável para uma faixa iniciando com intensidade menor a 1 m/s, sem perda de qualidade para o trabalho. Por esta razão, para fins experimentais, os intervalos de intensidade de corrente adotados foram alterados para valores menores (conforme o Quadro 4).

**Quadro 4: Categorias de Intensidade por faixas adaptada**

<b>Seguro</b>	<b>Atenção</b>	<b>Não operável</b>
$0 \leq i \leq 0,5$	$0,5 < i \leq 0,75$	$i > 0,75$
m/s		

Com os novos intervalos, foi possível obter uma quantidade de amostras maior para a classe “Não operável” e derivar conjuntos de treinamento com quantidades de amostras suficientes e balanceadas para a aplicação dos algoritmos de mineração.

#### 5.2.4 Construção do Modelo e Experimentos

A etapa de experimentos se dividiu em 2 abordagens: regressão e classificação. Na regressão foram construídos modelos de Redes Neurais, nos quais o algoritmo utilizado foi o MultilayerPerceptron (MLP). Para a classificação, foram utilizadas, além de Redes Neurais com MLP, algoritmos de Árvore de Decisão.

A realização dos experimentos consistiu em uma etapa iterativa onde cada execução do algoritmo compreendia uma combinação entre os parâmetros do algoritmo e ajustes no conjunto de treinamento.

O modelo construído para a rede neural constituiu-se das seguintes configurações:

- número de nós da camada escondida: variando de 8 até 37
- decay ativado
- taxa de aprendizado: 0.1, 0.01 e 0.001
- momento: 0,2
- número de iterações: 500
- tamanho do conjunto de validação: 40% do conjunto de treinamento
- Validation Threshold: 20

O número de iterações foi fixado em 500 pois foi verificado que a execução da rede neural sempre parava antes desse número de iterações. Antes de fixar o valor, foram realizados vários experimentos com números de iterações maiores.

Com relação ao conjunto de treinamento, é importante lembrar que quanto maior for  $k$ , o número de fatores na aplicação da SVD, melhor a qualidade do conjunto de dados. Desta forma, optou-se por não estipular um valor  $k$  fixo e basear todos os experimentos em conjuntos de treinamento de tamanhos variados de acordo com o valor de  $k$ . Todos os experimentos foram repetidos para 20, 25 e 30 fatores. Com essa variação foi possível avaliar, através da precisão do modelo, o melhor tamanho para o conjunto; isto é, o melhor valor de  $k$ .

Outra variação dos experimentos representou a adição de novos atributos no conjunto de treinamento, de maneira que, além da matriz com as informações temporais de  $d$ ,  $d-1$  e  $d-2$  para prever o ponto em  $d+1$ , utilizou-se dos valores das intensidades medidas nos  $n$  últimos dias do ponto a ser previsto. Este valor  $n$  também teve variações que corresponderam a 10, 5 e 3 dias. Com isso,  $n$  novas colunas foram acrescentadas a matriz, porém, isso ocorreu apenas após fazer as devidas transformações nos conjuntos de testes e treinamento SVD. Se a série dos  $n$  últimos dias medidos fosse incluída antes das transformações, o conjunto resultante perderia informação.

O resultado foram conjuntos no formato abaixo:

**Quadro 5: Formato do conjunto de dados com SVD**

Entrada				Saída
SVD			Sem SVD	
hoje (d-0)	ontem (d-1)	ante-ontem (d-2)	$n$ últimos dias do ponto	$d+1$

Considerando a geração de uma matriz através de uma grade com os pontos mais próximos do ponto a ser previsto (e sem os pontos distantes), obtém-se uma quantidade menor de dimensões de entrada: 27 atributos (9 pontos da grade x 3 dias). Nesse caso, uma opção a ser considerada foi a execução dos experimentos sem a aplicação do SVD. Testes com SVD também foram realizados e, nestes foram considerados de 10 a 25 fatores. Em ambos os casos, foram incluídos os  $n$  últimos dias medidos.

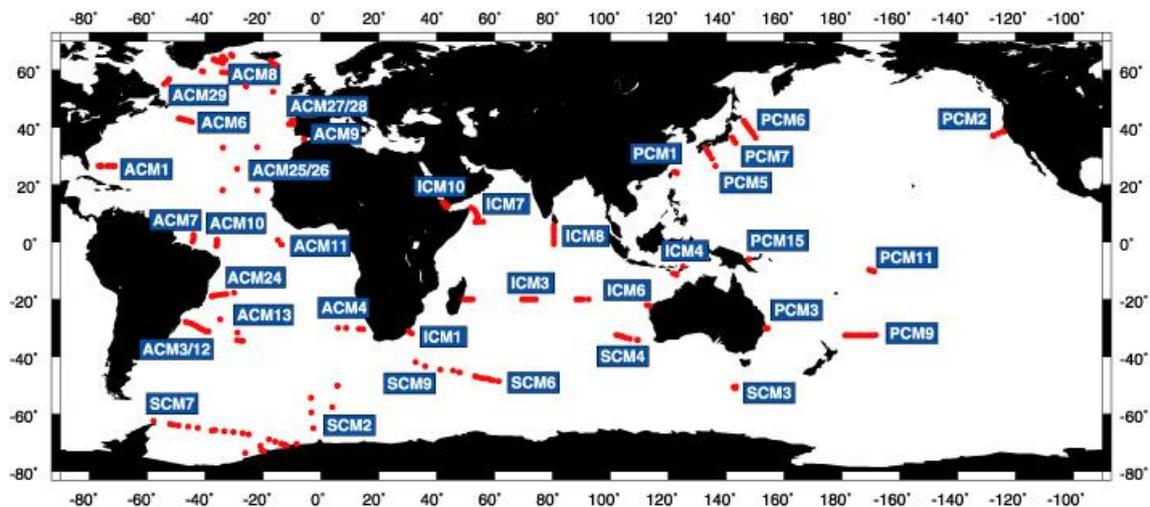
**Quadro 6: Formato do conjunto de dados sem SVD**

Entrada Sem SVD				Saída
hoje (d-0)	ontem (d-1)	ante-ontem (d-2)	n últimos dias do ponto	d+1

### 5.3 RECURSOS E FERRAMENTAS

#### 5.3.1 Bases de Dados

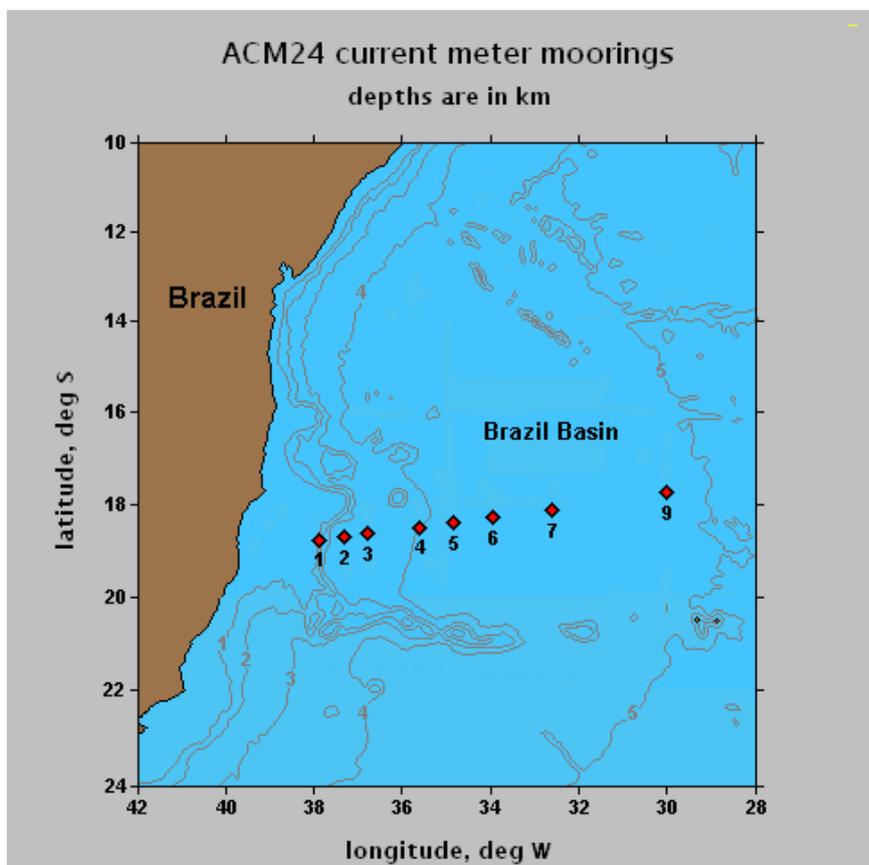
Dentre as bases de informações de correntes oceânicas disponíveis para a pesquisa encontram-se o Programa de Experimento de Circulação Oceânica Mundial (World Ocean Circulation Experiment - WOCE), o MyOcean e o Hycom. O WOCE (WOCE, 2013) foi um componente do Programa Internacional de Pesquisa Climática (World Climate Research Programme - WCRP) que usou recurso de aproximadamente 30 países para fazer observações *in-situ* e de satélite do oceano global. Essas observações ocorreram entre 1990 e 1998, seguindo com uma fase de análise e modelagem até o ano de 2002. O principal objetivo é que os dados resultantes sejam utilizados durante as décadas seguintes como recurso único para pesquisadores climáticos e cientistas marinhos.



**Figura 20: Localização dos Fundeios de Medição de Corrente (Fonte: WOCE)**

Dados de medição de correntes dos oceanos Atlântico, Índico, Pacífico e Meridional são disponibilizados em arquivos NetCDF ou ASCII com informações de

velocidade, direção e temperatura em intervalos de tempo de 1 hora. A utilização desta base foi inviabilizada devido à pequena quantidade de pontos medidos e a disposição geográfica dos mesmos, representados na Figura 20.



**Figura 21: Fundeios de Medição Localizados na região da Bacia de Campos (Fonte: WOCE)**

A Figura 21 demonstra que os pontos estão em uma direção perpendicular à CB. Dado o entendimento do perfil da Corrente do Brasil, a disposição geográfica favorável ao trabalho deveria ser no sentido da CB.

O My Ocean foi um projeto concedido pela Comissão Européia, iniciado em 2009, com término em março de 2012. O objetivo foi desenvolver e operar um sistema rigoroso, robusto e sustentável de Previsão e Monitoramento dos Oceanos, de modo a beneficiar diversas áreas de uso: segurança marítima, previsão de derramamento de óleo, gestão de recursos marinhos, alteração climática, previsão sazonal, atividades costeiras, pesquisas de gelo, qualidade da água e poluição (MY OCEAN). O catálogo interativo

disponibilizado na Internet permite extrair diversos parâmetros de informações oceânicas.

Por fim, o consórcio HYCOM (HYCOM, 2013) é um esforço multi-institucional patrocinado pelo Programa de Parceria Oceanográfica Nacional (National Ocean Partnership Program - NOPP) como parte do Global Ocean Data Assimilation Experiment (GODAE). O objetivo é desenvolver e avaliar um modelo oceânico homônimo ao consórcio, o HYbrid Coordinate Ocean Model (HYCOM).

HYCOM é um modelo sofisticado, de alta resolução, de circulação geral do oceano, que resolve equações primitivas da hidrodinâmica dos oceanos. É amplamente utilizado em diversas instituições para fins de pesquisa e previsão oceanográfica operacional. Trata-se de um conjunto de equações refinados ao longo de muitos anos, que descrevem os efeitos das marés, ventos, a rotação da terra, e muitos outros fatores sobre o fluxo de água (Marine Geospatial, 2013). Segundo (MARINHA, 2013), a principal característica do HYCOM está no emprego de coordenadas verticais híbridas, podendo ser níveis de pressão, isopicnais ou sigma (seguidoras de terreno); obtendo com isso, resultados mais realísticos nas simulações dos fenômenos oceânicos em regiões rasas e de mar aberto.

De acordo com (CPTEC), informações de satélite representam milhões de observações ao redor do globo, que são incorporadas, ou assimiladas, rotineiramente por grandes centros operacionais mundiais em seus modelos numéricos de previsão. Na assimilação, esquemas matemáticos complexos combinam resultados de modelos numéricos com dados observados para corrigir a solução do modelo e gerar valores mais próximos dos observados, de forma a melhorar a qualidade da previsão.

O modelo utiliza o sistema de assimilação de dados Navy Coupled Ocean Data Assimilation (NCODA). O NCODA usa o modelo como um palpite inicial em um esquema de Interpolação Ótima multi-variável (multi-variate Optimal Interpolation - MVOI) e assimila dados de altimetria e temperatura da superfície do mar obtidos de satélite, perfis verticais de temperatura e salinidade, flutuadores ARGO e boias oceanográficas. Em seguida, através de um sistema de assimilação de dados oceânicos modular (Modular Ocean Data Assimilation System - MODAS), informações de superfície são projetadas para dentro da coluna d'água.

O site disponibiliza acesso ao sistema de saída de previsões, quase em tempo real, baseado no HYCOM + NCODA. Dados diários são acessíveis em formato NetCDF. A saída é estruturada como uma série temporal de snapshots do estado da região simulada. O intervalo de tempo é de 1 dia e, a cada dia, é disponibilizado um conjunto de grades de 2D com parâmetros de superfície do mar (altura da superfície do mar, espessura da camada de mistura) e um conjunto de grades 3D, que especificam temperatura, salinidade, e vetores de corrente  $u$  e  $v$  em uma série de profundidades. São fornecidas 33 camadas de profundidade: 0, 10, 20, 30, 50, 75, 100, 125, 150, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1750, 2000, 2500, 3000, 3500, 4000, 4500, 5000 e 5500 metros.

O dado disponibilizado em arquivo NetCDF normalmente contém uma variável física de quatro dimensões de informações de corrente (salinidade, temperatura,  $u$  e  $v$ ), assim como variáveis auxiliares que representam as coordenadas (latitude e longitude) da grade, as 33 profundidades e o tempo. O período de tempo é dependente da escolha do usuário no momento da extração dos dados. A grade de dados também é definida pelo usuário, que especifica as coordenadas espaciais.

Neste trabalho a extração da grade de dados foi definida pelos parâmetros a seguir:

- Dimensões  $u$  e  $v$  selecionadas
- Latitude: entre -24.2790 e -22.0728,
- Longitude: entre -42.00000 e -40.00000
- Período: 2004 até 2010

Vale observar que neste trabalho o modelo é gerado a partir de informações geradas por outro modelo, ao invés de dados reais observados. Isso ocorre em função da falta de informações reais disponíveis publicamente. Além disso, não existem medições reais em uma cobertura espacial ampla como a disponível pelo Hycom. A orientação dos especialistas foi adotar os dados gerados pelo modelo Hycom como dados observados, de modo que isso não impactaria na qualidade do trabalho. Em uma aplicação real, uma estrutura de medição pode ser implantada em uma disposição espacial similar a utilizada pelo trabalho para tratar pontos de interesse específicos.

### 5.3.2 NetCDF

NetCDF (Network Common Data Form) é um conjunto de bibliotecas, interfaces e formatos de dados. Esses três componentes suportam a criação, acesso e compartilhamento de dados científicos orientados a conjuntos. As interfaces apoiam o acesso a dados e consiste em uma coleção de bibliotecas de acesso a dados distribuídas gratuitamente em C, Fortran, C++, Java, e outras linguagens. Essas bibliotecas suportam formatos de dados auto-descritivos e independentes de máquina para a representação de dados científicos. O arquivo no formato NetCDF é capaz de armazenar dados multidimensionais (variáveis) como temperatura, umidade, pressão, velocidade e direção do vento e correntes.

Em (UNIDATA, 2013) são enumeradas 6 características do formato NetCDF:

- **Auto-descritivo:** um arquivo NetCDF carrega informações sobre os dados que ele armazena, ou seja, dados e metadados estão juntos no mesmo arquivo.
- **Portável:** um arquivo NetCDF pode ser acessado por computadores com diferentes formas de armazenamento de tipos de dados (inteiro, caractere, ponto-flutuante).
- **Escalável:** Um pequeno subconjunto de um grande conjunto de dados pode ser acessado eficientemente.
- **“Anexável”:** Novos dados podem ser adicionados a um arquivo NetCDF adequadamente estruturado, sem copiar o conjunto de dados ou redefinir sua estrutura.
- **Compartilhável:** Um mesmo arquivo NetCDF pode ser acessado simultaneamente por um processo escritor e vários leitores.
- **Arquivável:** O acesso a todas as formas anteriores de dados NetCDF será suportado pelas versões atuais e futuras do software.

O quadro 7 mostra a estrutura de um documento extraído no formato NetCDF. Essa estrutura foi obtida a partir da biblioteca de acesso e manipulação de dados na ferramenta R, “ncdf”. O arquivo contém as coordenadas geográficas latitude e longitude (definidas pelas variáveis X e Y), a informação de tempo (expressa pela variável MT), os níveis de profundidade (variável Depth) e as componentes de corrente u e v. Estas

últimas são definidas no Quadro 7 como  $u[X, Y, Depth, MT]$  e  $v[X, Y, Depth, MT]$ , ou seja, o valor de  $u$  e  $v$  na posição do espaço  $X$  e  $Y$ , na profundidade  $Depth$  e no tempo  $MT$ .

**Quadro 7: Exemplo de estrutura de um documento NetCDF**

```
[1] "file glb_analysis.nc has 4 dimensions:"
[1] "MT      Size: 181"
[1] "Depth   Size: 33"
[1] "Y       Size: 31"
[1] "X       Size: 26"
[1] "-----"
[1] "file glb_analysis.nc has 4 variables:"
[1] "float u[X,Y,Depth,MT] Longname: u-veloc. [90.9H] Missval:1e+30"
[1] "float Latitude[X,Y]   Longname:Latitude Missval:1e+30"
[1] "float Longitude[X,Y]  Longname:Longitude Missval:1e+30"
[1] "float v[X,Y,Depth,MT] Longname: v-veloc. [90.9H] Missval:1e+30"
```

### 5.3.3 Ferramenta R

R é uma linguagem e ambiente de software livre integrado voltado para computação estatística e gráficos, com facilidades computacionais para manipulação de dados, cálculo e exibição gráfica. Dentre a variedade de técnicas estatísticas, estão modelagem linear e não linear, testes estatísticos clássicos, análise de séries temporais, classificação, clustering, etc (R PROJECT, 2013). O ambiente também disponibiliza recursos gráficos e é extensível por meio de pacotes que encapsulam bibliotecas para funções ou áreas de estudo específicas. A instalação do R inclui um conjunto de pacotes padrão, porém, outros pacotes são disponíveis na rede de distribuição do software e podem ser instalados conforme a necessidade do usuário.

A linguagem é amplamente utilizada para fins estatísticos e mineração de dados. Dentre os recursos importantes estão: manipulação eficaz de dados e facilidades de armazenamento; conjunto de operadores para cálculos sobre vetores e matrizes, em particular; uma coleção integrada de ferramentas intermediárias para análise de dados, facilidades gráficas para análise de dados e geração de imagens com qualidade para publicação; uma linguagem de programação que inclui estruturas condicionais, loops, funções recursivas definidas pelo usuário e facilidades de entrada e saída de dados.

O suporte dado pela ferramenta R se fez necessário durante a etapa de pré-processamento dos dados, compreendendo as seguintes tarefas: leitura das bases de dados no formato NetCDF, transformação e geração da massa de dados, aplicação da SVD e geração dos conjuntos de Treinamento e Testes.

#### 5.3.4 Weka

Weka (Waikato Environment for Knowledge Analysis) é uma pacote de software de código aberto que consiste em uma coleção de algoritmos de aprendizado de máquina para solucionar tarefas de mineração de dados. É um produto da Universidade de Waikato (Nova Zelândia), implementado em Java. Os algoritmos pode ser aplicados diretamente a um conjunto de dados ou chamados a partir do seu próprio código Java. No pacote de software estão disponibilizadas ferramentas para pré-processamento de dados, classificação, regressão, *clustering*, regras de associação e visualização.

De acordo com (WAIKATO, 2013), usuários do Weka incluem pesquisadores de aprendizado de máquina e cientistas industriais, sendo também amplamente utilizado para o ensino.

A ferramenta foi utilizada na fase experimental, na qual os algoritmos de aprendizado foram treinados, usando os conjuntos de dados fornecidos; seguindo com a etapa de validação, onde os modelos derivados do treinamento foram aplicados aos conjuntos de testes.

## Capítulo 6: Resultados e Avaliação Crítica

A fase experimental consistiu em uma série de experimentos utilizando as técnicas de Redes Neurais e Árvores de Decisão, aplicadas a conjuntos de treinamentos e em uma metodologia iterativa na qual as configurações dos conjuntos e dos modelos foram combinadas, formando um amplo espaço de opções a serem analisadas em busca da melhor solução.

Neste capítulo serão apresentados os resultados gerados a partir da execução dos experimentos. Estes resultados serão discutidos, selecionados e comparados. Por fim os melhores resultados serão submetidos a uma avaliação geral.

### 6.1 RESULTADOS – REGRESSÃO

Os experimentos cujos resultados são apresentados nesta sessão, têm como objetivo gerar um modelo capaz de fornecer previsões de 1 dia a frente para o valor numérico da intensidade da corrente.

#### 6.1.1 Experimento REG1:

- Conjunto de Treinamento: série de 346 amostras do ano de 2004
- Configuração do conjunto de treinamento: Grade 9x9, representando uma vizinhança com até 4 pontos distantes e um total de 80 vizinhos considerados (Figura 22). Para cada ponto são consideradas as medições do dia corrente e dos dois dias anteriores. A esses dados é aplicado o SVD, de modo que a entrada é composta pela aplicação do SVD com os n últimos dias do ponto analisado inclusos.

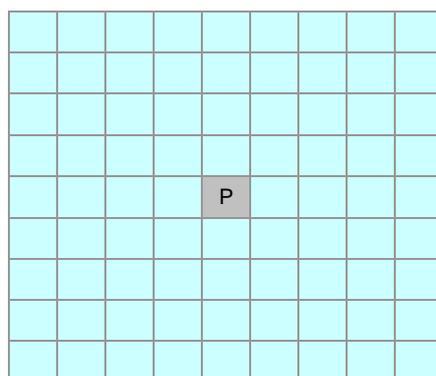


Figura 22: Grade 9x9 com o ponto central e vizinhança de até 4 pontos distantes

**Quadro 8: Formato do conjunto de dados - Experimento REG1**

Entrada				Saída
SVD		Sem SVD		
hoje (d-0)	ontem (d-1)	ante-ontem (d-2)	5 e 10 últimos dias do ponto	d+1

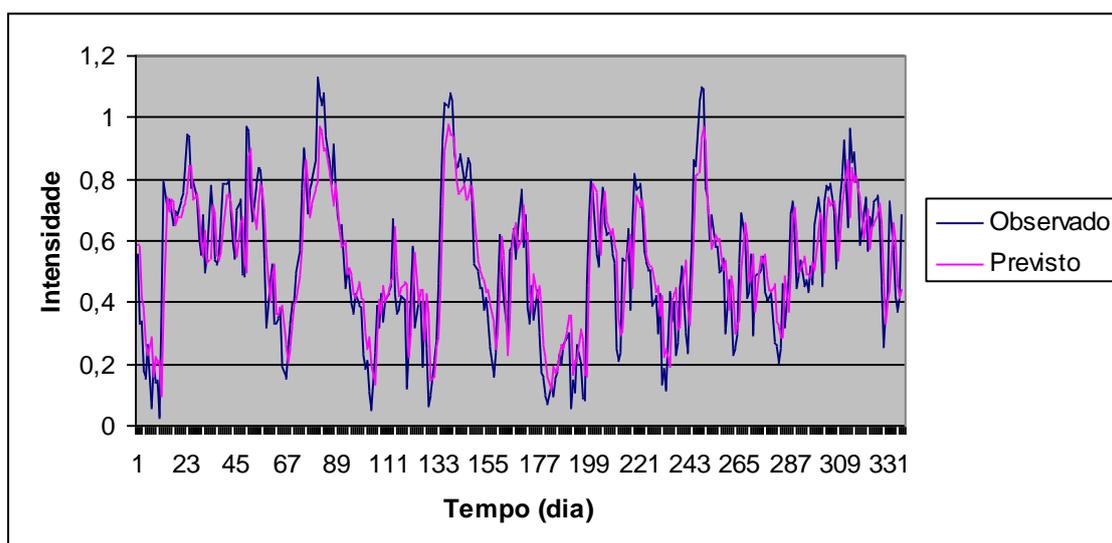
- Conjunto de Testes: série de 336 amostras do ano de 2005
- Parâmetros da rede:
  - número de nós da camada escondida: variando de 8 até 37
  - taxa de aprendizado: 0.1, 0.01 e 0.001
  - momento = 0,2
  - número de iterações = 500
  - tamanho do conjunto de validação = 50% do conjunto de treinamento
  - Validation Threshold = 20

O melhor resultado obtido neste experimento foi para um conjunto de treinamento formado por 20 fatores com a adição de 5 atributos que correspondem à série dos 5 últimos dias medidos para o ponto. O modelo aplicado ao conjunto descrito possui as configurações abaixo.

**Quadro 9: Modelo selecionado para o Experimento REG1**

MultilayerPerceptron
- 1 camada escondida, com 27 nós - taxa de aprendizado = 0,1 - momento = 0,2 - número de iterações = 500 - tamanho do conjunto de validação = 50% do conjunto de treinamento - Validation Threshold = 20

Para melhor visualizar a precisão do modelo, o gráfico de linha abaixo mostra a série prevista através do modelo selecionado, juntamente com a série observada (real). O gráfico evidencia que o modelo consegue prever valores de intensidade de corrente cuja série prevista, de fato, acompanha a série observada.



**Figura 23: Resultado do Experimento REG1 - Gráfico Comparativo das séries observada e prevista**

O resultado do modelo foi comparado com 2 linhas de bases: a) série temporal composta pela simples repetição do valor do dia anterior, e b) série constante formada unicamente pela média dos valores observados. Os erros resultantes de cada caso são listados a seguir:

**Tabela 2: Erros calculados para o modelo do Experimento REG1**

Medida de Erro	Valor
MAE – Série Prevista	0,10258
MAE – Série do Dia Anterior	0,10423
MAE – Série da Média	0,2010
RMSE – Série Prevista	0,1298

Os valores dos erros mostram uma diferença muito pequena em relação ao modelo e a aplicação do dia anterior, sugerindo um ganho não muito representativo. Na tentativa de melhorar o resultado, aumentou-se o conjunto de treinamento através de uma série composta pelos anos de 2004 e 2005. Os parâmetros referentes ao número de iterações e tamanho do conjunto de treinamento também foram ajustados. Os resultados são discutidos no experimento seguinte.

### 6.1.2 Experimento REG2:

- Conjunto de Treinamento: série de 682 amostras do ano de 2004 e 2005

- Configuração do conjunto de treinamento: Mesma configuração do Experimento REG1.

**Quadro 10: Formato do conjunto de dados - Experimento REG2**

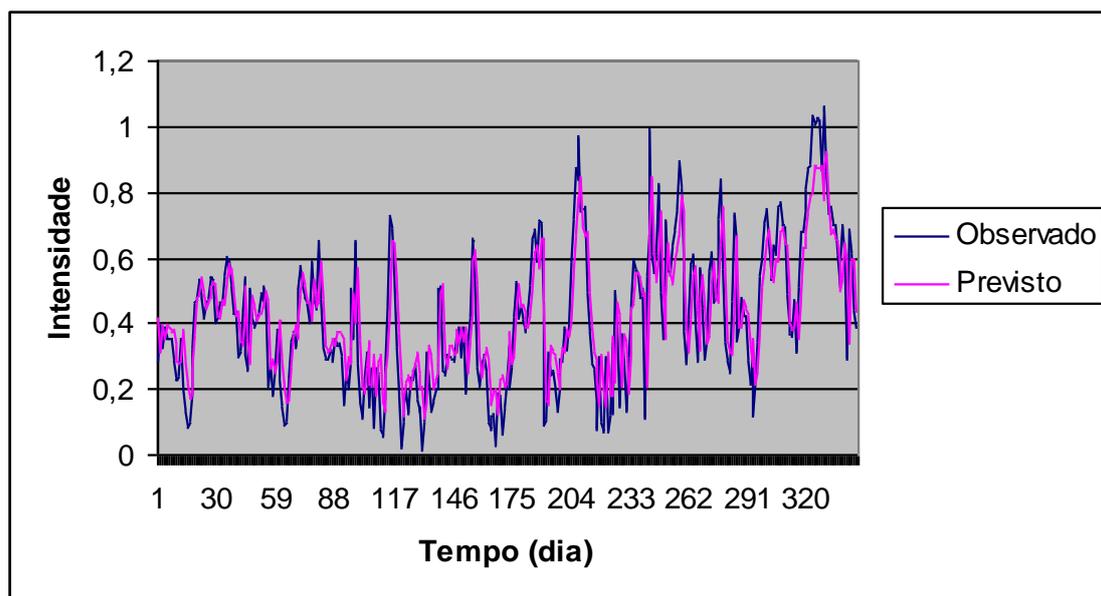
Entrada				Saída
SVD			Sem SVD	
hoje (d-0)	ontem (d-1)	ante-ontem (d-2)	5 e 10 últimos dias do ponto	d+1

- Conjunto de Testes: série de 344 amostras do ano de 2006
- Parâmetros da rede:
  - número de nós da camada escondida: variando de 8 até 37
  - taxa de aprendizado: 0.1, 0.01 e 0.001
  - momento = 0,2
  - número de iterações = 2000
  - tamanho do conjunto de validação = 40% do conjunto de treinamento
  - Validation Threshold = 20

O melhor resultado também foi usando um conjunto com 20 fatores no conjunto de treinamento, acrescentando os 5 últimos dias do ponto, e a rede neural montada com as configurações abaixo:

**Quadro 11: Modelo selecionado para o Experimento REG2**

MultilayerPerceptron
<ul style="list-style-type: none"> <li>- 1 camada escondida, com 35 nós</li> <li>- taxa de aprendizado = 0,001</li> <li>- momento = 0,2</li> <li>- número de iterações = 2000</li> <li>- tamanho do conjunto de validação = 40% do conjunto de treinamento</li> <li>- Validation Threshold = 20</li> </ul>



**Figura 24: Resultado do Experimento REG2 - Gráfico Comparativo das séries observada e prevista**

Obteve-se uma melhora na precisão, dada a pequena diminuição do erro. No entanto, o erro da série prevista ficou abaixo do erro da série formada pelo dia anterior e pela média, sendo que a diferença em relação ao do dia anterior ainda ficou muito pequena.

**Tabela 3: Erros calculados para o modelo do Experimento REG2**

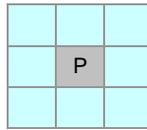
Medida de Erro	Valor
MAE – Série Prevista	0,1016
MAE – Série do Dia Anterior	0,1021
MAE – Série da Média	0,1762
RMSE – Série Prevista	0,1278

### 6.1.3 Experimento REG3:

No terceiro experimento foi formatado um conjunto de treinamento composto por uma grade com os pontos mais próximos do ponto a ser previsto sem a aplicação do SVD e com a inclusão os 5 últimos dias do ponto.

- Conjunto de Treinamento: série de 356 amostras do ano de 2004
- Configuração do conjunto de treinamento: Grade 3x3, representando uma vizinhança com os pontos mais próximos e um total de 8 vizinhos (Figura 24). A formação dos dados de entrada é análoga à realizada nos experimentos

anteriores, diferenciando pelo fato de serem adicionados apenas os 5 últimos dias do ponto analisado.



**Figura 25: Grade 3x3 com o ponto central e vizinhança com os pontos mais próximos**

**Quadro 12: Formato do conjunto de dados - Experimento REG3**

Entrada Sem SVD				Saída
hoje (d-0)	ontem (d-1)	ante-ontem (d-2)	5 últimos dias do ponto	d+1

- Conjunto de Testes: série de 346 amostras do ano de 2005
- Parâmetros da rede:
  - número de nós da camada escondida: variando de 8 até 37
  - taxa de aprendizado: 0.1, 0.01 e 0.001
  - momento = 0,2
  - número de iterações = 4000
  - tamanho do conjunto de validação = 40% do conjunto de treinamento
  - Validation Threshold = 20

O modelo de rede neural com 16 nós na camada oculta e taxa de aprendizado de 0,01 foi o que apresentou o menor erro, 0,097.

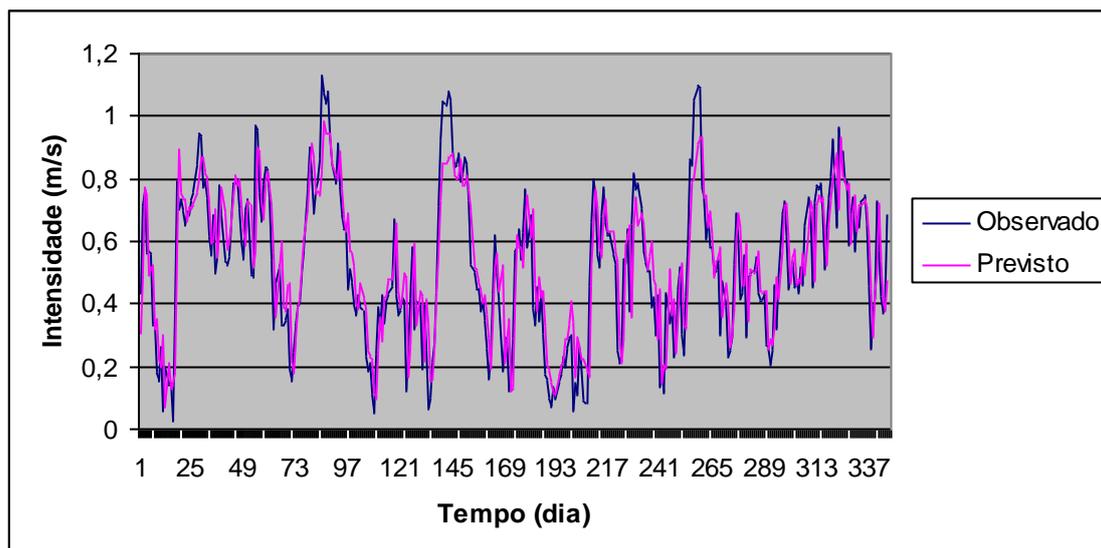


Figura 26: Resultado do Experimento REG3 - Gráfico Comparativo das séries observada e prevista

Quadro 13: Modelo selecionado para o Experimento REG3

MultilayerPerceptron
- 1 camada escondida, com 16 nós
- taxa de aprendizado = 0,01
- momento = 0,2
- número de iterações = 4000
- tamanho do conjunto de validação = 40% do conjunto de treinamento
- Validation Threshold = 20

Este resultado mostrou uma melhor taxa de acerto e ganho em relação a aplicação do modelo, quando comparado com as 2 linhas base.

Tabela 4: Erros calculados para o modelo do Experimento REG3

Medida de Erro	Valor
MAE – Série Prevista	0,0976
MAE – Série do Dia Anterior	0,1044
MAE – Série da Média	0,2012
RMSE – Série Prevista	0,1260

#### 6.1.4 Experimento REG4

No quarto experimento foi formatado um conjunto de treinamento composto por uma grade com os pontos mais próximos do ponto a ser previsto, com a aplicação do SVD e com a inclusão dos 5 últimos dias do ponto.

- Conjunto de Treinamento: série de 356 amostras do ano de 2004

- Configuração do conjunto de treinamento: Grade 3x3, representando uma vizinhança com os pontos mais próximos e um total de 8 vizinhos (Figura 25). A formação dos dados de entrada é análoga à do experimento anterior, porém neste caso, houve a aplicação do SVD, tendo sido realizados testes com 20 e 25.

**Quadro 14: Formato do conjunto de dados - Experimento REG4**

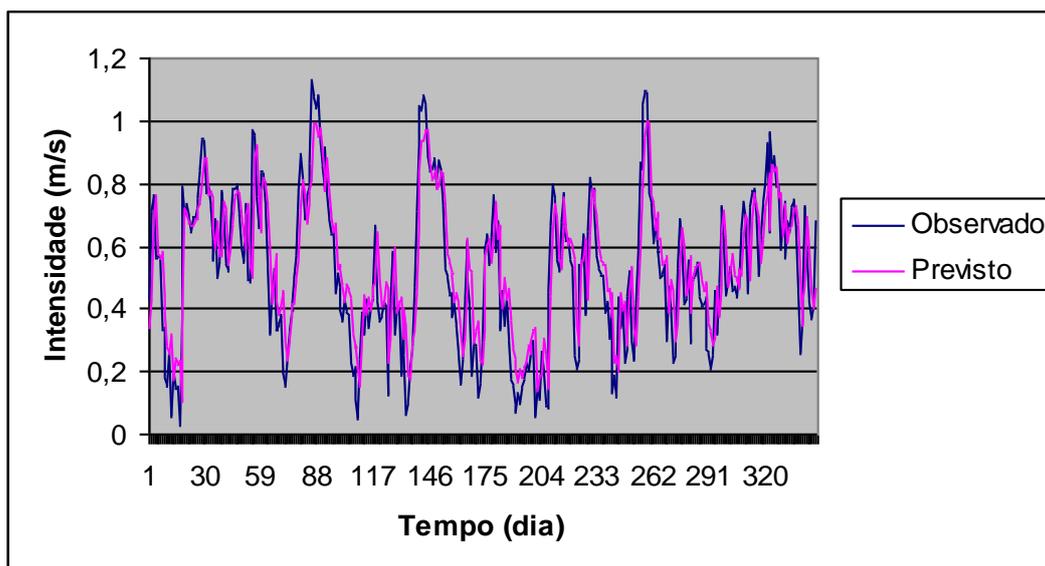
Entrada				Saída
SVD			Sem SVD	
hoje (d-0)	ontem (d-1)	ante-ontem (d-2)	5 últimos dias do ponto	d+1

- Conjunto de Testes: série de 346 amostras do ano de 2005
- Parâmetros da rede:
  - número de nós da camada escondida: variando de 8 até 37
  - taxa de aprendizado: 0.1, 0.01 e 0.001
  - momento = 0,2
  - número de iterações = 4000
  - tamanho do conjunto de validação = 40% do conjunto de treinamento
  - Validation Threshold = 20

O modelo de rede neural que apresentou menor erro é detalhado no quadro abaixo. Este resultado refere-se ao experimento realizado com um conjunto de treinamento formado por 20 fatores.

**Quadro 15: Modelo selecionado para o Experimento REG4**

MultilayerPerceptron
- 1 camada escondida, com 15 nós - taxa de aprendizado = 0,001 - momento = 0,2 - número de iterações = 4000 - tamanho do conjunto de validação = 40% do conjunto de treinamento - Validation Threshold = 20



**Figura 27: Resultado do Experimento REG4 - Gráfico Comparativo das séries observada e prevista**

Quando comparado com as duas linhas base, o resultado apresentou um erro melhor apenas em relação ao erro da série da média, tendo sido um pouco pior em relação ao erro da série do dia anterior. Como o conjunto de treinamento não é formado por um número grande de dimensões, a aplicação da SVD pode ter provocado uma perda de informações relevantes e isso impactou na qualidade do aprendizado.

**Tabela 5: Erros calculados para o modelo do Experimento REG4**

Medida de Erro	Valor
MAE – Série Prevista	0,1058
MAE – Série do Dia Anterior	0,1044
MAE – Série da Média	0,2012
RMSE – Série Prevista	0,1351

## 6.2 RESULTADOS – CLASSIFICAÇÃO ESCALA 20-20 (INTERVALOS DE 20 cm/s)

### 6.2.1 Experimento C1

Nesse experimento foi considerado um conjunto de treinamento como sendo a grade com os pontos mais próximos do ponto a ser previsto (grade representada pela Figura 25), sem SVD e adicionando-se os 5 últimos dias do ponto.

- Conjunto de Treinamento: 456 amostras derivadas de um balanceamento com 76 instâncias para cada classe e coletadas nos anos de 2004, 2005, 2006, 2007, 2008
- Configuração do conjunto de treinamento: Mesma configuração do Experimento REG3 (Quadro 12).
- Conjunto de Testes: série de 355 amostras do ano de 2009

#### a) Rede Neural

- Parâmetros da rede:
  - número de nós da camada escondida: variando de 8 até 37
  - taxa de aprendizado: 0.1, 0.01 e 0.001
  - momento = 0,2
  - número de iterações = 4000
  - tamanho do conjunto de validação = 40% do conjunto de treinamento
  - Validation Threshold = 20
- Resultados

**Quadro 16: Modelos de Rede Neural selecionados para o Experimento C1**

Modelo1		
MultilayerPerceptron		
- 1 camada escondida, com 9 nós		
- taxa de aprendizado = 0,01		
Instâncias Classificadas Corretamente	192	54.0845 %
Instâncias Classificadas Incorretamente	163	45.9155 %
Número Total de Instâncias	355	
Matriz de Confusão		
<pre> a  b  c  d  e  f  &lt;-- classified as 72  6  0  0  0  0   a = 20 40 52 17  4  0  0   b = 40 10 20 33 21  2  1   c = 60  0  2  5 31  3  4   d = 80  0  0  3 10  2 14   e = 100  0  0  0  1  0  2   f = 120           </pre>		
Modelo2		
MultilayerPerceptron		
- 1 camada escondida, com 11 nós		
- taxa de aprendizado = 0,01		
Instâncias Classificadas Corretamente	182	51.2676 %
Instâncias Classificadas Incorretamente	173	48.7324 %
Número Total de Instâncias	355	

Matriz de Confusão						
a	b	c	d	e	f	<-- classified as
72	6	0	0	0	0	a = 20
42	47	21	3	0	0	b = 40
10	21	31	21	3	1	c = 60
0	2	7	28	3	5	d = 80
0	0	2	11	2	14	e = 100
0	0	0	0	1	2	f = 120
Modelo3						
MultilayerPerceptron						
- 1 camada escondida, com 12 nós						
- taxa de aprendizado = 0,01						
Instâncias Classificadas Corretamente	177	49.8592 %				
Instâncias Classificadas Incorretamente	178	50.1408 %				
Número Total de Instâncias	355					
Matriz de Confusão						
a	b	c	d	e	f	<-- classified as
71	7	0	0	0	0	a = 20
45	43	22	3	0	0	b = 40
10	21	30	24	1	1	c = 60
0	3	6	28	4	4	d = 80
0	0	3	9	3	14	e = 100
0	0	0	1	0	2	f = 120
Modelo4						
MultilayerPerceptron						
- 1 camada escondida, com 10 nós						
- taxa de aprendizado = 0,01						
Instâncias Classificadas Corretamente	189	53.2394 %				
Instâncias Classificadas Incorretamente	166	46.7606 %				
Número Total de Instâncias	355					
Matriz de Confusão						
a	b	c	d	e	f	<-- classified as
72	6	0	0	0	0	a = 20
43	50	15	5	0	0	b = 40
10	20	30	25	1	1	c = 60
0	2	4	31	4	4	d = 80
0	0	2	10	4	13	e = 100
0	0	0	1	0	2	f = 120

## b) Árvores de Decisão

- Resultados

**Quadro 17: Modelos de Árvore de Decisão selecionados para o Experimento C1**

Modelo 5		
SimpleCart		
Instâncias Classificadas Corretamente	171	48.169 %
Instâncias Classificadas Incorretamente	184	51.831 %
Número Total de Instâncias	355	

Matriz de Confusão							
a	b	c	d	e	f	<-- classified as	
66	11	1	0	0	0	a = 20	
43	39	24	6	1	0	b = 40	
8	24	24	26	5	0	c = 60	
0	1	5	29	8	2	d = 80	
0	0	1	9	11	8	e = 100	
0	0	0	0	1	2	f = 120	
Modelo 6							
REPTree							
Instâncias Classificadas Corretamente						157	44.2254 %
Instâncias Classificadas Incorretamente						198	55.7746 %
Número Total de Instâncias						355	
Matriz de Confusão							
a	b	c	d	e	f	<-- classified as	
72	6	0	0	0	0	a = 20	
56	36	17	3	1	0	b = 40	
15	30	20	15	6	1	c = 60	
0	3	7	15	17	3	d = 80	
0	1	3	6	12	7	e = 100	
0	0	0	0	1	2	f = 120	
Modelo 7							
J48							
Instâncias Classificadas Corretamente						145	40.8451 %
Instâncias Classificadas Incorretamente						210	59.1549 %
Número Total de Instâncias						355	
Matriz de Confusão							
a	b	c	d	e	f	<-- classified as	
69	5	3	1	0	0	a = 20	
53	29	22	6	3	0	b = 40	
17	18	23	22	6	1	c = 60	
2	3	13	15	9	3	d = 80	
0	0	2	12	7	8	e = 100	
0	0	0	0	1	2	f = 120	
Modelo 8							
J48							
Instâncias Classificadas Corretamente						179	50.4225 %
Instâncias Classificadas Incorretamente						176	49.5775 %
Número Total de Instâncias						355	
Matriz de Confusão							
a	b	c	d	e	f	<-- classified as	
71	7	0	0	0	0	a = 20	
45	41	24	3	0	0	b = 40	
10	20	31	24	1	1	c = 60	
0	2	7	30	2	4	d = 80	
0	0	3	9	4	13	e = 100	
0	0	0	1	0	2	f = 120	

### 6.2.1 Experimento C2

A configuração do conjunto de treinamento utilizada para este experimento corresponde à grade com os pontos mais próximos do ponto a ser previsto (grade

representada pela Figura 25), com SVD e muitos fatores, incluindo os 5 últimos dias do ponto.

- Conjunto de Treinamento: 456 amostras derivadas de um balanceamento com 76 instâncias para cada classe e coletadas nos anos de 2004, 2005, 2006, 2007, 2008
- Configuração do conjunto de treinamento: Grade representada pela Figura 25 e formato do conjunto de dados representado pelo Quadro 14.
- Número de fatores: 20 e 25
- Conjunto de Testes: série de 355 amostras do ano de 2009

**a) Rede Neural**

- Parâmetros da rede:
  - número de nós da camada escondida: variando de 8 até 37
  - taxa de aprendizado: 0.1, 0.01 e 0.001
  - momento = 0,2
  - número de iterações = 4000
  - tamanho do conjunto de validação = 40% do conjunto de treinamento
  - Validation Threshold = 20
- Resultados

**Quadro 18: Modelos de Rede Neural selecionados para o Experimento C2**

Modelo1						
MultilayerPerceptron						
SVD com 25 fatores						
- 1 camada escondida, com 16 nós						
- taxa de aprendizado = 0,01						
Instâncias Classificadas Corretamente	89	25.0704 %				
Instâncias Classificadas Incorretamente	266	74.9296 %				
Número Total de Instâncias	355					
Matriz de Confusão						
a	b	c	d	e	f	<-- classified as
68	9	0	0	0	1	a = 20
92	11	5	0	0	5	b = 40
60	8	10	0	0	9	c = 60
26	1	6	0	0	12	d = 80
23	1	0	0	0	5	e = 100
3	0	0	0	0	0	f = 120

Modelo2		
MultilayerPerceptron		
SVD com 25 fatores		
- 1 camada escondida, com 23 nós		
- taxa de aprendizado = 0,01		
Instâncias Classificadas Corretamente	90	25.3521 %
Instâncias Classificadas Incorretamente	265	74.6479 %
Número Total de Instâncias	355	
Matriz de Confusão		
<pre> a b c d e f &lt;-- classified as 69 9 0 0 0 0   a = 20 95 8 6 0 0 4   b = 40 60 9 13 0 0 5   c = 60 25 1 14 0 0 5   d = 80 22 0 2 0 0 5   e = 100 3 0 0 0 0 0   f = 120 </pre>		

### b) Árvores de Decisão

- Resultados

**Quadro 19: Modelos de Árvore de Decisão selecionados para o Experimento C2**

Modelo 3		
LADTree		
SVD com 20 fatores		
Instâncias Classificadas Corretamente	162	45.6338 %
Instâncias Classificadas Incorretamente	193	54.3662 %
Número Total de Instâncias	355	
Matriz de Confusão		
<pre> a b c d e f &lt;-- classified as 73 4 1 0 0 0   a = 20 56 24 30 0 3 0   b = 40 13 13 43 0 18 0   c = 60 0 1 10 0 33 1   d = 80 0 0 4 0 22 3   e = 100 0 0 0 0 3 0   f = 120 </pre>		
Modelo 4		
SimpleCart		
SVD com 20 fatores		
Instâncias Classificadas Corretamente	141	39.7183 %
Instâncias Classificadas Incorretamente	214	60.2817 %
Número Total de Instâncias	355	
Matriz de Confusão		
<pre> a b c d e f &lt;-- classified as 73 4 1 0 0 0   a = 20 56 33 21 0 2 1   b = 40 13 29 27 0 14 4   c = 60 0 8 3 0 24 10   d = 80 0 2 2 0 6 19   e = 100 0 0 0 0 1 2   f = 120 </pre>		
Modelo 5		
SimpleCart		
SVD com 20 fatores		
Instâncias Classificadas Corretamente	156	43.9437 %
Instâncias Classificadas Incorretamente	193	54.3662 %

Número Total de Instâncias							355
Matriz de Confusão							
a	b	c	d	e	f	<-- classified as	
73	4	1	0	0	0		a = 20
56	24	30	0	3	0		b = 40
13	13	43	0	18	0		c = 60
0	1	10	0	33	1		d = 80
0	0	4	0	22	3		e = 100
0	0	0	0	3	0		f = 120

### 6.2.2 Experimento C3

O conjunto de treinamento utilizado neste experimento foi formado a partir de uma grade com vizinhança de 4 pontos (grade 9x9 representada pela Figura 22), com SVD aplicado e com os 5 últimos dias do ponto incluídos.

- Conjunto de Treinamento: 456 amostras derivadas de um balanceamento com 76 instâncias para cada classe e coletadas nos anos de 2004, 2005, 2006, 2007, 2008
- Configuração do conjunto de treinamento: Grade representada pela figura 22 e formato do conjunto de dados representado pelo Quadro 14.
- Número de fatores: 20 e 25
- Conjunto de Testes: série de 355 amostras do ano de 2009

#### a) Rede Neural

- Parâmetros da rede:
  - número de nós da camada escondida: variando de 8 até 37
  - taxa de aprendizado: 0.1, 0.01 e 0.001
  - momento = 0,2
  - número de iterações = 4000
  - tamanho do conjunto de validação = 40% do conjunto de treinamento
  - Validation Threshold = 20
- Resultados

**Quadro 20: Modelos de Rede Neural selecionados para o Experimento C3**

Modelo1
---------

<b>MultilayerPerceptron</b>		
SVD com 20 fatores		
- 1 camada escondida, com 21 nós		
- taxa de aprendizado = 0,01		
Instâncias Classificadas Corretamente	171	48.169 %
Instâncias Classificadas Incorretamente	184	51.831 %
Número Total de Instâncias	355	
Matriz de Confusão		
<pre> a b c d e f &lt;-- classified as 58 18 2 0 0 0   a = 20 31 51 17 14 0 0   b = 40 5 26 16 36 3 1   c = 60 0 1 0 36 7 1   d = 80 0 0 1 13 10 5   e = 100 0 0 0 1 2 0   f = 120 </pre>		
<b>Modelo 2</b>		
<b>MultilayerPerceptron</b>		
SVD com 20 fatores		
- 1 camada escondida, com 20 nós		
- taxa de aprendizado = 0,01		
Instâncias Classificadas Corretamente	170	47.8873 %
Instâncias Classificadas Incorretamente	185	52.1127 %
Número Total de Instâncias	355	
Matriz de Confusão		
<pre> a b c d e f &lt;-- classified as 58 18 2 0 0 0   a = 20 29 53 19 12 0 0   b = 40 5 28 15 36 2 1   c = 60 0 1 1 38 3 2   d = 80 0 0 1 16 6 6   e = 100 0 0 0 1 2 0   f = 120 </pre>		
<b>Modelo 3</b>		
<b>MultilayerPerceptron</b>		
SVD com 25 fatores		
- 1 camada escondida, com 27 nós		
- taxa de aprendizado = 0,01		
Instâncias Classificadas Corretamente	150	42.2535 %
Instâncias Classificadas Incorretamente	205	57.7465 %
Número Total de Instâncias	355	
Matriz de Confusão		
<pre> a b c d e f &lt;-- classified as 58 15 5 0 0 0   a = 20 34 37 20 20 2 0   b = 40 5 16 17 46 3 0   c = 60 0 2 2 33 4 4   d = 80 0 0 1 17 4 7   e = 100 0 0 0 0 2 1   f = 120 </pre>		
<b>Modelo 4</b>		
<b>MultilayerPerceptron</b>		
SVD com 20 fatores		
- 1 camada escondida, com 12 nós		
- taxa de aprendizado = 0,01		
Instâncias Classificadas Corretamente	157	44.2254 %
Instâncias Classificadas Incorretamente	198	55.7746 %

Número Total de Instâncias	355
Matriz de Confusão	
a b c d e f	<-- classified as
59 12 7 0 0 0	a = 20
33 33 29 18 0 0	b = 40
6 15 21 44 0 1	c = 60
0 0 2 39 1 3	d = 80
0 0 1 18 4 6	e = 100
0 0 0 1 1 1	f = 120

## b) Árvores de Decisão

- Resultados

**Quadro 21: Modelos de Árvore de Decisão selecionados para o Experimento C3**

Modelo 5		
J48		
SVD com 20 fatores		
Instâncias Classificadas Corretamente	161	45.3521 %
Instâncias Classificadas Incorretamente	194	54.6479 %
Número Total de Instâncias	355	
Matriz de Confusão		
a b c d e f	<-- classified as	
71 6 1 0 0 0	a = 20	
54 34 23 1 1 0	b = 40	
17 25 29 14 2 0	c = 60	
1 5 14 19 5 1	d = 80	
1 0 4 10 8 6	e = 100	
0 0 0 3 0 0	f = 120	
Modelo 6		
J48graft		
SVD com 20 fatores		
Instâncias Classificadas Corretamente	163	45.9155 %
Instâncias Classificadas Incorretamente	192	54.0845 %
Número Total de Instâncias	355	
Matriz de Confusão		
a b c d e f	<-- classified as	
72 5 1 0 0 0	a = 20	
54 34 23 2 0 0	b = 40	
17 25 32 12 1 0	c = 60	
1 5 17 18 3 1	d = 80	
1 0 4 13 7 4	e = 100	
0 0 0 3 0 0	f = 120	

## 6.3 RESULTADOS – CLASSIFICAÇÃO 3 ESCALAS

Os experimentos realizados para a classificação baseada na escala por segurança operacional (3 classes) seguiram o mesmo procedimento daqueles realizados na seção anterior.

### 6.3.1 Experimento C4

- Configuração: Grade com os pontos mais próximos do ponto a ser previsto, incluindo os 5 últimos dias do ponto, sem SVD
- Conjunto de Treinamento: 456 amostras derivadas de um balanceamento com 272 instâncias para cada classe e coletadas nos anos de 2004, 2005, 2006, 2007, 2008
- Configuração do conjunto de treinamento: Grade representada pela Figura 25 e formato do conjunto de dados representado pelo Quadro 12.
- Conjunto de Testes: série de 355 amostras do ano de 2009

#### a) Rede Neural

- Parâmetros da rede:
  - número de nós da camada escondida: variando de 8 até 37
  - taxa de aprendizado: 0.1, 0.01 e 0.001 e 0.001
  - momento = 0,2
  - número de iterações = 4000
  - tamanho do conjunto de validação = 40% do conjunto de treinamento
  - Validation Threshold = 20
- Resultados

**Quadro 22: Modelo de Rede Neural selecionado para o Experimento C4**

Modelo 1			
MultilayerPerceptron			
- 1 camada escondida, com 11 nós			
- taxa de aprendizado = 0,001			
Instâncias Classificadas Corretamente	292		82.2535 %
Instâncias Classificadas Incorretamente	63		17.7465 %
Número Total de Instâncias	355		
Matriz de Confusão			
a	b	c	<-- classified as
210	20	2	a = 1
16	60	15	b = 2
0	10	22	c = 3

#### b) Árvores de Decisão

- Resultados

**Quadro 23: Modelo de Árvore de Decisão selecionado para o Experimento C4**

Modelo 2			
LADTree			
SVD com 20 fatores			
Instâncias Classificadas Corretamente	284	80	%
Instâncias Classificadas Incorretamente	71	20	%
Número Total de Instâncias	355		
Matriz de Confusão			
a	b	c	<-- classified as
204	27	1	a = 1
15	60	16	b = 2
0	12	20	c = 3

### 6.3.2 Experimento C5

- Configuração: Grade com os pontos mais próximos do ponto a ser previsto, incluindo os 5 últimos dias do ponto, com SVD e muitos fatores
- Conjunto de Treinamento: 456 amostras derivadas de um balanceamento com 272 instâncias para cada classe e coletadas nos anos de 2004, 2005, 2006, 2007, 2008
- Configuração do conjunto de treinamento: Grade representada pela Figura 25 e formato do conjunto de dados representado pelo Quadro 14.
- Número de fatores: 20 e 25
- Conjunto de Testes: série de 355 amostras do ano de 2009

#### a) Rede Neural

- Parâmetros da rede:
  - número de nós da camada escondida: variando de 8 até 37
  - taxa de aprendizado: 0.1, 0.01 e 0.001 e 0.001
  - momento = 0,2
  - número de iterações = 4000
  - tamanho do conjunto de validação = 40% do conjunto de treinamento
  - Validation Threshold = 20

- Resultados

**Quadro 24: Modelos de Rede Neural selecionados para o Experimento C5**

Modelo 1			
MultilayerPerceptron			
SVD com 20 fatores			
- 1 camada escondida, com 24 nós			
- taxa de aprendizado = 0,001			
Instâncias Classificadas Corretamente	281	79.1549 %	
Instâncias Classificadas Incorretamente	74	20.8451 %	
Número Total de Instâncias	355		
Matriz de Confusão			
a	b	c	<-- classified as
225	5	2	a = 1
42	32	17	b = 2
4	4	24	c = 3

### *b) Árvores de Decisão*

- Resultados

**Quadro 25: Modelo de Árvore de Decisão selecionado para o Experimento C5**

Modelo 3			
SimpleCart			
SVD com 25 fatores			
Instâncias Classificadas Corretamente	294	82.8169 %	
Instâncias Classificadas Incorretamente	61	17.1831 %	
Número Total de Instâncias	355		
Matriz de Confusão			
a	b	c	<-- classified as
223	8	1	a = 1
29	51	11	b = 2
2	10	20	c = 3

### **6.2.3 Experimento C6**

- Configuração: Grade com vizinhança de 4 pontos (9x9), incluindo os 5 últimos dias do ponto, com SVD
- Conjunto de Treinamento: 456 amostras derivadas de um balanceamento com 272 instâncias para cada classe e coletadas nos anos de 2004, 2005, 2006, 2007, 2008
- Configuração do conjunto de treinamento: Grade representada pela figura 22 e formato do conjunto de dados representado pelo Quadro 14.

- Número de fatores: 20 e 25
- Conjunto de Testes: série de 355 amostras do ano de 2009

**a) Rede Neural**

- Parâmetros da rede:
  - número de nós da camada escondida: variando de 8 até 37
  - taxa de aprendizado: 0.1, 0.01 e 0.001
  - momento = 0,2
  - número de iterações = 4000
  - tamanho do conjunto de validação = 40% do conjunto de treinamento
  - Validation Threshold = 20
- Resultados

**Quadro 26: Modelo de Rede Neural selecionado para o Experimento C6**

Modelo 1		
MultilayerPerceptron		
SVD com 20 fatores		
- 1 camada escondida, com 24 nós		
- taxa de aprendizado = 0,001		
Instâncias Classificadas Corretamente	281	79.1549 %
Instâncias Classificadas Incorretamente	74	20.8451 %
Número Total de Instâncias	355	
Matriz de Confusão		
a	b	c <-- classified as
225	5	2   a = 1
42	32	17   b = 2
4	4	24   c = 3

**b) Árvores de Decisão**

- Resultados

**Quadro 27: Modelos de Árvore de Decisão selecionados para o Experimento C6**

Modelo 2		
LADTree		
SVD com 20 fatores		
Instâncias Classificadas Corretamente	283	79.7183 %
Instâncias Classificadas Incorretamente	72	20.2817 %
Número Total de Instâncias	355	

Matriz de Confusão			
a	b	c	<-- classified as
202	24	6	a = 1
12	53	26	b = 2
0	4	28	c = 3
Modelo 3			
REPTree			
SVD com 30 fatores			
Instâncias Classificadas Corretamente	291	81.9718 %	
Instâncias Classificadas Incorretamente	64	18.0282 %	
Número Total de Instâncias	355		
Matriz de Confusão			
a	b	c	<-- classified as
223	7	2	a = 1
29	46	16	b = 2
2	8	22	c = 3
Modelo 4			
LADTree			
SVD com 30 fatores			
Instâncias Classificadas Corretamente	287	80.8451 %	
Instâncias Classificadas Incorretamente	68	19.1549 %	
Número Total de Instâncias	355		
Matriz de Confusão			
a	b	c	<-- classified as
223	3	6	a = 1
29	36	26	b = 2
2	2	28	c = 3

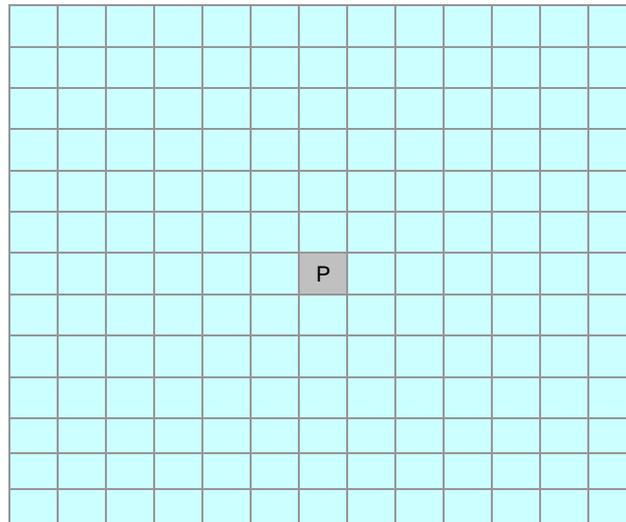
## 6.4 RESULTADOS – PREVISÃO 2 DIAS A FRENTE

Essa etapa da fase experimental buscou repetir os procedimentos realizados nos experimentos anteriores de regressão e classificação para o dia seguinte, tendo como objetivo aumentar a escala de previsão. Neste caso, a escala de previsão foi estendida para 2 dias à frente. Os resultados dessa etapa serão apresentados nas sessões seguintes.

### 6.4.1 Experimento REG5:

- Configuração: Grade com vizinhança de 6 pontos (13x13), incluindo os 3 últimos dias do ponto, com SVD
- Conjunto de Treinamento: série de 346 amostras do ano de 2004+2005
- Configuração do conjunto de treinamento: Grade 13x13, representando uma vizinhança com os pontos mais próximos e um total de 168 vizinhos (Figura 28).

A formação dos dados de entrada é análoga à realizada nos experimentos anteriores, diferenciando pelo fato de serem adicionados apenas os 5 últimos dias do ponto analisado após a aplicação do SVD.



**Figura 28: Grade 13x13 com o ponto central e vizinhança com os pontos mais próximos**

**Quadro 28: Formato do conjunto de dados - Experimento REG5**

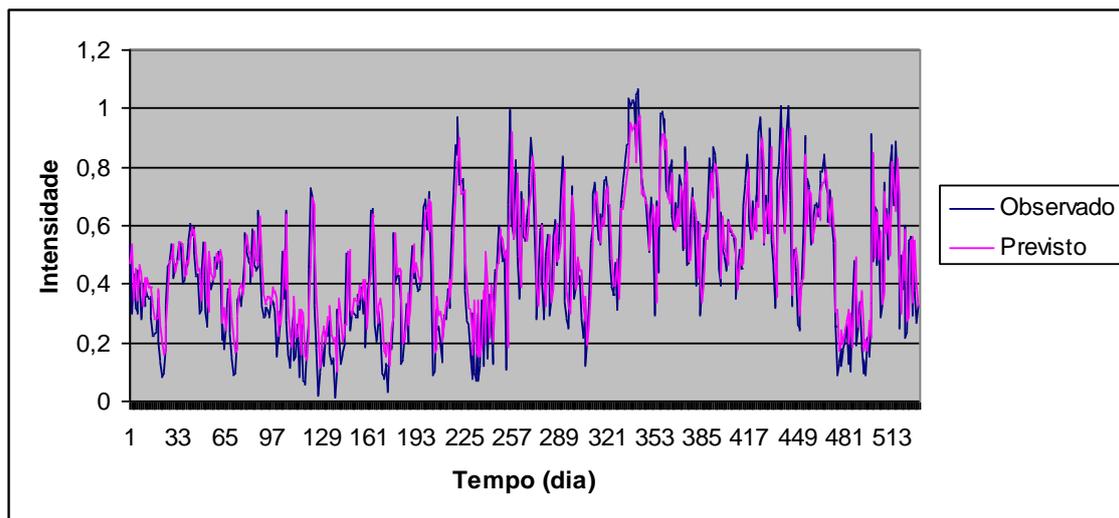
Entrada				Saída
SVD			Sem SVD	
hoje (d-0)	ontem (d-1)	ante-ontem (d-2)	3 últimos dias do ponto	d+2

- Conjunto de Testes: série de 336 amostras do ano de 2006+2007
- Parâmetros da rede:
  - número de nós da camada escondida: variando de 8 até 37
  - taxa de aprendizado: 0.1, 0.01 e 0.001
  - momento = 0,2
  - número de iterações = 4000
  - tamanho do conjunto de validação = 40% do conjunto de treinamento
  - Validation Threshold = 20

O melhor resultado obtido neste experimento foi para um conjunto de treinamento formado por 20 fatores e modelo com as configurações abaixo, cujo erro foi de 0,1088.

**Quadro 29: Modelo selecionado para o Experimento REG5**

Modelo 1
MultilayerPerceptron
SVD com 20 fatores
- 1 camada escondida, com 9 nós - taxa de aprendizado = 0,001 - momento = 0,2 - número de iterações = 4000 - tamanho do conjunto de validação = 40% do conjunto de treinamento - Validation Threshold = 20



**Figura 29: Resultado do Experimento REG5 - Gráfico Comparativo das séries observada e prevista**

O resultado do modelo foi comparado com 2 linhas de bases: a) série temporal composta pela simples repetição do valor do dia anterior, e b) série constante formada unicamente pela média dos valores observados. Os erros resultantes de cada caso são listados a seguir:

**Tabela 6: Erros calculados para o modelo do Experimento REG5**

Medida de Erro	Valor
MAE – Série Prevista	0,1088
MAE – Série do Dia Anterior	0,1127
MAE – Série da Média	0,1890
RMSE – Série Prevista	0,1381

#### 6.4.2 Experimento C7 - Classificação Escala 20-20 (Intervalos de 20 cm/s)

- Configuração: Grade com vizinhança de 6 pontos (13x13), incluindo os 3 últimos dias do ponto, com SVD
- Conjunto de Treinamento: 456 amostras derivadas de um balanceamento com 76 instâncias para cada classe e coletadas nos anos de 2004, 2005, 2006, 2007, 2008
- Configuração do conjunto de treinamento: Grade representada pela Figura 28 e formato do conjunto de dados representado pelo Quadro 28.
- Conjunto de Testes: série de 534 amostras do ano de 2009+2010

#### a) Rede Neural

- Parâmetros da rede:
  - número de nós da camada escondida: variando de 8 até 37
  - taxa de aprendizado: 0.1, 0.01 e 0.001
  - momento = 0,2
  - número de iterações = 4000
  - tamanho do conjunto de validação = 40% do conjunto de treinamento
  - Validation Threshold = 20
- Resultados

**Quadro 30: Modelos de Rede Neural selecionados para o Experimento C7**

Modelo 1						
MultilayerPerceptron						
SVD com 25 fatores						
- 1 camada escondida, com 12 nós						
- taxa de aprendizado = 0,01						
Instâncias Classificadas Corretamente	253	47.3783 %				
Instâncias Classificadas Incorretamente	281	52.6217 %				
Número Total de Instâncias	534					
Matriz de Confusão						
a	b	c	d	e	f	<-- classified as
98	9	4	1	0	0	a = 20
74	52	38	3	0	1	b = 40
15	31	61	13	2	3	c = 60
0	1	26	27	14	8	d = 80
0	0	8	5	7	24	e = 100
0	0	0	1	0	8	f = 120
Modelo 2						
MultilayerPerceptron						
SVD com 25 fatores						

- 1 camada escondida, com 10 nós						
- taxa de aprendizado = 0,01						
Instâncias Classificadas Corretamente	268	50.1873 %				
Instâncias Classificadas Incorretamente	266	49.8127 %				
Número Total de Instâncias	534					
Matriz de Confusão						
a	b	c	d	e	f	<-- classified as
96	15	0	1	0	0	a = 20
72	79	13	3	0	1	b = 40
13	53	38	18	0	3	c = 60
0	4	19	41	6	6	d = 80
0	0	7	13	6	18	e = 100
0	0	0	1	0	8	f = 120

### b) Árvores de Decisão

- Resultados

**Quadro 31: Modelo de Árvore de Decisão selecionado para o Experimento C7**

Modelo 3						
MultilayerPerceptron						
SVD com 25 fatores						
Instâncias Classificadas Corretamente	248	46.4419 %				
Instâncias Classificadas Incorretamente	286	53.5581 %				
Número Total de Instâncias	534					
Matriz de Confusão						
a	b	c	d	e	f	<-- classified as
100	9	2	1	0	0	a = 20
84	60	20	3	1	0	b = 40
20	59	13	28	5	0	c = 60
0	14	0	46	15	1	d = 80
0	3	2	9	24	6	e = 100
0	0	0	1	3	5	f = 120

### 6.4.3 Experimento C8 - Classificação 3 Escalas

- Configuração: Grade com vizinhança de 6 pontos (13x13), incluindo os 3 últimos dias do ponto, com SVD
- Conjunto de Treinamento: 456 amostras derivadas de um balanceamento com 272 instâncias para cada classe e coletadas nos anos de 2004, 2005, 2006, 2007, 2008
- Configuração do conjunto de treinamento: Grade representada pela figura 25 e formato do conjunto de dados representado pelo Quadro 28.
- Conjunto de Testes: série de 355 amostras do ano de 2009+2010

### a) Rede Neural

- Parâmetros da rede:
  - número de nós da camada escondida: variando de 8 até 37
  - taxa de aprendizado: 0.1, 0.01 e 0.001
  - momento = 0,2
  - número de iterações = 4000
  - tamanho do conjunto de validação = 40% do conjunto de treinamento
  - Validation Threshold = 20
- Resultados

**Quadro 32: Modelo de Rede Neural selecionado para o Experimento C8**

Modelo 1			
MultilayerPerceptron			
SVD com 25 fatores			
- 1 camada escondida, com 10 nós			
- taxa de aprendizado = 0,001			
Instâncias Classificadas Corretamente	434	81.2734 %	
Instâncias Classificadas Incorretamente	100	18.7266 %	
Número Total de Instâncias	534		
Matriz de Confusão			
a	b	c	<-- classified as
330	11	2	a = 1
46	61	31	b = 2
3	7	43	c = 3

### b) Árvores de Decisão

- Resultados

**Quadro 33: Modelo de Árvore de Decisão selecionado para o Experimento C8**

Modelo 2		
SimpleCart		
SVD com 20 fatores		
Instâncias Classificadas Corretamente	444	83.1461 %
Instâncias Classificadas Incorretamente	90	16.8539 %
Número Total de Instâncias	534	

Matriz de Confusão			
a	b	c	<-- classified as
328	14	1	a = 1
42	80	16	b = 2
2	15	36	c = 3

### 6.5 AVALIAÇÃO CRÍTICA

Nesta seção, os melhores resultados apresentados serão discutidos e comparados.

O Experimento REG1 obteve como melhor resultado um modelo de rede neural composto por 1 camada escondida de 27 nós, taxa de aprendizado de 0,1 e conjunto de validação sendo a metade do conjunto de treinamento. O gráfico que compara a série observada e a série prevista evidencia que o modelo consegue prever valores de intensidade cuja série prevista acompanha de forma bem próxima à série observada.

Em uma análise isolada na qual apenas essa visualização e avaliação da taxa de erro são consideradas, o modelo é considerado um bom preditor. Porém, quando esse resultado é comparado à série gerada por um modelo que simplesmente repete o valor do dia anterior, com erros de 0,10258 e 0,10423, respectivamente; verifica-se que o ganho da técnica proposta no trabalho não é tão significativo. O custo de usar um modelo que repete o valor do dia anterior é menor que o custo de utilizar a técnica de mineração de dados.

No Experimento REG2, onde houve um aumento no tamanho do conjunto de treinamento, verificou-se uma pequena redução na taxa de erro da série gerada pelo modelo proposto. Ainda assim, a diferença manteve-se pequena.

Os testes realizados com os primeiros experimentos mostraram como melhores resultados os modelos construídos com conjuntos de dados nos quais foram adicionados os 5 últimos dias do ponto analisado. Isso levou a conclusão de que os valores medidos menos recentes não contribuíam tão significativamente para o resultado quando o tamanho da série histórica aumentava para 10. Por isso, os experimentos seguintes adotaram unicamente a série histórica com os 5 últimos dias

Uma melhora no ganho da técnica proposta neste trabalho foi observada no

Experimento REG3, onde uma nova configuração da grade foi definida. O erro da série prevista caiu para 0,09760, contra 0,10445, erro da série do dia anterior. Já o Experimento REG 4 apresentou uma piora no resultado com um aumento no erro para 0,1058; superando o erro da linha base.

Os experimentos realizados para modelos de classificação na escala 20-20 também revelaram que a técnica configura-se com uma opção vantajosa, ao derivar um bom modelo de classificação.

Como pôde ser visto, o Experimento C1, caracterizado por uma grade com os pontos mais próximos, sem SVD e os 5 últimos dias do ponto; resultou em modelos nos quais as maiores taxas de acerto ficaram entre 50% e 54%. Um olhar focado apenas na medida de erro induz uma avaliação pessimista, dado que, de forma resumida, fala-se em um modelo que acerta metade da amostra. Todavia, quando se faz uma análise além da medida de erro, olhando para a matriz de confusão e considerando o contexto do problema, é possível concluir que existe uma eficácia na utilização do modelo. A matriz de confusão referente ao modelo 5, mostra uma concentração dos valores na diagonal da matriz. Os erros para cada uma das classes são listados abaixo.

**Tabela 7: Erros calculados por classe da matriz de confusão do modelo 5 – C1**

<b>Classe</b>	<b>Taxa de Acerto</b>
20	0,8461
40	0,3451
60	0,2758
80	0,6444
100	0,3793
120	0,6666

Apesar da baixa taxa de acerto do modelo, em relação às classes 40 e 60, as previsões concentraram-se na faixa considerada segura para a operação. O modelo obteve uma melhora na taxa de acerto para a classe 80 e, mesmo com os resultados incorretos, o modelo concentra a classificação na faixa segura e, em alguns casos tendendo para atenção. De qualquer forma, a probabilidade da classe 80 ser classificada como 120 (situação em que não é recomendável operar) é muito pequena. Caso essa probabilidade fosse alta, esse resultado seria considerado insatisfatório, uma vez que o falso positivo também gera prejuízo pois significa não produzir.

O Experimento C2 difere-se do anterior pelo fato de que houve a aplicação de SVD com muitos fatores utilizados para formar o conjunto de treinamento. No entanto, os resultados revelaram taxas de acerto muito pequenas, sendo 25.3521% a maior taxa. As matrizes de confusão reforçaram a baixa eficácia dos modelos, evidenciando uma desordem nas células com valores não nulos e uma concentração na primeira coluna (referente à classe 20), onde o esperado é uma concentração na diagonal.

No Experimento C3, onde foi utilizada uma grade maior de 4x4, obteve-se modelos com taxa de acerto nas proximidades dos 48%, mas observa-se uma desordem em relação às células não nulas das matrizes de confusão.

O quadro abaixo apresenta o resultado de um modelo de classificação que repete a classe do dia anterior em comparação com o resultado do Modelo 5 do Experimento C1.

**Quadro 34: Comparação do Modelo da Linha base para classes 20-20 com o Modelo 5 – C1**

<b>Modelo Linha base</b>		
Instâncias Classificadas Corretamente	196	0.55211 %
Instâncias Classificadas Incorretamente	159	0.44788 %
Número Total de Instâncias	355	
<b>Matriz de Confusão</b>		
a	b	c d e f <-- classified as
52	26	1 0 0 0   a = 20
24	61	25 2 1 0   b = 40
3	25	44 12 3 0   c = 60
0	1	13 23 7 1   d = 80
0	0	4 7 16 2   e = 100
0	0	0 1 2 0   f = 120
<b>Modelo 5 – Experimento C1</b>		
Instâncias Classificadas Corretamente	171	48.169 %
Instâncias Classificadas Incorretamente	184	51.831 %
Número Total de Instâncias	355	
<b>Matriz de Confusão</b>		
a	b	c d e f <-- classified as
66	11	1 0 0 0   a = 20
43	39	24 6 1 0   b = 40
8	24	24 26 5 0   c = 60
0	1	5 29 8 2   d = 80
0	0	1 9 11 8   e = 100
0	0	0 0 1 2   f = 120

Mesmo com 48,169% de taxa de acerto, o modelo do trabalho apresenta uma melhor concentração da diagonal da matriz e consegue melhor classificar situações extremas, acertando mais onde a classe é 120. Verifica-se que o modelo da linha base errou para os 3 casos onde a classe era 120, e uma das amostras foi classificada como

80, situação em que é possível operar. Esta classificação pode gerar prejuízos e danos, uma vez que se decide por operar quando, na realidade, as condições são extremas.

A terceira abordagem do trabalho também apresentou resultados favoráveis a utilização da técnica com modelos de classificação de maior acurácia. Os experimentos C4 e C5, caracterizados pela grade com os pontos mais próximos, apresentaram resultados parecidos.

As taxas de acerto dos modelos selecionados em ambos os casos variaram entre 80% e 82%. Mesmo com essa proximidade, os modelos resultantes do experimento C4 são considerados melhores quando se avalia a matriz de confusão sob o ponto de vista da acurácia do modelo em relação à classe 3 (não operável). No experimento C5, os modelos classificaram pelo menos 2 casos na classe 1 (seguro), quando a classificação correta é a 3 (não operável). Apesar de ser um número muito pequeno em relação ao total de amostras da classe 3 (32 amostras), o objetivo é zerar a quantidade de classificações desse tipo. Conforme já falado, classificações como estas colocam em risco a operação.

Concluindo a análise da terceira abordagem, o Experimento C6 também apresenta bons resultados, principalmente em relação à taxa de acerto para a classe 3. O modelo 2 apresentou taxa de acerto total de 79,7183%, sendo 87,06% para a classe 1, 58,24% para a classe 2 e 87,50% para a classe 3. Neste modelo não houve caso onde uma amostra da classe 3 (não operável) foi classificada como sendo da classe 1 (seguro).

Comparando os modelos selecionados com a linha base, observa-se que a eficácia da técnica supera a eficácia da classificação via repetição do dia anterior. A taxa de acerto do modelo 2 (Experimento C6) para a classe 1 é de 87,50%, enquanto que a taxa da linha base é de 62,50%. Além disso, o modelo conseguiu zerar os casos onde a classe é 3 e foi classificada como 1.

**Quadro 35: Comparação do Modelo da Linha Base para classes por nível de segurança com o Modelo 1 – C4 e Modelo 2 – C6**

Modelo Linha Base		
Instâncias Classificadas Corretamente	286	80,56338 %
Instâncias Classificadas Incorretamente	69	19,43662 %
Número Total de Instâncias	355	

Matriz de Confusão			
a	b	c	<-- classified as
209	22	1	a = 1
23	57	11	b = 2
1	11	20	c = 3
Modelo 1 - C4			
Instâncias Classificadas Corretamente		292	82.2535 %
Instâncias Classificadas Incorretamente		63	17.7465 %
Número Total de Instâncias		355	
Matriz de Confusão			
a	b	c	<-- classified as
210	20	2	a = 1
16	60	15	b = 2
0	10	22	c = 3
Modelo 2 - C6			
Instâncias Classificadas Corretamente		283	79.7183 %
Instâncias Classificadas Incorretamente		72	20.2817 %
Número Total de Instâncias		355	
Matriz de Confusão			
a	b	c	<-- classified as
202	24	6	a = 1
12	53	26	b = 2
0	4	28	c = 3

Cabe ressaltar que além da boa acurácia do modelo preditor, seu desempenho demonstra uma característica altamente vantajosa em relação aos modelos tradicionais de previsão de intensidade correntes oceânicas, que necessitam de um tempo muito maior (dias) para completar o processamento. Especialistas da área destacam este fator como um dos principais ganhos em se utilizar a técnica de mineração de dados, que apresentou tempos de processamento na média de 1 minuto.

Os trabalhos relacionados que tratam da predição de velocidade de correntes utilizam parâmetros como, por exemplo, ventos, ondas, pressão atmosférica, marés, etc, para prever o valor da velocidade da corrente em um ponto [lat, lon]. Esta metodologia confronta com o problema da ausência dados reais medidos para os parâmetros citados não só na Costa Brasileira, mas também em outras regiões oceânicas. A metodologia proposta neste trabalho supera o problema, pois permite prever o valor e a faixa da velocidade de correntes utilizando apenas um parâmetro, que é a própria velocidade.

## Capítulo 7: Conclusão

As atividades petrolíferas envolvem, além de um capital alto, diversos impactos ambientais. Pequenas falhas podem causar grandes vazamentos com consequências graves ao meio ambiente e prejuízos financeiros milionários. Períodos de inatividade em função de condições climáticas e oceânicas adversas também provocam prejuízos. Desta forma, torna-se essencial um planejamento cuidadoso das operações.

Na indústria de petróleo, carregamentos ambientais como ondas, ventos e correntezas interferem nas operações offshore. As operações envolvem elementos estruturais com grande sensibilidade a forças provocadas por esses carregamentos, pois causam movimentação dinâmica nas estruturas, interferindo nas operações e podendo resultar em danos por excesso de carga ou falhas por fadiga. Por isso, no planejamento de atividades offshore, fenômenos como ondas, ventos, correntezas, dentre outros, devem ser considerados.

Danos em estruturas ou falhas causadas por condições ambientais impróprias para a operação paralisam as atividades. Além disso, a ausência por informações confiáveis e a incerteza também são fatores que contribuem para a decisão mais conservadora de não operar. Informações sobre condições ambientais futuras direcionam para um nível adequado de conservadorismo do projeto, ajudam a decidir de maneira mais eficaz quando a operação deverá ser interrompida e a otimizar as janelas de inatividade.

Nesse contexto, a observação das condições meteo-oceanográficas e a capacidade de prever condições climáticas futuras são requisitos altamente relevantes, que direcionam o planejamento, projeto e operação em estruturas oceânicas. Métodos tradicionais de previsão de velocidade de correntes baseiam-se no fenômeno físico, incluindo fórmulas matemáticas complexas com um conjunto de parâmetros cujos dados reais não estão acessíveis ou disponíveis em qualquer região.

Nesta dissertação foi proposta uma metodologia baseada em Mineração de Dados para a geração de um preditor de intensidade de correntes oceânicas. A intenção do trabalho foi fornecer previsões de intensidade de correntes oceânicas para fins de antecipação de condições extremas que inviabilizem a operação segura. Desta forma, foi apresentado um modelo e a metodologia utilizada para gerar modelos similares com a

função de prever valores de intensidade e condições futuras para um ponto específico, com base em informações históricas e atuais do próprio ponto e sua vizinhança.

Foram propostos três modelos: a) preditor de valores de intensidade; b) classificador com escala de intensidade de 20 cm/s; e c) classificador com escalas baseadas em níveis de segurança. A estrutura de entrada dos conjuntos de dados para os três modelos é similar. A diferença está na saída onde o modelo preditor possui como rótulo um valor contínuo, que é valor da velocidade da corrente. No primeiro classificador, o valor contínuo foi discretizado em faixas com intervalos de 20 cm/s. O segundo classificador trabalha com 3 classes que correspondem aos níveis de segurança operacional: Seguro, Atenção e Não Operável.

Os resultados mostraram que a série prevista com o preditor acompanha a série observada com um erro médio absoluto de 0,09 m/s e raiz do erro médio quadrático de 0,12 m/s. O erro médio ficou em torno de 20% do valor absoluto das velocidades previstas para 2 dias. De acordo com especialistas, este resultado é considerado muito bom para pontos offshore.

Cabe destacar que, mais importante que o valor absoluto previsto pelo modelo de regressão, a previsão é considerada boa em todo o período analisado e o modelo sempre foi capaz de prever quando a intensidade ia aumentar ou diminuir; ou seja, cumpriu com um dos requisitos de grande utilidade para o planejamento de operações.

A mesma avaliação vale para os resultados dos modelos de classificação por níveis de segurança, que em média previu corretamente 80% dos casos. Estes resultados comprovam que a metodologia pode ser usada como uma ferramenta confiável para a previsão de intensidades de correntes.

As contribuições deste trabalho são relevantes para o ramo petrolífero pois: fornecem informações estratégicas para engenheiros e operadores no suporte a uma gama de serviços operacionais e processos de tomada de decisão, apoiando no planejamento e tomada de decisão; antecipam situações críticas que ponham em risco as estruturas e a própria operação; evitam prejuízos milionários e ocorrências de danos ambientais. Embora a maior motivação do trabalho estivesse na área de aplicação compreendida pelo setor petrolífero, vale lembrar que os requisitos por informações de correntes oceânicas estão presentes em outros ramos como indústria naval, pesca, serviços de busca e resgate.

Por fim, é possível visualizar algumas oportunidades para aperfeiçoar os resultados e aprofundar a pesquisa realizada, inclusive expandir a aplicação para outro domínio. Portanto, alguns pontos são propostos a seguir trabalhos futuros:

- ✓ Aplicar a metodologia para prever a direção da corrente: Conforme mencionado anteriormente, a direção de correntes também é uma informações importante e deve ser considerada. O movimento dos gelos à deriva, por exemplo, pode impactar nas operações. Em situações de vazamento, saber a direção da correntes é extremamente útil para definir as ações a serem tomadas.
- ✓ Utilizar o mesmo raciocínio para verificar se é possível gerar um modelo capaz de prever a velocidade de correntes em diferentes profundidades.
- ✓ Aumentar sucessivamente o intervalo da previsão e verificar o aumento do erro. O objetivo é alcançar um período máximo no qual o modelo é capaz de prever valores e condições futuras com uma acurácia aceitável perante o domínio do problema.
- ✓ Aumentar o raio de distância da vizinhança: Neste caso, pretende-se analisar até que raio de distância é possível utilizar as informações de correntes medidas na vizinhança para prever as condições futuras em um ponto específico.
- ✓ Aplicar a metodologia em outros pontos do espaço fora da Bacia de Campos para gerar novos modelos que sejam compatíveis com outros perfis de corrente.
- ✓ Aplicar o método em uma região caracterizada por correntes de intensidades mais altas (fora da costa brasileira) e alterar as faixas do classificador por nível de segurança, de modo que a classe “Não Operável” seja definida realmente por correntes acima de 1 m/s.
- ✓ Estender a pesquisa para o domínio de ondas: não menos importante, as ondas oceânicas, quando atingem níveis extremos, também impactam rigorosamente nas instalações e operações offshore. A análise de extremos no domínio de ondas também é investigada pela literatura e a antecipação de condições críticas é fundamental. A previsão de informações de ondas também envolve um conjunto de parâmetros, de modo que a ausência de

dados reais também é uma realidade neste domínio. A aplicação de uma metodologia com os mesmos fundamentos da apresentada neste trabalho pode trazer resultados satisfatórios e representar uma opção promissora para a previsão de condições futuras de ondas.

## REFERÊNCIAS

ALTUNKAYNAK, Abdüsselam. **Adaptive estimation of wave parameters by Geno-Kalman filtering**. Ocean Engineering. Elsevier Science B.V., Amsterdam, v. 35, p. 1245 - 1251. 2008.

American Bureau of Shipping. **Subsea Riser Systems: Guide for Building and Classing**. American Bureau of Shipping (ABS), 2006.

AMERICAN PETROLEUM INSTITUTE. **2000 Joint Association Survey on Drilling Costs**. American Petroleum Institute, Washington, DC., 2001.

ARAÚJO, Bruno Allison; SILVA, Antonio Almeida; FARIAS, Marconi Andrade. **Análise de Resultados Experimentais de Ensaios em Corpos de Prova de Fadiga em Materiais para Dutos API 5L X60**. In: PDPETRO, 4., Campinas, SP. 2007.

AUGUSTO, Oscar Brito; ANDRADE, Bernardo Luís Rodrigues; ROSSI, Ronaldo Rosa. **Otimização de recursos para a operação de instalação de âncoras de equipamentos offshore**. Pesqui. Oper., Rio de Janeiro , v. 22, n. 3, jul. 2002 . Disponível em <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0101-74382002000300002&lng=pt&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0101-74382002000300002&lng=pt&nrm=iso)>. acessos em 27 out. 2013. <http://dx.doi.org/10.1590/S0101-74382002000300002>.

AYDOĞNA, Burak; AYAT, Berna; ÖZTÜRK, Mehmet N.; ÇEVİK, Esin Özkan; YÜKSEL, Yalçın. **Current velocity forecasting in straits with artificial neural networks, a case study: Strait of Istanbul**. Ocean Engineering, v. 37, p. 443–453, 2010.

BAKER, KIRK. **Singular Value Decomposition Tutorial**. Disponível em <[http://www.ling.ohio-state.edu/~kbaker/pubs/Singular\\_Value\\_Decomposition\\_Tutorial.pdf](http://www.ling.ohio-state.edu/~kbaker/pubs/Singular_Value_Decomposition_Tutorial.pdf) >. acessos em 3 jul. 2013.

BARBARIOLI , Angelo Barcelos; RIBEIRO, César Henrique de Assis; COUTO, Luiz Silva do. **Projeto OCEANOP: Informações Meteorológicas Para Região Oceânica Do Brasil**. Disponível em: < <http://www.cbmet.com/cbm-files/14-c13ababdaa144454ee48de978830bd36.pdf> >. Acesso em: jun/2011.

BARBOSA, Rômulo Lima. **Análise de Forças de Contato e Desgaste Interno em**

**Riser Marinho Devido ao Contato com Tubos de Perfuração.** 2007. 123 f. Dissertação. COPPE, UFRJ, Rio de Janeiro, 2007.

BASTOS, Glaudson M. **Algumas Aplicações Práticas da Tecnologia Data Mining.** Disponível em: <  
[http://www.sebraepb.com.br:8080/bte/download/Tecnologia%20da%20Informa%E7%E3o/162\\_1\\_arquivo\\_datamining.pdf](http://www.sebraepb.com.br:8080/bte/download/Tecnologia%20da%20Informa%E7%E3o/162_1_arquivo_datamining.pdf) >. Acesso em: jun/2011.

BREIMAN, L.; FRIEDMAN, J.; OLSHEN, R.; STONE, C.. **Classification and Regression Trees.** Wadsworth Int. Group, 1984.

CAETANO NETO, Ernesto S.; INNOCENTIN, Valdir, ROCHA, Rosmeri Porfírio da. **Um sistema de previsão de tempo e de ondas oceânicas para o Atlântico Sul.** Rcv. bras.occanogr., 44(1), 1996.

CAETANO NETO, Ernesto S.; INNOCENTIN, Valdir, ROCHA, Rosmeri Porfírio da. **Um sistema de previsão de tempo e de ondas oceânicas para o Atlântico Sul.** Revista Brasileira de Oceanografia, v. 44, n. 1, São Paulo, 1996.

CHAKRABARTI, Subrata K. **Handbook of Offshore Engineering.** Elsevier, 2005. v.2.

CROUT, Richard; CONLEE, Don; GILHOUSEN, Dave; BOUCHARD, Richard; GARCIA, Mike; DEMARCO, Frank; LIVINGSTON, Mark; COOPER, Cort; RAYE, Robert. **Real-time oil platform ocean current data in the Gulf of Mexico: An IOOS industry partnership success story.** In: Symposium on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface (IOAS-AOLS), 10., 2006.

CYBENKO, G. **Approximation by superpositions of a sigmoidal function.** Mathematics of Control, Signals, and Systems, v. 2, p. 303-314, 1989.

DET NORSKE VERITAS. **Rules For Offshore Installations.** Det Norske Veritas, Norway, jul. 1995.

Ewing, J. A. **Wind, Wave and Current Data for the Design of Ships and Offshore Structures.** Marine Structures, v. 3, p. 421-459, 1990.

FARRANT, Tim; JAVED, Khalid. **Minimising the effect of deepwater currents on drilling riser operations.** In: Offshore Engineering Limited Woking, 2., 2001, Surrey, UK. 2001.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. **Knowledge Discovery and Data Mining: Towards a Unifying Framework**. In: International Conference on Knowledge Discovery and Data Mining (KDD-96), 2., 1996, Portland, Oregon, 1996.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. **From data mining to knowledge discovery: an overview**. AI Magazine, v. 17, 1996.

FERNÁNDEZ, Eloi Fernández y; MUSSO, Bruno. **Oportunidades e Desafios da Agenda de Competitividade para Construção de uma Política Industrial na Área de Petróleo: Propostas para um Novo Ciclo de Desenvolvimento Industrial**. In: Fórum Nacional, 23., 2011, Rio de Janeiro, 2011.

FISCHER, J; FLEMMING, N. C. **Operational Oceanography: Data Requirements Survey**. EuroGOOS, Southampton Oceanography Centre, Southampton. n.12, 1999.

FRIEDMAN, Jerome H. **Data Mining and Statistics: What's the Connection?**. In: Symposium on the Interface - Computer Science and Statistics. 29. 1997.

GILBERT, Trevor. **Maritime Response Operations - Requirements for Metocean Data and Services**. In: Conference and Workshop on Meteorological and Oceanographic Services for Marine Pollution Emergency Response Operations. July 13-17, 1998 Townsville, Australia, 1998.

GODOI, Sueli Susana de. **Dinâmica Quase-Geostrófica do Sistema Corrente do Brasil no Embaimento de São Paulo (23,5°- 27° S)**. 2005. 157 f. Dissertação. USP, São Paulo, 2005.

GREEN, J.A. Mattias; STIGEBRANDT, Anders. **Statistical models and distributions of current velocities with application to the prediction of extreme events**. Estuarine, Coastal and Shelf Science, v. 58, p. 601–609, 2003.

HAN, Jiawei; KAMBER, Micheline; PEI, Jian. **Data Mining: Concepts and Techniques**. 2 ed. Elsevier, 2006.

HARIHARAN, Madhu; CERKOVNIK, Mark E.; THOMPSON, Hugh M.. **The Significance of Low Velocity Near Bottom Currents on the In-Line Vortex-Induced Vibration Response of Rigid Subsea Jumpers**. 2H Offshore, set/2004.

HAZA, Angelique; GRIFFA, Annalisa; MARIANO, Arthur; MOLCARD, Anne;

ÖZGÖKMEN, Tamay; PITERBARG, Leonid; POJE, Andrew; POULAIN, Pierre; RICSSEN, Michel. **Particle Prediction**. Disponível em: <<http://www.rsmas.miami.edu/LAPCOD/research/prediction/>>. Acesso em: 27 out. 2013.

HSIAO, Chih-Tsung; HWANG, Ching-Her. **Study on The Current Velocity Prediction By Artificial Neural Network at The Entrance of Hualien Port of Taiwan**. In: International Conference on Natural Computation (ICNC), 6., 2010.

HYCOM. **About Us**. Disponível em: <<http://hycom.org/about>>. Acesso em: 27 out. 2013.

IQDOUR, Radouane; ZEROUAL, Abdelouhab. **The MLP Neural Networks for Predicting Wind Speed**. In: Second International Symposium on Communications, Control and Signal Processing, Marrakech, Morocco, 2006.

JENKINS, P.B.; CROCKFORD, A.L. **Drilling Costs**. Society of Petroleum Engineers. SPE-European Meeting, London, 1975.

SOUSA E SILVA, José Alisson; SILVA, Antônio Almeida; FARIAS, Marconi de Andrade. **Desenvolvimento e construção de bancada de teste para validação de modelos de danos sob corrosão-fadiga em dutos e estruturas “Offshore”**. In: Congresso Brasileiro de P&D em Petróleo e Gás, 3., 2005, Salvador, Bahia, 2005.

JUNG, G.,B., TOLDO JR., E.E. **Análise das forçantes da direção superficial da corrente longitudinal na praia de Tramandaí, RS**. In: Simpósio sobre Ondas, Marés, Engenharia Oceânica e Oceanografia por Satélite (OMAR-SAT), 8., 2009, Arraial do Cabo, 2009.

KAISER, Mark J.; PULSIPHER, Allan G.. **The impact of weather and ocean forecasting on hydrocarbon production and pollution management in the Gulf of Mexico**. Energy Policy, 35, 966-983, 2007.

KANG, Moon-Kyung; LEE, Hoonyol; YANG, Chan-Su; YOON, Wang-Jung. **Estimation of Ocean Current Velocity in Coastal Area Using Radarsat-1 SAR Images and HF-radar Data**. In: IEEE International Geoscience and Remote Sensing Symposium (IGARSS), v. 1, Boston, 2008, 413-416.

KITE-POWELL, H.L. **Economic considerations in the design of ocean observing systems**. Oceanography, v. 22, n. 2, p. 44-49, 2009.

LE CUNFF, C.; BIOLLEY, F.; FONTAINE, E. ; ÉTIENNE, S.; FACCHINETTI,

M.L..**Vortex-Induced Vibrations of Risers: Theoretical, Numerical and Experimental Investigation**. Oil & Gas Science and Technology, v. 57, n. 1, p. 59-69, 2002.

LIANG, W.-D.; TANG, T.Y.; YANG, Y.J.; KO, M.T.; CHUANG, W.-S. **Upper-ocean currents around Taiwan**. Deep-Sea Research II. v. 50, p. 1085-1105, 2003.

LIMEBURNER, R.; BEARDSLEY, R. **Ross Ice Shelf 2010 M2 Moored Velocity Profiles**. 2011.

MACLACHLAN, Malcolm. **An Introduction to Marine Drilling**. 1. ed. Oilfield Publications Limited, 1999.

MAIMON, Oded; ROKACH, Lior. **Data Mining and Knowledge Discovery Handbook**. 2. ed. Springer. 2010.

MARINHA. **Modelo HYCOM**. Disponível em: <<http://www.mar.mil.br/dhn/chm/meteo/prev/modelos/hycom.htm>>. Acesso em: 27 out. 2013.

MIGUENS, Altineu Pires. **Navegação: A Ciência e a Arte: Navegação costeira, estimada e em águas restritas**. *Marés e correntes de maré; correntes oceânicas*. Marinha do Brasil, v.1, 1993. p. 227-274.

MOOERS, Christopher N. K.. **Coastal Ocean Prediction**. v. 56, 1999.

MY OCEAN. **The Project**. Disponível em: <<http://www.myocean.eu/web/2-the-project.php>>. Acesso em: 27 out. 2013.

NOAA Ocean Explorer. **Multimedia Discovery Missions: Lesson 8 - Ocean Currents**. National Oceanic and Atmospheric Administration. Disponível em: <<http://oceanexplorer.noaa.gov/edu/learning/>>. Acesso em: 11 jul. 2012.

PETROBRAS. **Bacia de Campos comemora 35 anos pronta para os desafios do futuro**. Disponível em: <<http://fatosedados.blogspotrobras.com.br/2012/08/15/bacia-de-campos-comemora-35-anos-pronta-para-os-desafios-do-futuro/>>. Acesso em: 3 jun. 2013.

PICCININI, Flávio Costa. **Carregamento Hidrodinâmico sobre uma Estrutura Cilíndrica Esbelta Proveniente de Ondas e Correntes Oceânicas**. Revista da Engenharia de Instalações no Mar, Macaé, n. 1, jan./jul. 2008.

OFFSHORE ENGINEER. Offshore Engineer: Technological turning points. Abr. 2002.

PICCININI, Flávio Costa. **Carregamento Hidrodinâmico sobre uma Estrutura Cilíndrica Esbelta Proveniente de Ondas e Correntes Oceânicas**. Revista da Engenharia de Instalações no Mar. n. 1, Jan./Jul. 2008.

PIRES, Tiago Ancelmo de Carvalho; CAVALCANTI, Dayse Duarte; SILVA, José Jéferson do Rêgo; SENA, Luiz Augusto Cabral Vitória. **Um método de análise simplificada do risco de falha em estruturas metálicas devido ao impacto térmico de um jet fire proveniente de blowout em poços de petróleo**. In: SIMPEP, 13., 2006, Bauru, São Paulo. 2006.

PRESTA, Roberto; SOUSA, Fernando Jorge Mendes de; ELLWANGER, Gilberto Bruno; VASCONCELOS, José Marcio do Amaral; SILVA NETO, Severino Fonseca da. **Análise de fadiga em estruturas de Heliponto**. In: Congresso Nacional de Transporte Aquaviário, Construção Naval e Offshore, 22., SOBENA, 2008.

PYLE, Dorian. **Business Modeling and Data Mining**. 1. ed. Elsevier. 2003.

QUINLAN, J. R.. **C4.5: Programs for Machine Learning**. Morgan Kaufmann, Los Altos, 1993.

QUINLAN, J.R.. **Induction of decision trees**. Machine Learning 1, 81-106, 1986.

SHANKS, F.E. **High-Current, Deepwater Drilling Experiences**. Society of Petroleum Engineers. Drilling Conference, New Orleans, Louisiana, 1985.

R PROJECT. **What is R?**. Disponível em: <<http://www.r-project.org/>>. Acesso em: 6 jun. 2013.

SHANKS, F.E.; HAMMETT, D.S.; ZINKGRAF, H.L. **Experience drilling in deepwater high current areas**. In: Offshore Technology Conference, Houston, Texas, 1979.

SHILLING, R.; BRENDLING, B. **First-Year Simulation of Drilling and Completion Operations on the Atlantis Project**. In: Offshore Technology Conference (OTC 18122), 2006, Houston, Texas, 2006.

SILVEIRA, Ilson Carlos Almeida da. **O Sistema Corrente do Brasil na Bacia de Campos, RJ**. 2007. 181 f. Tese. USP, São Paulo, 2007.

SMIRREN, J.R. van; ROMEO, John; VOGEL, Michael. **Developments in Acoustic**

**Doppler Current Profiling and its Technical Application in Deepwater Drilling.** Offshore Technology Conference, Houston, 1999.

**SOS-OC. Ocean Currents from space: User requirements for Ocean Currents – WP2 report.** Satellite Observing Systems Ltd (SOS-OC). 2002.

SPRING, Walt; GU, George; GRAY, Mike; MIDDLEHURST, Bill; STRITTO, Frank J. Dello. **Deepwater Ocean Currents West of Britain: Measurements & Criteria Development.** Offshore Technology Conference. Houston, 1999.

THOMAS, J. E. **Fundamentos de Engenharia do Petróleo.** 1. ed. Rio de Janeiro: Editora Interciência, 2001.

TSUKADA, Raphael Issamu; YAMAMOTO, Marcio; MOROOKA, Celso Kazuyuki. **Análise de Operação da Instalação de Risers e Descida de BOP em Sondas Flutuantes de Perfuração.** In: PDPETRO, 4., 2007. Campinas, 2007.

TURBAN, Efraim; ARONSON, Jay E.; LIANG, Ting-Peng. **Decision support systems and intelligent systems.** 7. ed. Prentice Hall, 2005.

UNIDATA. **NetCDF FAQ.** Disponível em: <<http://www.unidata.ucar.edu/software/netcdf/docs/faq.html#usage>>. Acesso em: 27 out. 2013.

UNIVERSITY OF WAIKATO. **Map of ocean gyres.** The Ocean in Action. University of Waikato. Disponível em: <<http://www.sciencelearn.org.nz/Contexts/The-Ocean-in-Action/Sci-Media/Images/Map-of-ocean-gyres/>>. Acesso em: 11 jul. 2012.

VALADÃO, Ana Maria Fernandes. **Análise de Fadiga de estruturas Offshore tipo topside - Estudo de Caso.** 2007. 85 f. Monografia. UFRJ, Rio de Janeiro, 2011.

WAIKATO. **Machine Learning Group.** Disponível em: <<http://www.cs.waikato.ac.nz/ml/index.html>>. Acesso em: 6 jun. 2013.

WOCE. **The World Ocean Circulation Experiment.** Disponível em: <<http://woce.nodc.noaa.gov/wdiu/>>. Acesso em: 27 out. 2013.