



ANÁLISE DA COMUNIDADE RECIFAL DA FOZ DO RIO DOCE APÓS O DESASTRE DE MARIANA

André Machado Rocha

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

Orientadores: Fabiano Lopes Thompson
Laura Silvia Bahiense da Silva
Leite

Rio de Janeiro
Março de 2018

ANÁLISE DA COMUNIDADE RECIFAL DA FOZ DO RIO DOCE APÓS O
DESASTRE DE MARIANA

André Machado Rocha

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
PRODUÇÃO.

Examinada por:

Prof. Fabiano Lopes Thompson, D.Sc.

Prof. Laura Silvia Bahiense da Silva Leite, D.Sc.

Prof. Marta Lima de Queirós Mattoso, D.Sc.

Prof. Carlos Eduardo de Rezende, D.Sc.

RIO DE JANEIRO, RJ – BRASIL
MARÇO DE 2018

Rocha, André Machado

Análise da Comunidade Recifal da Foz do Rio Doce após o Desastre de Mariana / André Machado Rocha. – Rio de Janeiro: UFRJ / COPPE, 2018.

XI, 87 p.: il.; 29,7cm.

Orientadores: Fabiano Lopes Thompson

Laura Silvia Bahiense da Silva Leite

Dissertação (mestrado) – UFRJ / COPPE / Programa de Engenharia de Produção, 2018.

Referências Bibliográficas: p. 80 – 87.

1. Comunidade recifal da foz do Rio Doce. 2. Desastre de Mariana. 3. Métodos de aprendizado estatístico. 4. Redes neurais. 5. Métodos baseados em árvores. 6. Métodos aditivos. I. Thompson, Fabiano Lopes *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Produção. III. Título.

Agradecimentos

Primeiramente agradeço a minha mãe Ana e ao meu pai Carlos pelos anos ininterruptos de apoio e carinho incondicionais.

Ao meu irmão Gabriel, ouvinte, parceiro e companheiro de longos anos.

Aos meus orientadores Fabiano Thompson e Laura Bahiense, pela atenção, paciência e auxílio nos momentos de necessidade.

Aos companheiros do SAGE pelo aprendizado constante ao longo desses dois anos.

Ao professor Rogério (*in memoriam*) por abrir as portas do SAGE, e ensinar o valor do pensamento crítico para a formação profissional e pessoal.

Ao meu avô João (*in memoriam*), exemplo de humildade e força de espírito.

Agradeço ao Programa de Engenharia de Produção da COPPE/UFRJ pelo conhecimento adquirido ao longo dos anos, suporte acadêmico e aprendizado pessoal.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo suporte financeiro que viabilizou esse trabalho.

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

ANÁLISE DA COMUNIDADE RECIFAL DA FOZ DO RIO DOCE APÓS O DESASTRE DE MARIANA

André Machado Rocha

Março/2018

Orientadores: Fabiano Lopes Thompson
Laura Silvia Bahiense da Silva Leite

Programa: Engenharia de Produção

O desabamento da barragem de Mariana jogou milhões de toneladas de lama tóxica no Rio Doce, e conseqüentemente no mar do Espírito Santo. Esse incidente já é considerado o maior desastre ecológico de que se tem notícia no Brasil.

Ao norte do deságüe do Rio Doce encontra-se o Banco de Abrolhos, um dos maiores santuários ecológicos do país. Inicialmente, se pensava que os rejeitos tóxicos tomariam o caminho do sul, em direção ao litoral carioca, porém novos indícios apontam que uma parte dos resíduos tomaram o caminho do norte, chegando ao sul de Abrolhos, onde se localizam recifes recém-descobertos.

Os sistema coralíneo da foz do Rio Doce apresenta diversas espécies de corais em risco de extinção, e apresentava sinais de vitalidade. Neste trabalho, com base nas amostras coletadas, foi realizada uma análise da comunidade recifal da foz do Rio Doce através da aplicação de modelos de aprendizado estatístico, que vêm sendo mais utilizados na Biologia conforme a demanda por informação aumenta, especialmente pela sua capacidade de identificar relações não-lineares.

Os resultados apontaram as variáveis que estariam correlacionadas à cobertura de corais nos recifes, em especial a correlação negativa da combinação dos colchões de algas epilíticas e cianobactérias com a cobertura coralínea. Também foram detectadas indícios de influências positivas de determinadas concentrações de esponjas e *CCA* para o desenvolvimento de espécies de corais.

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

ANALYSIS OF REEF COMMUNITY LOCATED NEAR RIO DOCE'S MOUTH
AFTER MARIANA'S DISASTER.

André Machado Rocha

March/2018

Advisors: Fabiano Lopes Thompson
Laura Silvia Bahiense da Silva Leite

Department: Production Engineering

The collapse of the Mariana dam has dumped millions of tons of toxic mud in the Rio Doce, and consequently to Espírito Santo's sea. This incident is already considered the biggest ecological disaster in Brazil.

On the north of the river mouth is localized the Abrolhos bank, one of the greatest ecological sanctuaries on the country. Initially was thought that the tailings would go southward, toward the coast of Rio de Janeiro state, but new evidences indicate that part of the waste reached Abrolhos bank, exactly where newly discovered reefs are located.

The coralline system near the mouth of the Rio Doce presents several species of corals in danger of extinction, and presents ecological index that could indicate signs of vitality. In this work, based on sampled data, an analysis of the reef community located near Rio Doce mouth was developed through models of statistical learning, which are being more used in biological studies as information disponibility increases, especially for capacity of identification of non-linear relations.

The results pointed out the relevance of other groups of species, especially the negative influence of the combination of epilithic algae and cyanobacteria mattresses for healthy coral cover. Also, evidences were found that for certain levels, the concentration of sponges and *CCA* could benefit coral species.

Sumário

Lista de Figuras	ix
Lista de Tabelas	xi
1 Introdução	1
1.1 Motivações	3
1.2 Objetivos	4
1.3 Limitações	5
1.4 Estrutura	5
2 Revisão Bibliográfica	6
2.1 Recifes de Corais, relevância e ameaças	6
2.2 Caracterização do Banco de Abrolhos	9
2.3 Modelos de Aprendizado de Máquina para Recifes	15
2.4 Recifes da Foz do Rio Doce e o desastre de Mariana	17
3 Método	21
3.1 Índices de Diversidade	21
3.2 Métodos de Ordenamento	22
3.2.1 Análise dos Componentes Principais	23
3.2.2 Ordenação Multidimensional não Métrica	25
3.3 Métodos de Aprendizado Estatístico	26
3.3.1 Regressão Linear e Regressão Aditiva	27
3.3.2 Redes Neurais	29
3.3.3 Modelos de Árvore	35
4 Experimentos Computacionais	40
4.1 Dados	40
4.2 Desastre de Mariana e os recifes de Conceição da Barra	43
4.2.1 Diversidade de espécies	50
4.2.2 Análises de Ordenação	52
4.3 Modelos de Previsão	56

4.3.1	Tratamento de Dados	56
4.3.2	Regressão Linear	59
4.3.3	Modelos Aditivos	60
4.3.4	Redes Neurais	61
4.3.5	Árvores de Crescimento Progressivo	62
4.3.6	Comparação de Resultados	64
5	Discussões	70
6	Conclusão	76
6.1	Proposta de trabalhos futuros	78
6.2	Considerações Finais	78
	Referências Bibliográficas	80

Lista de Figuras

2.1	Corais afetados pelo branqueamento.	9
2.2	Três megahabitats do banco de Abrolhos.	10
2.3	Foto de um Rodolito.	11
2.4	Corais localizados em zonas não protegidas.	12
2.5	Corais localizados em zonas marinhas protegidas.	12
2.6	Recifes localizados na região centro-norte do Banco de Abrolhos e suas respectivas áreas de proteção.	13
2.7	Comparação da análise de sensibilidade para quatro métodos que buscam prever a cobertura de corais no Alasca.	17
2.8	Mapa da região do banco de Abrolhos com destaque para os novos recifes descobertos, ao norte da foz do Rio Doce.	19
2.9	Caminho da lama do desastre de Marina ao chegar no mar da costa do Espírito Santo. Imagens de 05/01/2016 e 06/01/2016 respectivamente.	20
3.1	Exemplo de regressão cúbica com os <i>joelhos</i> em destaque.	28
3.2	Unidade Neural Básica.	31
3.3	Exemplo de um Redes Neural completa.	32
3.4	Árvore de Regressão, devido as muitas ramificações pode se tornar difícil de interpretar.	36
3.5	Comportamento do erro em relação ao número de árvores. Para o dados de treinamento há uma expectativa de decrescimento, já para os dados de validação, após alcançar um mínimo há um comportamento oscilante.	38
4.1	Programa <i>Coral Count Extention</i>	41
4.2	Imagens do desastre no Rio Doce.	43
4.3	Imagem MODIS 16 de junho de 2016.	44
4.4	Comparação da análise de peixes antes e depois do desastre de Mariana, por família e por espécie.	45
4.5	Corais da região da foz do Rio Doce.	46
4.6	Algumas espécies encontradas nos recifes da foz do Rio Doce.	46

4.7	Análise APC dos recifes da foz do Rio Doce, comparação entre recifes norte e sul e por tipo de habitat.	48
4.8	Abundância relativa das espécies nos corais da foz do Rio Doce.	49
4.9	Número de espécies identificadas nos recifes de Abrolhos.	50
4.10	Índice de Entropia de Shannon dos recifes de Abrolhos.	51
4.11	Índice de Equitatividade de Pielou dos recifes de Abrolhos.	52
4.12	nMDS dos habitats recifais de Abrolhos.	53
4.13	Espécies mais presentes em cada Recife.	55
4.14	Particionamento de Garson. Indica a importância relativa de cada variável explicativa no modelo.	62
4.15	Particionamento da importância de cada variável explicativa no modelo de árvores de crescimento progressivo.	63
4.16	Análise de sensibilidade das variáveis de cobertura no modelo 1, para cada um dos métodos utilizados.	68
4.17	Análise de sensibilidade das variáveis MODIS no modelo 1, para cada um dos métodos utilizados.	69
5.1	Análise de sensibilidade da cobertura de corais para Esponjas e CCA nos modelos de redes neurais e árvores de crescimento progressivo . . .	73
5.2	Variabilidade da análise de sensibilidade de Cobertura de Corais x Esponjas. Método de Redes Neurais.	74

Lista de Tabelas

3.1	Variáveis a serem definidas em um modelo de Redes Neurais.	33
3.2	Formulação da Matriz P	34
3.3	Matriz Q e soma das colunas.	35
4.1	Recifes do Rio Doce utilizados para coleta e informações sobre os locais.	41
4.2	Locais usados para comparação do bentos, com a política de preservação adotada, pontos de coleta e anos em que foram realizadas.	42
4.3	Resumo dos Dados coletados pelo MODIS.	42
4.4	Tipos de Habitat no Recifes Analisados.	47
4.5	Corais agrupados.	57
4.6	Médias e desvios padrões das variáveis utilizadas nas modelagem por Recife.	58
4.7	Variáveis explicativas para cada um dos modelos formulados tendo como saída a cobertura dos recifes de corais.	58
4.8	Modelo 1 em regressão linear- Resumo dos coeficientes para cada variável, e seu valor de significância.	59
4.9	Modelo 1 em métodos aditivos- Resumo dos coeficientes para cada variável, e seu valor de significância.	60
4.10	Variáveis definidas no modelo de Redes Neurais.	61
4.11	Variáveis definidas no modelo de Árvores de Crescimento Progressivo.	63
4.12	R^2 dos modelos utilizados. Resultado para a modelagem com todas as variáveis explicativas.	65
4.13	Modelo 2 - R^2 dos modelos utilizados. Resultado para a modelagem com todas as variáveis explicativas.	66

Capítulo 1

Introdução

Recifes de corais estão entre os ecossistemas mais importantes do planeta. Quase 25% da vida marinha conhecida se localiza perto de alguma formação recifal, enquanto estas ocupam menos de 1% do espaço dos oceanos. A produção primária de energia em um conjunto de recifes é comparada ao de uma floresta amazônica [65].

No âmbito socioeconômico, milhões de pessoas se beneficiam diretamente dos recifes, seja através da pesca, como fonte de alimentos, ou como atividade comercial e turismo [39]. Há também serviços indiretos prestados, como por exemplo, a proteção do litoral contra a força das ondas, que é dissipada pela barreira natural. As áreas ocupadas por recifes de corais podem ser consideradas valiosas em termos econômicos para nossas sociedades [40, 66].

Em todo o mundo, as ações humanas sobre o meio ambiente são sentidas, e com os recifes a situação é ainda mais complicada, já que compõem um dos ecossistemas mais frágeis apesar de toda sua riqueza. Um trabalho das Nações Unidas do ano de 2004 estimou que por volta de 10% de todos os recifes já estariam degradados além do ponto de recuperação, e pelo menos mais 30% estariam em processo de deterioração [50].

O possível colapso dos recifes de corais vem de uma combinação de fatores locais e globais. A nível mundial, as alterações climáticas resultam em acidificação do oceano, mudança de temperatura da água e espalhamento de doenças como o branqueamento. Ações de comunidades que vivem nas cercanias do oceano, como a sobre-exploração da pesca, práticas destrutivas de uso e poluição também interferem na saúde dos ecossistemas recifais [6, 72].

A degradação dos recifes interfere no custo de oportunidade para as comunidades locais. A sobre-exploração de sua capacidade e a falta de cuidados na preservação podem levar à deterioração da capacidade econômica da região, e também acarretar carências nutritivas. Por isso, é crucial o desenvolvimento de políticas públicas que promovam o uso sustentável desses ecossistemas [40].

O Brasil é um país reconhecido pela grande diversidade de ecossistemas em

seu território. A região do sul da Bahia apresenta um rico conjunto de habitats, garantindo uma elevada biodiversidade, em especial para o ambiente marinho, que conta com rios, mangues, praias, estuários, recifes de coral e ilhas oceânicas [48].

É nessa região que se localiza o banco de Abrolhos, uma plataforma continental que possui em sua máxima extensão até 200 km adentro do oceano Atlântico. Os recifes da região são considerados a zona mais rica em termos de corais de todo o Atlântico Sul. As formações coralíneas da região são considerados os mais preservados de todo o litoral brasileiro, apesar dos danos já causados [11].

Atualmente, a região conta com um mosaico, formado em parte por áreas que teoricamente seriam completamente protegidas, as chamadas reservas marinhas. Contudo, alguns espaços são destinados ao extrativismo sustentável por comunidades locais, e a maior parte do banco de Abrolhos não conta com nenhum tipo de fiscalização especial [12].

As formações recifais mais conhecidas de Abrolhos ficam na parte norte e central do banco. A região sul, por outro lado, nunca foi vista como uma prioridade para a pesquisa na região. Entretanto, com a descoberta de novas formações coralíneas no norte do Espírito Santo, essa situação vem mudando. Em 2014, uma expedição encontrou recifes de corais em uma área ainda compreendida pelo banco de Abrolhos, próximos à cidade de Conceição da Barra, que anteriormente se pensava ser formada basicamente por sedimentos não-consolidados [45].

Com a presença de grande número de espécies de peixes em risco de extinção, o Espírito Santo é tido como um dos estados federativos de maior relevância para a conservação da biodiversidade. Apesar da já destacada importância, a área ainda é pouco protegida por zonas marinhas [45].

Na região de Conceição da Barra, em particular, podia-se encontrar à época um número relativamente grande de espécies de peixes, algumas inclusive ameaçadas de extinção. Foram também encontrados recifes povoados por corais saudáveis, com poucas evidências de branqueamento. Também foi possível observar danos aparentemente causados pelas âncoras e redes de pescaria nos corais, denunciando a influência antropogênica [45].

O rompimento da barragem do fundão no Rio Doce já é considerado uma das maiores tragédias ambientais e humanas ocorridas no país. Dezesesseis dias após o derramamento de milhares de litros de rejeitos de minério, a lama tóxica atingiu o mar [46]. O deságue aconteceu a setenta e quatro quilômetros das novas formações recifais descritas por MAZZEI *et al.*[45], e havia indícios de que a lama poderia ter alcançado o ecossistema.

Novos dados foram coletados nos recifes de corais de Conceição da Barra em novembro de 2016, e posteriormente em 2017, respectivamente um e dois anos após o acidente. Com análises relativas ao número e tipos de espécies de peixes encontradas,

qualidade da água e sobre o bentos das formações, as expedições trouxeram indícios de que a lama poderia ter alcançado o ecossistema do sul do banco de Abrolhos.

O bentos pode ser entendido como a composição de espécies que habitam a superfície de um recife. Sendo um dos habitats mais ricos do oceano em termos de nutrientes, é de se esperar uma grande disputa pelo espaço disponível. Influências externas podem ter papel significativo para alteração do equilíbrio vigente no local.

Nesse trabalho, realizou-se a análise desses novos dados coletados para esse ambiente com o objetivo de posicionar a composição biológica dos recifes de Conceição da Barra frente a outros habitats recifais localizados no banco de Abrolhos, buscando possíveis explicações para as variações biológicas entre as regiões.

O desenvolvimento tecnológico torna cada vez maior a quantidade e o nível de informação disponível para as áreas biológicas, ao mesmo tempo que aumenta a necessidade da utilização de métodos quantitativos mais sofisticados para a interpretação dos fenômenos biológicos [29].

Além disso, na biologia, as maiorias das interações entre os organismos apresentam padrões não-lineares de comportamento. Contudo, a maior parte dos estudos utiliza métodos de regressão linear múltipla para analisar as reações de causa e efeito. Essa metodologia permite constatar a relevância de variáveis explicativas, porém analisa o comportamento da variável dependente de forma generalista, sendo difícil analisar as nuances dessas interações. Como consequência, vem crescendo a aplicação de metodologias de aprendizado de máquina para estudo dos recifes de corais. [14, 29, 55]

Este trabalho faz uma comparação entre os métodos de regressão tradicionais e os modelos de aprendizado de máquina para a análise da comunidade recifal da foz do Rio Doce após o desastre de Mariana.

1.1 Motivações

Os corais são, ao mesmo tempo, uma das formações mais sensíveis e mais antigas existentes no mundo. O entendimento do desenvolvimento dessas estruturas ao longo do tempo pode revelar informações sobre a própria história do planeta em que vivemos, e também nos ajudar a fazer previsões sobre o futuro.

O banco de Abrolhos é uma formação extremamente rica em termos de biodiversidade, e muito do que já foi descoberto (e do que ainda está por se descobrir) é (será) utilizado por campos de pesquisa aplicados que abrangem a biotecnologia e a medicina. Para que isso ocorra, é preciso compreender o processo de formação e analisar as mudanças que ocorrem nesses ecossistemas.

Os estudos das características biológicas dos corais recém-descobertos próximo ao município de Conceição da Barra reforçam a importância desse ecossistema, ga-

rantindo que sejam adotadas estratégias adequadas de conservação e exploração dessas áreas.

Dezesseis dias após o derramamento de milhares de litros de rejeitos de minério, causado pelo rompimento da barragem do fundão no Rio Doce, a lama tóxica atingiu o mar, desaguando a setenta e quatro quilômetros das novas formações recifais descritas por [45].

Os novos dados coletados nos recifes de corais de Conceição da Barra em novembro de 2016, quase um ano após o acidente, trouxeram indícios de que a lama poderia ter alcançado o ecossistema do sul do banco de Abrolhos.

Fazia-se necessária, portanto, a análise dos novos dados coletados para esse ambiente com o objetivo de posicionar a composição biológica dos recifes de Conceição da Barra frente a outros habitats recifais localizados no banco de Abrolhos, buscando possíveis explicações para as variações biológicas entre as regiões.

A utilização de métodos de aprendizado estatístico já foi utilizada antes para corais brasileiros [14]. Porém ainda não há sido feita uma comparação sobre as ferramentas e acurácia da capacidade de um ou mais métodos, assim como feito em outros estudos internacionais [29, 59].

1.2 Objetivos

O objetivo geral deste estudo é analisar a saúde do banco recifal localizado no norte do litoral do Espírito Santo através da comparação entre a composição biológica dos recifes de Conceição da Barra frente a outros habitats recifais localizados no banco de Abrolhos, buscando possíveis explicações para as variações biológicas entre as regiões.

Um objetivo secundário deste estudo é analisar a chegada e o efeito da lama tóxica oriunda do desastre de Mariana na comunidade de recifes próximos a foz do Rio Doce. Dado o grande valor deste ecossistema, tanto do ponto de vista ecológico quanto socioeconômico, os resultados obtidos poderão ser utilizados para auxiliar no desenvolvimento de possíveis políticas de preservação.

Um forte indicador sobre a saúde de um ecossistema de recifes de corais é a própria condição das espécies coralíneas. Por essa razão, outro objetivo secundário é a análise da cobertura dos recifes no banco de Abrolhos, através da análise das principais variáveis explicativas que se relacionam com a abundância dos corais nos recifes, e de que forma se dão as relações entre essas variáveis. Para isso serão utilizados métodos de aprendizado de iterativo.

Aproveitando a utilização de métodos de aprendizado estatístico sobre os dados coletados, e suas análises derivadas, realiza-se também uma comparação entre os métodos. Observando as vantagens e desvantagens para caso específico.

1.3 Limitações

Os dados de composição bentônica são obtidos através de imagens de alta resolução sobre as formações recifais. Devido aos custos logístico, técnico e financeiro de se obter esse tipo de informação, os dados utilizados nesse estudo foram obtidos em pesquisas anteriores, pelos pesquisadores responsáveis, devidamente referenciados nesse trabalho.

O litoral do banco de Abrolhos é extenso e conta com outras formações recifais não apresentadas neste estudo. A capacidade de prever a cobertura coralínea em locais com menor informação, a partir de informações coletadas e a utilizando ferramentas de aprendizado estatístico é uma realidade, e já foi utilizada em outros estudos [59], porém este assunto não será abordado.

Outra possível utilização para métodos de aprendizado estatístico é a previsão futura dos pontos observados [28]. Porém este estudo focará apenas na relevância e sensibilidade das variáveis explicativas, e na acurácia dos métodos, sem a tentativa de analisar cenários futuros.

Este estudo não utilizará métodos de aprendizado de máquina na previsão de espécies em particular. Este assunto já foi discutido na literatura [14, 29, 41], e oferece muitas possibilidades, mas não será o foco desta pesquisa.

1.4 Estrutura

Na primeira seção é feita a introdução do tema, assim como a apresentação dos objetivos e motivações. É apresentado a relevância dos estudos de corais no mundo, assim como a importância biológica do banco de Abrolhos.

Na segunda seção os principais trabalhos sobre recifes de corais, a região de Abrolhos e a utilização de métodos de aprendizado na biologia são apresentados. A revisão da literatura foi feita buscando compreender e relacionar as principais ameaças aos ecossistemas que envolvem corais, caracterizar o banco de Abrolhos e analisar a utilização de ferramentas matemáticas para compreender relações ecológicas.

Na terceira seção é apresentada a metodologia deste estudo. A partir do referencial teórico, os métodos mais convenientes para cumprir os objetivos propostos a partir das informações obtidas são abordados.

Na quarta seção os resultados do estudo são apresentados. É feita também uma breve discussão das informações obtidas.

A quinta seção é reservada para a discussão dos resultados apresentados. Possíveis hipóteses são levantadas e é feita a comparação com a bibliografia.

A sexta seção é dedicada a conclusão do trabalho. Resumo dos resultados e da discussão são apresentados, assim como sugestões para trabalhos futuros.

Capítulo 2

Revisão Bibliográfica

Neste capítulo são apresentadas as referências utilizadas para o restante do estudo. Os conhecimentos adquiridos da bibliografia foram divididos em quatro seções. A primeira parte trata da relevância dos corais, tanto em sua importância biológica quanto social. A interação do homem com os recifes é abordada nesta seção, que apresenta as principais ameaças que este ecossistema vem sofrendo pela ação humana.

A segunda seção é destinada a região onde se localizam os corais da foz do Rio Doce, o banco de Abrolhos. O local se destaca como uma das maiores fontes de biodiversidade do Brasil, com uma presença rica e vasta de diversos grupos de espécies que variam desde os micróbios aos grandes mamíferos, como as baleias. Os sistemas recifais representam as áreas marinhas mais importantes da região em termos de diversidade biológica. A saúde dos ecossistemas coralíneos é um patrimônio que deve ser visto com prioridade para a formulação de políticas públicas de preservação.

A terceira parte é destinada a revisão de estudos estatísticos utilizando variáveis ecológicas. Além de testes estatísticos cresce no estudos biológicos o interesse por métodos que consigam captar além do impacto das variáveis explicativas sobre uma variável dependente, também a sensibilidade dessa interação.

Na quarta parte do capítulo as informações já conhecidas sobre os corais da foz do Rio Doce são apresentadas, os dados são relativos as primeiras expedições ao local. O conhecimento sobre os danos causados pelo desastre de Mariana também é explorado.

2.1 Recifes de Corais, relevância e ameaças

Sociedades se concentram em grande parte ao longo das costas. Existem previsões de que até 50 % da população mundial morará perto do litoral por volta de 2030. Devido a ocupação, as regiões costeiras estão entre as zonas ecológicas mais impactadas pela ação humana[3].

O valor total estimado de ganhos apenas com o turismo em recifes de coral é de aproximadamente 36 bilhões de dólares no mundo inteiro[66]. A degradação local por si só diminui a capacidade do local oferecer serviços para a população, tais como comida, água e amenidades[3].

Com o aquecimento global e o conseqüente aumento dos níveis do oceanos, recifes ocupados por corais podem ter um papel ainda mais destacado de defesa da costa contra mudanças climáticas. Com o aumento previsto dos nível dos oceanos, é esperado uma maior amplitude e força de ondas nas costas, os recifes podem aliviar o impacto dessa movimentação [24]. Apesar disso, impactos causados podem fragilizar a capacidade dos ecossistemas de resistirem a desastres naturais[3].

O declínio na camada coralínea dos recifes é associado as ações do ser humano sobre o globo. Os impactos ocorrem em escala global, como os efeitos do aquecimento dos mares, e na escala local, como a disseminação de doenças causadas por patógenos, sobrepesca e outras interações humanas diretas. Como em outros ambientes onde há disputa por espaço, a diminuição dos corais em um recife resulta por um aumento de outras espécies, em geral algas ou cianobactérias [65][72].

Estudos realizados na República Dominicana indicaram uma cobertura de algas nos recifes maior em regiões próximas a comunidades do que em áreas de preservação. A atuação humana pode ter grande influência para a proliferação de algas em recifes [22].

Para a região do caribe há evidências que em recifes localizados próximos a costa há menor abundância de peixes, indícios de danos físicos aos corais por âncoras, e maior número de espécies coralíneas infectadas por doenças em comparação com áreas protegidas. Os resultados obtidos foram semelhantes em outras áreas desprotegidas do caribe [22].

A sobrepesca de peixes herbívoros causa um rápido desequilíbrio ecológico com a mudança de fase dos ambientes recifais. Com a degradação das condições locais, há uma substituição gradativa de corais construtores por algas epilíticas e macroalgas, com a diminuição da biomassa de peixes, que são predadores naturais dessas espécies vegetais, não há um elemento natural que contenha esse crescimento [72].

A mesma interação ecológica é vista na região do banco de Abrolhos, cerca das reservas marinhas existem evidências de concentração de peixes maior do que em outras regiões desprotegidas . Áreas protegidas funcionariam como uma maneira eficiente de garantir a diversidade de espécies tanto de peixes quanto de corais, ao mesmo tempo em que seria uma fonte de renda para comunidades locais, funcionando como uma espécie de santuário [10].

Em um cenário de disputas biológicas do bentos o aumento das algas seria apenas uma consequência da degradação dos corais. Porém indícios sugerem que esse processo contém um ciclo de auto reforço, o crescimento das espécies de algas leva a

um aumento no carbono orgânico dissolvido e do açúcar no ambiente. Bactérias que carregam agentes patogênicos aos corais se alimentam dessas substâncias, crescendo sua proporção na comunidade, gerando assim um ciclo de reforço negativo para as espécies coralinas [65].

O estudo confirma que durante o processo de degradação das espécies coralinas nos recifes, não apenas a comunidade microbiana local sofre mudanças estruturais, como também essas mesmas mudanças aceleram o processo. Novos aspectos das comunidades microbianas diminuem a resistência das espécies de corais, tornando-as mais suscetíveis aos próprios agentes patogênicos. A presença de algas e outros patógenos em grandes quantidades não só indica o estado ruim de saúde dos recifes, como pode significar que este está em um processo de degradação [65].

As cianobactérias tem maior capacidade de resistir a ações antrópicas, por isso há uma tendência do aumento de sua cobertura em corais degradados. Em geral, com a crescente presença dos colchões existe uma tendência a diminuição da camada de corais e um aumento da presença de macroalgas, esponjas e algas epilíticas [9].

A nível global, cresce o número de casos de branqueamento dos corais. Em 2016 ocorreu no mundo uma epidemia dessa doença como nunca antes havia sido registrado. Suspeita-se que o evento foi impulsionado pelo recorde de temperaturas no mundo em 2015-2016. Outros fatores que poderiam impulsionar a propagação da doença como a qualidade da água e as atividades pesqueiras, apresentaram pouca relevância estatística nos resultados medidos [39].

Para as espécies atingidas pela doença a expectativa de recuperação dos corais é de 10 a 15 anos, para indivíduos de rápido crescimento. No caso de morte por branqueamento é esperado que a recuperação leve décadas. Com o aquecimento global cada vez mais se mostrando uma realidade, é possível imaginar que as ondas de branqueamento se tornem mais frequentes, causando cada vez mais danos aos corais [39]. Na figura 2.1 uma colônia de recifes branqueados.



Figura 2.1: Corais afetados pelo branqueamento.

Fonte: HUGHES *et al.*[39]

2.2 Caracterização do Banco de Abrolhos

O Banco dos Abrolhos é reconhecido como uma área relevante para preservação da biodiversidade nacional. O governo brasileiro considera a área como “extremamente elevada” para prioridade à conservação[47].

Com a utilização de um sonar foi revelado recentemente a estrutura de composição do banco de Abrolhos, a cobertura de recifes total chegou a 8845 km^2 , maior do que era esperado, e também revelou o maior banco de rodolitos do mundo chegando a 20904 km^2 . Há ainda um megahabitat composto por sedimentos não consolidados[49]. As divisões podem ser vistas na figura 2.2.

Próximo ao litoral as áreas são ocupadas em geral por sedimentos não consolidados, sendo esta a região de menor atividade biológica, e onde há maior atividade comercial humana [49].

As formações de rodolitos estão localizadas na faixa final do banco, mais próximo ao oceano atlântico. Ocupando uma extensão de 43 % do espaço total do banco de Abrolhos, o equivalente a grande barreira de corais na Austrália. Essa é a maior área ocupada por rodolitos conhecida no mundo [49] .

Rodolitos são algas calcárias grandes produtoras de CaCO_3 . No banco de Abrolhos a quantidade de carbonato de cálcio produzida pelas espécies forma um dos maiores bancos dessa substância, favorecendo o desenvolvimento dos corais. A engenhosidade das formações deste grupo também fornece micro habitats para peixes invertebrados, favorecendo a biodiversidade da região e aumentando a oferta de pescado disponível para as comunidades locais [4]. Um Rodolito pode ser visto na figura

2.3.

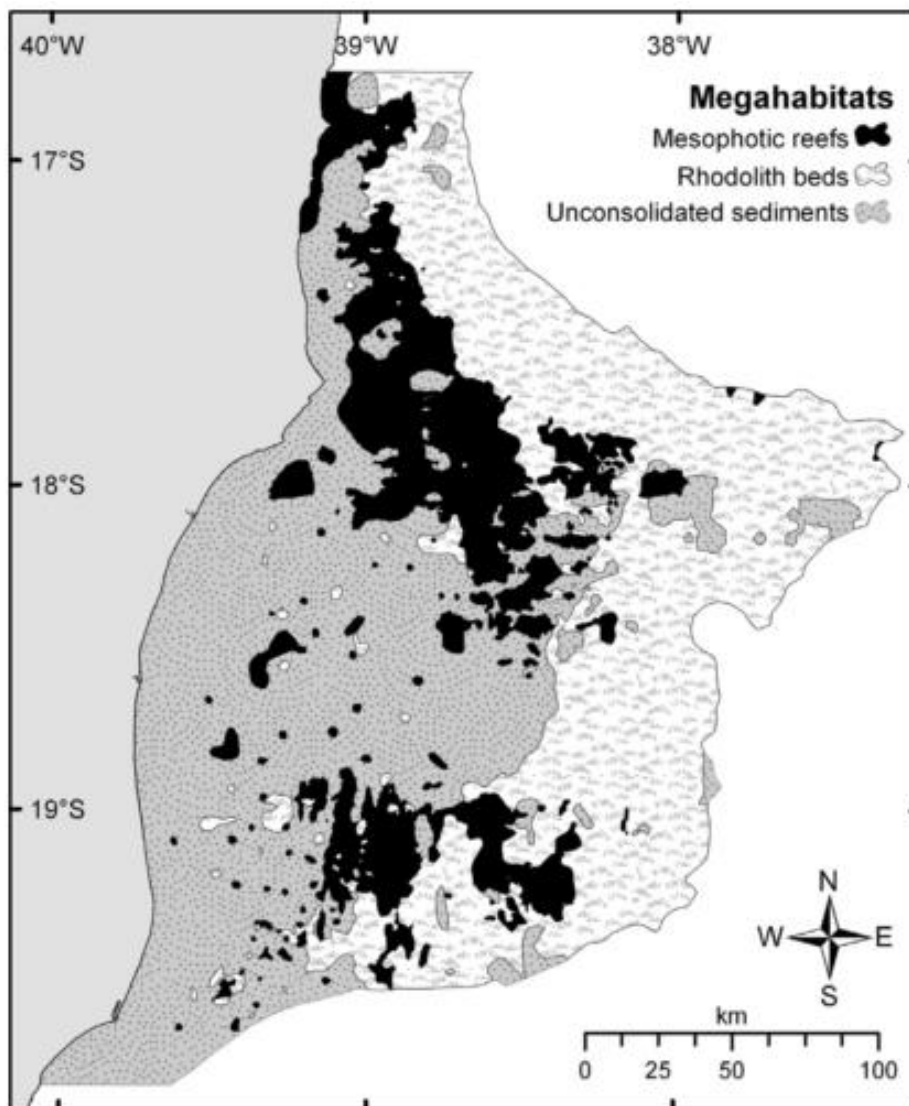


Figura 2.2: Três megahabitats do banco de Abrolhos.

Parte preta simboliza as formações recifais, cinza escuro é apresentada a região composta por sedimentos não-consolidados e em cinza claro o banco de Rodólitos.

Fonte: MOURA *et al.*[49]

As formações recifais estão localizadas em sua maioria ocupando a faixa longitudinal central, há um aparente hiato no meio da cadeia. Enquanto os outros tipos de megahabitats se caracterizam como estruturas de continuidade e configuração topográfica mais simples, os recifes, se caracterizam por sua intermitência e complexidade [49].

Nos recifes de corais ocorrem as formações mais ricas em termos biológicos. Todas as formações se encontram em distâncias da costa que variam de 5 a 65 km. Porém há uma grande diferença de composição do bentos entre os grupos de recifes, fatores como incidência de luz, profundidade e inclinação do solo podem ter influência sobre

esses habitats [31].



Figura 2.3: Foto de um Rodolito.

Fonte: AMADO *et al.*[4]

Algas do tipo epilíticas são o grupo mais presente dentre as espécies que habitam os recifes de Abrolhos, podendo chegar a ocupar mais da metade da cobertura total do bentos. Sua presença é associada a uma diminuição da complexidade dos recifes, assim como a diminuição dos habitats disponíveis para biodiversidade coralina[70].

As análises metagenômicas das algas apresentaram estabilidade, independente do ponto de recolhimento, que abrangeu uma vasta área do banco de Abrolhos. Além disso, aparentam maior estabilidade em relação a mudanças no ambiente e crescem com maior velocidade que os organismos coralíneos [70].

O aparecimento das algas poderia desfavorecer a diversidade de espécies. Análises sugerem uma relação entre a presença de peixes e uma cobertura coralina mais saudável, indicando o efeito negativo da sobrepesca sobre os corais. A redução da quantidade de herbívoros eleva naturalmente a população de macroalgas, já que sem seus predadores naturais podem se expandir com maior eficiência [10]. A diferença visual entre os corais que se localizam em regiões com diferentes níveis de preservação, pode ser observados nas figuras 2.4 e 2.5.



Figura 2.4: Corais localizados em zonas não protegidas.

Fonte: BRUCE *et al.*[10]



Figura 2.5: Corais localizados em zonas marinhas protegidas.

Fonte: BRUCE *et al.*[10]

Diferentes recifes da região de Abrolhos foram analisados temporalmente utilizando a cobertura do bentos no estudo de FRANCINI-FILHO *et al.*[31]. Buscando entender as possíveis ações humanas ou naturais que contribuíram para as mudanças observadas. Amostras foram recolhidas em lugares preservados, preservados com múltiplos usos e não preservados [31].

O mapa da região pode ser observado na figura 2.6. As linhas contínuas ao redor das formações indicam que é uma área marinha protegida. Linhas pontilhadas mostram áreas sobre proteção de áreas extrativistas. Recifes que não possuem nenhum tipo de circunferência estão em áreas desprotegidas.

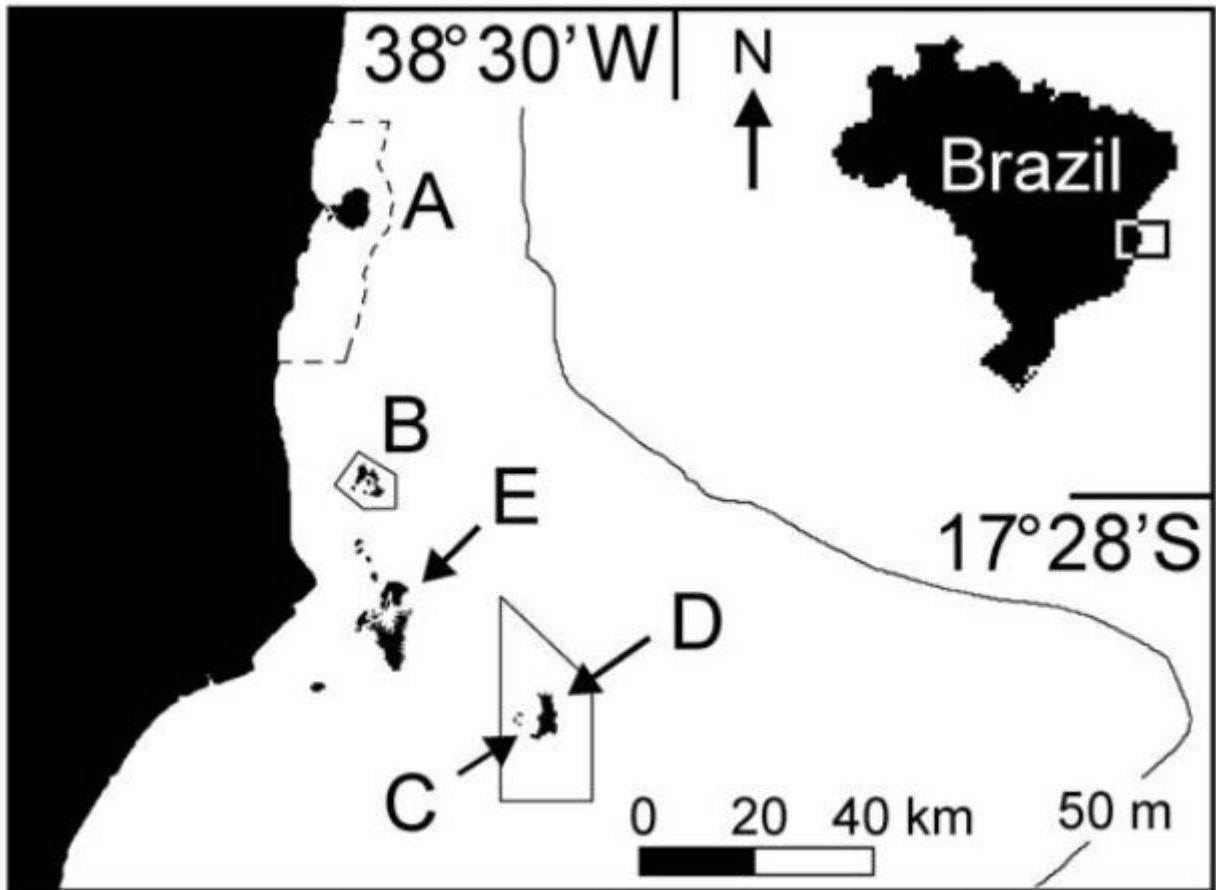


Figura 2.6: Recifes localizados na região centro-norte do Banco de Abrolhos e suas respectivas áreas de proteção.

A - Itacolomis, recifes destinados a múltiplos usos B - Timbebas, zona protegida ,
 C - Arquipélago de Abrolhos, zona protegida , D - Parcel de Abrolhos, zona
 protegida. E – Recifes costeiros desprotegidos.

Fonte: FRANCINI-FILHO *et al.*[31]

Análises de recifes do banco de Abrolhos foram desenvolvidas para os períodos de 2003-2005 e 2006-2008 a partir das fotos tiradas do bentos. Para utilização dos testes de ANOVA, as porcentagens analisadas dos bentos foram convertidas na raiz do número original .O método de análise de variância foi proposto para analisar variações na composição do bentos em diferentes dimensões [31].

O primeiro teste foi utilizado para determinar variações temporais nos padrões encontrados. Apesar de certas espécies apresentarem aumento de importância relativa com o passar do tempo, os resultados não indicaram mudanças estatisticamente consideráveis entre os anos observados [31].

O segundo teste de ANOVA foi utilizado para determinar diferenças entre os tipos de habitats. Há indicações de que existe uma mudança estatística de composição entre o topo e a parede dos recifes, essa variação pode ser explicada pela diferença

de profundidade e recebimento de luz de cada um dos locais[31].

Um terceiro teste ANOVA foi realizado para identificar mudanças significativas entre as localizações onde foram encontrados os recifes. Novamente há pouca evidência estatística, a composição do bentos em diferentes regiões apresentou pouca variação[31].

Foram utilizados métodos de ordenação para localização das amostras em duas dimensões. Os resultados também indicaram uma maior relevância do tipo de habitat em relação a avaliação temporal e por recife. Os modelos utilizados foram a Ordenação Multidimensional não métrica e a Análise da Correspondência Canônica [31].

Os efeitos da preservação e principalmente da não preservação podem ser observados nos recifes. Nas áreas preservadas há maior presença de peixes e uma menor concentração de algas. Além disso, nas áreas mais suscetíveis a ação humana, foram encontrados maior concentração de vírus patogênicos aos corais, enquanto nas áreas protegidas, há maior abundância de genes relacionados a fotossíntese [10].

A principal atividade humana na região atualmente é a indústria pesqueira, é disseminada por um largo território, com estimativa de ocupar mais de 25 % apenas da área de sedimentos não consolidados. As pescaria é uma das formas da população nativa conseguir seu sustento, na década de 90 começaram a surgir pressões locais pelo estabelecimento de reservas extrativistas da região [49].

Em menor escala também há no banco outras indústrias como as indústrias de óleo e gás, mineração e dragagem. Tais atividades possuem potencial econômico, ao mesmo tempo em que oferecem um grande risco a saúde do ecossistema. Esses setores ocupam hoje por volta de 1 % da área total do banco [49].

Hoje além das reservas extrativistas, também existem as zonas marinhas protegidas, que buscam proteger as formações recifais mais importantes do ponto de vista biológico. Nessas regiões não há permissão para pesca, algumas das poucas atividades permitidas são a pesquisa e o turismo, com restrições. No total, o espaço ocupado pelos parques e reservas não chega a 0,5 % da área total do banco [49].

Já foi proposto estruturação de um plano abrangente para preservação do bioma da região. Levando em conta as necessidades das comunidades locais e também possíveis interesses econômicos na área, poderia ao mesmo tempo ajudar na manutenção das estruturas biológicas e resolver conflitos pendentes. Esse plano deveria ser elaborado em uma agenda transparente, que inclua em sua composição, cientistas, ONGs, representantes das indústrias e locais [49].

2.3 Modelos de Aprendizado de Máquina para Recifes

Com o aumento do número de informações disponível, cresceu a necessidade por novos métodos científicos que consigam explorar essa abundância, com isso aumentou a utilização de inteligência artificial, cujas técnicas conseguem analisar situações complexas de forma mais acurada e interpretável que métodos de estatística convencionais [61].

Métodos de aprendizado de máquina tem a flexibilidade para compreender problemas complexos e estrutura-los, tornando-os uma ferramenta adequada para o entendimento e previsão de interações ecológicas, superando métodos convencionais de regressão [56].

As principais características do método de aprendizado de máquina destacada, é a capacidade de analisar variáveis que tem comportamento não linear e a capacidade de realizar classificações em tempo real. Uma das vantagens da abordagem é que o processo de desenvolvimento "aprende" a partir da informação disponível, portanto o modelo é adaptável para novas situações sem a exigência de grande esforço [21].

Segunda vantagem, não é necessário a formulação de um modelo a priori, nem precisam ser levantadas hipóteses sobre a distribuição dos dados, isso pode facilitar em especial problemas complexos em que há pouca informação. Uma última vantagem, essa do ponto de vista computacional, é que as análises podem ser feitas em paralelo, exigindo pouco esforço computacional, embora alguns métodos estatísticos convencionais também obtenham ganhos com esse paralelismo [21].

Até o surgimento de novas tecnologias que permitiram uma abundância de informações que não era possível anteriormente, os estudos em recifes se restringiam a analisar dados obtidos no campo. Com o aumento do número de informações disponível, cresceu a necessidade por novos métodos científicos que consigam explorar essa abundância, com isso aumentou a utilização de inteligência artificial, cujos métodos conseguem analisar situações complexas de forma mais robusta e interpretável que métodos de estatística convencionais [61]. A ênfase deste estudo será em ferramentas que utilizam árvores de regressão ou redes neurais.

O método de árvores de crescimento progressivo (*Boosted Regressions Tree*) já foi utilizado na modelagem de riquezas de espécies em diversos estudos. Inclusive para a previsão de uma espécie coralínea no banco de Abrolhos [14].

Nesta pesquisa o modelo apontou que a espécie *Scolymia wellsi* (*Mussidae*) tem preferência por temperaturas mais altas e por ambientes menos iluminados.

Outra ferramenta de de aprendizado de máquina que é utilizada para previsão de riqueza de espécies é a Rede Neural. A modelagem de concentração de cianobactérias foi realizada anteriormente em FLOMBAUM *et al.*[28], que utilizou o modelo de rede

neural para a análise de para dois tipos de espécies, *Prochlorococcus* e *Synechococcus*. Juntos os organismos desse grupo são responsáveis por uma fração importante da produção primária marinha mundial. O modelo nesse caso foi realizado com a abundância das bactérias em nível microbiano, e teve como entradas os padrões abióticos locais de várias amostragens ao redor mundo.

Foram comparados 3 métodos estatísticos para analisar a questão, redes neurais, modelos de regressão paramétricos e não paramétricos . Os resultados do modelo de aprendizado de máquina foi considerado tão satisfatório quando o modelo de regressão não-paramétrico, e melhor do que o modelo paramétrico. As variáveis que se mostraram mais determinantes foram a temperatura e a radiação ativa fotosintética[28].

A capacidade de redes neurais em prever resultados foi utilizada pelo estudo para projetar o futuro das concentrações das cianobactérias nos oceanos. Os resultados indicam um aumento no futuro, principalmente em decorrência do aquecimento global [28].

A utilização de métodos de aprendizado de máquina na previsão da camada coralínea foram utilizados em ROOPER *et al.* [59]. Utilizando informações bióticas e abióticas o trabalho buscou calcular a probabilidade da existência de corais no golfo do Alasca.

Na pesquisa do golfo do Alasca foram comparados 4 métodos, regressão linear múltipla, regressão aditiva, florestas aleatórias e árvores de crescimento progressivo. Os métodos de floresta superaram os métodos de regressão, em se tratando de capacidade de previsão[59].

Por outro lado, os modelos de regressão apresentaram resultados de correlação entre as variáveis explicativas e dependentes considerados mais coerentes. Os gráficos de interação entre variáveis dos modelos de floresta apresentaram descontinuidades que podem atrapalhar a interpretação [59]. Um exemplo dessa análise está na figura 2.7.

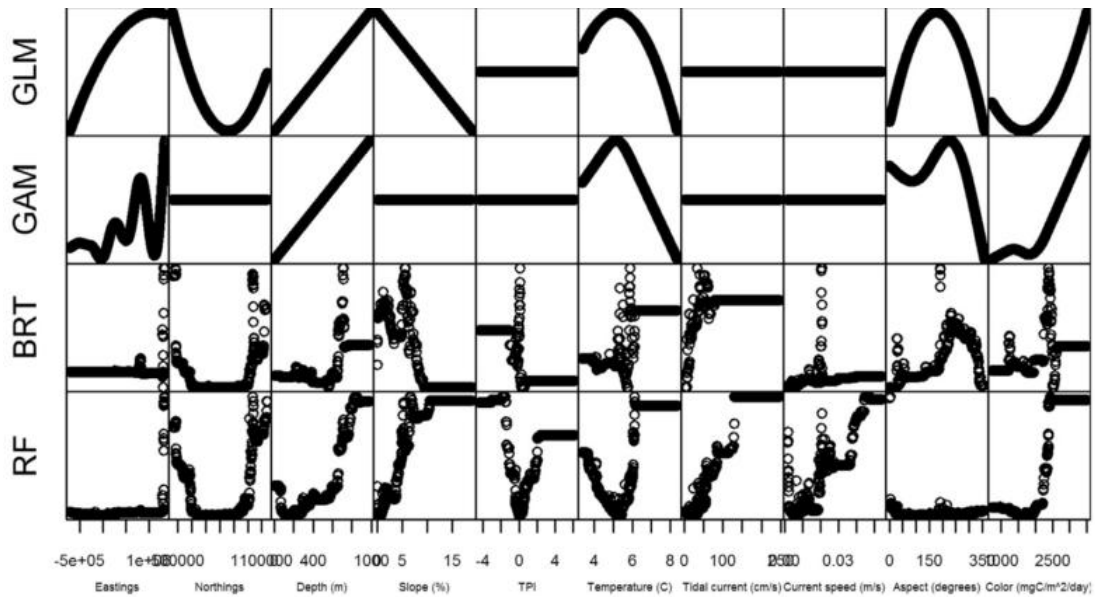


Figura 2.7: Comparação da análise de sensibilidade para quatro métodos que buscam prever a cobertura de corais no Alasca.

GL- Regressão Linear Múltipla; GAM - Modelos aditivos; BRT - Árvore de Crescimento Progressivo; RF - Florestas Aleatórias.

Fonte: ROOPER *et al.*[59]

Enquanto os métodos de floresta utilizados por ROOPER *et al.*[59] demonstraram maior consistência na previsão das espécies, sua capacidade de interpretação dos resultados não foi clara da mesma forma. Os gráficos da figura 2.7 representam a análise de sensibilidade de cada uma das variáveis, que são representadas nas colunas, para cada um dos métodos, representados nas linhas.

O primeiro ponto a se levar em consideração na comparação entre os métodos é que os resultados nem sempre são semelhantes para cada variável. Outro ponto é que os métodos de floresta podem apresentar análises com padrão não identificável. Não necessariamente os saltos apresentados nos gráficos de agrupamento de árvores significam limites biológicos com grande impacto [59]. Esse cuidado na interpretação dos resultados dos modelos de aprendizado estatístico já foi levantado por outros autores [56].

2.4 Recifes da Foz do Rio Doce e o desastre de Mariana

As colônias de corais encontradas na região sul do banco de Abrolhos, próximas ao deságue do rio Doce, aparentavam saúde quando descritas por MAZZEI *et al.* Foram encontrados grupos das seguintes espécies *Montastraea cavernosa*, *Mussismis-*

lia harttii, *Mussismilia braziliensis* e *Millepora alcicornis*. Apenas alguns espécimes de *Montastraea cavernosa* apresentava sinais de branqueamento. [45].

A cobertura coralínea nativa estimada visualmente através de imagens segundo dados recém-publicados para essa área de estudo está entre 10 e 20 % da área total dos recifes. A espécie *Palythoa caribaeorum* teve uma área grande de cobertura, chegando a 10 % de alguns recifes [45].

Em relação a composição das espécies no bentos, a estrutura recifal do complexo de ambientes coralíneos do sul do Banco dos Abrolhos é bastante distinta da encontrada tanto nos litorais do sul do Espírito Santo quanto do sul da Bahia [45].

A presença de poucos predadores de níveis tróficos superiores sugere que a nível atual de pesca está acima da capacidade de recomposição do estoque natural. Também foram encontrados danos aparentemente causados pelas âncoras nos corais e redes de pescaria, denunciando a influência antropogênica. Apesar da região costeira do Espírito Santo ser um área de grande importância para a conservação, com a presença de grande número de espécies de peixes em risco de extinção, a área ainda é pouco protegida por zonas marinhas [45].

O mapa da região pode ser visto na figura 2.8, com destaque para a foz do Rio Doce. Em novembro de 2015 uma barragem localizada próxima a cidade de Mariana, no estado de Minas Gerais, se rompeu, derramando cinquenta milhões de metros cúbicos de rejeitos de minérios de ferro. Os compostos despejados eram formados principalmente por lama tóxica, e seu despejo no Rio Doce causou uma grande tragédia ambiental e humanitária [46].

A lama derramada viajou por mais de 650 km pela bacia do Rio Doce, causando danos massivos aos ecossistemas da região. Tentativas foram feitas para conter o volume de rejeitos ao longo do rio, porém com pouca efetividade. Após 16 dias os rejeitos chegaram ao mar, onde é esperado que seus impactos grandes e de longa duração. Em janeiro de 2016 a área marinha total com presença de lama do desastre era estimada em 7000 km^2 [46].

O desastre ambientais afetou a vida de mais de um milhão de pessoas, em quarenta e um municípios que beiram o rio. As principais perdas em termos de serviços ecológicos foram a diminuição do acesso a água limpa para consumo direto e agricultura, limitação dos recursos pesqueiros e a capacidade de geração de energia elétrica[26].

O riscos ambientais no médio e longo prazo não são desprezíveis, principalmente devido as concentrações de ferro encontradas na água. É possível que em situações chuva intensa ocorra a remobilização de sedimentos nocivos que se encontram no fundo do rio, liberando metais que afetam o ecossistema por muitos anos[37].

Tanto em relação ao volume de rejeitos despejados quanto em termos de área total afetada, desastre ocorrido em Marina pode ser considerado o maior do mundo.

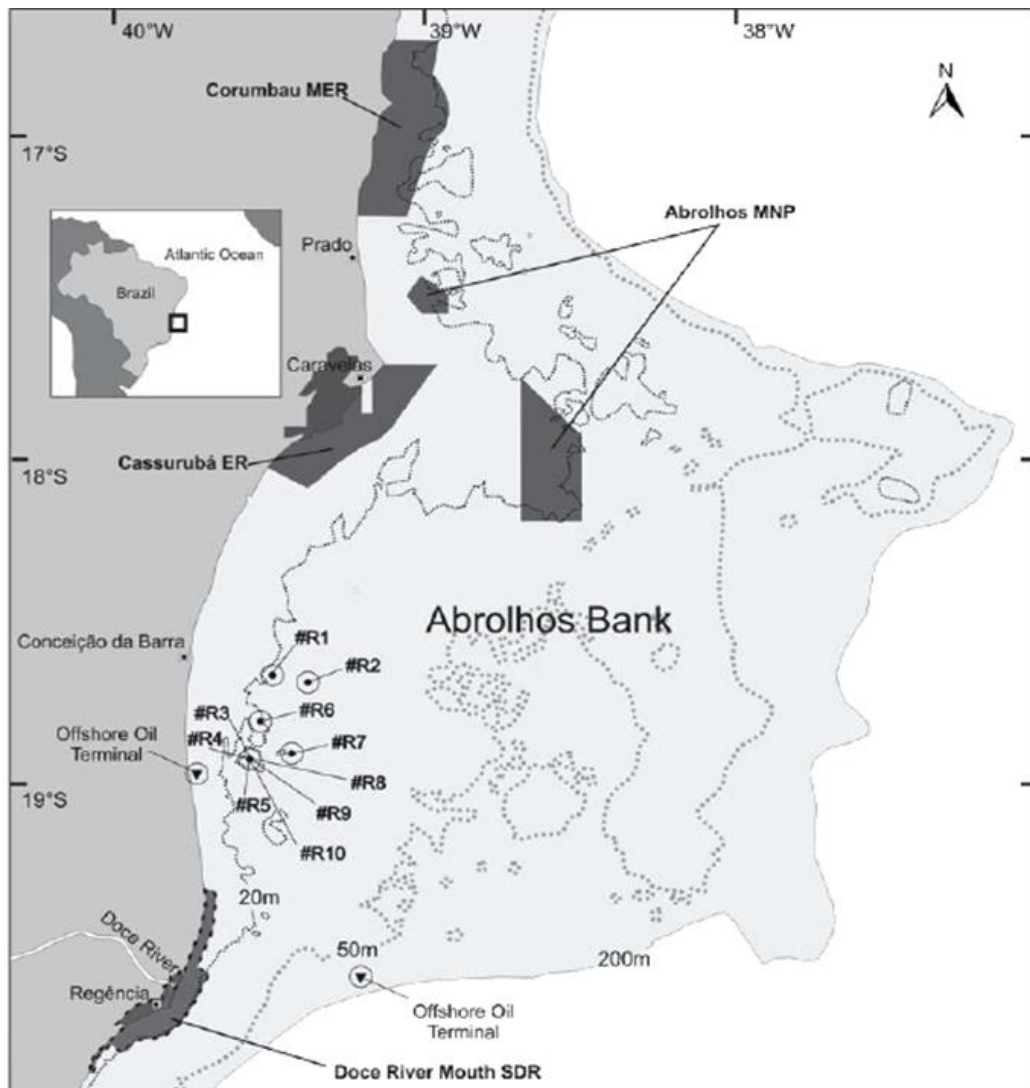


Figura 2.8: Mapa da região do banco de Abrolhos com destaque para os novos recifes descobertos, ao norte da foz do Rio Doce.

Fonte: MAZZEI *et al.*[45]

Um ano após a tragédia a empresa responsável pelo ocorrido ainda não realizou compromissos assumidos pela companhia, como a remoção da lama e o monitoramento da dispersão dos resíduos. Por outro lado, assiste-se a relaxação da lei, diminuição das penas dos envolvidos, e a demora na realização de medidas de recuperação [20].

Inicialmente se pensava que o maior fluxo do desastre teria se dirigido ao Rio de Janeiro, deixando os corais do banco de Abrolhos intactos[44]. Porém novos indícios podem elucidar se os rejeitos chegaram até a parte do banco que fica no Espírito Santo, atingindo os corais descritos por MAZZEI *et al.*, imagens na figura 2.9.

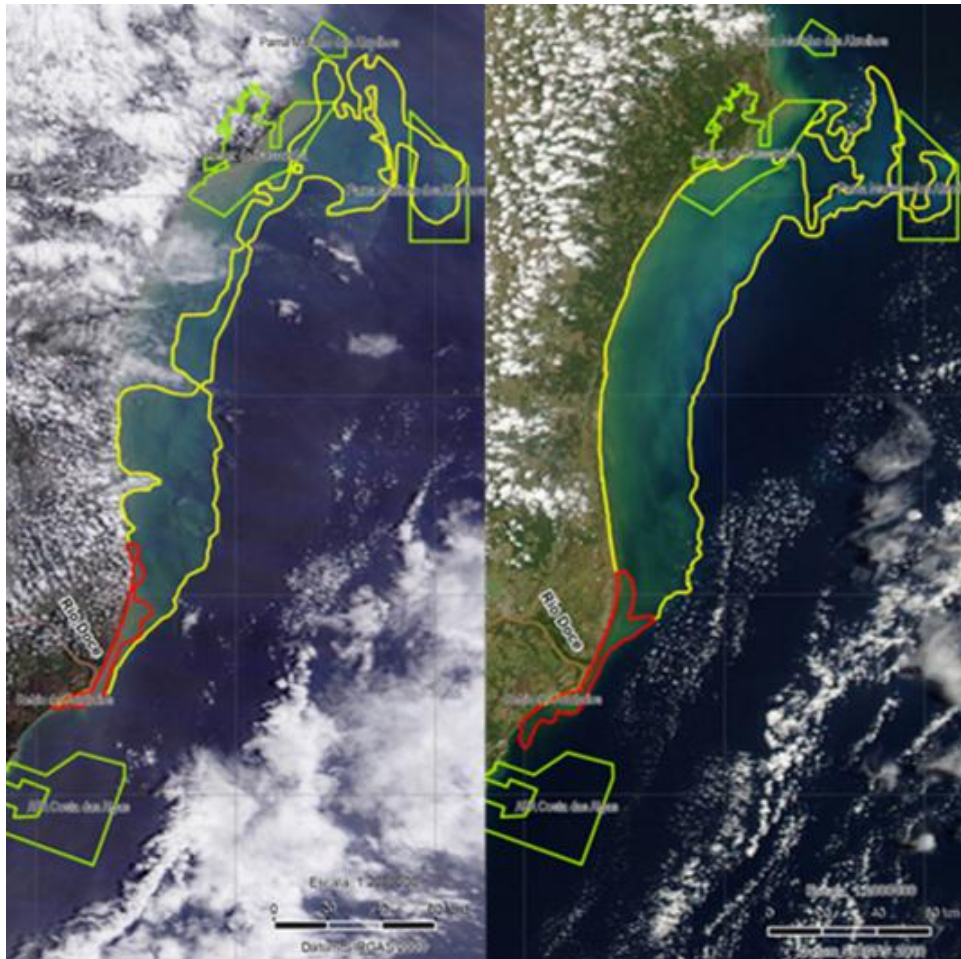


Figura 2.9: Caminho da lama do desastre de Marina ao chegar no mar da costa do Espírito Santo. Imagens de 05/01/2016 e 06/01/2016 respectivamente.

Polígonos vermelhos representam sedimentos de maior concentração de resíduos, polígonos amarelos sedimentos de menor concentração e nos polígonos verdes estão representadas as áreas de conservação nacional.

Imagens obtidas através do sistema MODIS.

Fonte:MAZZEI *et al.*[45]

Na figura 2.9 estão as imagens observadas a partir de sensoriamento remoto. Há indícios de que a dispersão dos sedimentos se deu tanto para o sul, quanto para o norte da foz do Rio Doce, chegando até mesmo ao banco de Abrolhos. Até onde se sabe sobre a região, os recifes de Conceição da Barra são a primeira formação que apresenta corais no trajeto dos rejeitos [45].

Capítulo 3

Método

Em decorrência da crise ambiental mundial cresce o número de pesquisas voltadas ao meio ambiente. Essas pesquisas têm como característica avaliar e monitorar as características da biodiversidade, por meio de métodos que envolvem coleta, análise, simulação, modelagem, visualização e compartilhamento das informações recolhidas [69].

Levantar as hipóteses e entender os resultados é uma questão delicada para análises biológicas. Nem sempre se um certo fator apresenta maior relevância para a variação das amostras significa necessariamente que este seja o principal fator de variância[53].

Nesta seção são apresentados os dados coletados e a forma de tratamento. Os métodos que foram utilizados para análise foram definidos, com ênfase na parte teórica.

3.1 Índices de Diversidade

Os índices expostos foram baseados em LEGENDRE & LEGENDRE [42].

Um dos parâmetros para medir a riqueza biológica de determinado sítio é a diversidade de espécies, feita através da contagem simples de tipos diferentes de indivíduos encontrados em determinado local.

Pode ocorrer que em determinados espaços biológicos poucos indivíduos estejam presentes de forma muito abundante, enquanto outras espécies se façam presentes escassamente ou com um único representante. Uma alternativa a utilização de índices que combinem os dois resultados. H pode ser utilizado para calcular a dispersão de espécies de acordo com a equação 3.1. Onde p_i é a frequência relativa da espécie i em um universo de q espécies identificadas. Essa forma de cálculo recebe o nome de *Entropia de Shannon*.

$$H = - \sum_{i=1}^q p_i \log p_i \quad (3.1)$$

Na equação 3.1, $H = 0$ representa a situação em uma espécie domina todo o ambiente. Quanto maior o número de espécies encontradas q e maior a uniformidade de suas distribuições, maior será o valor assumido pelo índice.

Outros autores consideram que a medida da diversidade deve ser interpretada a partir de dois fatores separados. Um desses fatores é o número de espécies encontradas, como descrito a pouco. O outro é um índice que calcula a uniformidade da distribuição de seres no ambiente. Diferentemente do índice H que é uma composição das duas medidas, a visão em separado desses valores pode fornecer outras informações relevantes.

A forma mais simples de calcular a equitatividade compara o resultado encontrado de entropia, que está descrito com o valor máximo possível para o número de espécies encontradas. O valor máximo para o índice H é encontrado quando todas as espécies estão igualmente distribuídas, de acordo com a equação 3.2.

$$H_{max} = - \sum_{i=1}^q \frac{1}{q} \log \frac{1}{q} = \log q \quad (3.2)$$

A partir do valor da entropia máxima, o índice de equitatividade de Pielou pode ser calculado pela equação 3.3.

$$J = \frac{H}{H_{max}} = \left(- \sum_{i=1}^q p_i \log p_i \right) / \log q \quad (3.3)$$

Neste estudo ambas as metodologia foram aplicadas para análise do bentos, os métodos fornecem informações diferentes a serem interpretadas. Uma composta por uma análise combinada do número de espécies com a equitatividade da distribuição, e outra feita a partir da entropia.

3.2 Métodos de Ordenamento

Em ecologia as amostras normalmente são ordenadas em função do número m de espécies, em uma matriz de n dimensões, onde cada linha representa uma amostra. Técnicas de ordenação visam representar em poucas dimensões dados multidimensionais, de forma a fornecer o máximo de informação possível sobre a semelhança entre os pontos [68].

A transposição de dados multiespaciais para em espaços reduzidos começam pela ordenação dos pontos no espaço multidimensional. Assim feita a representação

dos pontos em poucas dimensões, de forma a preservar da melhor forma possível a distância existente no multiverso [42].

Por essa razão surgiram os gráficos bivariados, que incluem em sua composição as duas dimensões de maior interesse para o estudo, representando o maior grau de variação possível. Também é a entender o quanto da variação total está representada nas dimensões utilizadas [42].

Assim como para os índices de diversidade, a matemática das análises de ordenamento foram baseados em Legendre & Legendre[42].

3.2.1 Análise dos Componentes Principais

Cada método de ordenação tem sua metodologia voltada a maximizar um tipo de resultado. O método de Análise dos Componentes Principais (ACP) tem como prioridade transpor com a maior eficiência possível as distâncias euclidianas no multiverso para o universo bidimensional.

Em uma matriz de m dimensões é possível retirar m eixos principais. O primeiro eixo principal é o que representa por si só a maior variação possível das amostras, e os eixos subsequentes representam a máxima variação desconsiderando os eixos já obtidos.

O método para obtenção dos eixos principais pode ser encontrado resolvendo a equação 3.4. Os auto vetores u_k representam os eixos principais da matriz S .

Onde S representa a matriz de dispersão, que pode se encontrada a partir dos dados originais Y , seguindo a equação 3.5.

Por fim, a variável λ_k é a matriz de autovalores da matriz de dispersão e pode ser encontrada pela equação 3.6.

$$(S - \lambda_k I)u_k = 0 \quad (3.4)$$

$$S = \frac{1}{n-1}[y - \bar{y}]'[y - \bar{y}] \quad (3.5)$$

$$|S - \lambda_k I| = 0 \quad (3.6)$$

Esse conjunto de equações tem algumas características que o torna interessante para a análise ecológica de acordo com Numerical Ecology.

i. Como a matriz S é simétrica, os autovalores encontrados u_k são ortogonais, e consequentemente linearmente independentes entre si.

ii. Os autovalores λ_k representam a representação da variância obtida em cada

um dos eixos principais.

iii. Consequentemente aos itens i. e ii., é possível em poucas dimensões representar uma grande parcela da variância total dos pontos, ao mesmo tempo em que sabe-se a quantidade dessa variação que é representada.

Podem ser calculados quantos auto vetores u_k quanto houver dimensões na matriz original, representadas pelo número de variáveis descritivas m . Esse auto vetores depois de normalizados formam as colunas da matriz U .

Cada uma das amostras representadas por uma linha na matriz original terá um posicionamento no novo plano desenvolvido pelo método de ACP. Os novos sistemas são obtidos a partir da equação 3.7.

$$F = [y - \bar{y}]U \quad (3.7)$$

A qualidade da representação pode ser obtida pela equação 3.8. Onde p é o número de dimensões obtidas e m é o número total de variáveis, sendo assim $p \leq m$. A razão obtida pode ser considerada como um comparativo ao coeficiente R^2 da regressão.

$$\left(\sum_{i=1}^p \lambda_k\right) / \left(\sum_{i=1}^m \lambda_k\right) \quad (3.8)$$

A ACP fornece também a importância de cada uma das variáveis explicativas na formação do gráfico bidimensional, que é obtido a partir da equação 3.9. A matriz resultante dessa multiplicação apresenta a importância relativa de cada uma das variáveis descritivas para a formação dos eixos. Os resultados obtidos na equação permitem analisar quais as variáveis mais significativas para as distâncias entre os pontos observados.. A matriz Λ é composta pela diagonal dos autovalores, que são obtidos pela equação 3.6.

$$U\Lambda^{1/2} \quad (3.9)$$

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & 0 & \lambda_n \end{bmatrix} \quad (3.10)$$

Para adicionas os vetores representativos da importância das variáveis explicativas. Os elementos da matriz calculada como $u_{jk}\sqrt{\lambda_k}$ são as projeções de cada variável j em cada eixo principal k , no gráfico ACP são representadas pelas setas.

Em um gráfico de ACP os ângulos formados pelas setas descritivas das variáveis são projeções de suas covariâncias reais. O ângulo formado por uma variáveis descritiva j com um dos eixos k é encontrado pela equação 3.11. Os valores de s_j são as diagonais da matriz S encontrada na equação 3.5.

$$\arccos[u_{jk}\sqrt{\lambda_k}/s_j] \quad (3.11)$$

Portanto o cálculo do ACP fornece dois tipos de informações, a primeira é relativa ao ordenamento dos pontos e as distâncias relativas. A segunda se refere a contribuição das variáveis para essa organização e uma projeção das correlações existentes entre os vetores descritivos.

O ACP é utilizado com as variáveis de maior relevância, aquelas espécies que estiveram mais presentes nos amostras selecionadas, assim como entender a contribuição das variáveis para a organização do gráfico.

3.2.2 Ordenação Multidimensional não Métrica

O método de Ordenação Multidimensional não-Métrico, conhecido por sua sigla em inglês nMDS, tem como sua prioridade manter os agrupamentos de pontos da forma mais fiel possível. Existe uma outra lógica na elaboração dos planos em relação ao ACP, que são calculados em cima dos eixos principais, enquanto no nMDS a principal parte do processo consiste em ajustar a distâncias entre os pontos.

O método de ACP utiliza auto vetores para calcular os eixos que maximizam a variância da ordenação. Os eixos do nMDS são arbitrários e não necessariamente tem variação decrescente. Podendo ser rodados, centrados ou invertidos[68]. O índice que determina a melhor representação dos eixos é o *stress*, que representa a percentagem de divergência entre as distâncias da matriz original e no espaço reduzido.

Os passos para desenvolvimento de uma análise nMDS são:

1) Escolher o número de dimensões m que serão utilizadas na ordenação.

2) Formular a configuração inicial dos pontos, podendo ser utilizado a disposição espacial final de outro modelo, como por exemplo o ACP. Para evitar soluções que são mínimos locais distantes dos mínimos globais, o modelo é utilizado diversas vezes com diferentes ordenações iniciais.

3) Calcular as distâncias d_{ij} no espaço ordenado, comparando todos os pontos par-a-par, uma das formas de realizar o cálculo é a partir das distâncias euclidianas.

4) Comparar as distâncias originais organizadas na matriz D_{ij} com as obtidas no ordenamento d_{ij} a partir de uma regressão. As distâncias previstas entre os pontos são previstas pelo método de regressão e formam a matriz \hat{d}_{ij})².

5) O método para comparar a adequação da alocação dos pontos é pela fórmula do *stress*, que compara as distâncias obtidas pelo nMDS com as distâncias esperadas pela regressão do passo 5, de acordo com a equação 3.12.

$$S = \sqrt{\sum_{i,j} (d_{ij} - \hat{d}_{ij})^2 / \sum_{i,j} d_{ij}^2} \quad (3.12)$$

6) Mover as variáveis de forma a diminuir o índice de *stress*. O método utilizado é um algoritmo descendente, calculando cada uma das derivadas parciais para cada uma das distâncias d_{ij} .

7) Repetir os passos 3 a 6 até atingir um número pré-determinado de iterações ou a convergência dos valores.

Apesar de serem da mesma família de métodos o ACP e o nMDS tem objetivos distintos. A utilização de cada metodologia neste trabalho tem finalidades diferentes. O nMDS é utilizado para comparar a distribuição espacial dos pontos com todas as dimensões de variáveis colhidas no bentos.

3.3 Métodos de Aprendizado Estatístico

Métodos de aprendizado estatístico ou aprendizado de máquina, estão dentro de uma área de rápido crescimento em especial na bioinformática. A partir de modelos não-lineares e complexos é possível identificar as relações presentes na ecologia e realizar previsões a respeito[56].

A partir dos dados coletados de bentos e por sensoriamento remoto quatro métodos terão suas capacidades de previsões testadas, e seus métodos de análise comparados.

Para facilitar a comparação, é proposto a transformação das variáveis dependentes e explicativas [54]. No caso das primeiras, a transformação deve converter os valores para o intervalo [0,1], de acordo com a fórmula 3.13. Para o caso das variáveis explicativas, a conversão é feita para os valores normais pela equação 3.14.

$$r_n = \frac{y_n - \min(Y)}{\max(Y) - \min(Y)} \quad (3.13)$$

$$z_n = \frac{x_n - \bar{X}}{\sigma_n} \quad (3.14)$$

Tal transformação não tem como objetivo melhorar o desempenho do modelo. A razão da transformação é padronizar a entrada e saídas de dados, e assim ter mais segurança para realizar as análises. Para facilitar a interpretação nos gráficos gerados pelos modelos é feito um retorno para os dados originais.

3.3.1 Regressão Linear e Regressão Aditiva

Nesta seção são apresentadas de forma sucinta a teoria desenvolvida sobre Regressões lineares múltiplas e Modelos aditivos gerais. A base teórica da seção é WOOD & AUGUSTIN [71].

Um método de regressão linear tradicional tem a estrutura da equação 3.15. A variável dependente é representada por y , as p variáveis explicativas são representadas por x_1, x_2, \dots, x_p , β_0 é o intercepto, $\beta_1, \beta_2, \dots, \beta_p$ são os coeficientes de regressão e e representa o termo de erro

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + e \quad (3.15)$$

Para o caso dos modelos generalizados é feita uma transformação da variável dependente y em uma função do tipo $g(\mu)$, como na equação 3.16. Essa transformação é uma função linear dos parâmetros.

$$g(\mu) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (3.16)$$

Nos modelos aditivos aditivas há uma substituição dos coeficientes de regressão $\beta_1, \beta_2, \dots, \beta_p$ por funções de regressão $s_1(x_1) + s_2(x_2) + \dots + s_p(x_p)$. Unindo a lógica da generalização e funções aditivas o resultado pode ser visto na equação 3.17.

$$g(\mu) = \beta_0 + s_1(x_1) + s_2(x_2) + \dots + s_p(x_p) \quad (3.17)$$

A vantagem da utilização deste tipo de modelo para modelar uma variável dependente é que existe uma maior flexibilidade para analisar o comportamento entre as variáveis. Ao invés de uma correlação apenas linear de cada variável explicativa com a variável dependente, é possível nesse caso estabelecer relações mais sensíveis

Cada uma das subfunções que compõe o modelo aditivo generalizado é composto por um grupo de funções, como na equação 3.18. Uma das formas possíveis de modelar essa relação é através de funções polinomiais como na equação 3.19. Os coeficientes $\alpha_1, \alpha_2, \dots, \alpha_p$ representam o peso de cada grau de potência no modelo final.

$$s(x_i) = \sum_{j=1}^n \alpha_j b_j(x_i) \quad (3.18)$$

$$s(x) = \alpha_1 + \alpha_2 x + \alpha_3 x^2 + \dots \alpha_n x^{n-1} \quad (3.19)$$

Porém a base polinomial não está dentre os mais utilizados na abordagem aditiva generalizada. Existe uma tendência que com o aumento das dimensões do polinômio leva ao aumento da correlação entre as funções, resultando em aumento de variância nos resultados.

Um método bastante popular e que será usado neste estudo é o de curvas cúbicas. No modelo de curvas cúbicas são utilizado constantes que chamadas de *joelhos da função* $x_j^* : j = 1, 2, 3..m$. Enquanto na equação 3.18 cada um dos termos b_j representava uma um grau de potência , no novo modelo ele representará de acordo com a equação 3.20.

$$b_j(x) = |x_i - x_j^*|^3 \quad (3.20)$$

Sendo i o número de variáveis explicativas e j o número de *joelhos* escolhidos para representar a função. Como podemos imaginar, o posicionamento dessas constantes é relevante para a determinação do modelo. Na figura 3.1 um exemplo do posicionamento e sua importância.

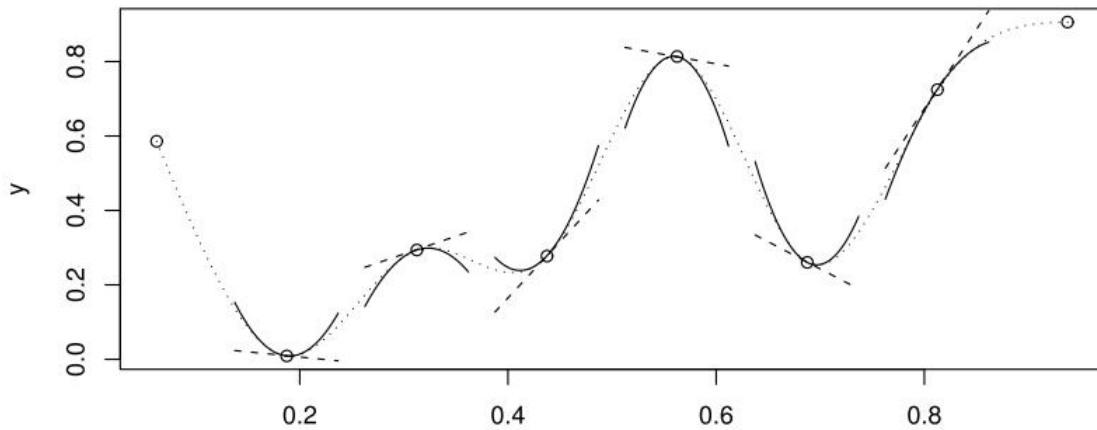


Figura 3.1: Exemplo de regressão cúbica com os *joelhos* em destaque.

WOOD & AUGUSTIN [71]

Normalmente modelos iterativos usam um grande número de *joelhos* para a função, quanto maior a quantidade menor o impacto da localização. Porém a utilização irrestrita de pontos de retorno pode levar a um sobreajuste dos dados, com

a função se tornando excessivamente curvilínea.

Assim como regressões lineares a modelagem de funções aditivas buscam minimizar o erro quadrático entre os resultados previstos pelo modelo e os valores da variável dependente observados. Para evitar um sobreajuste dos dados no modelo estudado nessa seção é adicionada uma penalidade a função de minimização da equação 3.21.

$$\sum_{i=1}^n (f(x_i) - y_i)^2 + \lambda \int [f''(x)^2] dx \quad (3.21)$$

Podem ocorrer em modelos aditivos um sobre ajuste dos dados, gerando funções que realizam grandes oscilações. Para evitar funções demasiadamente curvas, é utilizada uma penalidade representada por $\lambda \int [f''(x)^2] dx$.

O valor da constante λ representa a magnitude da punição a curvatura, um valor muito alto tendencia aproxima o modelo de uma reta, enquanto um valor baixo não oferece resistência as curvas. Há portanto um custo de oportunidade entre curvatura e aproximação dos pontos no modelo.

O valor do λ utilizado é obtido de forma empírica, geralmente por validação cruzada dos dados. Sendo o valor escolhido, aquele que minimiza o erro quadrático dos dados que foram separados inicialmente para validação.

3.3.2 Redes Neurais

A rede neural é um método de aprendizado de máquina, pois o modelo se ajusta conforme seus próprios dados através de iterações, assim como sistemas biológicos. É possível estabelecer um modelo que adote um comportamento a partir dos exemplos de entrada, tendo como referência os dados de saída reais.

De forma a modelar um processo de redes neurais mais preciso, em que o resultados encontrados condigam com os resultados reais, foi desenvolvido o procedimento iterativo denominado *back-propagation*, que ajusta os pesos das relações neurais a partir do erro encontrado . O erro nesse caso é a diferença entre os resultados originais e os obtidos pelo modelo. O objetivo é encontrar uma metodologia eficiente e generalista, que faça com que as relações se modifiquem de forma natural a partir vetores que mostram os resultados reais [60].

Os métodos desse tipo em geral se dividem em supervisionados e não supervisionados. Um método supervisionado é aquele em que o resultado obtido por cada iteração pode ser comparado com o valor esperado real, em geral padrões em que há resultado conhecido são utilizados para treinar o modelo, que depois pode ser testado em padrões desconhecidos ou utilizado para classifica-los. Por outro lado, métodos não supervisionados são aqueles em que não padrão de saída conhecido,

nesse caso, a modelagem pode servir como um método de agrupamento [21].

A rede neural é composta por camadas. A primeira cujo número de células é igual ao número de parâmetros de entrada, a última, que corresponde ao número de saídas do modelo, e as camadas intermediárias que também são chamadas de camadas ocultas. Uma das grandes dificuldades é que as metodologias de redes neurais utilizam neurônios intermediários para cálculos, aumentando o número de relações que devem ser consideradas ao se readequar os pesos [60].

A capacidade preditiva dos métodos de redes neurais é conhecida [54][34]. Alguns autores se preocuparam em desenvolver métodos para avaliar a importância relativa das variáveis no resultado. Para as ciências biológicas em geral é mais importante entender a correlação dos fenômenos observados do que sua capacidade de prever resultados a partir de padrões de entrada. A discussão e utilização dos métodos de análise da rede é um desafio que pode render bons resultados.

Modelagem de uma Rede Neural

A modelagem da rede neural deste estudo teve como base o estudo de FARACO *et al.*[25].

A figura 3.2 representa uma unidade neural, esse é o formato mais simples para o modelo. Esta é composta por entradas v_1, v_2, \dots, v_n , a unidade de processamento $\sum i$, uma função de ativação linear f_i e uma saída O_i . Cada uma das j entradas recebe um valor de entrada v_{ij} , que é ponderado por um peso w_{ij} , esses pesos podem ser excitatórios ou inibitórios. A unidade de processamento agrega todas as entradas e adiciona um valor limiar de θ_i , que adiciona um desvio estocástico ao resultado, que é então processado pela função de ativação f_i , responsável pela saída dos dados O_i .

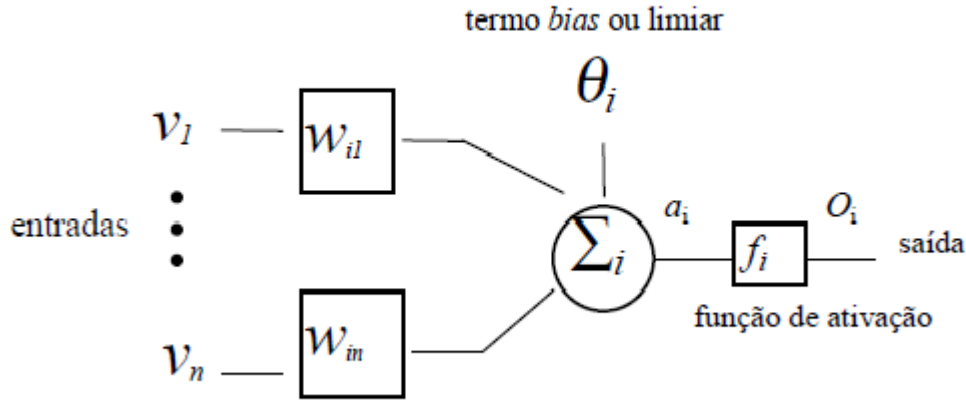


Figura 3.2: Unidade Neural Básica.

Fonte: FARACO *et al.*[25]

Em uma modelagem de redes neurais o peso das conexões entre os neurônios é a chave para entender as relações entre entradas e saídas. As contribuições relativas de cada uma das variáveis dependem impreterivelmente das conexões a qual estão relacionadas. Quanto maior o peso da conexão, maior a força explicativa da variável. Na figura 3.3 as cores e larguras das relações indicam a importância de cada uma das variáveis explicativas sobre os neurônios intermediários e destes com o resultado final.

A saída dos dados pode também ser descrita como nas equações 3.22 e 3.23. Sendo a_i a resposta do modelo de agregação de redes neurais, e também a entrada para a função de saída f_i .

$$O_i = f(a_i) \quad (3.22)$$

$$a_i = \sum_j w_{ij}v_j - \theta_i \quad (3.23)$$

A partir de um processo contínuo de entrada de dados, verificação dos resultados de saída e ajuste dos componentes internos, o modelo se desenvolveria até o ponto ser capaz de reconhecer os padrões e prever as saídas com uma diferença tolerável de ε em relação aos resultados reais.

O cálculo do erro do modelo é calculado através da equação 3.24, onde E é o erro do modelo, calculado a partir de $\sum_{i,p} E_i^p$, que representa a soma de todos os erros das i saídas possíveis para os p padrões existentes. Estes são obtidos após a comparação das saídas previstas pelo modelo O_i^p e o resultado real d_i^p .

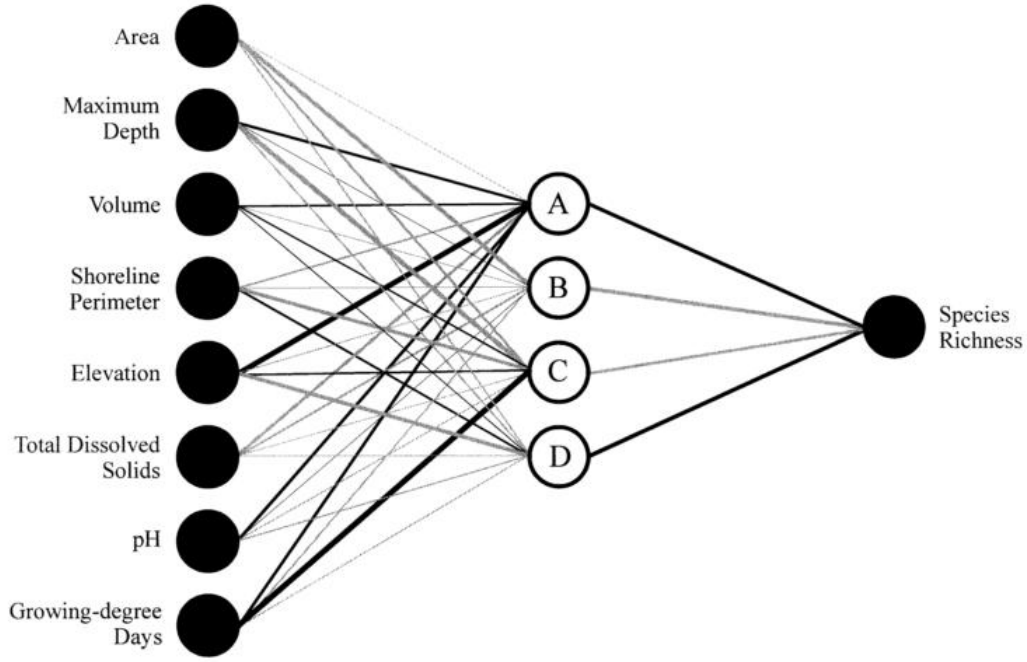


Figura 3.3: Exemplo de um Redes Neural completa.
Fonte:OLDEN *et al.*[54]

$$E = \sum_{i,p} E_i^p = \frac{1}{2} \sum_{i,p} (d_i^p - O_i^p)^2 \quad (3.24)$$

A saída de dados esperada, de acordo com o modelo é calculada pela equação 3.25. O_i é o resultado esperado para cada uma das i saídas para cada padrão p .

$$O_i^p = f(a_i^p) = f\left(\sum_j w_{ij}v_j^p - \theta_i\right) \quad (3.25)$$

O algoritmo *back-propagation* atua em cima dos pesos das conexões neurais w_{ij} , alterando o valor de cada uma, a partir do seu impacto no resultado. Quanto maior o impacto no erro, maior será a variação do peso, que aumenta em caso de diminuição, e sobe em caso de aumento. É utilizada uma constante γ , que é pode ser entendida como o passo que é dado em cada iteração. A variação do peso de cada conexão é definido pela equação 3.26.

$$\Delta w_{ij} = -\gamma \frac{\partial E_i^p}{\partial w_{ij}} \quad (3.26)$$

O tamanho do passo γ pode ter grande impacto no modelo, um valor muito alto pode impedir que se encontre um valor ótimo para os neurônios, enquanto um pequeno pode aumentar significativamente o esforço computacional. Para melhorar

o desempenho do modelo, o autor utiliza a técnica de *momentum* para modificar o tamanho do passo. A técnica consiste em aplicar um constante μ . Essa técnica altera o gradiente da iteração $t + 1$, adicionando uma parcela da modificação na iteração anterior, de acordo com a equação 3.27 .

$$\delta w_{ij}(t + 1) = -\left(\gamma \frac{\partial E_i^p}{\partial w_i}\right) + \mu \Delta w_{ij}(t) \quad (3.27)$$

Outro fator de preocupação, é que o mínimo local de uma variável, ou seja, aquele peso que garante seu peso ótimo para o modelo esteja distante de seu peso ótimo global. A experiência mostrou que dificilmente os resultados se estagnarão em um ponto distante do mínimo global, a não ser em casos em que o modelo se tornou demasiado complexo, e que existe um número muito extenso de conexões [60].

Para evitar que a rede neural se torne enviesada para determinada amostra, de forma que se afaste da sua capacidade de prever novos resultados, costuma-se dividir os dados de entrada em dois grupos, um de treino e um para controle. Após a utilização dos dados de treinamento para *calibragem* dos pesos das conexões, o modelo é testado com o segundo grupo de dados, separados do anterior chamado de dados de validação. É no grupo de controle que se analisa a real capacidade do modelo.

Símbolo	Nome
n	Variáveis de entrada
j	Número de neurônios intermediários
γ	Passo
μ	<i>Momentum</i>
-	Número de Iterações
-	Fração dos dados para Treinamento
E_{min}	Erro Mínimo

Tabela 3.1: Variáveis a serem definidas em um modelo de Redes Neurais.

Fonte: Elaboração Própria.

As variáveis a serem consideradas na elaboração do modelo estão dispostas na tabela 3.1. As variáveis ótimas são obtidas de forma empírica com testes sendo realizados para cada uma, os melhores resultados são selecionados para serem utilizadas no modelo final.

Com os resultados dos novos dados satisfatórios as próximas etapas buscam entender a contribuição das variáveis explicativa e a influência de cada uma sobre a variável dependente. Os métodos selecionados para tal missão são expostos nas próximas subseções.

Participação dos Pesos

Com a utilização crescente de redes neurais para modelagem de dados nasce a demanda por métodos que explicassem a participação de cada variável descritiva.

A análise de participação dos pesos em uma rede neural foi proposto inicialmente por GARSON [33], e depois aprimorado por GOH[35], sendo esta versão do método a que é mais comumente utilizada nos estudos científicos.

O modelo de redes neurais de três camadas possui duas matrizes de pesos, a primeira faz a ligação dos dados de entrada com a camada dos neurônios, e a segunda relacionada essa camada intermediária com o resultado esperado. Os pesos das relações podem ser positivos ou negativos, indicam o tipo de correlação existente. Quando maior o peso absoluto de cada uma das relações, maior é a influência sobre o resultado final.

O método proposto por GOH[35] possui as seguinte etapas de elaboração. As matrizes de ligação da primeira com a segunda camada, e dessa camada intermediária serão denominadas de I e O , respectivamente. Essas informações são obtidas diretamente do modelo de redes neurais, após o término da etapa de treino.

1) Obter o módulo dos valores presentes na matrizes I e O .

2) Calcular os pesos finais de cada uma das variáveis de entrada para o resultado final, multiplicando o valor das ligações dos dados de entrada com o valor da conexão desse neurônio intermediário com o resultado final. Exemplo na tabela 3.2.

P	Entrada1	Entrada2	Entrada3
Neurônio 1	$i_{11} * o_1$	$i_{12} * o_1$	$i_{13} * o_1$
Neurônio 2	$i_{21} * o_2$	$i_{22} * o_2$	$i_{13} * o_2$
Neurônio 3	$i_{31} * o_3$	$i_{32} * o_3$	$i_{13} * o_3$
Neurônio 4	$i_{41} * o_4$	$i_{42} * o_1$	$i_{43} * o_4$

Tabela 3.2: Formulação da Matriz P .

Fonte:GOH[35] adaptado.

3) A partir da matriz P calcular os pesos relativos de cada uma das variáveis de entrada em cada neurônio para obter a matriz Q . Exemplo para o neurônio 1, $Q_{11} = P_{11}/(P_{11} + P_{12} + P_{13})$.

4) Calcular a soma das colunas, obtendo o vetor S que representa a soma de cada entrada para o resultado final. Exemplo na tabela 3.3.

P	Entrada1	Entrada2	Entrada3
Neurônio 1	q_{11}	q_{12}	q_{13}
Neurônio 2	q_{21}	q_{22}	q_{23}
Neurônio 3	q_{31}	q_{32}	q_{33}
Neurônio 4	q_{41}	q_{42}	q_{43}
Soma	$S_1 = \sum_{i=1}^4 q_{i1}$	$S_2 = \sum_{i=1}^4 q_{i2}$	$S_3 = \sum_{i=1}^4 q_{i3}$

Tabela 3.3: Matriz Q e soma das colunas.

Fonte:GOH[35] adaptado.

5) A última etapa é a normalização em 1 para a importância de cada uma das variáveis de entrada, um exemplo para a Entrada 1 $G_1 = S_1 * 100 / (S_1 + S_2 + S_3)$.

O resultado final do método proposto nesta subseção leva em consideração os valores absolutos dos pesos, portanto não permite avaliar que tipo de impacto cada uma das variáveis exerce sobre o resultado final. Para analisar de que forma os dados se relacionam é utilizado uma análise de sensibilidade.

3.3.3 Modelos de Árvore

Modelos de árvores de regressão são úteis para estudos ecológicos. Porém modelos de árvores simples podem apresentar algumas dificuldades em lidar com dados mais robustos. Ao invés de construir árvores muito ramificadas e portanto difíceis de interpretar, conclui-se que a agregação de diversas árvores levaria a melhores resultados de regressão.

A literatura possui diversos exemplos de modelos, como os métodos de árvores ensacadas e florestas aleatórias. Neste estudo será utilizado o método de árvores de crescimento progressivo, que mostrou bons resultados na literatura e já foi utilizado para os recifes de Abrolhos em CONI *et al.*[14].

Árvores de Regressão simples

A construção de uma árvore de regressão e sua aplicação para dados ecológicas é baseada em DE'ATH & FABRICIOUS [18].

Uma árvore de regressão é construída através da divisão sistemática do grupo de dados original. As divisões são feitas através de um único critério, baseado em uma única variável explicativa. O critério escolhido para divisão deve tornar ambos os grupos homogêneos entre si na melhor medida possível.

A partir da primeira separação, os novos grupos formados devem passar por novas divisões. Cada um dos grupos é caracterizado pelas médias e desvios dos valores encontrados. O objetivo de cada nova divisão é minimizar a impureza dentro de cada

grupo, uma das formas mais usadas para calcular essa medida é a soma da variância em relação a média. Um exemplo de árvore de regressão pode ser encontrado na figura 3.4.

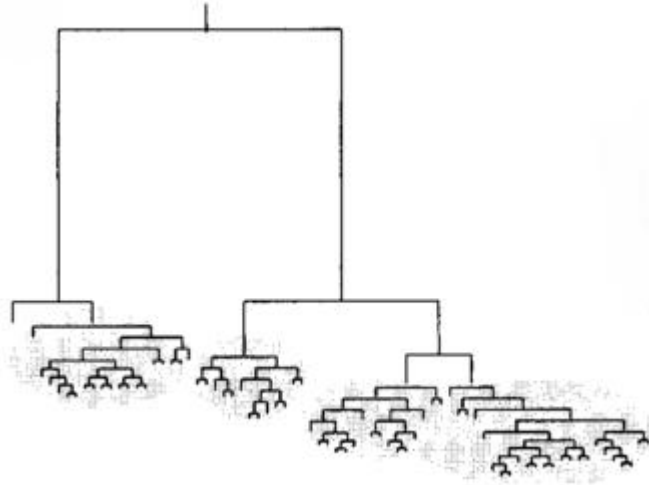


Figura 3.4: Árvore de Regressão, devido as muitas ramificações pode se tornar difícil de interpretar.

Fonte:DE'ATH & FABRICIOUS [18]

A medida de impureza de um grupo pode ser calculada de diversas formas em uma árvore de regressão, sendo duas delas as formas mais comuns. A primeira é através da soma dos quadrados, similar ao método dos mínimos quadrados da regressão linear. A segunda é pelo módulos dos desvios sobre a média dos grupos, esse método é mais robusto, porém pode causar distorções quando a média é muito próxima de zero.

O tamanho das árvores é definido pelo número de grupos finais. Árvores muito grandes podem ter maior capacidade de explicar os resultados, porém também dificultam a interpretação dos resultados.

Uma forma natural de se pensar para determinar o tamanho de uma árvore poderia ser através de um coeficiente de melhora. A partir do momento em que o crescimento da árvore não gerasse um aumento no coeficiente de explicação maior que a medida pré-estabelecido, a árvore estaria finalizada.

BREIMAN *et al* cita duas desvantagens desse raciocínio. Caso o coeficiente de corte seja muito baixo, o resultado será uma árvore muito grande. Caso esse coeficiente seja grande, o potencial da árvore não será plenamente atingido.

Para resolver esse problema, existem três passos a seguir para a construção de uma árvore, de acordo com BREIMAN *et al*. O primeiro passo é a construção de árvores excepcionalmente grande.

O segundo passo é conhecido como *poda* da árvore. Ao invés de parar um pro-

cesso sem saber qual o tamanho ótimo, uma árvore sobre dimensionada é produzida, a partir daí, galhos que não colaborarem significativamente para a melhora de desempenho são excluídos.

O terceiro e último passo é a seleção de um modelo ótimo através de validação cruzada. Os melhores modelos selecionados no passo dois para cada tamanho tem sua capacidade de análise medida através de dados não utilizados inicialmente.

Árvores de crescimento gradativo

O método para construção de modelos utilizando agrupamento de árvores foi baseado em DE'ATH[19] e ELITH *et al.*[23].

Os modelos de árvore simples tem como fraquezas a capacidade de fazer previsões de resultados e dificuldade de interpretação, principalmente no caso de árvores com mais ramos. Para superar isso foram desenvolvidos modelos de florestas, que reúnem um grande número de submodelos. A ideia é que muitas árvores menores são mais eficientes para previsão e interpretação do que uma árvore muito grande.

O primeiro método aditivo de árvores surgiu em 1996 com o software chamado AdaBoost de FREUND & SCHAPHIRE[32]. A lógica por trás de modelos aditivos de árvores é bastante similar a de regressões aditivas vistos na seção.

Cada uma das amostras é classificado pela sequência de árvores, formando assim os pesos de cada uma delas para o resultado final, quanto maior o índice de erro de uma determinada amostra, maior o peso que será dado aquela árvore que acertar este corretamente. Matematicamente os modelos podem ser definidos pela equação 3.28.

$$f(x) = \sum_m f_m(x) = \sum_m \beta_m b(x; \gamma_m) \quad (3.28)$$

Onde $f_m(x)$ representa o resultado de cada uma das m árvores, β_m o peso de cada nó em cada árvore e γ_m os parâmetros utilizados na elaboração de cada modelo individual. Assim a partir da definição de duas variáveis.

No modelo proposto para construção das árvores de crescimento progressivo as árvores vão sendo adicionadas ao modelo, e iterativamente os pesos de cada árvore são reajustados, de forma a minimizar a diferença entre o o resultado do modelo e a realidade dos dados $\delta L(y, f(x))$. Os passos para obtenção do modelo final estão a seguir, a variável n representa o número de árvores totais que compõe o modelo agregado.

- 1) Comece com $f(x) = 0$
- 2) Para m variando de 1 até n :
 - a) Calcular a função de resíduos entre a função f e o resultado real $r = -([\delta L(y, f(x))]/[\delta f(x)])$

b) Modelar uma nova árvore de regressão para obter o coeficiente y_m para $\beta(x; \gamma_m)$

c) Estimar os β_m para a função após a inclusão da nova árvore, minimizando a função $L(y, f_{m-1}(x) + \beta(x; \gamma_m))$

d) Atualizar a função $f_m(x) = f_{m-1}(x) + \beta_m(x; \gamma_m)$

3) Atualize a função $f(x) = \sum_m f_m(x)$

Assim a partir dos passos acima é feito um processo iterativo que determina as árvores utilizadas no modelo final e o peso relativo de cada uma delas. Faltam apenas definir o número total de árvores no conjunto, representado pela variável n .

A variável n deve representar o tamanho do conjunto que melhor se adequa ao grupo de dados estudado. Curiosamente um número maior de modelos simples não necessariamente implica em um conjunto melhor para os dados de validação. O motivo disso é que um número muito grande de árvores pode tornar o modelo final sobreajustado aos dados de treinamento. Assim o número n ideal deve ser testado empiricamente, em busca do melhor resultado. O comportamento explicado pode ser observado na figura 3.5.

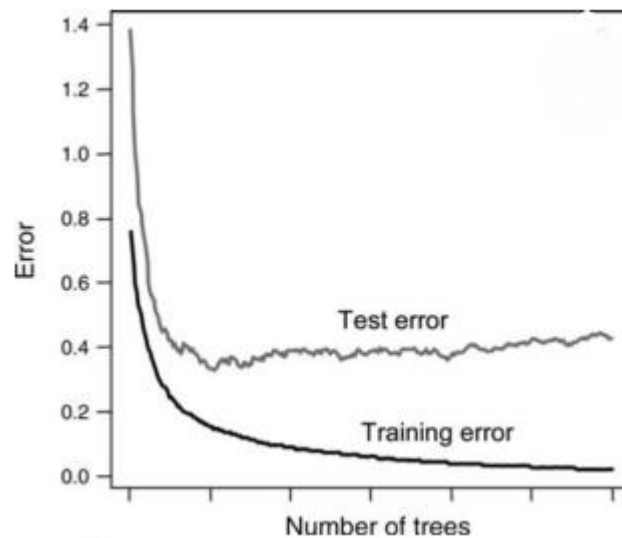


Figura 3.5: Comportamento do erro em relação ao número de árvores. Para os dados de treinamento há uma expectativa de decréscimo, já para os dados de validação, após alcançar um mínimo há um comportamento oscilante.

Fonte: DE'ATH [19]

O modelo de árvores de crescimento progressivo proposto é eficiente para análise dos dados ecológicos. Tem potencial de previsão superior ao das árvores de regressão singulares, e não oferece a dificuldade de interpretar resultado desses métodos.

As principais ferramentas oferecidas pelo pacote gbm disponível para R são:

i) Importância relativa para o modelo de cada variável.

ii) Gráficos de dependência parcial. Funcionam como uma análise de sensibilidade, correlacionam variações na variável dependente a partir de um intervalo de cada variável explicativa. Também podem ser utilizados em três dimensões, com a variável resposta atendendo a intervalos de duas variáveis explicativas.

Em relação a outros métodos de aprendizado de máquina, como as florestas aleatórias e árvores ensacadas, o desempenho foi superior principalmente para regressões. Em comparação com as redes neurais, DE'ATH[19] argumenta que os resultados são mais interpretáveis.

Apesar de todas as vantagens dos modelos de florestas, há uma ressalva quanto a capacidade de modelar relações. Caso as correlações entre as variáveis explicativas e dependentes sejam lineares, ou algo próximo a isso, os resultados serão inferiores aos encontrados em modelos de regressão tradicionais.

Capítulo 4

Experimentos Computacionais

4.1 Dados

Extrair informações relevantes de amostras obtidas do oceano é um desafio considerável. É custosa a obtenção de amostras que podem ter extraídos dados, e são muitas as variáveis de influência para o modelo final. Antes da viagem, já devem estar definidos que tipo de dados serão coletados, como será feita a amostragem, e a forma de organização dos dados brutos [67].

Boas práticas no manejo das informações são fundamentais estudos relacionados ao oceano. Permitem a integração, comparação e reutilização dos dados recolhidas. Recomendações devem ser seguidas para garantir a confiabilidade e a melhor extração possível de informações [67].

É importante que as informações obtidas por meio dos métodos sejam confiáveis, para que essa condição seja atendida é importante garantir o máximo de qualidade na obtenção dos dados [69].

A expedição que coletou os dados utilizados neste estudo ocorreu entre os dias 11 a 19 de setembro de 2016. O objetivo era analisar os impactos causados pelo desastre de Mariana, tanto no Rio Doce quanto no ambiente marinho.

Percorrendo aproximadamente 1900 km, foram totalizados 9 pontos amostrados entre a cidade de Mariana e o mar, sendo 5 pontos no Rio Doce, 2 pontos no estuário e 2 pontos no mar. A coleta no mar foi realizada em dois pontos que continham recifes coralinos nas proximidades da cidade de Barra Nova, na parte Sul do Banco de Abrolhos. A tabela 4.1 apresenta informações sobre os pontos coletados, a referência dos dados é o mapa apresentado por MAZZEI *et al.*[45], que está disposto na figura 2.8.

Ponto	Recife	Prof. Topo(m)	Prof. Fundo (m)	Distância da Costa(km)	Distância do Rio Doce (km)
Recife 1 Sul	R4	12	24	17,5	76
Recife 2 Norte	R6	13.5	23	28	81

Tabela 4.1: Recifes do Rio Doce utilizados para coleta e informações sobre os locais.
Fonte:Relatório COPPE 20275.

Bentos é a composição de indivíduos em uma superfície recife. Compreende todos os organismos visíveis que disputam esse espaço. Dentre as principais grupos desse ambiente no banco de Abrolhos se destacam espécies de corais, algas, esponjas e cianobactérias [45].

A identificação de cada espécie presente na formação foi feita através da metodologia de foto-quadrados, que consiste na obtenção de imagens de alta resolução em pontos aleatórios dos habitats observados [31].

Cada uma das fotos representa uma amostra, que é composta por um mosaico formado por 15 quadrados. Dentro de cada uma dos quadrados, são selecionados 20 pontos, que são analisados pelo programa *Coral Point Count Extention*, que pode ser visto na figura 4.1. O objetivo do processo é identificar os organismos bentônicos e qual a fração do espaço ocupada por cada grupo de espécies.

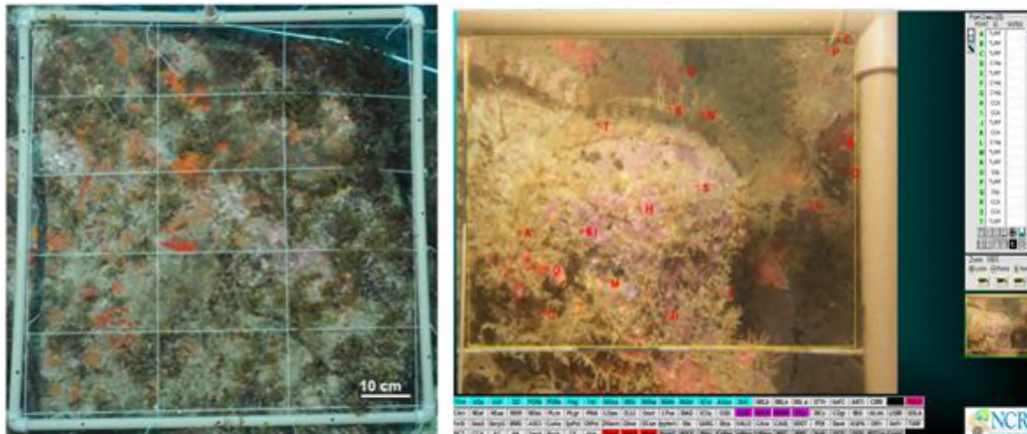


Figura 4.1: Programa *Coral Count Extention*.
Fonte:Relatório COPPE 20275.

Para efeito de comparação da comunidade recifal da foz do Rio Doce serão utilizados outros do banco de Abrolhos, expostos na tabela 4.2. A coluna Indicativo se refere as referências encontrados no mapa de Abrolhos visto na figura 2.6, como os corais do Espírito Santo não foram mapeados na figura em questão, estes não possuem menção na tabela.

Indicativo	Local	Política de Preservação	Pontos de Coleta	Anos de Coleta
-	Espírito Santo	Desprotegido	2	2016, 2017
B	Timbebas	Protegido	3	2006-2008, 2014
D	Parcel de Abrolhos	Protegido	5	2006-2008, 2013-2014
E	Paredes	Desprotegido	4	2006-2008, 2013-2014

Tabela 4.2: Locais usados para comparação do bentos, com a política de preservação adotada, pontos de coleta e anos em que foram realizadas.

Fonte:Elaboração Própria

Os locais selecionados para comparação tiveram coletas em diversos pontos, em diferentes anos. Para determinadas análise é utilizada apenas a coleta mais recente, pela proximidade temporal com os dados obtidos no Espírito Santo, para outras, todo o volume de informações é utilizado, será explicitado a abrangência temporal para cada método.

Também são utilizados dados obtidos através de sensoriamento remoto, a partir da plataforma MODIS. As variáveis coletadas estão resumidas na tabela 4.3.

Símbolo	Varáveis Coletadas por Sensoriamento Remoto
CHLA	Concentração de clorofila A
PAR	Radiação fotossintética
POC	Concentração de carbono orgânico
NSST	Temperatura noturna na superfície
KD	Coefficiente de difusão 490 nm (Turbidez)

Tabela 4.3: Resumo dos Dados coletados pelo MODIS.

Fonte: Elaboração Própria

Variáveis físico-químicas que possam exercer influência sobre os recifes foram acrescentadas aos modelos de matemáticos. Como as coletas de bentos ocorreram em meses diferentes, a informação de sensoriamento remoto foi utilizada calculando-se a média dos últimos doze meses anteriores a coleta.

Portanto a análise da cobertura de corais da foz do Rio Doce contará com três tipos de dados; a informações de cobertura do bentos, as características físicas do local, e as informações físico-químicas coletadas a partir de sensoriamento remoto. Cada modelo visto nas próximas seções utilizará parte das variáveis.

4.2 Desastre de Mariana e os recifes de Conceição da Barra

O desastre de Mariana se propagou por toda a extensão restante do Rio Doce até o mar, ligando o sinal de alerta sobre a biodiversidade da região. Os impactos podem ser notados visualmente como pode ser observado na figura 4.2



Figura 4.2: Imagens do desastre no Rio Doce.

Fonte: Relatório COPPE-20275.

Imagens com provável transporte da lama para o norte da desembocadura, baseado em informações sobre o vento, foram submetidos à análise de agrupamento. A pluma foi identificada como a área CL5, que está em cinza figura 4.3 os ventos em direção noroeste da região levaram os resíduos da foz do Rio Doce até a localização dos recifes, que estão marcados com pequenas cruces, próximos a região azul-clara da imagem.

A excursão chegou no local dos recifes aproximadamente 3 meses após a chegada de resíduos do desastre de Mariana. Em relação a excursão anterior em 2014 [45] é

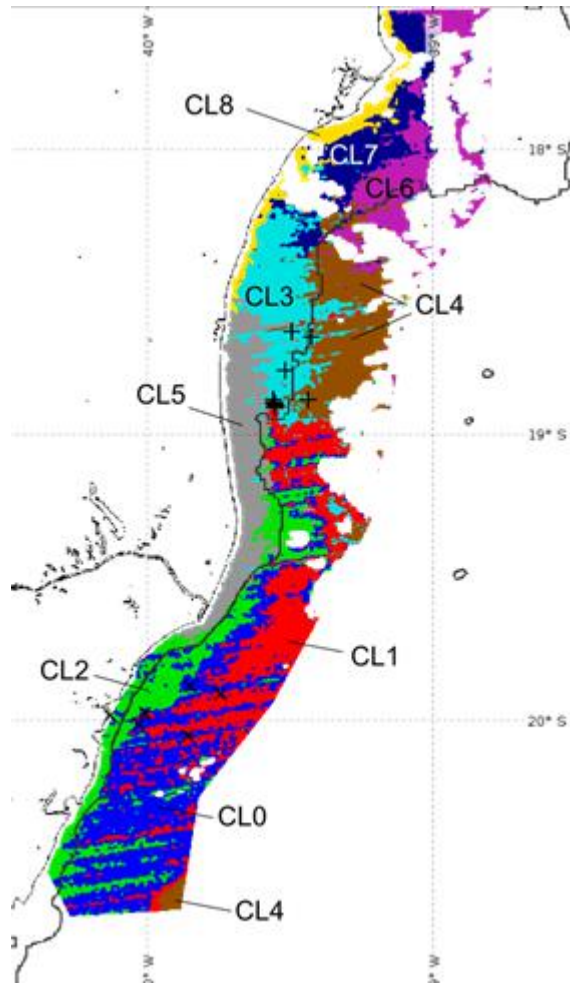


Figura 4.3: Imagem MODIS 16 de junho de 2016.
Fonte: Relatório COPPE-20275.

possível perceber uma diminuição das espécies identificadas de peixes e de famílias. Essa discrepância foi mais sensível nos corais localizados ao sul, como visto na figura 4.4

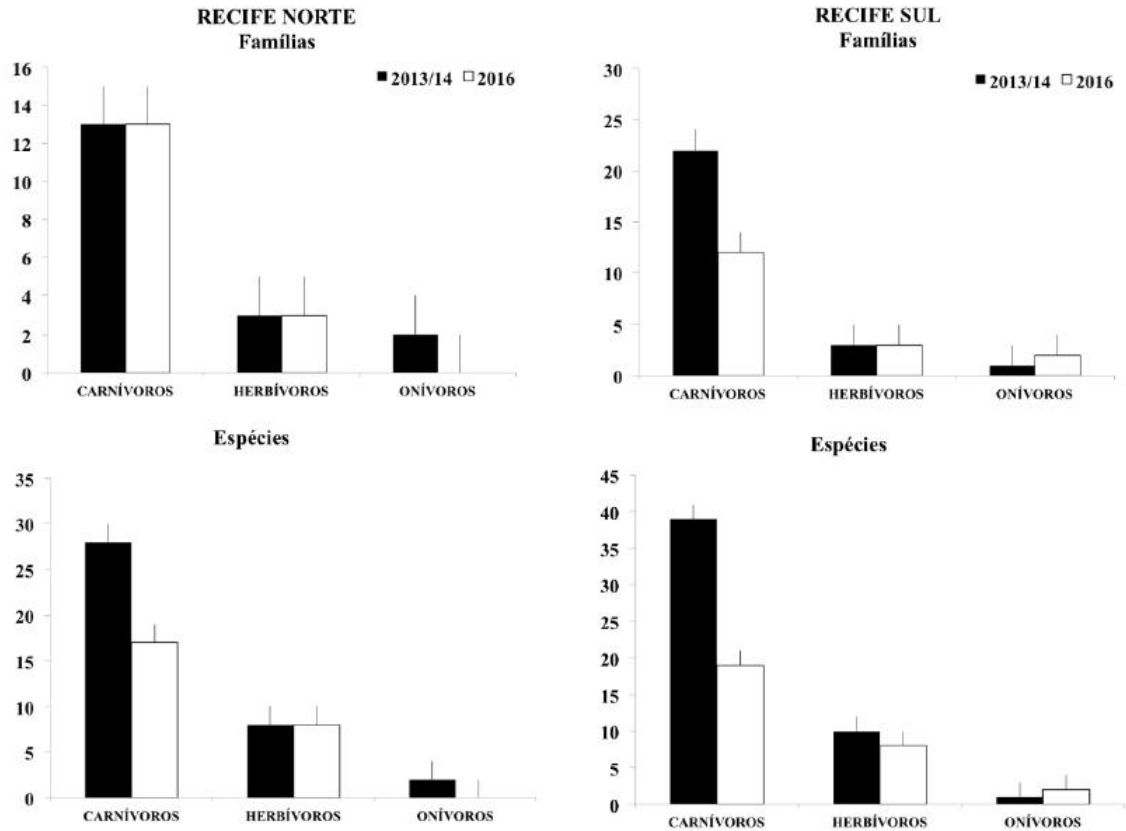


Figura 4.4: Comparação da análise de peixes antes e depois do desastre de Mariana, por família e por espécie.

Fonte: Relatório COPPE-20275.

O estudo foi conduzido na porção sul do Banco dos Abrolhos, litoral norte do Espírito Santo. Os recifes da região são em geral formados por algas calcárias e corais construtores, similares aos pináculos-chapeirões, típico do norte do Banco dos Abrolhos, Bahia [45].

Na figura 4.5 uma foto de um recife composto principalmente por *Mussimilia braziliensis*. Algumas das espécies estão identificadas na figura 4.6.



Figura 4.5: Corais da região da foz do Rio Doce.
Fonte: Relatório COPPE-20275.

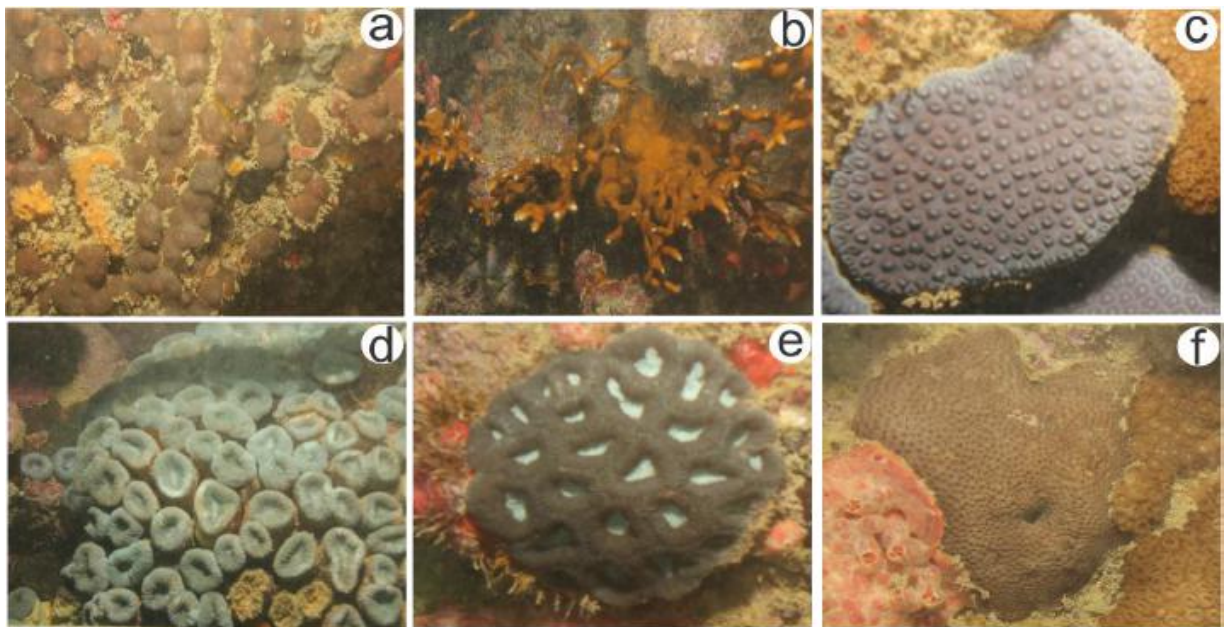


Figura 4.6: Algumas espécies encontradas nos recifes da foz do Rio Doce.
a) *Madracis decactis*, b) hidrocoral *Millepora alcicornis*, c) *Montastrea cavernosa*,
d) *Mussismilia hartii*, e) *Mussismilia hispida*, f) *Siderastrea stellata*. Fonte:
Relatório COPPE-20275.

Os recifes recém-descobertos foram analisados por meio da caracterização bentônica da região neste estudo. Para o cálculo dos índices de diversidade, gráficos bivariados, e gráficos de barra foi realizada utilizando o software R versão 3.2.1. O pacote *vegan* auxiliou as análises de ordenamento, enquanto os gráficos foram elaborados com o suporte fornecido pelo pacote *ggplot2*.

Também no software R foram implementadas as regressões lineares múltiplas, os modelos aditivos e a árvore de crescimento progressivo, utilizando respectivamente o pacotes básico, *mgcv* e *gbm*. Para os gráficos foram utilizados ferramentas dos mesmos pacotes.

O método de Redes Neurais foi implementado no software *MATLAB* versão R2017a, com o auxílio do pacote *NNTools*. O particionamento dos pesos e a análise de sensibilidade foram realizados por códigos no próprio software.

Como apontado por estudos da região o tipo de habitat tem grande influência sobre a formação do bentos. Por essa razão as amostras foram divididas entre topo e parede, conforme a tabela 4.4.

Hab	Habitat
PA	Parede
TP	Topo

Tabela 4.4: Tipos de Habitat no Recifes Analisados.

Fonte: Elaboração Própria.

A análise dos componentes principais foi utilizada para avaliar as diferenças da formação entre os dois pontos de coleta dos recifes da foz do Rio Doce. O resultado está na figura 4.7.

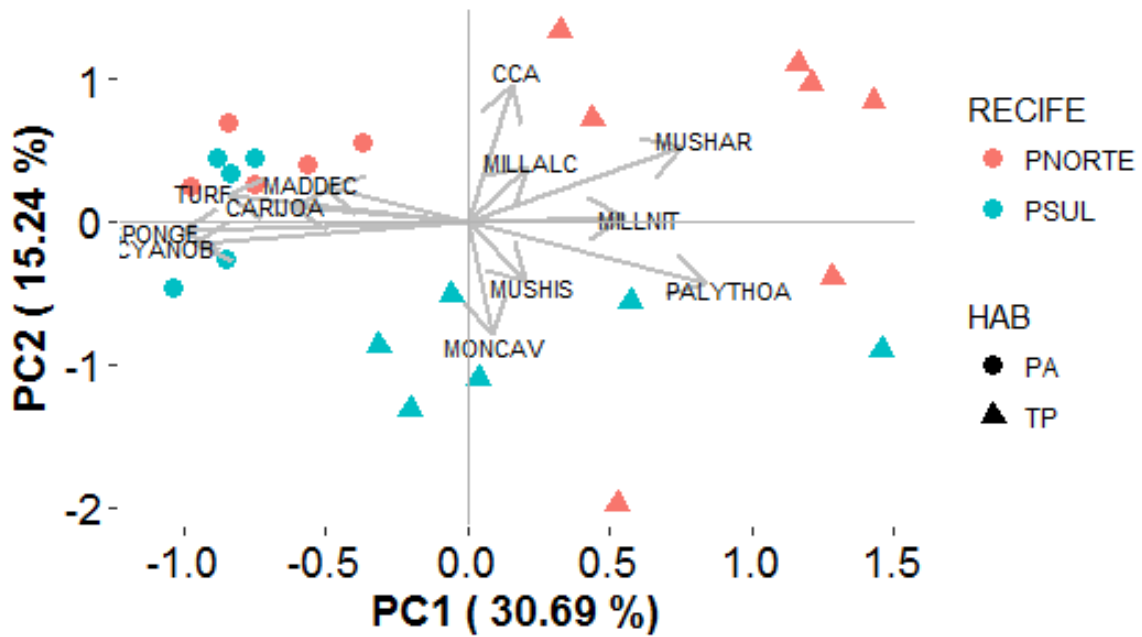


Figura 4.7: Análise APC dos recifes da foz do Rio Doce, comparação entre recifes norte e sul e por tipo de habitat.

Fonte: Elaboração Própria.

Os resultados indicam uma separação clara entre os tipos de habitat. No ambiente das paredes dos recifes há uma maior concentração de cianobactérias, esponjas, e algas.

Em relação aos topos das formações, há uma maior cobertura de espécies corálinas. Os recifes do sul tiveram maior abundância dos corais do tipo *Mussismilia hispida* e *Montastrea cavernosa*. Nos corais do norte foi detectada uma maior cobertura das espécies *Mussismilia hartii* e da alga calcária *Crustose coralline algae*. Em ambos os ambientes do topo há uma maior incidência da espécie *Palythoa sp.*

Na figura 4.8 estão dispostos as espécies mais abundantes encontradas nos corais do Espírito Santo. Os dados estão agrupados pelo local de coleta, recifes do norte e do sul, e pelo tipo de habitat. A linha que sai do topo da coluna da espécie representa o desvio padrão do local.

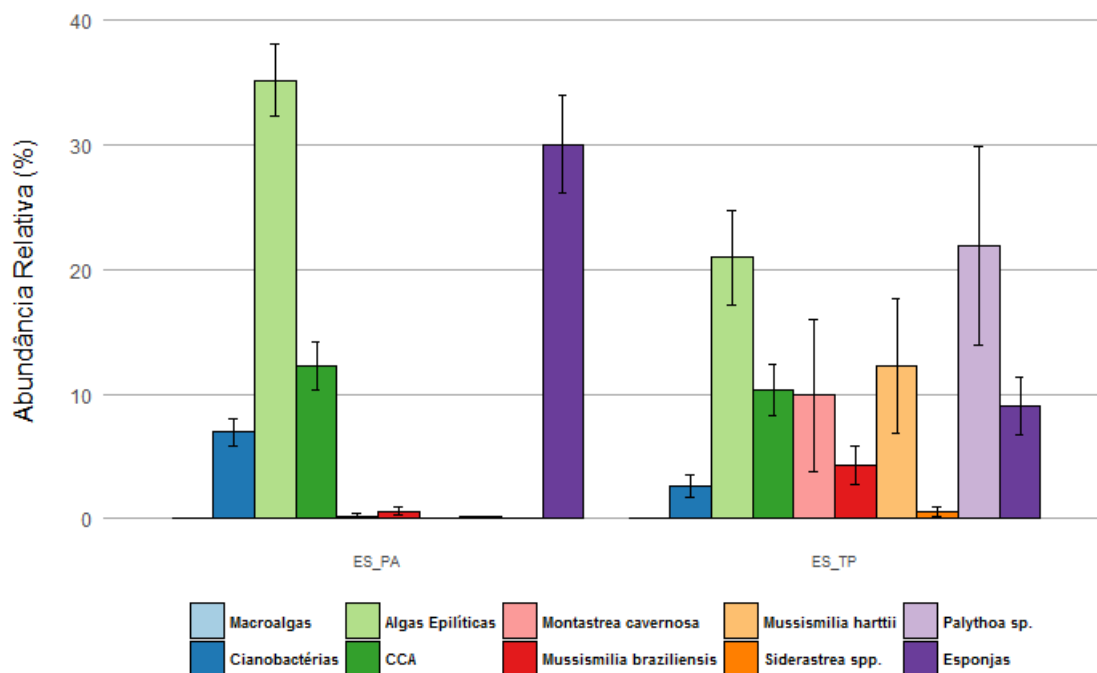


Figura 4.8: Abundância relativa das espécies nos corais da foz do Rio Doce.

Fonte: Elaboração Própria.

As algas foram a espécie mais predominante na parede do recife norte, enquanto no recife sul as esponjas foram o grupo mais abundante, com as algas em segundo. Aparentemente há mais algas e esponjas nas paredes dos recifes do que no topo. A espécie *Carijoa riisei* só foi encontrada nas paredes, chegando a cobrir um décimo desse território no recife norte.

As cianobactérias também estiveram mais presentes nas paredes dos recifes do que em seus respectivos topos. Sua presença foi maior nos recifes sul para os dois habitats.

No topo dos recifes foi observado uma maior presença das espécies coralíneas. As espécies mais dominantes foram *Mussismilia hartii*, *Montastrea cavernosa*, *Millepora alcicornis* e *Palythoa Caribaeorum* no recife norte. Nas formações do sul a espécie *Palythoa Caribaeorum* tem maior predominância, chegando a quase um terço da cobertura total do local, espécies de corais como *Montastrea cavernosa* e *Montastrea braziliensis* também aparecem com destaque. Há um decaimento do sul em relação ao norte para as espécies *Mussismilia hartii* e *Millepora alcicornis*, sendo que esta não foi detectada nos recifes mais próximos ao rio Doce.

Em 2017 uma nova excursão ao local extraiu novas informações sobre os corais de Conceição da Barra. É possível observar se há alguma deterioração causada pela ação de longo prazo da lama tóxica que chegou aos recifes. Para comparar a composição dos corais da foz do Rio Doce em ambos os anos são utilizados outros pontos anteriormente colhidos no banco de Abrolhos.

4.2.1 Diversidade de espécies

Para comparar a composição dos recifes da foz do Rio Doce são utilizados outros pontos anteriormente colhidos no banco de Abrolhos. Os anos de coletas analisados foram apenas os mais recentes dentre os disponíveis.

A figura 4.9 apresenta o número de espécies obtidos em cada amostra dos locais coletados. Cada ponto representa o valor encontrado para cada amostra. Os dados foram divididos de acordo com o recife onde foram coletados e de acordo com o tipo de habitat. O centro da barra simboliza a média das amostras, os extremos representam o primeiro quartil e o 3 quartil dos dados. As linhas de cada barra apresentam o valor máximo e o mínimo dentre amostras coletadas. Esse padrão é utilizado para interpretar os dados dos três índices utilizados.

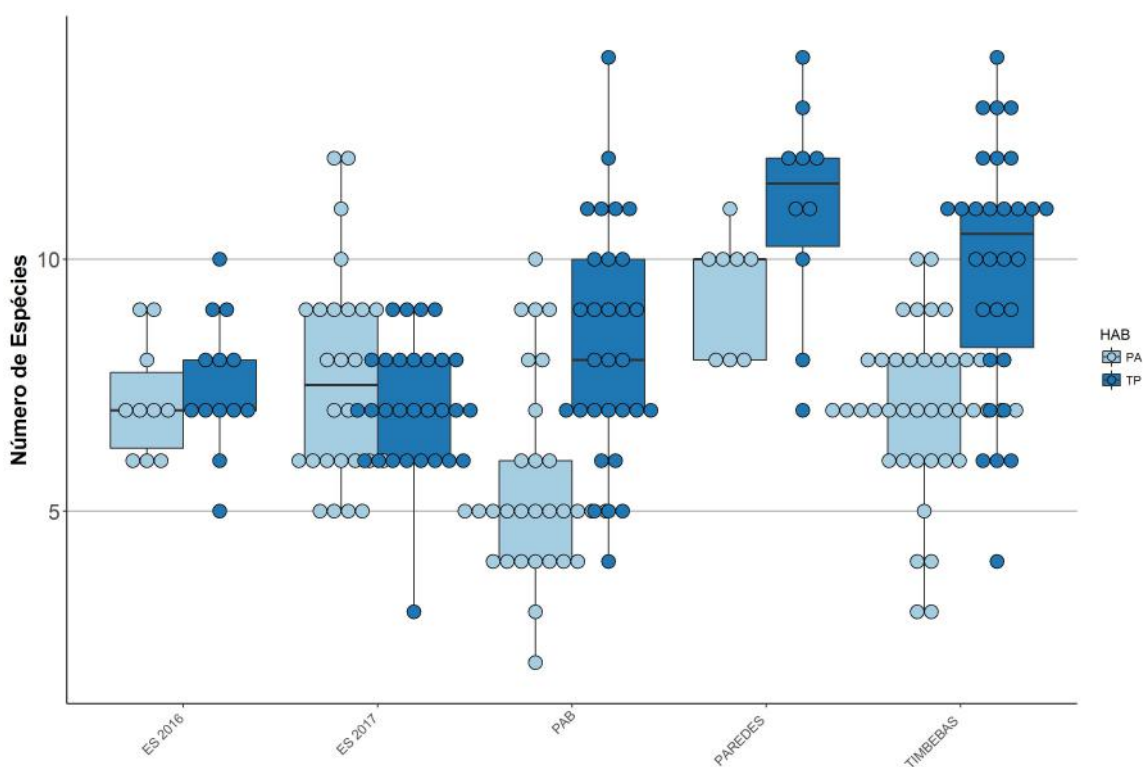


Figura 4.9: Número de espécies identificadas nos recifes de Abrolhos.

Fonte: Elaboração Própria.

O número de espécies encontradas no topo dos recifes foi superior ao número das paredes na maioria dos recifes observados. Tanto para o topo dos recifes quanto para a parede, os recifes localizados em Paredes apresentaram o maior número de espécies. Nos recifes da foz do Espírito Santo, ocorre a menor diferença entre os dois tipos de *habitat*, enquanto em Parcel de Abrolhos (PAB) há a maior discrepância dentre os locais amostrados.

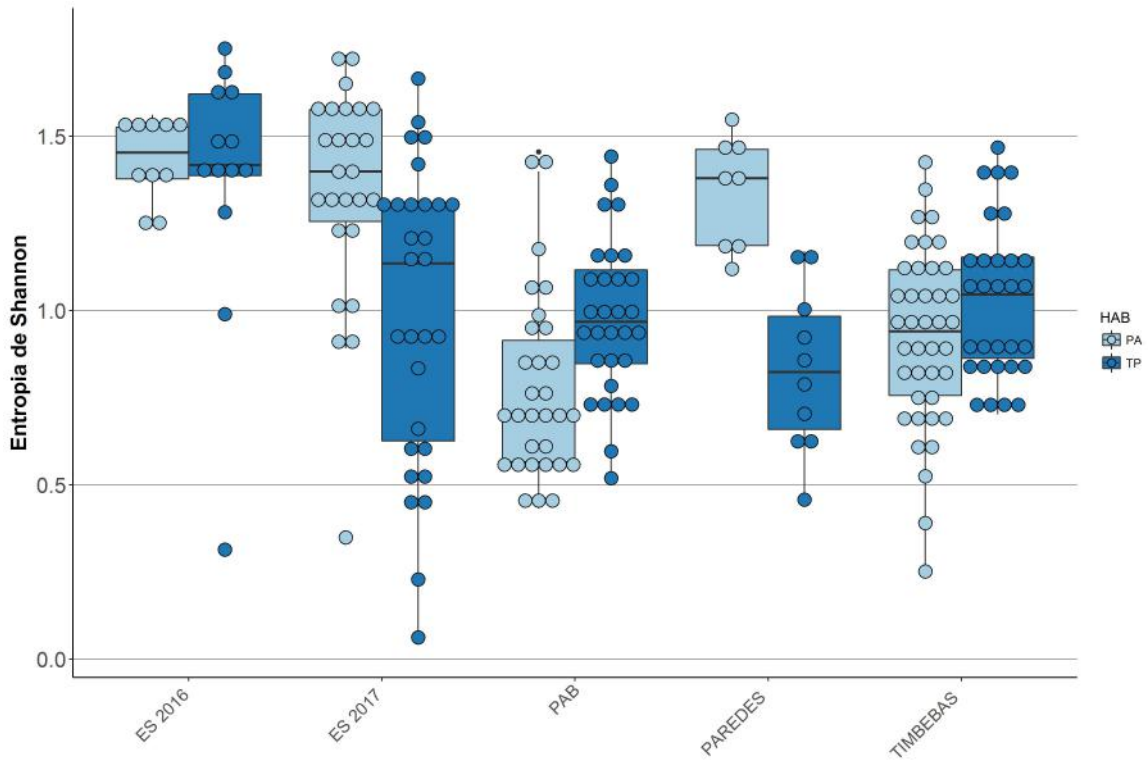


Figura 4.10: Índice de Entropia de Shannon dos recifes de Abrolhos.

Fonte: Elaboração Própria.

Na figura 4.10 é apresentado o índice de Entropia de Shannon. Os recifes do Espírito Santo apresentam o maior valor de entropia média, tanto para o topo quanto para a parede. Nos recifes da foz do Rio Doce foi detectado o maior desvio, com pontos acima de 1,5 e outros abaixo de 0,5. Parcel de Abrolhos apresentou resultados com características semelhantes a Timbebas, porém médias um pouco menores, enquanto Paredes apresentou um padrão diferente, com os valores das coletas nas paredes apresentando maior valor que no topo das formações.

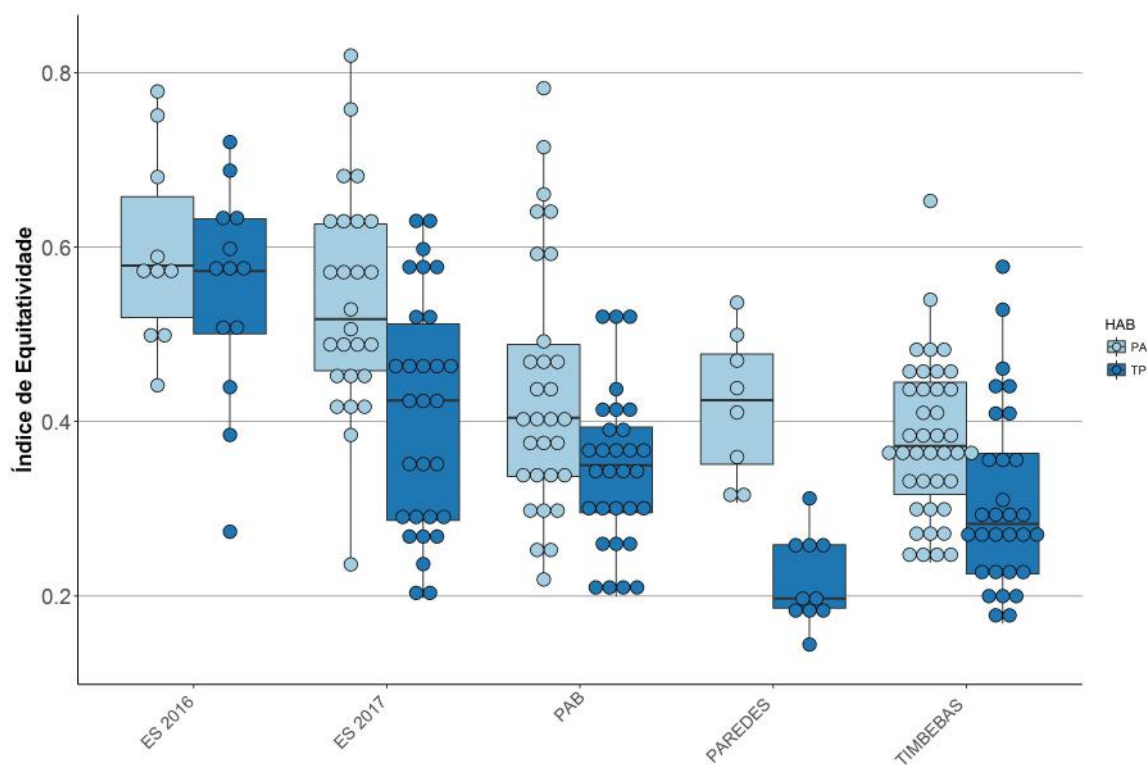


Figura 4.11: Índice de Equitatividade de Pielou dos recifes de Abrolhos.

Fonte: Elaboração Própria.

Na figura 4.11 está representado o índice de Equitatividade de Pielou. Um valor próximo de 1 (um) indica simetria perfeita na distribuição das espécies encontradas, enquanto o valor zero apresenta uma espécie dominando todo o ambiente. Há uma indicação de que as amostras obtidas nos recifes do Espírito Santo possuem uma distribuição de espécies mais equilibrada que nos outros pontos. Porém os resultados das amostras de 2017 foram menores em relação a 2016, embora ainda acima dos dados de outros recifes.

Em Parcel de Abrolhos e Timbebas as médias de equitatividade dos dois tipos de habitat foram próximas, em ambos os locais houve uma variância grande do índice. Em Paredes novamente o resultado entre os tipos de *habitats* apresentaram a maior variação dentre os locais avaliados

4.2.2 Análises de Ordenação

O gráfico de nMDS é observado na figura 4.12. A rotina do *software* utilizou dois tipos de transformação nos dados, primeiro foi realizado a dupla normalização de *Wisconsin*, e depois foram calculados as raízes quadradas dos resultados. No total foram realizadas vinte iterações para os dados, cada uma com um ordenamento inicial. O melhor resultado obteve o valor do índice $stress=0,2694338$.

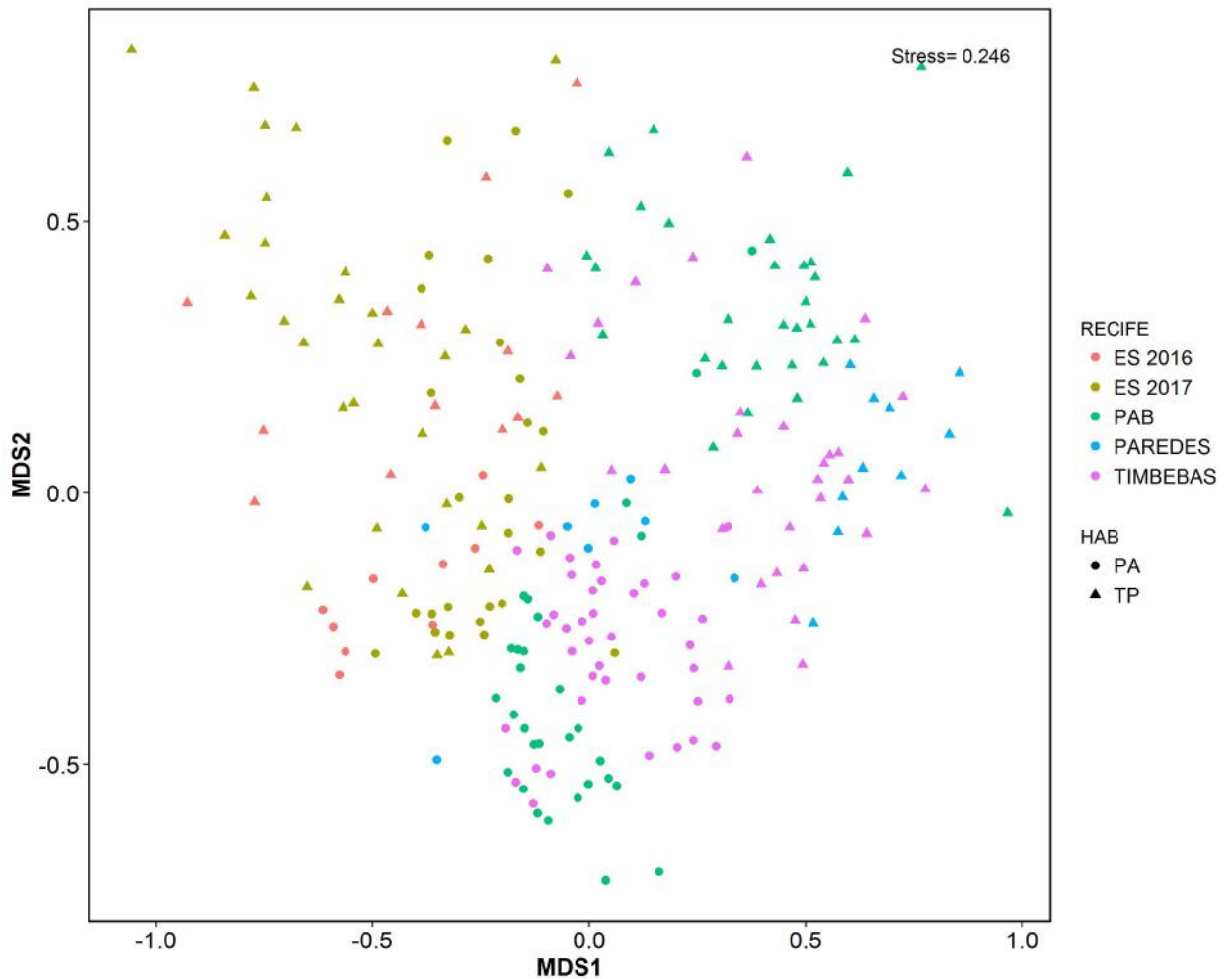


Figura 4.12: nMDS dos habitats recifais de Abrolhos.

Fonte: Elaboração Própria.

A análise gráfica dos pontos pode ser observado na figura 4.12. Os pontos relativos aos recifes da foz do Rio Doce apresentaram um relativo agrupamento por tipo de *habitat*, com um distanciamento entre os dois grupos. Os pontos da parede se acomodaram ao norte do gráfico. As amostras do topo se localizam na faixa vertical esquerda. Em comparação aos outros recifes, as amostras do Espírito Santo parecem se concentrar nas partes periféricas da figura, além disso apresentam menor distinção entre o topo e parede.

Em relação aos outros locais os pontos dos recifes do Espírito Santo parecem estar mais deslocados. Os pontos da parede estão um pouco acima das paredes dos outros locais. Porém a maior diferença aparece no *habitat* do topo do recifes. Enquanto nos outros pontos há uma clara concentração dos pontos, principalmente aa direita do gráfico, nos recifes localizados próximos a foz do Rio Doce os pontos aparecem divididos entre a parte de cima e a parte de baixo do lado esquerdo do

gráfico

Os recifes de Parcel de Abrolhos, Paredes e Timbebas apresentaram um comportamento semelhante em relação a divisão por habitats. Os pontos relativos a parede estão na parte esquerda do gráfico, e os pontos relativos ao topo ficam no parte direita da imagem.

Dentre estes três grupos de recifes, em Parcel de Abrolhos há uma maior diferença entre os dois grupos de *habitat*. Há poucos pontos da região ocupando a região central do gráfico quando comparamos com Paredes e Timbebas.

Os recifes de Timbebas ocupam em sua maior parte a região central do gráfico. Ainda assim, há uma distinção entre os pontos da parede e do topo.

A análise de abundância relativa foi realizada em paralelo com o objetivo de entender as razões que levaram as similaridades e dissimilaridades observadas. Além de permitir entender em parte o ordenamento dos pontos, há também a informação sobre o tipos de espécie que resultou na variação.

Foram consideradas apenas as dez espécies mais representativas, considerando todos os pontos de coleta. A partir das médias de abundância em cada um dos locais foi composto o gráfico de barras da figura 4.13. Os *habitats* foram novamente divididos entre topo e parede. Para cada um dos pontos de coleta, foi analisada a abundância relativa das principais frequências, a diferença que falta para a completude de cada barra é composto por outras espécies.

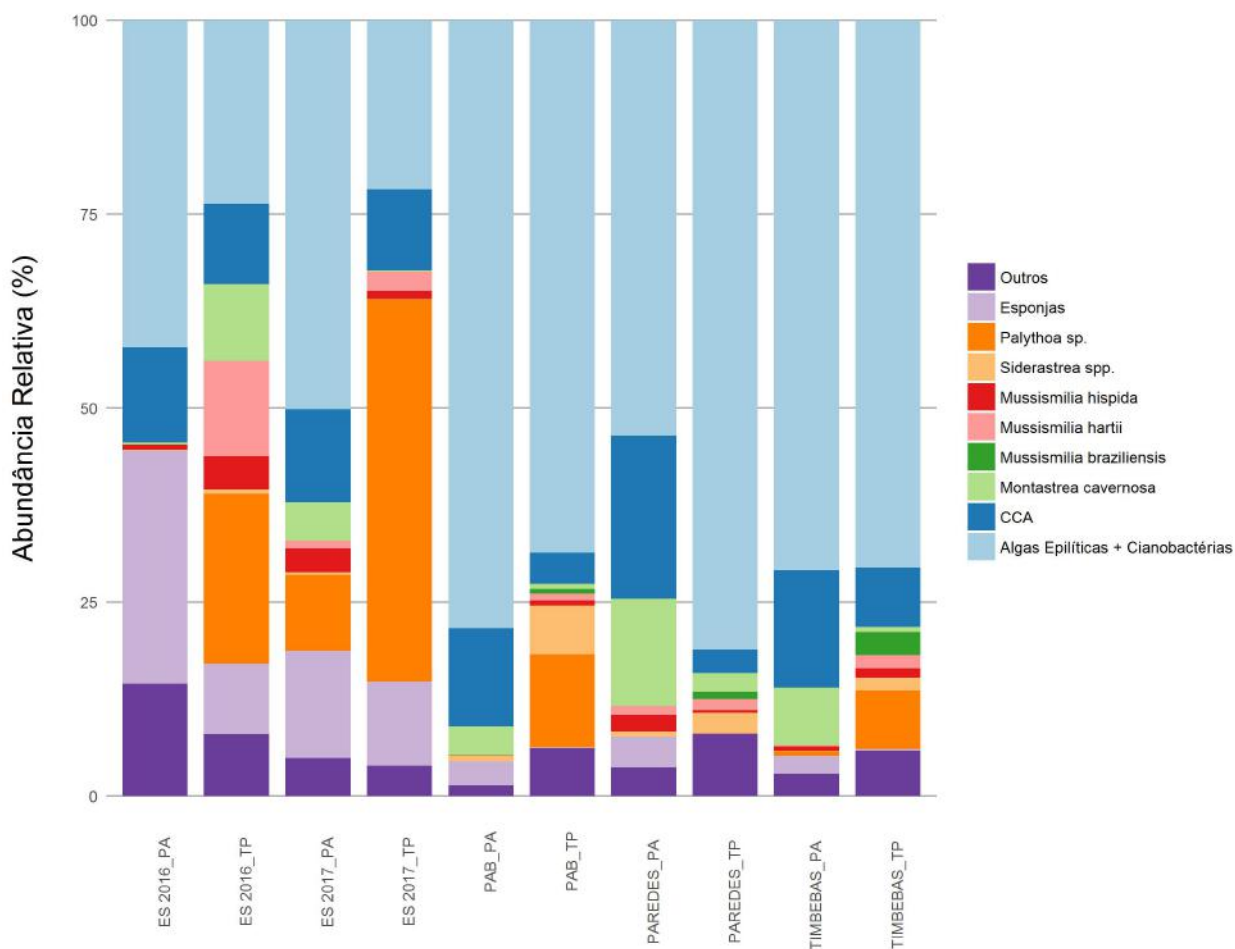


Figura 4.13: Espécies mais presentes em cada Recife.

Fonte: Elaboração Própria.

Os resultados indicam que a principal espécie que ocupa os recifes são as algas epilíticas. Em Parcel de Abrolhos, Paredes e Timbebas essa dominância fica mais evidente.

Nos recifes da foz do Rio Doce as algas epilíticas são também a espécie mais presente na parede, porém não ocorre o mesmo no topo, onde o coral *palythoa sp.* é a espécie que ocupa a maior área de cobertura. O mesmo padrão pode ser observado tanto em 2016 quanto em 2017, porém no ano posterior a dominância dessas duas espécies é mais evidente.

Como os índices de diversidade indicaram, nos recifes da foz do Rio Doce há a presença de bom número de espécies, e com distribuições mais equitativas do que em outros recifes. As dez espécies de maior recorrência em todos os locais representam menos da abundância total nos recifes do Espírito Santo do que nos outros sítios.

Há uma maior recorrência de esponjas nos recifes do Espírito Santo do que nos

outros locais, em especial no topo, onde quase 30 % do espaço total é coberto com esse grupo. Além das esponjas cianobactérias também tem maior presença nesses recifes do que nos outros locais analisados.

É destacado a cobertura de espécies coralíneas nos recifes do Espírito Santo. No ambiente da parede, mais de 10% é ocupado pelo recife *Mussismilia Hartii*, e por volta de 30 % pelo *Palythoa sp.*.

4.3 Modelos de Previsão

Nesta seção são apresentados os modelos de previsão da camada coralínea. Em um primeiro momento o tratamento dos dados é discutido em conjunto com as primeiras análises. Depois os métodos utilizados são apresentados, com suas características individuais. Ao final, os resultados são comparados.

4.3.1 Tratamento de Dados

Um dos maiores indicadores da saúde de um sistema recifal é a cobertura de espécies coralíneas no bentos local. A importância de espécies de corais na composição dos recifes já foi discutida nesse estudo, por essa razão os modelos de previsão discutidos nessa seção terão como seu objeto principal, a saída do modelo, a cobertura de corais.

O objetivo do modelo é captar as variáveis que tiveram mais influência para a cobertura coralínea. Os modelos foram testados com as variáveis explicativas oriundas as próprias informações do bentos, informações físicas sobre os corais como a divisão por tipo de *habitat* e dados obtidos por sensoriamento remoto. A cobertura dos corais foram agrupados em uma variável dependente, as espécies encontradas estão listadas na tabela 4.5.

Um resumo com as variáveis utilizadas e a descrição dos valores utilizados pode ser visto na tabela 4.6. No total mais de 400 amostras foram analisadas nos modelos.

As variáveis que tem como unidade a porcentagem de cobertura foram obtidas a partir da metodologia de foto-quadrados do bentos recifal nos locais. As variáveis profundidade, distância da costa e nível de proteção foram obtidas nessas mesmas excursões que coletaram dados do bentos.

As variáveis explicativas clorofila A, coeficiente de difusão (simboliza a turbidez), carbono orgânico dissolvido, radiação fotossintética e temperatura noturna foram obtidas a partir da observação remota dos recifes pelo serviço MODIS. Os valores utilizados para cada variável representam a média dos últimos 12 meses antes da fotografia do bentos.

A tabela 4.6 indica que nos recifes localizados no norte do Espírito Santo há

Corais Agrupados
<i>Agaricia agaricites</i>
<i>Agaricia fragilis</i>
<i>Favia gravida</i>
<i>Carijoa riisei</i>
<i>Favia leptophylla</i>
<i>Madracis decactis</i>
<i>Meandrina braziliensis</i>
<i>Millepora alcicornis</i>
<i>Millepora nitida</i>
<i>Montastrea cavernosa</i>
<i>Muricea flamma</i>
<i>Muricea sulphurea</i>
<i>Mussismilia braziliensis</i>
<i>Mussismilia hartii</i>
<i>Mussismilia hispida</i>
<i>Neospongodes atlântica</i>
<i>Palythoa caribaeorum</i>
<i>Phyllogorgia dilatata</i>
<i>Plexaurella grandiflora</i>
<i>Plexaurella regia</i>
<i>Porites astreoides</i>
<i>Porites branneri</i>
<i>Scolymia wellsii</i>
<i>Siderastrea spp.</i>
<i>Zoanthus sp.</i>

Tabela 4.5: Corais agrupados.
 Fonte: Elaboração Própria.

uma maior cobertura de corais em relação aos outros pontos. Além disso, não há evidências de macroalgas e também uma menor cobertura de algas epilíticas, são três indicadores de boa qualidade dos recifes. Há um maior destaque para a concentração de esponjas, aparentemente maior do que nos outros locais.

Os dados de entrada e saída passaram pelo processo de normalização de acordo com as equações 3.13 e 3.14. Essa transformação não se refletiu em melhores resultados, assim como previsto, porém a razão de sua utilização é auxiliar o momento de análise.

Os dados de cobertura de corais foram modelados duas vezes. No primeiro modelo foram utilizados todas as variáveis disponíveis, já o segundo foi realizado apenas com as variáveis físico-químicas, sem levar em consideração a cobertura de outras espécies. O resumo dos modelos está na tabela 4.7. As variáveis explicativas foram utilizadas em outros estudos anteriores [14].

Variável	ESPÍRITO SANTO	PAB	PAREDES	TIMBEBAS
Variável Dependente				
Corais (% de cobertura)	38,8 ± 27,2	18,6 ± 21,2	20,4 ± 18	16,9 ± 19
Variáveis Explicativas				
Algas Epilíticas (% de cobertura)	31,7 ± 16	59,4 ± 22,5	50,3 ± 19,5	52,7 ± 23,8
CCA (% de cobertura)	11,2 ± 6,8	14,9 ± 13,7	13,7 ± 13	22,9 ± 19,2
Esponjas (% de cobertura)	18,2 ± 14,6	1,2 ± 2,4	1,4 ± 2,5	0,9 ± 1,6
Clorofila A (mg m ⁻³)	0,3 ± 0	0,3 ± 0	1,7 ± 0,1	0,5 ± 0
Coefficiente de difusão (490 nm)	0,1 ± 0	0 ± 0	0,1 ± 0	0,1 ± 0
Carbono orgânico dissolvido (mg m ⁻³)	95,5 ± 0	80 ± 3,1	245,5 ± 12,8	129,5 ± 5,6
Radiação Fotossintética (mol m ⁻²)	46,2 ± 0	45,3 ± 1	43,8 ± 0,9	45,1 ± 1,2
Temperatura noturna (°C)	26 ± 0	25,6 ± 0,2	25,5 ± 0,2	25,5 ± 0,1
Profundidade (m)	14,3 ± 2,2	11,8 ± 4,4	4,2 ± 1,9	7,7 ± 2,5
Distância da costa (km)	23 ± 5,4	61,5 ± 0,1	14,3 ± 0	17,9 ± 1
Proteção	1	3	1	2

Tabela 4.6: Médias e desvios padrões das variáveis utilizadas nas modelagem por Recife.

Fonte: Elaboração Própria.

Variáveis Explicativas	Modelo 1	Modelo 2
Algas Epilíticas	Sim	Não
CCA	Sim	Não
Esponjas	Sim	Não
Clorofila A	Sim	Sim
Kadmium	Sim	Sim
Carbono orgânico dissolvido	Sim	Sim
Radiação Fotossintética	Sim	Sim
Temperatura noturna	Sim	Sim
Profundidade	Sim	Sim
Distância da costa	Sim	Sim
Proteção	Sim	Sim

Tabela 4.7: Variáveis explicativas para cada um dos modelos formulados tendo como saída a cobertura dos recifes de corais.

Fonte: Elaboração Própria.

Nas próximas subseções os quatro modelos escolhidos para desenvolvimento são abordados, analisando tanto sua formulação quanto os resultados preliminares. Na

última subseção deste capítulo é feita a comparação entre eles.

4.3.2 Regressão Linear

O método utilizado foi de regressão múltipla utilizado linear, ou seja, medindo a capacidade dos pontos se ordenarem de acordo com combinações de retas simples. O modelo teórico utilizado é apresentado em 4.3.2. A saída y representa a camada coralínea, o fator β é o intercepto e os α_i representam as inclinações de reta para cada variável explicativa.

$$y = \beta + \alpha_1 \text{Algas} + \alpha_2 \text{CCA} + \alpha_3 \text{Esponjas} + \alpha_4 \text{Habitat} + \alpha_5 \text{CHLOR}_A + \alpha_6 \text{KD} + \alpha_7 \text{POC} + \alpha_8 \text{PAR} + \alpha_9 \text{NSST} + \alpha_{10} \text{PROF} + \alpha_{11} \text{DIST} + c(4.1)$$

Os resultados da regressão no R são apresentados na tabela 4.8. A magnitude e sinal dessa correlação é indicada pela segunda coluna. Nas duas últimas colunas está indicado a significância dessa relação, pois nem sempre os dados estão claramente estatisticamente relacionados.

	Estimate	Std. Error	Valor t	Pr(<— t —)	
(Intercept)	19.7619	0.5608	35.238	<2e-16	***
Algas	-14.0206	0.5864	-23.908	<2e-16	***
CCA	-9.4959	0.5114	-18.569	<2e-16	***
Esponjas	-2.0087	0.3716	-5.405	1.09e-07	***
CHLOR _a	83.0769	33.3973	2.488	0.0132	*
KD	-101.2470	48.3567	-2.094	0.0369	*
POC	13.2011	19.2586	0.685	0.4934	
PAR	0.3746	0.7740	0.484	0.6286	
NSST	-0.8942	0.7088	-1.262	0.2078	
PROF	-0.2345	0.7368	-0.318	0.7505	
DIST	-4.8059	2.4545	-1.958	0.0509	.

Tabela 4.8: Modelo 1 em regressão linear- Resumo dos coeficientes para cada variável, e seu valor de significância.

Nível de significância para cada variável a partir do valor p: .- 10%, *-5%, **-1%, ***-0,1%

Fonte: Elaboração Própria.

O valor p indica a significância da relação, quanto menor, maior é a probabilidade da correlação entre os dados ser relevante. O número de asteriscos na última coluna representa quando aceitar essa relação para diversos padrões, a quantidade de asteriscos representa o rigor para aceitação da correlação real.

Os dados apontaram como relevantes as quatro variáveis utilizadas relativas a cobertura, macroalgas, Algas, Esponjas e *CCA*. Todas apresentaram relações negativas.

4.3.3 Modelos Aditivos

Os modelos aditivos foram utilizados de acordo com o método exposto na seção 3.3.1. O método de estimativa dos parâmetros utilizou os chamados *joelhos* como parâmetros a serem estimados. Algumas variáveis explicativas não obtiveram graus de liberdade suficiente para estimar suas funções de correlação. Estas variáveis foram utilizadas como fatores binários. A equação utilizada está na equação 4.3.3.

As variáveis explicativas *Habitat*, não possuíam graus de liberdade suficientes para serem incorporadas ao modelo aditivo como funções. Isso ocorreu porque há pouca variação para esses dois tipos de dados, que se comportaram melhor quando tratadas como variáveis binárias, por isso foram incorporadas ao modelo como fatores. Aconteceu o mesmo com as variáveis de profundidade, distância da costa e proteção.

$$y = s(\text{Algas}) + s(\text{CCA}) + s(\text{Esponjas}) + s(\text{CHLOR}_A) + s(\text{KD}) + s(\text{POC}) + s(\text{PAR}) + s(\text{NSST}) + c(4.2)$$

O resultados do modelo no R são apresentados na tabela 4.9. As duas últimas colunas novamente apresentam a significância estatística dessa regressão.

	edf	Ref.df	F	valor-p	
s(Algas)	3.815	4.776	151.747	<2e-16	***
s(CCA)	4.944	6.062	64.151	<2e-16	***
s(Esponjas)	1.000	1.000	25.053	8.27e-07	***
s(CHLOR _a)	1.000	1.000	0.016	0.89829	
s(KD)	1.000	1.000	0.000	0.99582	
s(POC)	1.000	1.000	0.090	0.76378	
s(PAR)	3.125	3.721	0.389	0.84688	
s(NSST)	8.250	8.702	2.434	0.00841	**
s(PROF)	7.998	8.466	23.657	<2e-16	***

Tabela 4.9: Modelo 1 em métodos aditivos- Resumo dos coeficientes para cada variável, e seu valor de significância.

Nível de significância para cada variável a partir do valor p: .- 10%, *-5%, **-1%, ***-0,1%

Fonte: Elaboração Própria.

As variáveis explicativas que descrevem cobertura de corais no recifes novamente

apresentaram uma relação estatisticamente relevante com probabilidade de erro menor que 0.1%. A temperatura e a profundidade também apresentaram significância na correlação com a cobertura de corais.

Diferentemente da regressão linear múltipla o quadro que apresenta a significância das relações entre as variáveis explicativas e dependentes não oferece informação sobre a magnitude nem direção do impacto das variáveis. Para isso é preciso recorrer a análise de sensibilidade, que é apresentada mais a frente.

4.3.4 Redes Neurais

As principais variáveis a serem determinadas em um modelo de redes neurais são: o número de neurônios intermediários, são recomendados que variem de duas a três vezes a quantidade de variáveis de entrada, o número de iterações e a combinação do passo com o *momentum*, que determina a direção e comprimento da variação dos pesos.

Essas variáveis são determinadas de forma empírica. Ou seja, o modelo é produzido diversas vezes, com diferentes parâmetros em cada uma e os melhores resultados são anotados para serem utilizados no modelo final.

Os parâmetros utilizados no modelo final foram obtidos de modo empírico.

Símbolo	Nome	Padrão Utilizado
n	Variáveis de entrada	13
j	Número de neurônios intermediários	35
γ	Passo	0.03
μ	<i>Momentum</i>	0.5
-	Número de Iterações no treinamento	5000
-	Fração dos dados para Treinamento	85%
E_{min}	Erro Mínimo	$10e - 25$

Tabela 4.10: Variáveis definidas no modelo de Redes Neurais.

Fonte: Elaboração Própria.

O modelo de Redes Neurais é muito utilizado para previsão de resultados dado padrões de entrada. Porém não é possível informar as razões que levaram aos resultados não podem ser interpretadas de forma natural, é necessário a implementação de análises do modelo [54].

O método de particionamento de Garson leva em consideração o peso das conexões para calcular o peso de cada variável de entrada no modelo final. Os resultados da figura 4.14 foram obtidos com o método apresentado na seção 3.4.2.

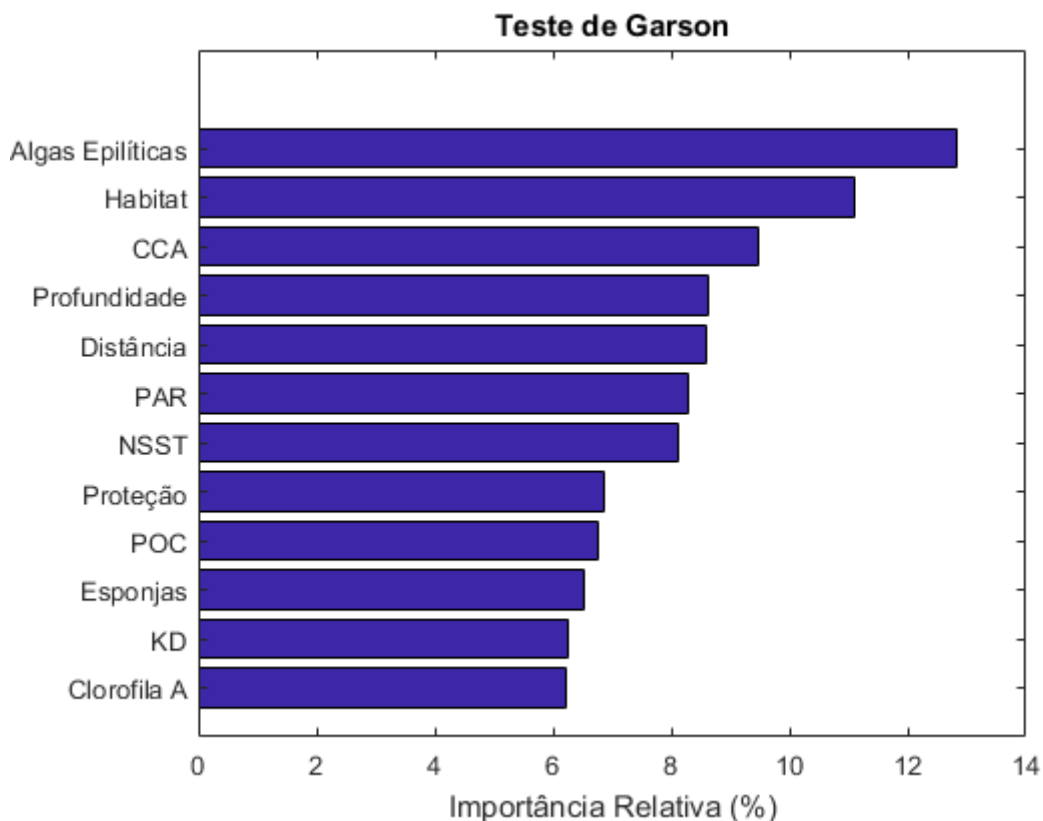


Figura 4.14: Particionamento de Garson. Indica a importância relativa de cada variável explicativa no modelo.

Fonte: Elaboração Própria.

Os resultados do particionamento indicaram que as variáveis mais importantes para a cobertura de recifes com espécies coralíneas foram as algas epilíticas, com certa vantagem sobre as demais, em torno de 13% do total.

Em seguida num grupo que tem sua participação individual variando entre 8 e 10 %, estão as variáveis profundidade, *habitat* e CCA. Outras variáveis físico químicas tem sua participação relativa variando entre 6 e 8 %.

No método de particionamento de Garson não há distinção entre o tipo de impacto causado por cada variável. Para avaliar isso é também necessário a utilização de análise de sensibilidade.

4.3.5 Árvores de Crescimento Progressivo

O modelo de árvores de crescimento progressivo tem duas variáveis que se apresentam com maior peso para o modelo. A primeira é a complexidade de cada árvore individual, uma árvore mais ramificada pode oferecer mais informações, porém tem maior custo computacional e pode acabar prejudicando o modelo como um todo.

A segunda variável é o número de árvores por grupo. As árvores foram agrupadas

em associações de setenta unidades, e a capacidade de previsão é analisada por uma validação interna do modelo. A partir do resultado obtido, um novo grupo de setenta é formado de forma a melhorar o resultado obtido.

Símbolo	Nome	Padrão Utilizado
n	Variáveis de entrada	13
j	Número de árvores por grupo	70
γ	Passo	0.01
-	Complexidade da árvore	7
-	Validação Interna	90%

Tabela 4.11: Variáveis definidas no modelo de Árvores de Crescimento Progressivo.
Fonte: Elaboração Própria.

O modelo de árvores de crescimento progressivo assim como as redes neurais não apresenta um valor p para analisar o índice de significância de cada relação entre as variáveis explicativas e dependentes.

Os métodos de análises incluem a importância relativa de cada variável, método similar ao particionamento de Garson, que no caso do modelo em questão é apresentado na figura 4.15.

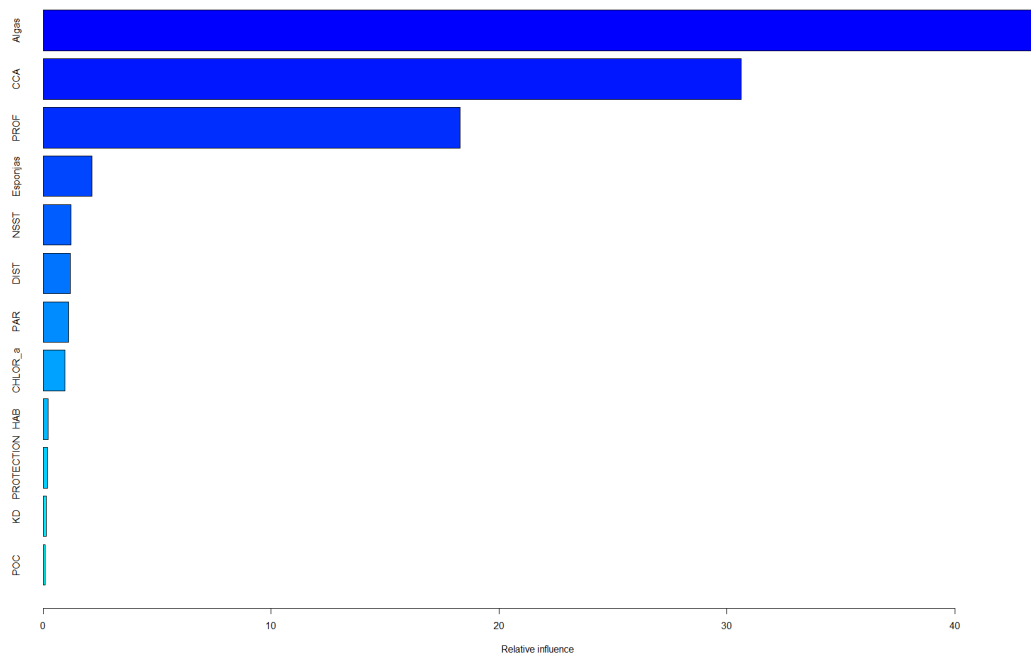


Figura 4.15: Particionamento da importância de cada variável explicativa no modelo de árvores de crescimento progressivo.

Fonte: Elaboração Própria.

As variáveis de cobertura novamente apresentaram maior relevância, com a combinação de algas epilíticas e cianobactérias representando a maior influência com mais de 50 % de importância relativa. *CCA* também apresentou relevância acima de 30%.

A profundidade aparece como a quarta força do modelo, com influência chegando a quase %, maior inclusive que o índice de cobertura de esponjas.

As outras variáveis físico-químicas apresentaram menor papel nesse modelo do que nas redes neurais. Nesse método houve uma maior concentração de importância para as variáveis mais relevantes, com um grande número de variáveis apresentando participação final pequena em relação ao modelo.

4.3.6 Comparação de Resultados

Como visto na literatura as funções principais de modelos que correlacionam o comportamento de uma variável dependente são a previsão futura de dados e a análise das relações. Nessa seção as ferramentas utilizadas na literatura por cada método são apresentadas.

As formas de comparação entre os métodos utilizadas neste estudo seguem o modelo adotado em outros trabalhos [29][59][56]. Um resumo dos métodos abaixo:

i) Comparação da capacidade de previsão tanto dos dados de treino, quanto dos dados de validação cruzada. Esses dados são separados do grupo inicial, e são chamados de dados de validação do modelo.

ii) Análise de sensibilidade das variáveis, os modelos são simulados para cada uma das variáveis explicativas, com estas variando durante toda sua abrangência de resultados, enquanto isso as outras variáveis de entrada são mantidas com seus valores constantes. É utilizado apenas o modelo 1 da tabela 4.7 para a análise de sensibilidade, de forma a permitir uma maior profundidade.

Em relação a previsão, uma medida da capacidade de um modelo pode ser dado pela regressão dos dados reais com as saídas de um grupo de dados separados, esse indicador é analisado pelas tabelas 4.12 e 4.13. Existem outras formas de se trabalhar essa capacidade, porém este não é o principal foco do estudo, algumas outras formas serão sugeridas ao final do estudo.

Os métodos desenvolvidos realizaram a modelagem em duas fases. Primeiramente uma parcela de cerca de 80 % dos dados foi separada de forma aleatória, esse grupo foi utilizado para o treinamento dos modelos. O restante das informações foi utilizada na segunda etapa, a validação cruzada.

A comparação através de validação cruzada dos dados permite analisar a capacidade dos modelos de prever resultados tanto nos dados de treino, como nos dados de validação.

Os modelos paramétricos foram simulados uma vez cada, por estabelecerem os melhores parâmetros que se encaixam aos dados, seus resultados se mantêm constantes quando desenvolvidos diversas vezes. Já os modelos não-paramétricos no seu resultado final dependem de sua configuração inicial, que é determinada de forma aleatória, dessa forma seus resultados apresentam variância. Por essa razão as modelagens desse segundo caso devem ser rodadas diversas vezes, para uma análise mais abrangente.

As modelagens pelos métodos de regressão linear e aditiva são chamadas de paramétricas. São elaborados de forma a interpretar as relações entre as variáveis a partir de modelos pré-definidos, restando apenas o cálculo dos parâmetros que melhor se adequem a situação.

Já os modelos de redes neurais e árvores de crescimento progressivo são ditos não paramétricos. Essas modelagens tentam *aprender* da melhor forma comportamento das variáveis dependentes, sem se basear em um modelo *a priori*. Nesses tipos de modelagem, é necessário definir orientações para o modelo, que funcionam como instruções para o processo iterativo. Os padrões utilizados foram determinados de forma empírica a partir dos melhores resultados obtidos e estão descritos nas tabelas 4.10 e 4.11.

Capacidade de previsão

Os quatro modelos foram desenvolvidos com os dados para treinamento e depois testados novamente com os dados de validação. Para ambos os casos os resultados esperados pelos modelos foram comparados com a cobertura reais de corais a partir de uma regressão simples. Os resultados obtidos pelos quatro modelo estão na tabela 4.12, quanto mais próximo o valor está da unidade, melhor é a regressão.

Método		R^2 -a Treino	R^2 -a Valida
Regressão Linear Múltipla		0.6706	0.591
Método Aditivo		0.788	0.7334
Redes Neurais	Média	0.9676 ± 0.0054	0.6488 ± 0.0690
	Máximo	0.9778	0.7785
	Mínimo	0.9510	0.4101
Árvore Cresc. Progressivo	Média	0.9598 ± 0.0229	0.8202 ± 0.0025
	Máximo	0.9862	0.8246
	Mínimo	0.904	0.8148

Tabela 4.12: R^2 dos modelos utilizados. Resultado para a modelagem com todas as variáveis explicativas.

Fonte: Elaboração Própria.

Nos dados de treinos do modelo 1 todas os métodos apresentaram grande acurácia de previsão dos dados de treino. O pior resultado ocorreu no método de regressão linear múltipla. As redes neurais e o modelos de árvores apresentaram a melhor estimativa de dados, chegando próximo a unidade.

Em relação aos dados de validação, os modelos aditivos se destacaram junto com o método de árvores, inclusive para o primeiro caso o desempenho foi superior a própria eficiência em ler os dados de treino. A regressão linear múltipla vem a seguir, com desempenho levemente superior aos dados de treinamento. As redes neurais obtiveram a menor média de acurácia, porém devido a variação dentro do próprio modelo, as melhores modelagens chegaram próximas aos valores obtidos pelos métodos mais eficientes.

Na tabela 4.13 são apresentados os resultados para o modelo 2, que não utiliza coberturas de outros grupos de espécies como variável explicativa para a abundância das espécies coralíneas.

Método		R^2 -a Treino	R^2 -a Valida
Regressão Linear Múltipla		0.1442	0.1696
Método Aditivo		0.383	0.2229
Redes Neurais	Média	0.4399 ± 0.0000	0.3538 ± 0.0025
	Máximo	0.4399	0.3545
	Mínimo	0.4399	0.3536
Árvore Cresc. Progressivo	Média	0.2845 ± 0.0156	0.1554 ± 0.02152
	Máximo	0.3171	0.1998
	Mínimo	0.2424	0.0997

Tabela 4.13: Modelo 2 - R^2 dos modelos utilizados. Resultado para a modelagem com todas as variáveis explicativas.

Fonte: Elaboração Própria.

Nesse modelo a capacidade de previsão de todos os métodos é muito menor. A maior perda se deu no método de regressão linear múltipla. Os métodos de redes neurais e modelos aditivos obtiveram os melhores resultados. As árvores de crescimento progressivo, que obtiveram o melhor desempenho na tabela 4.12, dessa vez tiveram uma queda expressiva de acurácia em comparação as ferramentas que obtiveram resultados mais eficientes.

Uma das explicações possíveis para o desempenho é concentração da importância das variáveis de cobertura para o método de árvores do modelo 1. As variáveis que analisam a capacidade de cobertura no bentos ocuparam mais de 90% de importância relativa no modelo. Quando essas variáveis foram retiradas, o modelo perdeu em parte sua capacidade de previsão, demonstrando menos capacidade para se adequar

aos dados mais restritos.

Análise de Sensibilidade

Os resultados da análise de Sensibilidade dos métodos são relativos ao Modelo 1, que conta com todas as variáveis. Na primeira figura aparecem os resultados das variáveis de cobertura para todos os modelos, e na segunda a análise dos parâmetros físico químicos.

O método de análise foi o mesmo para todos as ferramentas utilizadas. Foram simulados cem resultados para cada variável explicativa, acompanhando toda a abrangência do espaço coletado dessa variável, enquanto isso as demais variáveis são mantidas na sua média.

O eixo y dos gráficos representa a cobertura de corais estimada. A dimensão varia de 0% a 100%, que representam, nenhuma evidência de cobertura coralínea e cobertura total, respectivamente. Algumas simulações estipularam valores negativos ou maiores que a unidade aos resultados, estes não são representados nos gráficos, pois representam situações irreais.

As análises de sensibilidade para as variáveis de cobertura são apresentadas na figura 4.16.

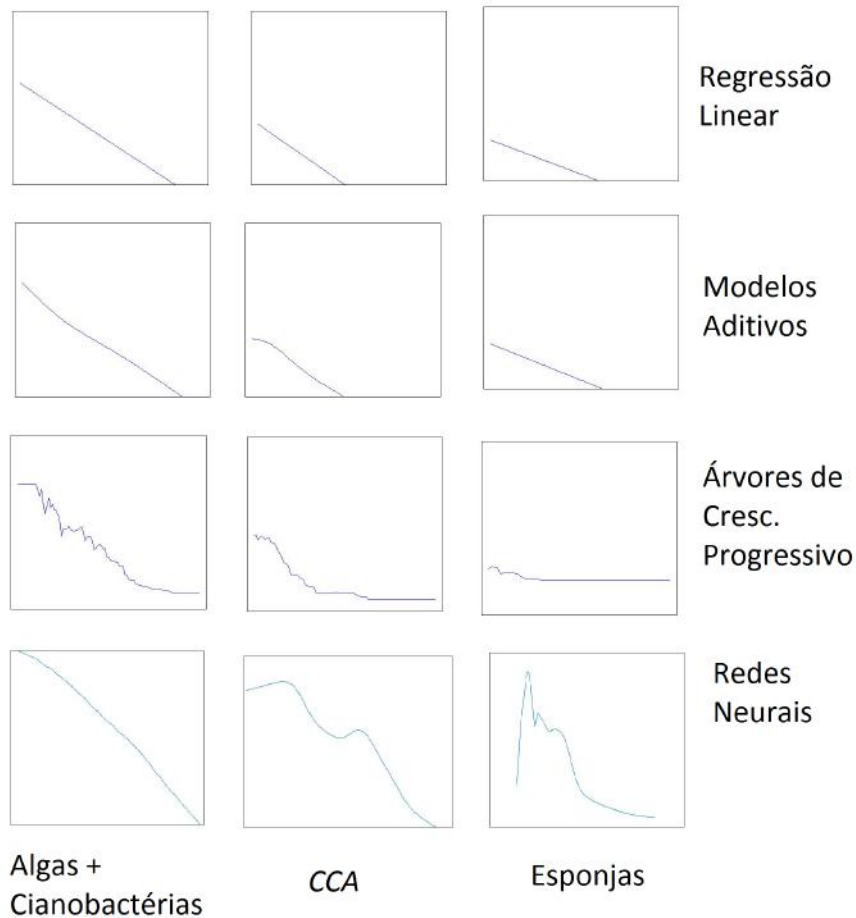


Figura 4.16: Análise de sensibilidade das variáveis de cobertura no modelo 1, para cada um dos métodos utilizados.

Fonte: Elaboração Própria.

Os resultados da análises de sensibilidade das variáveis de cobertura apresentaram um correlação predominante negativa para todas as espécies. Esse resultado era de se esperar, pois há uma disputa por espaço no bentos, quanto maior o espaço ocupado por um grupo, menos há disponível para o crescimento de espécies coralíneas.

Dentre os grupos analisados chama a atenção o comportamento das esponjas. Em todas as ferramentas essa espécie apresentou a menor correlação negativa, e na análise das redes neurais há identificação de reforço positivo no aumento desse grupo para a cobertura de corais.

A análise de sensibilidade das variáveis MODIS são apresentadas na figura 4.17.

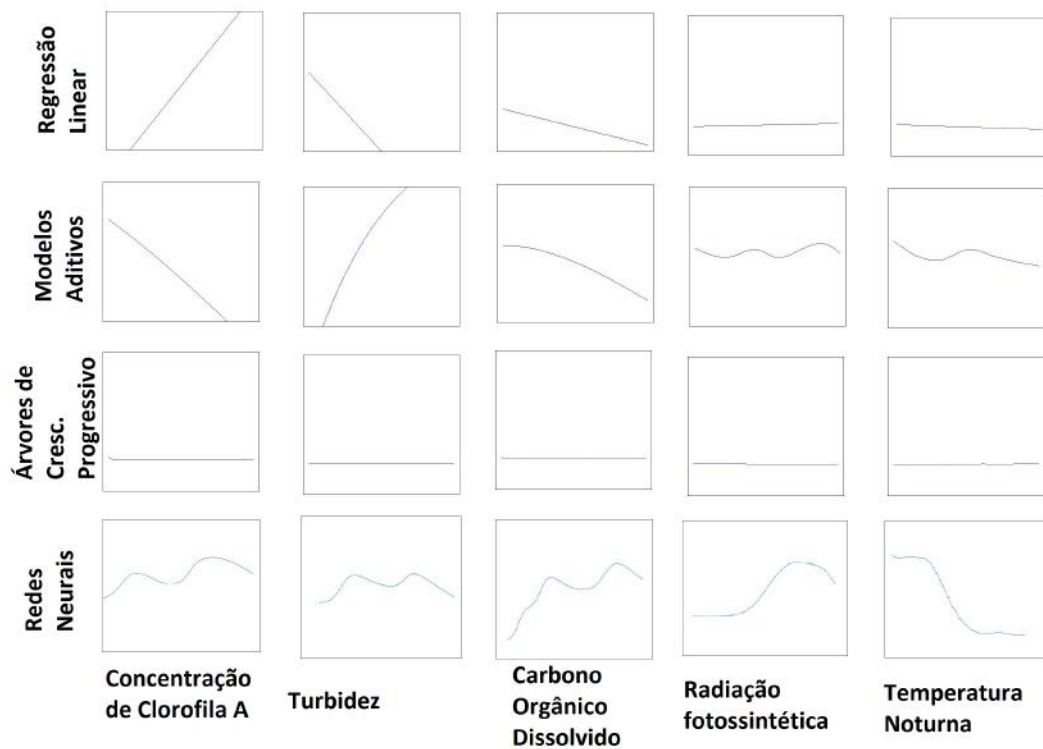


Figura 4.17: Análise de sensibilidade das variáveis MODIS no modelo 1, para cada um dos métodos utilizados.

Fonte: Elaboração Própria.

Graficamente há a interpretação de que o aumento da temperatura noturna causa influência negativa na camada coralínea, porém nos modelos aditivos não há significância estatística que sustente essa afirmação, para esse método.

Em relação a outras variáveis, há um comportamento dúbio sobre a influência, determinados métodos apresentam um tipo de comportamento, enquanto em outros o resultado é oposto. Essas variáveis em geral apresentaram pouca representatividade nos modelos.

Capítulo 5

Discussões

Os resultados apresentados ajudam a traçar um panorama da situação atual dos recifes que se localizam na foz do Rio Doce, sendo possível comparar com outros ecossistemas recifais também localizados em Abrolhos. Os modelos de aprendizado estatístico desenvolvidos para a região buscaram justificar fatores que tem maior relevância para a cobertura de espécies coralíneas.

A região da foz do Rio Doce sofre o impacto da ação antrópica há muitos anos, principalmente com a ação de comunidades locais que dependem dos recifes para sustento. Com o desastre de Mariana se torna ainda mais importante a observação do comportamento biológico do complexo recifal.

Em termo de volume de rejeitos quanto em magnitude de área afetada o desastre ocorrido em Marina pode ser considerado o maior do mundo. O total de construções danificadas foi de 806, sendo pelo menos 218 completamente destruídas. O distrito mais afetado foi Bento Rodrigues, que se localiza a 6 km da barragem do fundão, lá aproximadamente 84 % das residências foram inutilizadas [20].

A expectativa inicial era de que a lama oriunda da quebra da barragem não chegasse aos recifes de Abrolhos [44]. Porém novos indício obtidos a partir de dados de sensoriamento remoto, que podem ser vistos na figura 4.3, apontam que a parcela da lama que se dirigiu ao norte da foz do Rio Doce alcançou a região dos corais que se localizam perto de Conceição da Barra.

A análise do bentos mostrou uma diferença significativa nos novos recifes entre a composição dos recifes por tipo de *habitat*. Na parede há um maior concentração de cianobactérias, esponjas e algas epilíticas. No topo há maior cobertura de corais. Essa diferença sustenta os resultados encontrados por FRANCINI-FILHO et al[31] no banco de Abrolhos.

Nos corais que se situam mais próximo a foz do Rio Doce existem evidências de uma maior concentração de cianobactérias e dos corais do tipo *Palythoa caribaeorum*. Esses corais estariam em um estado de degradação mais avançado do que os do norte.

Nos corais norte há maior concentração de espécies endêmicas, em destaque a

cobertura do coral *Mussimilia hartii*, que chega a ocupar uma área média de mais 20 % nas amostras obtidas, considerando o topo do recifes. Essa espécie está na lista vermelha de ameaça global de extinção da IUCN [1], reforçando ainda mais a importância de preservação dos recifes do Espírito Santo.

Em relação a outros locais com concentração de banco de corais em Abrolhos, o número de espécies encontradas nos recifes da foz do Espírito Santo está próximo a média. Porém nos índices de entropia de Shannon e Equitatividade de Pielou houve um destaque para os recifes próximos a foz do Rio Doce. Como os índices são relacionados, a hipótese é que há uma distribuição mais homogênea das espécies no Espírito Santo, sem tanta dominância de apenas um grupo sobre os demais.

Em relação a análise bidimensional, há uma concentração dos pontos dos recifes do foz do Rio Doce isolados a esquerda da figura. Enquanto os outros locais tiveram seus pontos dispersos de forma mais concêntrica, os pontos dos recifes do Espírito Santo apresentaram um leve desvio a esquerda, e menor divisão entre os tipos de *habitat*.

Nos corais do Espírito Santo foi observada uma concentração maior de cianobactérias e de esponjas em relação aos outros locais. Ao mesmo tempo houve um menor domínio de algas epilíticas. A preocupação é especial com os colchões de cianobactérias que estão presentes nos recifes.

Este tipo de cobertura é composto por diversos tipos de bactérias pertencentes ao mesmo grupo e dentre essas espécies de micro-organismos algumas podem ser potencialmente perigosas para os corais saudáveis, expondo-os ao risco de transmissão de doenças patógenas [13]. A abundância de cianobactérias está relacionada a proximidade de áreas urbanizadas, a temperatura e até mesmo com a força das ondas nos locais. Com a ação do homem sobre o globo, existe uma tendência de aumento de cobertura de cianobactérias nos recifes[9].

Também ocorreram mudanças no litoral do Espírito Santo entre 2016 e 2017. Há indícios de aumento da cobertura de algas epilíticas nos recifes e também de abundância da espécie coralínea *palythoa*. Porém comparando o ano posterior ao desastre com os recifes localizados no banco de abrolhos, ainda há menor cobertura de algas epilíticas, maior cobertura coralínea, e maior equitatividade de espécies nos recifes do litoral capixaba, oferecendo poucos indícios de que a lama contaminou severamente os recifes.

Foram gerados modelos baseados em métodos de aprendizado estatístico para analisar quais seriam principais fatores de influência sobre a camada de corais em recifes de Abrolhos. As principais formas de avaliação dos modelos foram a capacidade de previsão da camada coralínea, tanto com o grupo de dados de treino como por validação cruzada, e a análise de sensibilidade das variáveis explicativas.

A utilização dos próprios dados de cobertura e fatores físico-químicos, modelo

1, obteve como resultado um R^2 com valores elevados em comparação a outros trabalhos para todos os modelos observados. No modelo 2, quando foram utilizados apenas dados de sensoriamento remoto para fazer a previsão dos recifes a capacidade de previsão do R^2 ajustado caiu consideravelmente, chegando a um nível parecido com outros trabalhos que utilizavam métodos semelhantes [29, 59].

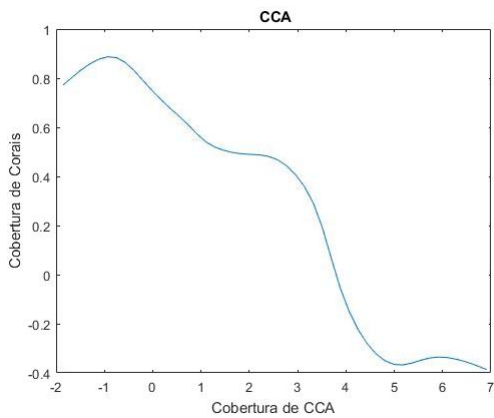
A observação remota é uma forma interessante para observar mudanças ambientais nos ecossistemas e pode fornecer informações importantes para acompanhamento de recifes de corais. O custo benefício do acompanhamento remoto é interessante caso seja feito de forma eficiente, pois permite investimentos direcionados aos resultados, otimizando a utilização dos recursos de preservação [7]. Porém os resultados apresentados na tabela 4.13 demonstraram uma baixa capacidade dos métodos, os motivos podem ser a vasta abrangência temporal e espacial dos dados remotos, já que foram coletados a série anual de um espaço muito maior que a localização dos recifes.

No segundo modelo também, há um decréscimo acentuado em especial do método de árvores de crescimento progressivo, que no modelo 1 havia obtido o melhor desempenho. Esse método já foi utilizado com eficiência para a previsão de espécies de corais [14]. Para a modelagem 2 neste estudo obtiveram melhor resultado os métodos aditivos e as redes neurais.

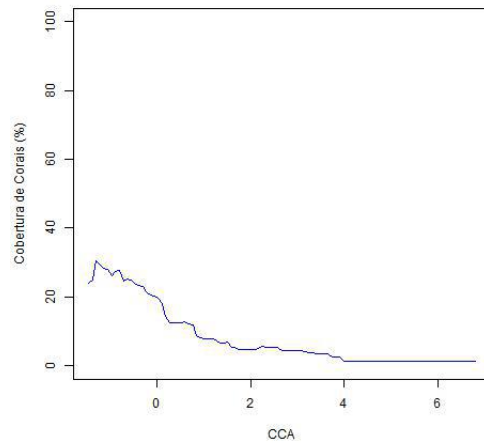
Nem sempre o mesmo método é o mais eficiente para a previsão de dados com estruturas e profundidades diferentes. Apesar do ótimo desempenho no modelo 1, o método de árvores ficou abaixo de outros no modelo 2, com dados mais restritos.

Em relação ao tipo de impacto causado pelas variáveis, os modelos apontaram uma relação predominantemente negativa entre a cobertura de espécies coralíneas e outros grupos presentes no bentos. Esse tipo de relação já era esperada, visto que o bentos é um ambiente de disputa natural de espaço entre as espécies.

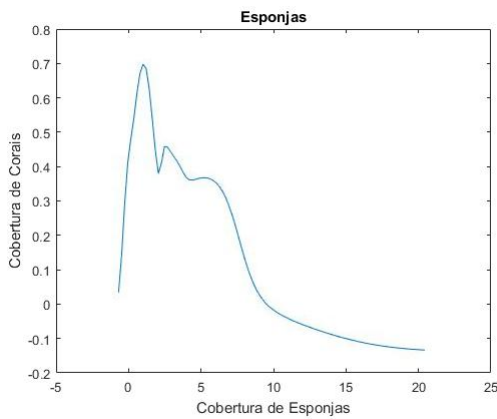
Os modelos de redes neurais e de árvores indicaram que a relação entre a cobertura de espécies e a cobertura de esponjas e *CCA* apresentaram relação negativa mais amena se comparado ao de algas epilíticas. Inclusive em determinadas circunstâncias o aumento desses grupos poderia levar a um crescimento da cobertura coralínea no bentos, como pode ser visto na figura 5.1.



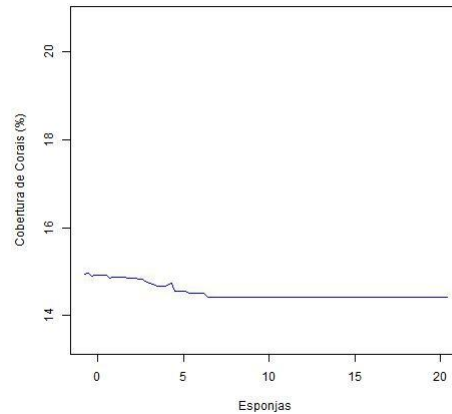
(a) Sensibilidade CCA em Redes Neurais.



(b) Sensibilidade CCA no modelo de árvores.



(c) Sensibilidade de Esponjas em Redes Neurais.



(d) Sensibilidade de Esponjas no modelo de árvores.

Figura 5.1: Análise de sensibilidade da cobertura de corais para Esponjas e CCA nos modelos de redes neurais e árvores de crescimento progressivo .

Fonte: Elaboração própria.

Em uma primeira instância espécies coralíneas disputam espaço no bentos com espécies de esponjas. Porém indícios apontam que este grupo realiza um importante trabalho no ecossistema recifal, retendo grande parte da matéria orgânica presente no ambiente como observado por DE GOIEJ *et al.*. Oceanos em geral são grandes desertos em termos de diversidade, o trabalho de retenção da matéria orgânica seria um dos pilares para a grande diversidade encontrada nos recifes [17].

Embora não sejam imunes a agente patógenos e doenças, as esponjas demonstraram maior resiliência a mudanças ambientais como o aumento da temperatura, acidificação do oceano e eutrofização da água. Esses organismos podem inclusive

aumentar sua abundância em um cenário de aquecimento global, com o a diminuição da competição por espaço nos recifes [2].

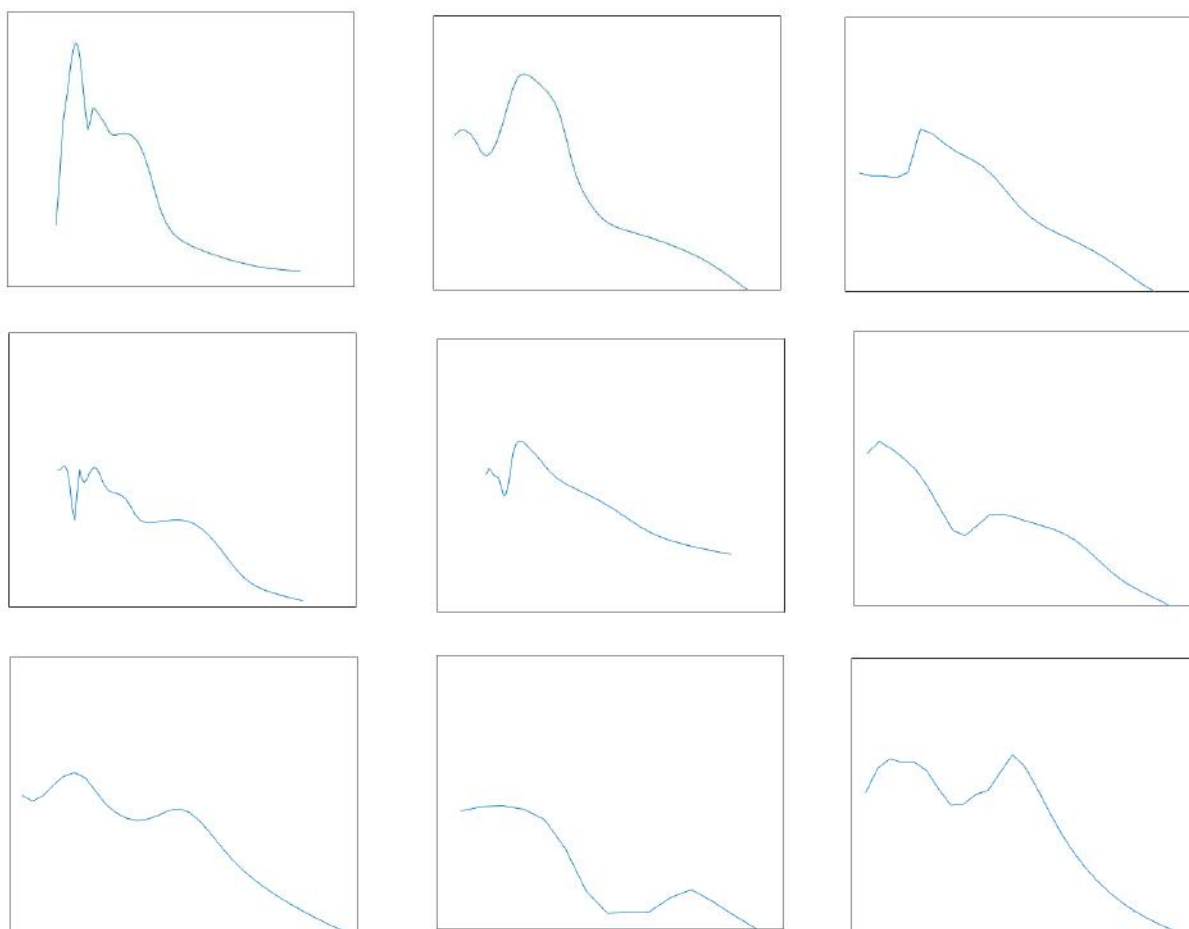


Figura 5.2: Variabilidade da análise de sensibilidade de Cobertura de Corais x Esponjas. Método de Redes Neurais.

Fonte: Elaboração própria.

Os resultados de variabilidade da análise de sensibilidade das esponjas podem ser vistos na figura 5.2. Apenas foram considerados os melhores resultados do modelo de redes neurais para montagem desse quadro, aqueles em que o R^2 foi superior a 0.7.

Como o método de Redes Neurais apresentou a maior variabilidade de resultados, como pode ser visto no quadro 4.12, os resultados da análise de sensibilidade também foram rodados diversas vezes, afim de garantir que o padrão de crescimento de corais com o aumento de cobertura de esponjas fosse considerável. Os resultados apontam relações diferentes entre cada modelo de redes neurais, porém em

todos é possível identificar essa correlação positiva em determinados concentrações de esponjas. Reforçando a teoria proposta por DE GOEIJ *et al.*[17].

Apesar do excelente desempenho no modelo 1, as árvores apresentaram como particularidade negativa a concentração da importância relativa em poucas variáveis, que pode ser vista na figura 4.15. As quatro variáveis de entrada principais concentraram mais de volta de 90% do poder de explicação do modelo, restando pouca relevância para as demais. Apesar de não comprometer a capacidade de previsão dos dados, afetou de certa forma o poder de explicação dessas variáveis, como pode ser visto na figuras 4.16 e 4.17, poucas variáveis de entrada puderam ser de fato ter seus impactos avaliados.

Por fim, uma comparação entre os métodos de regressão e de aprendizado estatístico pode levar em conta a utilização ou não do valor p , representando a significância estatística dos efeitos individuais das variáveis de influência. Métodos de regressão tradicionalmente levam em conta a utilização desse índice para a analisar os reais fatores de importância, enquanto métodos de aprendizado de máquina não o utilizam.

A utilização do valor p pode ser visto como uma vantagem para os métodos de regressão tradicionais, porém a própria utilização desse índice é questionada por outros autores. Para alguns pesquisadores, a utilização desse valor é superestimada, e seu resultado não representaria tanta certeza em relação a causa e efeito real de um determinado fator de influência [52].

A utilização nos modelos de aprendizado iterativo de cálculos para determinar a importância das variáveis explicativas são uma possível alternativa a utilização do valor- p . Em lugar de traçar uma linha divisória que determina se existe ou não influência real, as influências sobre a variável dependente são divididas entre os diversos fatores de entrada.

Capítulo 6

Conclusão

Este estudo teve como foco os corais situados no norte do Espírito Santo, próximos ao município de Conceição da Barra. Descobertos a poucos anos, essas formações se localizam no sul do banco de Abrolhos, em uma porção da região que foi afetada pelo desastre ocorrido em Mariana [45].

O objetivo geral deste estudo era analisar a saúde do banco de recifes localizados próximos ao município de Conceição da Barra. Estudos do bentos nos recifes do Espírito Santo mostraram um recife com grande diversidade de espécies, incluindo cobertura sudável de corais nativos ameaçados de extinção, como o *Mussimilia Hartii* e o *Mussimilia Braziliensis*.

Dentre todos os recifes localizados no banco de Abrolhos, foram encontrados a menor proporção de algas e macroalgas naqueles localizados no próximos ao município de Conceição da Barra. Esse fato em geral é um indicativo de boa saúde dos corais. Além disso, há indicação de que os corais do Espírito Santo possuem a maior cobertura coralínea dentre todos os recifes, e as espécies se distribuem de forma mais homogênea, sem predominância de uma sobre as outras.

Um dos objetivos secundários do estudo é a analisar a possível chegada dos rejeitos oriundos do desastre de Mariana. O despejo de toneladas de lama tóxica no rio Doce foi um acidente sem precedentes na história do Brasil. O tamanho dos danos econômicos, sociais e ambientais ainda está para ser calculado [26].

Inicialmente se pensava que a lama não chegaria até os recifes de Abrolhos, e que a dispersão levaria os rejeitos do derramamento para o sul, na direção do estado do Rio de Janeiro [44]. Porém evidências coletadas no MODIS indicam que uma parte do conteúdo despejado seguiu na direção norte, chegando até o banco de Abrolhos, onde os recifes descritos por MAZZEI *et al.* estão localizados.

As informações coletadas sobre as concentrações de peixes em 2016, um ano após o acidente, mostram que há uma menor quantidade de espécies e de famílias no período da chegada dos resíduos tóxicos, é possível que os peixes tenham migrado da região afetada. Sendo predadores naturais das algas, peixes exercem um papel

importante no equilíbrio ecológico do bentos recifal.

Porém as análises numéricas do bentos dos corais do Espírito Santo de 2016 e 2017 mostram corais mais saudáveis do que outras amostras dos bancos de Abrolhos. Apesar de uma possível diminuição do número de espécies de 2016 para 2017, as informações disponíveis não permitem afirmar que estatisticamente houve degradação dos corais.

Outro objetivo secundário foi a utilização de métodos de aprendizado de máquina para avaliar a cobertura de corais no recifes. Prever a distribuição de espécies marinhas é um dos pilares para elencar políticas de proteção, em especial para espécies coralíneas, cujos ecossistemas sensíveis a impactos locais e globais [59].

Os resultados dos métodos de previsão foram satisfatórios no modelo 1, que contava com todas as variáveis explicativas disponíveis. Já no modelo 2, quando não foram introduzidas as variáveis de cobertura, o desempenho foi inferior para todos os modelos.

Esse decréscimo da acurácia é natural, visto que outras variáveis de cobertura estão intrinsecamente relacionadas a cobertura de corais no recifes, disputando o mesmo espaço. Mesmo assim, avaliar a cobertura de corais a partir apenas do sensoriamento remoto pode se tornar uma ferramenta valiosa para o acompanhamento dos ecossistemas recifais, o torna-se bem menor em relação a visitas físicas aos locais [7, 59].

A combinação de algas epilíticas com cianobactérias foi indicada como estatisticamente relevante nos dois modelos de regressão, e assumiu o maior peso na importância relativa nos modelos de aprendizado iterativo. O fato destas espécies apresentarem a maior cobertura relativa em praticamente todos os recifes, destaca sua importância nestes ambientes.

Em uma escala local, a abundância de cianobactérias e algas epilíticas apresentou impacto significativo sobre a camada coralínea, em concordância com outros estudos. Logo, estratégias de preservação para a região de Abrolhos devem incluir medidas que regulem a cobertura dessas espécies que oferecem risco às espécies de corais.

Há indicações que para outras espécies que habitam o bentos, *CCA* e esponjas determinadas concentrações dessas espécies podem até mesmo auxiliar o crescimento de espécies coralíneas, suas concentrações portanto podem indicar boas condições de saúde dos recifes. Tal fato foi constatado por DE GOIEJ *et al.*, ressaltando a importância das esponjas nos ciclos de captação de nutrientes.

Por fim a comparação dos métodos estatísticos foi o último objetivo secundário proposto. No modelo 1 as previsões para o grupo de validação apresentaram alta correspondência com os dados reais. Nesse cenário o método de árvores de crescimento progressivo obtiveram o melhor desempenho.

Já no modelo 2, a acurácia de todos os métodos caiu consideravelmente. Tal

resultado já era previsto, pois os outros dados de cobertura que foram retirados do modelo 1 para o modelo 2 apresentam correlação direta com a cobertura coralínea. O desenvolvimento de métodos de previsão das espécies de corais que não contém dados de cobertura de outras espécies se tornaria uma ferramenta interessante para o acompanhamento de políticas de preservação.

6.1 Proposta de trabalhos futuros

Dentre as propostas advindas da elaboração deste estudo deve estar o acompanhamento da condição dos recifes de Conceição da Barra. Os reflexos do desastre que sucedeu em Marina ainda podem não ter atingido seu ápice, e a deterioração do ecossistemas recifal da região é uma possibilidade.

Os métodos de aprendizado estatístico oferecem um leque de opções ao estudo dos recifes. Previsões para outros grupos relevantes devem ser consideradas, dentre eles esponjas, algas epilíticas, cianobactérias, *CCA* e espécies específicas de corais.

Métodos de previsão podem também oferecer suporte para o acompanhamento da saúde dos ecossistemas. Outras variáveis podem fortalecer a capacidade do modelo. A influência das comunidades nativas da região no assoreamento de rios próximos aos recifes também pode ser um fator de importância sobre a camada coralínea.

Como demonstrado em outros estudos, peixes herbívoros são predadores naturais de algas e cianobactérias, e sua biomassa em geral está associada a uma diminuição de sua cobertura. A proteção dos parques marinhos que protegem a pesca predatória dessas espécies, oferece maior cobertura de espécies de corais nos recifes do que outras regiões [10].

O desenvolvimento de novas ferramentas que auxiliem os modelos de previsão também é recomendado. É possível o desenvolvimento de novas formas de modelagem, mais adequadas ao problema estudado.

6.2 Considerações Finais

A preservação do ecossistema de recifes localizados próximos a Conceição da Barra também deve levar em conta eventos como a tragédia de Mariana, que afetou de forma drástica o Rio Doce. Não se sabe se esses ecossistemas podem se recuperar do desastre anterior e ainda há o risco de novos incidentes semelhantes. É necessário que controle da exploração de minas subterrâneas e outras atividades exploratórias seja feito de forma a preservar os interesses da sociedade em detrimento do lucro privado.

A preservação da biodiversidade brasileira deveria ser visto como prioridade. O valor revertido em serviços econômicos para a região é incalculável, e provavelmente

é maior do que o das atividades econômicas predatórias realizadas, além do próprio valor intrínseco de todas as formas de vidas que ali desfrutam sua existência.

Referências Bibliográficas

- [1] 2016, “The IUCN Red List of Threatened Species”, *Science*, , n. 2016-1.
- [2] ACHLATIS, M., ZANDE, R. M. V. D., SCHÖNBERG, C. H. L., et al., 2017, “Sponge bioerosion on changing reefs : ocean warming poses physiological constraints to the success of a photosymbiotic excavating sponge”, , n. August, pp. 1–13.
- [3] ADGER, W. N., 2005, “Social-Ecological Resilience to Coastal Disasters”, *Science*, v. 309, n. 5737, pp. 1036–1039. doi: 10.1126/science.1112122.
- [4] AMADO-FILHO, G. M., MOURA, R. L., BASTOS, A. C., et al., 2012, “Rhodolith Beds Are Major CaCO₃ Bio-Factories in the Tropical South West Atlantic”, *Plos One*, v. 7, n. 4, pp. 5–10. doi: 10.1371/journal.pone.0035171.
- [5] AMBIENTE, M. D. M., *Panorama da conservação dos ecossistemas costeiros e marinhos no Brasil*. ISBN: 9788577381425.
- [6] BISHOP, J., 2010, “The economics of ecosystems and biodiversity: TEEB for business”, *Initiatives*, p. 27.
- [7] BOUMA, J. A., KUIK, O., DEKKER, A. G., 2011, “Science of the Total Environment Assessing the value of Earth Observation for managing coral reefs : An example from the Great Barrier Reef”, *Science of the Total Environment, The*, v. 409, n. 21, pp. 4497–4503. ISSN: 0048-9697.
- [8] BREIMAN, L., 2001, “Randomforest2001”, pp. 1–33. ISSN: 1098-6596. doi: 10.1017/CBO9781107415324.004.
- [9] BROCKE, H. J., POLERECKY, L., DE BEER, D., et al., 2015, “Organic matter degradation drives benthic cyanobacterial mat abundance on caribbean coral reefs”, *PLoS ONE*, v. 10, n. 5, pp. 1–19. doi: 10.1371/journal.pone.0125445.

- [10] BRUCE, T., MEIRELLES, P. M., GARCIA, G., et al., 2012, “Abrolhos bank reef health evaluated by means of water quality, microbial diversity, benthic cover, and fish biomass data”, *PLoS ONE*, v. 7, n. 6. ISSN: 19326203. doi: 10.1371/journal.pone.0036687.
- [11] CASTRO, C. B., PIRES, D., 2001, “Brazilian Coral Reefs: What we already know and what is still missing”, *Bulletin of Marine Science*, v. 69, n. 2, pp. 357–371.
- [12] CASTRO, C. B., SEGAL, B., PIRES, D. O., et al., 2005, “Distribution and Diversity of Coral Communities in the Abrolhos Reef Complex, Brazil”, pp. 19–39. ISSN: 0085-2538. doi: 10.1896/ci.cabs.2005.rap.
- [13] CHARPY, L., CASARETO, B. E., LANGLADE, M. J., et al., 2011, “Cyanobacteria in Coral Reef Ecosystems: A Review”, *Journal of Marine Biology*, v. 2012, pp. doi:10.1155/2012/259571. doi: 10.1155/2012/259571.
- [14] CONI, E. O. C., FERREIRA, C. M., MEIRELLES, P. M., et al., 2017, “Modeling abundance, growth, and health of the solitary coral *Scolymia wellsii* (Mussidae) in turbid SW Atlantic coral reefs”, *Marine Biology*, v. 164, n. 4, pp. 66. ISSN: 0025-3162. doi: 10.1007/s00227-017-3090-4.
- [15] COSTANZA, R., ARGE, R., GROOT, R. D., et al., 1997, “The value of the world ’ s ecosystem services and natural capital”, *Nature*, v. 387, n. May, pp. 253–260. ISSN: 0028-0836. doi: 10.1038/387253a0.
- [16] CUTLER, D. R., EDWARDS, T. C., BEARD, K. H., et al., 2007, “Random Forests for Classification in Ecology”, *Ecology*, v. 88, n. 11, pp. 2783–2792. ISSN: 0012-9658. doi: 10.1890/07-0539.1.
- [17] DE GOEIJ, J. M., VAN OEVELEN, D., VERMEIJ, M. J., et al., 2013, “Surviving in a marine desert: The sponge loop retains resources within coral reefs”, *Science*, v. 342, n. 6154, pp. 108–110. doi: 10.1126/science.1241981.
- [18] DE’ATH, G., FABRICIUS, K. E., 2000, “Classification and regression trees: A powerful yet simple technique for ecological data analysis”, *Ecology*, v. 81, n. 11, pp. 3178–3192. ISSN: 00129658. doi: 10.1890/0012-9658(2000)081[3178:CARTAP]2.0.CO;2.
- [19] DE’ATH, G., 2007, “Boosted trees for ecological modeling and prediction”, *Ecology*, v. 88, n. 1, pp. 243–251. doi: 10.1890/0012-9658(2007)88[243:BTFFEMA]2.0.CO;2.

- [20] DO CARMO, F. F., KAMINO, L. H. Y., JUNIOR, R. T., et al., 2017, “Fundão tailings dam failures: The environment tragedy of the largest technological disaster of Brazilian mining in global context”, *Perspectives in Ecology and Conservation*, v. 15, n. 3, pp. 145–151. ISSN: 25300644. doi: 10.1016/j.pecon.2017.06.002.
- [21] DONALD S. FRANKEL, S. L. F., 1996, “Application of neural networks to flow cytometry data analysis and real-time cell classification”, *Cytometry*, v. 23, n. 4, pp. 290–302.
- [22] EASTWOOD, E. K., CLARY, D. G., MELNICK, D. J., 2017, “Coral reef health and management on the verge of a tourism boom: A case study from Miches, Dominican Republic”, *Ocean & Coastal Management*, v. 138, pp. 192–204. ISSN: 09645691. doi: 10.1016/j.ocecoaman.2017.01.023.
- [23] ELITH, J., LEATHWICK, J. R., HASTIE, T., 2008, “A working guide to boosted regression trees”, *Journal of Animal Ecology*, v. 77, n. 4, pp. 802–813. ISSN: 00218790. doi: 10.1111/j.1365-2656.2008.01390.x.
- [24] ELLIFF, C. I., SILVA, I. R., 2017, “Coral reefs as the first line of defense : Shoreline protection in face of climate change”, *Marine Environmental Research*, v. 127, pp. 148–154. ISSN: 0141-1136. doi: 10.1016/j.marenvres.2017.03.007.
- [25] FARACO, R. A. L., PIRES, P. C., CRUZ, F. R. B., 1998, “Minimização do erro no algoritmo *back-propagation* aplicado ao problema de manutenção de motores”, .
- [26] FERNANDES, G. W., GOULART, F. F., RANIERI, B. D., et al., 2016, “Deep into the mud: ecological and socio-economic impacts of the dam breach in Mariana, Brazil”, *Natureza e Conservacao*, v. 14, n. 2, pp. 35–45. ISSN: 16790073. doi: 10.1016/j.ncon.2016.10.003.
- [27] FERNANDES, G. W., GOULART, F. F., RANIERI, B. D., et al., 2016, “Deep into the mud: ecological and socio-economic impacts of the dam breach in Mariana, Brazil”, *Natureza e Conservacao*, v. 14, n. 2, pp. 35–45. ISSN: 16790073. doi: 10.1016/j.ncon.2016.10.003.
- [28] FLOMBAUM, P., GALLEGOS, J. L., GORDILLO, R. A., et al., 2013, “Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*”, *Pnas*, v. 110, n. 24, pp. 9824–9829. ISSN: 1091-6490. doi: 10.1073/pnas.1307701110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1307701110.

- [29] FRANÇA, S., CABRAL, H. N., 2015, “Predicting fish species richness in estuaries: Which modelling technique to use?” *Environmental Modelling and Software*, v. 66, pp. 17–26.
- [30] FRANCINI-FILHO, R. B., MOURA, R. L., THOMPSON, F. L., et al., 2008, “Diseases leading to accelerated decline of reef corals in the largest South Atlantic reef complex (Abrolhos Bank, eastern Brazil)”, *Marine Pollution Bulletin*, v. 56, n. 5, pp. 1008–1014. ISSN: 0025326X. doi: 10.1016/j.marpolbul.2008.02.013.
- [31] FRANCINI-FILHO, R. B., CONI, E. O. C., MEIRELLES, P. M., et al., 2013, “Dynamics of Coral Reef Benthic Assemblages of the Abrolhos Bank, Eastern Brazil: Inferences on Natural and Anthropogenic Drivers”, *PLoS ONE*, v. 8, n. 1, pp. 1–12. ISSN: 19326203. doi: 10.1371/journal.pone.0054260.
- [32] FREUND, Y., SCHAPIRE, R. E., 1995, “A decision-theoretic generalization of on-line learning and an application to boosting”, v. 139, pp. 23–37. ISSN: 00220000. doi: 10.1007/3-540-59119-2_166.
- [33] GARSON, G. D., 1991, “Interpreting neural-network connection weights.” *Ai Expert*, v. 6(7), n. 11, pp. 47–51.
- [34] GEVREY, M., DIMOPOULOS, I., LEK, S., 2003, “Review and comparison of methods to study the contribution of variables in artificial neural network models”, *Ecological Modelling*, v. 160, n. 3, pp. 249–264. ISSN: 03043800. doi: 10.1016/S0304-3800(02)00257-0.
- [35] GOH, A. T. C., 1995, “Back-propagation neural networks for modeling complex systems”, *Artificial Intelligence in Engineering*, v. 9, n. 3, pp. 143–151. ISSN: 09541810. doi: 10.1016/0954-1810(94)00011-S.
- [36] GOMES, L. E. D. O., CORREA, L. B., SÁ, F., et al., 2017, “The impacts of the Samarco mine tailing spill on the Rio Doce estuary, Eastern Brazil”, *Marine Pollution Bulletin*, v. 120, n. 1-2, pp. 28–36. ISSN: 18793363. doi: 10.1016/j.marpolbul.2017.04.056.
- [37] HATJE, V., PEDREIRA, R. M. A., DE REZENDE, C. E., et al., 2017, “The environmental impacts of one of the largest tailing dam failures worldwide”, *Scientific Reports*, v. 7, n. 1, pp. 10706. ISSN: 2045-2322. doi: 10.1038/s41598-017-11143-x.
- [38] HUGHES, T. P., 1994. “Catastrophes, Phase Shifts.pdf”. ISSN: 0036-8075.

- [39] HUGHES, T., KERRY, J., ÁLVAREZ-NORIEGA, M., et al., 2017, “Global warming and recurrent mass bleaching of corals”, *Nature*. ISSN: 0028-0836. doi: 10.1038/nature21707.
- [40] KITTINGER, J. N., TENEVA, L. T., KOIKE, H., 2015, “From reef to table: Social and ecological factors affecting coral reef fisheries, artisanal seafood supply chains, and seafood security”, *PLoS ONE*, v. 10, n. 8, pp. 1–24.
- [41] LEATHWICK, J. R., ELITH, J., FRANCIS, M. P., et al., 2006, “Variation in demersal fish species richness in the oceans surrounding New Zealand: An analysis using boosted regression trees”, *Marine Ecology Progress Series*, v. 321, pp. 267–281. ISSN: 01718630. doi: 10.3354/meps321267.
- [42] LEGENDRE, P., LEGENDRE, L., 1998, *Numerical Ecology*. Elsevier.
- [43] LEK, S., BELAUD, A., DIMOPOULOS, Y., et al., 1995, “Improved estimation, using neural networks, of the food consumption of fish populations”, v. 46 (01).
- [44] MARTA-ALMEIDA, M., MENDES, R., AMORIM, F. N., et al., 2016, “Fundão Dam collapse: Oceanic dispersion of River Doce after the greatest Brazilian environmental accident”, *Marine Pollution Bulletin*, v. 112, n. 1-2, pp. 359–364. ISSN: 18793363. doi: 10.1016/j.marpolbul.2016.07.039.
- [45] MAZZEI, E. F., BERTONCINI, A. A., PINHEIRO, H. T., et al., 2016, “Newly discovered reefs in the southern Abrolhos Bank, Brazil: Anthropogenic impacts and urgent conservation needs”, *Marine Pollution Bulletin*, v. 114, n. 1, pp. 123–133. ISSN: 0025326X. doi: 10.1016/j.marpolbul.2016.08.059.
- [46] MIRANDA, L. S., MARQUES, A. C., 2016, “Hidden impacts of the Samarco mining waste dam collapse to Brazilian marine fauna – an example from the staurozoans (Cnidaria)”, *Biota Neotropica*, v. 16, n. 2, pp. 1–3. ISSN: 1676-0611. doi: 10.1590/1676-0611.
- [47] MMA, M. D. M. A., 2010, *Áreas Aquáticas Protegidas como Instrumento de Gestão Pesqueira*. Brasília. ISBN: 9788577381425.
- [48] MMA, M. D. M. A., 2010, *Panorama da conservação dos ecossistemas costeiros e marinhos no Brasil*. Brasília. ISBN: 9788577381425.
- [49] MOURA, R. L., SECCHIN, N. A., AMADO-FILHO, G. M., et al., 2013, “Spatial patterns of benthic megahabitats and conservation planning in the

Abrolhos Bank”, *Continental Shelf Research*, v. 70, pp. 109–117. ISSN: 02784343. doi: 10.1016/j.csr.2013.04.036.

- [50] NATIONS, U., PROGRAMME, E., 2004, “Successes and Challenges in the Management of Coral Reef Marine Protected Areas”, , n. 176.
- [51] NEVES, A. C. D. O., NUNES, F. P., DE CARVALHO, F. A., et al., 2016, “Neglect of ecosystems services by mining, and the worst environmental disaster in Brazil”, *Natureza e Conservacao*, v. 14, n. 1, pp. 24–27. ISSN: 16790073. doi: 10.1016/j.ncon.2016.03.002.
- [52] NUZZO, R., 2014, “Statistical errors: P values, the ”gold standard” of statistical validity, are not as reliable as many scientists assume.” *Nature*, v. 506, n. 7487, pp. 150–152.
- [53] O’HARA, T. D., SCHLACHER, T. A., ROWDEN, A. A., et al., 2016, “Data Analysis Considerations”. In: Clark, M. R., Consalvey, M., Rowden, A. A. (Eds.), *Biological Sampling in the Deep Sea*, Wiley Blackwell, cap. 17, pp. 386–403.
- [54] OLDEN, J. D., JACKSON, D. A., 2002, “Illuminating the ”black box”: A randomization approach for understanding variable contributions in artificial neural networks”, *Ecological Modelling*, v. 154, n. 1-2, pp. 135–150. ISSN: 03043800. doi: 10.1016/S0304-3800(02)00064-9.
- [55] OLDEN, J. D., JOY, M. K., DEATH, R. G., 2004, “An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data”, *Ecological Modelling*, v. 178, n. 3-4, pp. 389–397. ISSN: 03043800. doi: 10.1016/j.ecolmodel.2004.03.013.
- [56] OLDEN, J. D., LAWLER, J. J., POFF, N. L., et al., 2014, “Machine Learning Methods Without Tears : A Primer for Ecologists MACHINE LEARNING METHODS WITHOUT TEARS : A PRIMER FOR ECOLOGISTS”, v. 83, n. 2, pp. 171–193.
- [57] PAERL, H. W., PAUL, V. J., 2012, “Climate change: Links to global expansion of harmful cyanobacteria”, *Water Research*, v. 46, n. 5, pp. 1349–1363. doi: 10.1016/j.watres.2011.08.002.
- [58] PRASAD, A. M., IVERSON, L. R., LIAW, A., 2006, “Newer classification and regression tree techniques: Bagging and random forests for ecological prediction”, *Ecosystems*, v. 9, n. 2, pp. 181–199. ISSN: 14329840. doi: 10.1007/s10021-005-0054-1.

- [59] ROOPER, C. N., ZIMMERMANN, M., PRESCOTT, M. M., 2017, *Deep Sea Research Part I: Oceanographic Research Papers*, v. 126, n. January, pp. 148–161.
- [60] RUMELHART, 1986, “Learning representations by back-propagating errors”, *Nature*, v. 323, n. 9.
- [61] SALCEDO-SANZ, S., CUADRA, L., VERMEIJ, M. J., 2016, “A review of Computational Intelligence techniques in coral reef-related applications”, *Ecological Informatics*, v. 32, pp. 107–123. ISSN: 15749541. doi: 10.1016/j.ecoinf.2016.01.008.
- [62] SANTOS, E. C., ARMAS, E. D., CROWLEY, D., et al., 2014, “Artificial neural network modeling of microbial community structures in the Atlantic Forest of Brazil”, *Soil Biology and Biochemistry*, v. 69, pp. 101–109. ISSN: 00380717. doi: 10.1016/j.soilbio.2013.10.049.
- [63] SCHAPIRE, R. E., 2003, “The boosting approach to machine learning: an overview”, *Nonlinear Estimation and Classification*, v. 171, pp. 149–171. ISSN: 09300325. doi: 10.1.1.24.5565.
- [64] SEGURA, F. R., NUNES, E. A., PANIZ, F. P., et al., 2016, “Potential risks of the residue from Samarco’s mine dam burst (Bento Rodrigues, Brazil)”, *Environmental Pollution*, v. 218, pp. 813–825. doi: 10.1016/j.envpol.2016.08.005.
- [65] SILVEIRA, C. B., CAVALCANTI, G. S., WALTER, J., et al., 2017, “Microbial processes driving coral reef organic carbon flow and productivity”, *FEMS Microbiology Reviews*, v. Accepted, n. April, pp. 1–21. ISSN: 1574-6976. doi: 10.1093/femsre/fux018.
- [66] SPALDING, M., BURKE, L., WOOD, S. A., et al., 2017, “Mapping the global value and distribution of coral reef tourism”, *Marine Policy*, v. 82, n. January, pp. 104–113. doi: 10.1016/j.marpol.2017.05.014.
- [67] STOCKS, K. I., STOUT, N. J., SHANK, T. M., 2016. “Information and Management Strategies for Deep-Sea Biology” . .
- [68] VALENTIN, J. L., *Ecologia Numérica, uma Introdução a Análise Multivariada de Dados Ecológicos* . Segunda edição ed. Brasília, Editora Interciência.
- [69] VEIGA, A. K., CARTOLANO, E. A., SARAIVA, A. M., 2014, “Data quality control in biodiversity informatics: The case of species occurrence

data”, *IEEE Latin America Transactions*, v. 12, n. 4, pp. 683–693. ISSN: 15480992. doi: 10.1109/TLA.2014.6868870.

- [70] WALTER, J. M., TSCHOEKE, D. A., MEIRELLES, P. M., et al., 2016, “Taxonomic and functional metagenomic signature of turfs in the Abrolhos reef system (Brazil)”, *PLoS ONE*, v. 11, n. 8, pp. 1–20. ISSN: 19326203. doi: 10.1371/journal.pone.0161168.
- [71] WOOD, S. N., AUGUSTIN, N. H., 2002, “GAMs with integrated model selection using penalized regression splines and applications to environmental modelling”, *Ecological Modelling*, v. 157, n. 2-3, pp. 157–177.
- [72] ZANEVELD, J. R., BURKEPILE, D. E., SHANTZ, A. A., et al., 2016, “Overfishing and nutrient pollution interact with temperature to disrupt coral reefs down to microbial scales”, *Nature Communications*, v. 7, n. May, pp. 11833.