

Spring 5-6-2019

## **Bias Reduction in Machine Learning Classifiers for Spatiotemporal Analysis of Coral Reefs using Remote Sensing Images**

Justin J. Gapper  
*Chapman University*, [gappe102@mail.chapman.edu](mailto:gappe102@mail.chapman.edu)

Follow this and additional works at: [https://digitalcommons.chapman.edu/cads\\_dissertations](https://digitalcommons.chapman.edu/cads_dissertations)



Part of the [Environmental Indicators and Impact Assessment Commons](#), [Longitudinal Data Analysis and Time Series Commons](#), [Oceanography Commons](#), and the [Statistical Models Commons](#)

---

### **Recommended Citation**

J. J. Gapper, "Bias reduction in machine learning classifiers for spatiotemporal analysis of coral reefs using remote sensing images," Ph.D. dissertation, Chapman University, Orange, CA, 2019. <https://doi.org/10.36837/chapman.000078>

This Dissertation is brought to you for free and open access by the Dissertations and Theses at Chapman University Digital Commons. It has been accepted for inclusion in Computational and Data Sciences (PhD) Dissertations by an authorized administrator of Chapman University Digital Commons. For more information, please contact [laughtin@chapman.edu](mailto:laughtin@chapman.edu).

Bias Reduction in Machine Learning Classifiers for  
Spatiotemporal Analysis of Coral Reefs using Remote Sensing Images

A Dissertation by

Justin J. Gapper

Chapman University

Orange, CA

Schmid College of Science and Technology

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Computational and Data Sciences

May 2019

Committee in charge:

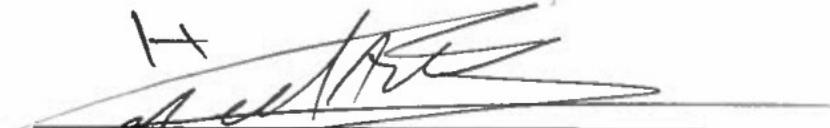
Hesham El-Askary, Ph.D., Chair

Erik Linstead, Ph.D.

Thomas Piechota, Ph.D.

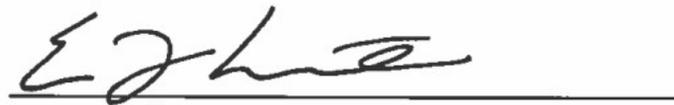


The dissertation of Justin J. Gapper is approved.



---

Hesham El-Askary, Ph.D., Committee Chair



---

Erik Linstead, Ph.D



---

Thomas Piechota, Ph.D

May 2019

Bias Reduction in Machine Learning Classifiers for  
Spatiotemporal Analysis of Coral Reefs using Remote Sensing Images

Copyright © 2019

by Justin J. Gapper

## ACKNOWLEDGEMENTS

I would first like to acknowledge and extend my utmost gratitude to the chair of my advisory committee, Hesham El-Askary, Ph.D. Incorporating his original ideas and recommendations were essential to the success of this research. Dr. El-Askary's constant feedback and encouragement made completing this work possible. In addition, Dr. El-Askary's work ethic and passion had an immense impact on this work as well as me personally. Finally, I appreciate Dr. El-Askary's honest candor and insistence on producing only the highest quality scientific research.

I would also like to thank Dr. Linstead and Dr. Piechota for their continual input and support. Dr. Linstead provided valuable feedback on my code and algorithmic approaches for which I am appreciative. Dr. Piechota supplied scientific contributions particularly with respect to climatology and providing feedback on responses to journal reviewer comments.

I am indebted to and infinitely grateful for my family. First, for the patience and understanding of Judah and Titus during this intense time of work, study, and research. I hope this can be an illustration of perseverance and grit for my sons as they continue to grow and learn. Most of all I want to expound upon the appreciation I have for my beautiful and brilliant wife, Sarah. I want to thank her for her love, support, hard work, and dedication throughout this journey. Sarah and I are an inseparable team achieving these accomplishments jointly as one unit. Without her partnership I would not have been able to complete this research or attain any of these successes.

Finally, I would like to thank God for providing me with the grit, knowledge, and creativity needed to complete this scientific research of His incredible creation.

# DEDICATION

To Sarah, Judah, and Titus

“The only thing you have to know is you can learn anything.”

“If any of you lacks wisdom, you should ask God,  
who gives generously to all without finding fault, and it will be given to you”

James 1:5

# **ABSTRACT**

Bias Reduction in Machine Learning Classifiers for  
Spatiotemporal Analysis of Coral Reefs using Remote Sensing Images

by Justin J. Gapper

This dissertation is an evaluation of the generalization characteristics of machine learning classifiers as applied to the detection of coral reefs using remote sensing images. Three scientific studies have been conducted as part of this research: 1) Evaluation of Spatial Generalization Characteristics of a Robust Classifier as Applied to Coral Reef Habitats in Remote Islands of the Pacific Ocean 2) Coral Reef Change Detection in Remote Pacific Islands using Support Vector Machine Classifiers 3) A Generalized Machine Learning Classifier for Spatiotemporal Analysis of Coral Reefs in the Red Sea. The aim of this dissertation is to propose and evaluate a methodology for developing a robust machine learning classifier that can effectively be deployed to accurately detect coral reefs at scale. The hypothesis is that Landsat data can be used to train a classifier to detect coral reefs in remote sensing imagery and that this classifier can be trained to generalize across multiple sites. Another objective is to identify how well different classifiers perform under the generalized conditions and how unique the spectral signature of coral is as environmental conditions vary across observation sites. A methodology for validating the generalization performance of a classifier to unseen locations is proposed and implemented (Controlled Parameter Cross-Validation,). Analysis is performed using satellite imagery from nine

different locations with known coral reefs (six Pacific Ocean sites and three Red Sea sites). Ground truth observations for four of the Pacific Ocean sites and two of the Red Sea sites were used to validate the proposed methodology. Within the Pacific Ocean sites, the consolidated classifier (trained on data from all sites) yielded an accuracy of 75.5% (0.778 AUC). Within the Red Sea sites, the consolidated classifier yielded an accuracy of 71.0% (0.7754 AUC). Finally, long-term change detection analysis is conducted for each of the sites evaluated. In total, over 16,700 km<sup>2</sup> was analyzed for benthic cover type and cover change detection analysis. Within the Pacific Ocean sites, decreases in coral cover ranged from 25.3% reduction (Kingman Reef) to 42.7% reduction (Kiritimati Island). Within the Red Sea sites, decrease in coral cover ranged from 3.4% (Umluj) to 13.6% (Al Wajh).

# TABLE OF CONTENTS

	<u>Page</u>
<b>ACKNOWLEDGEMENTS .....</b>	<b>IV</b>
<b>DEDICATION.....</b>	<b>VI</b>
<b>ABSTRACT.....</b>	<b>VII</b>
<b>TABLE OF CONTENTS .....</b>	<b>IX</b>
<b>LIST OF TABLES .....</b>	<b>XII</b>
<b>LIST OF FIGURES .....</b>	<b>XIII</b>
<b>1. INTRODUCTION.....</b>	<b>1</b>
1.1 Goals of this Research .....	1
1.2 2030 Agenda for Sustainable Development .....	3
1.3 Paris Agreement.....	10
1.4 Sendai Framework for Disaster Risk Reduction.....	11
1.5 World Economic Forum: The Global Risks Report .....	12
<b>2. WATER COLUMN CORRECTION/DEPTH INVARIANT INDEX.....</b>	<b>15</b>
2.1 Introduction.....	15
2.2 Data Used.....	17
2.3 Methodology.....	19
2.3.1 Cloud and Quality Mask .....	20
2.3.2 Water Mask.....	21
2.3.3 Atmospheric Correction.....	21
2.3.4 Water Column Correction.....	22
2.4 Output Parameters.....	25
2.4.1 Image Correction and Deep-Water AOI Parameters .....	25
2.4.2 Ratio of Attenuation.....	27
2.5 Results and Discussion .....	30
2.6 Conclusion .....	39
<b>3. FIRST STUDY: EVALUATION OF SPATIAL GENERALIZATION CHARACTERISTICS OF A ROBUST CLASSIFIER AS APPLIED TO CORAL REEF HABITATS IN REMOTE ISLANDS OF THE PACIFIC OCEAN .....</b>	<b>41</b>
3.1 Introduction.....	41
3.2 Materials and Methods.....	44
3.2.1 Remote Sensing Data.....	44

3.2.2	Study Sites .....	46
3.3	Methodology .....	49
3.3.1	Cloud and Quality Mask .....	51
3.3.2	Water Mask .....	52
3.3.3	Atmospheric Correction .....	52
3.3.4	Water Column Correction .....	53
3.4	Results .....	54
3.4.1	Generalization Performance by Site .....	54
3.4.2	Quantitative Assessment of Site-Specific Generalization .....	64
3.4.3	Robust Combined Model .....	67
3.5	Discussion .....	69
3.5.1	Spectral Signature Generalization Properties for Coral Reef Classification .....	69
3.5.2	Methodology Benefits and Challenges .....	71
3.6	Conclusion .....	71
<b>4.</b>	<b>SECOND STUDY: CORAL REEF CHANGE DETECTION IN REMOTE PACIFIC ISLANDS USING SUPPORT VECTOR MACHINE CLASSIFIERS</b> .....	<b>74</b>
4.1	Introduction .....	74
4.2	Materials and Methods .....	78
4.2.1	Satellite Data Used .....	78
4.2.2	Ground Truth Data Used for Training and Validation .....	80
4.2.3	Sites .....	82
4.3	Methodology .....	84
4.3.1	Cloud Mask .....	86
4.3.2	Land Mask .....	87
4.3.3	Atmospheric Correction and Water Column Correction .....	87
4.3.4	SVM Site Application, Validation, and Change Analysis .....	88
4.4	Results .....	91
4.4.1	Classification Accuracy by Site .....	91
4.4.2	Consolidated Model Robust to Site-Specific Bias .....	94
4.5	Change Detection Analysis .....	97
4.5.1	Palmyra Atoll .....	97
4.5.2	Kingman Reef .....	100
4.5.3	Baker Island Atoll .....	103
4.5.4	Howland Island .....	106
4.5.5	Tabuaeran Island .....	109
4.5.6	Kiritimati Island .....	112
4.6	Discussion .....	116
4.7	Conclusion .....	119

<b>5. THIRD STUDY: A GENERALIZED MACHINE LEARNING CLASSIFIER FOR SPATIOTEMPORAL ANALYSIS OF CORAL REEFS IN THE RED SEA</b>	<b>121</b>
5.1 Introduction.....	121
5.2 Materials and Methods.....	123
5.2.1 Study Area .....	123
5.2.2 Data Used and Preprocessing Steps.....	124
5.2.3 Preprocessing .....	125
5.2.4 Generalized Machine Learning Classifier .....	126
5.3 Results.....	130
5.3.1 Classification Accuracy by Site.....	130
5.3.2 Consolidated Model Robust to Site-Specific Bias.....	133
5.3.3 Change Detection Analysis.....	135
5.4 Discussion.....	145
5.5 Conclusion .....	147
<b>6. CONCLUSION .....</b>	<b>149</b>
<b>7. REFERENCES.....</b>	<b>154</b>

# LIST OF TABLES

	<u>Page</u>
Table 1-1: SDG 13 Targets and Indicators .....	5
Table 1-2: SDG 14 Targets and Indicators .....	8
Table 2-1: Selected Scenes for Study .....	18
Table 3-1: Selected Scenes for Study .....	45
Table 3-2: Selected Scenes for Study, Additional Information .....	46
Table 3-3: Assessment Metrics for Evaluation of Model Performance for Each Site and Consolidated Input. ....	66
Table 3-4: Confusion Matrices by Site and Consolidated Inputs. ....	66
Table 4-1: Image Capture Date and Ground Truth Observation Period for each Location. ....	79
Table 4-2: Satellite Data Summary.....	80
Table 4-3: SVM Classifier Performance Assessment by Site. ....	93
Table 4-4: Confusion Matrices by Site and Consolidated Inputs. ....	93
Table 4-5: Controlled Parameter Cross-Validation (CPCV) procedure results by site. ...	96
Table 4-6: Change Detection Analysis by Site. ....	115
Table 4-7: Classification accuracy of select additional learning algorithms. ....	117
Table 5-1: Summary of Data Used. ....	125
Table 5-2: SVM Classifier Performance Assessment by Site and Consolidated Model. ....	132
Table 5-3: Confusion Matrix by Site and the Consolidated Model. ....	133
Table 5-4: Change Detection Analysis by Site. ....	145

# LIST OF FIGURES

	<u>Page</u>
Figure 1-1: World Economic Forum Global Risks Landscape.....	13
Figure 2-1: Scene masking, atmospheric correction, and water column correction process flow. ....	20
Figure 2-2: Top: Comparison of mean deep-water radiance for Landsat 8 band 2, band 3, and band 4 for each of the 29 scenes selected for analysis. Bottom: Comparison of standard deviation of deep-water radiance. ....	26
Figure 2-3: Comparison of variance in radiance for each band. Certain scenes show more variance due to unique geographic characteristics.....	29
Figure 2-4: Comparison of covariance of spectral radiance between bands. ....	29
Figure 2-5: Comparison of the atmospheric and surface reflection correction constant (a) for each scene and pair of bands. ....	29
Figure 2-6: Comparison of ration of attenuation for scene and pair of bands. ....	30
Figure 2-7: Left to right: water masked scene, band 2/3 depth invariant index, band 2/4 depth invariant index, band 3/4 depth invariant index. Top to bottom: 013-055 Panama Coiba National Park, 013-057 Colombia Malpelo Fauna and Flora Sanctuary, 015-043 USA Everglades National Park, 016-053 Costa Rica Area de Conservación Guanacaste, 016-056 Costa Rica Cocos Island National Park. ....	33
Figure 2-8: Left to right: water masked scene, band 2/3 depth invariant index, band 2/4 depth invariant index, band 3/4 depth invariant index. Top to bottom: 018-047 Mexico Sian Ka’an, 018-048 Belize Belize Barrier Reef Reserve System, 018-060 Ecuador Galápagos Islands, 034-047 Mexico Archipiélago de Revillagigedo, 038-039 Mexico Sea of Cortez.....	34
Figure 2-9: Left to right: water masked scene, band 2/3 depth invariant index, band 2/4 depth invariant index, band 3/4 depth invariant index. Top to bottom: 069-063 Kiribati Phoenix Islands Protected Area, 073-042 USA Papahānaumokuākea, 083-074 France Lagoons of New Caledonia, 085-082 Australia Lord Howe Islands, 087-069 Solomon Islands East Rennell. ....	35
Figure 2-10: Left to right: water masked scene, band 2/3 depth invariant index, band 2/4 depth invariant index, band 3/4 depth invariant index. Top to bottom: 091-075 Australia Great Barrier Reef, 104-041 Japan Ogasawara Islands, 106-055 Palau Rock Islands Southern Lagoon, 114-066 Indonesia Komodo National Park, 114-080 Australia Ningaloo Coast.....	36

Figure 2-11: Left to right: water masked scene, band 2/3 depth invariant index, band 2/4 depth invariant index, band 3/4 depth invariant index. Top to bottom: 115-078 Australia Shark Bay, Western Australia, 116-054 Philippines Tubbataha Reefs Natural Park, 123-065 Indonesia Ujung Kulon National Park, 126-046 Viet Nam Gulf of Tonkin, 159-051 Yemen Socotra Archipelago. ....	37
Figure 2-12: Left to right: water masked scene, band 2/3 depth invariant index, band 2/4 depth invariant index, band 3/4 depth invariant index. Top to bottom: 161-067 Seychelles Aldabra Atoll, 167-079 South Africa Simangaliso Wetland Park, 170-047 Sudan Sanganeb Marine National Park, 213-063 Brazil Brazilian Atlantic.....	38
Figure 3-1 The location of the four sites (1) Palmyra Atoll (2) Kingman Reef (3) Baker Island Atoll (4) Howland Island .....	46
Figure 3-2 Scene masking, atmospheric correction, and water column correction process flow. ....	51
Figure 3-3: Palmyra Atoll: Plot of the posterior probability for belonging to the coral class for each pixel in the Palmyra Atoll site.....	55
Figure 3-4: Palmyra Atoll predicted class membership based on posterior probabilities.	55
Figure 3-5: Kingman Reef: Plot of the posterior probability for belonging to the coral class for each pixel in the Kingman Reef site.....	57
Figure 3-6: Kingman Reef predicted class membership based on posterior probabilities. ....	57
Figure 3-7: Baker Island Atoll: Plot of the posterior probability for belonging to the coral class for each pixel in the Baker Island Atoll site.....	60
Figure 3-8: Baker Island Atoll predicted class membership based on posterior probabilities.....	61
Figure 3-9: Howland Island: Plot of the posterior probability for belonging to the coral class for each pixel in the Howland Island site.....	63
Figure 3-10: Howland Island predicted class membership based on posterior probabilities.....	64
Figure 3-11: ROC Curve indicating model performance of the algorithm when applied to the consolidated ground truth set (AUC = 0.7298).....	69
Figure 4-1: The relative location of each of the six sites (1) Palmyra Atoll, (2) Kingman Reef, (3) Baker Island Atoll, (4) Howland Island, (5) Tabuaeran Island, and (6) Kiritimati Island.....	84

Figure 4-2: SVM classifier training and change analysis process flow for Palmyra Atoll, Kingman Reef, Baker Island Atoll, and Howland Island. ....	85
Figure 4-3: Robust SVM Classifier training and change analysis process flow for Tabuaeran Island and Kiritimati Island.....	86
Figure 4-4: Performance evaluation of the combined classifier using ROC Curve and AUC (0.778). ....	96
Figure 4-5: Posterior probability map for the Palmyra Atoll area of interest (top, 2001 and bottom, 2015).....	98
Figure 4-6: Predicted class map for the Palmyra Atoll area of interest (top, 2001 and bottom, 2015).....	99
Figure 4-7: Difference in predicted class membership map for the Palmyra Atoll area of interest for 2001 initial state compared to 2015 final state.....	99
Figure 4-8: Posterior probability map for the Kingman Reef area of interest (top, 2001 and bottom, 2015).....	101
Figure 4-9: Predicted class probability map for the Kingman Reef area of interest (top, 2001 and bottom, 2015). ....	102
Figure 4-10: Difference in predicted class membership map for the Kingman Reef area of interest for 2001 initial state compared to 2015 final state.....	103
Figure 4-11: Posterior probability map for the Baker Island Atoll area of interest (top, 2002 and bottom, 2014). ....	104
Figure 4-12: Predicted class map for the Baker Island Atoll area of interest (top, 2002 and bottom, 2014).....	105
Figure 4-13: Difference in predicted class membership map for the Palmyra Atoll area of interest for 2002 initial state compared to 2014 final state.....	106
Figure 4-14: Posterior probability map for the Howland Island area of interest (top, 2001 and bottom, 2015). ....	107
Figure 4-15: Predicted class map for the Howland Island area of interest (top, 2001 and bottom, 2015).....	108
Figure 4-16: Difference in predicted class membership map for the Howland Island area of interest for 2001 initial state compared to 2015 final state.....	109
Figure 4-17: Posterior probability map for the Tabuaeran Island area of interest (top, 2000 and bottom, 2014). ....	110

Figure 4-18: Predicted class map for the Tabuaeran Island area of interest (top, 2000 and bottom, 2014).....	111
Figure 4-19: Difference in predicted class membership map for the Tabuaeran Island area of interest for 2000 initial state compared to 2014 final state.....	112
Figure 4-20: Posterior probability map for the Kiritimati Island area of interest (top, 2002 and bottom, 2014). .....	113
Figure 4-21: Predicted class map for the Kiritimati Island area of interest (top, 2002 and bottom, 2014).....	114
Figure 4-22: Difference in predicted class membership map for the Kiritimati Island area of interest for 2002 initial state compared to 2014 final state.....	115
Figure 5-1: Relative location of the three locations under study (1) Gulf of Aqaba Location, (2) Umluj Location, (3) Al Wajh Location.....	124
Figure 5-2: SVM Classifier training (Gulf of Aqaba), validation (Umluj), and application of the robust classifier (Al Wajh).....	129
Figure 5-3: Temporal change detection analysis process flow. ....	130
Figure 5-4: Performance evaluation of the robust, combined classifier using ROC Curve and AUC for the Gulf of Aqaba and consolidated model.....	135
Figure 5-5: Posterior probability map for the Gulf of Aqaba area of interest (top, 2000 and bottom, 2018). .....	137
Figure 5-6: Predicted class membership map for the Gulf of Aqaba area of interest (top, 2000 and bottom, 2018). .....	138
Figure 5-7: Change detection analysis between 2000 and 2018 of the Gulf of Aqaba area of interest. ....	139
Figure 5-8: Posterior probability map for the Umluj area of interest (left, 2000 and right, 2018). .....	140
Figure 5-9: Predicted class membership map for the Umluj area of interest (left, 2000 and right, 2018).....	141
Figure 5-10: Change detection analysis between 2000 and 2018 of the Umluj area of interest.....	142
Figure 5-11: Posterior probability map for the Al Wajh area of interest (left, 2000 and right, 2018). .....	143
Figure 5-12: Predicted class membership map for the Al Wajh area of interest (left, 2000 and right, 2018).....	144

Figure 5-13: Change detection analysis between 2000 and 2018 of the Al Wajh area of interest..... 144

# 1. Introduction

## 1.1 Goals of this Research

Previous research concerning the analysis of coral reefs using remote sensing data has been limited to in-situ analysis. That is, these studies have isolated specific reefs or a small area of interest (AOI) in order to perform benthic habitat cover detection using remote sensing data. These studies achieve a remarkable level of accuracy particularly if based on high resolution imagery. Often, these analyses can achieve upwards of 90% class prediction accuracy when compared to ground truth observation data [1]. The recent abundance of high-resolution satellite platforms enables these scientific studies and their results are quite promising with respect to analyzing the individual reefs they target for the time period in which remote sensing data from their platform of choice is available. However, there are two challenges faced by these studies. First, they rely upon high resolution remote sensing data and second, they are spatially limited in scope.

Recent advances in technology have enabled high-resolution remote sensing imagery. While these satellites enable benthic habitat classification with a high degree of accuracy, they do not enable the historical archive necessary for long-term temporal change detection analysis. The Landsat missions, on the other hand, afford a rich archive of historical imagery. This data enables long-term change analysis, defined here as greater than 10-years. Yet, the Landsat platform is limited to medium resolution data both now

and historically. Even though the technology for much higher resolution sensors is available, Landsat has maintained a 30m x 30m pixel resolution. This was intentionally done in order to maintain parity with previous missions. As a result, the Landsat platform is the most common source of remote sensing data for change detection analysis of coral reefs as well as all other benthic and land cover types [1].

Secondly, previous research has been limited in scope spatially. An abundance of scientific studies are available which have isolated a specific reef or small AOI within which benthic cover type detection or change detection analysis is performed. These in-situ analyses are, by definition, limited in scope. Studies conducted in this way often yield accuracies, measured by the percentage of pixel cover types matching ground truth observations, of 85% or greater. Yet, the isolation of a specific area for both training and analysis necessarily mean that these analyses are often overfit to that specific location. That is, the classifiers used to evaluate the reef are both trained and tested using data from the same location. As a result, these classifiers will not perform well if applied to a new location. This is because the training and testing methodology employed to create the models causes them to be significantly overfit to the local conditions represented in the respective AOI. In this way, in-situ analyses rely upon site-specific biases that prevent them from generalizing to new locations. These classifiers memorize the site-specific geomorphology, fauna, and other local conditions at the specific site. Therefore, while the classifiers serve the purpose as it pertains to the specific location under analysis, they will not generalize to new locations.

The goal of this research is to develop a robust machine learning classifier that can generalize spatially beyond the scope of previous in-situ type analyses. The proposed methodology includes using remote sensing data from multiple sites and the associated ground truth data in order to develop a robust classifier that can generalize beyond any single AOI. In order to measure the effectiveness of the generalized classifier, the Controlled Parameter Cross-Validation (CPCV) evaluation procedure is proposed. This methodology accounts for site-specific information that may bias the results of standard train/test split or cross-validation methods and provides a more accurate assessment of how well the classifier is generalizing to new data.

## **1.2 2030 Agenda for Sustainable Development**

The 2030 Agenda for Sustainable Development was adopted by all United Nations (UN) Member States in 2015. This agenda seeks to build on the Millennium Development Goals and complete what they did not achieve. The agenda includes 17 Sustainable Development Goals (SDGs), 169 targets, and 232 indicators all aimed at establishing principles of sustainable development in national policies. The research proposed in this dissertation addresses two of these goals. In particular, SDG 13 concerning climate action and SDG 14 concerning life below water.

The objective of SDG 13 is to take urgent action to combat climate change and its impacts. To achieve this, SDG 13 asserts five different targets. Table 1-1 shows each of these targets and the associated indicator(s). Goal 13.3 asserts, “Improve education, awareness-raising and human and institutional capacity on climate change mitigation, adaptation, impact reduction and early warning.” The research proposed in this dissertation

performs two functions to contribute to this goal. First, it informs countries on mitigation, adaptation, and impact reduction data from which curricula can be based. This contributes to the first indicator, “Number of countries that have integrated mitigation, adaptation, impact reduction and early warning into primary, secondary and tertiary curricula.” Curricula must be based on scientific study and results. This research provides a methodology for developing results with respect to coral reef extent as well as evaluation of results from the proposed methodology. The research proposed in this dissertation addresses the second indicator, “Number of countries that have communicated the strengthening of institutional, systemic and individual capacity-building to implement adaptation, mitigation and technology transfer, and development actions.” By informing countries, particularly those with significant coastal habitats, of their exposure to climate change this research enables those countries to more effectively strengthen capacity-building to implement adaptation and mitigation actions. In addition, this research contributes to the indicators behind target 13.B, “Promote mechanisms for raising capacity for effective climate change-related planning and management in least developed countries and small island developing States, including focusing on women, youth and local and marginalized communities” by informing small developing island states such as Kiribati of their exposure to climate-change. Kiribati is a small island country with significant populations on Kiritimati Island and Tabuaeran Island both of which are included in this research. Informing this small island developing State of their exposure to climate-change risk addressed the associate SDG indicator, “Number of least developed countries and small island developing States that are receiving specialized support, and amount of support, including finance, technology and capacity-building, for mechanisms for raising

capacities for effective climate change-related planning and management, including focusing on women, youth and local and marginalized communities” by directly informing at least one such island country. In addition to these direct implications of this research there are several indirect impacts. For example, this research can be used to inform target 13.2, “Integrate climate change measures into national policies, strategies and planning” by informing policy makers of the specific, localized impact associated with their directives.

Table 1-1: SDG 13 Targets and Indicators

<b>Targets</b>	<b>Indicators</b>
13.1 Strengthen resilience and adaptive capacity to climate-related hazards and natural disasters in all countries	<p>13.1.1 Number of deaths, missing persons and persons affected by disaster per 100,000 people</p> <p>13.1.2 Number of countries with national and local disaster risk reduction strategies</p> <p>13.1.3 Proportion of local governments that adopt and implement local disaster risk reduction strategies in line with national disaster risk reduction strategies</p>
13.2 Integrate climate change measures into national policies, strategies and planning	13.2.1 Number of countries that have communicated the establishment or operationalization of an integrated policy/strategy/plan which increases their ability to adapt to the adverse impacts of climate change, and foster climate resilience and low greenhouse gas emissions development in a manner that does not threaten food production (including a national adaptation plan, nationally determined contribution, national communication, biennial update report or other)

Table 1-1 (cont.): SDG 13 Targets and Indicators

Targets	Indicators
13.3 Improve education, awareness-raising and human and institutional capacity on climate change mitigation, adaptation, impact reduction and early warning	<p>13.3.1 Number of countries that have integrated mitigation, adaptation, impact reduction and early warning into primary, secondary and tertiary curricula</p> <p>13.3.2 Number of countries that have communicated the strengthening of institutional, systemic and individual capacity-building to implement adaptation, mitigation and technology transfer, and development actions</p>
13.A Implement the commitment undertaken by developed-country parties to the United Nations Framework Convention on Climate Change to a goal of mobilizing jointly \$100 billion annually by 2020 from all sources to address the needs of developing countries in the context of meaningful mitigation actions and transparency on implementation and fully operationalize the Green Climate Fund through its capitalization as soon as possible	13.A.1 Mobilized amount of United States dollars per year starting in 2020 accountable towards the \$100 billion commitment
13.B Promote mechanisms for raising capacity for effective climate change-related planning and management in least developed countries and small island developing States, including focusing on women, youth and local and marginalized communities	13.B.1 Number of least developed countries and small island developing States that are receiving specialized support, and amount of support, including finance, technology and capacity-building, for mechanisms for raising capacities for effective climate change-related planning and management, including focusing on women, youth and local and marginalized communities

SDG 14 aims to conserve and sustainably use the oceans, seas, and marine resources for sustainable development. Table 1-2 shows each of the targets of SDG 14 and the associated indicator(s). Of the 10 SDG targets set to achieve this goal, at least three are directly related to the research presented in this dissertation. First, target 14.2 states, “By 2020, sustainably manage and protect marine and coastal ecosystems to avoid significant

adverse impacts, including by strengthening their resilience, and take action for their restoration in order to achieve healthy and productive oceans.” This target is enabled by identifying coastal ecosystems most susceptible to adverse impacts thereby enabling local governments to take action and protect their valuable coastal resources. The indicator associated with this target is 14.2.1, “Proportion of national exclusive economic zones managed using ecosystem-based approaches” which can only be enabled by identifying the ecosystems impacted by said approaches. Furthermore, target 14.5 calls for, “By 2020, conserve at least 10 per cent of coastal and marine areas, consistent with national and international law and based on the best available scientific information.” The scientific information delivered in these studies directly impacts this target and informs the associated indicator, “Coverage of protected areas in relation to marine areas,” by identifying the extent of one of the key marine areas that is called to be protected. Finally, target 14.A indicates, “Increase scientific knowledge, develop research capacity and transfer marine technology, taking into account the Intergovernmental Oceanographic Commission Criteria and Guidelines on the Transfer of Marine Technology, in order to improve ocean health and to enhance the contribution of marine biodiversity to the development of developing countries, in particular small island developing States and least developed countries.” The methodology proposed in this research directly contributes to this target through increasing scientific knowledge and developing research capacity which improves ocean health and marine biodiversity to Kiribati and similar small island developing States. In addition to these direct influences, the research presented in this dissertation address multiple targets and indicators associated with SDG 14 indirectly.

These indirect implications as well as the direct consequences of this research aggregate to a significant total influence that is ambitions with far-reaching impact.

Table 1-2: SDG 14 Targets and Indicators

<b>Targets</b>	<b>Indicators</b>
14.1 By 2025, prevent and significantly reduce marine pollution of all kinds, in particular from land-based activities, including marine debris and nutrient pollution	14.1.1 Index of coastal eutrophication and floating plastic debris density
14.2 By 2020, sustainably manage and protect marine and coastal ecosystems to avoid significant adverse impacts, including by strengthening their resilience, and take action for their restoration in order to achieve healthy and productive oceans	14.2.1 Proportion of national exclusive economic zones managed using ecosystem-based approaches
14.3 Minimize and address the impacts of ocean acidification, including through enhanced scientific cooperation at all levels	14.3.1 Average marine acidity (pH) measured at agreed suite of representative sampling stations
14.4 By 2020, effectively regulate harvesting and end overfishing, illegal, unreported and unregulated fishing and destructive fishing practices and implement science-based management plans, in order to restore fish stocks in the shortest time feasible, at least to levels that can produce maximum sustainable yield as determined by their biological characteristics	14.4.1 Proportion of fish stocks within biologically sustainable levels
14.5 By 2020, conserve at least 10 per cent of coastal and marine areas, consistent with national and international law and based on the best available scientific information	14.5.1 Coverage of protected areas in relation to marine areas

Table 2-2 (cont.): SDG 14 Targets and Indicators

Targets	Indicators
<p>14.6 By 2020, prohibit certain forms of fisheries subsidies which contribute to overcapacity and overfishing, eliminate subsidies that contribute to illegal, unreported and unregulated fishing and refrain from introducing new such subsidies, recognizing that appropriate and effective special and differential treatment for developing and least developed countries should be an integral part of the World Trade Organization fisheries subsidies negotiation</p>	<p>14.6.1 By 2020, prohibit certain forms of fisheries subsidies which contribute to overcapacity and overfishing, eliminate subsidies that contribute to illegal, unreported and unregulated fishing and refrain from introducing new such subsidies, recognizing that appropriate and effective special and differential treatment for developing and least developed countries should be an integral part of the World Trade Organization fisheries subsidies negotiation</p>
<p>14.7 By 2030, increase the economic benefits to Small Island developing States and least developed countries from the sustainable use of marine resources, including through sustainable management of fisheries, aquaculture and tourism</p>	<p>14.7.1 Sustainable fisheries as a percentage of GDP in small island developing States, least developed countries and all countries</p>
<p>14.A Increase scientific knowledge, develop research capacity and transfer marine technology, taking into account the Intergovernmental Oceanographic Commission Criteria and Guidelines on the Transfer of Marine Technology, in order to improve ocean health and to enhance the contribution of marine biodiversity to the development of developing countries, in particular small island developing States and least developed countries</p>	<p>14.A.1 Proportion of total research budget allocated to research in the field of marine technology</p>
<p>14.B Provide access for small-scale artisanal fishers to marine resources and markets</p>	<p>14.B.1 Progress by countries in the degree of application of a legal/regulatory/policy/institutional framework which recognizes and protects access rights for small-scale fisheries</p>

Table 2-2 (cont.): SDG 14 Targets and Indicators

Targets	Indicators
14.C Enhance the conservation and sustainable use of oceans and their resources by implementing international law as reflected in UNCLOS, which provides the legal framework for the conservation and sustainable use of oceans and their resources, as recalled in paragraph 158 of The Future We Want	14.C.1 Number of countries making progress in ratifying, accepting and implementing through legal, policy and institutional frameworks, ocean-related instruments that implement international law, as reflected in the United Nation Convention on the Law of the Sea, for the conservation and sustainable use of the oceans and their resources

### 1.3 Paris Agreement

The Paris Agreement was signed on April 22, 2016. The goal of this doctrine is to keep the increase in global average temperature to well below 2°C. Associated with this agreement are several articles stipulated in order to achieve the goal. First Article 8 states, “Parties recognize the importance of averting, minimizing and addressing loss and damage associated with the adverse effects of climate change, including extreme weather events and slow onset events, and the role of sustainable development in reducing the risk of loss and damage.” The deterioration of coral reefs as a result of temperature increases is a significant impact related to slow onset events. The research presented in this dissertation addresses this aspect of the article by informing the loss and damage to coral reefs related to temperature change. Second article 12 asserts, “Climate change education, training, public awareness, public participation and public access to information (Art 12) is also to be enhanced under the Agreement” and core to this is information regarding the impact of climate-change on coral reefs. Definitively measuring this impact significantly enhances the public awareness, public participation, and public access as it enables both the availability of data and the visibility into who and what is being impacted. The research methods proposed in this research directly contribute to this article of the Paris Agreement.

Finally, Article 14 of the Paris Agreement decrees, “A global stocktake, to take place in 2023 and every 5 years thereafter, will assess collective progress toward meeting the purpose of the Agreement in a comprehensive and facilitative manner. Its outcomes will inform Parties in updating and enhancing their actions and support and enhancing international cooperation.” This is directly supported by the research proposed in this dissertation which performs a significant stock-take of Pacific Ocean and Red Sea coral as well as enables a global stock-take of coral reefs.

#### **1.4 Sendai Framework for Disaster Risk Reduction**

The Sendai Framework for Disaster Risk Reduction aims for: “The substantial reduction of disaster risk and losses in lives, livelihoods and health and in the economic, physical, social, cultural and environmental assets of persons, businesses, communities and countries.” To this end, the research presented in this dissertation addresses the environmental assets of persons, businesses, communities, and countries. Specifically, the primary resource of many countries is their marine environment and, in particular, the coastal reefs marine environment. This ecosystem provides many communities with a local resource for income (tourism or otherwise) and sustenance (through fishing, research, or otherwise). Therefore, the health of this local resource is critical to the lives, livelihoods, and health of local communities economically, physically, socially, culturally, and environmentally. Particularly with respect to the Kiritimati Island and Tabuaeran Island communities as well as the small Red Sea villages which rely on the coral reefs for subsistence, the results of this research are imperative to inform policy and decision making.

## **1.5 World Economic Forum: The Global Risks Report**

The Global Risks Report is an annual evaluation by the World Economic Forum which looks at the foremost risks facing the world [2]. These risks range from weapons of mass destruction to terrorist attacks to financial crisis. There are several methods by which the World Economic Forum reports on these risks. One such methodology is a visualization called the Global Risks Landscape. This visualization plots each primary global risk on a scale by likelihood on the x-axis and impact on the y-axis. Therefore, global risks that are higher on the plot represent a larger impact while global risks that are to the right of the plot represent a larger likelihood. In 2019, all environmental risks were represented in the first quadrant as both most likely and most impactful, Figure 1-1.

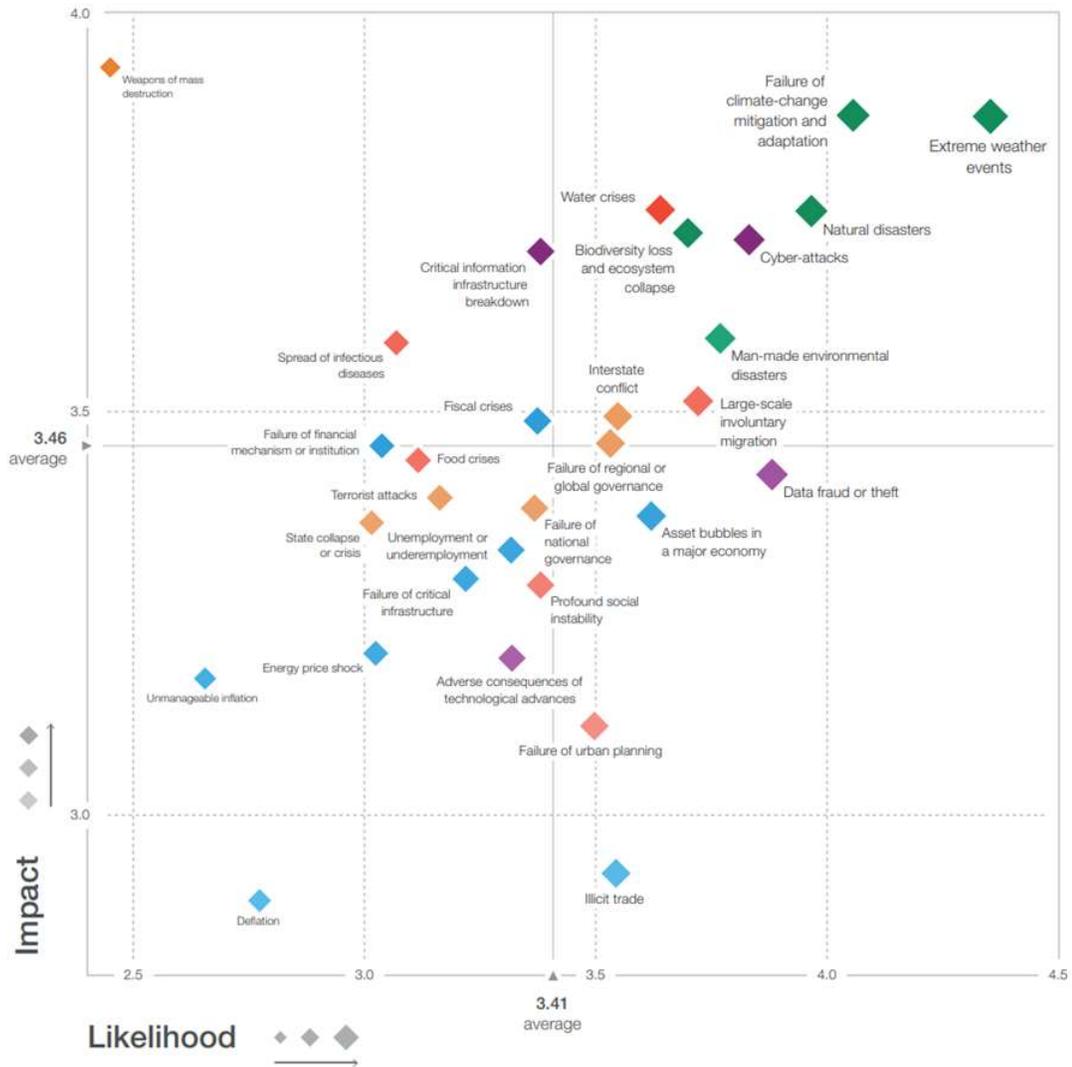


Figure 1-1: World Economic Forum Global Risks Landscape.

Furthermore, failure of climate-change mitigation and adaptation was both the second most likely risk and second highest impact risk. This risk represents the failure of governments and businesses to recognize and react to climate change resulting in catastrophic loss of biodiversity. Since 1970 species abundance is down by 60% and the loss of biodiversity is affecting health and socioeconomic development with implications for productivity and even regional security.

This dissertation directly addresses the failure of climate-change mitigation and adaptation risk outlined in the World Economic Forum Global Risks Report. Specifically, the research proposed identifies the impact that climate-change is having within coral reef ecosystems of both the Red Sea and Pacific Ocean. Coral reefs are the most biodiverse marine ecosystems and among the most biodiverse ecosystems in the world. Therefore, this environment is of utmost important when considering the threat of catastrophic loss of biodiversity. This research is an evaluation of the changing cover across a large extent of coral reef environments and therefore provides valuable information for the governments and businesses addressed in the Global Risks Report. Providing this insight provides data for these organizations to call attention to the need for climate-change mitigation and adaptation. Furthermore, this research is foundational to the global mapping and evaluation of coral reefs. Enabling a robust classifier that can generalize beyond site-specific deployment is the key to understanding and informing climate-change mitigation strategies on a global scale.

# 2. Water Column Correction/Depth Invariant Index

## 2.1 Introduction

Coral reefs are among the most complex and diverse marine ecosystems in the world [3]. However, these delicate ecosystems are under extreme threat due to numerous environmental and anthropogenic forces. Ocean acidification and mass bleaching events leading to large scale coral death is well documented [4] [5] [6] [7] [8] [9] [10] [11] [12]. Therefore, monitoring of these ecosystems is critical to inform policy and decision making for all agencies and at all levels. A comprehensive plan for evaluating the health of these delicate ecosystems is a complex endeavor that can only be achieved through the combined efforts of both detailed in-situ analyses and efficient, large scale analyses. Traditionally, reefs have been monitored using expensive and tedious underwater surveying techniques [13] [14] [15] [16] [17] [18] [19] [20] that, by definition, cannot cover large areas [21]. These traditional techniques have several drawbacks that restrict their use and the relevance of outcomes. These are (1) cost-related: detailed, continuous monitoring of coral reefs by field survey is expensive and substantial reef areas are located in developing countries with limited resources; (2) scale-related: reefs are highly heterogeneous systems [22] [23], therefore, even with sufficient resources, monitoring programs provide scattered information in time and space, with some areas being more intensively sampled than others

and less easily accessible areas being under-sampled; and (3) focus-related: most field monitoring programs are focused on the state variables describing some of the biological components of the reef system and are not linked explicitly to the identification of stressors or processes [4]. As such, satellite observations serve as a useful mean of timely and cost effective global monitoring and surveying of large and remote coral reef areas globally that could otherwise not be achieved [24] [25] [26]. The most common sensors suitable for subsurface, ocean floor cover identification are SPOT High-Resolution Visible (HRV), Landsat Multispectral Scanner (MSS), Thematic Mapper (TM), Enhanced Thematic Mapper Plus (ETM+), Operational Land Imager (OLI), IKONOS, Advanced Airborne Hyperspectral Imaging System (AAHIS), Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), and Sentinel-2 [4] [27] [28] [29] [30] [31] [32] [33] [34]. In addition, it is noteworthy that there are many more satellites available that are capable of providing remote sensing data for coral reef analysis [25] [26] [35]. Visible spectrum is known to be useful in mapping of subsurface habitats [36] [37] [38]. This is owed to the fact that wavelengths (400nm-600nm) have ~15-3m penetration through clear waters, depending on turbidity and water quality [39]. However, this penetration depth is wavelength dependent as it increases with longer wavelengths [40]. As a result, the blue spectral bands (400nm) attenuate more slowly than red spectral bands (600nm) [1]. Moreover, underwater marine environment detection doesn't only come with spatial and spectral limitations challenges, but also the confounding influence variable depth on bottom reflectance, and disturbances due to turbidity of the water column [41] [42] [43] [44]. These factors significantly influence the spectral reflectance of the seafloor, thereby causing identical bottom types to exhibit substantially different characteristics in remote sensing data [45] [46]. Therefore,

water column and atmospheric corrections are needed in order to accurately detect the existence of coral within a pixel [47] [48] [49].

In this study, we considered 29 different scenes with known benthic regions. We implemented a process which accounts for any interferences in the image (clouds, dropped pixels, etc), applied a water mask, corrected for atmospheric obstructions, then calculated each of the depth invariant indices across the image. To better understand the output, we plotted each of the depth invariant indices as an image and observed the differences between the plots. This analysis showed that depth invariant index can be applied to a variety of ocean scenes throughout the world, with limited variance in the resulting output parameters. Furthermore, the results from this analysis can be used as inputs to a classification algorithm in order to rapidly identify benthic zone cover types, enabling large-scale, multi-temporal change detection.

## **2.2 Data Used**

Images from Landsat-8 OLI with 30m spatial resolution were used to conduct the survey. The visible bands were used due to their water column penetration properties with band 1 corresponding to 0.433-0.453 $\mu\text{m}$ , band 2 corresponding to 0.450-0.515 $\mu\text{m}$ , band 3 corresponding to 0.525-0.600 $\mu\text{m}$ , and band 4 corresponding to 0.630-0.680 $\mu\text{m}$ . It is noteworthy that the previous visible bands have the spectral range that can help in identifying water-land interface. In addition, band 5, which represents near infrared (NIR), was used to identify areas of full wavelength absorption for water masking. Scenes were selected based on the existence of corals in benthic zones. Landsat 8 images were first filtered by cloud cover which was restricted to less than 10% for each scene. The remaining

images were then visually inspected to determine which were most appropriate, including factoring in the location of clouds and other disturbances in the locations of known corals. In total 29 different scenes were selected. A listing of each site and the associated Landsat scene Row-Path reference is provided in Table 2-1:

Table 2-1: Selected Scenes for Study

Row-Path	Description	Image Capture Date	Row-Path	Description	Image Capture Date
013-055	Panama Coiba National Park	2/13/2017	091-075	Australia Great Barrier Reef	9/24/2016
013-057	Colombia Malpelo Fauna and Flora Sanctuary	4/2/2017	104-041	Japan Ogasawara Islands	8/3/2016
015-043	USA Everglades National Park	11/2/2014	106-055	Palau Rock Islands Southern Lagoon	3/13/2017
016-053	Costa Rica Area de Conservación Guanacaste	4/7/2017	114-066	Indonesia Komodo National Park	3/29/2014
016-056	Costa Rica Cocos Island National Park	3/22/2017	114-080	Australia Ningaloo Coast	2/17/2017
018-047	Mexico Sian Ka'an	7/18/2014	115-078	Australia Shark Bay, Western Australia	4/29/2017
018-048	Belize Belize Barrier Reef Reserve System	4/5/2017	116-054	Philippines Tubbataha Reefs Natural Park	4/28/2014
018-060	Ecuador Galápagos Islands	3/4/2017	123-065	Indonesia Ujung Kulon National Park	10/3/2013
034-047	Mexico Archipiélago de Revillagigedo	3/4/2017	126-046	Viet Nam Gulf of Tonkin	10/8/2013
038-039	Mexico Gulf of California	5/19/2017	159-051	Yemen Socotra Archipelago	5/3/2017
069-063	Kiribati Phoenix Islands Protected Area	12/11/2013	161-067	Seychelles Aldabra Atoll	2/10/2017
073-042	USA Hawaiian Archipelago	3/2/2016	167-079	South Africa iSimangaliso Wetland Park	6/28/2017
083-074	France Lagoons of New Caledonia	3/11/2017	170-047	Sudan Sanganeb Marine National Park and Dungonab Bay	
085-082	Australia Lord Howe Island Group	4/5/2015	213-063	Brazil Brazilian Atlantic Islands	
087-069	Solomon Islands East Rennell	5/10/2017			

## 2.3 Methodology

The overall layout of the processing approach applied to the 29 selected scenes is shown in Figure 2-1. This approach can be broken down into three major components: preprocessing the image (cloud, quality, and land/water masking), atmospheric correction, and water column correction. Scenes were selected to minimize the presence of clouds in the image. Pixels that still suffer cloud cover or other obstructions were then identified by leveraging the Landsat BQA band and masked accordingly. This is followed by creating a water mask by applying a threshold to the NIR band. This is carried out because the water body and corals have similar spectral reflectance, which may lead to misclassification in water/coral areas. A deep-water AOI was selected to be used in atmospheric correction via the dark-pixel subtraction method [37] [50] [51]. These adjustments were applied to each of the visible bands before the depth invariant indices were calculated. Analysis was conducted using the open source R programming language and environment for statistical computing [52].

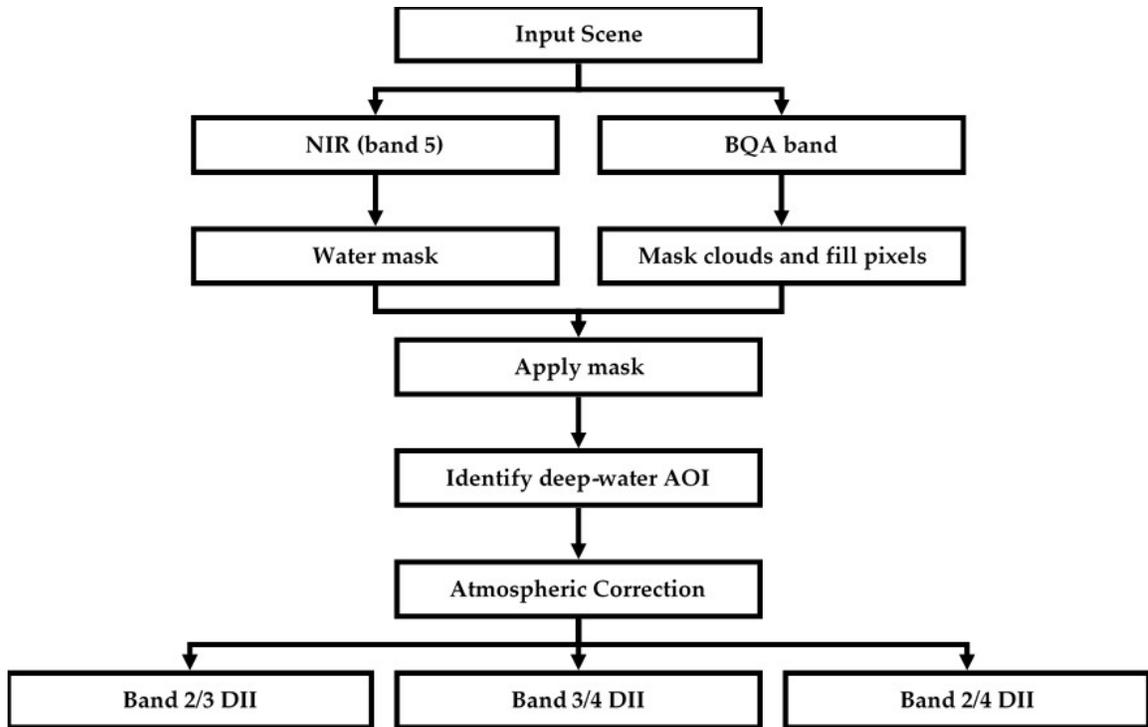


Figure 2-1: Scene masking, atmospheric correction, and water column correction process flow.

### 2.3.1 Cloud and Quality Mask

The first step taken in preprocessing the selected images was to identify and mask cloud cover. Great care was taken to select scenes with minimal amount of cloud cover, but in many instances, no available image was completely clear of cloud cover. Masking of clouds was performed by leveraging the Landsat 8 Level 1 product quality band (BQA) [53]. This band includes values for each pixel that, when converted to binary, indicate any potential disturbances with respect to the pixel and a rough approximation of confidence that that condition exists [54]. Converting each observation in this band and thresholding provides a mask to account for some of these conditions. Bits 14 and 15 indicate the likelihood of cloud interference while bits 12 and 13 indicate the possibility of cirrus cloud interference. There are several possible ways in which bias from cloud obstruction can

contaminate an analysis. The obvious entry point is disrupting the reflectance of a pixel or group of pixels. In addition to directly influencing surface reflectance, cloud cover present in the deep-water AOI can alter the values used for atmospheric correction applied to the image through the dark-pixel subtraction method [55]. As a result, this initial step of masking cloud interference is a critical preprocessing step.

### **2.3.2 Water Mask**

The study of DII related metrics across large scenes requires preprocessing that includes masking land as well as clouds. This step is imperative, as including land pixels will severely distort the DII parameters when calculated across the scene. The water mask was created by leveraging the Landsat 8 NIR sensor. This sensor measures light between 0.851 and .0879 micrometers. Water absorbs light in these wavelengths, therefore, it is a good candidate for discerning water from land in any given scene [56]. As in the visible bands, the Landsat 8 NIR band (band 5) is at 30m resolution. A threshold was applied to the NIR band pixel values of each scene. The plots were then evaluated visually to determine the most appropriate cutoff for separation of land and water. A mask was then created for pixels determined to be water.

### **2.3.3 Atmospheric Correction**

In the visual bands, 90% of the at sensor reflectance depends on atmospheric and water surface properties [57]. Therefore, atmospheric correction is first performed using the dark pixel subtraction method [58]. This method selects areas of the scene with water known to be deep enough for the visible bands to fully attenuate. Signal received from

these areas are comprised of atmospheric radiance and surface reflectance, thereby isolating the impact of these elements. Assuming the atmospheric and water surface conditions generalize to the rest of the scene (i.e. uniform throughout the area of interest), the mean deep-water radiance at sensor can be leveraged to correct for the effect of atmospheric and surface reflectance interferences [1] [59] [60]. Depths greater than 50m will assure that the visible wavelengths have fully attenuated [61]. In addition, two standard deviations are subtracted to account for possible sensor noise [62]. It is important to highlight the assumption that conditions are uniform across each scene. In addition, because of this assumption, the deep-water AOI selected should appear in the same scene that is being analyzed. This will minimize the possibility of unintended bias that may be introduced by leveraging a deep-water AOI of another image and that, to the greatest extent possible, the effect of full attenuation of the wavelengths is isolated. In preprocessing, the deep-water AOI was selected through visual inspection. In some instances, references to external sources were made to verify the appropriate depth.

### **2.3.4 Water Column Correction**

As light penetrates water, the intensity decreases exponentially with increasing depth. The rate of attenuation is wavelength-dependent and has a severe effect on remotely sensed data of aquatic habitats [1]. Therefore, water column correction is appropriate for imagery with multiple water-penetrating spectral bands [51] [63]. Within these visible spectral bands, longer wavelength blue bands attenuate less rapidly than shorter-wavelength red bands. Therefore, the spectral radiances recorded at sensor are dependent on both the subsurface strata reflectance and depth. The confounding influence of depth

can create significant distortions in the subsurface reflectance. Since most marine habitat-mapping exercises are only concerned with mapping benthic features, it is advantageous to remove the influence of variable depth. The radiance at sensor in band  $i$  ( $L_i$ ) can be expressed as [1] [64] [65] [66]:

$$L_i = L_{si} + a_i \cdot r_i \cdot e^{-fk_i z} \quad (1)$$

Where the following are represented:

$L_{si}$ : the mean deep-water radiance in band  $i$

$a_i$ : a constant for band  $i$  accounting for atmospheric effects and water surface reflection

$r_i$ : the bottom reflectance

$f$ : a geometric factor to account for path length through water (set to two for a two-flow model)

$k_i$ : the coefficient of attenuation of band  $i$  (to account for various interferences suspended in the water and scattering due to turbidity) [1] [66]

$z$ : depth

Applying natural logarithms and rearranging equation (1) generates an atmospherically corrected radiance for band  $i$  that varies linearly according to depth [1]:

$$\ln(L_i - L_{si}) = \ln(a_i \cdot r_i) - 2k_i z \quad (2a)$$

Similarly, the equation can be applied to band  $j$ :

$$\ln(L_j - L_{sj}) = \ln(a_j \cdot r_j) - 2k_j z \quad (2b)$$

Equation (2a) can be rearranged to determine the bottom reflectance  $r_i$ :

$$\ln(r_i) = \frac{\ln(L_i - L_{si}) + 2k_i z}{a_i} \quad (3)$$

Yet this equation presents us with three unknown variables. Namely, the constant, the coefficient of attenuation, and depth ( $a_i$ ,  $k_i$ , and  $z$  respectively). However, by leveraging the ratio of attenuation coefficients between each pair of bands we can avoid calculating estimates of  $k$  for each band [63] [64]. The ratio of attenuation between bands  $i$  and  $j$  can be determined using the following equation:

$$\frac{k_i}{k_j} = a + \sqrt{(a^2 + 1)} \quad (4)$$

where

$$a = \frac{\sigma_{ii} - \sigma_{jj}}{2\sigma_{ij}} \quad (5)$$

and  $\sigma_{ii}$  is the variance of band  $i$  and  $\sigma_{ij}$  is the covariance between bands  $i$  and  $j$ .

Therefore, using these equations, the depth invariant index (DII) can be calculated for any given scene without any external references to additional data.

$$DII_{ij} = \ln(L_i - L_{si}) - \left[ \frac{k_i}{k_j} \cdot \ln(L_j - L_{sj}) \right]$$

## **2.4 Output Parameters**

We will now discuss in detail the output and related parameters that result from applying the above methodology to the 29 Landsat 8 scenes.

### **2.4.1 Image Correction and Deep-Water AOI Parameters**

A section of each scene representing deep-water was selected for atmospheric correction using dark-pixel subtraction. The mean deep-water AOI for each scene is shown in Figure 2-2. The highest mean deep-water values for were observed in the Sea of Cortez, Mexico, and the Sudan Sanganeb Marine National Park areas (row-path 038-039 and 170-047, respectively). These two scenes were unique in that they are located in large gulfs. As a result, they are both protected from the influence of currents and associated turbidity. In addition, the deepest water in these scenes is likely shallower than that of other scenes. Finally, both locations are in areas known to be somewhat arid environments with favorable conditions for reducing atmospheric interference.

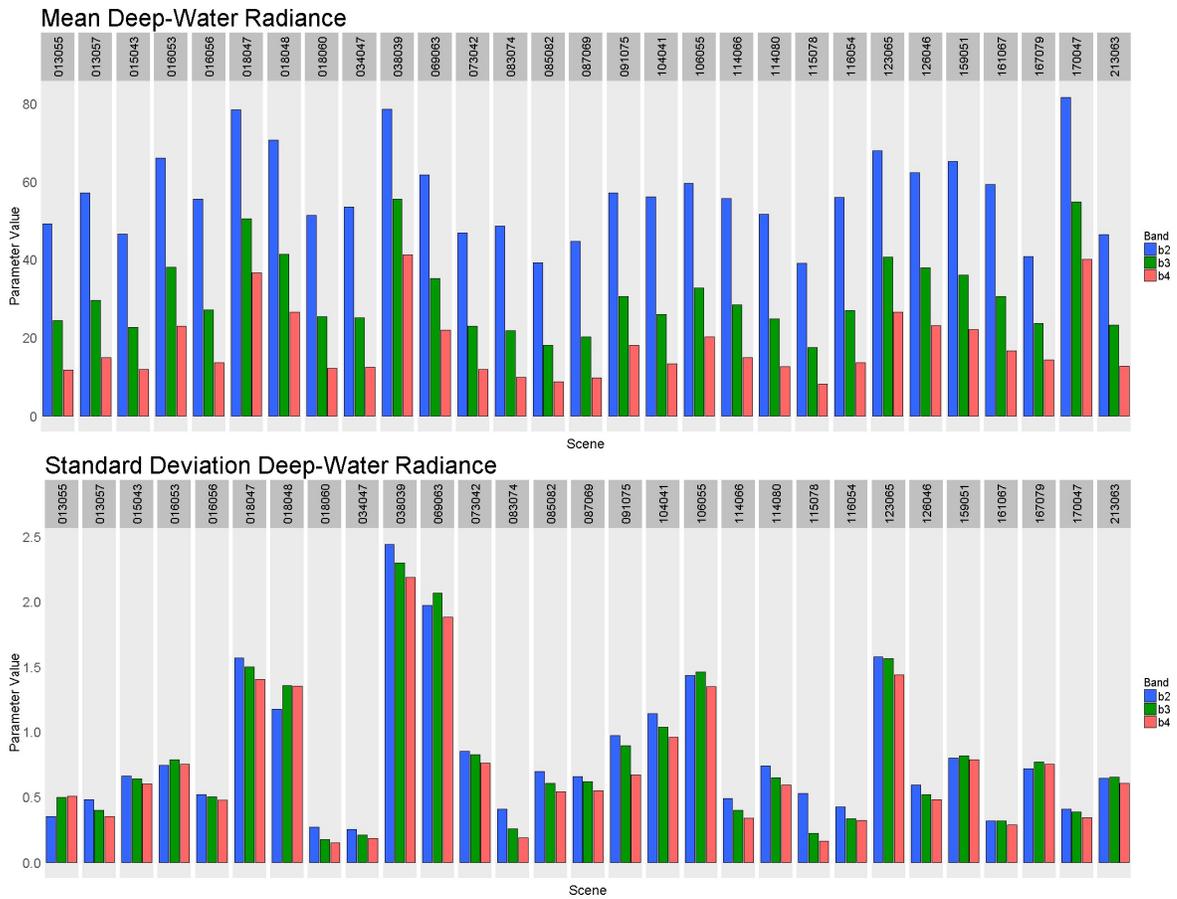


Figure 2-2: Top: Comparison of mean deep-water radiance for Landsat 8 band 2, band 3, and band 4 for each of the 29 scenes selected for analysis. Bottom: Comparison of standard deviation of deep-water radiance.

The standard deviation measures were consistent across many of the scenes. The Sea of Cortez, Mexico, showed the largest standard deviation which could also be a product of the location being in a large gulf. In addition, the Phoenix Islands in Kiribati and Ijung Kulon National Park in Indonesia (row-path 123-065) showed higher variance. This could be due to clouds in the selected area. Clouds were masked from the image and an attempt was made to avoid them when selecting the deep-water AOI in each scene.

## 2.4.2 Ratio of Attenuation

The ratio of attenuation is calculated from variance and covariance of the bands. The variance and covariance were calculated based on the atmospherically corrected and masked water pixels only. Figure 2-3 and figure 2-4 display the resulting variance and covariance for each scene. Several scenes stand out as having very high variance in radiance. This is due to unique geographical characteristics of those regions. The Everglades National Park in Florida, USA, (row-path 015-043) is known for its expansive wetlands. This unique aquatic environment contains very shallow waters mixed with a high concentration of vegetation. Regions such as this will absorb wavelengths in the NIR spectrum and therefore are appropriately classified as water. Furthermore, by the very nature of the environment, its spectral radiance exhibits a high degree of variance as an abundance of very shallow cover is detected within the image. Similarly, Australia's Shark Bay exhibits a high degree of variance likely due to the spectral radiance and ecology of very shallow waters.

The covariance of band radiation reveals similar characteristics to the variance of the bands. That is, scenes with regions dominated by shallow waters such as the Everglades National Park in Florida, USA, and Shark Bay, Australia, have a high degree of variance in spectral radiance between the bands. This is an expected result as a larger variety of vegetation has evolved to live in these aquatic environments.

With the variance and covariance known, we can calculate the constant used to account for atmospheric effects and water surface reflection ( $a$ ) for each scene. This constant is estimated as the difference between the covariance of each band pair divided

by two times the covariance between the bands. The resulting value for the constant ( $a$ ) for each scene is presented in Figure 2-5. The largest value for the constant ( $a$ ) was derived from the image of Hawaiian Archipelago in the USA. This could be consistent with the geography of that specific scene which features mostly deep ocean with two reefs. The two reefs are the Maro Reef and smaller Raita Bank which both are at a unique depth in which visible wavelengths in the red band will fully attenuate but the blue band will not. Furthermore, the rest of the scene is very deep-water. Therefore, red light fully attenuates across nearly the entire image contributing a very low variance. Yet, the blue band has comparably higher variance since it is able to return light from the Maro Reef and Raita Bank. This difference combined with the generally deep characteristic of the rest of the image results in a high value for the constant ( $a$ ).

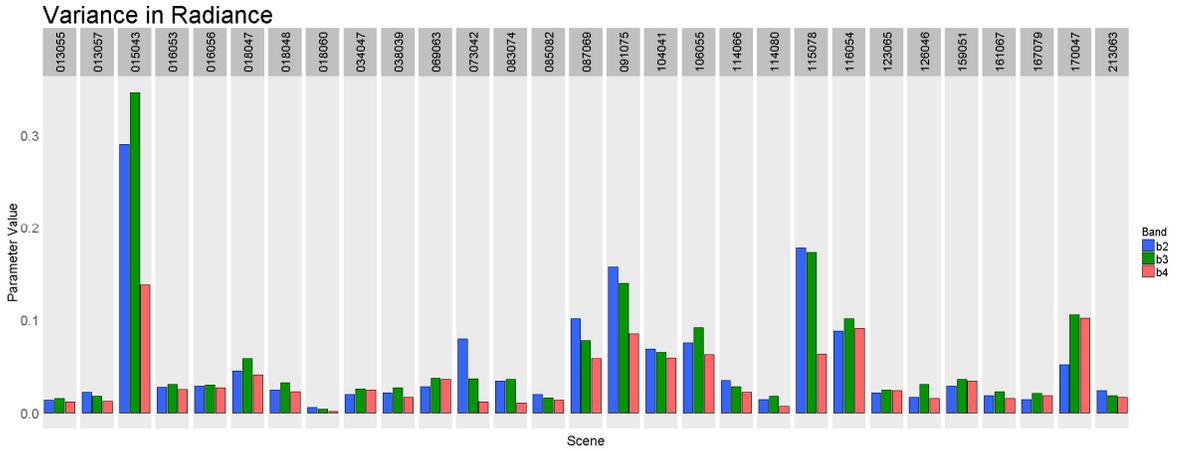


Figure 2-3: Comparison of variance in radiance for each band. Certain scenes show more variance due to unique geographic characteristics.

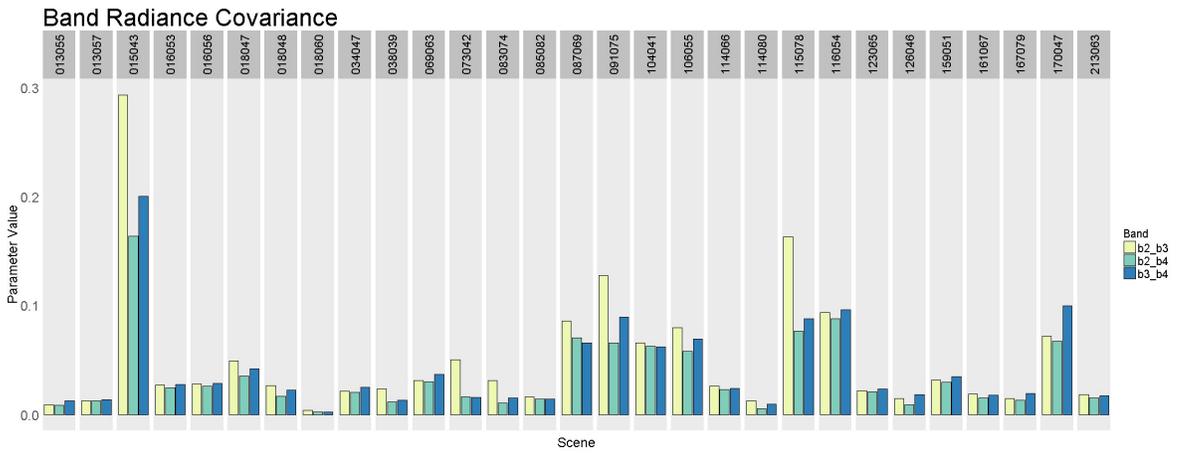


Figure 2-4: Comparison of covariance of spectral radiance between bands.

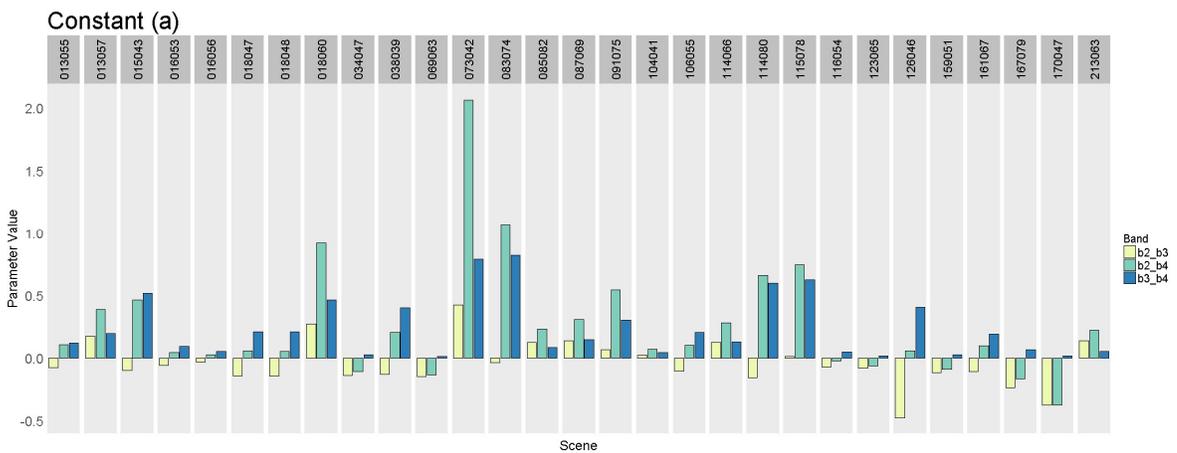


Figure 2-5: Comparison of the atmospheric and surface reflection correction constant (a) for each scene and pair of bands.

Finally, the ratio of attenuation is very similar for all scenes with few exceptions. Most notable is the Hawaiian Archipelago which we discussed above. The larger values for the ratio of attenuation in other scenes are due to similar reasons. Specifically, each of these scenes have geographic features which reside beyond the threshold to which all the visible bands can penetrate. As a result, the DII results in three distinct regions. First that to which all visible bands can penetrate and return information to the sensor. Second, regions in which one of the bands fully attenuates. This results in only partial information being returned to the sensor. There are regions in which all visible bands fully attenuate and for these regions no useful information regarding the ocean floor can be returned. Figure 2-6 displays the Ratio of Attenuation for each of the scenes analyzed.

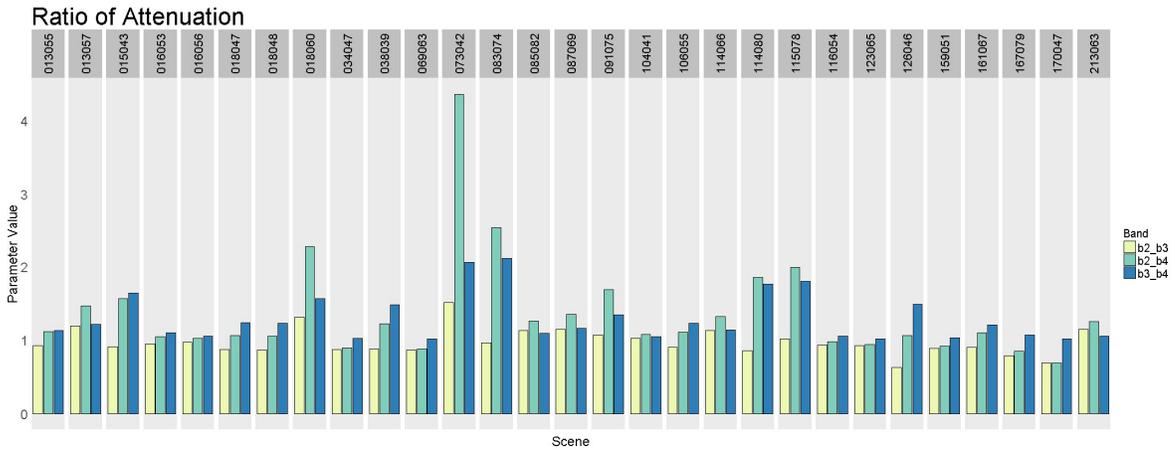


Figure 2-6: Comparison of ration of attenuation for scene and pair of bands.

## 2.5 Results and Discussion

The calculated depth invariant indexes can be plotted for each scene (Figure 2-7 through Figure 2-12) for visual inspection. Each scene can support three depth invariant indices corresponding with band 2/band 3, band 2/band 4, and band 3/band 4. These maps are a depth invariant view of the ocean floor characteristics. However, there are several

limitations to what can be viewed, and if light from one or both band pairs fully attenuate, the ocean floor cannot be analyzed. Light in the blue spectrum (0.452-0.512 $\mu\text{m}$ ) can penetrate water the furthest while light with shorter wavelengths attenuate faster. It is estimated that the blue band will fully attenuate in water that is approximately 21.4m deep while light in the red spectrum (0.636-0.590 $\mu\text{m}$ ) will fully attenuate in water that is 5.2m deep. The impact of this can be observed in several of the images in which the band 2 and band 3 DII shows variance in the ocean floor while the DII based on band 3 and band 4 do not. This is a result of band 3 and band 4 fully attenuating in water that is greater than 16.8m deep. Water this deep absorbs both green and red wavelengths and therefore does not return useful information regarding the ocean floor. This threshold can clearly be seen in several of the images. Figure 2-7 includes a plot of each of the DIIs related to the Sea of Cortez, Mexico (row-path 038-039). This scene includes a series of shoals at varying depths extending out from the Baja California shore. The shallowest of these shoals ranges from 8m to 15m. At this depth red light fully attenuates, however, green light does not. As a result, a gradient corresponding with the depth at which red light fully attenuates can be observed in both the band 3-band 4 (green-red) and band 2-band 4 (blue-red) DII plots. However, this same gradient does not exist in the band 2-band 3 (blue-green) DII plot as the wavelengths do not fully attenuate and information regarding reflectance of the ocean floor is returned to the sensor. It is also worth noting that in the band 2-band 3 plot a certain mixture of index values are presented that appear unique compared to that of the DII plots using bands that have attenuated. Similarly, the image of the Gulf of Tonkin shows stratification in the band 3-band 4 DII image but not in the band 2-band 3 DII image. As in the image of the Sea of Cortez, this region is characterized by a long benthic zone that

extends out from the shore. The depth of this underwater feature ranges from less than 5m to approximately 20m. Therefore, while the band 2-band3 DII appropriately shows depth invariant information regarding the ocean floor, the band-3-band 4 image can only provide information regarding the areas for which light in the red wavelengths has not fully attenuated. This results in the stratification that can be observed in the band 3-band 4 image.

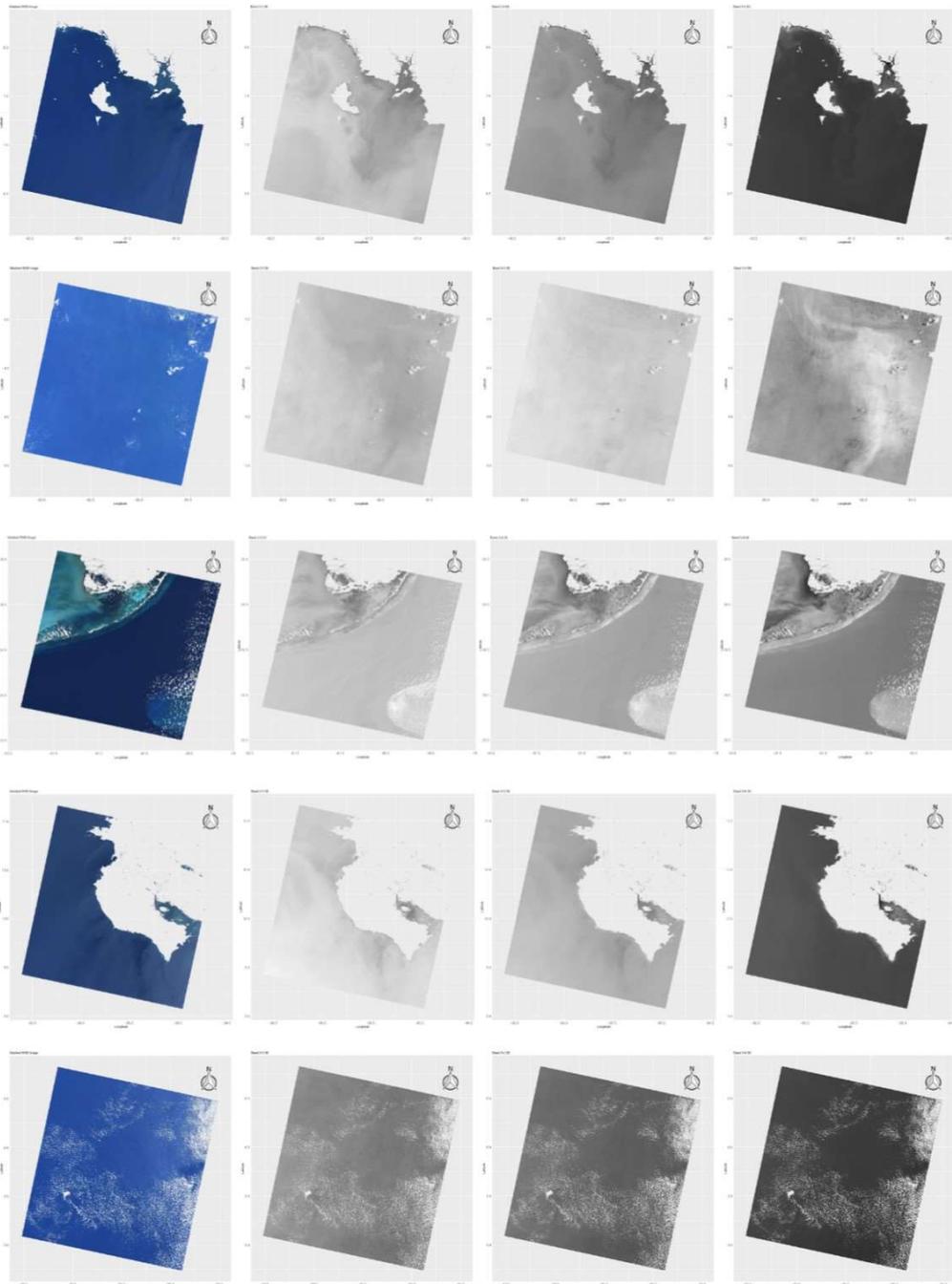


Figure 2-7: Left to right: water masked scene, band 2/3 depth invariant index, band 2/4 depth invariant index, band 3/4 depth invariant index. Top to bottom: 013-055 Panama Coiba National Park, 013-057 Colombia Malpelo Fauna and Flora Sanctuary, 015-043 USA Everglades National Park, 016-053 Costa Rica Area de Conservación Guanacaste, 016-056 Costa Rica Cocos Island National Park.

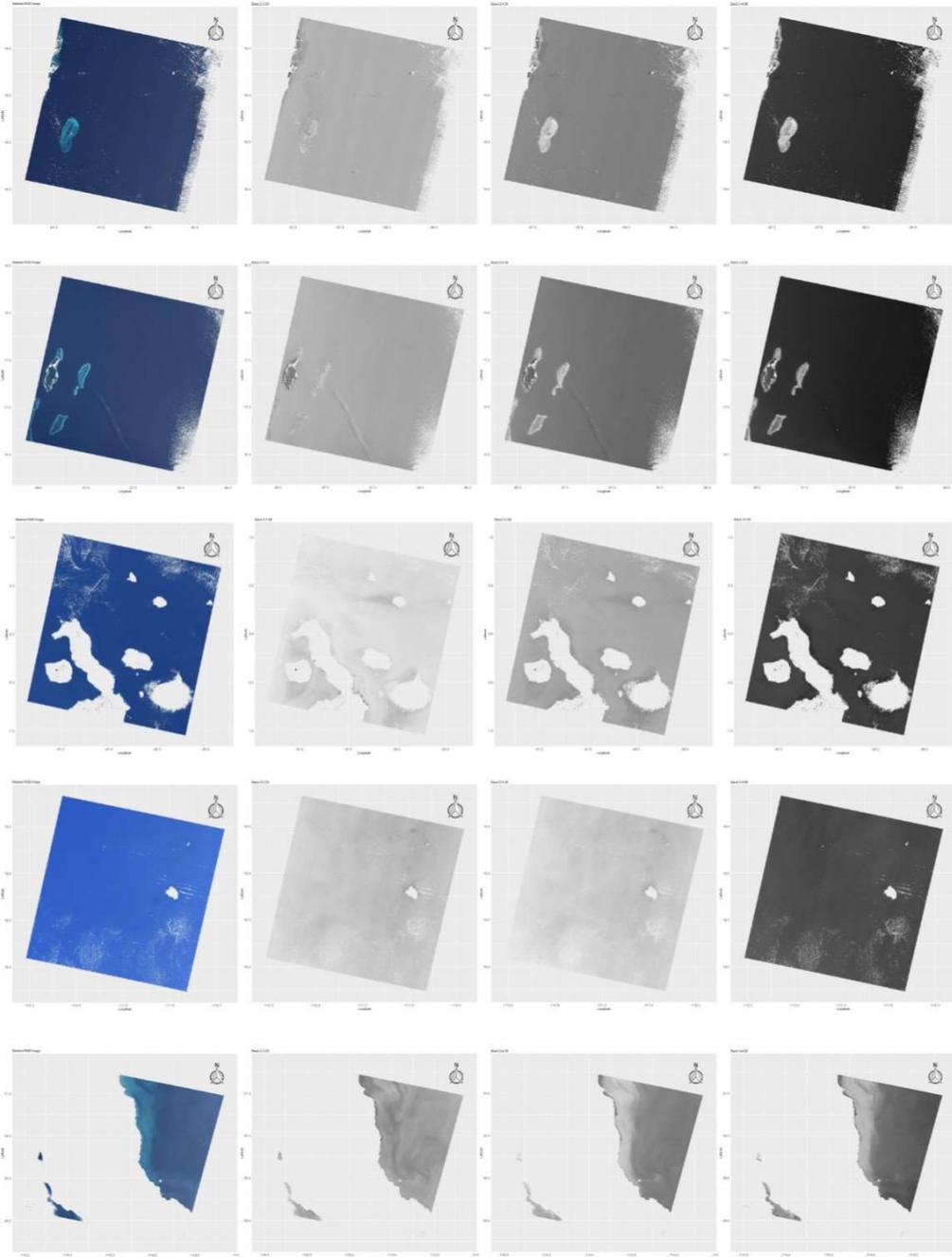


Figure 2-8: Left to right: water masked scene, band 2/3 depth invariant index, band 2/4 depth invariant index, band 3/4 depth invariant index. Top to bottom: 018-047 Mexico Sian Ka'an, 018-048 Belize Belize Barrier Reef Reserve System, 018-060 Ecuador Galápagos Islands, 034-047 Mexico Archipiélago de Revillagigedo, 038-039 Mexico Sea of Cortez.

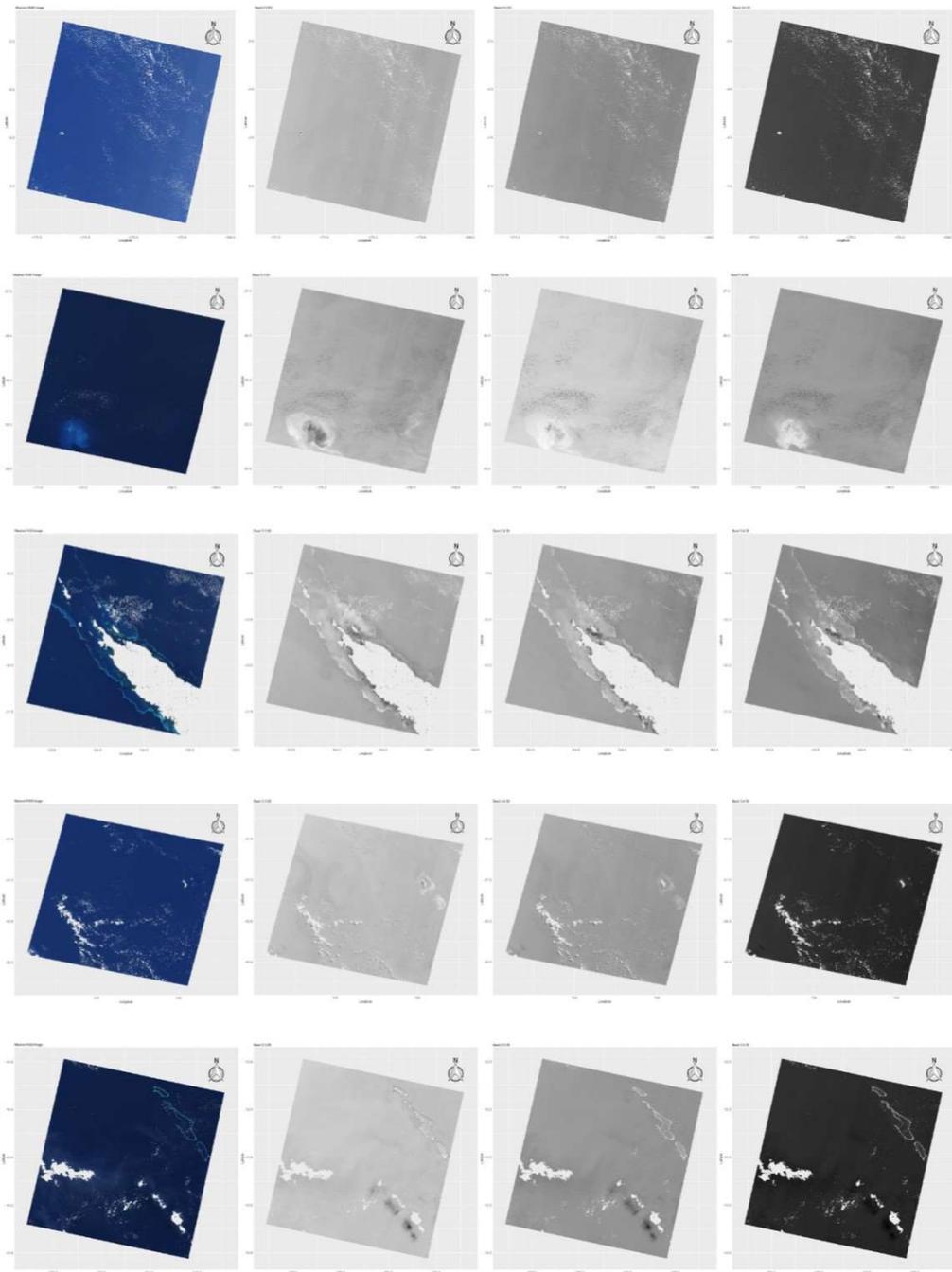


Figure 2-9: Left to right: water masked scene, band 2/3 depth invariant index, band 2/4 depth invariant index, band 3/4 depth invariant index. Top to bottom: 069-063 Kiribati Phoenix Islands Protected Area, 073-042 USA Papahānaumokuākea, 083-074 France Lagoons of New Caledonia, 085-082 Australia Lord Howe Islands, 087-069 Solomon Islands East Rennell.

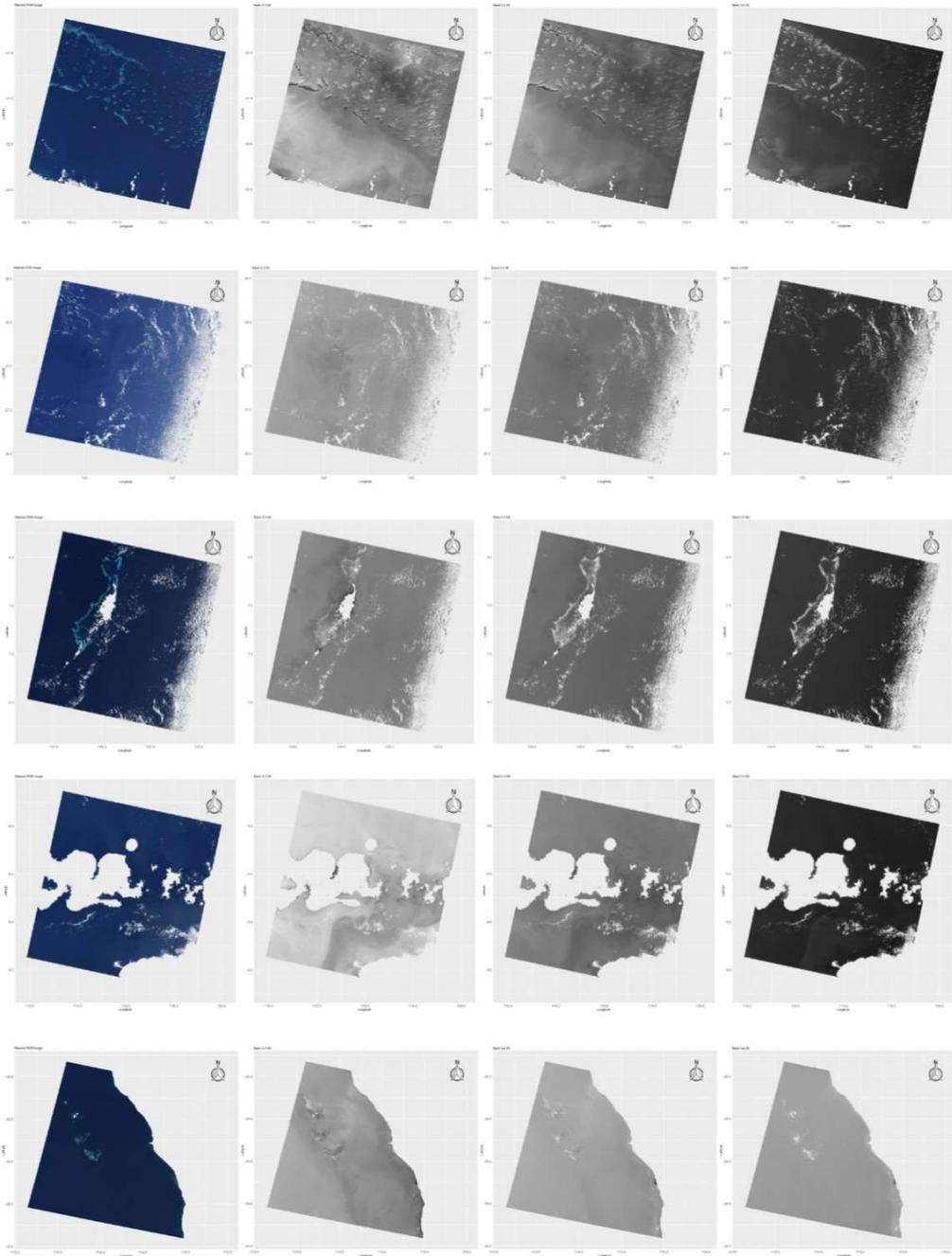


Figure 2-10: Left to right: water masked scene, band 2/3 depth invariant index, band 2/4 depth invariant index, band 3/4 depth invariant index. Top to bottom: 091-075 Australia Great Barrier Reef, 104-041 Japan Ogasawara Islands, 106-055 Palau Rock Islands Southern Lagoon, 114-066 Indonesia Komodo National Park, 114-080 Australia Ningaloo Coast.

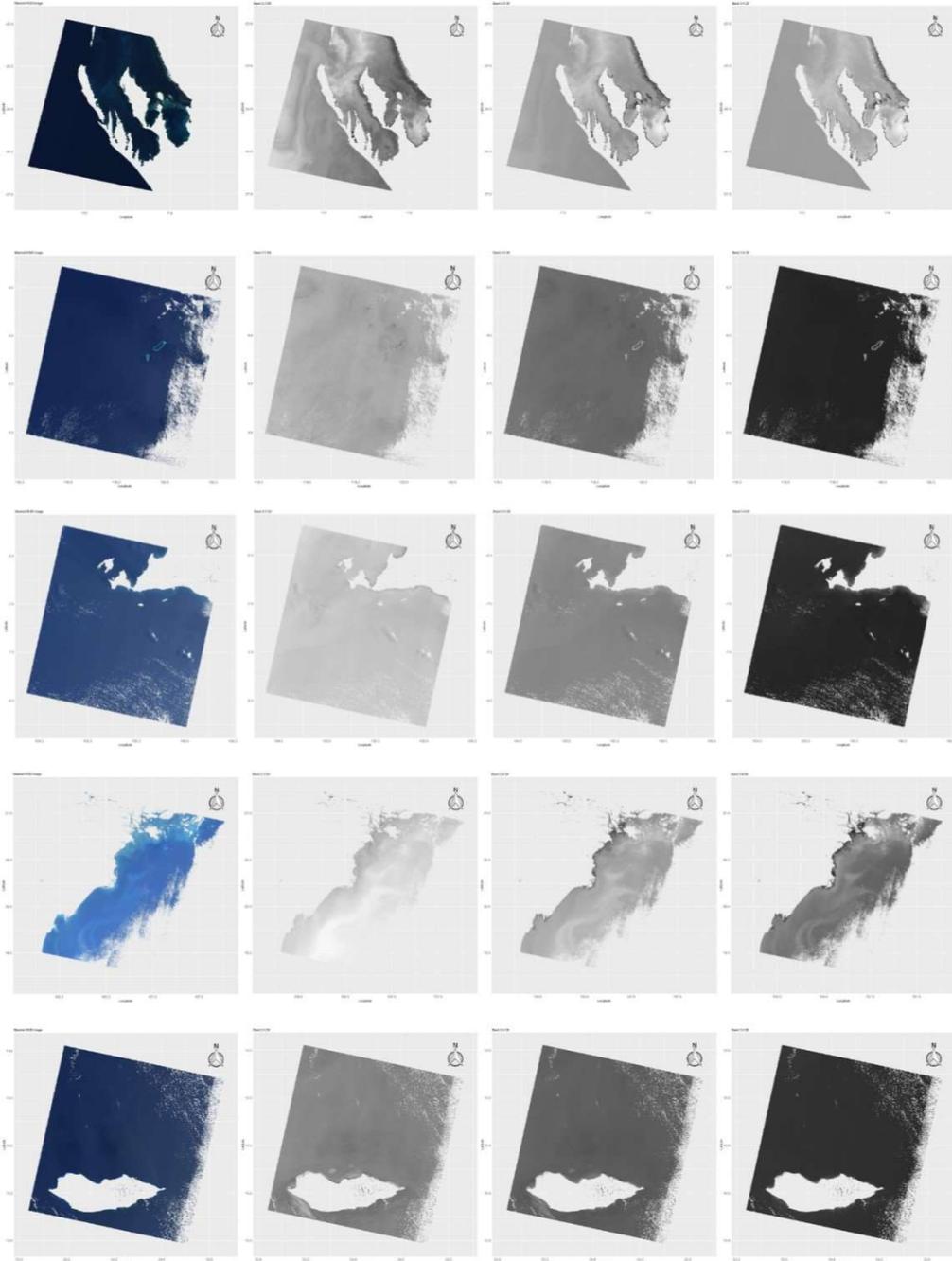


Figure 2-11: Left to right: water masked scene, band 2/3 depth invariant index, band 2/4 depth invariant index, band 3/4 depth invariant index. Top to bottom: 115-078 Australia Shark Bay, Western Australia, 116-054 Philippines Tubbataha Reefs Natural Park, 123-065 Indonesia Ujung Kulon National Park, 126-046 Viet Nam Gulf of Tonkin, 159-051 Yemen Socotra Archipelago.

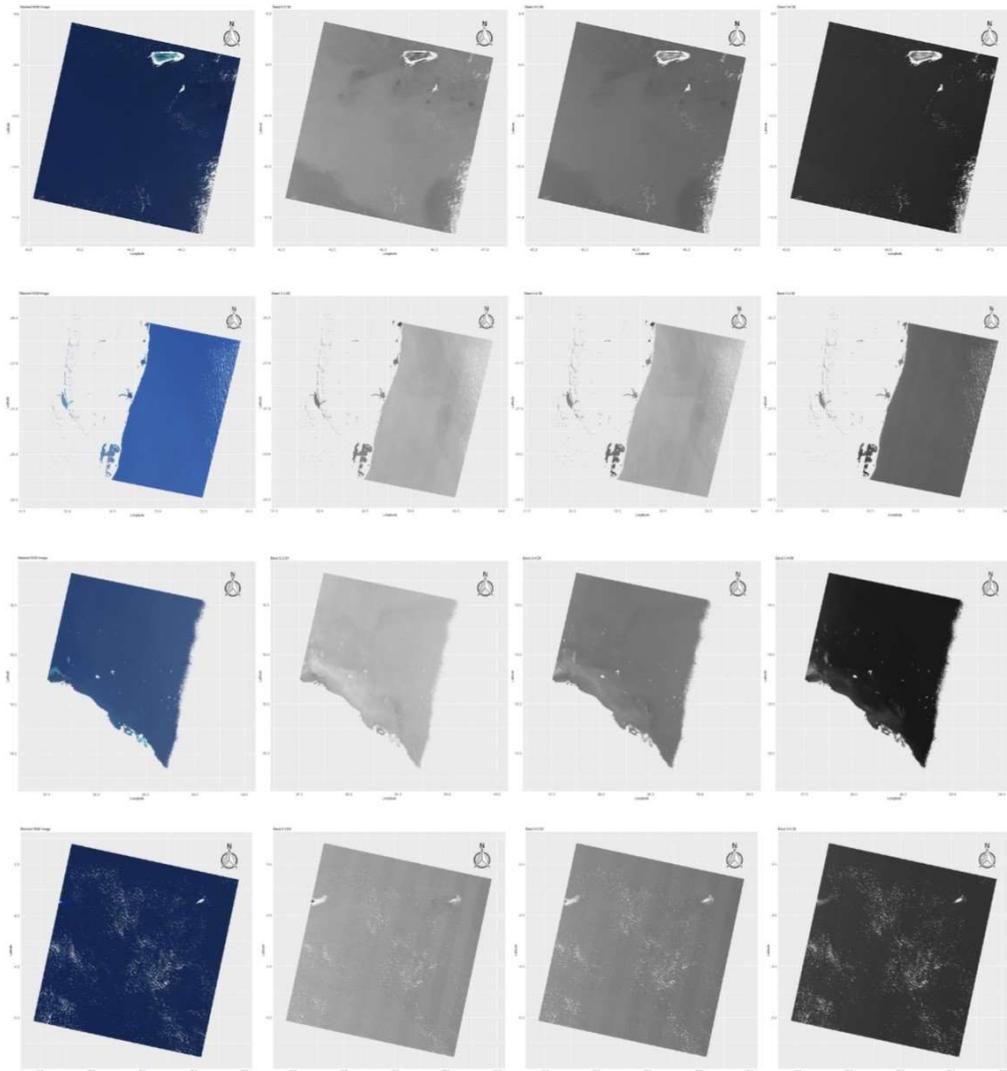


Figure 2-12: Left to right: water masked scene, band 2/3 depth invariant index, band 2/4 depth invariant index, band 3/4 depth invariant index. Top to bottom: 161-067 Seychelles Aldabra Atoll, 167-079 South Africa Simangaliso Wetland Park, 170-047 Sudan Sanganeb Marine National Park, 213-063 Brazil Brazilian Atlantic.

## 2.6 Conclusion

This research presented the application of DII to a diverse collection of scenes on a global scale. The results show that DII can be applied to a variety of scenes and return information regarding subsurface ground cover in shallow benthic zones. We analyzed each factor that needed to be controlled to minimize inaccuracies and bias. This included masking of clouds, dropped, and fill pixels using the Landsat BQA band. Deploying DII across large scenes meant that we needed to apply calculations to all portions of the image containing water and therefore created a water-mask using the Landsat NIR band. We further preprocessed the images to adjust for atmospheric disturbances using the dark-pixel subtraction method. This was deployed by interactively selecting an area of the image known to be deep to represent our deep-water AOI. We then analyzed the results of our calculations. Specifically, we looked at the differences across scenes in the mean deep-water radiance, deep-water radiance standard deviation, and final dark-pixel adjustment values. These steps concluded our preprocessing and we plotted the results for each scene including the deep-water AOI selected. We proceeded to calculate DII for each of the scenes. Deployment of DII at this scale has not been performed before. We computed and reviewed each of the contributing parameters including the radiance variance and covariance for each band, the atmospheric and water surface reflection correction constant (a), and the ratio of attenuation coefficients for each pair of bands. We analyzed 29 scenes comparing each of these parameters across each of the scenes and noted any large differences. Finally, we calculated the DII for each pair of bands and each scene and plotted the results as a map using the mask calculated during preprocessing. We then compared this to the preprocessed RGB plots using the visible bands and compared the results to

geological features known to exist. We looked deeper into some of the stratification that was observed and discussed the attenuation of visible wavelengths as they pass through water. Specifically, we noted stratification in the depth index which incorporated light within the red wavelength and that this same feature did not appear in the other indices. Furthermore, we were able to correlate the occurrence of the stratification with the approximate depth at each location in which the visible wavelengths for each respective band was fully attenuating. We identified three distinct regions within the DII information. First, regions for which information is returned to the sensor regarding the ocean floor by all three visible bands. Second, regions in which only one of the visible bands for a given index returns useful information to the sensor regarding the ocean floor. And, finally, the third region in which all the visible bands fully attenuate in the water and no useful information regarding the ocean floor is returned to the sensor

A general global decline in coral abundance has been observed and attributed to many ecological and man-made factors [5] [6] [7] [8] [9] [10]. Additional study and testing will need to be conducted in order to validate the appropriateness of applying a subsurface cover detection algorithm to the DII corrected scenes for change detection. Preliminary evidence suggests the strong possibility of being able to differentiate major subsurface bottom types [67] [68] [69] [70] [71], however, the generalizability of these results across scenes will need to be ascertained. In addition, further research into the use of the Landsat dark-blue (band 1) and its attenuation properties should be investigated.

# **3. First Study: Evaluation of Spatial Generalization Characteristics of a Robust Classifier as Applied to Coral Reef Habitats in Remote Islands of the Pacific Ocean**

## **3.1 Introduction**

There have been a multitude of studies that leverage remote sensing data to evaluate reef health on an in-situ basis. These studies isolate a single location and provide in depth analysis of reef health for that site. Many of these studies are temporal, also called spectral signature generalization and expansion [72]. These studies provide invaluable insight into the impact of climate change and the resulting site-specific environmental transformation. While these studies consistently reiterate the spectral signature generalization properties as they relate to a single site temporally, they do not address the spatial generalization properties of the spectral signature. If the existing in-situ studies are considered temporal studies, the progression based on spatial generalization can be considered a longitudinal or spatial study. While temporal studies have been conducted to evaluate the generalization properties of remote sensing information across seasons and years, evaluation of the spectral signature generalization properties across various proximities will measure the

larger impact of disparate environmental conditions spatially in addition to variations in the spectral reflectance of coral ecosystems themselves. This longitudinal analysis will account for any location-based bias that previous in-situ analysis could not account for. Studying this bias and how well information contained within spectral reflectance generalizes spatially, in combination with the existing knowledge of how well spectral reflectance generalizes temporally, will significantly enhance our scientific understanding of detecting coral using remote sensing data. Furthermore, this knowledge will augment the existing research to enable temporal analysis on a larger scale than previous in-situ efforts. This research has been conducted with these far-reaching goals in mind. Specifically, we have focused on Landsat data since there is a rich anthology of historical scenes which can be analyzed. We have limited our usage of the existing Landsat bands only to those which have been available historically and not the coastal aerosol band which is only available on the most recent Landsat 8 mission.

Coral environments are among the most diverse environments in the world and therefore the spectral radiance can have some degree of variation even within a single pixel as the coral species varies within the location. In addition, coral environments are often heterogeneous, exhibiting a complex mixture of bottom types within a single pixel. This poses further challenges with using remote sensing data to detect coral, are the diversity of coral species and cover within each pixel, and the resolution available from a given spectrometer. These factors weaken the ability of a classifier to accurately determine the existence of coral. This is because satellites equipped with lower resolution spectrometers will tend to capture more heterogeneity within each pixel, and in turn deliver more mixed information regarding spectral reflectance to the classifier.

In this study, we evaluate the spatial generalization principles of the spectral signature across sites and scenes when applied to the detection of coral reefs. We considered four different locations across three different Landsat 8 scenes in which coral reefs are known to exist. Our implementation accounts for obstructions in the image (clouds, shadows, dropped pixels, etc.), applies a water mask, corrects for atmospheric obstructions and sun glint, then calculates depth invariant indices across the image. Using the corrected depth invariant outputs of these preprocessing steps, we then applied linear discriminant analysis (LDA) to predict the presence of coral. After evaluating the accuracy of this in-situ analysis, we trained an LDA algorithm for application to coral cover type in another location within the same scene. Finally, we applied the linear discriminant analysis function to several sites in different Landsat scenes and analyzed the results. The methodology is a representation of the common data and applied science practice of splitting a dataset into training and testing sets. Many in-situ, temporal analyses perform train and test splits when creating a classifier as it is the most appropriate thing to do, however, there is an inherent bias between train and test samples of a single site simply due to the common environmental factors influencing the model input features. This bias is more pronounced when samples are close in proximity to each other. This results in a model that is trained and assessed using only localized observations from a single site and will not generalize well longitudinally. The reason is those observations do not adequately represent observations from other locations due to fluctuations in environmental conditions including water turbidity, the diversity of marine life, and the bottom cover type itself, as well as unique geomorphological features that may be specific to a single location. Training a model on a given site and then evaluating performance on another site will eliminate this

localized bias. As the distance between training and testing observations increases, the influence of environmental factors on both train and test data from that unique site is reduced or eliminated. The result is a novel model that is trained on the diversity of environments and conditions enabling the output to be more robust than in-situ analysis to location bias and therefore more scalable.

In this scientific study the first section will discuss the materials and methods used in the study including a description of the data used and the geomorphology of each site as well as the steps taken to preprocess the satellite images. The Materials and Methods Section is followed by a review of the per site results as well as a quantitative assessment of the site-specific generalization performance and evaluation of a robust model constructed using consolidated information from all the sites. The Discussion Section of the study includes an examination of the spectral signature generalization properties as it pertains to coral reef detection, an appraisal of the benefits and challenges of the methods evaluated in the research as well as a proposal for future work in the area. The study closes with a discussion regarding conclusions and outcomes resulting from this research.

## **3.2 Materials and Methods**

### **3.2.1 Remote Sensing Data**

Images from Landsat-8 OLI with 30m spatial resolution were used in the analysis. The visible bands were used due to their water column penetration properties, with band 2 corresponding to 0.450-0.515 $\mu\text{m}$  (blue), band 3 corresponding to 0.525-0.600 $\mu\text{m}$  (green), and band 4 corresponding to 0.630-0.680 $\mu\text{m}$  (red). Band 5, which represents NIR, was used to identify areas of full wavelength absorption for water masking. For each location

of the analysis, Landsat 8 images produced within 6 months of the ground truth observation date were used. Scenes were restricted to no more than 10% cloud cover and the remaining images were then inspected to determine which were most appropriate, including factoring in the location of clouds and other disturbances in the observation areas. In total three different scenes were selected in the Pacific Ocean for our experiment as shown in Figure 3-1. A listing of each site and the associated Landsat scene Row-Path reference is provided in Table 3-1. Additional detail regarding the sites and associated ground truth observation data can be found in Table 3-2. It is noteworthy that the first two sites are within the same path/row scene, hence the three scenes here are representing four sites.

Table 3-1: Selected Scenes for Study

<b>Location (Figure 3-1)</b>	<b>Path/ Row</b>	<b>Location</b>	<b>Image Capture Date</b>	<b>Area of Interest Size</b>
1	065/056	Palmyra Atoll	5/27/2015	19x5-km
2	065/056	Kinman Reef	5/27/2015	16x8-km
3	060/073	Baker Island Atoll	8/20/2014	9x11-km
4	059/074	Howland Island	1/18/2015	9x8-km

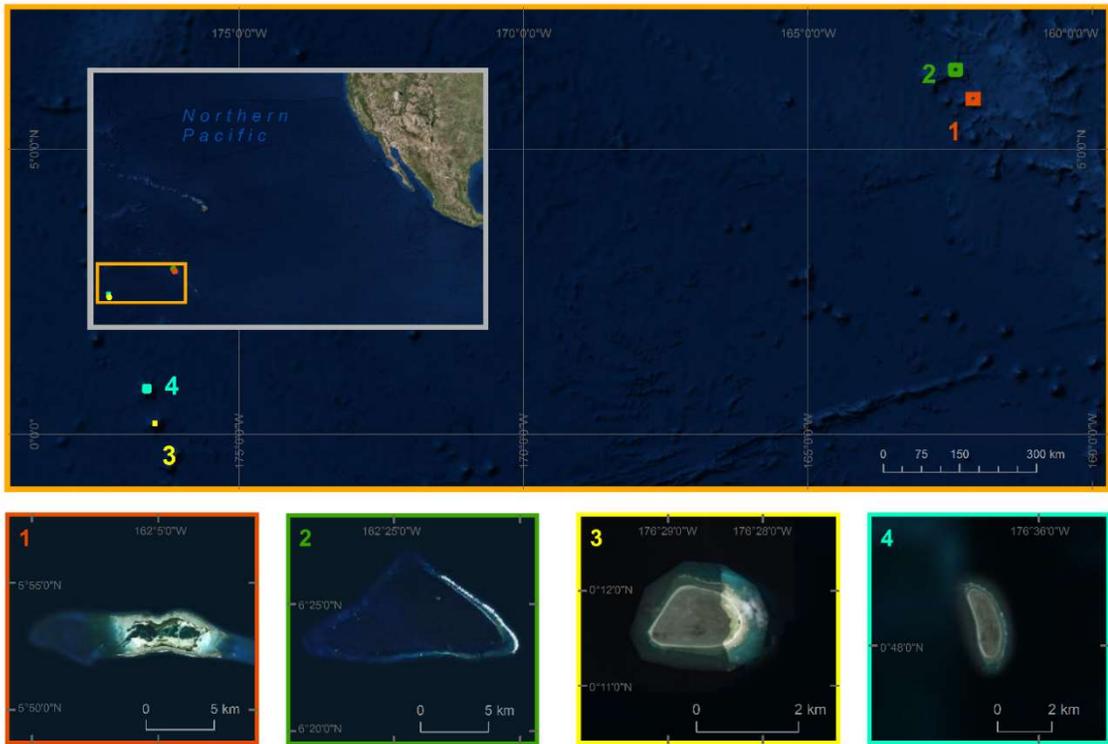


Figure 3-1 The location of the four sites (1) Palmyra Atoll (2) Kingman Reef (3) Baker Island Atoll (4) Howland Island

Table 3-2: Selected Scenes for Study, Additional Information

Location	Number of Observations	Distance from Training Site (Palmyra Atoll)
Palmyra Atoll	82	-
Kingman Reef	57	67.9-km
Baker Island Atoll	26	1,720.0-km
Howland Island	30	1,710.4-km

### 3.2.2 Study Sites

#### Palmyra Atoll

Palmyra Atoll is a 20km long elliptical reef located at 5°52'N 162°6'W within the Northern Line Island chain. It contains elongated terraces that extend 3-5km off both the east and west ends of the atoll the depths of which range from 7-25m [73]. Benthic

environments in the backreef are generally characterized by high rugosity, continuous reefs consisting of >50% live coral, interspersed with large, dead, standing corals [74]. Benthic fore reef habitats are dominated by hard coral and crustose coralline algae, together comprising 48% of all surfaces [75]. Halimeda and turf algae dominate fore reef benthic algal assemblages [76]. The lagoon is heavily degraded, characterized by high turbidity, sedimentation, and a benthos dominated by sponges with very few corals observed [77]. Coral reefs are concentrated on the westward reef terrace with algae accumulation primarily along the reef crest and within the lagoon. Over 50% of the hard coral at the site is comprised of *Montipora*, *Porites*, *Pocillopora*, and *Pavona* genera [78]. The isolated area of interest containing the Palmyra atoll with masked land and cloud pixels is shown in Figure 3-1.

### **Kingman Reef**

Kingman Reef is the northern-most reef in the Line Island chain located 68km northwest of the Palmyra Atoll at 6°23'N 162°25'W. The atoll is triangular, stretching 18km east-west and 9km north-south [79] with shallow (<2m) reefs along the southern and northern sides that are connected by a deeper reef (>20m) along the western terrace. The atoll contains two small rubble islands near the eastern ends of the shallow reefs but lacks permanent emergent land. The lagoon is predominantly deep (> 30m) with large patch reefs that range from 50 to 200m in diameter and extend to within 2 to 10m of the surface. The lagoon side of the reef crest is comprised of a steeply sloped back reef habitat. The fore reef is consistent along the northern and southern coasts originating with a gradually sloped terrace extending 30 to 60m from the reef crest with a drop-off beginning at ~20m

deep [80]. The most commonly occurring hard coral genera within Kingman Reef are Porites, Pocillopora, Acropora, and Favia which comprise >50% of all hard coral cover at the location [78]. Figure 3-1 exhibits the Kingman Reef area of interest with masked cloud cover. In addition, a small portion of exposed rubble can be observed on the eastern portion of the northern reef crest.

### **Baker Island Atoll**

Baker Island Atoll is an outlier island of the Phoenix Island Archipelago and classified as a low reef island. The shallow marine benthic habitats consist of fringing reef crests, shallow back reefs, steep fore reefs, spurs-and-grooves, and small reef terraces. The west, north, and south of the island consists of steep reef slopes that descend to great depths [81]. The easterly side of the island is characterized by spurs-and-grooves and oligotrophic waters off reef terraces [82]. However, the island's proximity to the equator (0°12'N 176°29'W) causes it to be influenced by both the westward-flowing Southern Equatorial Current at the surface and the strong eastward-flowing Equatorial Undercurrent, resulting in nutrient rich topographic upwelling on the western side of the island [83]. Acropora comprise >60% of coral observed at the location while Fungia and Pocillopora are also common genera [83]. Figure 3-1 includes a display of the Baker Island Atoll area of interest with land and cloud mask applied.

### **Howland Island**

Howland Island is located at 0°48'N 176°37'W, just 66km northwest of Baker Island, however they fall in two different Landsat 8 scenes. As a result, they share many common

environmental features, but the two sites are captured in separate path-row images. Like Baker Island, Howland Island is also classified as a low reef island and considered an outlier island of the Phoenix Island Archipelago. The Island's geomorphology consists of a narrow, shallow fringing reef and a steep slope that descends to great depths just off the coast. The western coast of the island is sandy and low while the waves and trade winds have caused the eastern side to be more abrupt and covered with coral rubble [84]. Given the island's proximity to the equator, it also is impacted by topographic upwelling of rich nutrients as the Southern Equatorial Current and Equatorial Undercurrent encounter the abrupt west slope of the island [83]. However, the hard coral species that exist in the habitat are somewhat different from what can be observed at Baker Island. The most abundant genera are Pocillopora, Pavona, Porites, Montipora, and Fungia which is more similar to the representation found at Palmyra Atoll [83]. Figure 3-1 includes a presentation of the Howland Island area of interest isolated for study and processing with a land and cloud mask applied.

### **3.3 Methodology**

The overall layout of the processing approach applied to the three selected scenes is shown in Figure 3-2. The approach can be broken down into six major components: preprocessing the image for land, water, and cloud masking, atmospheric and water column correction, LDA model training, LDA model application, model performance evaluation, and analysis of model generalization properties. These steps were repeated for each of the four sites within the three scenes. Scenes were selected to minimize the presence of clouds in the image. Pixels that still suffer cloud cover or other obstructions were then identified

and masked. This is followed by creating a water mask by applying a threshold to the NIR band. The reflectance cutoffs employed for each scene in this study ranged from 0.31 to 0.325. The water body and corals have similar spectral reflectance, which may lead to misclassification in water/coral areas. A deep-water AOI was selected to be used in atmospheric correction via the dark-pixel subtraction method [37] [50] [51]. An LDA application was then implemented and trained on the Palmyra Atoll site data. LDA is a frequently applied algorithm in studies involving classification based on remote sensing data. The algorithm attempts to discriminate discrete classes using a linear combination of continuous independent variables. In this way the algorithm characterizes coral pixels based on a pooled covariance matrix of the pixel values and prior probabilities of the classification groups. A decision boundary between the classes is determined and observations are assigned to the class from which it has the smallest squared deviance [85] [86]. Accuracy of the algorithm was evaluated using leave one out cross-validation. The model trained using Palmyra Atoll site observations was then applied to Kingman Reef site data (also within Path/Row 065/056) and the resulting performance was evaluated. Finally, the same model was applied to sites in several different scenes namely Baker Island Atoll and Howland Island. The resulting model performance was measured, analyzed, and implemented using the open source R programming language and environment for statistical computing [52].

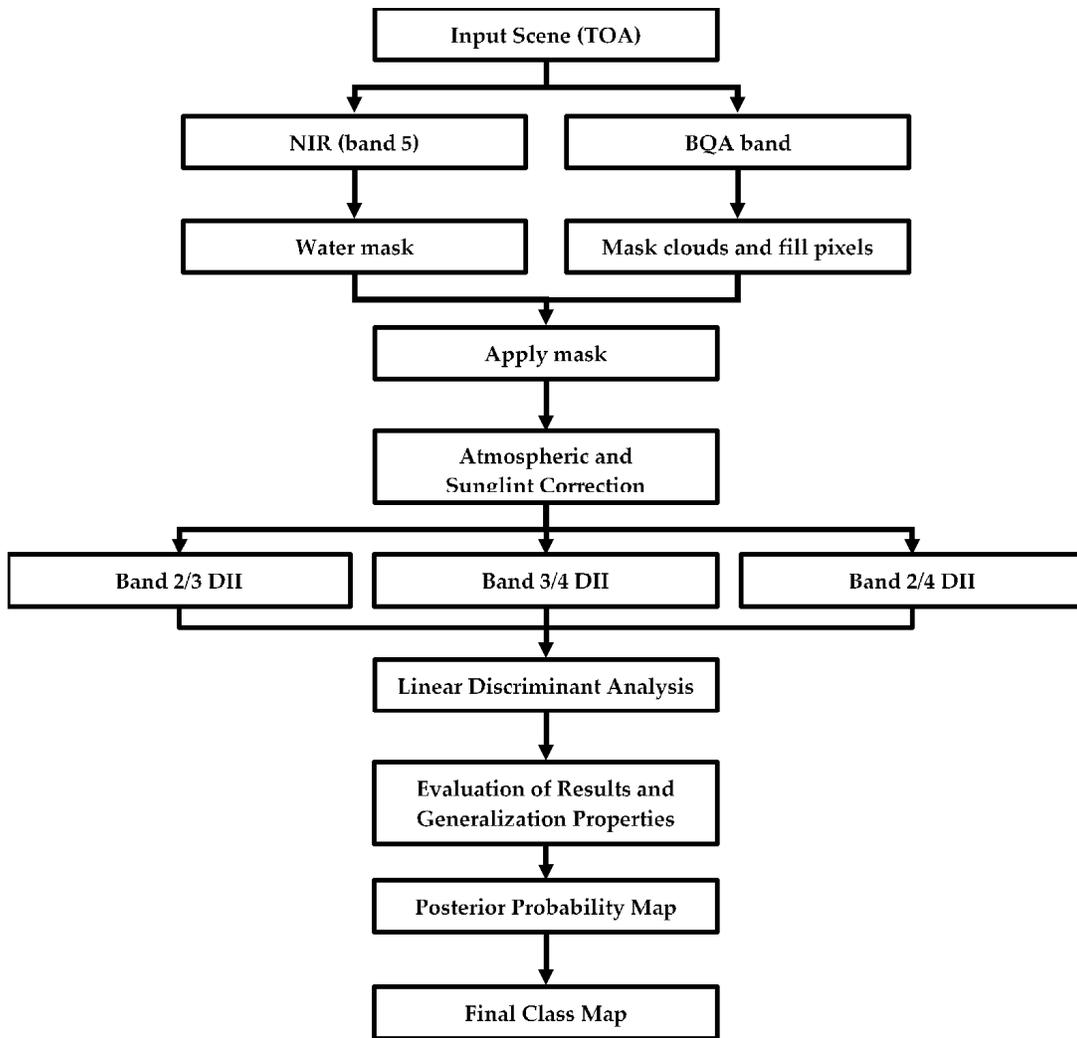


Figure 3-2 Scene masking, atmospheric correction, and water column correction process flow.

### 3.3.1 Cloud and Quality Mask

The first step taken to preprocess the selected images was the identification and masking of pixels obscured by cloud cover and shadows. While scenes were selected to minimize the amount of cloud cover none of the images were completely devoid of clouds. Given the location of the sites, the Pacific Ocean near the equator, and the six-month window constraint, the probability that a satellite image would be taken on a cloud free day is small. Masking of clouds was performed by leveraging the Landsat 8 product quality

band [53] which provides a per pixel approximation of confidence that a given condition exists [54]. There are several possible ways in which bias from cloud obstruction can contaminate an analysis. The obvious entry point is disrupting the reflectance of a pixel or group of pixels. In addition to directly influencing surface reflectance, cloud cover present in the deep-water AOI can alter the values used for atmospheric correction applied to the image through the dark-pixel subtraction method [55]. As a result, this initial step of masking cloud interference is a critical preprocessing step.

### **3.3.2 Water Mask**

The study of DII related metrics across large scenes requires preprocessing that includes masking land in addition to clouds and shadows. This step is imperative as including land pixels can distort the DII parameters when calculated across a scene. The water mask was created by leveraging the Landsat 8 NIR sensor. This sensor measures light between 0.851 and 0.0879  $\mu\text{m}$ . Water absorbs light in these wavelengths, therefore, it is a good candidate for discerning water from land in any given scene [56]. As in the visible bands, the Landsat 8 NIR band (band 5) is at 30m resolution. A threshold was applied to the NIR band pixel values of each scene. The plots were then evaluated visually to determine the most appropriate cutoff for separation of land and water. A mask was created for pixels determined to be water based on this threshold value.

### **3.3.3 Atmospheric Correction**

In the visual bands, 90% of the at sensor reflectance depends on atmospheric and water surface properties [57]. Therefore, atmospheric correction is first performed using the dark

pixel subtraction method [58]. This method selects areas of the scene with water known to be deep enough for the visible bands to fully attenuate. Signal received from these areas are comprised of atmospheric radiance and surface reflectance, thereby isolating the impact of these elements. Assuming the atmospheric and water surface conditions generalize to the rest of the scene (i.e. uniform throughout the area of interest), the mean deep-water radiance at sensor can be leveraged to correct for the effect of atmospheric and surface reflectance interferences [1] [59] [60]. Depths greater than 50m will assure that the visible wavelengths have fully attenuated [61]. In addition, two standard deviations are subtracted to account for possible sensor noise [62]. It is important to highlight the assumption that conditions are uniform across each scene. In addition, because of this assumption, the deep-water AOI selected should appear in the same scene that is being analyzed. This will minimize the possibility of unintended bias that may be introduced by leveraging a deep-water AOI of another image and that, to the greatest extent possible, the effect of full attenuation of the wavelengths for each scene is uniquely isolated.

### **3.3.4 Water Column Correction**

As light penetrates water, the intensity decreases exponentially with increasing depth. The rate of attenuation is wavelength-dependent and has a severe effect on the remote sensing based detection of aquatic habitats [1]. Therefore, water column correction is appropriate for imagery with multiple water-penetrating spectral bands [51] [63]. Within these visible spectral bands, longer-wavelength blue bands attenuate less rapidly than shorter-wavelength red bands. Therefore, the spectral radiances recorded at sensor are dependent on both the subsurface strata reflectance and depth. The confounding influence

of depth can create significant distortions in the subsurface reflectance. Since most marine habitat-mapping exercises are only concerned with mapping benthic features, it is advantageous to remove the influence of variable depth [1] [63] [64] [65] [66].

## **3.4 Results**

Using the masked, corrected scenes, we developed a classifier to identify the existence of coral in a pixel. We applied the predictor to an alternate site within the same Landsat scene (Kingman Reef). Finally, we applied the predictor to multiple sites in different Landsat scenes (Baker Island Atoll and Howland Island).

### **3.4.1 Generalization Performance by Site**

#### **Palmyra Atoll**

The model was initially trained and applied to data from the Palmyra Atoll site of Landsat Path-Row scene 065/056. This site contained a total of 82 unique observation points. Of these, 66 were in unobscured pixels. For each valid observation point the corresponding pixel index values were extracted from the masked, atmospherically corrected, and water column corrected scene based on location. The extracted DII values were then matched to the ground truth class. The LDA model was trained on these observations.

For each pixel, the posterior probability of that pixel belonging to the coral class was calculated. The resulting posterior probability for each pixel belonging to the coral class is shown in Figure 3-3. The resulting map of predicted values was analyzed and compared to

known geomorphology of the site for context validation. Figure 3-4 displays the final class predictions by pixel based on the trained model.

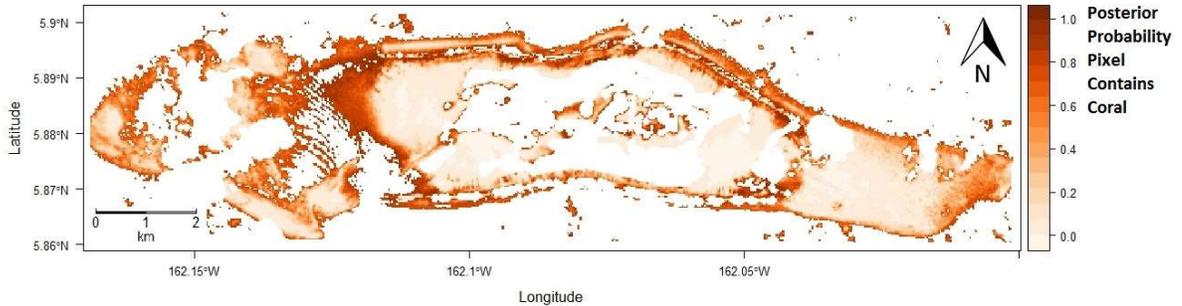


Figure 3-3: Palmyra Atoll: Plot of the posterior probability for belonging to the coral class for each pixel in the Palmyra Atoll site.

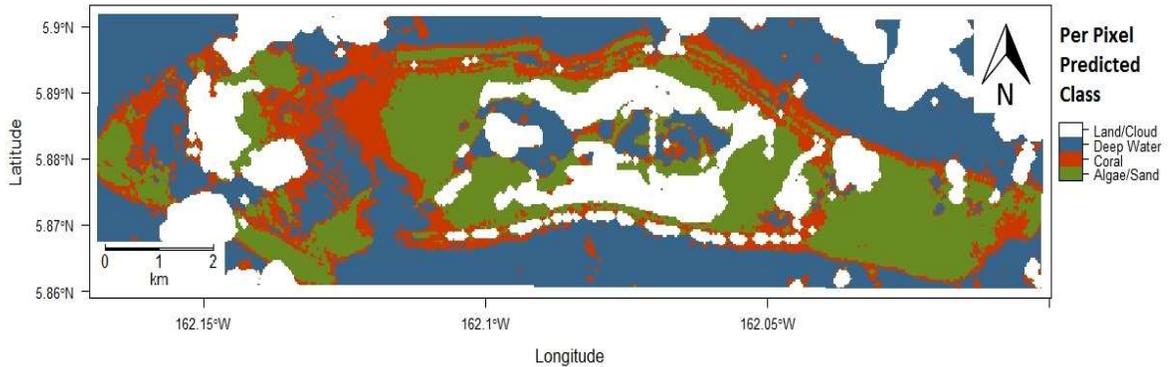


Figure 3-4: Palmyra Atoll predicted class membership based on posterior probabilities.

### Kingman Reef

The first evaluation of how well a supervised classifier can generalize across multiple sites was evaluated by applying the model trained on the Palmyra site data to labeled data from another location within the same Landsat scene. This was done by predicting labeled data from the Kingman Reef using the algorithm trained on data from the Palmyra Atoll. The Kingman Reef is a small reef near Palmyra Atoll therefore the two locations reside within the same Landsat image. Sites within the same scene will have

similar atmospheric and water conditions, an assumption that previous in-situ studies have relied upon heavily. Water turbidity, in particular, is an important assumption as water column correction has yet to produce a proven method for correcting such interferences or even evaluating its impact on predicting subsurface benthic habitat bottom type.

There were 57 bottom type observations available for the Kingman Reef site. Of these, 42 could be matched to valid pixels in the masked, atmospherically corrected, and water column corrected scene. The trained model correctly predicted 78.57% of the observations in the Kingman Reef site. This is evidence that the model built using water column corrected indexes generalizes well to sites within the same scene. The decrease in accuracy from Palmyra Atoll to Kingman Reef of less than 1% is strong evidence for the within scene homogeneity assumption relied on so heavily in previous in-situ studies. The strong performance is a reflection of the location-based environmental conditions that the two sites have in common. Specifically, because the two sites are within close proximity, they share similar atmospheric and geomorphological conditions. In addition, the marine species are likely to be more homogeneous compared to sites separated by a greater distance. Finally, water conditions at each site is likely to be similar but not identical. Site-specific variation in water conditions can occur within a scene due to differences in how tides, currents, and other natural phenomena interact with site-specific geomorphological characteristics. The results indicated that the consolidated impact of all these factors amounts to less than 2% decrease in the accuracy of predicting bottom type between two sites within the same Landsat image. The resulting posterior probability that a given pixel belongs to the coral class are presented in Figure 3-5 and the overall class predictions are presented in Figure 3-6.

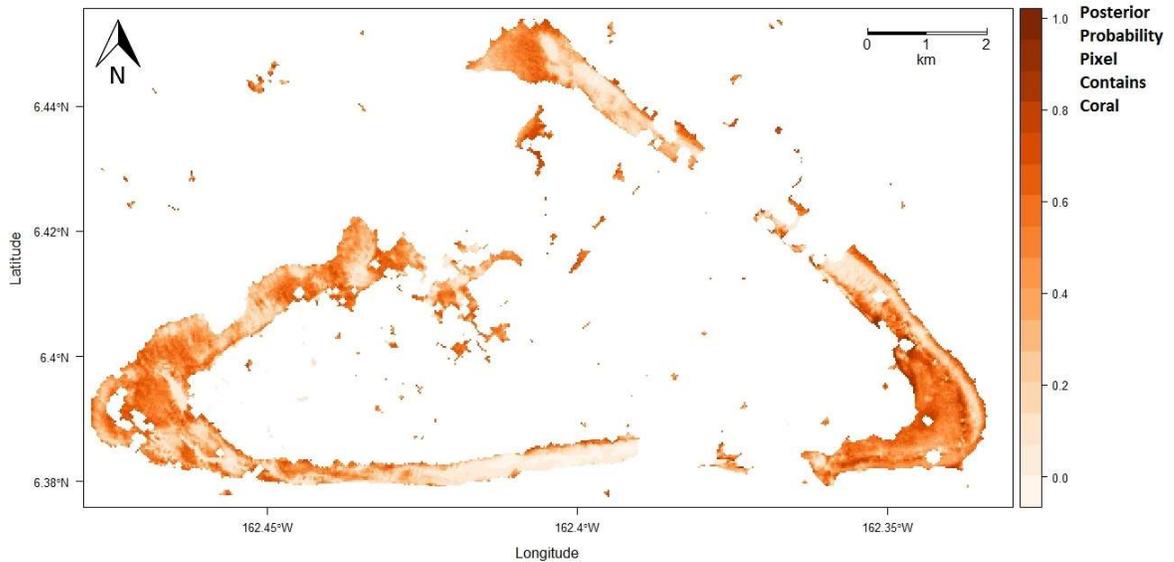


Figure 3-5: Kingman Reef: Plot of the posterior probability for belonging to the coral class for each pixel in the Kingman Reef site.

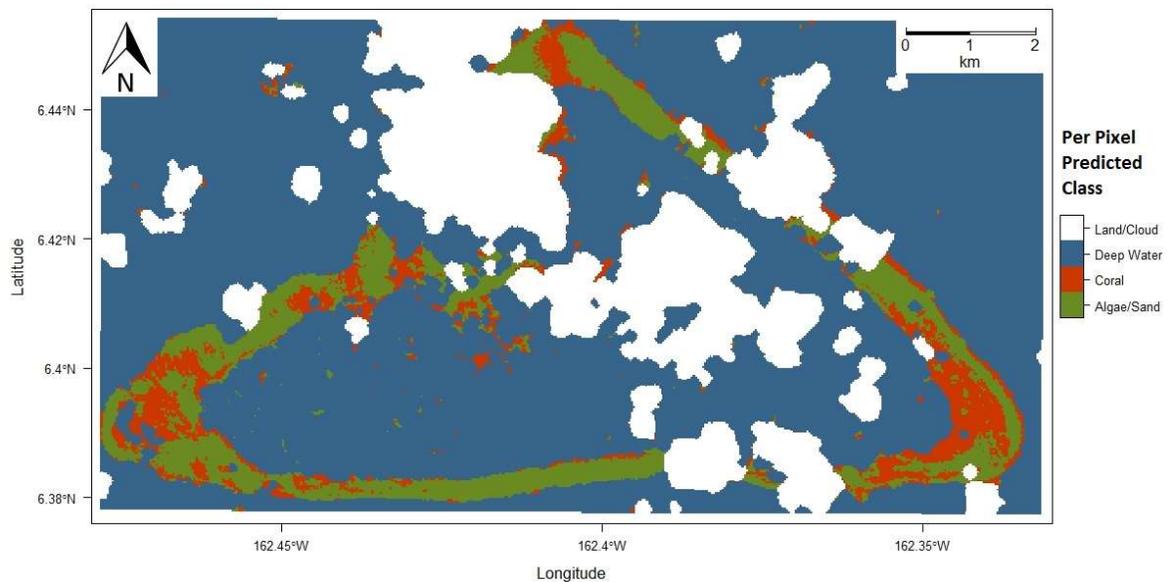


Figure 3-6: Kingman Reef predicted class membership based on posterior probabilities.

### **Baker Island Atoll**

The ability for the classifier to generalize to a different scene is an important result. The conditions between scenes can change significantly. In addition, the variation of marine life species represented in different sites can alter the observed DII pixel value.

Even more importantly, there are likely to be variances in how reefs naturally formed due to alterations in the site-specific geomorphology. Even with a uniform approach to atmospheric and water column correction, the environmental changes across scenes can be substantial.

The ground truth data from the Baker Island Atoll location included 26 bottom type observations all of which were associated with a valid DII pixel value. The classifier trained on Palmyra Atoll data was able to correctly classify 69.23% of these observations. The decrease in accuracy can be attributed to changes in site-specific conditions between Baker Island Atoll and Palmyra Atoll. Differences in how currents and tides impact water turbidity as well as other environmental impacts reduce the ability for observations to appropriately represent other locations. Previous in-situ studies have relied on the assumption of uniform water conditions. This is because, while there may be significant obstruction of light due to matter floating in the water, it is not likely to be substantially different across a single scene. Therefore, because all observations are distorted by approximately the same amount, it only represents a uniform amount of noise across all pixels in a given scene. However, when data from one scene is used to evaluate a different scene, these changes in localized or image specific conditions become more pronounced. For example, changes in water conditions including turbidity can vary within a scene but is likely to be more similar between two sites within a scene compared to two sites located in separate images altogether. First, scenes in two separate images are likely to be separated by greater distance than those in the same scene. This gives rise to greater environmental fluctuation between sites that are separated by enough distance to be in two separate Landsat images. There are multiple similar reasons for this fluctuation in water conditions.

Namely, the impact of currents and tides can vary from site to site. The unique location of Baker Island Atoll relative to the equator compared to Palmyra Atoll and Kingman Reef expose it to more topographic upwelling of nutrient rich waters which in turn impact the type of coral at the site and hinder the generalization of the trained model. In addition, the images of two different scenes were captured on different times and dates that likely had varying weather and ocean conditions. The consolidated impact of variation on accuracy due to the limited observations and to generalizing beyond the in-situ image is now measured. For example, the total impact of site-specific environmental conditions between Palmyra Atoll and Baker Island Atoll represents an 11% decrease in accuracy between the two sites. The resulting per pixel posterior probabilities for the bottom type to be coral and the per pixel predicted class are presented in Figures 3-7 and 3-8 respectively.

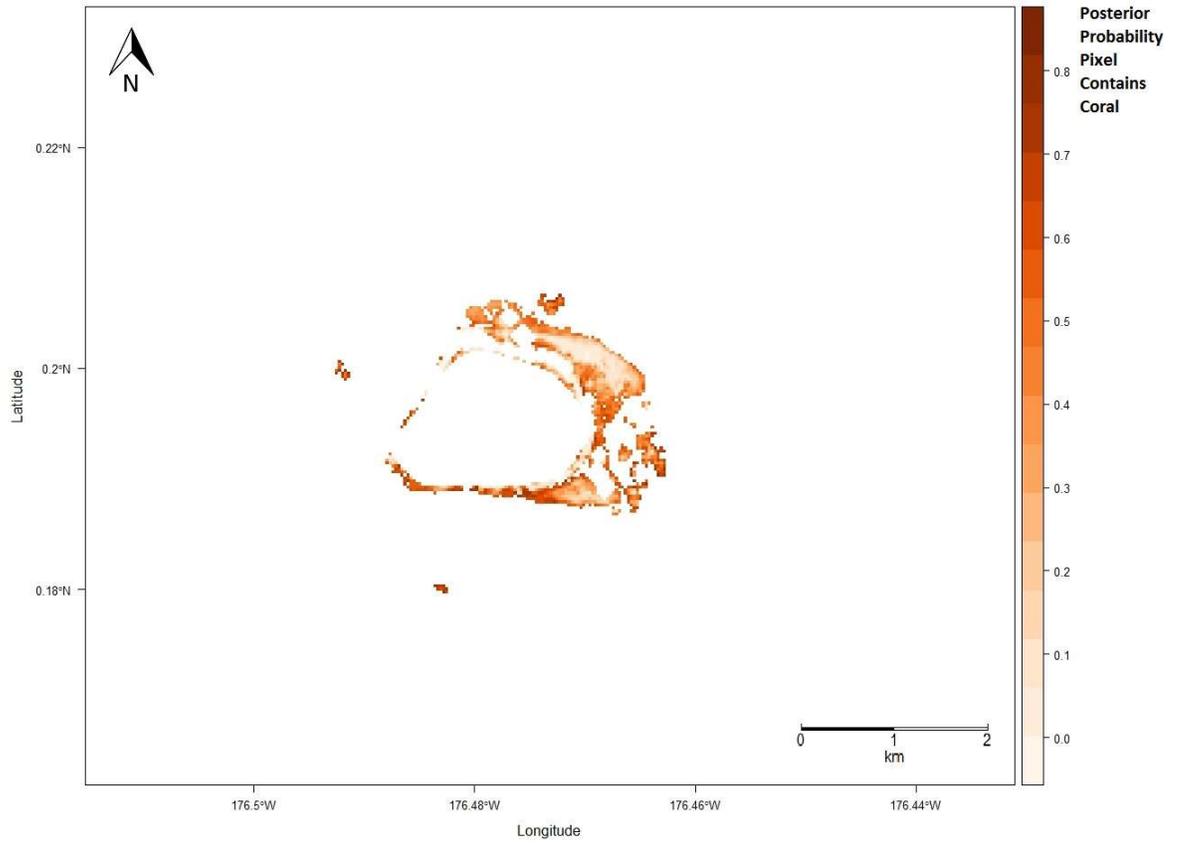


Figure 3-7: Baker Island Atoll: Plot of the posterior probability for belonging to the coral class for each pixel in the Baker Island Atoll site.

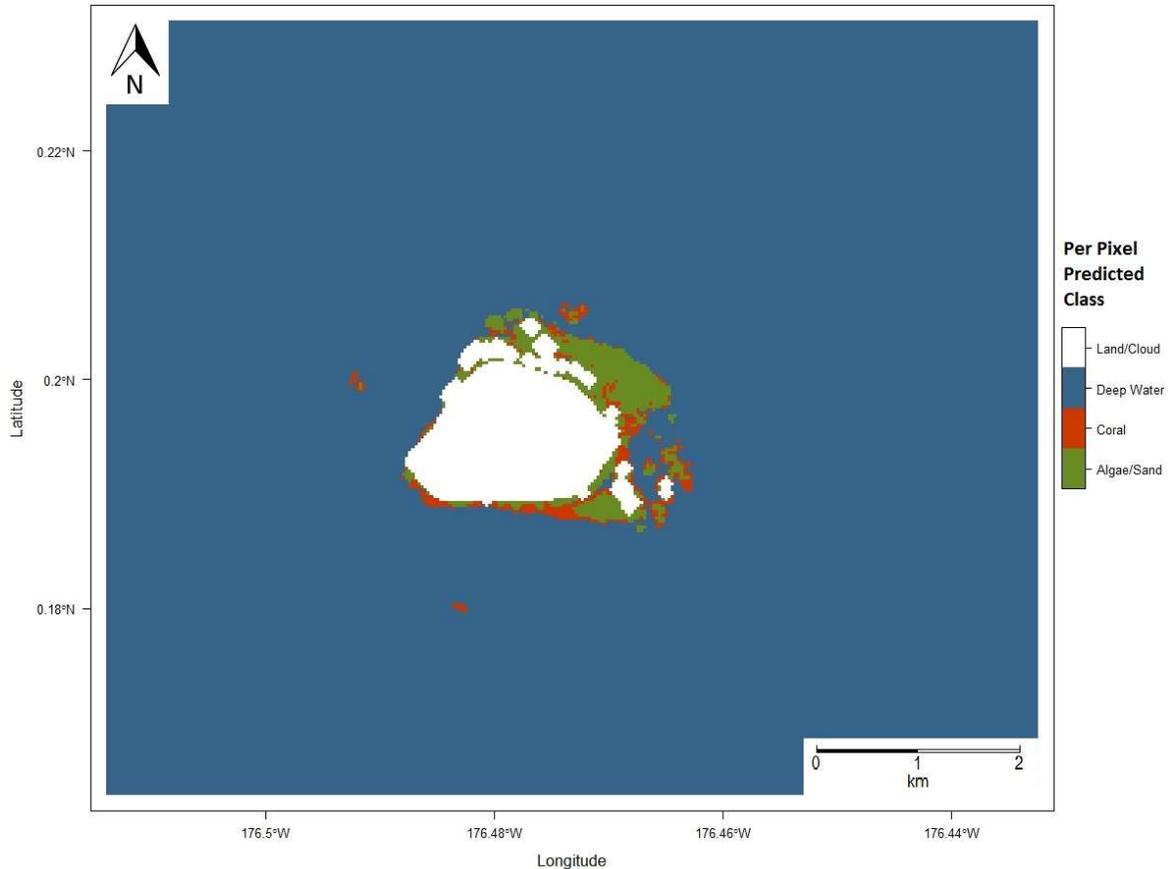


Figure 3-8: Baker Island Atoll predicted class membership based on posterior probabilities.

### Howland Island

Given the proximity and geomorphologic similarities between Howland Island and Baker Island, it can be assumed that the environmental conditions that exist within the two sites are similar relative to some of the other locations. Therefore, the accuracies produced when the Palmyra Atoll model was applied to each site are similar but not identical. This result is informative in isolating the impact on generalization due to environmental variation associated with location vs the impact on generalization due to the timing in which the Landsat image was captured. There were 30 observations of bottom type at the Howland Island site of which 28 could be assigned valid depth invariant pixel values. When

applied to the Howland Island site the model correctly predicted 71.43% of the observations.

Figure 3-9 presents the per pixel posterior probability that the pixel contains coral. Because the two sites are not close enough to reside within the same Landsat scene, they do not get the benefit of uniform interference across the scenes. The result is accuracy scores that are comparable due to environmental and species similarities within the two sites but not identical. The consolidated effect of these variations in conditions between the Palmyra Atoll training data and Howland Island test site is an 8.87% decrease in accuracy of bottom type prediction between the two sites. Figure 3-10 presents the final class predictions for each pixel in the Howland Island area of interest.

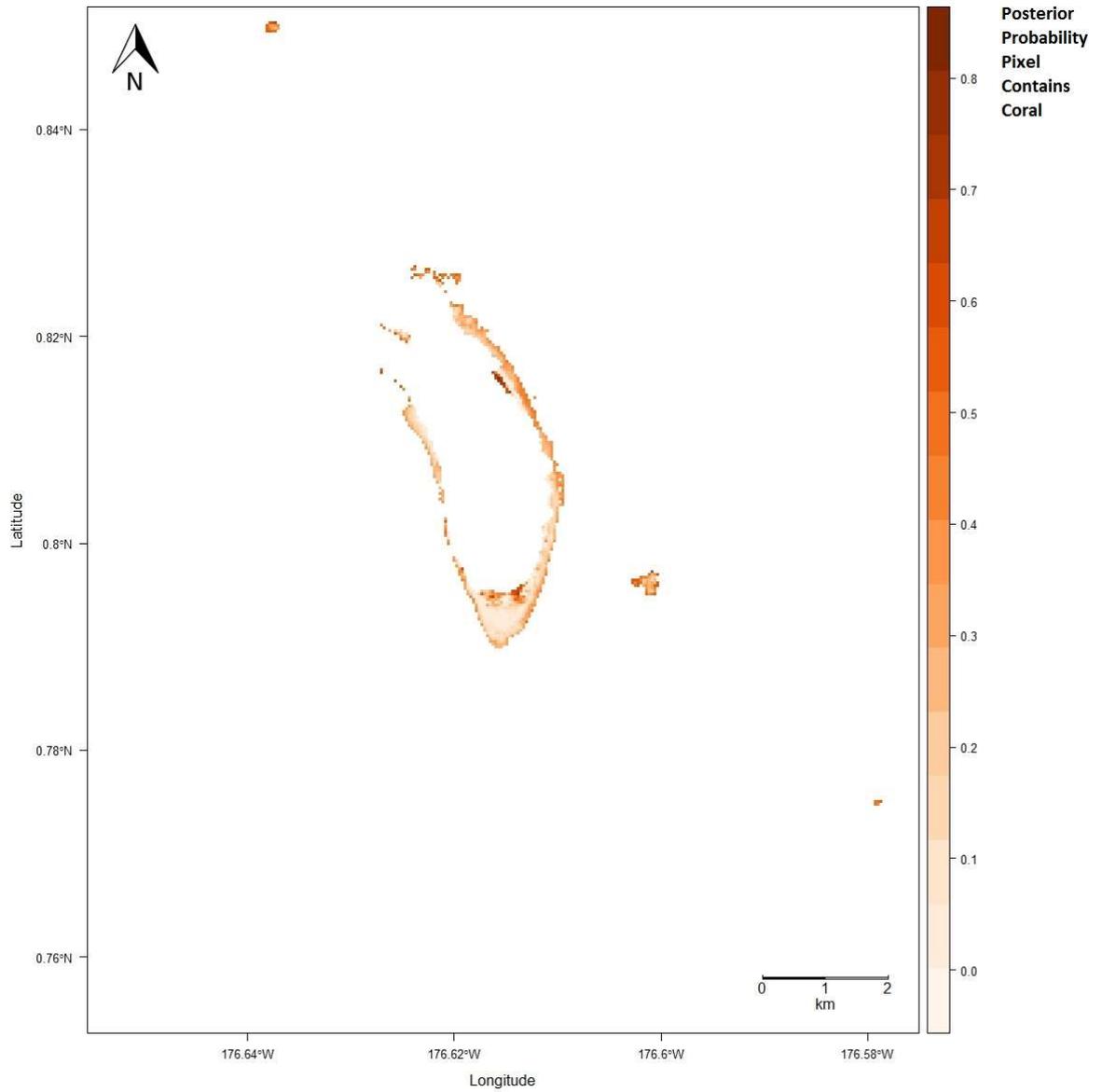


Figure 3-9: Howland Island: Plot of the posterior probability for belonging to the coral class for each pixel in the Howland Island site.

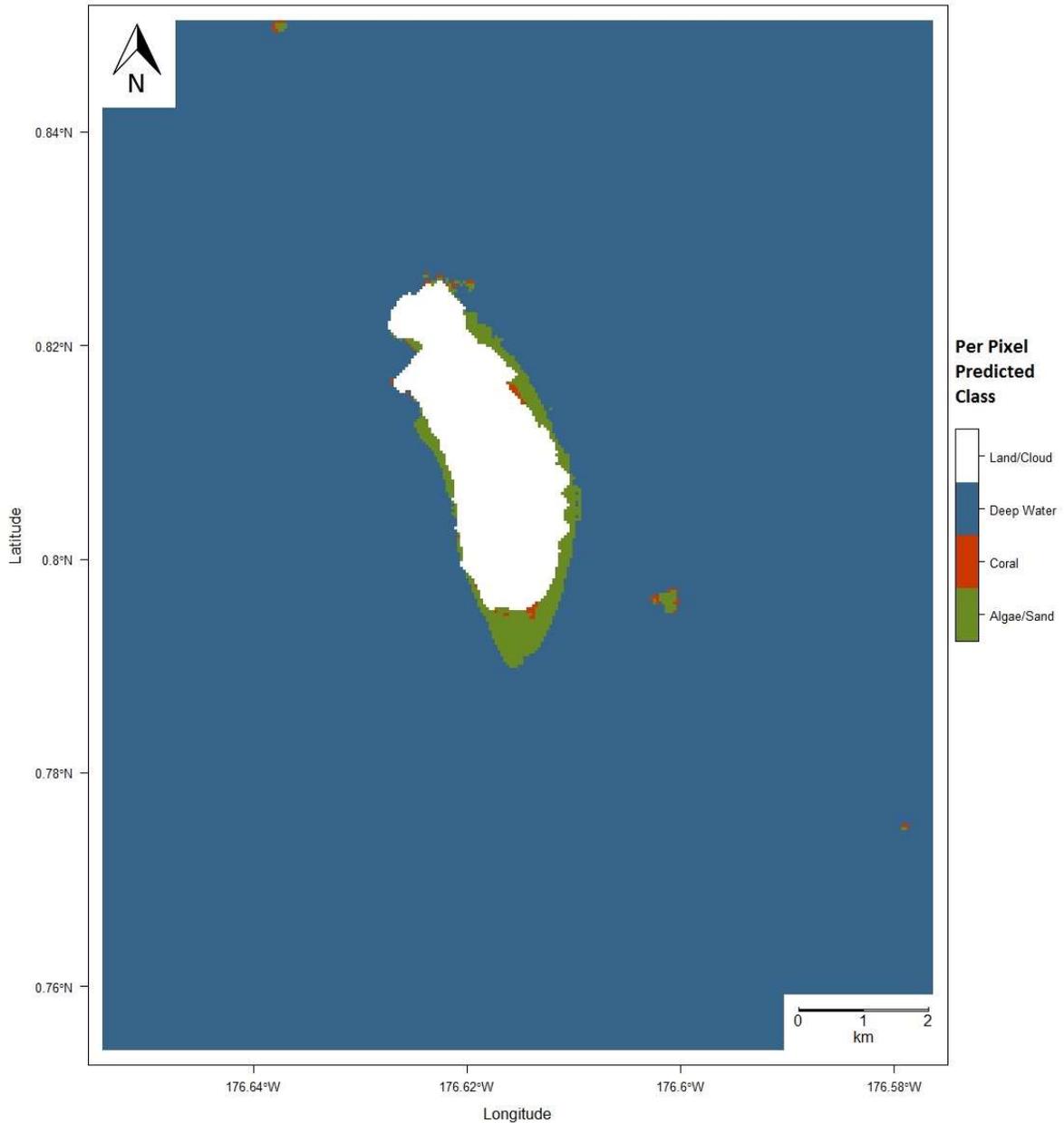


Figure 3-10: Howland Island predicted class membership based on posterior probabilities.

### 3.4.2 Quantitative Assessment of Site-Specific Generalization

The algorithm correctly classified 80.30% of the observations within the Palmyra Atoll site and obtained a precision of 0.7800 and recall of 0.9512. Precision and recall are very common statistical measure for measuring type I and type II error rates. Precision is

a measure of Type I errors and commonly thought of as a measure of the exactness of an algorithm. Recall is a measure of Type II errors and is often thought of as a measure of how completely the algorithm retrieves information. The harmonic mean of precision and recall is known as the F-measure. The results of the model application to the Palmyra Atoll site yielded an F-measure of 0.8571. Finally, specificity, which is an indication of the algorithm's ability to differentiate between true negative observations, was 0.5600 when evaluated against the Palmyra Atoll truth data.

When applied to the Kingman Reef site, the algorithm obtained similar performance to that observed in the Palmyra Atoll site. Specifically, the Kingman Reef application accurately classified 78.57% of the observations with precision and recall of 0.8276 and 0.8571, respectively. This indicates that the model outperformed the original site in type I error when applied to the Kingman site but committed more type II errors as well. The model obtained a specificity of 0.6429 and an overall F-measure of 0.8421 both of which are similar results to those obtained from the Palmyra site. The similarity between the two sites can be attributed to their close proximity and commonality in environmental conditions.

Evaluation of the model performance when applied to the Baker Island Atoll site revealed 69.23% of the ground truth observations can be correctly classified. Further evaluation of the algorithm results show that the algorithm yielded precision and recall of 0.6522 and 1.0000, respectively. The F-measure for the application was 0.7895 which was similar to the measure produced in the evaluations of previous sites. Specificity equal to 0.2727 was obtained.

The Howland Island application produced more type II errors than observed in other scenes and, as a result, had a low recall score of 0.5333. Specificity scored higher than the applications to other sites with a score of 0.9231 due to fewer type I errors. Due to the low type I error rate, precision was high at 0.8889. The F-measure was the lowest of all applications at 0.6667. These results are summarized in Table 3-3 followed by the detailed confusion matrices in table 3-4.

Table 3-3: Assessment Metrics for Evaluation of Model Performance for Each Site and Consolidated Input.

	<b>Palmyra Atoll</b>	<b>Kingman Reef</b>	<b>Baker Island Atoll</b>	<b>Howland Island</b>	<b>Consolidated Sites</b>
Accuracy	80.30%	78.57%	69.23%	71.43%	74.07%
Precision	0.7800	0.8276	0.6522	0.8889	0.7244
Recall	0.9512	0.8571	1.0000	0.5333	0.9293
Specificity	0.5600	0.6429	0.2727	0.9231	0.4444
F-measure	0.8571	0.8421	0.7895	0.6667	0.8142

Table 3-4: Confusion Matrices by Site and Consolidated Inputs.

		<b>Ground Truth Labels</b>	
		<b>Coral</b>	<b>Not Coral</b>
<b>Predicted Class</b>	<b>Palmyra Atoll</b>		
	<b>Coral</b>	39	11
	<b>Not Coral</b>	2	14
	<b>Kingman Reef</b>		
	<b>Coral</b>	24	5
	<b>Not Coral</b>	4	9
	<b>Baker Island Atoll</b>		
	<b>Coral</b>	15	8
	<b>Not Coral</b>	0	3
	<b>Howland Island</b>		
	<b>Coral</b>	8	1
	<b>Not Coral</b>	7	12
	<b>Consolidated Sites</b>		
	<b>Coral</b>	92	35
	<b>Not Coral</b>	7	28

### 3.4.3 Robust Combined Model

The accuracy for any given site is constrained by how well the observed values used to train an algorithm represent the general population, and how well the algorithm itself can correctly model the relationship between inputs and the target variable. One way to ensure that the observation data adequately represents the population is to increase the breadth of observations used to train the predictor. Having analyzed the variation in accuracy across the various sites, we retrained the model using observations from all four sites. The resulting model was evaluated using leave one out cross-validation (LOOCV) for the classification of 127 coral pixels and 35 non-coral pixels. Of the 127 coral predictions 92 were correct and 35 were incorrect. Of the 35 not-coral predictions, 28 were correct and 7 were incorrect which corresponds to an accuracy of 74.07%. The classifier yielded a precision of 0.7244 and recall of 0.9293 indicating good performance on type II errors and marginal performance against type I errors. The model yielded a specificity of 0.4444 and the final F-measure was 0.8142.

These results expose the novel finding that a model can be trained on data from multiple Landsat sites and yield robust predictions of coral. The associated confusion matrix can be found in table 3-4. The resulting model yielded strong results to data collected from multiple sites and multiple scenes. This is a demonstration of the robustness of Landsat data to generalize across scenes. However, the model did not obtain the same accuracy produced using data from a single scene or even the model applied to an alternate site within the same scene. This is due to the impact of variation in environmental conditions between scenes. Most notably, there are often significant differences in water

conditions and turbidity across various locations. However, the impact of these differences can be mitigated as the more sites are considered for training the model. In addition, there are variances in the marine species that exist across locations. These two factors confound the inputs for a predictive model and lead to lower accuracies. Previous research has relied on assumptions of homogeneity. Comparing the results of the Palmyra Atoll and Kingman Reef sites within the same scene confirms this assumption although with a slight drop in accuracy. Figure 3-11 identifies the receiver operating characteristic (ROC) curve resulting from the model developed using the consolidated site data. This is an important visualization used to identify performance as the power as a function of the type I error or recall (true positive rate) as a function of  $1 - \textit{Specificity}$  (false positive rate). The area under curve (AUC) is an important diagnostic for evaluating model performance related to the ROC curve. In this instance the value of the AUC was 0.7298.

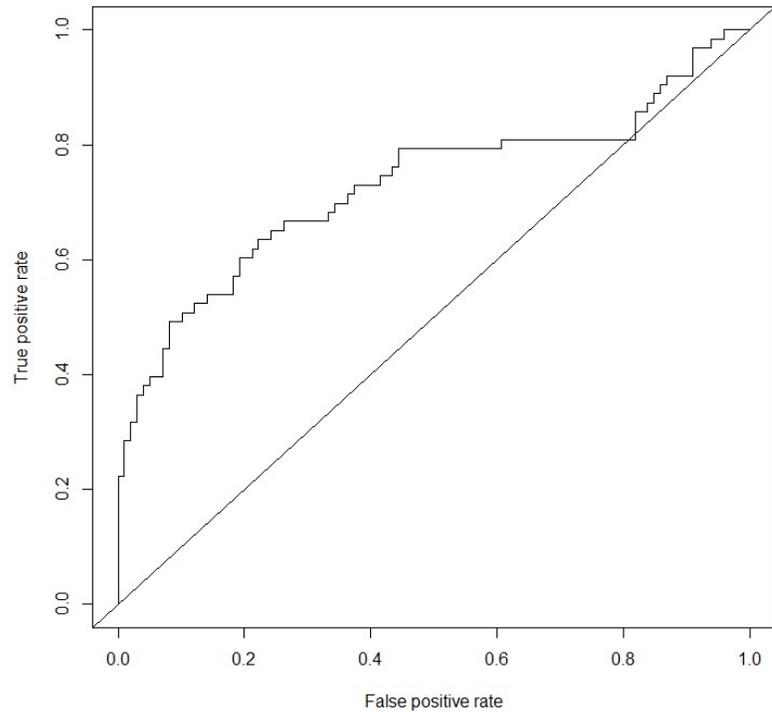


Figure 3-11: ROC Curve indicating model performance of the algorithm when applied to the consolidated ground truth set (AUC = 0.7298)

## 3.5 Discussion

### 3.5.1 Spectral Signature Generalization Properties for Coral Reef

#### Classification

The ability for a supervised classifier to generalize across sites is a critical outcome of this research. Previous research has concerned analysis and classifiers strictly limited to in-situ sites. This research advances the field of subsurface identification by evaluating the ability for Landsat information to generalize across sites and scenes. We measured this by locating five different sites known to contain benthic areas with coral reefs. We then obtained corresponding ground truth labels for a sampling of coral and non-coral pixels for each location. A classifier was trained using linear discriminant analysis to predict the

presence of coral. This model yielded leave one out cross-validation accuracy of 80.30% in the training site and only a small decrease in accuracy when applied to a remote site within the same scene. The accuracy of the model when applied to the Kingman Reef site was 78.57% demonstrating that while conditions can change across a given scene the impact is minimal. This validates the classic assumption made by previous in-situ studies of homogeneity of conditions across a given Landsat scene. We then applied the model to several other sites. Baker Island Atoll yielded an accuracy of 69.23%, and Howland Island produced an accuracy of 71.43%. These lower accuracies are due to two primary reasons. First, variances in the types of coral and algae species that live in a given ecosystem create small variations in the signal that is received by the satellite. This diversity of life is related to each ecosystem's adaptation to the surrounding geomorphic conditions. Changes in environmental conditions and related noise create disturbances that manipulate the signal received by the sensor. While atmospheric and light attenuation due to water column penetration can be corrected, the interference due to localized water turbidity cannot. This can account for a decrease in accuracy of predicting coral by more than 10% as shown in this study. The interference due to water turbidity is not uniform and some scenes are impacted more than others. Finally, a consolidated model was created using the strength of observations across all sites. Accuracy of 74.07% was calculated using leave one out cross-validation. The resulting model demonstrated a robustness to some of the perturbations mentioned.

### **3.5.2 Methodology Benefits and Challenges**

The outcomes presented in this study enable a solution to reduce location-based bias of the spectral signature of coral reefs. This represents a strong advantage over previous in-situ studies which require localized observation data in order to detect coral reefs. The results presented in this study indicate that the spectral signature information of coral contained in one site can be leveraged to evaluate another, unobserved site with up to 71% accuracy. Furthermore, the results of this study indicate that a robust model can be created by leveraging the consolidated information from several sites and produce accurate predictions for coral of up to 74%. These results are a key component required for the progression from in-situ analysis to large scale spatial analysis of coral reefs.

The primary challenge of the proposed method are the lower accuracy scores as compared to that of in-situ analysis. Previous in-situ analysis based on Landsat data generally obtain accuracies of up to 80% [1], as was obtained in the site-specific study of Palmyra Atoll here. The reduction in accuracy is an indication of variation in the geomorphology and ecology of the robust model. In-situ analyses incorporate this bias into the training of their predictors and therefore yield higher accuracies.

### **3.6 Conclusion**

The ability for a supervised classifier to generalize across sites is a critical outcome of this research. Previous research has concerned analysis and classifiers strictly limited to in-situ sites. This research advances the field of coral classification using remote sensing data by evaluating the ability for Landsat information to generalize across sites and scenes. This

outcome is part of a natural progression toward a global evaluation of coral health using satellite data both in the present and historically. Understanding the spectral generalization properties of coral enables more robust evaluation of many reefs which have not been studied before. Furthermore, this research provides a baseline accuracy for evaluating the presence of coral in locations for which observations are unavailable or have not been made. The classification accuracy of the model trained on a known site and applied to a new site is less than that of previous in-situ analysis because it does not incorporate the site-specific geomorphology and location-based bias. However, a model trained using data from multiple sites is more robust to these environmental variations and free of location specific bias. We measured the generalization criterion by locating four different sites known to contain benthic areas with coral reefs. We then obtained corresponding ground truth labels for a sampling of coral and non-coral pixels for each location. A classifier was trained using linear discriminant analysis to predict the presence of coral. This model yielded leave one out cross-validation accuracy of 80.30% in the training site and only a small decrease in accuracy when applied to a remote site within the same scene. The accuracy of the model when applied to the Kingman Reef site was 78.57% demonstrating that while conditions can change across a given scene the impact is minimal. This validates the classic assumption made by previous in-situ studies of homogeneity of conditions across a given Landsat scene. We then applied the model to several other sites. Baker Island Atoll yielded an accuracy of 69.23%, and Howland Island produced an accuracy of 71.43%. These lower accuracies are due to two primary reasons. First, variances in the types of coral and algae species that live in a given ecosystem create small variations in the signal that is received by the satellite. This diversity of life is related to each ecosystems'

adaptation to the surrounding geomorphic conditions. Even more, changes in environmental conditions and related noise create disturbances that manipulate the signal received by the sensor. While atmospheric and light attenuation due to water column penetration can be corrected, the interference due to localized water turbidity cannot. As seen in this study, this can account for a decrease in accuracy of coral prediction by more than 10%. Furthermore, the interference due to water turbidity is not uniform and some scenes are impacted more than others. Finally, a consolidated model was created using the strength of observations across all sites. Accuracy of 74.07% was calculated using leave one out cross-validation. The resulting model demonstrated a robustness to some of the perturbations mentioned.

# **4. Second Study: Coral Reef Change Detection in Remote Pacific Islands using Support Vector Machine Classifiers**

## **4.1 Introduction**

Coral reefs are among the most critical ecosystems in the world due to the role that they play in maintaining biodiversity and sustaining the lifecycle of so many marine species. Unfortunately, many large-scale mass mortality events associated with coral bleaching have been documented. These bleaching events are driven by a variety of anthropogenic and environmental influences [4] [5] [6] [7] [8] [9] [10] [11] [12].

A significant amount of previous research has been conducted to analyze coral reefs using remote sensing data [25] [26] [27] [28] [35]. These studies include leveraging multispectral sensors [29], hyperspectral sensors [30], and comparisons between the two [31]. In addition, studies have been conducted using satellites with a wide range of spatial resolutions from medium resolution Landsat based research [36] [61] to high-resolution Sentinel-2 based research [33] [34]. It has been shown that classifiers based on higher resolution platforms typically attain a greater degree of accuracy, often by more than 10%,

than lower resolution satellites [87]. This is due to the reduced within pixel mixing of benthic cover types when attempting to classify highly heterogeneous ecosystems such as coral reefs [88]. More recently, an object-based approach to coral cover detection has been explored [89] for geomorphological mapping and benthic community discrimination [90] [91] [92]. While the object-based approach is promising yielding accuracies greater than 90%, they do rely upon high-resolution imagery. It has been shown that the pixel-based approach yields no significant difference from the object-based approach when applied to medium-resolution satellites such as Landsat [93] [94]. The recent advancements in satellite technology have allowed high-resolution imagery to be readily available from multiple platforms. However, while these platforms show great promise for analyzing the state of benthic habitats currently and in the recent past, they lack the history for a longer-term perspective on change. Yet, the Landsat platform provides a rich inventory of historical images albeit at medium-scale resolution. What is more, the quality of the data provided by missions has been proven to be appropriate for temporal analysis [95] [96] [97] [98] [99] [100] [101] [102] [103] [104].

Technological advancements in remote sensing satellites has produced an extensive archive of images which are being used increasingly for scientific research on surface cover and cover change detection. The foremost example of this large historical archive of remote sensing images used for scientific research is the Landsat program which has been in operation since 1972 [72]. Due to the rich history and open source nature of Landsat, this platform has been used in the majority of change detection analysis [4] to detect the decline of coral habitats through temporal analysis even over extended time periods of 18 years or more [105]. Researchers suggest that historic Landsat imagery is the best available data

source for studies of historic changes in environments [106]. However, the majority of Landsat images have no corresponding ground truth data on which a classifier can be trained to identify cover. Therefore, each of these studies have been conducted using the well-known and well-studied convention of training a supervised classifier on the Landsat imagery captured within the same time period as the ground truth observations were made then applying that classifier temporally to historical images [72]. This methodology is known as signature extension [97] [98], or signature generalization [95] [96] and since the 1970's studies have been conducted exploring the methodology as applied to Landsat data [97] [99] using both individual Landsat satellites [47] as well as extended analysis over multiple Landsat satellites [107]. The scope of these temporal cover change detection studies vary from seagrass and coral reefs to mangroves [108] to forest and crop land [109]. Many studies have been conducted to validate this approach. Specifically, it has been established that there is no statistical difference ( $p = 0.303$ ) between changes in coral habitat areas as observed by the Coral Reef Evaluation and Monitoring Project (CREMP) and change detection analysis conducted using a combination of Landsat missions [105]. Additional validation studies demonstrated that change analysis conducted based on Landsat 5, Landsat 7, and Landsat 8 was accurate with an overall accuracy of  $88.9\% \pm 1.0\%$  and a kappa-coefficient of 0.86 [109]. This study expands upon the previous limited scope analyses by applying a classifier to two new sites [110] [111] [112] [113] [114]. Therefore, in addition to training a classifier and performing change detection analysis for each of four previously unanalyzed sites (Palmyra Atoll, Kingman Reef, Howland Island, and Baker Island Atoll), this study leveraged a robust classifier longitudinally applied to two additional sites (Tabuaeran Island and Kiritimati Island) enabling them to be evaluated

for live coral coverage change detection as well. The validation of this methodology was conducted by isolating the ground truth observations of each location as test data (i.e. the target site). Ground truth data from the three neighboring sites (i.e. training sites) were used to train a classifier the performance of which was evaluated against the ground truth data of the target, test site. This procedure was repeated for each of the four locations for which ground truth data was available. A consolidated classifier was then trained using ground truth observations from all four sites and validated using cross-validation.

The objectives of this study are: (1) to propose a new methodology for training a coral cover classifier, (2) evaluate the effectiveness of this classifier to generalize to new data, and (3) leverage the robust classifier to conduct coral cover change detection. To conduct the study, we selected four Remote Pacific Island reefs for which benthic cover ground truth observations were available and obtained the related Landsat images captured within six months of the observation dates for each location. The reefs analyzed in this research have not been the subject of previous remote sensing studies for widespread change detection analysis.

Section Two of this dissertation begins with a discussion of the materials and methods used in the study. The section begins with a brief review of the data used followed by a per site description of the geomorphology and concludes with an explanation of the preprocessing steps taken to correct the remote sensing images. Following the Materials and Methods Section is an analysis of the per site performance evaluation of the algorithm and a quantitative assessment of the algorithm trained using the combined information from all sites with ground truth observations and applied longitudinally to the two new

locations. This is followed by the temporal change detection analysis for each site. The Discussion Section of this chapter includes an evaluation of the challenges and benefits of the methodology presented in the study as well as a proposal for future work on the subject. The chapter concludes with a discussion of the resulting outcomes and conclusions from this research.

## **4.2 Materials and Methods**

### **4.2.1 Satellite Data Used**

Remote sensing data from Landsat-8 OLI and Landsat 7 ETM with 30m spatial resolution were used for this analysis. Landsat 7 was launched on April 15, 1999 and the subsequent Landsat 8 satellite was launched on February 11, 2013. Both Landsat 7 and Landsat 8 have a 16-day revisit cycle and capture scenes that are approximately 170km north-south by 183km east-west. Landsat 8 consist of nine spectral bands while Landsat 7 consists of seven. Due to their water column penetration properties, the visible bands from each scene were used as inputs to calculate the DII on which the classifier was trained. On the Landsat 8 remote sensing platform, the visible bands used were Band 2 (0.45-0.51), Band 3 (0.53-0.59) and Band 4 (0.64-0.67). The coastal aerosol band was excluded from this analysis despite its water penetrating properties because there is no corresponding Landsat 7 band capturing light in the same wavelength range. The Landsat 7 visible bands used for this analysis were Band 1 (0.45-0.52), Band 2 (0.52-0.60), and Band 3 (0.63-0.69). The NIR band was also leveraged to identify areas of full wavelength absorption for water masking. Landsat 8 images captured within six months of the ground truth observation date were selected for all sites in which ground truth data was available [115] [116]. Table 4-1

provides detailed information regarding the accuracy of the data with respect to the ground truth observation periods and associated image capture dates. Landsat 7 images of the same sites captured at least 12 years before the Landsat 8 images were also obtained for temporal change analysis. All Landsat images obtained were categorized Tier 1 data products radiometrically calibrated and orthorectified using ground control points and digital elevation model (DEM) data to Level-1 Precision and Terrain (L1TP). The data were determined to have well-characterized radiometry within image-to-image tolerances of  $\leq 12\text{m}$  radial root mean square error (RMSE) [117] [118]. L1TP products are the highest quality Level-1 products produced by Landsat and considered suitable for pixel-level time series analysis [95] [96] [97] [98] [99] [100] [101] [102] [103] [104].

Table 4-1: Image Capture Date and Ground Truth Observation Period for each Location.

Site	Ground Truth Observation Period	Landsat 8 Image Capture Date	Difference
Palmyra Atoll	4/15/15 - 4/22/15	5/27/15	<2 months
Kingman Reef	4/24/15 - 4/28/15	5/27/15	<2 months
Baker Island Atoll	2/8/15 - 2/11/15	8/20/14	<6 months
Howland Island	2/3/15 - 2/7/15	1/18/15	<1 months

Two additional sites (Tabuaeran Island and Kiritimati Island) were also selected for longitudinal analysis using the classifier. These two additional sites are large and contain small amounts of human population. Tabuaeran Island has never been the subject of a coral classification and change detection analysis. Specific sites within the Kiritimati Island area of interest were recently studied for changes in reef structural complexity before and after a two year period [119], however, there has never been a large-scale mapping of the location using remote satellite imagery nor has a long-term (>10 years) analysis been

conducted at the site. All remote sensing images with no more than 10% cloud cover across the scene were reviewed. Each scene was visually inspected by a researcher to identify which was most appropriate for each site considering the location of disturbances including cloud cover and cloud shadows in the area of interest within the scene. Table 4-2 displays for each site the Landsat Path-Row, Landsat 8 OLI and Landsat 7 ETM scene capture data, area of interest total size, and total number of ground truth observations.

Table 4-2: Satellite Data Summary.

Site (Figure 4-1)	Path/ Row	Site	Latitude- Longitude	Final State Scene Capture Date	Initial State Scene Capture Date	Area of Interest Size	Number of Ground Truth Observations
1	065/056	Palmyra Atoll	5°52'N 162°6'W	5/27/2015	1/4/2001	22×7-km	82
2	065/056	Kingman Reef	6°23'N 162°25'W	5/27/2015	1/4/2001	19×11-km	57
3	073/060	Baker Island Atoll	0°12'N 176°29'W	8/20/2014	1/15/2002	9×8-km	26
4	074/059	Howland Island	0°48'N 176°37'W	1/18/2015	1/19/2001	5×5-km	30
5	063/057	Tabuaeran Island	3°51'N 159°21'W	2/3/2014	8/31/2000	20×18-km	*
6	061/059	Kiritimati Island	1°40'N 158°30'W	12/22/2014	10/10/2002	47×42-km	*

#### 4.2.2 Ground Truth Data Used for Training and Validation

Ground truth observations were obtained from the National Oceanic and Atmospheric Administration (NOAA) as part of the National Coral Reef Monitoring Program (NCRMP) [115]. This effort identified 39 islands and atolls across the U.S. Pacific territories (including Palmyra Atoll, Kingman Reef, Baker Island Atoll, and Howland Island) as part of a large scale, rapid ecological assessment (REA) of reef environments [120]. Within each reef location, a stratified random sampling of survey sites were selected to capture as wide a domain of the environment as possible based on reef zones (backreef, forereef, protected slope, and lagoon), depth zone (shallow, mid, and deep), and when applicable sector (i.e. section of coastline with broadly similar habitat, exposure, and and/or management status). The sampled site locations were selected prior to each survey mission

using geographic information system (GIS) substrate and strata maps maintained by NOAA Ecosystem Sciences Division (ESD). These substrate and strata maps were created using information from the National Centers for Coastal Ocean Science (NCCOS), reef zones and geomorphologic structures digitized from IKONOS satellite imagery and nautical charts, bathymetric data from the NOAA ESD-affiliated Pacific Islands Benthic Habitat Mapping Center, University of Hawai'i at Manoa, and prior knowledge gained from previous visits to the locations. Observation sites were preloaded into GPS units as waypoints for experienced [120] divers to survey. At each site, a 30m gray polyester transect line was laid across the substratum with markings at 7.5, 15, and 22.5m indicating each diver's observation point and the edges of their survey plots. Divers estimated the percentage cover by type including hard coral and recorded the observations on a data collection sheet [121]. Validation of estimates were performed through photo-quadrat images capturing the benthic habitat at 1m intervals across the 30m transect of the observation site. The images were annotated using CoralNet [122] [123], a deep convolutional neural network computer vision tool, in order to produce a highly consistent evaluation of percent cover, frequency of occurrence, benthic community taxonomic composition and relative generic richness. It has been shown that even experienced divers tend to underestimate hard coral cover (by -3%), and encrusting algae (-2.3%) and overestimate fleshy macroalgae (6.5%) [124]. CoralNet can effectively identify coral cover in images with 89.7% accuracy (Cohen's kappa) [123] and provides an unbiased evaluation of the benthic habitats across locations therefore the information captured via photo-quadrat was relied on for training and validation purposes in this study. In total, 8,825 images characterizing the benthic cover of 308 sites were captured and analyzed across the

four island locations [115]. Of the 308 sites, 162 could be clearly be assigned as the ground truth observation for benthic cover within a pixel of the 30m resolution Landsat data.

### **4.2.3 Sites**

The six sites used for this study were chosen based on the locations of known coral reefs in the Pacific Ocean and relative proximity to each other. Palmyra Atoll and Kingman Reef are both located in the Northern Line Island chain with Kingman Reef only 69km northwest of Palmyra Atoll. Both locations can be observed within the same Landsat scene. Palmyra Atoll is characterized by elongated terraces that extend 3 to 5km to the east and west. The depths of these terraces range from 7 to 25m [73]. The northern-most reef in the Line Island chain is Kingman Reef. This triangular atoll extends 9km north to south and 18km east to west [79]. A deeper reef (>20 m) extends along the western terrace and the southern and northern sections consist of shallow reefs (<2 m). The lagoon is mostly deep (>30m) and contains large patch reefs with a diameter of 50-200 m, 2-10m under the surface. Images of both Palmyra Atoll and Kingman Reef can be found in Figure 4-1.

Howland Island and Baker Island Atoll are both low reef, outlier islands of the Phoenix Island Archipelago. Baker Island Atoll is characterized by steep reef slopes to the west, north, and south which drop to substantial depths [81]. To the east, the island consists of spur and groove geomorphic features and oligotrophic waters off reef terraces [82]. Howland Island is 66km northwest of Baker Island Atoll. Although closer in proximity to each other than Kingman Reef and Palmyra Atoll, these two locations reside within the extent of two separate Landsat scenes. Due to each islands' proximity to the equator, they are subjected to the westward-flowing Southern Equatorial Current at the surface while the

strong eastward-flowing Equatorial Undercurrent flows below. The result is a nutrient rich topographic upwelling on the western sides of each island [83]. The waters around Howland Island descend to great depths just off the coast with a narrow, shallow fringing reef surrounding the island. The eastern side of the island is abrupt and covered with coral rubble due to erosion caused by waves and strong trade winds. The western portion of the island is low and sandy [84]. Images of these two islands can be found in Figure 4-1.

Tabuaeran Island (also called Fanning Island) and Kiritimati Island (also called Christmas Island) are both raised coral atolls. Images of these two islands can be found in Figure 4-1. The land area of Tabuaeran Island covers less than 34km<sup>2</sup> and has a population of approximately 2,000 [125]. The lagoon of Tabuaeran Island is shallow (<15 m) and spans an area of 110. There is a single deep pass to the west through which 95% of the lagoon/ocean water exchange occurs [126] as well as two shallow (<1 m) passes to the southeast and north [125]. This geomorphology results in clear water in the vicinity of the deep pass and turbid water at the north and south basins of the lagoon. As a result, much of the coral within the lagoon is located in deeper pools near the primary inlet [127] and are comprised of coral fauna that is unique from the surrounding reef. The eastern portion of the island is characterized by a broad shelf that extends seaward for several hundred meters [127]. Kiritimati Island's land mass covers 363km<sup>2</sup> making it the largest coral atoll in the world [128]. The lagoon is large covering 328km<sup>2</sup>. The eastern end of the lagoon is comprised of several hundred smaller landlocked lagoons [129]. The island population has been increasing and in 2015 it was reported that 6,447 inhabitants were living on the island [130]. A narrow reef flat runs around most of the island and sand plains occur to the northwest and southeast. Both Tabuaeran Island and Kiritimati Island encounter southeast

trade winds for much of the year resulting in a relatively arid environment and highly variable precipitation [127] [131].



Figure 4-1: The relative location of each of the six sites (1) Palmyra Atoll, (2) Kingman Reef, (3) Baker Island Atoll, (4) Howland Island, (5) Tabuaeran Island, and (6) Kiritimati Island.

### 4.3 Methodology

Each location required preprocessing including application of cloud and land masks, atmospheric correction, and water column correction. First, both the Landsat 7 and Landsat 8 image for each location was preprocessed for analysis. This included applying a water mask, cloud/fill pixel mask, and sun glint correction. The dark-pixel subtraction method was applied to correct for atmospheric contributions [37] [50] [51]. Finally, the

preprocessing steps concluded with a calculation of the per pixel DII for water column correction. Each of the four initial sites for which ground truth observations were available were subjected to these preprocessing steps. The resulting DII values were used as the inputs to train our machine learning classifiers the output of which was the probability that a pixel belonged to the coral class.

$$P(\text{Pixel}_{\text{coral}} | \text{DII}_{\text{blue,green}}, \text{DII}_{\text{blue,red}}, \text{DII}_{\text{green,red}})$$

The processing steps applied to Palmyra Atoll, Kingman Reef, Baker Island Atoll, and Howland Island are detailed in Figure 4-2. The processing steps applied to the additional two locations, Tabuaeran Island and Kiritimati Island are detailed in Figure 4-3.

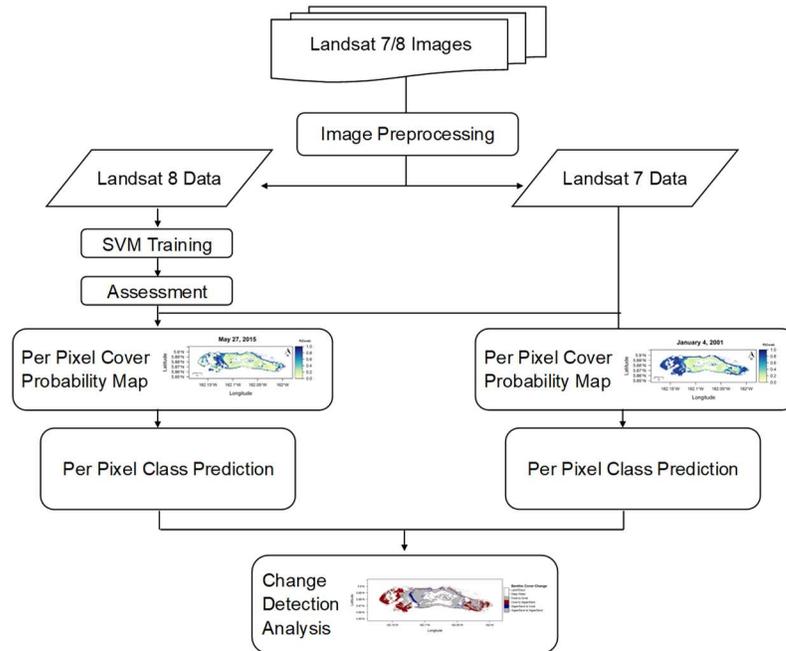


Figure 4-2: SVM classifier training and change analysis process flow for Palmyra Atoll, Kingman Reef, Baker Island Atoll, and Howland Island.

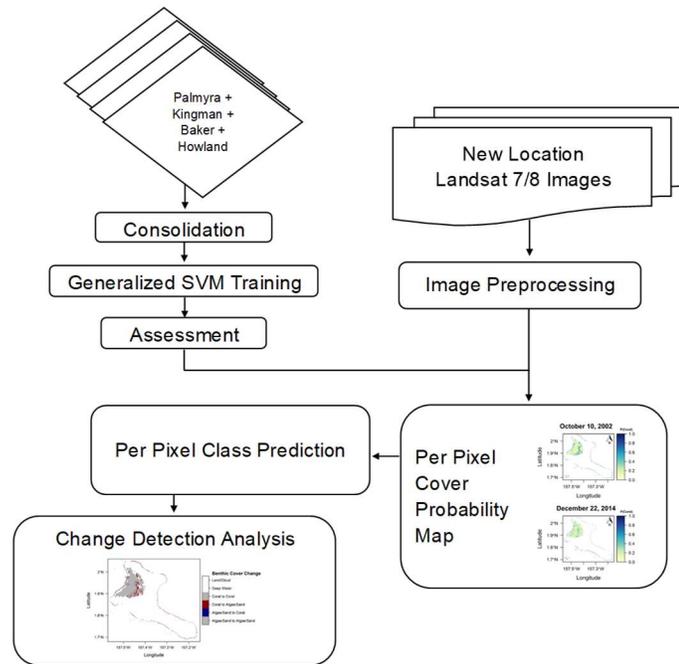


Figure 4-3: Robust SVM Classifier training and change analysis process flow for Tabuaeran Island and Kiritimati Island.

### 4.3.1 Cloud Mask

The initial preprocessing step included masking pixels obscured by clouds and their shadows. The Landsat images were selected from scenes with less than 10% cloud cover to minimize cloud and obstructions particularly in the coral reef areas of interest within each scene. However, given the locations of our sites and recency requirement to the time period of observation data, some amount of cloud cover was unavoidable. As a result, a cloud and pixel quality mask applied to both the initial state image and final state image of each scene. This was performed by leveraging the Landsat 7 and Landsat 8 BQA band [51] [53]. Furthermore, when conducting the change detection analysis step, both the initial state image mask and final state image mask were applied to both images of each scene. This important step was performed in order to isolate the change in coral population apart from the confounding effects of differences in quality between the two images.

### **4.3.2 Land Mask**

The next preprocessing step was to prepare a mask for land pixels. The Landsat NIR band captures light between 0.851 and 0.0879  $\mu\text{m}$ . Light in the wavelength range does not penetrate water making it an ideal candidate for delineating water from land [54]. The mask was developed by applying a threshold to the NIR band pixel values. Plots were created for each area of interest and inspected to identify the most appropriate cutoff for the mask.

### **4.3.3 Atmospheric Correction and Water Column Correction**

Atmospheric and water surface properties account for 90% of at sensor reflectance in the visual bands [56]. Therefore, atmospheric correction was performed to remove this interference via the dark pixel subtraction method [57]. An area of deep-water (>50 m) in which visible wavelengths have fully attenuated was selected from within each scene [58]. Under the assumption that the atmospheric and water surface conditions are uniform across the scene, the mean deep-water radiance at sensor was leveraged to correct for the effect of atmospheric and surface reflectance interferences [1] [58] [59] [60]. Due to the full attenuation of visible light, signal received from the deep-water area of interest is comprised of atmospheric radiance and surface reflectance. This allows the impact of these factors to be isolated for correction [55]. Two standard deviations are subtracted from mean radiance at sensor to account for possible sensor noise [62].

The intensity of visible light decreases exponentially as it penetrates water. The attenuation rate is wavelength-dependent and severely impacts the study of ocean habitats

using remote sensing data [61]. Remote sensing imagery that contains several water-penetrating (visible) bands can be leveraged for water column correction [51]. Shorter-wavelength light will attenuate less rapidly than longer-wavelength light. As a result, at sensor the spectral radiances are comprised of both depth and the composition of the subsurface strata. This represents a confounding influence that can create a material shift in the subsurface reflectance. Therefore, it is advantageous to remove the influence of variable depth in the study of benthic features [1]. In this study, the well-known and broadly used water column correction method developed by Lyzenga [43] [63] [64] [65] [66] was applied to each image.

#### **4.3.4 SVM Site Application, Validation, and Change Analysis**

In the analysis of Palmyra Island Atoll, Kingman Reef, Howland Island Atoll, and Baker Island, the preprocessing steps were followed by training of a Support Vector Machine (SVM) classifier [132] against each of the Landsat 8 images and evaluation of results using ground truth data. Beginning with these four locations for which ground truth observation data was available, we trained an SVM classifier on the three Landsat 8 DII ( $DII_{\text{blue,green}}$ ,  $DII_{\text{blue,red}}$ ,  $DII_{\text{green,red}}$ ) values to predict the observed pixel class. Tuning of the algorithm was performed using cross-validation. A radial basis function kernel outperformed other kernel methods and the optimal cost (regularization) and gamma (influence) parameters were determined for each site based on optimal accuracy and generalization criteria. Once the optimal hyperparameters for the SVM algorithm were obtained, the algorithm was applied to determine the posterior probability of each pixel within the Landsat 8 area of interest for each of the four locations. A map of the posterior

probability that each pixel contained coral was produced and the final class for each pixel was determined by thresholding the posterior probability at 50%. The SVM classifier was then applied to the applicable, preprocessed Landsat 7 scene for each location. Posterior probability and predicted class maps were then produced for each of the Landsat 7 scenes as well. A per pixel comparison of the predicted class was made for all pixels in each area of interest of each scene in order to determine the change in coral cover for the given time periods.

The analysis of the four initial sites was followed by analysis of two more remote locations (Tabuaeran Island and Kiritimati Island). These unique locations were selected on the basis that they have never been the subjects of coral cover change detection analysis that maps the entire ecosystem using remote satellite imagery [119]. A SVM classifier was trained using the combined inputs (truth labels and DII values calculated from the corresponding Landsat 8 scene) from all four sites therefore enabling the resulting algorithm to more adeptly generalize beyond site-specific biases of previous in-situ analysis [133]. Validation of this approach was performed in two different ways. First, the accuracy of the model trained on the consolidated data was assessed through cross-validation. While this methodology enabled an overview of the model performance it does not fully account for the location-based bias. Specifically, training the SVM algorithm on a sampling from all sites will incorporate some amount of site-based bias into the resulting model. While this evaluation method represents an improvement to historical in-situ analysis, it is not a full analysis of the generalization properties of the classifier. Therefore, an additional evaluation procedure was deployed to test the ability of the classifier to generalize to each location independently. This second method of evaluation was

performed by iteratively selecting the ground truth observations from one of the four island locations to withhold as test cases. A model was then trained on the ground truth data from the remaining three sites. Finally, the resulting trained classifier was applied to predict the coral cover classification of the ground truth observations from the remaining location that was withheld from training as test data. The result of this CPCV method is a more accurate evaluation of the generalization properties of the machine learning algorithm when applied to new data. In this way, the CPCV process controls for much of the location-based bias of in-situ analysis and therefore is a more accurate assessment of the SVM classifier performance. Table 4-5 displays the results of the CPCV test procedure.

Using the consolidated ground truth observations from all the available sites enabled a more diverse set of inputs to train the machine learning algorithm. As a result, the classifier produced was capable of generalizing more effectively to additional sites [133]. The algorithm was tuned using cross-validation and the optimal model selected for application to Tabuaeran Island and Kiritimati Island areas of interest. Landsat 8 scenes from 2014 and Landsat 7 scenes from 2000 and 2002 were obtained and preprocessed. The trained classifier was then applied to these sites longitudinally using the model trained across the consolidated truth data from Palmyra Atoll, Kingman Reef, Baker Island Atoll, and Howland Island. The per pixel posterior probability map was created and the final classification of each pixel determined using a 50% threshold. Change detection analysis was then conducted by comparing the initial and final classification of each pixel. Implementation and analysis were performed using the open source R programming language and environment for statistical computing [52].

## **4.4 Results**

### **4.4.1 Classification Accuracy by Site**

The preprocessed DII values were used to train SVM models [86] [134] for each of Palmyra Island, Kingman Island, Baker Island Atoll, and Howland Island. In this manner, the DII values represent the features of the SVM model and the corresponding observed benthic cover type the response (e.g. live coral or not). The posterior probability for each pixel containing live coral cover was computed and a mapping of these posterior probabilities was developed for each location. A threshold was then applied to the probabilities in order to identify the most likely class and mapped for reference. Contextual validation was performed by comparing the resulting maps of predicted probabilities against the known geomorphology of each site. Finally, the trained SVM algorithm was applied to the initial state Landsat 7 image for each location and the same posterior probability and predicted class maps were generated. Table 4-3 summarizes the results of the accuracy assessment for each location and Table 4-4 provides the resulting confusion matrices for each location.

The Palmyra Atoll site contained a total of 82 unique ground truth observations. 66 of these observations were in unobscured pixels for which valid DII data could be precisely identified by pixel location matching the recorded GPS coordinates. The SVM model trained on this data correctly classified 87.9% of the ground truth observations within the area of interest. This model was further evaluated using precision and recall as common and well-known statistical measures for type I and type II errors. Precision measures the impact of type I error and therefore is an evaluation of the classifier's ability to be exact.

Recall measures the impact of type II error and therefore is an evaluation of how completely the classifier retrieves information. The Palmyra Atoll model obtained a precision of 0.837 and recall of 1.000. Another common measure of the effectiveness of a classifier is the harmonic mean of precision and recall, known as the F-measure. The F-measure of the SVM classifier trained and tuned against the Palmyra Atoll site data was 0.911. Specificity is a measure of the classifier's capacity for differentiating between true negative observation data. As applied to the Palmyra Atoll site, the SVM classifier obtained a specificity of 0.680.

The Kingman Reef area of interest contained 57 ground truth observations for which 42 valid pixel values were unobscured. The classifier trained on this data yielded an overall accuracy of 85.7%. The resulting precision was measured at 0.824 and recall was measured at 1.000. This was an indicated that the model was committing type I errors but no type II errors. The model obtained an overall F-measure of 0.903 and specificity of 0.571.

There were 26 bottom type observations for the Baker Island Atoll site all of which had corresponding pixel values that were valid. The SVM model trained on this data obtained an accuracy of 69.2%. This lower accuracy is likely due to the inherent limitations of the observation data and mixing of bottom types within the pixels due to unique coral fauna and geomorphology. The model yielded a precision and recall of 0.706 and 0.800, respectively. The F-measure was calculated at 0.750 while specificity was observed at 0.546.

The Howland Atoll site area of interest contained 30 observations for which 28 valid pixels were obtained. The SVM model trained on this data yielded an accuracy of 82.1%. The resulting precision and recall were measured to be 0.813 and 0.867, respectively which indicates that the model is robust to both type I and type II errors. Furthermore, the classifier yielded an F-measure of 0.839 and specificity of 0.769.

Table 4-3: SVM Classifier Performance Assessment by Site.

	Palmyra Atoll	Kingman Reef	Baker Island Atoll	Howland Island
Accuracy	87.9%	85.7%	69.2%	82.1%
Precision	0.837	0.824	0.706	0.813
Recall	1.000	1.000	0.800	0.867
Specificity	0.680	0.571	0.546	0.769
F-measure	0.911	0.903	0.750	0.839

Table 4-4: Confusion Matrices by Site and Consolidated Inputs.

		Ground Truth Labels	
		Coral	Not Coral
<b>Predicted Class</b>		<b>Palmyra Atoll</b>	
	<b>Coral</b>	41	8
	<b>Not Coral</b>	0	17
		<b>Kingman Reef</b>	
	<b>Coral</b>	28	6
	<b>Not Coral</b>	0	8
		<b>Baker Island Atoll</b>	
	<b>Coral</b>	12	5
	<b>Not Coral</b>	3	6
		<b>Howland Island</b>	
	<b>Coral</b>	13	3
	<b>Not Coral</b>	2	10
		<b>Consolidated Sites</b>	
	<b>Coral</b>	91	32
<b>Not Coral</b>	8	31	

#### **4.4.2 Consolidated Model Robust to Site-Specific Bias**

Once the classifier for each site was trained and deployed, a combined classifier was built using the consolidated information of all valid observation data from the four sites (Palmyra Atoll, Kingman Reef, Baker Island Atoll, and Howland Island). Combining the inputs from all sites enabled the development of an algorithm more robust to site-specific bias. The limitation or elimination of this bias allowed the resulting model to more effectively generalize to other sites for which the training sites were representative. Without the benefit of site-specific bias, the resulting accuracy is likely to suffer because the classifier no longer had the benefit of accounting for local conditions and geomorphology in training. However, the resulting algorithm did gain the benefit of being robust to these site-specific conditions and therefore could be applied to additional sites. When the SVM algorithm was applied to the consolidated observations of all sites it yielded an accuracy of 75.3%. The resulting precision was 0.740 and recall was 0.919. Specificity was measured to be 0.492 and the F-measure was 0.820. These figures are detailed in Table 4-5.

As expected, the classifier did not obtain as high an accuracy when assessed using the CPCV method. As discussed in the Materials and Methods Section, this was a result of the further isolation and reduction of site-specific bias during the training of the algorithm. The CPCV method resulted in an accuracy of 78.8% when tested against the Palmyra Atoll ground truth data. This corresponds to a precision of 0.776, recall of 0.927, and specificity of 0.560. The F-measure was 0.844. The Kingman Reef site evaluated using the CPCV process obtained an accuracy of 81.0%, precision of 0.776, recall of 0.927, and specificity

of 0.560. The corresponding F-measure was calculated at 0.867. The Baker Island Atoll and Howland Island results from the CPCV method show a decrease in accuracy to 65.4% and 67.9%, respectively. This decrease is related to the distance between the majority of the ground truth observations (i.e. Palmyra Atoll and Kingman Reef). This geographic distance resulted in ground truth observations that do not represent the test site as well due to differences in marine species, environmental conditions, and geomorphology. Baker Island Atoll and Howland Island were disproportionately impacted by this due to disproportionate representation in the ground truth observations. Some stratified random sampling of the locations was performed during CPCV assessment to control for this disproportionate representation, however, data size constraints limited the extent to which sampling methods could be applied. The Baker Island Atoll assessment yielded a precision of 0.750, recall of 0.600, and specificity of 0.769. The Howland Island test produced a precision of 0.750, recall of 0.600, and specificity of 0.769. The CPCV procedure obtained an F-measure of 0.769 and 0.667 for Baker Island Atoll and Howland Island, respectively. The assessment of the SVM classifier trained on the consolidated ground truth observations from all sites yielded an accuracy of 75.3%. The corresponding precision was 0.740, recall was 0.919, and specificity was 0.492. The F-measure was calculated as 0.820. These figures are summarized in Table 4-5. Figure 4-4 contains the ROC curve for the classifier resulting from the model developed using the consolidated site data. This is a common visualization is used to evaluate the overall performance of a classifier. It plots the model's recall (true positive rate) as a function of  $1 - \textit{Specificity}$  (false positive rate) [135]. An additional classifier diagnostic measure related to the ROC curve is the AUC. The AUC

resulting from the combined classifier was 0.778. Table 4-5 is a summary of the results from the CPCV procedure.

Table 4-5: Controlled Parameter Cross-Validation (CPCV) procedure results by site.

	<b>Palmyra Atoll</b>	<b>Kingman Reef</b>	<b>Baker Island Atoll</b>	<b>Howland Island</b>	<b>Consolidated Sites</b>
Accuracy	78.8%	81.0%	65.4%	67.9%	75.3%
Precision	0.776	0.813	0.625	0.750	0.740
Recall	0.927	0.929	1.000	0.600	0.919
Specificity	0.560	0.571	0.182	0.769	0.492
F-measure	0.844	0.867	0.769	0.667	0.820

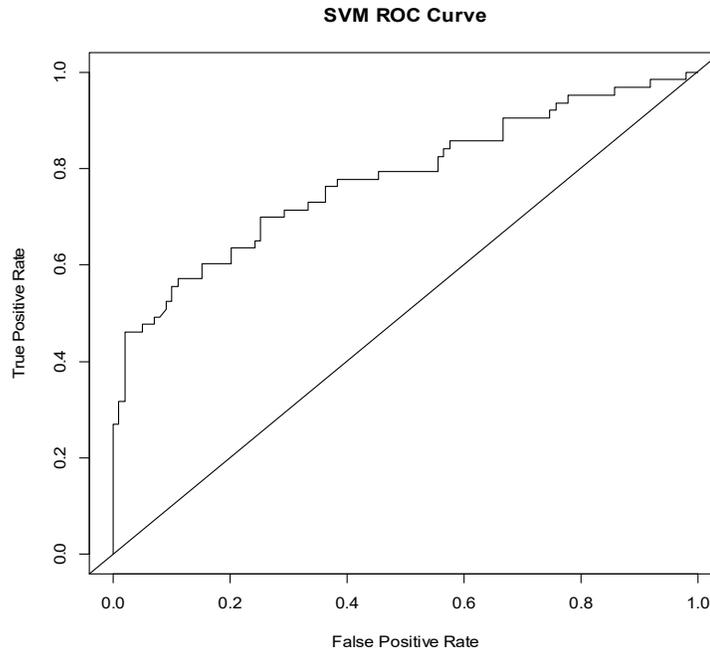


Figure 4-4: Performance evaluation of the combined classifier using ROC Curve and AUC (0.778).

## **4.5 Change Detection Analysis**

### **4.5.1 Palmyra Atoll**

The baseline image of the Palmyra Atoll site used for the change detection analysis in this study was captured in 2001 and contained 24,092 pixels identified as coral which corresponds to approximately 21.68km<sup>2</sup>. In 2015 the same location contained 16,410 coral pixels or approximately 14.77km<sup>2</sup> of coral cover. This represents a 31.9% net decrease in coral coverage over the 14-year period. 9,850 (8.87km<sup>2</sup>) of the coral pixels in 2001 became algae by 2015. Furthermore, in 2001 52.8% of the area surveyed was identified as coral compared to 35.9% in 2015. A minor amount of area (4.8% of pixels) that was algae in the earlier image was determined to be coral in the second image. See figure 4-5 for the Palmyra Atoll posterior probability map for 2001 (top) and 2015 (bottom). Figure 4-6 maps the coral cover change when the 2001 initial state classification map is compared to the 2015 final state classification map. Figure 4-7 maps the cover type classification change between the two images. Table 4-6 depicts the per class results of the temporal change detection analysis.

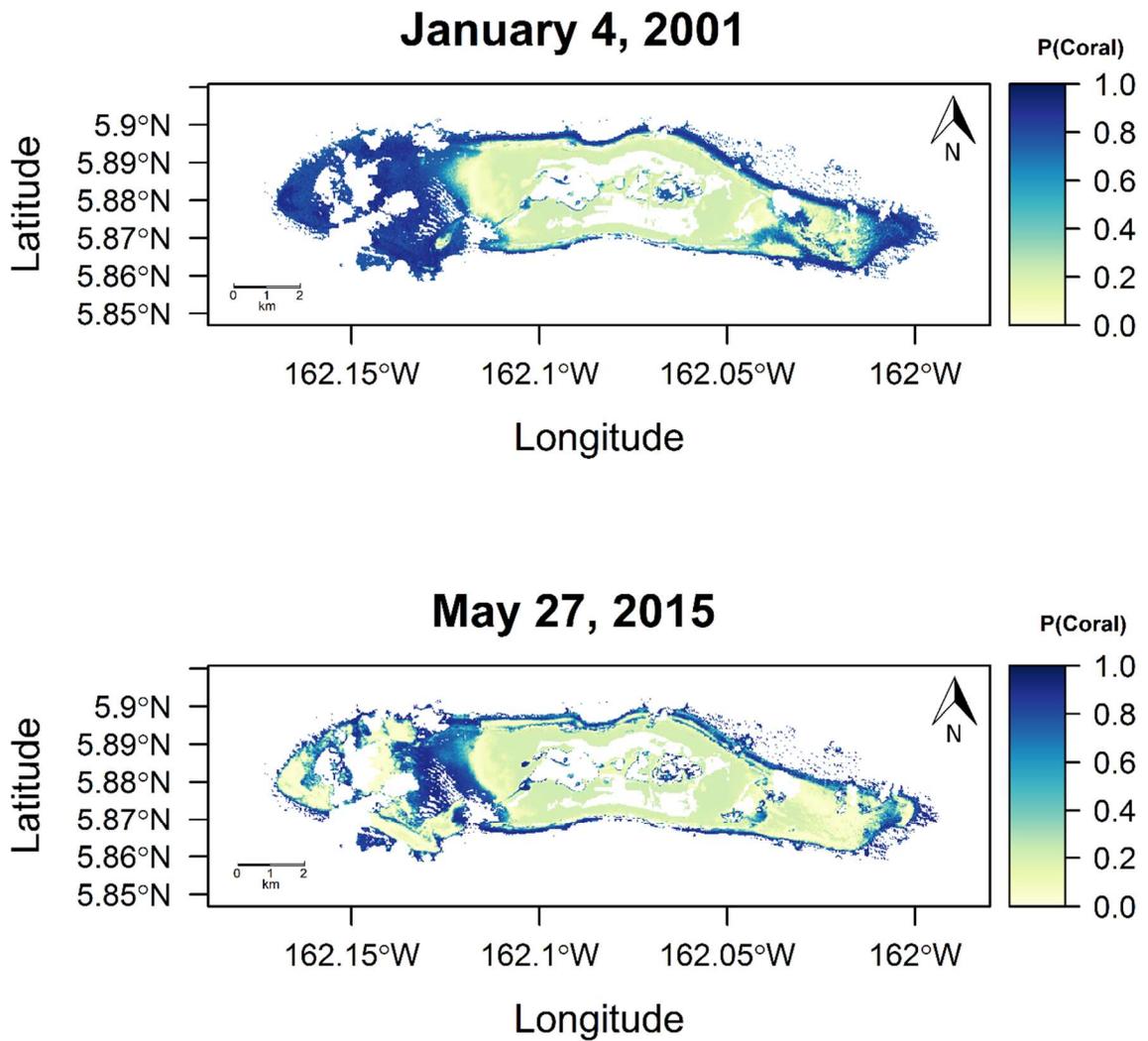


Figure 4-5: Posterior probability map for the Palmyra Atoll area of interest (top, 2001 and bottom, 2015).

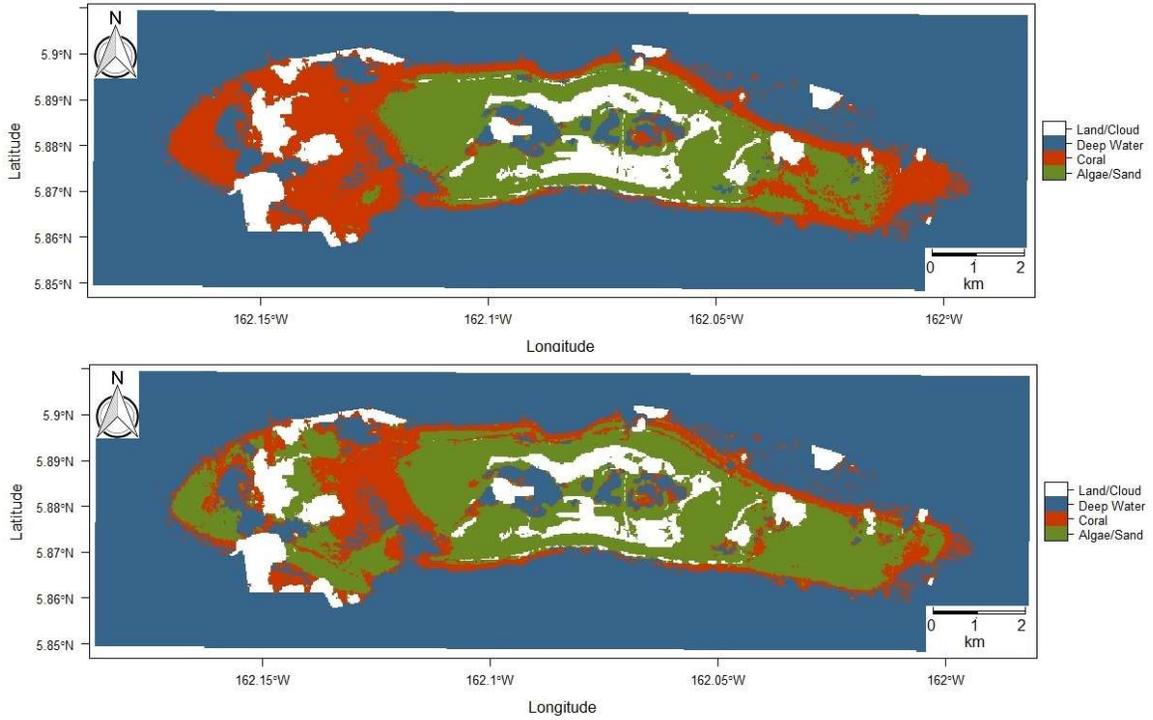


Figure 4-6: Predicted class map for the Palmyra Atoll area of interest (top, 2001 and bottom, 2015).

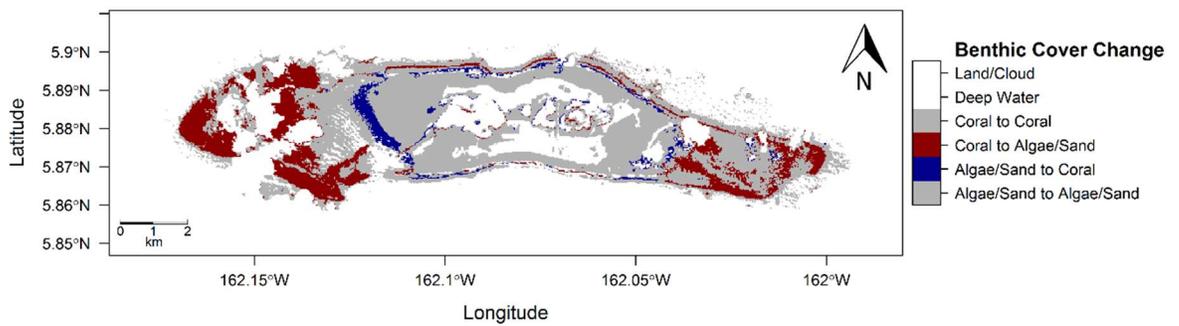


Figure 4-7: Difference in predicted class membership map for the Palmyra Atoll area of interest for 2001 initial state compared to 2015 final state.

## 4.5.2 Kingman Reef

The Kingman Reef site revealed some of the smallest declines in coral of the locations surveyed. The initial image of this location was captured in 2002. 33,294 coral pixels were identified as containing live coral cover in the initial class image which corresponds to approximately 29.96km<sup>2</sup>. The final status image was captured in 2015 and when analyzed revealed 24,860 pixels with live coral cover (22.37km<sup>2</sup>). This represents a 25.3% net decrease during the 14-year period between the image capture dates. 85.1% of valid, shallow area the initial state image contained coral while 63.5% of the final image contained coral. Figure 4-8 represents the mapping of Kingman Reef posterior probabilities for each pixel belonging to the coral class both for 2001 (top) and 2015 (bottom). Figure 4-9 maps the coral cover change when the 2001 initial state classification map is compared to the 2015 final state classification map. Figure 4-10 maps the cover type classification change between the two images. Table 4-6 is a summary of the results of the change detection analysis between the initial state image and final state image.

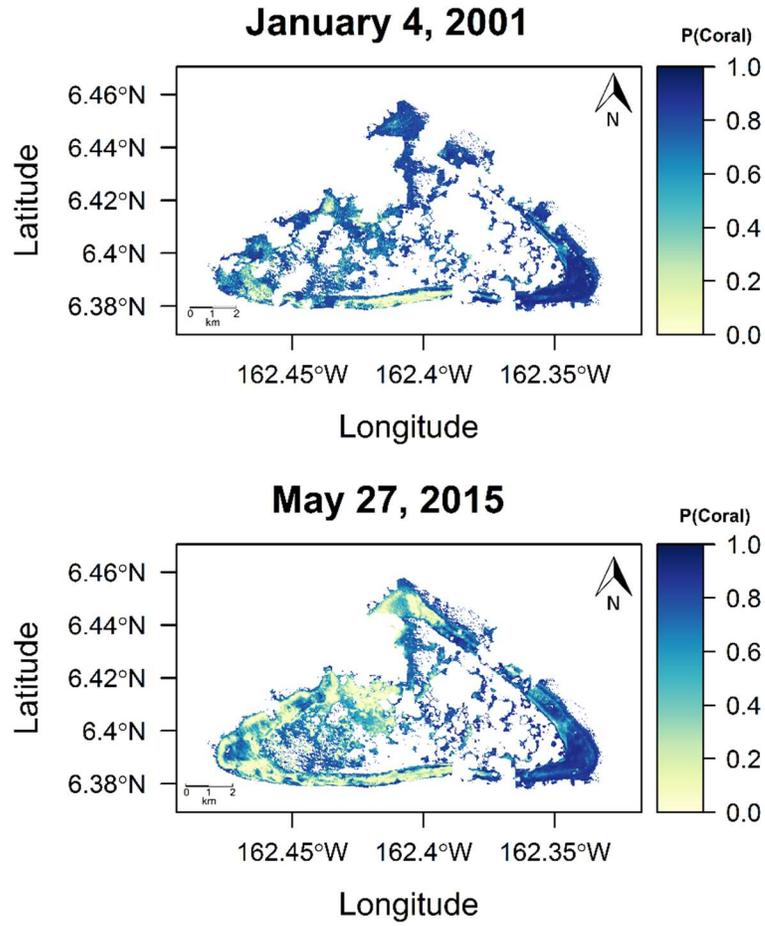


Figure 4-8: Posterior probability map for the Kingman Reef area of interest (top, 2001 and bottom, 2015).

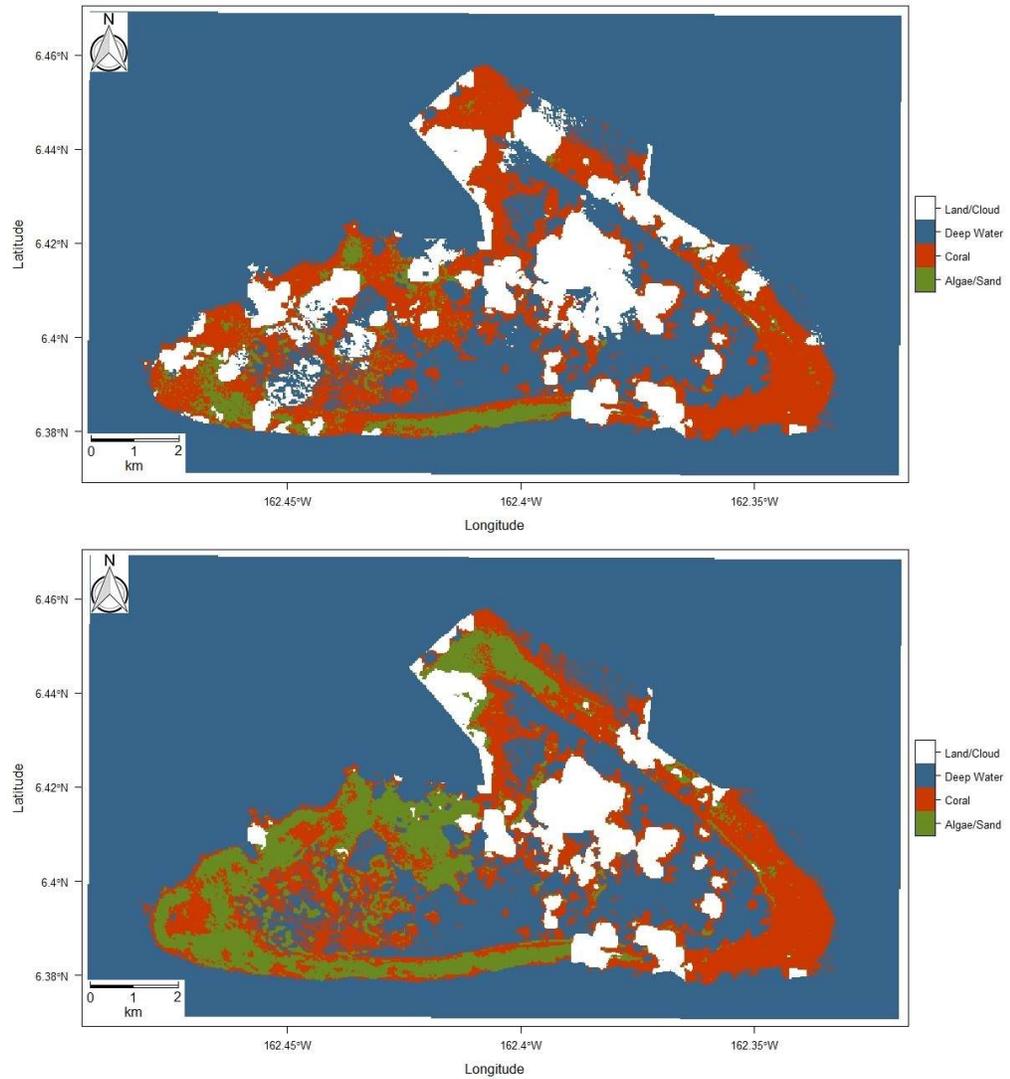


Figure 4-9: Predicted class probability map for the Kingman Reef area of interest (top, 2001 and bottom, 2015).

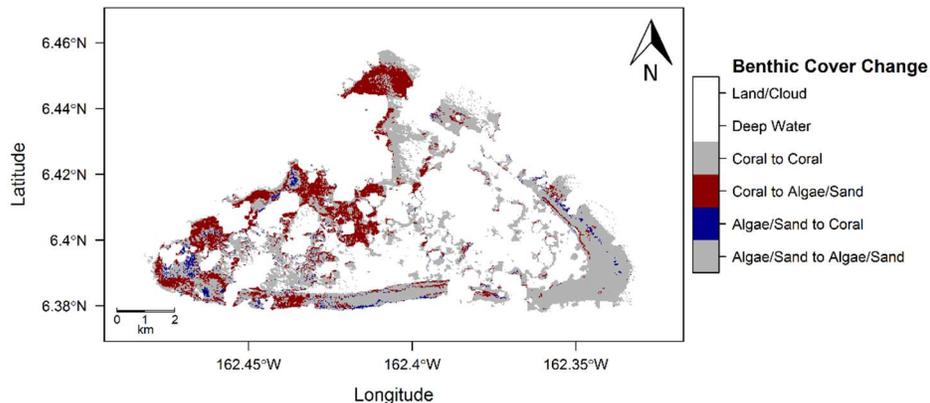


Figure 4-10: Difference in predicted class membership map for the Kingman Reef area of interest for 2001 initial state compared to 2015 final state.

### 4.5.3 Baker Island Atoll

The second largest percentage decrease in coral between the initial and final state images was observed at the Baker Island Atoll site. The image of the scene used to identify the initial state of live coral cover at the site was obtained in 2002 and the final image was obtained in 2014 representing a 13-year observation period. Initially 465 pixels or 0.42km<sup>2</sup> were classified as coral compared to 280 or 0.25km<sup>2</sup> in the final image. This corresponds to a 39.8% net decrease during the observed period. Furthermore, 22.0% of the shallow, valid pixels in the initial image were classified as coral compared to 13.3% in the final image. Figure 4-11 represents the mapping of posterior probabilities for each pixel belonging to the coral class for Baker Island Atoll beginning with 2002 (top) and the final class in 2014 (bottom). Figure 4-12 maps the coral cover change when the 2002 initial state classification map is compared to the 2014 final state classification map. Figure 4-13 maps the cover type classification change between the two images. Table 4-6 contains a summary of the change detection analysis for the Baker Island Atoll site.

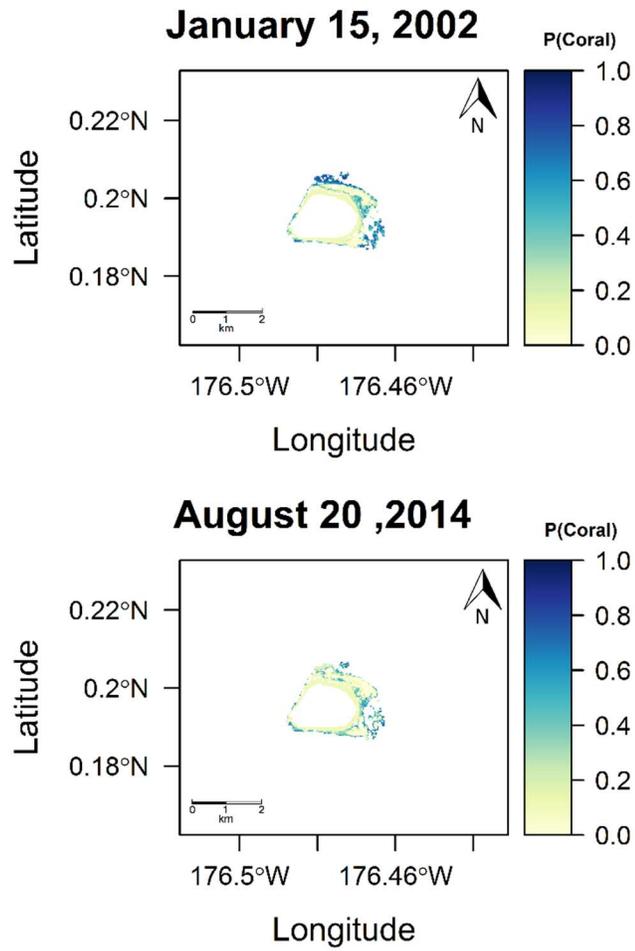


Figure 4-11: Posterior probability map for the Baker Island Atoll area of interest (top, 2002 and bottom, 2014).

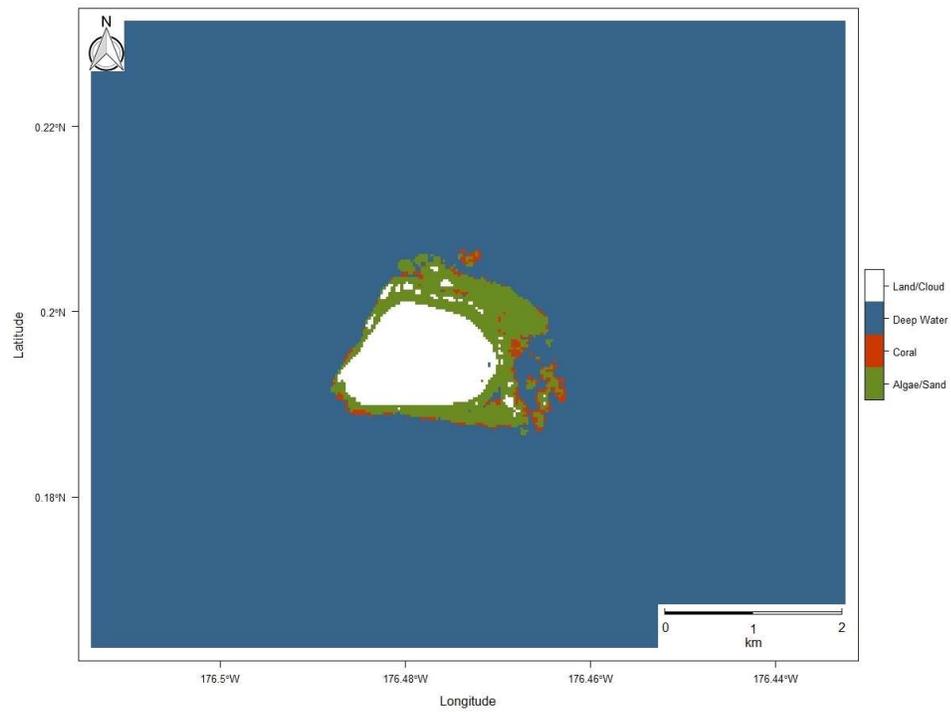
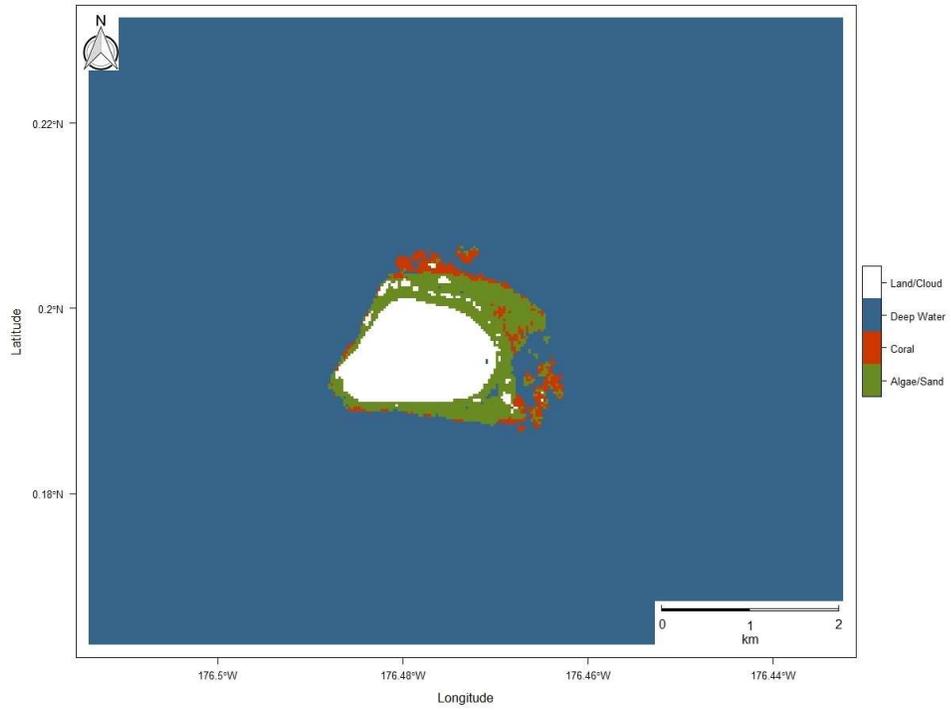


Figure 4-12: Predicted class map for the Baker Island Atoll area of interest (top, 2002 and bottom, 2014).

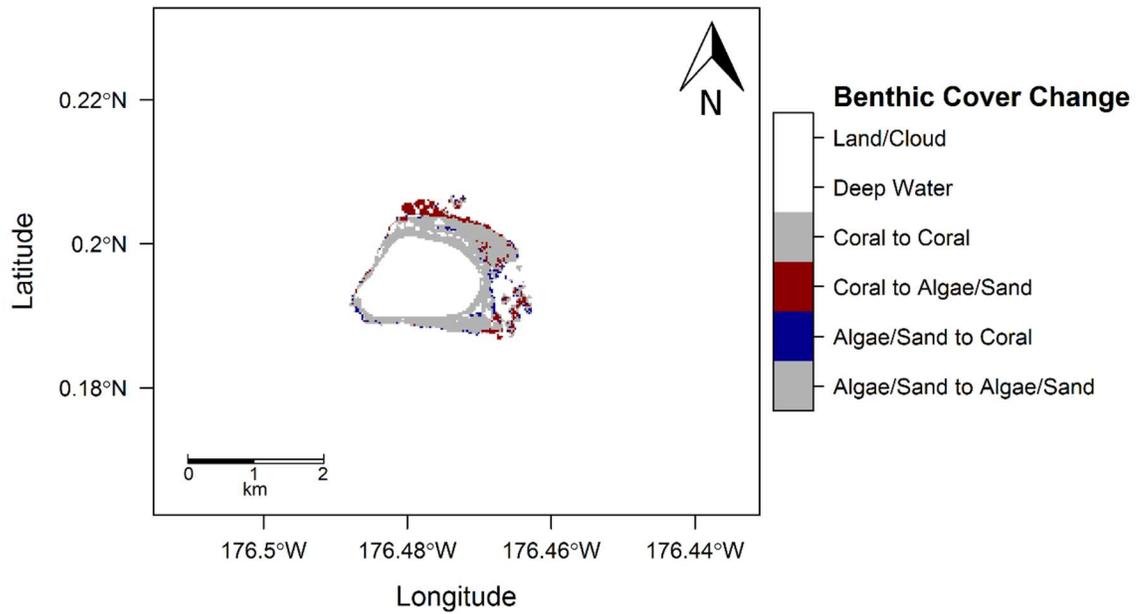


Figure 4-13: Difference in predicted class membership map for the Palmyra Atoll area of interest for 2002 initial state compared to 2014 final state.

#### 4.5.4 Howland Island

Howland Island was the smallest area of interest analyzed. During the 14-years between the initial state image and final state image this area incurred the smallest decrease in coral of all sites analyzed. The initial scene was captured by Landsat 7 in 2001 and was found to contain 252 coral pixels (0.23km<sup>2</sup>). The final state image was captured by Landsat 8 in 2015 and contained 188 coral pixels (0.17km<sup>2</sup>). This represents a 25.4% decrease over the observed time period. 21.5% of the initial image was classified as coral cover while 16.0% of the final image was classified as coral. Figure 4-14 displays the posterior probabilities for each pixel belonging to the coral class for 2001 (top) and 2015 (bottom). Figure 4-15 maps the coral cover change when the 2001 initial state classification map is compared to the 2015 final state classification map. Figure 4-16 maps the cover type

classification change between the two images. The results of the change detection analysis are detailed in Table 4-6.

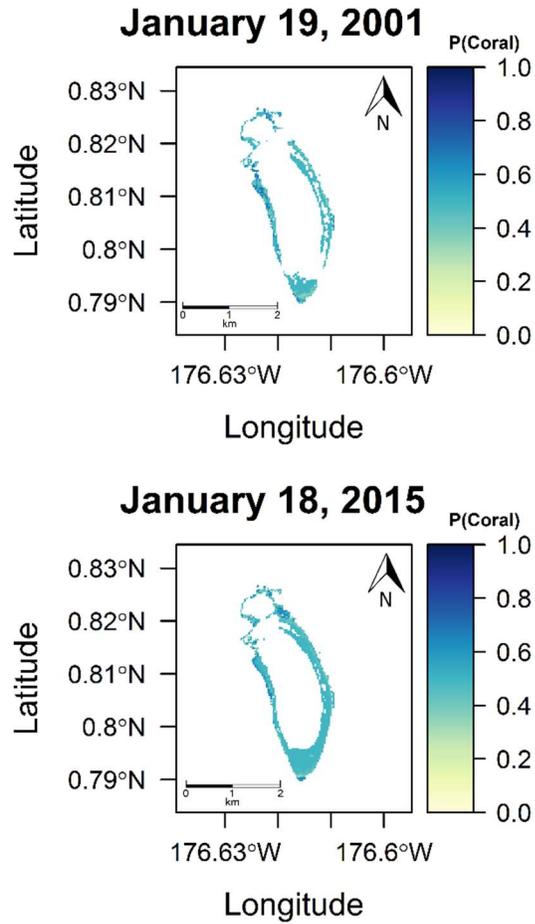


Figure 4-14: Posterior probability map for the Howland Island area of interest (top, 2001 and bottom, 2015).

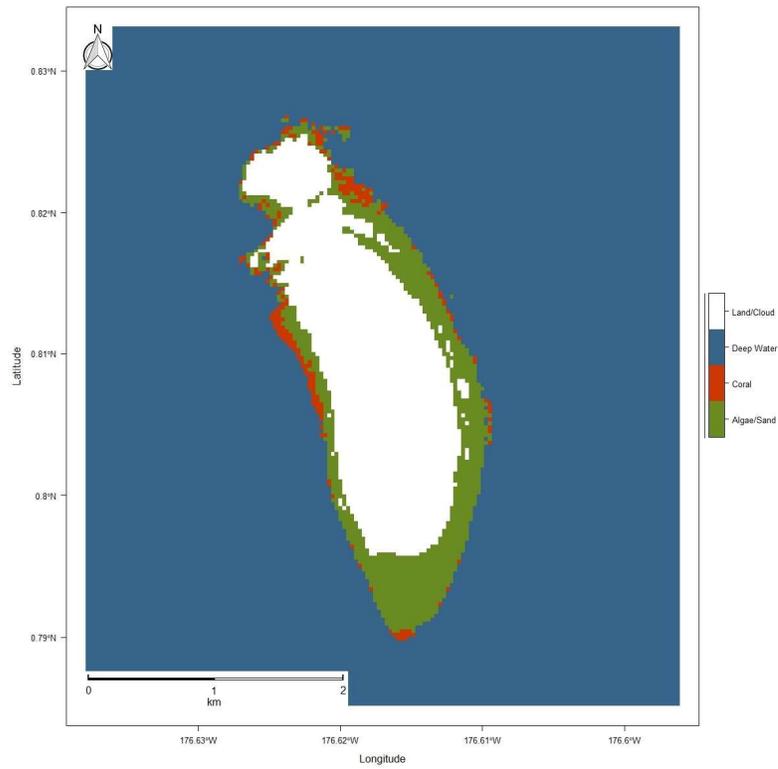
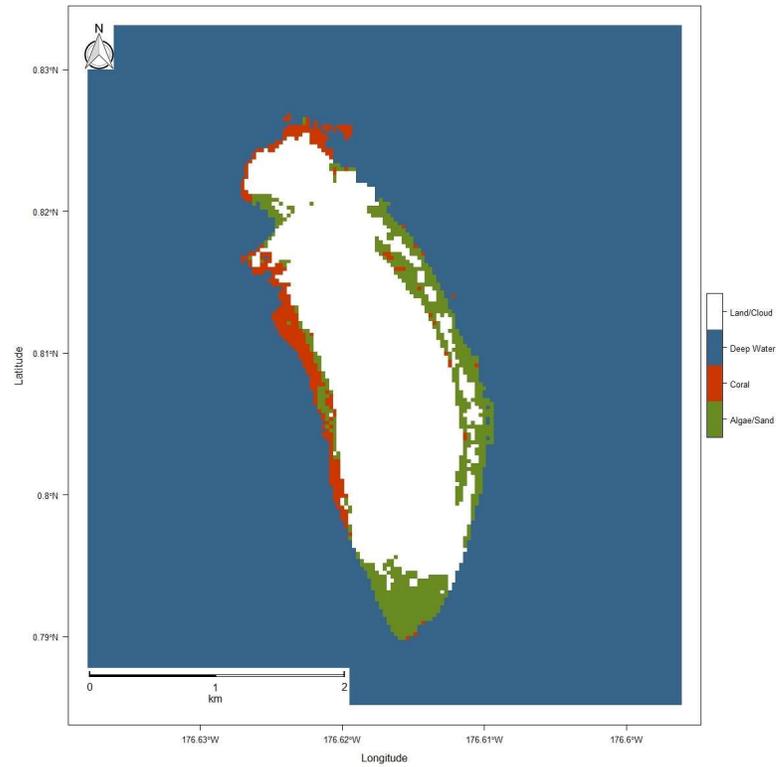


Figure 4-15: Predicted class map for the Howland Island area of interest (top, 2001 and bottom, 2015).

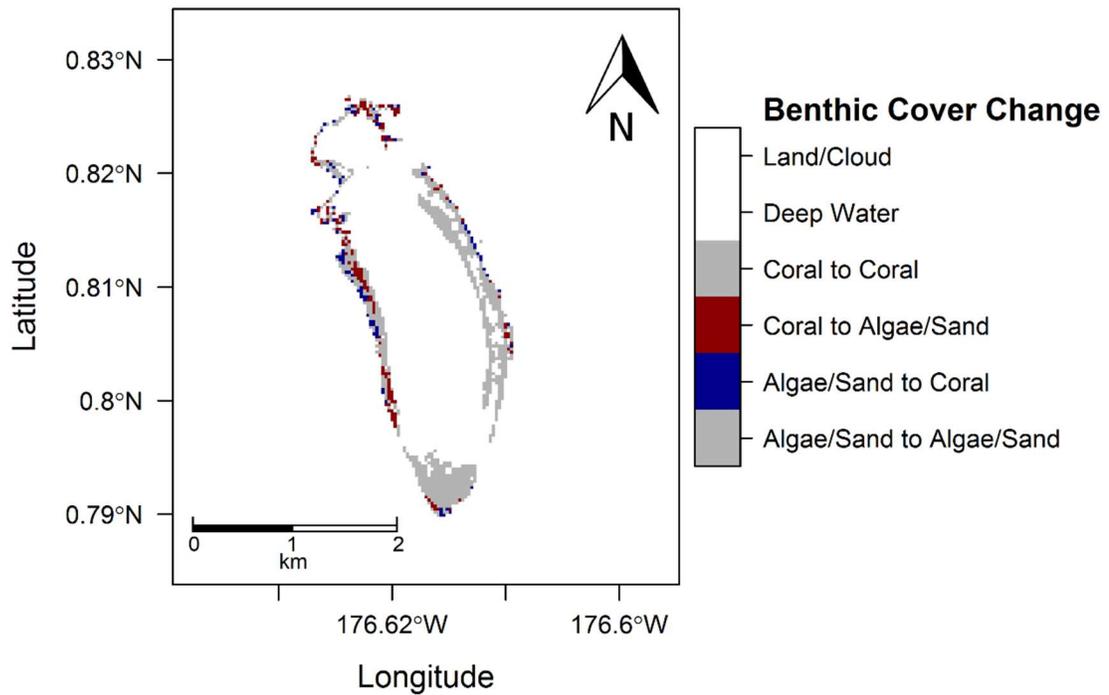


Figure 4-16: Difference in predicted class membership map for the Howland Island area of interest for 2001 initial state compared to 2015 final state.

#### 4.5.5 Tabuaeran Island

Tabuaeran Island represented the largest site evaluated for coral change detection. The initial image of this site was captured in 2000 and the final image was captured in 2014. The 13-year observation period yielded a 35.2% reduction in coral coverage within the area of interest. The initial state image contained 5,089 coral pixels or 4.58km<sup>2</sup>. The final state image contained 3,298 coral pixels or 2.97km<sup>2</sup>. While the area of interest for this site was somewhat large (94,731 pixels) the actual coral coverage within the site was somewhat small. Within the initial image 5.4% of the area contained live coral cover while only 3.5% of the final image represented coral. The posterior probability maps for 2000

(top) and 2014 (bottom) are detailed in Figure 4-17. Figure 4-18 maps the coral cover change when the 2000 initial state classification map is compared to the 2014 final state classification map. Figure 4-19 maps the cover type classification change between the two images. The results of the temporal change detection analysis can be viewed in Table 4-6.

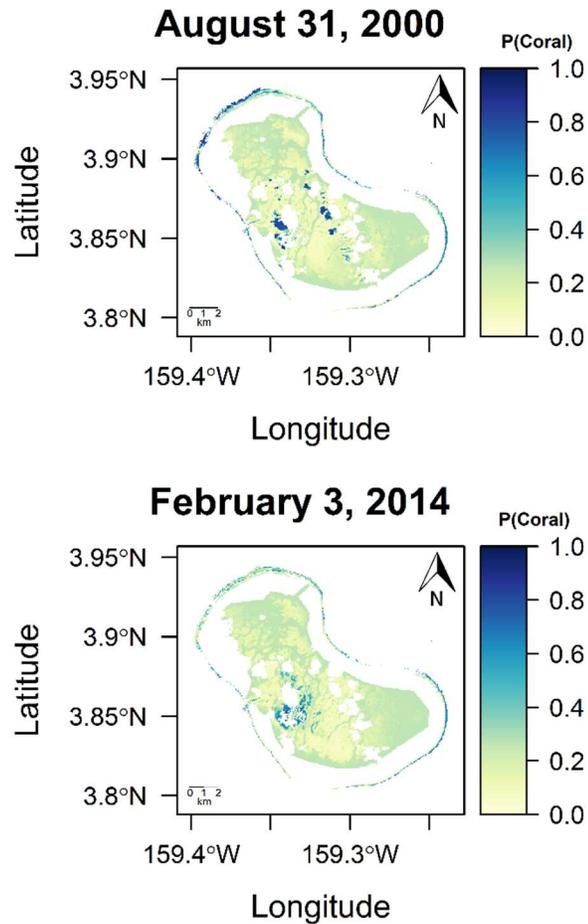


Figure 4-17: Posterior probability map for the Tabuaeran Island area of interest (top, 2000 and bottom, 2014).

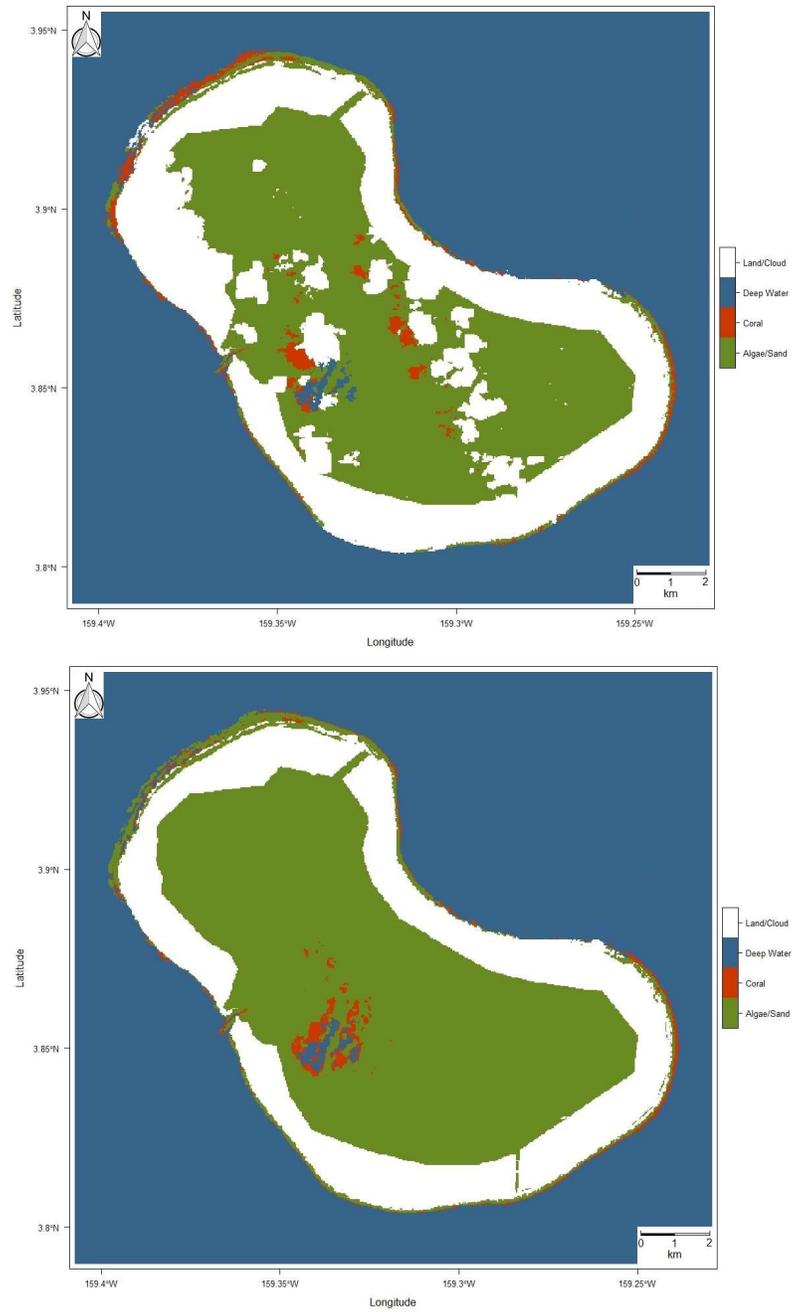


Figure 4-18: Predicted class map for the Tabuaeran Island area of interest (top, 2000 and bottom, 2014).

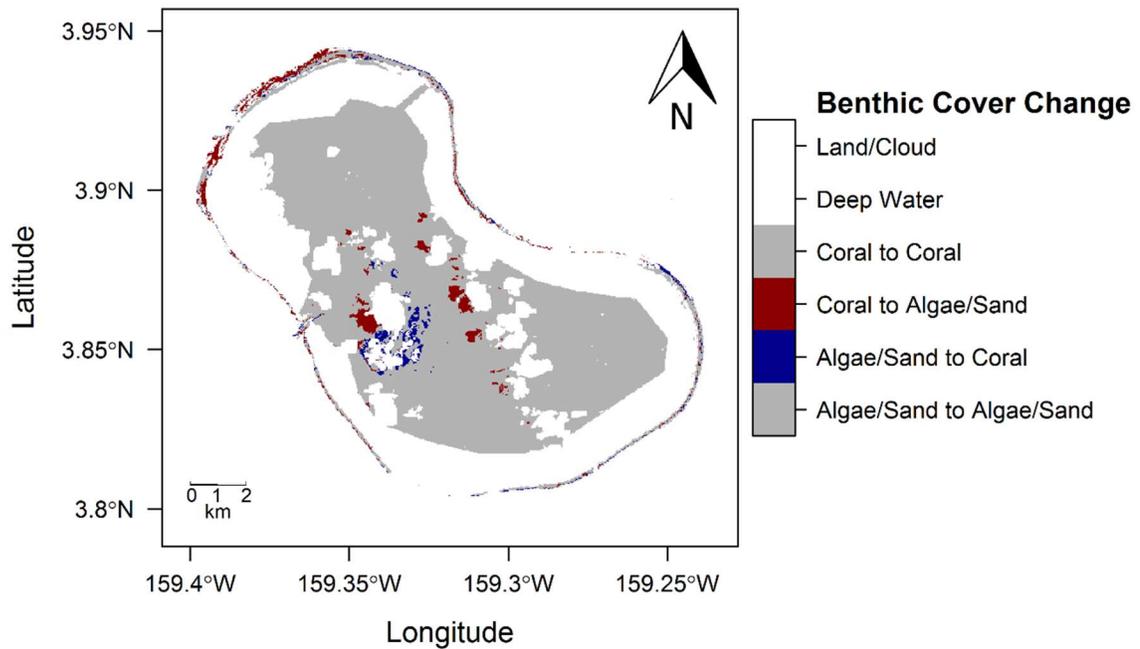


Figure 4-19: Difference in predicted class membership map for the Tabuaeran Island area of interest for 2000 initial state compared to 2014 final state.

#### 4.5.6 Kiritimati Island

The largest decrease in live coral cover between the initial observation date and final observation date was observed in the Kiritimati Island location. In this location there was a net decrease in coral cover of 42.7% between the initial state image captured in 2002 and the final state image captured in 2014. The initial image contained 29,303 pixels with live coral cover (26.37km<sup>2</sup>) while the final image contained 16,804 pixels with live coral cover (15.12km<sup>2</sup>). Furthermore, coral covered 19.0% of the initial image and only 11.0% of the final image. This site was also the largest area of interest analyzed covering 153,408 valid, shallow water pixels. The posterior probability mapping for Kiritimati Island for 2002 (top) and 2014 (bottom) can be found in Figure 4-20. Figure 4-21 maps the coral cover change when the 2002 initial state classification map is compared to the 2014 final

state classification map. Figure 4-22 maps the cover type classification change between the two images. Table 4-6 contains the results of the change detection analysis for Kiritimati Island.

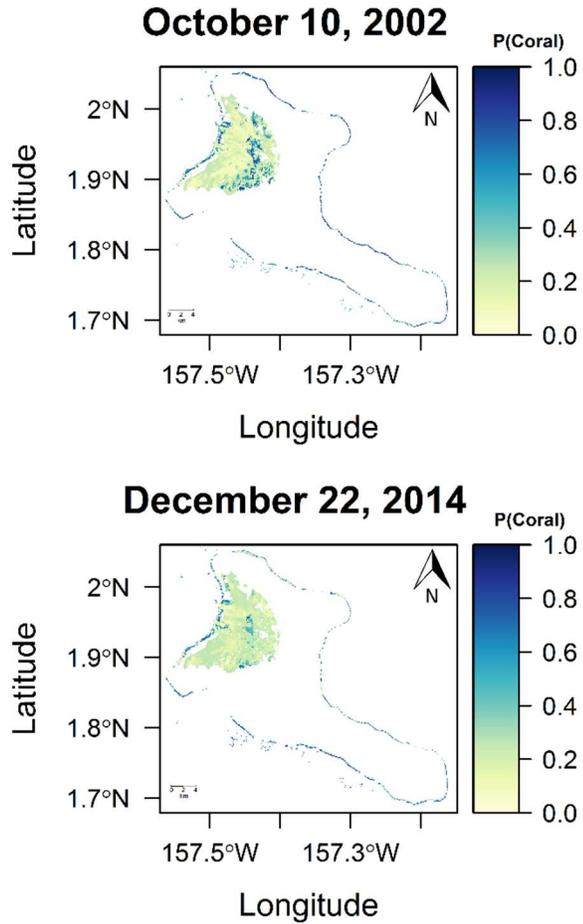


Figure 4-20: Posterior probability map for the Kiritimati Island area of interest (top, 2002 and bottom, 2014).

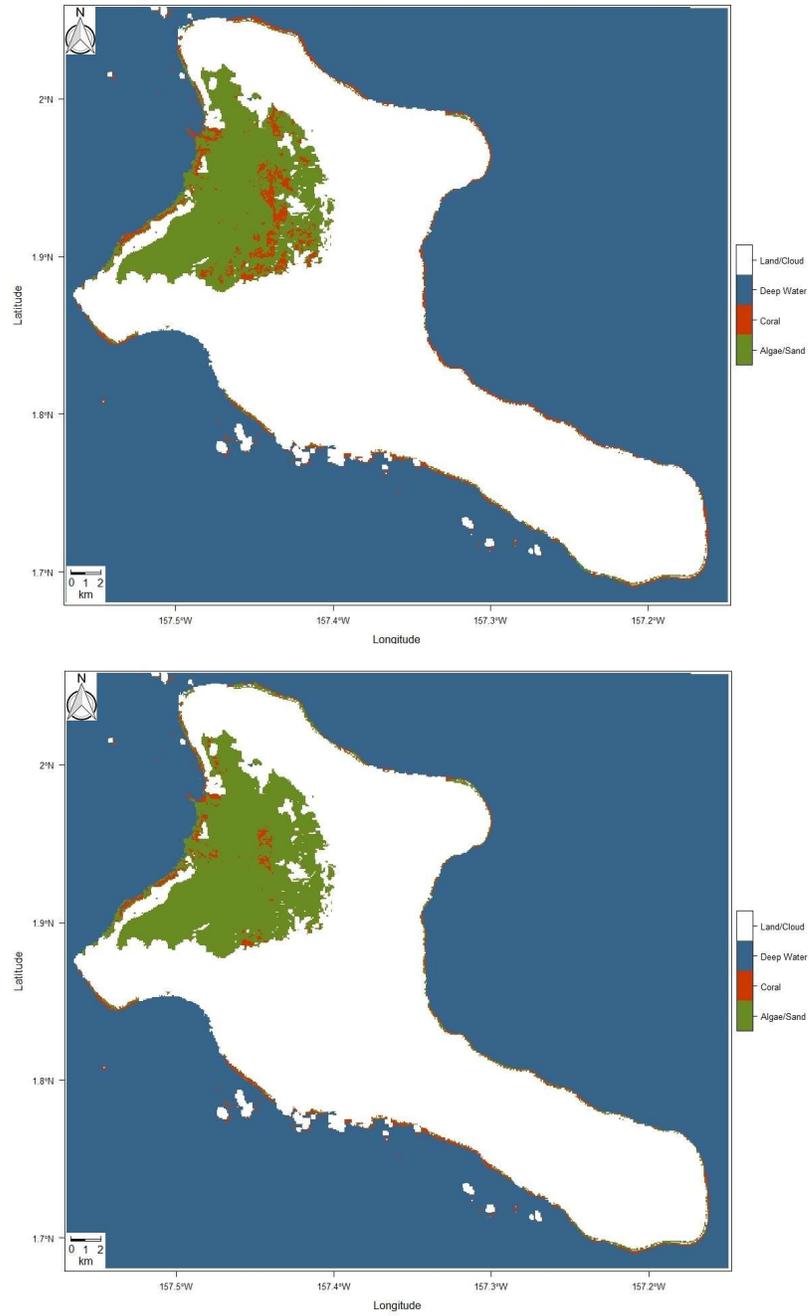


Figure 4-21: Predicted class map for the Kiritimati Island area of interest (top, 2002 and bottom, 2014).

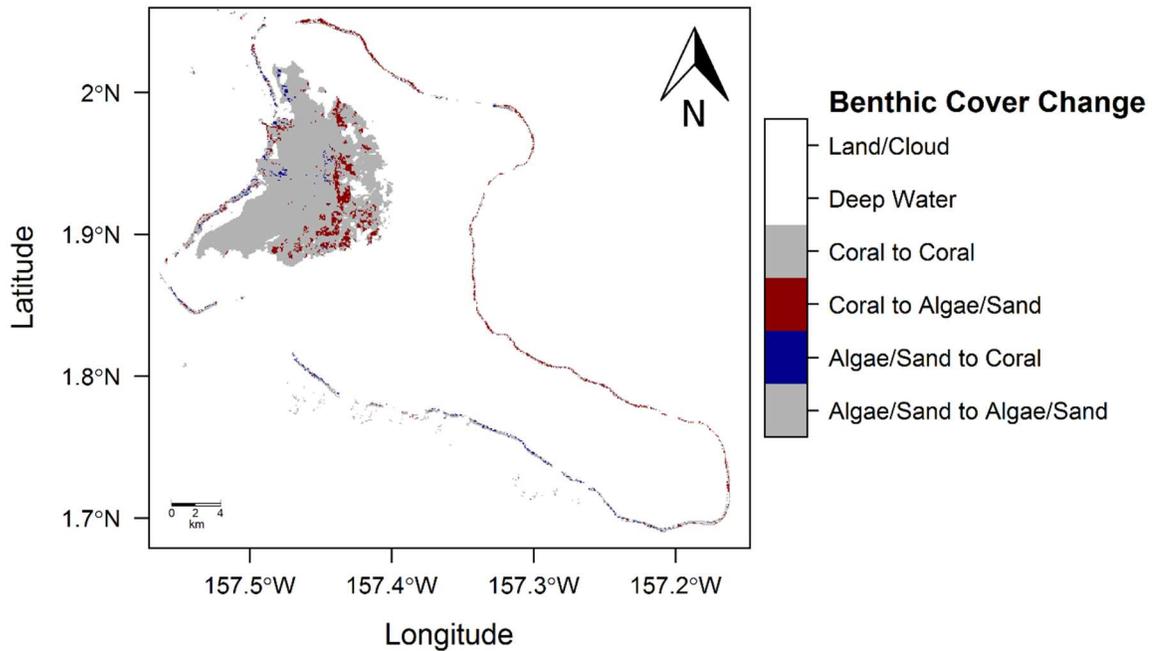


Figure 4-22: Difference in predicted class membership map for the Kiritimati Island area of interest for 2002 initial state compared to 2014 final state.

Table 4-6: Change Detection Analysis by Site.

		Initial Class (pixel count)		Initial Class ( $km^2$ )	
		Coral	Not Coral	Coral	Not Coral
<b>Final Class</b>		<b>Palmyra Atoll</b>		<b>Palmyra Atoll</b>	
	<b>Coral</b>	14,242	2,168	12.82	1.95
	<b>Not Coral</b>	9,850	19,395	8.87	17.46
		<b>Kingman Reef</b>		<b>Kingman Reef</b>	
	<b>Coral</b>	23,642	1,218	21.28	1.10
	<b>Not Coral</b>	9,652	4,627	8.69	4.16
		<b>Baker Island Atoll</b>		<b>Baker Island Atoll</b>	
	<b>Coral</b>	154	126	0.14	0.11
	<b>Not Coral</b>	311	1,520	0.28	1.37
		<b>Howland Island</b>		<b>Howland Island</b>	
	<b>Coral</b>	94	94	0.08	0.08
	<b>Not Coral</b>	158	828	0.14	0.75
		<b>Tabuaeran Island</b>		<b>Tabuaeran Island</b>	
	<b>Coral</b>	1,544	1,754	1.39	1.58
	<b>Not Coral</b>	3,545	87,888	3.19	79.10
		<b>Kiritimati Island</b>		<b>Kiritimati Island</b>	
<b>Coral</b>	13,290	3,514	11.96	3.16	
<b>Not Coral</b>	16,013	120,105	14.41	108.53	

## 4.6 Discussion

The methodology presented in this study revealed a significant decrease in live coral cover of multiple coral reef sites in the Pacific Ocean. This was shown using change detection analysis as applied to six different locations comparing an initial state image to a final state image with a 12 to 14-year timespan between the two. The methodology leveraged a robust SVM classifier to evaluate each site. The benefit of the SVM classifier over other classifiers such as LDA, QDA, and regression with regularization ( $\ell_1$  and  $\ell_2$  penalty) is the ability for the algorithm to be trained even with limited data as well as the ability for the algorithm to efficiently generalize. These traits are a result of the model design. Specifically, the algorithm fits a maximal-margin hyperplane to separate the dependent feature classes. The exact placement of this hyperplane is based on the observations that are closest to the classifier decision boundary called the support vectors. The maximal-margin enables the trained model to efficiently generalize and the support vectors delineating the maximal-margin enables the model to be trained on a small amount of ground truth observations in a low dimensional feature space. As a result, the SVM model tends to generalize to new data more effectively than other classifiers. Table 4-7 summarizes the classification accuracy of several alternative learning algorithms.

The primary benefit of the approach proposed in this section as well as other sections of this dissertation is the ability for a machine learning classifier trained on ground truth observations in several locations to generalize to additional locations. This capability was demonstrated through the CPCV process which yielded up to 81% accuracy when tested against ground truth observations from a location that was withheld from the model

training set. This is due to the reduction in site-specific bias such as local geomorphology, environmental condition, and unique coral fauna enabled by training based on ground truth data from multiple sites. Specifically, training based on a wider domain of ground truth observations which span multiple locations allows for a broader array of locations to be effectively represented. The implications of this is an expanded breadth of coral cover classification beyond isolated in-situ analysis. This is a powerful tool that should be leveraged to reduce the cost of monitoring the health of our coral reefs while expanding the coverage of coral reef monitoring. The natural evolution of in-situ change detection analysis to change detection analysis at scale will rely on the concept of generalizing a robust classifier longitudinally and the CPCV assessment as asserted in this research.

Table 4-7: Classification accuracy of select additional learning algorithms.

	<b>Palmyra Atoll</b>	<b>Kingman Reef</b>	<b>Baker Island Atoll</b>	<b>Howland Island</b>	<b>Consolidated Sites</b>
LASSO Regression	67.7%	68.3%	56.0%	51.9%	64.6%
Ridge Regression	67.7%	68.3%	56.0%	51.9%	64.6%
Logistic Regression	78.5%	82.3%	61.5%	71.4%	72.7%
QDA	76.9%	78.0%	61.5%	64.3%	73.3%

The primary challenge of the proposed method is that, while the classifier can generalize more effectively, its overall accuracy is lower than in-situ models. In-situ analysis based on Landsat data generally obtain accuracies of up to 80% or higher [1]. The robust, consolidated model developed in this research did not achieve as high of an accuracy. This is a natural result when the location-based bias of in-situ analysis are stripped away. Rather, the model can generalize more appropriately to additional sites as a direct result of averting biases such as local geomorphology and ecology. In-situ analyses incorporate this bias into the training of their predictors, and therefore yield higher accuracies but cannot generalize to new locations.

An additional challenge of the proposed method is that it relies upon medium resolution Landsat data. While the Landsat missions have proven to be an excellent data source for long-term temporal change analysis [95] [96] [97] [98] [99] [100] [101] [102] [103] [104], the comparatively low resolution represents a challenge to accuracy benthic cover classification. The 30m resolution of the Landsat platform means that highly heterogeneous environments such as coral reefs are difficult to classify due to the degree of within pixel mixing. Alternative higher resolution satellite platforms such as Sentinel-2 can be used to obtain benthic cover classification accuracies greater than studies based on Landsat images often showing as much as a 10% improvement [87]. In addition, new approaches based on high-resolution imagery show promising results. Most recently, research on object-based classification using high-resolution imagery from QuickBird and WorldView-2 have obtained classification accuracies over 90% [88]. However, it has been shown that these same methods deliver similar results to pixel based benthic cover classification when applied to medium resolution images such as those produced by the Landsat missions [93] [94]. Furthermore, the deployment of high-resolution remote sensing platforms is constrained by the progression of technological advances. As a result, these high-resolution platforms are more a recent development compared to the rich history provided by Landsat and any long-term (>10-years) evaluation of live coral cover change necessitates that the research be conducted using Landsat imagery. Similarly, the lack of a coastal aerosol band on missions prior to Landsat 8 is an unfortunate challenge given the water column penetrating properties of light in this channel.

## 4.7 Conclusion

This study measured the decline in live coral cover of six Remote Pacific Island sites over the past 12 to 14 years. This was achieved by building a classifier for each of four locations (Palmyra Atoll, Kingman Reef, Baker Island Atoll, and Howland Island), applying the classifier to the initial state image (captured between 2000 and 2002) and final state image (captured between 2014 and 2015), then conducting a per-pixel change detection analysis. The SVM algorithm as applied to these four sites achieved a classification accuracy as high as 87.9% (Palmyra Atoll site). When applied to Kingman Reef, Baker Island Atoll, and Howland Island the model achieved accuracies of 85.7%, 69.23, and 82.1% respectively. The results of the change detection analysis revealed an overall decline of coral coverage in these sites by as much as 38.8%. Within the Palmyra Atoll area of interest, a decline in coral from 2001 to 2015 of 31.9% was observed. This represents a net decrease of  $6.91 \text{ km}^2$  of coral. The Kingman Reef site and Howland Island sites incurred the least severe decline by percentage. However, these sites still incurred more than a 20% drop in coral cover. Within Kingman Reef a decline in coral cover of 25.3% was observed or  $7.59 \text{ km}^2$ . Similarly, within the Howland Island site coral declined 25.0% or  $0.06 \text{ km}^2$ . Coral coverage within the Baker Island Atoll area of interest declined by 39.8% or  $0.17 \text{ km}^2$ . Further analysis was conducted of two additional sites, Tabuaeran Island and Kiritimati Island. These sites were analyzed by developing an SVM classifier using the combined observation data from all four sites. Because this consolidated model was trained using information drawn from multiple sites, it is more robust to site-specific biases that in-situ models rely on. This methodology was validated using a CPCV process of training a model using the combined ground truth data from three sites and applying it

to the third. In this way, the CPCV assessment is an indication of how well the SVM classifier can generalize to data from new sites. The results of the CPCV assessment showed that the SVM classifier obtained an accuracy of 78.8% when tested against the Palmyra Atoll ground truth observations, 81.0% when tested against the Kingman Reef ground truth observations, 65.4% when tested against the Baker Island Atoll ground truth observations, and 67.9% when tested against the Howland Island ground truth observations. An SVM classifier was created using the consolidated ground truth data from all four sites and evaluated using cross-validation. The resulting algorithm correctly classified 75.3% across all ground truth observations across the four training sites. This consolidated classifier was then deployed to conduct a change detection analysis on the Tabuaeran Island and Kiritimati Island sites. The analysis revealed a 35.2% reduction in coral cover at Tabuaeran Island (1.61km<sup>2</sup>) when the 2000 initial state image was compared to the 2014 final state image. Kiritimati Island incurred the largest decrease in live coral coverage of all sites included in this study. In this location the analysis revealed a reduction in coral coverage of 11.25km<sup>2</sup> or 42.7%.

# **5. Third Study: A Generalized Machine Learning Classifier for Spatiotemporal Analysis of Coral Reefs in the Red Sea**

## **5.1 Introduction**

Coral reefs comprise less than 0.2% of the oceans yet contain 35% of all known marine species [136]. This remarkable concentration of biodiversity makes coral reefs an exceptionally unique ecosystem. However, this ecosystem is in decline. The effect of global bleaching events on coral reefs as well as the impact of local stressors are well documented [5] [6] [7] [8] [9] [10] [11] [12] [110] [111] [112] [137] [138]. In 2015 the NOAA declared the third global bleaching event. As a result, coral bleaching has now become the main driver of coral reef degradation globally [113]. This has led some researchers to project the total loss of this critical environment [114] [139]. Therefore, identifying reefs that may be more resilient to the impact of climate change due to local geomorphology or unique coral species biodiversity is now a scientific imperative.

Landsat is a commonly used remote sensing platform for change analysis of all types. The analysis of coral reefs is no exception. An abundance of studies that isolate a single location for change detection analysis [36] [61] [106] [107]. The framework of these

in-situ studies are common. First, an initial state and a final state image are selected for the location. Next, a classifier is trained on the final state image then applied to the initial state image. Finally, a per-pixel or similar [88] [90] [91] [92] change analysis is conducted. This well-established framework is effective for identifying changes in coral reefs for that isolated, in-situ location [43]. Yet, the cost of obtaining ground truth data on which to train a classifier limits the scope of this approach. That is, each change analysis study requires an associated set of ground truth observations on which the classifier can be trained. The alternative is to deploy an unsupervised learner, however accuracies of these methods are often inferior to supervised methods. In order to expand the in-situ change detection approach beyond the limited scope of individual locations, a new methodology to generalize the results longitudinally must be developed.

This study leveraged a robust machine learning classifier trained using the combined information of two large areas of interest in order to conduct a change analysis of the coral cover in the Red Sea. The spatial and temporal extent undertaken by this study has never been accomplished before. This research builds upon the previous in-situ methodology to produce a classifier that is robust to location specific biases. While removing these localized biases reduces the performance within the training location, it enables the classifier to generalize more effectively to new locations. In this way, the generalized classifier is robust to overfitting site-specific conditions as is the case with an in-situ approach. In addition to expanding the boundaries of spatial constraints, this research aims to evaluate a longer time period than previous research. Very little research has been conducted to evaluate the change in coral reefs over an 18-year period using remote sensing data [105]. Furthermore, given the frequency and severity of coral

bleaching events over the time period, this research is a critical check point for identifying how coral reefs in the Red Sea are surviving under the threat of climate change.

Section Two of this chapter is a discussion of the materials and methods used in this research. This includes a review of the data used as well as a description of the selected locations. The Materials and Methods Section is followed by Section Three which includes an evaluation of the per-site performance of the classifier and a quantitative assessment of the generalized classifier. The Results Section concludes with a review of the results from the change detection analysis for each site. Section Four includes a discussion and evaluation of the methodology undertaken in the research including benefits and challenges as well as a proposal for future work in the area. The study concludes with a review of the resulting outcomes and conclusions resulting from this research.

## **5.2 Materials and Methods**

### **5.2.1 Study Area**

The first location evaluated in this study was the coastal region near the city of Gulf of Aqaba. This area of interest spans from approximately 35°20'E to 34°35'E and 27°30'N to 28°50'N, including Sanafir Island. This location is one of the northmost coral reefs in the world. The second location studied was a group of reefs immediately off the coast of Umluj from approximately 36°55'E to 37°20'E and 24°35'N to 25°20'N. Both locations are in a unique environment for coral reefs. First, the Red sea is encompassed by desert and therefore receives very little water from rivers. Due to this, there is little fluctuation in salinity, temperature, and water quality within the coral reefs of the Red Sea [140]. Figure 5-1 identifies each of the three Red Sea AOIs and their relative location.

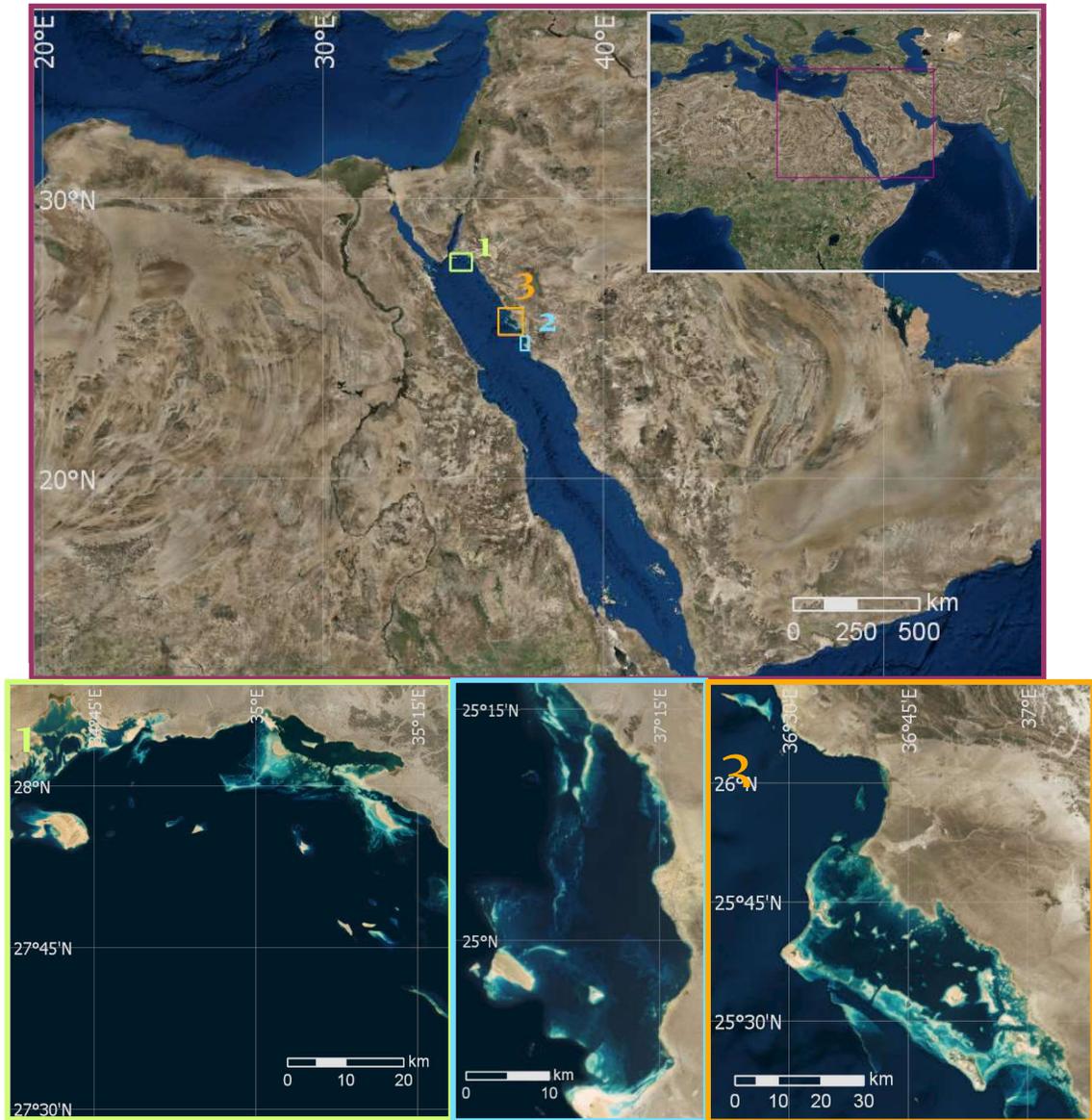


Figure 5-1: Relative location of the three locations under study (1) Gulf of Aqaba Location, (2) Umluj Location, (3) Al Wajh Location

### 5.2.2 Data Used and Preprocessing Steps

Images captured by Landsat 8 OLI and Landsat 7 ETM captured in 2018 and 2000, respectively were used to evaluate the locations of this study. Both Landsat missions have at least three visible bands with 30m spatial resolution. For this study, Landsat 8 band 2 (0.450-0.515 $\mu$ m), band 3 (0.525-0.600 $\mu$ m), and band 4 (0.630-0.680 $\mu$ m) were used due to

their water column penetrating properties. The coastal aerosol, band 1 was not used since there is no comparable band present in the Landsat 7 data. Landsat 7 band 2 (0.45-0.51  $\mu\text{m}$ ), band 3 (0.53-0.59  $\mu\text{m}$ ) and band 4 (0.64-0.67  $\mu\text{m}$ ) were used to analyze the previous condition of each site. Each Landsat mission is on a 16-day revisit cycle. 2018 images were selected from all available Landsat scenes to minimize the amount of cloud cover and obscurities with particular care given to the specific area of interest within the scene. Similarly, the 2000 images were selected from all available Landsat 7 images after taking into account the location of cloud cover and other obscurities. For each of the sites, Table 5-1 displays the Landsat Path-Row, Landsat 8 OLI and Landsat 7 ETM scene capture data, area of interest dimensions, and total number of ground truth observations.

Table 5-1: Summary of Data Used.

Site (Figure 5-1)	Path/ Row	Site	Latitude- Longitude	Final Scene Date	Baseline Scene Date	AOI Dimensions	Ground Truth Points
1	174/041	Gulf of Aqaba	27°57'N 34°50'E	11/7/2018	3/18/2000	65.7×64.7-km	1,085
2	172/043	Umluj	25°00'N 37°10'E	12/11/2018	2/17/2000	55.3×28.4-km	196
3	172/042	Al Wajh	25°35'N 36°48'E	2/26/2018	6/24/2000	104.1×77.7-km	-

### 5.2.3 Preprocessing

For each location a Landsat Tier 1 data product from 2000 and 2018 was obtained. All images underwent radiometric characterization and geometric correction as L1TP corrected data. The data were determined to have well-characterized radiometry within image-to-image tolerances of  $\leq 12\text{m}$  radial RMSE [117] [118]. Clouds and other obscured pixels were masked from the images by leveraging the Landsat BQA band [53] [54]. A water mask was developed using the respective Landsat NIR bands. Light in the NIR

spectral range (0.851 and 0.0879- $\mu\text{m}$ ) does not penetrate water therefore this band was used to identify areas of full wavelength absorption and develop the water mask [56]. Atmospheric correction was performed by dark-pixel subtraction method followed by sun glint correction [57] [58] [1] [59] [60] [55] [62]. Water column correction was then applied by calculating the per pixel DII according to the method of Lyzenga [63] [64] [65] [66]. The method of Lyzenga leverages the ratio of attenuation coefficients between each pair of bands thereby avoiding the need to calculate estimates of  $k$  for each band directly [63] [64]. The DII is calculated accordingly with the following equation as discussed in Chapter One:

$$DII_{ij} = \ln(L_i - L_{si}) - \left[ \frac{k_i}{k_j} \cdot \ln(L_j - L_{sj}) \right]$$

These DII values resulting from the preprocessing steps were then used as inputs for the machine learning classifier.

#### **5.2.4 Generalized Machine Learning Classifier**

The DII values resulting from the preprocessed images were first used to train a classifier for the Gulf of Aqaba site using data only from that location. In this way, the Gulf of Aqaba location served as a baseline to evaluate how well the algorithm is generalizing to additional sites. First, a stratified random sampling of the DII and observation data was performed to correct for class imbalance. A SVM algorithm was then applied to the sampled DII values derived from the Landsat 8 image of the Gulf of Aqaba location. Accuracy was evaluated using ground truth observations. The algorithm was tuned using cross-validation to obtain the optimal cost and gamma values taking into account both

accuracy and generalizability. A radial kernel function outperformed other kernel methods. Care was taken not to overfit the model to the Gulf of Aqaba site. Additional features were tried such as entropy, however, while the classification accuracy for this specific location went up, it was discovered that the additional features could not generalize to new data. Once the optimal hyperparameters were derived, the tuned SVM algorithm was evaluated based on six different measures: accuracy (percentage of ground truth observations correctly classified), precision (1 – user error), recall (producer error), specificity, F-measure, and kappa-coefficient. Once the accuracy assessment was completed, the SVM model was applied to the DII values from the entire Landsat 7 and Landsat 8 scenes. A map of coral cover based on the posterior probability of each pixel was developed and applying a threshold to these probabilities derived a predicted class which also was mapped. Coral cover change detection analysis was then conducted for the Gulf of Aqaba site through a per-pixel comparison of the initial class derived from the Landsat 7 image and the final class derived from the Landsat 8 image.

The algorithm developed using the DII and ground truth observations trained on the Gulf of Aqaba site was then deployed to the truth observations of the Umluj location. In this way, the research process resembles the well-known data science practice of train and test split and a true evaluation of the model robustness can be determined. The same assessment criteria used to evaluate the model as applied to the Gulf of Aqaba site was also used to evaluate the performance against the Umluj site. Once the model was evaluated, a per pixel map of the posterior probability that the pixel belonged to the coral class was created for Umluj. A threshold was applied to derive the per pixel predicted class and associated map. Using these data, a coral cover change analysis was conducted by

comparing the initial state classifications derived using the Landsat 7 data to the final state classifications derived using the Landsat 8 data on a per pixel basis.

Following the model evaluation as applied to the Gulf of Aqaba location and Umluj location, a robust machine learning classifier was developed using the combined ground truth observations of both sites. This classifier was trained using the consolidated ground truth observations as the dependent variable and associated DII pixel values from both Gulf of Aqaba and Umluj as the independent variables. The resulting robust classifier was then applied to change detection analysis of the Al Wajh location by comparing the per pixel predicted class using the DII values derived from the Landsat 8 image to those of the Landsat 7 image.

Using the consolidated information from both sites allowed the resulting classifier to more effectively generalize to new information from additional sites [133]. This result is due to the decrease in site-specific bias that results from training the algorithm on data that is more representative of the greater ecology and geomorphological conditions of the region. The result is an algorithm that is more representative of the region rather than the specific location-based bias that in-situ analyses rely upon. Training the SVM using data from data that is representative of the region rather than a specific reef enables larger scope analysis of coral reefs than has previously been conducted. Figure 5-2 outlines the processing steps for training the SVM classifier, validating the classifier generalization, and application of the robust classifier. Figure 5-3 outlines the processing steps taken to conduct the change detection analysis. Implementation and analysis were performed using the open source R programming language and environment for statistical computing [52].

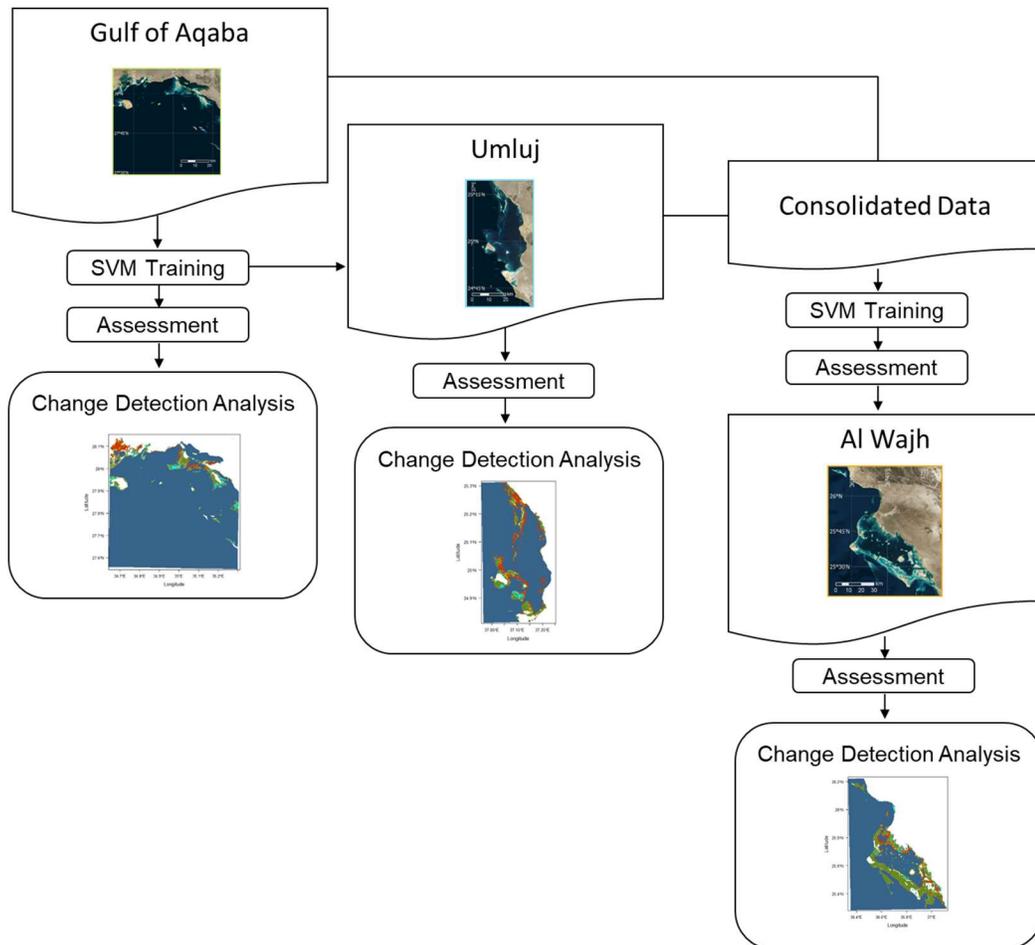


Figure 5-2: SVM Classifier training (Gulf of Aqaba), validation (Umluj), and application of the robust classifier (Al Wajh).

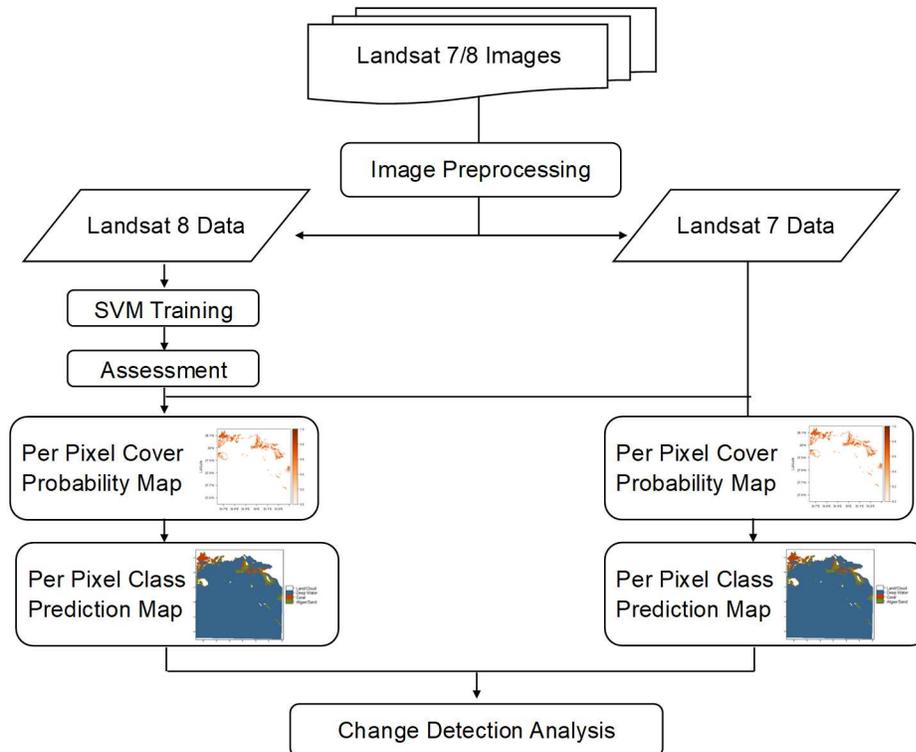


Figure 5-3: Temporal change detection analysis process flow.

## 5.3 Results

### 5.3.1 Classification Accuracy by Site

The generalization of the SVM model was evaluated using six different metrics. First, the raw accuracy was calculated. This is simply the percentage of ground truth observations which were correctly classified. While this is an extremely common metric for accuracy evaluation, it is really only a superficial evaluation of model performance. In addition to accuracy, precision and recall were calculated. Precision and recall are statistical measures of performance with respect to type I (false positives) and type II errors (false negatives), respectively. Specifically, precision is a measure of the classifier's ability to be exact while recall is a measure of how complete the results are. Precision is equal to  $(1 - \text{user error})$  and recall is equivalent to  $(\text{producer error})$ . Specificity, also called

the true negative rate, is another measure of accuracy which evaluates how many of the true negatives are correctly identified. Another common measure of accuracy is the F-measure. The F-measure is the harmonic mean of precision and recall which is an effective tool for evaluating the balance between these two measures and overall accuracy of the model. Finally, the kappa-coefficient was measured as the ratio of correct classification with respect to baseline agreement. While this metric has been questioned particularly in scenarios in which there is significant class imbalance (as is the case with most coral cover detection analysis), this metric is still informative in conjunction with the previously stated metrics. As a result, the kappa-coefficient was also used in evaluation of how well the SVM machine learning algorithm generalized. Table 5-2 is a representation of the results from each of these calculations as applied to both the Gulf of Aqaba location data and Umluj location data.

The first location evaluated, Gulf of Aqaba, contained 1,085 coral cover, ground truth observation points. A stratified random sampling of these observations was applied to reduce class imbalance and ensure a more robust classification of pixel coral cover. The sampled data contained a total of 404 observations 78.22% of which were correctly classified by the tuned SVM classifier. The algorithm also obtained a precision of 0.7664 and recall of 0.8119. The F-measure was calculated at 0.7885 and specificity measured 0.7525. The kappa-coefficient yielded by the model as applied to the Gulf of Aqaba location was 0.5644.

Applying the algorithm trained and tuned on the Gulf of Aqaba location data to the truth observations from the Umluj location yielded an accuracy of 72.73%. This decrease

in accuracy is expected due to the loss of site-specific bias which in-situ analysis rely upon. Rather, resulting from the construct of building a robust algorithm, the principal goal of this research was to produce a machine learning classifier that could generalize beyond the scope of the limited in-situ analysis. Therefore, the decrease in accuracy is acceptable given the challenge of generalization to new datasets. As with the Gulf of Aqaba location, stratified random sampling was used to adjust for class imbalance in the Umluj observations. A total of 44 ground truth observations were selected. The model tuned on the Gulf of Aqaba data and applied to the Umluj truth data yielded a precision and recall of 0.7500 and 0.6818, respectively. In addition, the model yielded a specificity of 0.7727 and recall of 0.6818. The resulting F-measure was 0.7143 showing a good balance between precision and recall. Lastly, the kappa-coefficient was calculated to be 0.4545. Table 5-2 contains a summary of these figures for comparison. In addition, Table 5-3 identifies the confusion matrices for each site as well as the results from the robust model trained using the consolidated data from both locations.

Table 5-2: SVM Classifier Performance Assessment by Site and Consolidated Model.

	<b>Gulf of Aqaba</b>	<b>Umluj</b>	<b>Consolidated Model</b>
Accuracy	78.22%	72.73%	70.98%
Precision	0.7664	0.7500	0.6992
Recall	0.8119	0.6818	0.7366
Specificity	0.7525	0.7727	0.6830
F-measure	0.7885	0.7143	0.7174
Kappa	0.5644	0.4545	0.4196

Table 5-3: Confusion Matrix by Site and the Consolidated Model.

		<b>Ground Truth Labels</b>	
		<b>Coral</b>	<b>Not Coral</b>
<b>Predicted Class</b>		<b>Gulf of Aqaba</b>	
	<b>Coral</b>	164	50
	<b>Not Coral</b>	38	152
		<b>Umluj</b>	
	<b>Coral</b>	15	5
	<b>Not Coral</b>	7	17
		<b>Consolidated Model</b>	
	<b>Coral</b>	165	71
	<b>Not Coral</b>	59	153

### 5.3.2 Consolidated Model Robust to Site-Specific Bias

Once the generalization criteria between the Gulf of Aqaba and Umluj site was measured, a robust SVM algorithm was trained using the consolidated information from both sites. The input features for this model were the stratified random sample of per pixel DII values from both the Gulf of Aqaba and Umluj locations which were used to predict the associated ground truth cover type classifications. Combining the data from both locations results in an algorithm that is able to generalize more adequately to information from new locations. However, the ability for the classifier to generalize comes with a small decrease in accuracy. This decrease is because the algorithm is no longer allowed the benefit of using site-specific biases such as water turbidity, local geomorphology, and the marine fauna that live in the specific location. These site-specific biases are what many in-situ analysis rely upon and, therefore, these models actually overfit to the data of a single location. As a result, while these classifiers perform well for the given location in which they are trained they cannot generalize to new data beyond the limited scope, in-situ study. The robust machine learning classifier is trained using inputs from multiple sites. As a result, the inputs are more representative of the benthic habitats of an entire region rather

than a single, exclusive reef within the region. Training the model using inputs that are representative of a larger area will enable the algorithm to more appropriately generalize to reefs throughout that region although at the cost of accuracy for any single reef within the area. The robust SVM classifier correctly classified 318 of the 448 ground truth pixels selected from the stratified random sampling of data from both the Gulf of Aqaba and Umluj locations. 70.98% of the pixels were correctly classified. The precision and recall of the model were 0.6992 and 0.7366, respectively. The model attained a specificity of 0.6830 and an F-measure of 0.7174. The kappa-coefficient was calculated to 0.4196. The results of this assessment are included in Table 5-2 as well as the confusion matrix in Table 5-3. The ROC of the consolidated model can be observed and compared to the performance of the model based on the Gulf of Aqaba location data in Figure 5-4. The resulting AUC of the consolidated model was 0.7754.

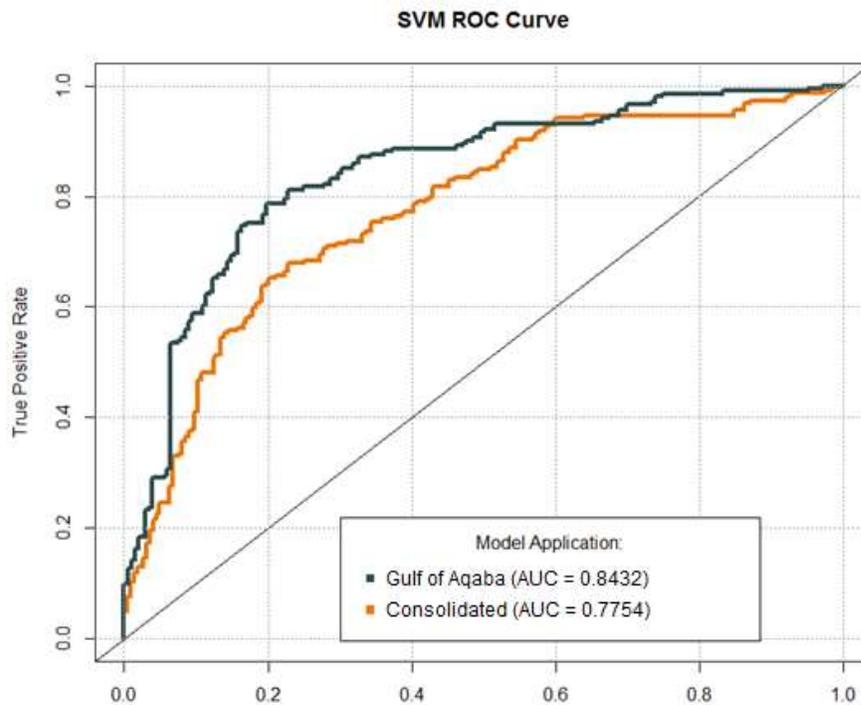


Figure 5-4: Performance evaluation of the robust, combined classifier using ROC Curve and AUC for the Gulf of Aqaba and consolidated model.

### 5.3.3 Change Detection Analysis

#### Gulf of Aqaba Location

The initial Landsat 7 image of the Gulf of Aqaba location was captured on March 18, 2000. This image was compared to a Landsat 8 image of the same location captured on November 7, 2018. The initial image contained 147,014 pixels identified as containing coral by the trained SVM classifier. In addition, 135,288 pixels were identified as algae, sand, and other benthic cover types. This corresponds to 132.31km<sup>2</sup> of coral. The 2018, final state image contained 130,225 coral pixels and 152,077 non-coral pixels. This represents 117.20km<sup>2</sup> of coral which is a reduction of 11.4% over the 18-year period. Furthermore, in the initial state image coral represented 52.1% of the shallow benthic

cover, however, in 2018 this was reduced to 46.1%. Figure 5-5 presents the posterior probability map for the initial, 2000 image (top) and final, 2018 image (bottom). Figure 5-6 presents the associated predicted class map based on the trained SVM algorithm for 2000 (top) and 2018 (bottom). Figure 5-7 is an image fusion, difference map between the initial per pixel predicted class membership and final predicted class membership. A summary of the results of the change detection analysis can be found in Table 5-4.

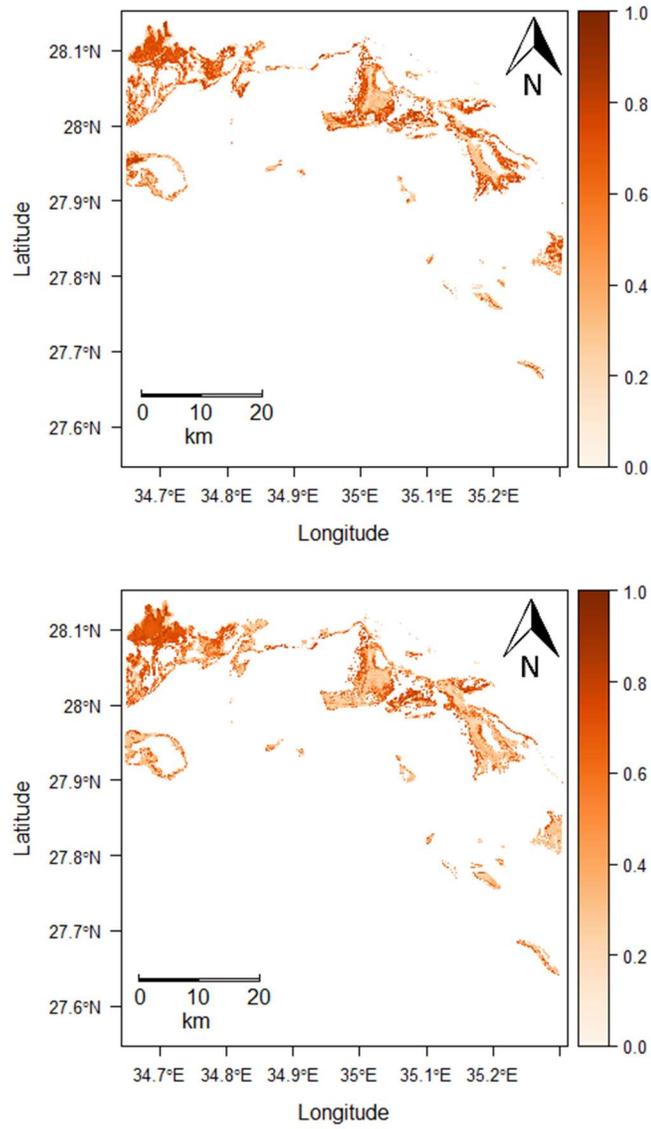


Figure 5-5: Posterior probability map for the Gulf of Aqaba area of interest (top, 2000 and bottom, 2018).

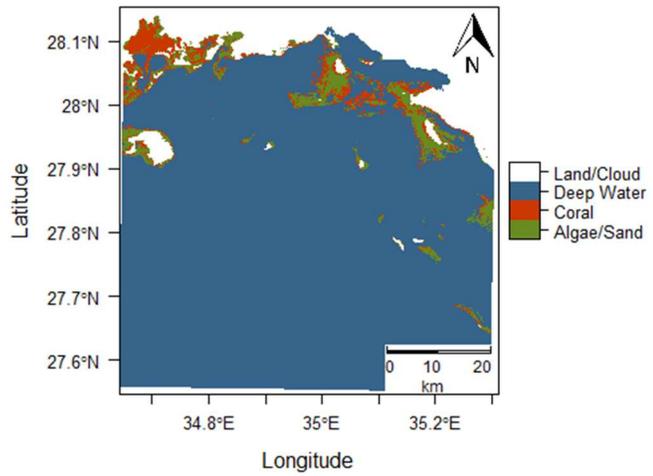
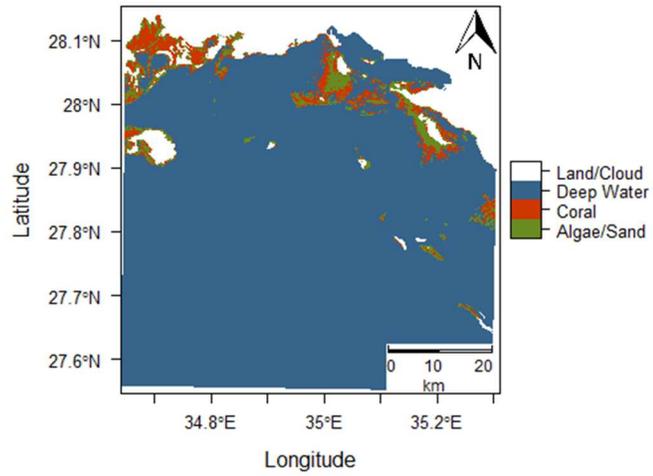


Figure 5-6: Predicted class membership map for the Gulf of Aqaba area of interest (top, 2000 and bottom, 2018).

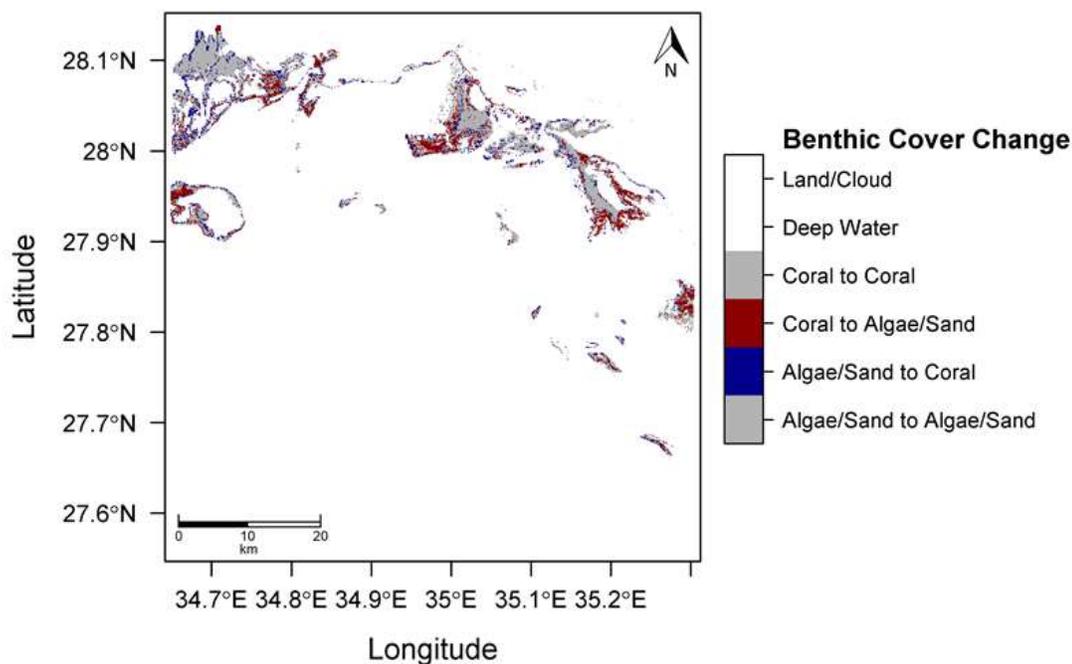


Figure 5-7: Change detection analysis between 2000 and 2018 of the Gulf of Aqaba area of interest.

### Umluj Location

The Umluj location incurred a smaller decline in coral cover than the other location. The initial Landsat 7 image of the location was captured on February 17, 2000. This image was compared to the Landsat 8 image captured on December 11, 2018. The initial, baseline image contained 113,284 pixels (101.96km<sup>2</sup>) classified as coral using the trained SVM classifier. This was compared to the final image which contained 109,443 pixels (98.50km<sup>2</sup>) classified as coral. This represents a decrease in coral cover of 3.4% over the 18-year period. In the initial state image 53.5% of the shallow water pixels were identified as containing coral compared to 51.7% of the final state image. A mapping of the Umluj location posterior probabilities for each pixel belonging to the coral class for both the 2000 (left) and 2018 (right) images are displayed in Figure 5-8. Figure 5-9 is a map of the per

pixel predicted class for both the baseline image (left) and final state image (right). A per pixel difference map between the initial state classification and final state classification of each pixel is presented in Figure 5-10. A summary of the results of the change detection analysis can be found in Table 5-4.

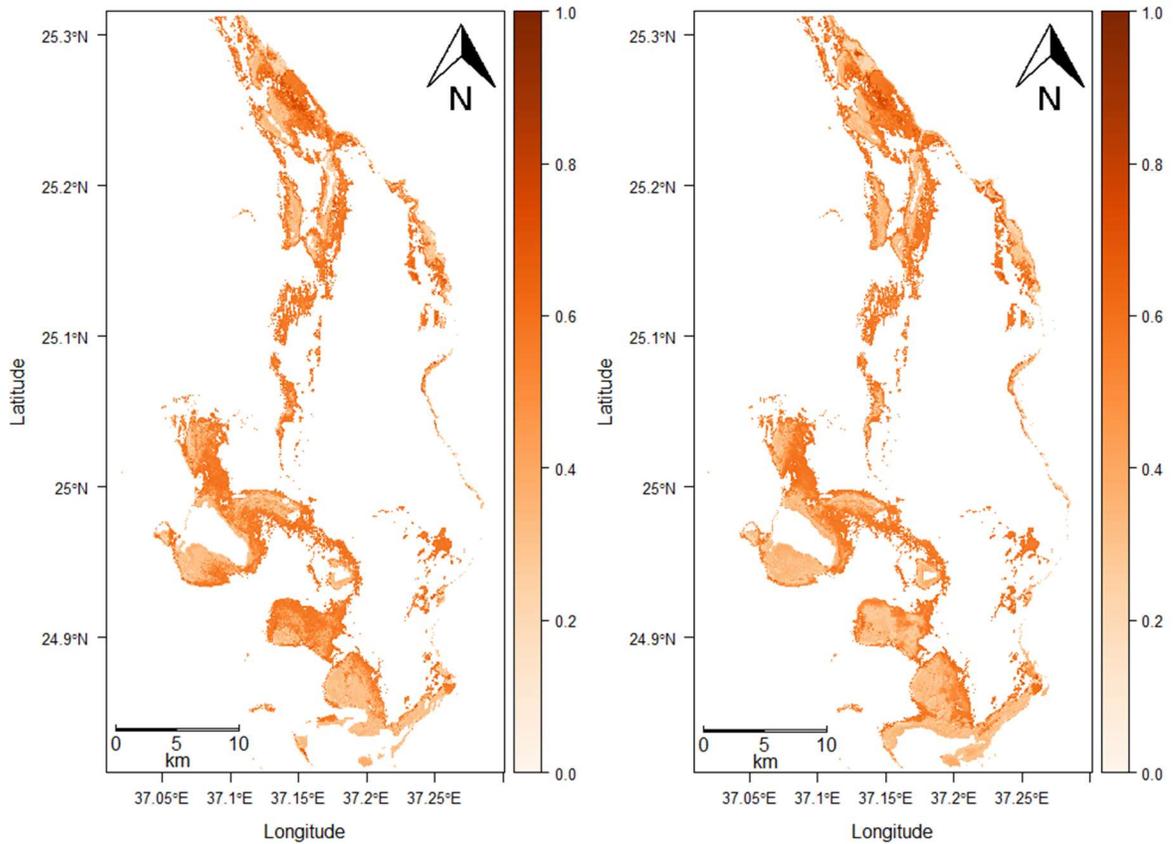


Figure 5-8: Posterior probability map for the Umluj area of interest (left, 2000 and right, 2018).

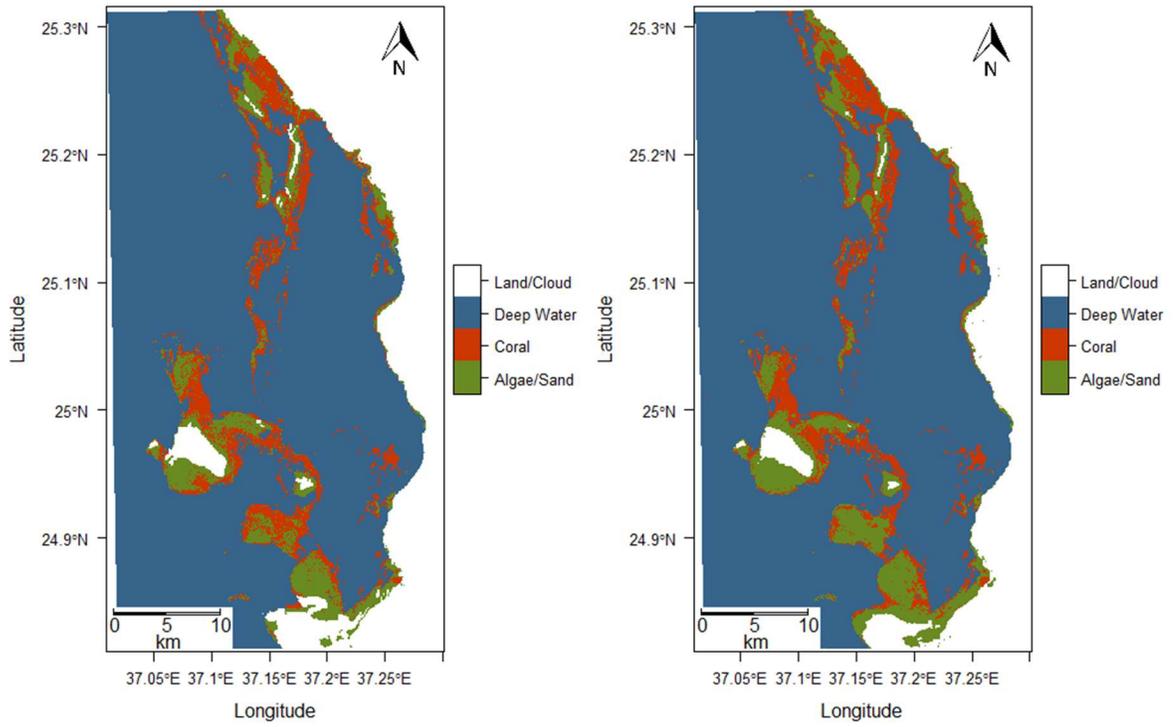


Figure 5-9: Predicted class membership map for the Umluj area of interest (left, 2000 and right, 2018).

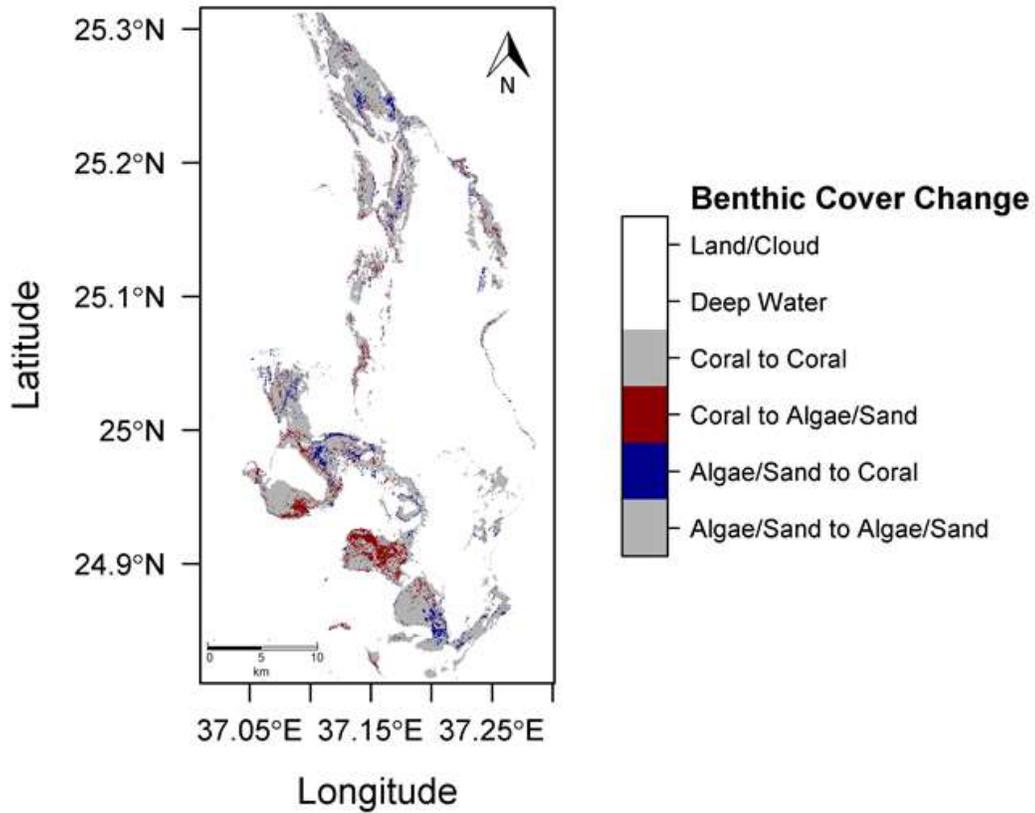


Figure 5-10: Change detection analysis between 2000 and 2018 of the Umluj area of interest.

### Al Wajh Location

The Al Wajh location sustained a decrease in coral cover of 13.6%. The initial evaluation of Al Wajh using the Landsat image from 2000 revealed 294,501 pixels (265.05km<sup>2</sup>) identified as containing coral. The final state image captured 18 years later revealed 254,567 coral pixels (229.11km<sup>2</sup>). The 18-year change in this location represents a 13.6% decrease in coral coverage. Furthermore, 28.2% of the shallow benthic area in the 2000 image contained coral compared to 24.4% of the 2018 image. A map of the initial state (left) and final state (right) posterior probabilities that each pixel contains coral can

be viewed in Figure 5-11. The resulting per pixel classification for 2000 (left) and 2018 (right) is displayed in Figure 5-12. Figure 5-13 displays the per pixel difference between the initial state class and final state class for the Al Wajh location. A summary of the results of the change detection analysis can be found in Table 5-4.

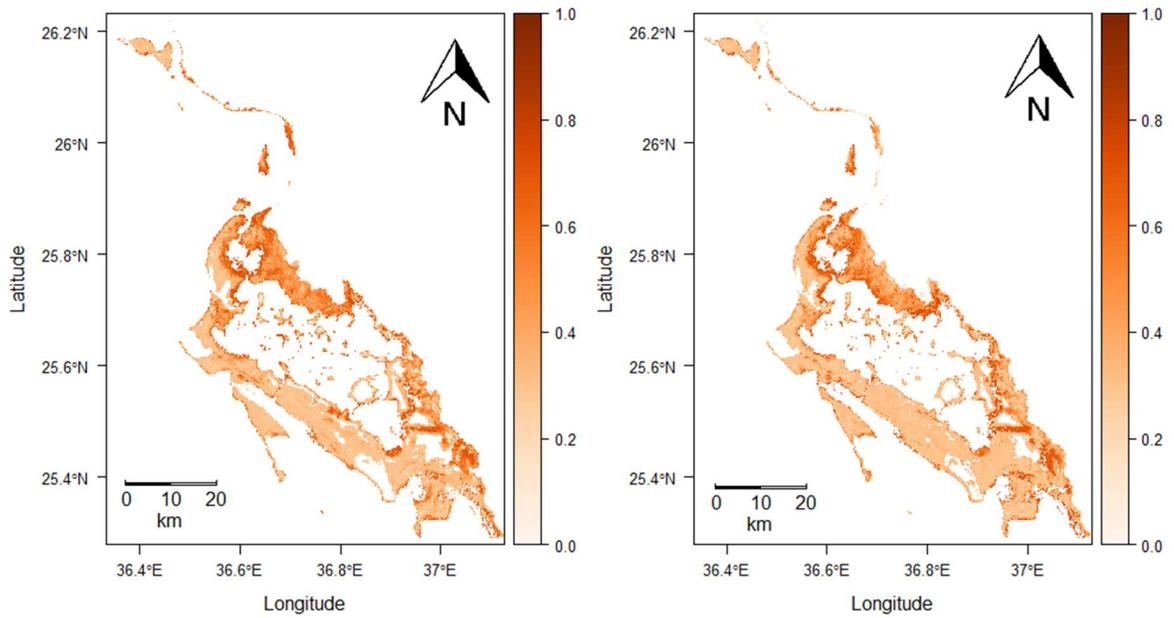


Figure 5-11: Posterior probability map for the Al Wajh area of interest (left, 2000 and right, 2018).

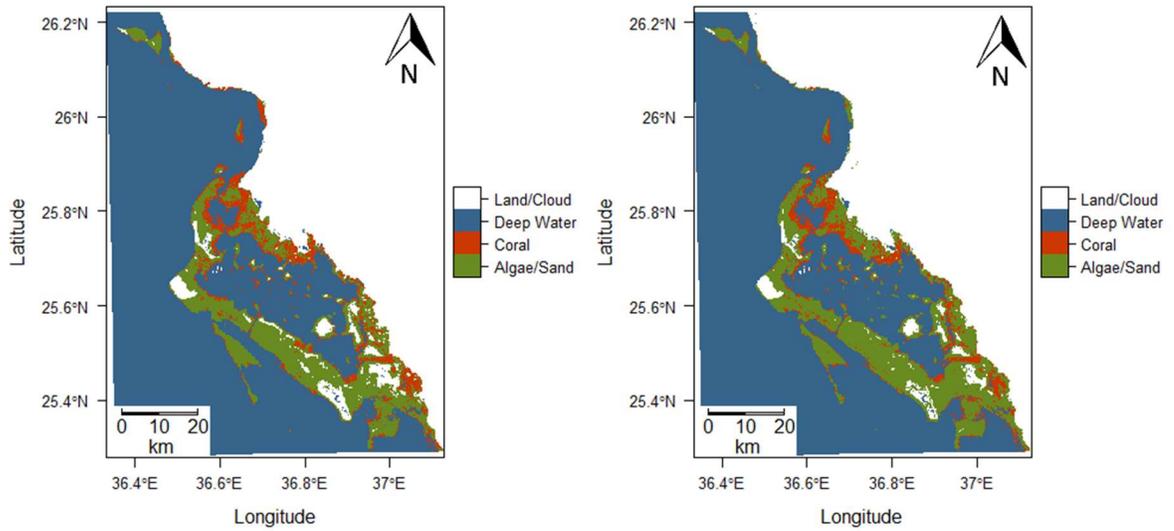


Figure 5-12: Predicted class membership map for the Al Wajh area of interest (left, 2000 and right, 2018).

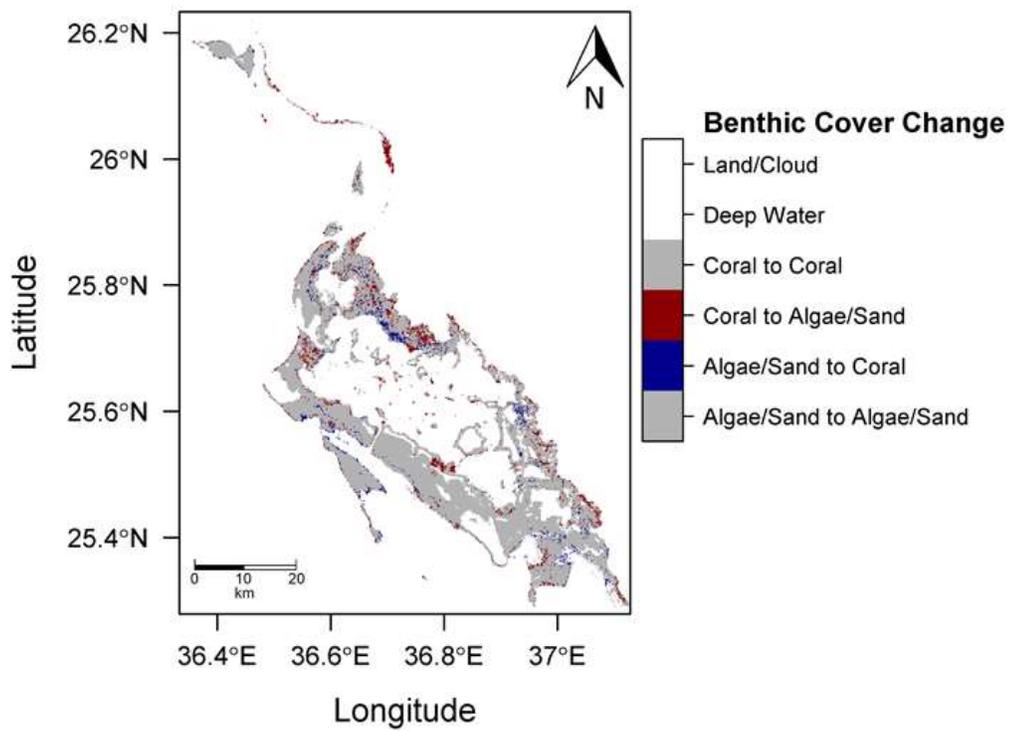


Figure 5-13: Change detection analysis between 2000 and 2018 of the Al Wajh area of interest.

Table 5-4: Change Detection Analysis by Site.

		Initial Class (pixel count)		Initial Class (km <sup>2</sup> )	
		Coral	Not Coral	Coral	Not Coral
Final Class		Gulf of Aqaba		Gulf of Aqaba	
	Coral	86,843	43,382	78.16	39.04
	Not Coral	60,171	91,906	54.15	82.72
		Umluj		Umluj	
	Coral	93,299	16,144	83.97	14.53
	Not Coral	19,985	82,151	17.99	73.94
		Al Wajh		Al Wajh	
	Coral	208,448	46,119	187.60	41.51
	Not Coral	86,053	702,610	77.45	632.35

## 5.4 Discussion

Identifying a methodology for developing a robust classifier that can generalize to new locations as applied to coral reefs in the Red Sea is the key outcome of this research. The methodology proposed here used remote sensing information (DII values) and ground truth observations from the Gulf of Aqaba in order to train a coral benthic cover classifier. The classification algorithm used was SVM with a radial basis kernel. That classifier was then deployed to predict benthic cover types in a new location within the Red Sea (Umluj) in order to evaluate the generalization properties. Validation of this procedure was done by assessing the results of the classifier against ground truth observations from the Umluj location. Finally, a robust classifier was trained using consolidated data from both the Gulf of Aqaba and Umluj. This robust classifier was validated using cross-validation techniques. The trained robust classifier was then applied to a third, unobserved location Al Wajh.

A temporal coral cover change detection analysis was then conducted. Landsat 7 images captured in 2000 were acquired of each of the three locations (Gulf of Aqaba, Umluj, and Al Wajh). These initial state images captured approximately 18 years prior to

the Landsat 8, final state images were used to identify the change in coral cover over time. The per pixel benthic habitat cover type for each of the three locations was determined using the trained SVM classifiers. The previously trained SVM classifiers were applied to the 2000 images for both the Gulf of Aqaba and Umluj locations. A per pixel comparison between the 2000, initial state image and the 2018, final state image was performed to identify the change in extent over the 18-year period. The consolidated, robust classifier was then applied to the 2000 image of the Al Wajh location and a per pixel change detection analysis conducted for that site as well.

The benefits of the methodology presented in this study is the ability for the machine learning classifier to generalize to new locations. This capability was evaluated by deploying the classifier trained on the Gulf of Aqaba location against the ground truth observations of the Umluj location. The ability of a classifier trained using this methodology to generalize to larger areas enables broader analysis of coral reef extent. In this way, the analysis of coral reefs using remote sensing data can scale beyond the scope of previous in-situ analysis. In fact, the natural progression of in-situ analysis is to study larger areas spatially as well as temporally. The methodology proposed in this research enables the expansion of machine learning classifiers to generalize to larger extents with a limited increase in resources. Therefore, using a broader training set which adequately represents the coral reefs present in multiple locations enables the development of a classifier that can generalize beyond in-situ style analysis.

The primary challenge of the proposed methodology is that the classifier trained on a diversity of coral reef locations will not achieve the same accuracy of in-situ models.

This is an expected result since the classifier is trained to represent multiple locations rather than the limited scope, single location of in-situ style analysis. This is because previous in-situ analysis rely upon localized environmental conditions for increased accuracy. In-situ type classifiers account for site-specific geomorphology and local coral fauna in addition to the benthic cover type for classification. Therefore, these algorithms are actually relying on site-specific bias and therefore do not perform as well when these conditions change. However, the in-situ style classifiers do perform exceptionally well within the specific location that they are trained. As a result, the primary challenge to the generalized training method evaluated in this research is an increase in robustness and overall accuracy across all locations evaluated at the expense of accuracy for each individual location if they had been subjected to an in-situ style analysis.

## **5.5 Conclusion**

This study performed temporal change detection analysis across three Red Sea locations over an 18-year period. First, a classifier was trained on a single location, the Gulf of Aqaba. This classifier achieved a raw accuracy as measured by the percentage of correctly classified pixels compared to the ground truth observations, of 78.22%. This trained classifier was then applied to predict coral cover in a second location, Umluj, in order to evaluate the generalization properties of the algorithm. 72.73% of the pixels within this new location were correctly classified compared to ground truth observations of this second location. The ground truth observations from both of these locations was then consolidated and a new, robust classifier developed and applied to a third site, Gulf of Aqaba. A change detection analysis was then performed. Landsat 7 images from 2000 were

acquired for each of the three locations. The machine learning classifiers were then applied to these earlier imaged to determine the per pixel benthic type. Classification maps for each of the locations and for each location were developed and a per pixel change detection analysis for the 18-year period was conducted. The results of this analysis showed a decrease in coral cover at the Gulf of Aqaba location of 117.20km<sup>2</sup> (11.4%), a decrease in coral at the Umluj location of 3.46km<sup>2</sup> (3.4%), and a decrease in coral at the Al Wajh location of 35.94km<sup>2</sup> (13.6%).

## 6. Conclusion

This research proposes a novel methodology for developing a generalized machine learning classifier that can evaluate coral cover in multiple locations. Previous research in the area has focused on in-situ style analysis that are inherently limited in scope. Furthermore, these in-situ analyses tend to suffer from site-specific bias that results in high accuracy for the specific location under study but poor generalized performance. The deployment process proposed in this research expands the scope of previous work to evaluate larger extents of marine habitat. The validity of the proposed methodology was evaluated in three studies.

The first scientific study evaluated the generalized performance across four Remote Pacific Island locations. In this study, an LDA classifier was trained on a single location, Palmyra Atoll. The performance of the model was evaluated against ground truth observations and a classification accuracy of 80.3% was obtained. This trained algorithm was then deployed to three additional sites in order to evaluate the model's capacity to generalize to new data. The generalized accuracy was measured using several metrics including accuracy (percentage of ground truth observations correctly classified), precision, recall, specificity, and F-measure. When applied to Kingman Reef, Baker Island Atoll, and Howland Island the classifier obtained an accuracy of 78.6%, 69.2%, and 71.4%, respectively.

The second scientific study built on the previous study in three major ways: 1) additional machine learning classifiers were evaluated 2) two additional sites were evaluated 3) temporal change detection analysis was conducted for each location. First, the more advanced machine learning classifiers evaluated included logistic regression, regularization methods ( $\ell_1$  and  $\ell_2$  penalty), QDA, and SVM. The SVM classifier with a radial kernel function outperformed the others. The maximal margin hyperplane approach was found to both handle the data size of the study and the goal of producing a generalized classifier well. In-situ style analysis of each of these sites conducted with the SVM classifier produced accuracies of 87.9%, 85.7%, 69.2% and 82.1% for Palmyra Atoll, Kingman Reef, Baker Island Atoll, and Howland Island respectively. In order to evaluate the generalization of the SVM classifier a variant to the traditional cross-validation technique was proposed. Under this alternative, CPCV process, rather than randomly selecting observations for each of the folds as in traditional cross-validation, the folds are selected by a common parameter (ie location). In this way, three islands were used to train a classifier for a fourth island. This process was repeated for each of the islands to develop an evaluation of generalized performance free from site-specific bias. The CPCV procedure yielded an accuracy of 78.8% for Palmyra Atoll, 81.0% for Kingman Reef, 65.4% for Baker Island Atoll, and 67.9% for Howland Island. Following this, a classifier was trained on the consolidated observations from all four locations which yielded an accuracy of 75.3%. This robust classifier trained on the combined information from each of the initial four sites was then deployed to evaluate the coral extent in two additional sites, Tabuaeran Island and Kiritimati Island. A temporal change detection analysis was then conducted for each of the six sites. Images captured 12 to 14-years prior to the final

state image of each location were used to evaluate the initial extent of coral. The machine learning classifiers developed for each location were applied to the initial state image for each of the four initial sites (Palmyra Atoll, Kingman Reef, Baker Island Atoll, and Howland Island) and the robust classifier was applied to the final two sites (Tabuaeran Island and Kiritimati Island). A per pixel change detection analysis was then performed for each of the six locations by comparing the benthic cover in the initial image to that of the final image. The result of this analysis showed a decrease in coral of 31.9% (6.91km<sup>2</sup>) at the Palmyra Atoll location, 25.3% (7.59km<sup>2</sup>) at the Kingman Reef location, 40.5% (0.17km<sup>2</sup>) at the Baker Island Atoll location, 26.1% (0.06km<sup>2</sup>) at the Howland Island location, 35.2% (1.61km<sup>2</sup>) at the Tabuaeran Island location, and 42.7% (11.25km<sup>2</sup>) at the Kiritimati Island location.

The third scientific study included in this dissertation evaluated the performance of the proposed methodology for training a robust machine learning classifier using three Red Sea locations. This study first trained a classifier using ground truth observations and Landsat 8 satellite imagery of the Gulf of Aqaba. The performance of this classifier was evaluated using accuracy (percentage of ground truth observations correctly classified), precision, recall, specificity, F-measure, and Kappa-coefficient. As in the previous study, an SVM classifier using a radial basis function was used. When applied to the Gulf of Aqaba ground truth observations, 78.2% of all observations were correctly classified. The generalized accuracy of this machine learning algorithm was then evaluated by applying the trained model to ground truth observations of a second site, the Umluj location. 72.7% of the ground truth observations of this site were correctly classified using the machine learning algorithm trained using data from the Gulf of Aqaba location. A robust classifier

was then developed using the consolidated ground truth information from both the Gulf of Aqaba and Umluj location. This classifier correctly classified 71.0% of all observations across both locations. Landsat 7 imaged captured in 2000 were then acquired as initial state observations for temporal change detection analysis. The classifiers were deployed to produce a benthic cover type map which was compared to the 2018 benthic cover type map for each location. A per pixel change detection analysis was conducted. The results of this analysis showed a reduction in coral cover of 11.4% (15.11km<sup>2</sup>) within the Gulf of Aqaba location, 3.4% (3.46km<sup>2</sup>) within the Umluj location, and 13.6% (35.94km<sup>2</sup>) within the Al Wajh location.

This research proposes a methodology for developing a robust machine learning classifier which can generalize spatially and temporally for the evaluation of coral cover. Future research is needed to evaluate the extent to which the proposed methodology can generalize. In particular, an analysis of how distance impacts generalization of the classifier is needed. Furthermore, how well a classifier can generalize to reefs in different oceans needs to be explored. In addition, a similar analysis to this exploring other ecosystems such as mangroves and seagrass should be investigated. Finally, on January 6, 2009 the Pacific Remote Islands Marine National Monument was established by the United States of America [141]. This is the largest fully protected marine area in the world and encompasses Palmyra Atoll, Kingman Reef, Baker Island Atoll, and Howland Island. A further analysis comparing these reefs before and after this event can identify the environmental impact of this governmental policy and inform future policy decisions.

In conclusion, this research successfully achieved its goal of finding a rigorous, reliable, scientific, and mathematically sound methodology for developing a machine learning classifier which reduces site-specific bias and applying this process to spatiotemporal analysis of coral reefs using remote sensing imagery.

## 7. References

- [1] E. P. Green, P. J. Mumby, A. J. Edwards and C. D. Clark, "Remote Sensing Handbook for Tropical Coastal Management," A. J. Edwards, Ed., Paris, UNESCO, 2000.
- [2] "The Global Risks Report 2019," World Economic Forum, 2019. [Online]. Available: <https://www.weforum.org/reports/the-global-risks-report-2019>.
- [3] C. Jessen and C. Wild, "Herbivory Effects on Benthic Algal Composition and Growth on a Coral Reef Flat in the Egyptian Red Sea," *Marine Ecology Progress Series*, 2013, p. 476: 9–21. doi: 10.3354/meps10157.
- [4] J. Hedley, C. Roelfsema, I. Chollett, A. Harborne, S. Heron, S. Weeks, W. Skirving, A. Strong, C. Eakin, T. Christensen, V. Ticzon, S. Bejarano and P. Mumby, "Remote sensing of coral reefs for monitoring and management: a review," *Remote Sensing*, 2016, p. 8:118 doi: 10 3390/rs8020118 .
- [5] R. B. Aronson and W. F. Precht, "White-Band Disease and the Changing Face of Caribbean Coral Reefs," *Hydrobiologia*, 2001, pp. 25–38, doi:10.1023/A:1013103928980.
- [6] P. W. Glynn, "Coral Reef Bleaching: Ecological Perspectives," *Coral Reefs*, 1993, pp. 1–17, doi:10.1007/BF00303779.
- [7] J. W. McManus, R. B. Reyes and C. L. Nañola, "Effects of Some Destructive Fishing Methods on Coral Cover and Potential Rates of Recovery," *Environmental Management*, 1997, pp. 69–78, doi: 10.1007/s002679900006.
- [8] E. Pennisi, "Survey confirms coral reefs are in peril," *Science*, 2002, p. 1622b–1623b.
- [9] L. Burke, K. Reytar, M. Spalding and A. Perry, "Reefs at Risk Revisited," Washington, DC, USA, World Resources Institute, 2011, p. 114.

- [10] C. A. Mora, "A clear human footprint in the coral reefs of the Caribbean," in *Proceedings of the Royal Society of London*, 2008.
- [11] O. Hoegh-Guldberg, J. P. Mumby, J. A. Hooten, S. R. Steneck, P. Greenfield and E. Gomez, "Coral reefs under rapid climate change and ocean acidification," *Science*, 2007, p. 1737–1742.
- [12] J. M. Pandolfi, R. H. Bradbury, E. Sala, T. P. Hughes, K. A. Bjorndal, R. G. Cooke, D. McArdle, L. McClenachan, M. J. H. Newman and G. Paredes, "Global trajectories of the long-term decline of coral reef ecosystems," *Science*, 2003, p. 955–958.
- [13] C. Wilkinson, "Status of Coral Reefs of the World," Townsville, QLD, Australia, Australian Institute of Marine Science, 1998, p. 184.
- [14] C. Wilkinson, "Status of Coral Reefs of the World," Townsville, QLD, Australia, Australian Institute of Marine Science, 2000, p. 363.
- [15] C. Wilkinson, "Status of Coral Reefs of the World," Townsville, QLD, Australia, Australian Institute of Marine Science, 2002, p. 378.
- [16] C. Wilkinson, "Status of Coral Reefs of the World," Townsville, QLD, Australia, Australian Institute of Marine Science, 2004, p. 301.
- [17] C. Wilkinson, "Status of Coral Reefs of the World," Townsville, QLD, Australia, Global Coral Reef Monitoring Network and Reef and Rainforest Research Center, 2008, p. 296.
- [18] T. A. Gardner, I. M. Cote, J. A. Gill, A. Grant and A. R. Watkinson, "Long-term region-wide declines in Caribbean corals," *Science*, 2003, pp. 301, 958–960.
- [19] J. F. Bruno and E. R. Selig, "Regional decline of coral cover in the Indo-Pacific: Timing, extent, and subregional comparisons," *PLoS ONE*, 2007, pp. 2, e711.
- [20] J. B. C. Jackson, M. K. Donovan, K. L. Cramer and V. V. Lam, "Status and Trends of Caribbean Coral Reefs: 1970–2012," IUCN Gland, Switzerland, Global Coral Reef Monitoring Network, 2014.
- [21] "Mapping Coral Reefs From Space," PCRf (Planetary Coral Reef Foundation Overview organization), 2002. [Online]. Available: <http://www.pcrf.org/>.
- [22] P. J. Edmunds and J. F. Bruno, "The importance of sampling scale in ecology: Kilometer-wide variation in coral reef communities," *Marine Ecology Progress Series*, 1996, pp. 143, 165–171.

- [23] T. P. Hughes, A. H. Baird, E. A. Dinsdale, N. A. Moltschaniwskij, M. S. Pratchett, J. E. Tanner and B. L. Willis, "Patterns of recruitment and abundance of corals along the Great Barrier Reef," *Nature*, 1999, pp. 397, 59–63.
- [24] M. El-Raey, A. A. Kader, S. Nasr and a. H. E. Gamily, "Remote Sensing and GIS for an Oil Spill Contingency Plan, Ras-Mohammed, Egypt," *International Journal of Remote Sensing*, 1996, p. 17 (11): 2013–2026 doi: 10.1080/01431169608948756 .
- [25] T. Kutser, A. G. Dekker and W. Skirving, "Modeling Spectral Discrimination of Great Barrier Reef Benthic Communities by Remote Sensing Instruments," *Limnology and Oceanography*, 2003, p. 48 (1): 497–510 doi: 10.4319/lo.2003.48.1\_part\_2.0497.
- [26] P. J. Mumby, E. P. Green, A. J. Edwards and C. D. Clark, "The cost-effectiveness of remote sensing for tropical coastal resources assessment and management," *Journal of Environmental Management*, 1999, p. 157–166.
- [27] J. M. Atkinson, P. G. Lucey, G. J. Taylor, J. Porter, S. Dollar and S. Andre, "CRESPO: Coral Reef Ecosystem Spectro-Photometric Observatory, Concept Study Report to the University Earth System Science Program National Aeronautics and Space Administration," Honolulu, HI, USA, University of Hawaii, 2001.
- [28] E. J. Hochberg and M. J. Atkinson, "Capabilities of Remote Sensors to Classify Coral, Algae, and Sand as Pure and Mixed Spectra," *Remote Sensing of Environment*, 2003, pp. 85(2): 174–189 doi: 10.1016/S0034-4257(02)00202-X .
- [29] P. Capolsini, S. Andréfouët, C. Rion and C. Payri, "A comparison of Landsat ETM+, SPOT HRV, IKONOS, ASTER, and airborne MASTER data for coral reef habitat mapping in South Pacific islands," *Canadian Journal of Remote Sensing*, 2007, pp. 29, 187–200 .
- [30] H. T. Kobryn, K. Wouters, L. E. Beckley and a. T. Heege, "Ningaloo reef: Shallow marine habitats mapped using a hyperspectral sensor," *PLoS ONE*, 2013, pp. 8, e70105 .
- [31] S. R. Phinn, E. Hochberg and a. C. M. Roelfsema, "Airborne photography, multispectral and hyperspectral remote sensing on coral reefs in Coral Reef Remote Sensing," J. A. Goodman, S. R. Phinn and S. Purkis, Eds., Berlin, Germany, Springer, 2013, p. 3–25.
- [32] C. M. Roelfsema, M. Lyons, E. M. Kovacs, P. Maxwell, M. I. Saunders, J. Samper-Villarreal and S. R. Phinn, "Multi-temporal mapping of seagrass cover, species and biomass: A semi-automated object based image analysis approach," *Remote Sensing of Environment*, 2014, pp. 150, 172–187 .

- [33] D. Traganos, B. Aggarwal and D. Poursanidis, "Towards Global-Scale Seagrass Mapping and Monitoring Using Sentinel-2 on Google Earth Engine: The Case Study of the Aegean and Ionian Seas," *Remote Sensing*, 2018, p. 1227.
- [34] D. Traganos, D. Poursanidis and B. Aggarwal, "Estimating Satellite-Derived Bathymetry (SDB) with the Google Earth Engine and Sentinel-2," *Remote Sensing*, 2018, p. 859.
- [35] H. Yamano, "Multispectral Applications," in *Coral Reef Remote Sensing: A Guide for Multi-Level Sensing Mapping and Assessment*, J. Goodman, S. Purkis and S. R. Phinn, Eds., Berlin, Germany, Springer, 2013, pp. 51-78.
- [36] H. El-Askary, S. H. A. El-Mawla, J. Li, M. M. El-Hattab and M. El-Raey, "Change detection of coral reef habitat using Landsat-5 TM, Landsat 7 ETM+ and Landsat 8 OLI data in the Red Sea," *Int J Remote Sens*, 2014, pp. 35, 2327–2346 .
- [37] T. Matsunaga, A. Hoyano and a. Y. Mizukami, "Monitoring of Coral Reefs on Ishigaki Island in Japan Using Multitemporal Remote Sensing Data," *Hyperspectral Remote Sensing of the Ocean*, 2001, p. 4154: 212–222 doi: 10 1117/12 411677 .
- [38] W. Ahmad and D. T. Neil, "An Evaluation of Landsat Thematic Mapper (TM) Digital Data for Discriminating Coral Reef Zonation: Heron Reef (GBR)," *International Journal of Remote Sensing*, 1994, p. 15 (13): 2583–2597 doi:10 1080/01431169408954268 .
- [39] D. L. B. Jupp, "Background and extensions to depth of penetration (DOP) mapping in shallow coastal waters," in *Proceedings of the Symposium on Remote Sensing of the Coastal Zone*, Gold Coast, Queensland, 1988.
- [40] J. P. Mumby, W. Skirving, A. E. Strong, J. T. Hardy, E. F. LeDrew, E. J. Hochberg, R. P. Stumpf and L. T. David, "Remote Sensing of Coral Reefs and their Physical Environment," *Marine Pollution Bulletin*, 2004, p. 48 (3–4): 219–228.
- [41] J. P. Mumby, C. D. Clark, E. P. Green and A. J. Edwards, "Benefits of water column correction and contextual editing for mapping coral reefs," *International Journal of Remote Sensing*, 1998, pp. 19(1) 203-210.
- [42] A. P. Cracknell, M. Ibrahim and J. McManus, "Use of satellite and aircraft data for bathymetry studies," in *Proceedings of 13th Annual Conference of the Remote Sensing Society, University of Nottingham*, Nottingham, 1987.
- [43] E. P. Green, P. J. Mumby, A. J. Edwards and a. C. D. Clark, "A review of remote sensing for the assessment and management of tropical costal resources," *Coastal Management*, 1996, pp. 24, 1-40 .

- [44] S. Andréfouët, E. J. Hochberg, C. Chevillon, F. E. Muller-Karger, J. C. Brock and C. Hu, "Multi-scale remote sensing of coral reefs," in *Remote Sensing of Coastal Aquatic Environments*, R. L. Har, X. Miller, C. E. D. Castillo and B. A. Mckee, Eds., Dordrecht, The Netherlands, Springer, 2005, p. 297–315.
- [45] A. J. M. Zainal, D. H. Dalby and I. S. Robinson, "Monitoring of marine ecological changes on the east coast of Bahrain with Landsat TM," *Photogrammetric Engineering and Remote Sensing*, 1993, p. 59:415–421.
- [46] P. N. Bierwirth, T. J. Lee and R. V. Burne, "Shallow sea-floor reflectance and water depth derived by unmixing multispectral imagery," *Photogrammetric Engineering and Remote Sensing*, 1993, p. 59:331–338.
- [47] S. Andréfouët, F. E. Muller-Karger, E. J. Hochberg, C. Hu and K. L. Carder, "Change detection in shallow coral reef environments using Landsat 7/ETM+ data," *Remote Sensing of Environment*, 2001, p. 78, 150–162.
- [48] S. Andréfouët, P. Kramer, D. Torres-Pulliza, K. Joyce, E. Hochberg, R. Garza-Perez, P. Mumby, B. Riegl, H. Yamano, W. White, M. Zubia, J. Brock, S. Phinn, A. Naseer, B. Hatcher and F. Muller-Karger, "Multi-sites evaluation of IKONOS data for classification of tropical coral reef environments," *Remote Sensing of Environment*, 2003, pp. 128-143.
- [49] P. J. Mumby and A. J. Edwards, "Mapping marine environments with IKONOS imagery: enhanced spatial resolution can deliver great thematic accuracy," *Remote Sensing of Environment*, 2002, p. 82:248–257.
- [50] A. Edwards, "Applications of Satellite and Airborne Image Data to Coastal Management," Paris, France, UNESCO, 1999.
- [51] S. Maritorena, "Remote Sensing of the Water Attenuation in Coral Reefs: A Case Study in French Polynesia," *International Journal of Remote Sensing*, 1996, p. 17 (1): 155–166 doi: 10.1080/01431169608948992 .
- [52] R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- [53] D. P. Roy, J. S. Borak, S. Devadiga, R. E. Wolfe, M. Zheng and J. Descloitres, "The MODIS Land product quality assessment approach," vol. 83, *Remote Sensing of Environment*, 2002, p. 62–76.
- [54] S. A. Ackerman, K. I. Strabala, C. C. Frey, R. A. Frey, C. C. Moeller and L. E. Gumley, "Discriminating clear sky from clouds with MODIS," *Journal of Geophysical Research*, 1998, pp. 103, 32141 – 32157.

- [55] P. S. Chavez, "An improved dark-object subtraction technique for atmospheric scattering correction for multispectral data," *Remote Sens. Environ.*, 1988, pp. 24, 459–479.
- [56] P. Manavalan, P. Sathyanath and G. L. Rajegowda, "Digital image analysis techniques to estimate waterspreadfor capacity evaluations of reservoirs," *Photogramm. Eng. Remote Sens.*, 1993, p. 1389–1395.
- [57] A. A. Gitelson and K. Ya Kondratyev, "Optical models of mesotrophic and eutrophic water bodies," *Int. J. Remote Sens.*, 1991, pp. 373-385.
- [58] P. S. Chavez, "Image-Based Atmospheric Corrections—Revisited and Improved," *Photogramm. Eng. Remote Sens.*, 1996, p. 1025–1036.
- [59] S. Tassan, "Modified Lyzenga's method for macroalgae detection in water with non-uniform composition," *International Journal of Remote Sensing*, 1996, pp. 1601-1607.
- [60] H. R. Gordon and A. Morel, "Remote Assessment of Ocean Color for Interpretation of Satellite Visible Imagery," New York, NY, USA, Springer, 1983.
- [61] T. Vanderstraete, R. Goossens and T. K. Ghabour, "Coral Reef Habitat Mapping in The Red Sea (Hurghada, Egypt) Based on Remote Sensing," *EARSel eProceedings*, 2004, p. 191–207.
- [62] R. A. Armstrong, "Remote sensing of submerged vegetation canopies for biomass estimation," *Int. J. Remote Sens.*, 1993, pp. 621-627.
- [63] D. Lyzenga, "Remote Sensing of Bottom Reflectance and Water Attenuation Parameters in Shallow Water Using Aircraft and Landsat Data," *International Journal of Remote Sensing*, 1981, pp. 71-82. Doi: 10.1080/01431168108948342.
- [64] D. Lyzenga, "Passive remote sensing techniques for mapping water depth and bottom features," *Applied Optics*, 1978, pp. 379-383.
- [65] J. M. Paredes and R. E. Spero, "Water depth mapping from passive remote sensing data under a generalized ratio assumption," *Applied Optics*, 1983, pp. 1134-1135.
- [66] R. W. Gould, R. A. Arnone and M. Sydor, "Absorption, Scattering, and Remote Sensing Reflectance Relationships in Coastal Waters: Testing a New Inversion Algorithm," *Journal of Costal Research*, 2001, pp. 328-341.
- [67] T. Vanderstraete, R. Goossens and T. K. Ghabour, "The Use of Multi-Temporal Landsat Images for the Change Detection of the Coastal Zone Near Hurghada,

- Egypt," *International Journal of Remote Sensing*, 2006, pp. 3645–3655, doi: 10.1080/01431160500500342.
- [68] T. J. Malthus and a. P. J. Mumby, "Remote sensing of the coastal zone: An overview and priorities for future research," *International Journal of Remote Sensing*, 2003, p. 2805–2815.
- [69] S. J. Bainbridge and R. E. Reichelt, "An assessment of ground truth methods for coral reef remote sensing data," in *In Proceedings of the 6th International Coral Reef Symposium*, Townsville, Australia, 1988.
- [70] S. R. Phinn, C. M. Roelfsema and R. Stumpf, "Remote sensing: Discerning the promise from the reality," in *Integrating and Applying Science: A Handbook for Effective Coastal Ecosystem Assessment*, B. J. Longstaff, T. J. B. Carruthers, W. C. Dennison, T. R. Lookingbill, J. M. Hawkey, J. E. Thomas, E. C. Wicks and J. Woerner, Eds., Cambridge, MD, USA, IAN Press, 2010.
- [71] J. D. Hedley, C. Roelfsema, B. Koetz and S. Phinn, "Capability of the Sentinel 2 mission for tropical coral reef mapping and coral bleaching detection," *Remote Sensing of Environment*, 2012, p. 145–155.
- [72] A. Laborte, A. Maunahan and H. R., "Spectral Signature Generalization and Expansion Can Improve the Accuracy of Satellite Image Classification," *PLoS One*, 2010, p. e10516.
- [73] "Palmyra Atoll National Wildlife Refuge," Ramsar Site Information Services, 2018. [Online]. Available: [https://www.fws.gov/refuge/Palmyra\\_Atoll/](https://www.fws.gov/refuge/Palmyra_Atoll/).
- [74] L. M. Max, S. L. Hamilton, S. D. Gaines and R. R. Warner, "Benthic processes and overlying fish assemblages drive the composition of benthic detritus on a central Pacific coral reef," *Marine Ecology Progress Series*, 2013, p. 181–195.
- [75] S. A. Sandin, J. E. Smith, E. E. DeMartini and E. A. Dinsdale, "Baselines and degradation of coral reefs in the Northern Line Islands," *PLoS ONE*, 2008, p. 3(2):e1548.
- [76] D. J. McCauley, F. Micheli, H. S. Young and D. P. Tittensor, "Acute effects of removing large fish from a near-pristine coral reef," *Marine Biology*, 2010, p. 2739–2750.
- [77] I. S. Knapp, J. E. Maragos and P. Vroom, "Monitoring supports establishment of Pacific Remote Islands Marine National Monument," in *Proceedings of the 12th international coral reef symposium*, Cairns, Australia, 2012.

- [78] G. J. Williams, J. E. Smith, E. J. Conklin, J. M. Gove, E. Sala and S. A. Sandin, "Benthic communities at two remote Pacific coral reefs: Effects of reef habitat, depth, and wave energy gradients on spatial patterns," *PeerJ*, 2013.
- [79] "Kingman Reef," Washington, DC, USA, Office of Insular Affairs, Department of the Interior, 2009.
- [80] A. M. Friedlander, S. A. Sandin and E. E. DeMartini, Spatial patterns of the structure of reef fish assemblages at a pristine atoll in the central Pacific, *Marine Ecology Progress Series*, 2010, pp. 410: 219-231.
- [81] "Pacific Remote Islands National Wildlife Refuge Complex. Baker Island National Wildlife Refuge: Draft Comprehensive Conservation Plan and Environmental Assessment," Washington, DC, USA, Pacific Remote Islands National Wildlife Refuge Complex, 2007.
- [82] J. Maragos, J. Miller, J. Gove and B. Mundy, "US coral reefs in the Line and Phoenix Islands, central Pacific Ocean: history, geology, oceanography and biology," in *Coral reefs of the USA. Coral reefs of the world 1*, B. M. Riegl and R. E. Dodge, Eds., New York, Springer, 2008, pp. 595-641.
- [83] J. Miller, J. Maragos and R. Brainard, "The state of Coral Reef Ecosystems of the Pacific Remote Island Areas," in *The State of Coral Reef Ecosystems of the United States and Pacific Freely Associated States*, J. E. Waddell and A. M. Clarke, Eds., OAA Technical Memorandum NOS NCCOS, 2008, p. 353–386.
- [84] "Pacific Remote Islands National Wildlife Refuge Complex. Howland Island National Wildlife Refuge: Draft Comprehensive Conservation Plan and Environmental Assessment," Washington, DC, USA, Pacific Remote Islands National Wildlife Refuge Complex, 2007.
- [85] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, 1936, pp. 179-188.
- [86] T. Hastie, R. Tibshirani and J. Friedman, "The Elements of Statistical Learning: Data Mining, Inference and Prediction, 2nd ed," New York, NY, USA, Springer, 2009, pp. 106-111.
- [87] J. D. Hedley, C. Roelfsema, V. Brando, C. Giardino, T. Kutser, S. Phinn, P. J. Mumby, O. Barrilero, J. Laporte and B. Koetz, "Coral reef applications of Sentinel-2: Coverage, characteristics, bathymetry and benthic mapping with comparison to Landsat 8," *Remote Sens. Environ.*, 2018, p. 598–614 .

- [88] Z. Zhou, L. Ma, T. Fu, G. Zhang, M. Yao and M. Li, "Change Detection in Coral Reef Environment Using High-Resolution Images: Comparison of Object-Based and Pixel-Based Paradigms," *ISPRS Int. J. Geo-Inf.*, 2018, p. 441.
- [89] T. Blaschke, "Object based image analysis for remote sensing," *ISPRS J. Photogramm.*, 2010, pp. 2-16.
- [90] J. Leon and C. D. Woodroffe, "Improving the synoptic mapping of coral reef geomorphology using object-based image analysis," *Int. J. Geogr. Inf. Sci.*, 2011, p. 949–969.
- [91] S. R. Phinn, C. M. Roelfsema and P. J. Mumby, "Multi-scale, object-based image analysis for mapping geomorphic and ecological zones on coral reefs," *Int. J. Remote Sens.*, 2012, p. 3768–3797.
- [92] C. Roelfsema, S. Phinn, S. Jupiter, J. Comley and S. Albert, "Mapping coral reefs at reef to reef-system scales, 10s–1000s km<sup>2</sup>, using object-based image analysis," *Int. J. Remote Sens.*, 2013, p. 6367–6388.
- [93] L. Dingle Robertson and D. J. King, "Comparison of pixel- and object-based classification in land cover change mapping," *Int. J. Remote Sens.*, 2011, p. 1505–1529.
- [94] D. C. Duro, S. E. Franklin and M. G. Dubé, "A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery," *Remote Sens. Environ.*, 2012, p. 259–272.
- [95] M. Pax-Lenney, C. Woodcock, S. A. Macomber, S. Gopal and C. Song, "Forest mapping with a generalized classifier and Landsat TM data," *Remote Sensing of Environment*, 2001, pp. 241-250.
- [96] C. E. Woodcock, S. A. Macomber, M. Pax-Lenney and W. B. Cohen, "Monitoring large areas for forest change using Landsat: Generalization across space, time and Landsat sensors," *Remote Sensing of Environment*, 2001, pp. 194-203.
- [97] J. A. Quirein and M. C. Trichel, "Acreage estimation, feature selection, and signature extension dependent upon the maximum likelihood decision rule," in *Symposium on machine classification of remotely sensed data, Purdue University, West Lafayette, Indiana, USA*, 1975.
- [98] I. Olthof, C. Butson and R. Fraser, "Signature extension through space for northern landcover classification: A comparison of radiometric correction methods," *Remote Sensing of Environment*, 2005, pp. 290-302.

- [99] M. E. Bauer, J. E. Cipra, P. E. Anuta and J. B. Etheridge, "Identification and area estimation of agricultural crops by computer classification of Landsat MSS data," *Remote Sensing of Environment*, 1979, pp. 77-92.
- [100] F. G. Hall, D. E. Strebel, J. E. Nickeson and S. J. Goetz, "Radiometric rectification: Toward a common radiometric response among multirate, multisensor images," *Remote Sensing of Environment*, 1991, pp. 11-27.
- [101] M. Pax-Lenney and C. E. Woodcock, "Monitoring agricultural lands in Egypt with multitemporal Landsat TM imagery: How many images are needed?," *Remote Sensing of Environment*, 1997, pp. 522-529.
- [102] R. G. Congalton and K. Green, "Assessing the Accuracy of Remotely Sensed Data: Principles and Practices," Florida, USA, CRC/Lewis Press, 1999, p. 137.
- [103] S. K. Langley, H. M. Cheshire and K. S. Humes, "A comparison of single date and multitemporal satellite image classifications in a semi-arid grassland," *Journal of Arid Environments*, 2001, pp. 401-411.
- [104] T. G. Van Niel and T. R. McVicar, "Determining temporal windows for crop discrimination with remote sensing: A case study in south-eastern Australia," *Computers and Electronics in Agriculture*, 2004, pp. 91-108.
- [105] D. A. Palandro, S. Andréfouët, C. Hu, P. Hallock, F. Muller-Karger, P. Dustan, M. K. Callahan, C. Kranenburg and C. R. Beaver, "Quantification of two decades of shallow-water coral reef habitat decline in the Florida Keys National Marine Sanctuary using Landsat data (1984–2002)," *Remote Sens. Environ.*, 2008, pp. 3388-3399.
- [106] A. Knudby, C. Newman, Y. Shaghude and C. Muhando, "Simple and effective monitoring of historic changes in nearshore environments using the free archive of Landsat imagery," *Int. J. Appl. Earth Obs. Geoinform.*, 2010, p. S116–S122.
- [107] M. B. Lyons, C. M. Roelfsema and S. R. Phinn, "Towards understanding temporal and spatial dynamics of seagrass landscapes using time-series remote sensing," *Estuar. Coast. Shelf Sci.*, 2013, pp. 42-53.
- [108] W. Li, H. El-Askary, M. Qurban, J. Li, K. ManiKandan and T. Piechota, "Using multi-indices approach to quantify mangrove changes over the Western Arabian Gulf along Saudi Arabia coast," *Ecological Indicators*, 2019, pp. 734-745.
- [109] Z. Wang, W. Yao, Q. Tang, L. Liu, P. Xiao, X. Kong, P. Zhang, F. Shi and Y. Wang, "Continuous Change Detection of Forest/Grassland and Cropland in the Loess Plateau of China Using All Available Landsat Data," *Remote Sens.*, 2018, p. 1775.

- [110] M. D. Spalding and B. E. Brown, "Warm-water coral reefs and climate change," *Science*, 2015, pp. 769-771.
- [111] A. C. Baker, P. W. Glynn and B. Riegl, "Climate change and coral reef bleaching: An ecological assessment of long-term impacts, recovery trends and future outlook," *Estuar. Coast. Shelf Sci.*, 2008, pp. 435-471.
- [112] T. P. Hughes, J. T. Kerry, M. Alvarez-Noriega, J. G. Alvarez-Romero, K. D. Anderson and A. H. Baird, "Global warming and recurrent mass bleaching of corals," *Nature*, 2017, pp. 373-377.
- [113] T. P. Hughes, K. D. Anderson, S. R. Connolly, S. F. Heron, J. T. Kerry and J. M. Lough, "Spatial and temporal patterns of mass bleaching of corals in the Anthropocene," *Science*, 2018, pp. 80-83.
- [114] J. M. Pandolfi, S. R. Connolly, D. J. Marshall and A. L. Cohen, "Projecting coral reef futures under global warming and ocean acidification," *Science*, 2011, pp. 418-422.
- [115] "Coral Reef Ecosystem Program; Pacific Islands Fisheries Science Center National Coral Reef Monitoring Program: Benthic Cover Derived from Analysis of Benthic Images Collected during Stratified Random Surveys (StRS) across the Pacific Remote Island Areas s," NOAA's National Center for Environmental Information, 2016. [Online]. Available: <https://inport.nmfs.noaa.gov/inport/item/36157>.
- [116] "Pacific Island Benthic Habitat Mapping Center: Palmyra Atoll," NOAA's National Center for Environmental Information, 2018. [Online]. Available: <http://www.soest.hawaii.edu/pibhmc/cms/>.
- [117] "Landsat Project Science Office, Landsat 7 (L7) Data Users Handbook," June 2018. [Online]. Available: <https://www.usgs.gov/media/files/landsat-7-data-users-handbook>.
- [118] "Landsat Project Science Office, Landsat 8 (L8) Data Users Handbook," October 2018. [Online]. Available: <https://www.usgs.gov/media/files/landsat-8-data-users-handbook>.
- [119] J. Magel, J. Burns, R. Gates and J. Baum, "Effects of bleaching-associated mass coral mortality on reef structural complexity across a gradient of local disturbance," *Scientific Reports*, 2019.
- [120] "NOAA Rapid Ecological Assessment (REA) Survey Methodology: #3 Benthic Habitat Surveys Coral Reef Ecosystem Program (CREP)," [Online]. Available: [https://www.pifsc.noaa.gov/cred/survey\\_methods/fish\\_surveys/spc\\_benthic\\_method\\_training\\_2016\\_final\\_draft.pdf](https://www.pifsc.noaa.gov/cred/survey_methods/fish_surveys/spc_benthic_method_training_2016_final_draft.pdf).

- [121] A. Heenan, "Long-term monitoring of coral reef fish assemblages in the Western central pacific," *Sci. Data*, 2017.
- [122] O. Beijbom, P. J. Edmunds, D. I. Kline, G. B. Mitchell and D. Kriegman, "Automated Annotation of Coral Reef Survey Images," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Providence, Rhode Island, USA, 2012.
- [123] O. Beijbom, P. J. Edmunds, C. Roelfsema, J. Smith, D. I. Kline, B. Neal, M. J. Dunlap, V. Moriarty, T.-Y. Fan, C.-J. Tan, S. Chan, T. Treibitz, A. Gamst, B. G. Mitchell and D. Kriegman, "Towards automated annotation of benthic survey images: variability of human experts and operational modes of automation," *PLOS One*, 2015.
- [124] K. McCoy, I. Williams and A. Heenan, "A Comparison of Rapid Visual Assessments and Photo-Quadrat Analyses to Monitor Coral Reef Habitats NOAA Pacific Islands Fisheries Science Center," PIFSC Data Report, 2015.
- [125] "Tabuaeran," Office of Te Beretitenti - Republic of Kiribati Island Report Series, 2012.
- [126] B. Gallagher, K. Shimada, F. Gonzalez and E. Stroup, "Tides and Currents in Fanning Atoll Lagoon," *Pacific Science*, 1971, pp. 201-205.
- [127] J. Maragos, "Reef Corals of Fanning Island," *Pacific Science*, 1974, pp. 247-255.
- [128] "Kiritimati," Office of Te Beretitenti - Republic of Kiribati Island Report Series, 2012.
- [129] E. Lovell, T. Kirata and T. Tekinaiti, "Status report for Kiribati's coral reefs," Centre IRD de Nouméa, 2002.
- [130] "Report on the Kiribati 2010 Census of Population and Housing Vol 1 Basic Information and Tables," Bairiki, Tarawa, National Statistics Office Ministry of Finance and Economic Planning, 2010.
- [131] A. Anderson, P. Wallin, H. Martinsson-Wallin, B. Fankhauser and G. Hope, "Towards a First Prehistory of Kiritimati (Christmas) Island, Republic of Kiribati," *The Journal of the Polynesian Society*, 2000, pp. 273-294.
- [132] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel and F. Leisch, "e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.7-1," 2019. [Online]. Available:

<https://CRAN.R-project.org/package=e1071>  
<https://github.com/cran/e1071/blob/master/R/svm.R>.

- [133] J. G. Gapper, H. El-Askary, E. Linstead and T. Piechota, "Evaluation of Spatial Generalization Characteristics of a Robust Classifier as Applied to Coral Reef Habitats in Remote Islands of the Pacific Ocean," *Remote Sens.*, 2018, p. 1774.
- [134] G. James, D. Witten, T. Hastie and R. Tibshirani, "An Introduction to Statistical Learning with Applications in R, 1st ed," New York, NY, USA, Springer, 2013, pp. 337-366.
- [135] F. Provost and T. Fawcett, "Data Science for Business What You Need to Know About Data Mining and Data-Analytic Thinking, 1st ed," Sebastopol, CA, USA, O'Reilly, 2013.
- [136] M. L. Reaka-Kudla, "Biodiversity of Caribbean Coral Reefs In: Caribbean Marine Biodiversity: The Known and the Unknown," P. Miloslavich and E. Klein, Eds., DEStech Publications, 2005, p. 259–276.
- [137] T. P. Hughes, A. H. Baird, D. R. Bellwood, M. Card, S. R. Connolly and C. Folke, "Climate change, human impacts, and the resilience of coral reefs," *Science*, 2003, p. 929–933.
- [138] S. F. Heron, J. A. Maynard and C. Ruben van Hooidonk, "Warming trends and bleaching stress of the world's coral reefs 1985–2012," *Scientific Reports*, 2016.
- [139] O. Hoegh-Guldberg, J. C. Ortiz and S. Dove, "The future of coral reefs," *Science*, 2011, p. 1494.
- [140] F. J. Edwards, "Climate and Oceanography," in *Key Environments: Red Sea*, A. J. Edwards and S. M. Head, Eds., Oxford, Pergamon Press, p. 45–69.
- [141] "Presidential Proclamation 8336: Establishment of the Pacific Remote Islands Marine National Monument," *Weekly Comp, Pres, Docs.*, 12 January 2009. [Online]. Available: [www.presidentialdocuments.gov](http://www.presidentialdocuments.gov).